Corynebacterium Glutamicum: A Platform for Studying Actinobacterial Protein-*O*-Mannosylation and High-Yield Heterologous Protein Production

By

Hirak Saxena

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Microbiology and Biotechnology

Department of Biological Sciences

University of Alberta

ABSTRACT

Corynebacterium glutamicum ATCC 13032 is a generally regarded as safe (GRAS) soil actinobacterium that is widely utilized in industry and was originally used for the large-scale production of amino acids. Through both genetic and metabolic engineering, the organism has been developed into a biorefinery, producing value-added products such as polyamines, food-grade chemicals, medicines, branched-chain amino acids, and many more. In addition to its industrial applications, *C. glutamicum* has even been engineered to function as a "biocontainer" for the bioremediation of arsenic from contaminated soils and waters. With a well-developed molecular toolbox available for the organism, *C. glutamicum* has only recently begun to be investigated for its potential application as a recombinant expression host.

As *C. glutamicum* is capable of protein-*O*-mannosylation (POM), the organism serves as an ideal host system for the investigation of this process in actinobacteria. POM is a form of *O*-glycosylation that is ubiquitous throughout all domains of life and has been extensively characterized in eukaryotic systems. However, in prokaryotes, this process has only been investigated in the context of pathogenicity (in *Mycobacterium tuberculosis*) even though there are many non-pathogenic bacteria that are known to regularly carry out POM. To date, there is no consensus on what benefit POM imparts to the non-pathogenic bacteria that can perform it. Through the expression and native-like mannosylation of known actinobacterial mannoproteins produced recombinantly in *C. glutamicum*, this work shows that this bacterium can be utilized as a host system for the study of actinobacterial POM. The complementation of a POM deficient mutant of *C. glutamicum* (via knockout of the native GT-39, the enzyme responsible for the

initiation of POM) by other actinobacterial GT-39s provides evidence that these closely related enzymes may have different activities and substrate specificities for targets of POM. Moreover, evidence is presented suggesting POM does not only occur in a general secretory pathway (SEC)-dependent manner; it also occurs with twin-arginine translocase (TAT) and non-SEC secreted substrates in a specific and tightly regulated manner. These results highlight the need for further biochemical characterization of POM in these and other bacterial species to help elucidate the true nature of its biological functions.

There are several inherent benefits to using *C. glutamicum* as a protein production platform: low levels of cytoplasmic and extracellular proteases, secretion via well-characterized pathways (SEC and TAT), and its status as an endo-toxin free strain. These factors make *C. glutamicum* an attractive host system for the production of Carbohydrate Active EnZymes (CAZymes) that can be used to modify the glycan profiles of therapeutic proteins. The accurate glycosylation of protein therapeutics is currently one of the various stratagems to increase both their efficacy and serum half-life. It has been repeatedly demonstrated that the capping of *N*-linked glycans by terminal sialic (*N*-acetylneuraminic) acids significantly increases the *in-vivo* half-life of a given glycoprotein. Terminal glycosylation has proven to be difficult to reproduce in mammalian cell culture systems due to large variations in the efficiency of native sialyltransferases (STs) in various production strains, in addition to non-human cell lines incorporating a non-human analogue of sialic acid *N*-glycolylneuraminic acid (Neu5Gc) which is highly antigenic to humans; the incorporation of these antigenic epitopes into recombinant therapeutics must be avoided. To circumvent these complications, STs produced in prokaryotic expression systems have been employed to perform the final glycosylation step of *N*-linked glycans *in vitro*. Unlike enzymes produced in *E. coli*, STs produced in *C. glutamicum* would be

essentially endotoxin free – facilitating the final polishing of therapeutic proteins and decreasing associated costs. This work details the development of *C. glutamicum* as a recombinant host for the production of STs, showing that with minimal engineering and the co-expression of folding chaperones, the strain is capable of producing both prokaryotic and eukaryotic STs with comparable activities and yields to the traditional recombinant host *E. coli*. Overall, this work demonstrates that *C. glutamicum* shows great potential as a recombinant expression platform for many applications.

PREFACE

Some of the research conducted for this thesis forms part of an industrial research collaboration with the PlantForm Corporation, which originated at the University of Guelph, ON and is now located in Toronto, ON. The literature review in Chapter 1 and the concluding analysis in Chapter 5 are my original work. Chapter 2 of this thesis has been published as H. Saxena, et al., "Towards an Experimental System for The Examination of Protein Mannosylation in Actinobacteria", Glycobiology, vol. 33, issue 6, 512 – 524. I was primarily responsible for the data collection, analysis, and manuscript composition. N. Buenbrazo, W.Y. Song, C. Li, D. Brochu, A. Robotham, W. Ding, L. Tessier, and R. Chen assisted with the data collection, W.Y Song and J. Kelly contributed to manuscript edits, and W. Wakarchuk was the supervisory author and involved in concept formation and manuscript composition. Chapter 3 of this thesis has been submitted to the journal Applied and Environmental Microbiology. I was responsible for the experimental design, data collection, analysis, and manuscript composition. R. Patel and J. Kelly assisted in data collection, J. Kelly and W. Wakarchuk also contributed to manuscript edits. W. Wakarchuk was the supervisory author and also involved with concept formation. Chapter 4 of this thesis has been submitted to the journal Microbial Cell Factories. I was responsible for the experimental design, analysis, and manuscript composition. I was primarily responsible for data collection, with contributions from D. Siapuatco, N. Thompson, L. Leclaire, and M. Kirby. N. Thompson also contributed to manuscript composition and edits. W. Wakarchuk as the supervisory author contributed to manuscript edits and was involved in concept formation.

ACKNOWLEDGEMENTS

First and foremost, I want to sincerely thank my supervisor, mentor, and colleague Dr. Warren Wakarchuk. If not for your willingness to take a chance on an inexperienced prospective graduate student all those years ago, I would not have had the opportunity to pursue my curiosity and develop into the researcher and person that I am now. The years of mentorship, advice, and direction that you've given me have, and will continue to, guide me on my own mentorship journey. You started me down a path that has provided me with the sturdiest of foundations for whatever is to come next. Though sometimes painfully difficult, the journey has been one to look back fondly on.

To my committee members Dr. Lisa Stein and Dr. Chris Cairo, your guidance and opinions were invaluable and the time you offered me despite your countless responsibilities is sincerely appreciated.

To all my Wak Lab mates over the years, who are far too numerous to list; I appreciate every single one of you and will never forget the comradery we shared in our time together. The lab environment you contributed to – through collaboration and sharing your expertise and friendship – made the bad times bearable and the good times unforgettable. I wish all of you the best in whatever you choose to pursue.

To my friends and family, I want to extend a heartfelt thank you for their constant support and encouragement. If not for your constant beleaguering of me with "How much longer do you have left?" or "When will you be done?", I would have completed my research and dissertation much earlier.

Grim, the world just isn't the same without you in it. We miss you everyday.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2.

Chapter 3.

Chapter 4.

LIST OF FIGURES

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 5.

1. BACKGROUND

1. 1. Protein glycosylation

Protein glycosylation is the most abundant post-translational protein modification found in nature with significant inherent complexity due to the multistep nature of its biosynthesis, structural heterogeneity, and the diversity of sugar monomers and glycosidic linkages utilized (Figure 1.1). Conserved in archaea, bacteria, and eukaryotes, glycosylation has been broadly defined as the attachment of a carbohydrate moiety – or glycan – to proteins, lipids, or other organic molecules (1). Glycopeptide linkages can be categorized into specific groups based on the oligosaccharide attached and the nature of the bond formed, including phosphoglycosylation, glycation, *N*-, *O*-, *C*-, and *S*-linked glycosylation (2–7). It should be mentioned that individual proteins are not restricted to a singular form of glycosylation. In-fact, many proteins are often glycosylated at multiple sites using a variety of different glycosidic linkages (8, 9). The roles of glycosylation are many and far reaching (Figure 1.2), playing significant roles in protein folding and stability, cell-cell adhesion and signaling, metabolism, transport, and many viral and bacterial infection and pathogenesis mechanisms (1, 10–14).

N-linked glycosylation is a highly prevalent – and the most studied – form of glycosylation that is characterized by the linkage of sugar residues (both as mono- and polysaccharides) to proteins through the nitrogen atom of an Asn residue. This process occurs extensively in eukaryotes and archaea and has also been found in several eubacteria (11). *O*-linked glycosylation is the attachment of sugar molecules to a protein through the oxygen atom of hydroxyl groups belonging to Ser and/or Thr, and sometimes Tyr, residues. *O*-linked glycosylation is highly conserved, being found in both eukaryotes and bacteria (15, 16). To briefly summarize the remaining types of glycosylation; *C*-linked glycosylation is characterized

by the covalent attachment of a mannose residue to the indole ring of a Trp (17), *S*-linked glycosylation is characterized by the attachment of oligosaccharides to the sulfur atom of Cys residues (18, 19), and glycation refers to the non-enzymatic covalent attachment of reducing sugars to proteins and/or lipids (20). Glycation is also colloquially known as the Maillard reaction (21).

Of the various distinct forms of glycosylation, *N*-linked and *O*-linked glycosylation have been most extensively studied (Figure 1.3). However, *N*-linked glycosylation has historically garnered significantly more interest and research than its *O*-linked counterpart. While *O*-glycosylation has been investigated in various eukaryotes, there is still a sizeable gap in the understanding of the various forms of *O*-glycosylation on the bacterial side of life (22).

1. 2. *O*-glycosylation

Unlike eukaryotic *N*-linked glycosylation (where glycosylation is initiated by the addition of a pre-formed oligosaccharide core structure followed by the sequential addition or removal of certain individual sugar residues, each catalyzed by different enzymes), *O*-linked protein glycosylation in eukaryotes is initiated by the attachment of a single monosaccharide residue to the hydroxyl group of Ser or Thr (and occasionally Tyr) residues followed by the sequential addition of many diverse monosaccharide residues which are also regulated by different glycosyltransferases (1, 23, 24).

O-linked glycosylation in eukaryotes can be categorized into mucin and non-mucin types. Mucin types have *N*-acetylgalactosamine (GalNAc) as the monosaccharide attached directly to the protein, whereas non-mucin types can have *N*-acetylglucosamine (GlcNAc), fucose (Fuc), mannose (Man), glucose (Glc), xylose (Xyl), or galactose (Gal) as the monosaccharide attached

directly to the protein (25–27). *O*-linked glycoproteins can be large (>200 kDa) with the glycans themselves composing most of the glycoprotein's mass. *O*-glycans are commonly biantennary and exhibit comparatively less branching than the glycans on *N*-linked glycoproteins. *O*-linked glycans tend to be very heterogeneous, showing great structural and compositional diversity (Figure 1.4), and as such they are generally classified by their core structure (15, 28). It is now estimated that *O*-linked glycoproteins are far more numerous than originally thought (29).

In addition to the cellular functions glycosylation is normally involved in, *O*-linked glycosylation is also involved in hematopoiesis, inflammation response mechanisms, differentiation of the ABO blood group antigens (Figure 1.5), and others (16, 30–34). Since *O*-glycosylation has been found to play significant roles in the conformation, solubility, stability, and hydrophobicity of *O*-glycosylated proteins, understanding the *in vivo* enzymatic synthesis and structure of *O*-linked glycans is critical. It should also be mentioned that the study of *O*-glycosylation is further complicated by the fact that this glycomodification lacks a consensus sequence or a universal enzyme to release *O*-glycans from their parent proteins (1, 35–38).

1. 3. Bacterial *O*-glycans

Most of the preliminary findings of bacterial *O*-glycans came from work focused on the *Neisseria* species with protein *O*-linked glycosylation being first characterized in *Neisseria meningitidis* (39). The pilin protein of this organism was shown to be modified by a trisaccharide and a similar glycan was later identified in *N. gonorrhoeae*, subsequently leading to the identification of the *O*-oligosaccharyltransferase (OST) PglL responsible for the modification (40, 41). Following the elucidation of this glycan structure – Gal(β 1-4)-Gal(α 1-3)-2,4-diacetimido-2,4,6-trideoxyhexose – many more similarly modified proteins were rapidly

reported (42–44). This family of bacterial OSTs was found to be widespread amongst pathogenic bacteria such as *Pseudomonas aeruginosa* (named PilO) where the periplasmic transfer of *O*-glycan structures (*en-bloc* from a lipid linked carrier) suggested the OSTs were membrane bound or associated (45, 46).

The numerous *O*-mannosyl glycans found on the glycoproteins of members of the *Actinomycetes* were seen to be analogous to those found on yeast and fungi (Figure 1.6) – a finding first seen with the cell surface *O*-glycans of *Mycobacterium tuberculosis* – and unlike the eukaryotic *en bloc* transfer of these *O*-glycans, bacterial *O*-mannosyl glycans were found to be transferred in a processive manner (22, 47, 48). The comparatively shorter bacterial *O*-mannosyl glycans were composed of 2 – 4 mannose residues and found on a wide variety of proteins with roles played in several cellular functions, but the most significant finding was the impact protein-*O*-mannosylation (POM) played on the virulence (Figure 1.7) of *M. tuberculosis* (49). For this reason, there was much early work focused on *Mycobacterial O*-mannosyl glycans as they were an attractive avenue for vaccine development (50).

1. 4. POM in actinobacteria

POM is a highly conserved subset of *O*-glycosylation, being the only type of *O*-glycosylation found throughout fungi, bacteria, and eukaryotes (except plants). Much of the information on POM has been garnered from work done primarily in *S. cerevisiae*, facilitating the understanding (Figure 1.8) of how it can impact disease states in eukaryotes. A considerable proportion of the mannosylated proteins identified to date are either secreted or cell membrane associated proteins (14, 22, 51, 52). Overall, this modification is carried out similarly in

eukaryotes and prokaryotes (Figure 1.9) – the most significant differences will be discussed in Chapter 3.

The presence of only a single protein *O*-mannosyltransferase (PMT) responsible for the synthesis of *O*-mannosyl glycans has been confirmed both genetically and biochemically in other actinobacteria like *Mycobacterium smegmatis*, *Corynebacterium glutamicum, Streptomyces coelicolor, Cellulomonas fimi,* and *Cellulomonas flavigena* (Figure 1.10) with only a handful of common mannoproteins identified between them (47, 53–56). However, each species contains a much larger population of mannoproteins either with no orthologue in the other closely related actinobacteria or with a different mannosylation profile. In addition, the complete knockout of the glycosyltransferase family 39 (GT-39) PMT in some of these organisms results in no discernable phenotype other than the decrease in virulence of pathogenic species (49).

POM in actinobacteria is severely lacking detailed information both of the process itself (Figure 1.11) and of the function of the modification for the cell and the individual glycoprotein. While POM has been shown to have a significant effect on the virulence of some members, it is clearly not the only function of the modification as the much larger portion of non-pathogenic members also perform it. The fundamental question remains unanswered; what benefit(s) does POM impart to these organisms?

1. 5. Glycosylation and therapeutic proteins

Presently, there is great interest in the improvement of the therapeutic efficacy of protein drugs/therapeutics or biologics through the engineering of their physiochemical and pharmacological properties (57). As a significant proportion of these biologics are themselves glycoproteins, the directed and specific manipulation of a protein glycosylation patterns through

glycoengineering is an attractive avenue for development (58–60). It has been repeatedly demonstrated that proper glycosylation profiles can not only lead to improved *in vivo* half-lives (61–64) but can also enhance the therapeutic efficacies of many protein-based drugs (65–68).

All protein drugs are prone to one or more clearance mechanisms, negatively impacting their plasma persistence (69). Size-based kidney filtration and glycan-based receptor clearance are two "universal" mechanisms affecting the half-lives of most circulating proteins – though notably, not immunoglobulins of the IgG class (70). While glycosylation is capable of influencing clearance by both of these mechanisms, the presence of a single terminal sugar can influence the lectin receptor-mediated cellular uptake of glycoproteins as in the case of the hepatic lectin known as the asialoglycoprotein receptor (Figure 1.12), or Ashwell-Morell receptor (71). As sialic acids are the most prevalent terminal monosaccharide of most glycoproteins there have been significant efforts made to improve the sialylation levels of biologics made in cell culture (72–77).

Because of the drawbacks inherent to mammalian cell culture in the context of recombinant protein production and glycoengineering – which will be discussed later – many non-mammalian expression systems and *in vitro* methodologies have been developed to improve the sialylation profiles of biologics and even produce therapeutic glycoproteins in prokaryotes (78–83). The low cost and rapidly scalable nature of enzyme-based *in vitro* sialylation allow the strategy to be utilized in concert or as a replacement to *in vivo* sialyation, which requires significant efforts towards development and employment (84, 85). The production of the requisite sialyltransferases (ST) in a prokaryotic host would further improve the production pipeline of the *in vitro* sialyation methodology.

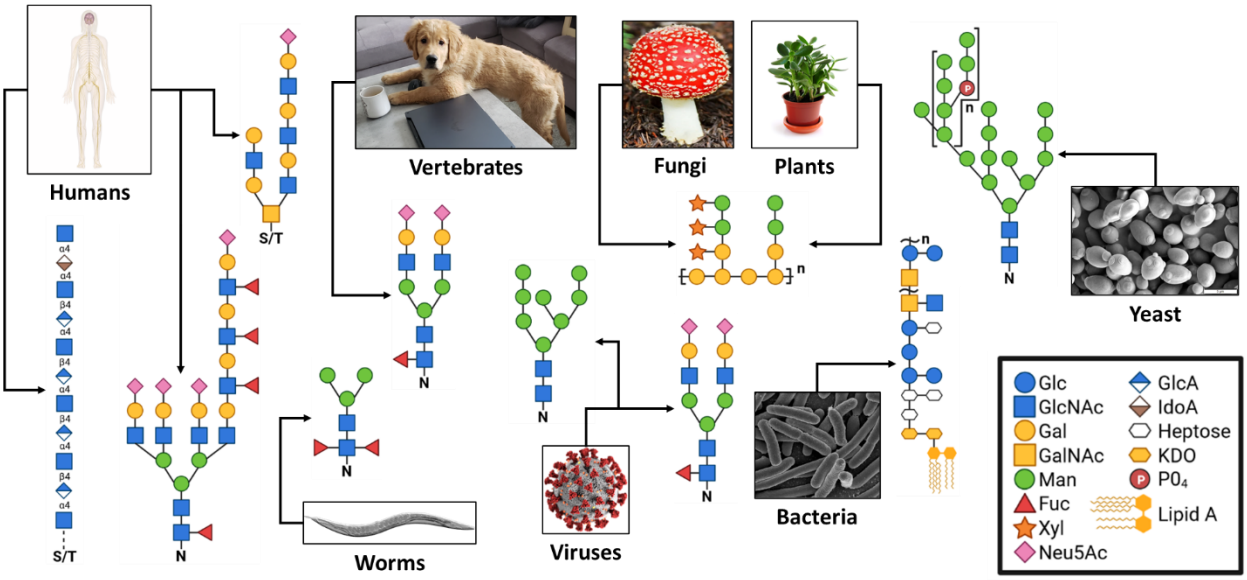


Figure 1. 1. Glycans are structurally diverse and representative of biodiversity.

The structural diversity of glycan structures synthesized across all domains of life highlights the complex nature of this post-translational modification. Structure can differ based on the monosaccharide residue, the configuration and sequence/position of the glycosidic linkages, and both the nature and location of the non-glycan moieties that oligosaccharides are attached to. While some glycan structures can be unique to the domains that they are present in (such as some glycan structures in plants and fungi), there can also exist significant overlap in glycan structures between the domains (such as some shared structures between bacteria, viruses, and eukaryotes). Monosaccharide abbreviations: *D*-Glucose (Glc), *N*-acetyl-*D*-glucosamine (GlcNAc), *D*-galactose (Gal), *N*-acetyl-*D*-galactosamine (GalNAc), *D*-mannose (Man), *L*-fucose (Fuc), *D*-xylose (Xyl), *N*-acetylneuraminic acid (Neu5Ac), *D*-glucuronic acid (GlcA), *L*-iduronic acid (IdoA), ketodeoxyoctonic acid (KDO).

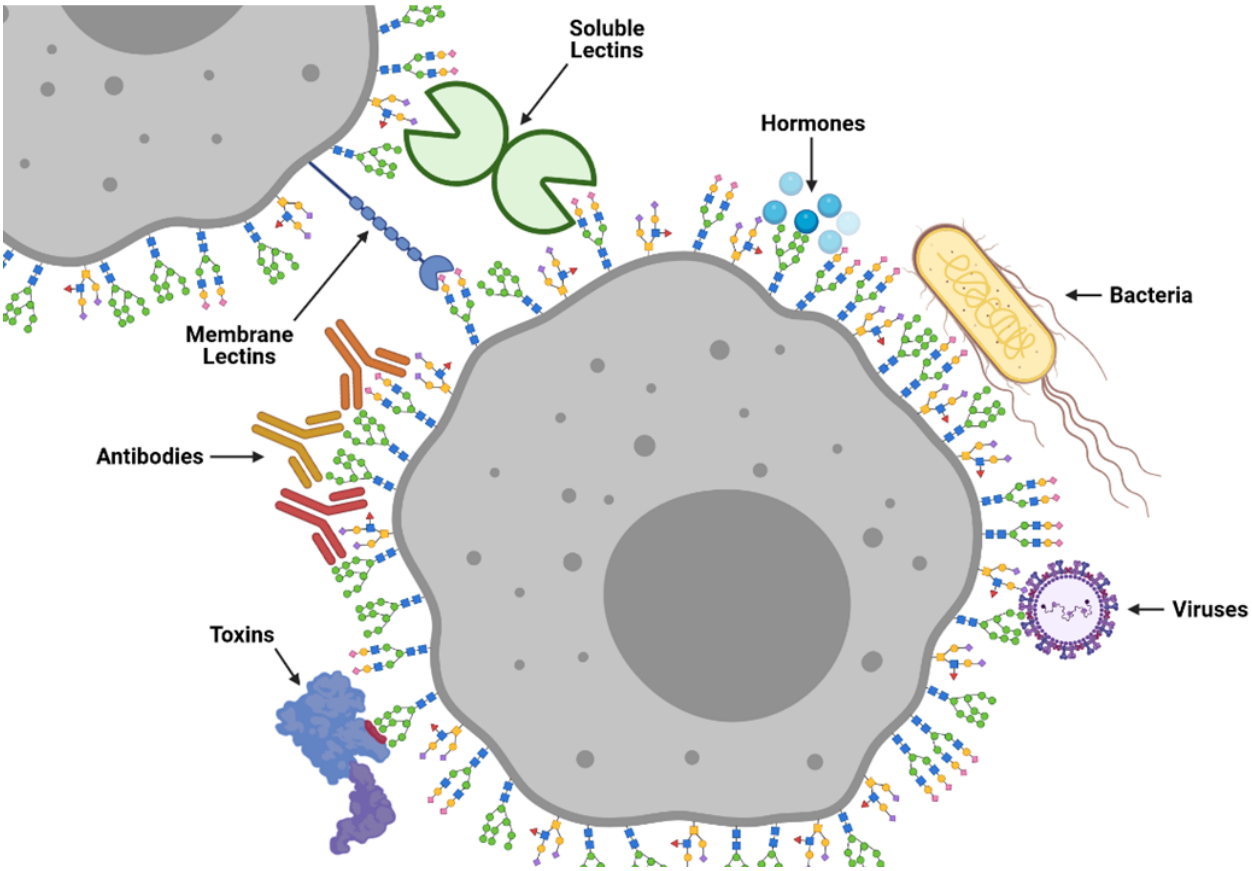


Figure 1. 2. Examples of the biological functions influenced by cell-surface glycosylation.

Glycans at the cell surface can act as targets for toxins, antibodies, lectins, hormones, bacteria, and viruses. Glycans (both intra- and extracellular) can also regulate cell-adhesion, recognition, metastasis, and development in addition to numerous receptor-ligand interactions.

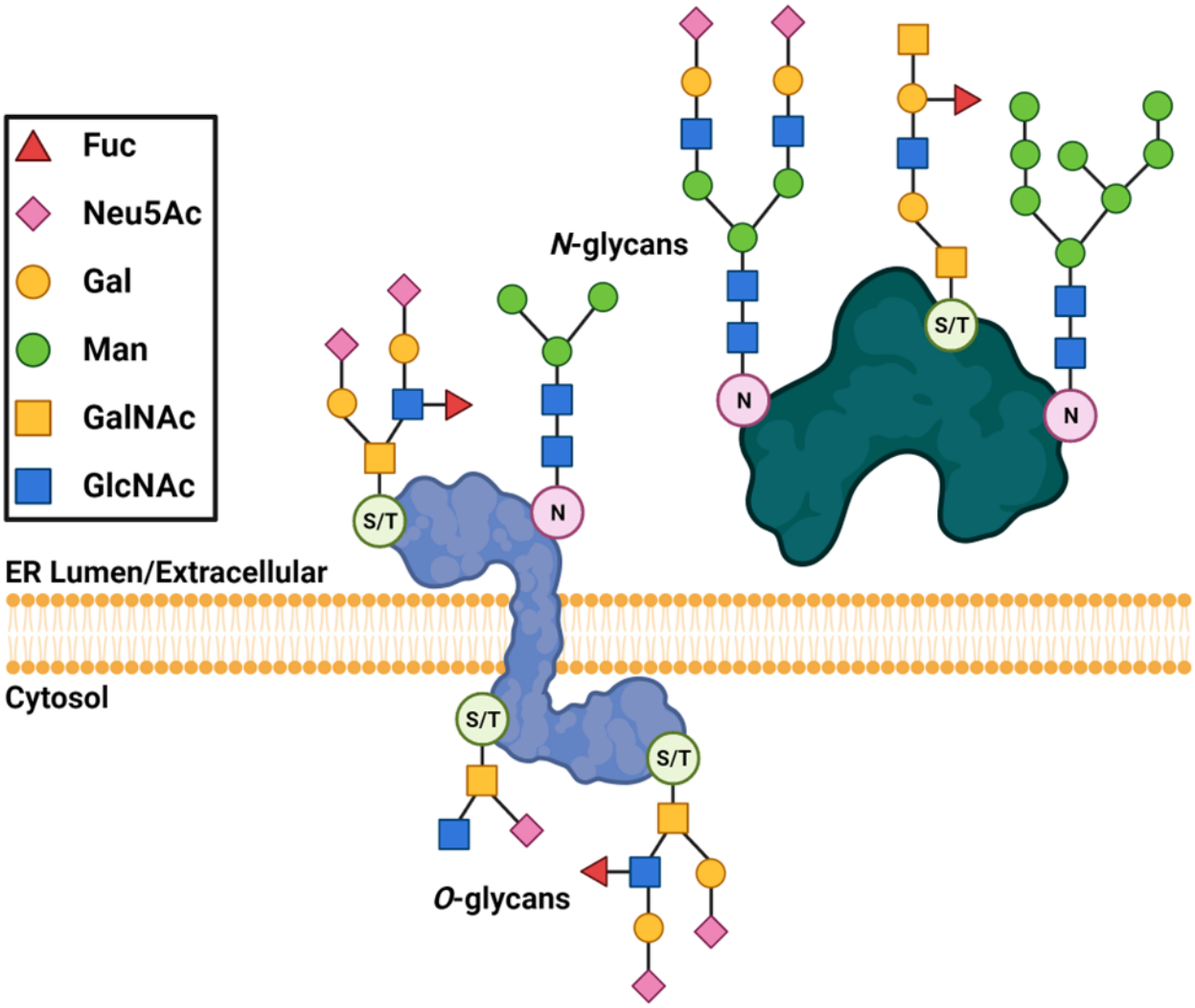


Figure 1. 3. Structural comparison of eukaryotic *N*- and *O*-linked glycans.

The most common types of protein glycosylation are *N*- and *O*-linked, characterized by their linkage to Asp (via a nitrogen atom) and Ser/Thr (via an oxygen atom) residues, respectively. *N*-linked glycans are pre-formed as oligomers prior to their transfer onto a nascent glycoprotein, followed by subsequent trimming and elongation. *O*-linked glycans are formed in a stepwise manner with the first sugar residue added directly to the nascent glycoprotein.

Adapted from: (1)

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (2017) Essentials of Glycobiology. *Cold Spring Harbor (NY)*.

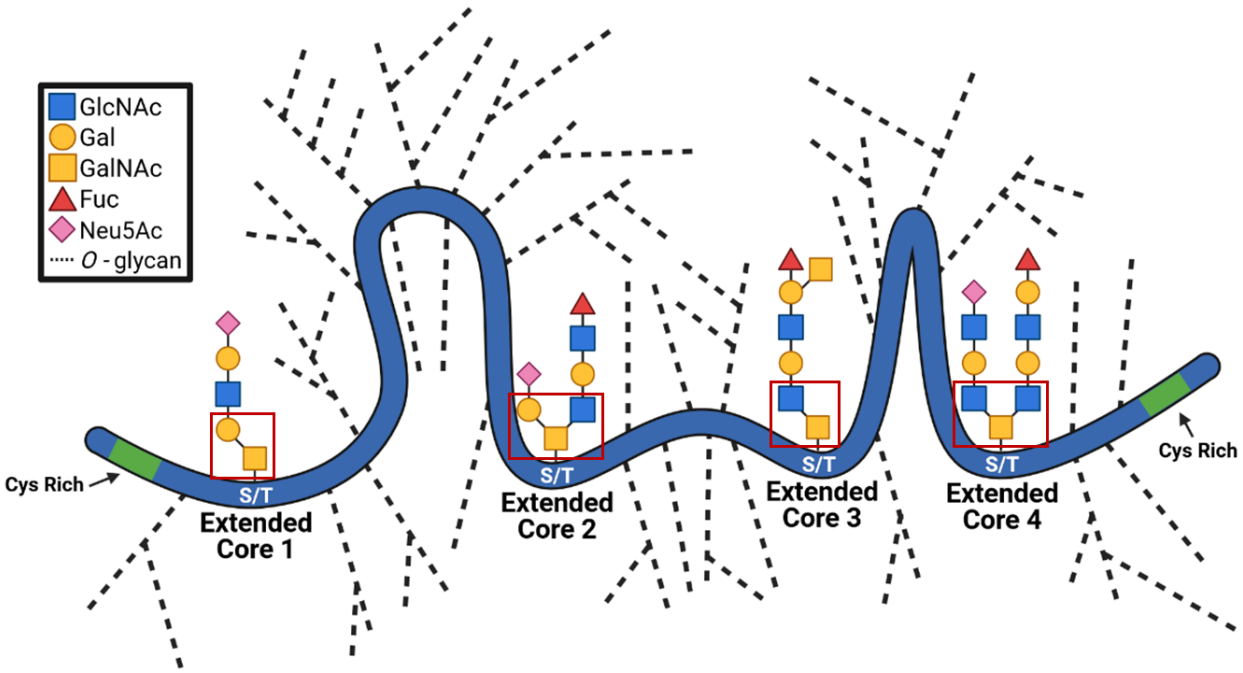


Figure 1. 4. Structural and compositional diversity in mucin-type *O*-glycans.

Mucins are highly *O*-glycosylated proteins, with the *O*-glycans contributing up to 80% of the weight of the mucin. Due to the heterogenous, diverse, and extensive nature of *O*-glycan structures, they are classified based on their core structures (the most predominant cores, 1 – 4 are shown in red boxes) which can be further extended with *N*-acetyl-*D*-glucosamine (GlcNAc), *D*-galactose (Gal), *N*-acetyl-*D*-galactosamine (GalNAc), *L*-fucose (Fuc), and *N*-acetylneuraminc acid (Neu5Ac).

Adapted from: (86)

González-Morelo, K. J., Vega-Sagardía, M., and Garrido, D. (2020) Molecular Insights Into O-Linked Glycan Utilization by Gut Microbes. *Front Microbiol*. 10.3389/fmicb.2020.591568.

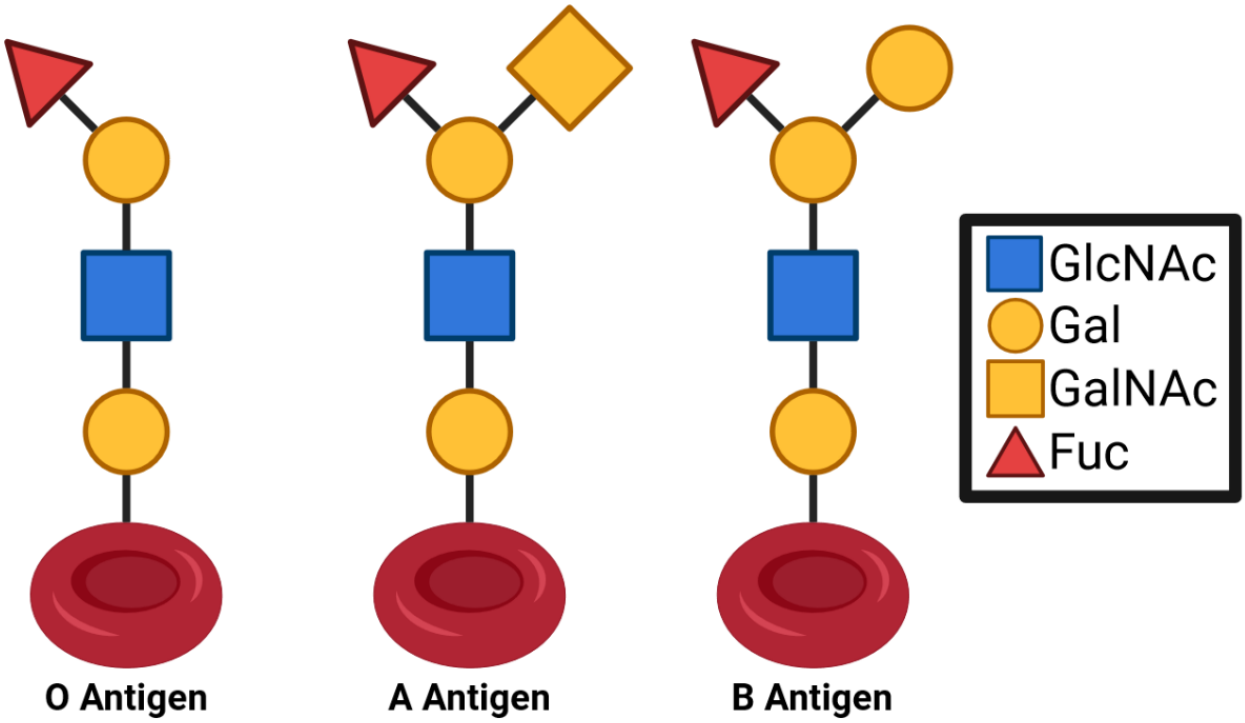


Figure 1. 5. The *O*-glycan structures presented on the surface of red blood cells determines blood type.

While type O, A, and B erythrocytes only display their own antigen, type AB erythrocytes display both the A antigen and B antigen.

Adapted from: (30)

Belický, Š., Katrlík, J., and Tkáč, J. (2016) Glycan and lectin biosensors. *Essays Biochem*. **60**, 37–47.

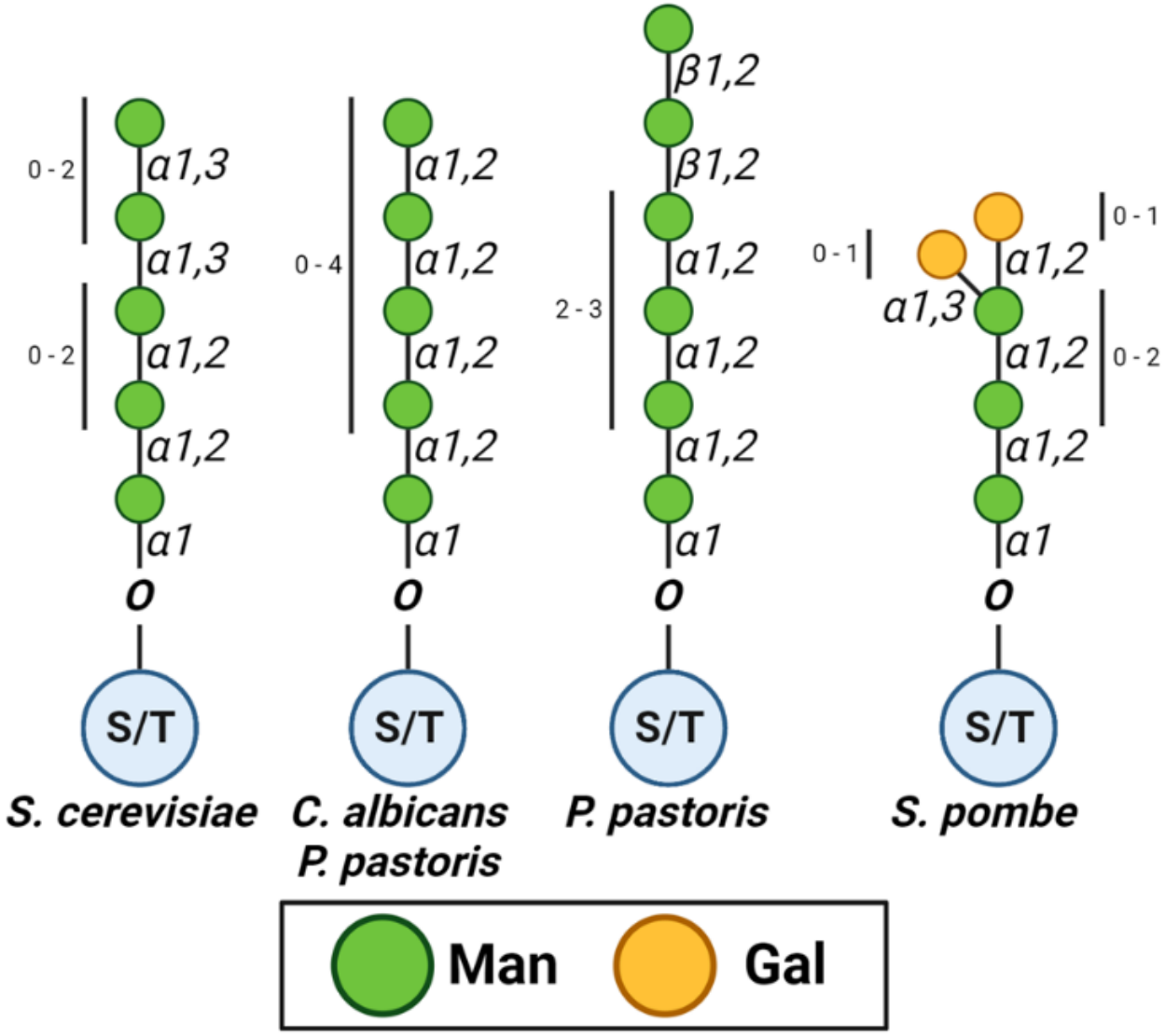


Figure 1. 6. Schematic representation of common *O*-mannosyl glycan structures in yeasts.

An α1,2-linked Man_{0-2}-Man-Ser/Thr is common in all yeasts, though subsequent α1,3 and α1,6 (not shown) linked mannoses are possible. This common *O*-mannosyl structure has also been identified in the *O*-mannose glycoproteins of actinobacteria.

Adapted from: (22)

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

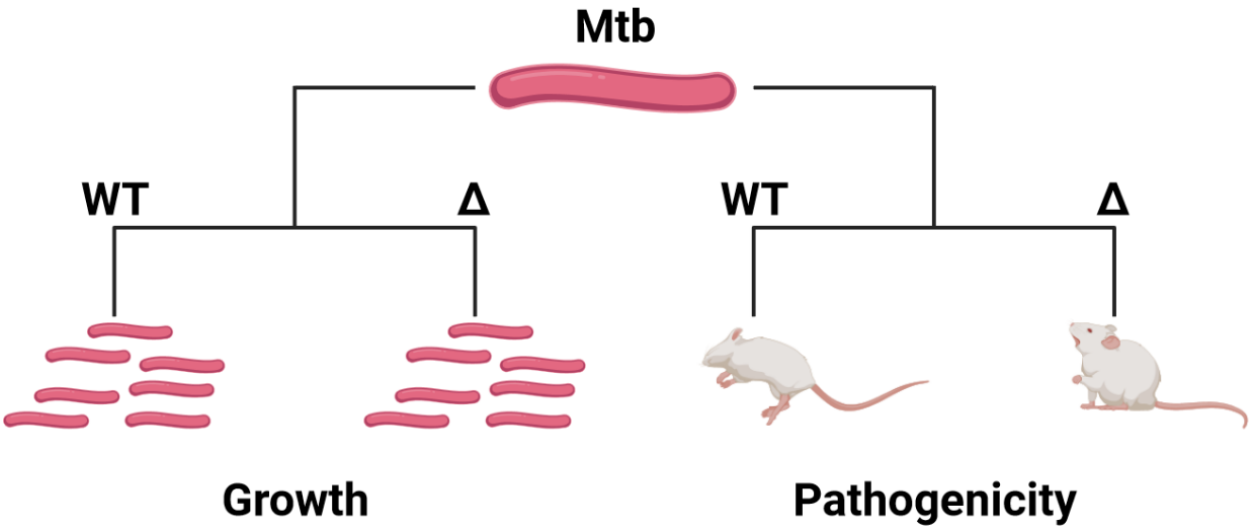


Figure 1. 7. Loss of protein-*O*-mannosylation in *M. tuberculosis* attenuates pathogenicity without impacting growth.

Knockout of the GT-39 enzyme (Δ) responsible for the initiating step of POM in *M. tuberculosis* (Mtb) results in no discernable growth phenotype and appears to be dispensable; however, this knockout dramatically attenuates pathogenicity in a mouse model. This finding showed the first potential link between (actino)bacterial POM and pathogenicity.

Adapted from: (49)

Liu, C. F., Tonini, L., Malaga, W., Beau, M., Stella, A., Bouyssié, D., Jackson, M. C., Nigou, J., Puzo, G., Guilhot, C., Burlet-Schiltz, O., and Rivière, M. (2013) Bacterial protein-O-mannosylating enzyme is crucial for virulence of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*. **110**, 6560–6565.

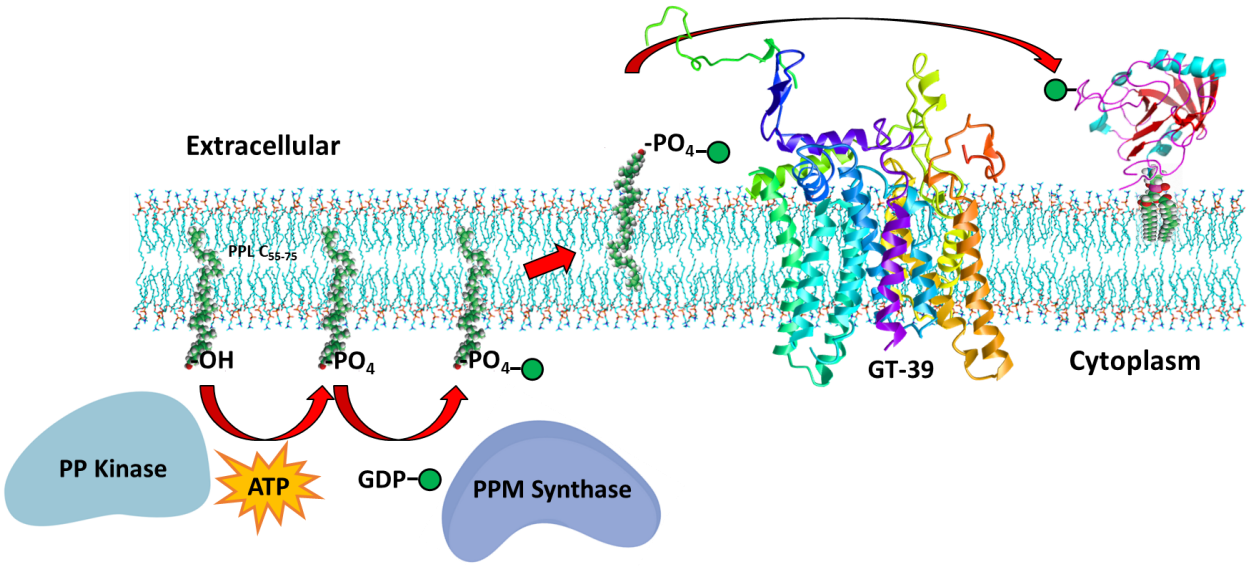


Figure 1. 8. Schematic representation of the hypothesized POM mechanism in bacteria.

As no biochemical studies have been performed with bacterial GT-39s, the generalized mechanism of POM in the domain of *Bacteria* has been inferred using data from the characterization of the process in *S. cerevisiae*. Initially, a polyprenol (PP) kinase uses ATP to phosphorylate a membrane anchored lipid donor (dolichol in eukaryotes, undecaprenol or polyprenol in prokaryotes) on the cytoplasmic face of the membrane. This phosphorylated lipid is then charged with a mannose (Man) monosaccharide (green circle) by a polyprenol phosphate mannose (PPM) synthase using GDP-Man. By a currently unknown mechanism, this charged glycolipid flips across the membrane and provides the GT-39 with an activated mannose to attach to suitable substrate proteins on S/T residues.

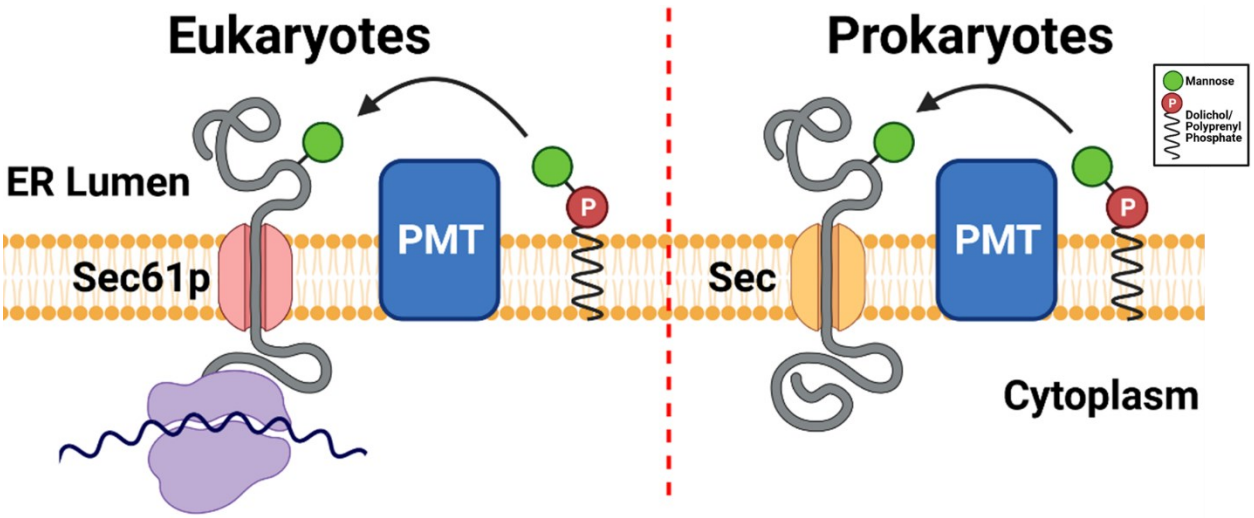


Figure 1. 9. POM is thought to occur in a SEC dependent manner in both eukaryotes and prokaryotes.

In eukaryotes (left), POM is known to occur as a protein is translocated via the SEC translocon across the ER membrane (and into the lumen). In prokaryotes (right), POM is thought to only occur in a similar SEC dependent manner after periplasmic export or secretion (dependent on the Gram-status of the organism). In both cases, mannose is provided to the protein-*O*-mannosyltransferase (PMT) by a lipid linked donor (dolichol phosphate in eukaryotes and polyprenylphosphate in prokaryotes). However, the current model of prokaryotic POM does not account for mannosylated cytoplasmic proteins, or mannosylated proteins secreted via translocons other than the SEC translocon.

Adapted from: (22, 87)

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

VanderVen, B. C., Harder, J. D., Crick, D. C., and Belisle, J. T. (2005) Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science*. **309**, 166–168.

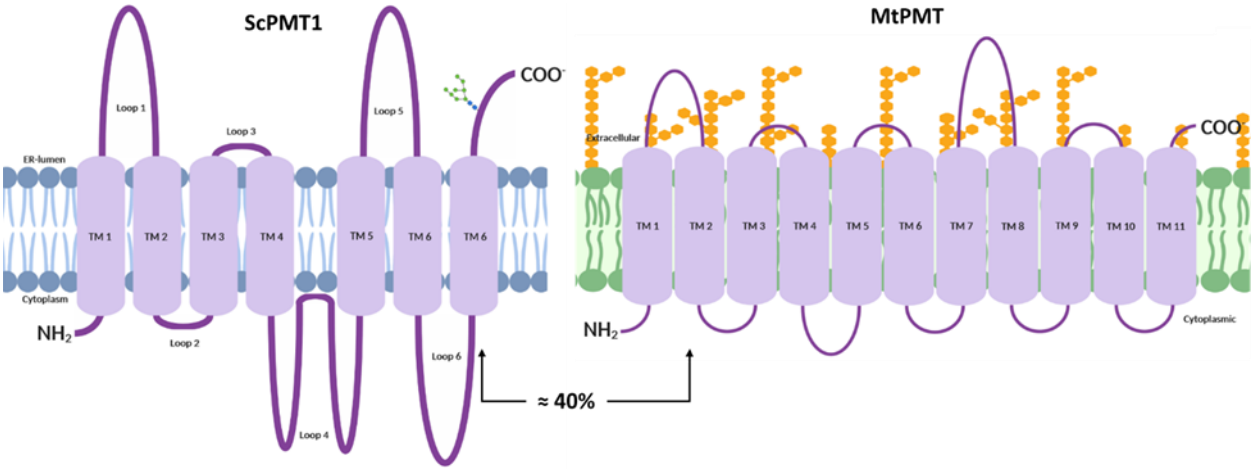


Figure 1. 10. TMHMM-2.0 predicted topologies of orthologous GT-39s from *S. cerevisiae* (ScPMT1) and *M. tuberculosis* (MtPMT).

A solved structure (3.20 Å) for the *S. cerevisiae* PMT-1 enzyme (complexed with PMT-2 and the sugar donor) was recently deposited into the Protein Database (PDB, 6P2R); however, there are currently no solved structures for GT-39 enzymes, or protein-*O*-mannosyltransferases (PMTs), of prokaryotic origin. Membrane topology predictions using hydropathy profiles – like TMHMM-2.0 – can therefore be used to approximate how bacterial GT-39s are presented *in vivo*. Both ScPMT1 and MtPMT are multipass transmembrane proteins, with the N-terminus on the cytosolic face and the C-terminus on the extracellular/ER luminal face. Both proteins also contain several cytoplasmic and extracellular/luminal loops with a conserved DE motif in loop 1 required for catalytic activity. These two GT-39s only share approximately 40% sequence homology at the amino acid level, suggesting the possibility for both structural and functional differences as evidenced by the number of transmembrane regions predicted to be contained in the MtPMT.

Adapted from: (87, 88)

VanderVen, B. C., Harder, J. D., Crick, D. C., and Belisle, J. T. (2005) Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science*. **309**, 166–168.

Girrbach, V., Zeller, T., Priesmeier, M., and Strahl-Bolsinger, S. (2000) Structure-function analysis of the dolichyl phosphate-mannose: Protein O-mannosyltransferase ScPmt1p. *Journal of Biological Chemistry*. **275**, 19288–19296.

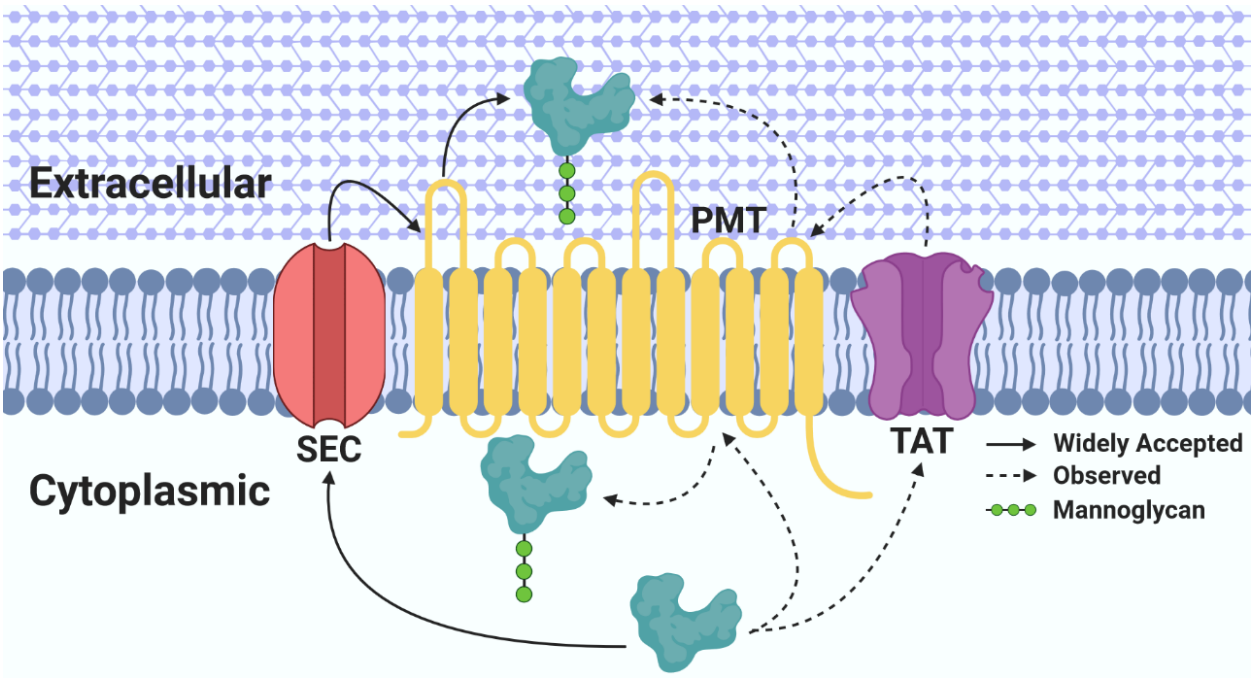


Figure 1. 11. Schematic diagram contrasting canonical POM in actinobacteria with experimental observations.

The widely accepted dogma of (actino)bacterial SEC-dependent POM dictates that target proteins must first be secreted through the SEC translocon, and subsequently are mannosylated by a GT-39/PMT enzyme on the extracellular face of the membrane. This hypothesis does not account for the number of mannoproteins identified to be TAT secreted or even present in the cytoplasm of actinobacteria.

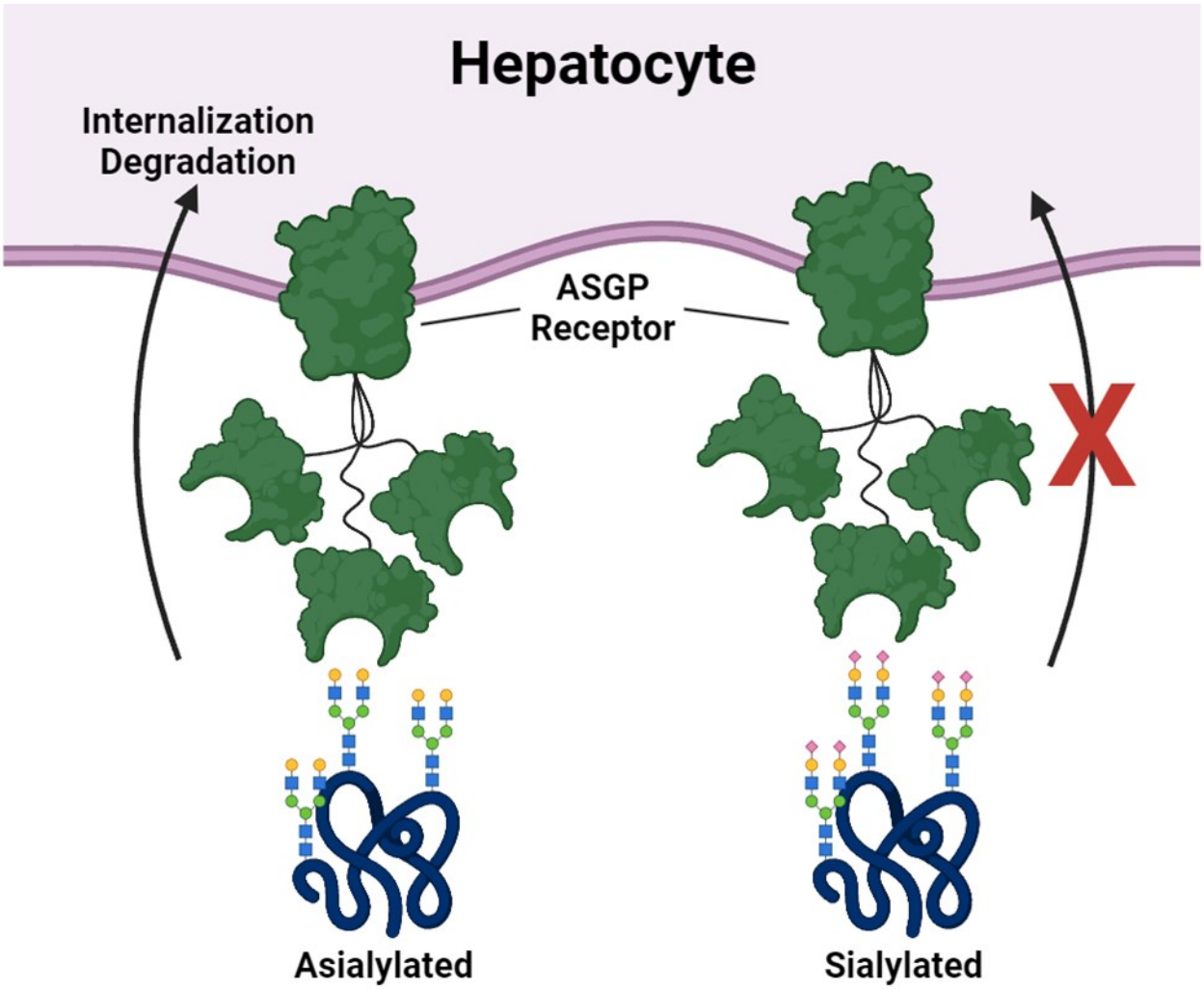


Figure 1. 12. Asialoglycoprotein internalization by hepatic asialoglycoprotein receptors (ASGP).

Terminal sialic acids on glycans act to mask serum glycoproteins from recognition by hepatic ASGP receptors – which recognize terminal galactose moieties on glycans – and are thusly protect them from uptake and degradation. As sialylated glycoproteins age, their sialic acid content progressively decreases making them a target for ASGP receptors – in this context, the sialic acid content of any given glycoprotein acts as a biological timer. Increasing the sialic acid content of therapeutic glycoproteins, either chemically or enzymatically, is an effective method towards improving their circulating half-life.

Adapted from: (89)

Maverakis, E., Kim, K., Shimoda, M., Gershwin, M. E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L. R., and Lebrilla, C. B. (2015) Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J Autoimmun*. **57**, 1–13.

2. TOWARDS AN EXPERIMENTAL SYSTEM FOR THE EXAMINATION OF PROTEIN MANNOSYLATION IN ACTINOBACTERIA

as written by Saxena, H., Buenbrazo, N., Song, W.Y., Li, C., Brochu, D., Robotham, A., Ding W., Tessier, L., Chen, R., Kelly, J., & Wakarchuk, W.

Glycobiology, 33(6):512-524 2023

2. 1. Chapter Overview

This chapter describes the identification of abundant POM in the cell-associated fraction of the actinobacteria *C. fimi* and *C. flavigena*. In addition to mapping the glycosylation sites of several glycoproteins, both cytoplasmic and TAT secreted mannoproteins were observed. These findings go against the collective consensus that POM only occurs in a SEC-dependent manner in actinobacteria and provides the rationale for the later investigation into the relationship between secretion pathways and POM. In addition, this chapter describes the validation of utilizing *C. glutamicum* as a recombinant host system for the study of actinobacterial POM. The accurate glycosylation of known actinobacterial mannoproteins, produced recombinantly in *C. glutamicum* showed that the organism could be purposed towards the *in vivo* study of actinobacterial POM.

2. 2. Abstract

The actinobacterial species *C. fimi* ATCC484 has long been known to secrete mannose-containing proteins, but a closer examination of glycoproteins associated with the cell has never been reported. Using ConA lectin chromatography and mass spectrometry we have surveyed the

cell associated glycoproteome from *C. fimi* and collected detailed information on the glycosylation sites of 19 cell-associated glycoproteins. In addition, we have expressed a previously known *C. fimi* secreted cellulase, Celf_3184, (formerly CenA), a putative peptide prolyl-isomerase, Celf_2022, and a penicillin binding protein, Celf_0189, in the mannosylation capable host, *C. glutamicum.* We found that the glycosylation machinery in *C. glutamicum* was able to use the recombinant *C. fimi* proteins as substrates and that the glycosylation matched closely that found on the native proteins when recovered from *C. fimi*. We are pursuing this observation as a prelude to dissecting the biosynthetic machinery and biological consequences of this protein mannosylation.

2. 3. Introduction

2. 3. 1. Actinobacterial protein-*O*-mannosylation

Protein mannosylation has been observed from bacteria to humans and appears to be a highly conserved protein modification that is essential in eukaryotes (22). Protein mannosylation in actinobacterial species was originally described over 40 years ago (90), however we still lack a basic understanding of the importance of this post-translational modification in prokaryotes. Mannosylated proteins have been discovered in many well-known actinobacteria including *Mycobacterium tuberculosis* (91–93), *Corynebacterium glutamicum* (56, 94), *Cellulomonas fimi* (95, 96), and *Streptomyces coelicolor* (97, 98). Functions for only a few of these mannosylated proteins have been discovered, making it difficult to decipher the biological significance of this modification within these bacteria.

The first mannosylated proteins characterized in actinobacteria were the secreted cellulases from *Cellulomonas* spp., where the glycosylation was shown to protect them from

proteolysis (96). In *Streptomyces sp.,* a cell surface phage receptor/phosphate binding protein was shown to require mannosylation to function (99). In *C. glutamicum,* a few secreted proteins were shown to be glycosylated, and one of them – the resuscitation-promoting factor – was shown to require glycosylation for function (56). *Mycobacterial* mannoproteins were first described as cell surface antigens (91). The list of mannosylated proteins identified from mycobacterial culture filtrates included approximately 70 proteins, composed mostly of putative lipoproteins, as well as periplasmic and secreted proteins (93). More recently, the lack of protein mannosylation was shown to influence laboratory growth rates and to decrease persistence of *M. tuberculosis* in a mouse model suggesting that, while not essential for laboratory growth, protein mannosylation has an important role to play in the virulence of this bacterium (49). Recently *S. coelicolor* was shown to have at least 38 mannoproteins, putative lipoproteins, many membrane proteins and possibly intracellular proteins (98). In addition, the lack of protein mannosylation has been shown to effect cell wall integrity in *S. coelicolor*, where cells lacking this PTM are more sensitive to a variety of cell wall active antibiotics (100), and subsequent work (98) showed a putative peptidoglycan remodeling protein was likely responsible for that phenotype. In all these species, the glycans are O-linked to Ser/Thr residues and they are short, 1-4 hexose chains, without branching (47, 101).

2. 3. 2. Protein-*O*-mannosyltransferases (PMTs)

The enzyme responsible for the initiation of protein mannosylation belongs to the highly conserved glycosyltransferase family GT-39 (102). Protein mannosylation takes place on serine and threonine residues but there is no consensus glycosylation site (sequon) as there is for *N*-linked glycosylation. This is typical for *O*-linked glycosylation where no sequence "rules" are

found for the various forms of *O*-glycosylation. This family of GT-39 enzymes has been studied mostly in eukaryotes where it is described as: Dol-P-Man:protein α-mannosyltransferase (PMT). The GT-39 family of glycosyltransferases are present in >3000 species of bacteria (from http://www.cazy.org/GT39_bacteria.html) as single copy genes compared to the normally, multicopy genes seen in eukaryotes. The PMT enzymes are multi-transmembrane domain proteins (103) which are predicted to have related topology in the membrane. The eukaryotic PMT enzymes use a lipid-linked mannose, dolichol phosphate-β-D-mannose, as a donor, rather than the simple sugar nucleotide, GDP-α-mannose (104). The charging of the lipid is performed by the GT-2 family enzyme, dolichol-phosphomannose synthase (105). In bacteria the donor for PMT is not a dolichol lipid, but the structurally related and terminally unsaturated C_{55} isoprenyl lipids (106), which are charged by a related GT-2 family enzyme prenyl-phospho-mannose synthase (PPM) (107).

Recently, a cryo-electron microscopy structure for a GT-39 hetero-dimer from *S, cerevisiae* was solved which revealed that these large proteins are multipass membrane proteins with an architecture related to oligosaccharyltransferases from both pro- and eukaryotes (108). The complex of PMT1 with PMT2 reveals much about substrate recognition and function of the enzyme as an inverting glycosyltransferase. The eukaryotic GT-39 enzymes are essential in multicellular organisms however the bacterial members of this enzyme class are dispensable for lab growth. The bacterial orthologues have not been studied with the same intensity except in *M. tuberculosis* where the mannoproteins appear to play a significant role in pathogenesis (49), but this leaves us with questions about the normal biology of this post-translational modification in the >3000 species of bacteria which carry the GT-39 enzyme. What does mannosylation

contribute to those proteins that possess it?

2. 3. 3 Use of *C. glutamicum*

In this manuscript, we present our preliminary survey of the mannoproteome from the cell associated fraction of *C. fimi* ATCC484, which has previously been shown to have this modification on selected secreted proteins (96). Our longer-term goal is to develop a recombinant system in *C. glutamicum* to systematically study the addition of mannose to various cell associated and secreted proteins to deduce the roles of the modification. *C. glutamicum* shows potential as a surrogate host to produce authentically glycosylated actinobacterial proteins due to their close genetic relatedness (Figure S2.1) and common mannosylation machinery. In addition, *C. glutamicum* has been widely utilized industrially to produce amino acids, nucleotides, and more recently enzymes – meaning that there is an established molecular toolbox for genetic engineering in this organism (109, 110). As a proof of concept, here we present the use of *C. glutamicum* as a surrogate actinobacterial host to produce authentically glycosylated proteins from *C. fimi.*

2. 4 Results

2. 4. 1 Lectin blotting to examine the complexity of the mannoproteome

We had previously investigated the secretome of *C. fimi* as a function of carbon source (111) and noticed that glycosylated cellular proteins were present. Previous research on *C. fimi* secreted cellulases had shown that they bound to the lectin ConA and contained mannose (95, 96). Therefore, we performed a lectin analysis for mannosylation of cellular proteins from *C. fimi* ATCC484 as well as the planned expression host *C. glutamicum* ATCC 13032. ConA lectin

blots of membrane fractions from both strains showed unique mannoprotein profiles, which are shown in Figure 2.1.

2. 4. 2. Glycoproteomics survey

Cellulomonas strains were grown on both carboxymethylcellulose (CMC) and soluble xylan. These growth conditions have been shown to induce different secreted CAZyme proteins (111) and so we anticipated that differences in cellular protein expression would be detected, especially for cell surface proteins involved in carbohydrate uptake. We performed an initial enrichment of cell extract proteins on ConA lectin resin, and selected fractions were analyzed by SDS-PAGE. Protein bands that were reactive by ConA lectin blotting were analyzed by in-gel tryptic digestion and nano LC-MS/MS. *C. fimi* proteins were identified by searching the peptide MS/MS spectra against the *C. fimi* protein sequence database using the Mascot™ search algorithm. Multiple proteins were identified in this manner (data not shown), but none could be confirmed as glycosylated proteins.

Next, selected ConA fractions were digested directly with trypsin and further enriched for glycopeptides by ion-pairing normal phase liquid chromatography HPLC (IP-NPLC) (112). Glycopeptide enriched IP-NPLC fractions were analyzed by nanoLC-MS/MS and matched to the proteins listed in Table 2.1. Most of the glycopeptides were identified in both the CMC and xylan IP-NPLC fractions, though their abundances tended to be greater in the latter. Many of the proteins from which the glycopeptides were derived were also identified in the in-gel tryptic digest analysis of the ConA reactive bands (Table 2.1). Those proteins that were not identified in the in-gel digest analysis could still have been present but at levels below the threshold for this

analysis (*i.e*., peptide ions were not intense enough to trigger MS/MS spectral acquisition). All the detected glycopeptides were modified exclusively with hexoses.

An example of the nanoLC-MS and -MS/MS spectra acquired for a tryptic glycopeptide derived from Extracellular Ligand Binding Receptor (ABC transporter substrate-binding protein, Celf_1830) is presented in Figure 2.2, panels A and B. This glycopeptide was identified in both the xylan and CMC fractions and was modified with 0, 1, or 2 hexoses. The nanoLC-MS/MS spectra for the remaining glycopeptides that were identified in this study are presented in Figure S2.2. Some of these glycopeptides are modified with up to 9 hexoses. The Celf_1830 glycopeptide modified with one hexose was also analyzed by LC-ETD-MS/MS on an ion trap mass spectrometer (Figure 2.2, panel C). ETD preserves delicate modifications such as *O*-linked glycans during the fragmentation process and it is often possible to identify the site(s) of glycosylation. An extensive z fragment ion series can be observed in the ETD spectrum which identifies the site of *O*-glycosylation as Thr35.

2. 4. 3. Bioinformatic analysis of *C. fimi* mannosylated proteins

BLAST analysis was performed with selected relatives of *C. fimi*, namely *C. flavigena*, *C. glutamicum*, *M. tuberculosis,* and *S. coelicolor.* *C. glutamicum* and *M*. *tuberculosis* have been shown to have secreted mannoproteins (56, 93). A mannoproteome from *S. coelicolor* has also been reported (98). This BLAST analysis revealed that several of the proteins are conserved across the five species (Table S1). This preliminary search for orthologues in organisms with demonstrated mannosylation suggests that in these taxa the role of the glycosylation on a particular protein might be similar, which gives us more protein targets to work with.

2. 4. 4. *Corynebacterium glutamicum* expression of *Cellulomonas* glycoproteins

Three *C. fimi* proteins were expressed both in *E. coli* and the surrogate expression host *C. glutamicum.* The protein sequences are shown in Figure S2.3. We chose Celf_3184 (previously known as CenA, or Cf_Cel6A), a secreted, highly glycosylated endoglucanase that was the first mannoprotein identified from *C. fimi* supernatants (96) and was previously expressed as a recombinant protein in *Brevibacterium lactofermentum* (113). Celf_2022 was selected as a second protein as it carried 6-8 hexose residues, lacks the predicted lipoprotein leader we saw on many other identified proteins (Table 2.2) and has a strong annotation as a peptidyl-prolyl cis-trans isomerase cyclophilin type. Based on the lack of predicted leader sequence we expected it to be cytoplasmic and, so far, orthologues have not been characterized from any *Cellulomonas* species. Our third mannoprotein, Celf_0189, was annotated as a penicillin binding protein with a putative lipoprotein leader sequence (Table 2.2, Figure S2.2).

Based on a SignalP5.0 analysis (Figure S2.3), the three proteins should all have different cellular locations. Celf_3184 has a strong TAT secretion leader, the Celf_2022 has no strong predicted leader sequence, and the Celf_0189 has a 50% probability of a lipoprotein leader sequence. These predictions are mostly consistent with the protein expressed in *C. glutamicum*, although we find Celf_2022 to be mostly a secreted protein and not cytoplasmic as we originally thought.

2. 4. 5. Expression and characterization of the glycoproteins

2. 4. 5. 1. Celf_3184

Previous work with Celf_3184 (CenA) (96) had shown the linker to be glycosylated between the carbohydrate binding module (CBM) and the catalytic domain. This protein is also

natively secreted though the TAT secretory pathway and was thus a good protein to measure the secretion and glycosylation potential of the *C. glutamicum* expression system. We compared the expression levels and localization of Celf_3184 in *E. coli* and *C. glutamicum*. The protein was recovered from the cytoplasm of *E. coli* and in the culture supernatant from *C. glutamicum*. Only the *C. glutamicum*-expressed protein was ConA reactive (Figure 2.3), and intact LC-MS analysis indicated that it was extensively modified with hexoses (Figure 2.4). NanoLC-MS/MS of the chymotryptic digest determined that the expected signal peptide was removed (data not shown) and that glycosylation was associated with the "PT" linker region (Figure S2.4A).

2. 4. 5. 2. Celf_2022

Initially, prediction for the localization of Celf_2022 was performed using the SingalP 3.0 server. This prediction failed to accurately detect a leader sequence for Celf_2022, which was, thus, assumed to be localized to the cytoplasm. However, a large amount of Celf_2022 was harvested from the culture supernatant with little being recovered from within the *C. glutamicum* cells, suggesting that the protein is actively secreted. With the release of SignalP 5.0, a signal peptide not belonging to either the Sec or TAT export pathways was suggested (Figure S2.3) corroborating our findings that the protein is actively secreted. The *C. glutamicum*-derived protein was larger in mass relative to the protein expressed in *E. coli* and it reacted with ConA, indicating glycosylation (Figure 2.3). Intact mass analysis of Celf_2022 expressed in *E. coli* determined its molecular weight to be 21,474 Da (Figure S2.5), corresponding to the loss of the first 53 amino acids from the N-terminus, which was confirmed by nanoLC-MS/MS (data not shown). Intact mass analysis of the *C. glutamicum*-expressed protein revealed that it was modified with hexoses and that it also appeared to have lost the first 53 amino acids (Figure

S2.6, panel B). Analysis of the glycopeptides enriched from the tryptic digest of *C. glutamicum*-expressed Celf_2022 (Figure S2.4, panel B) determined that the hexoses were attached in the same region as observed in the original glycoproteomics survey of *C. fimi* (Table 2.1).

2. 4. 5. 3. Celf_0189

ConA blotting confirmed that Celf-0189 expressed in *C. glutamicum* was glycosylated (Figure 2.5). It was not possible to obtain an intact mass spectrum for the putative penicillin binding protein as it could not be maintained in solution without the aid of detergents that are incompatible with LC-MS. Instead, the protein was digested with endoproteinase Glu-C and glycopeptides were enriched from the digest using Polyhydroxyethyl A media (112). LC-MS/MS analysis (Figure S2.6, panel a) revealed a number glycopeptide peaks, all of which were derived from the same region of the protein. The peak at 14.6 minutes, for example, was dominated by a triply charged ion at *m/z* 1123.51 (Figure S2.6, panel B). MS/MS analysis identified the ion as ^{568}TTATPAAPSADATARPQVPLTTEE^{591} modified with 6 hexoses (Figure S2.6, panel C). This peptide sequence also aligns with that of the Celf_0189 tryptic glycopeptides identified in original *C. fimi* study described above (Table 2.1, Figure S2.2, panel B).

2. 4. 6. HILIC-HPLC analysis of released *O*-glycans

Significant glycosylation was detected in all the recombinantly expressed proteins, we wanted to assess the size of the attached glycans. We used chemical release and tagging of the *O*-linked glycans followed by analysis on HILIC HPLC (Figure 2.6) with fluorescence detection (HILIC-FLD). We made standards from both glucosides and mannosides (mono, di- and tri-saccharides) which showed we could differentiate α-glucosides from α-mannosides. When

compared to standards, we observed mostly disaccharides of mannose on both Celf_2022 and Celf_3184, with a possible trace of trisaccharide from Celf_3184. Presence of mannose in these glycans is shown through the digestion products from α-mannosidase is shown in Figures S2.8 and S2.9.

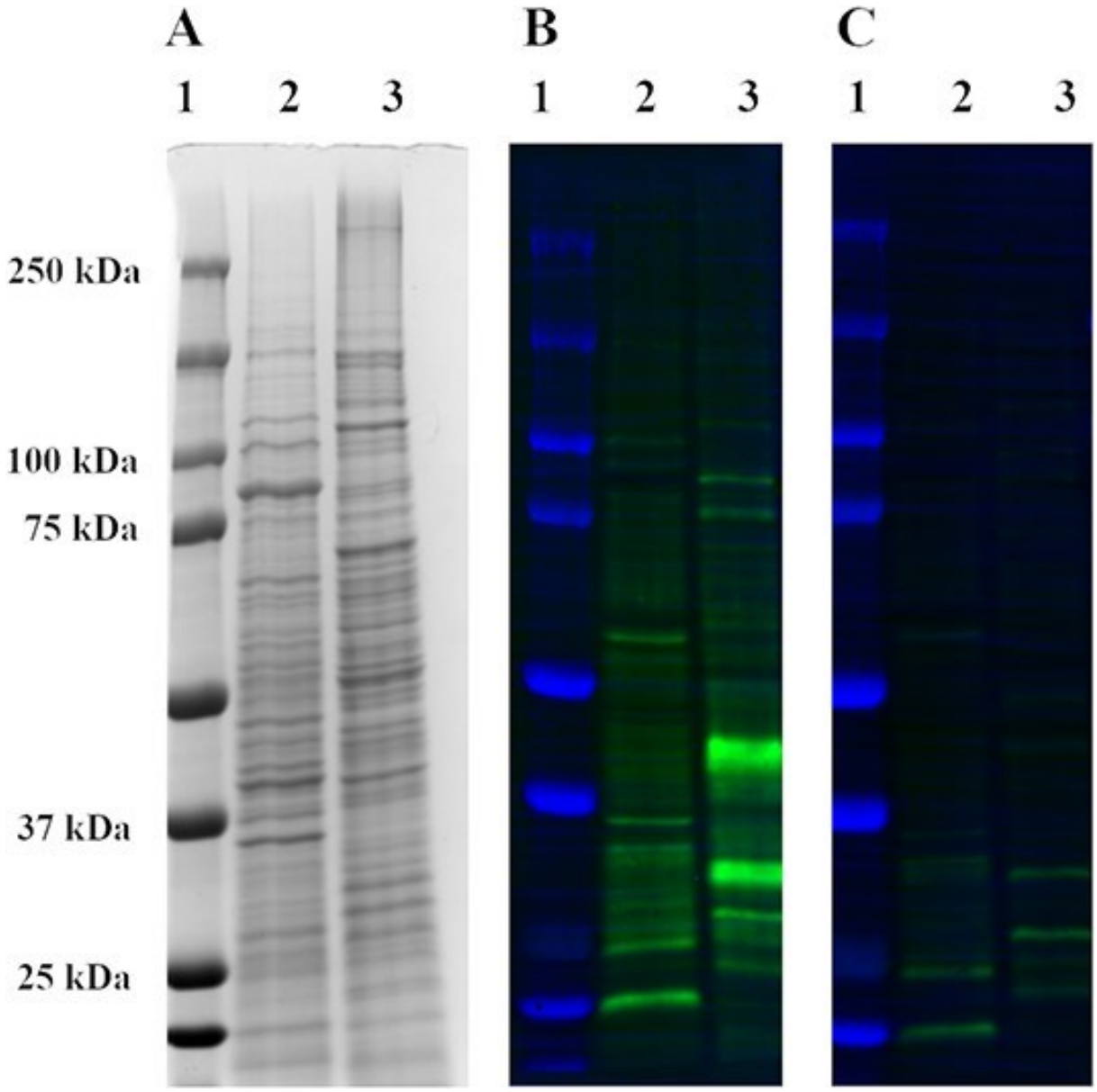


Figure 2. 1. Detection of ConA reactive proteins in cell associated proteins from *C. fimi* ATCC 484 and *C. glutamicum* ATCC 13032.

(A) Coomassie stained 7-12% gradient SDS-PAGE. (B) Western blot image using ConA-FITC conjugate (green) and colorimetric showing protein ladder (blue, C). Western blot image with 500 mM of α-methyl-glucopyranoside. Lane 1, ladder, lane 2, *C. fimi* ATCC 484 membrane fraction, lane 3 *C. glutamicum* ATCC 13032 membrane fraction. Molecular weight standards are the Bio-Rad All Blue ladder.

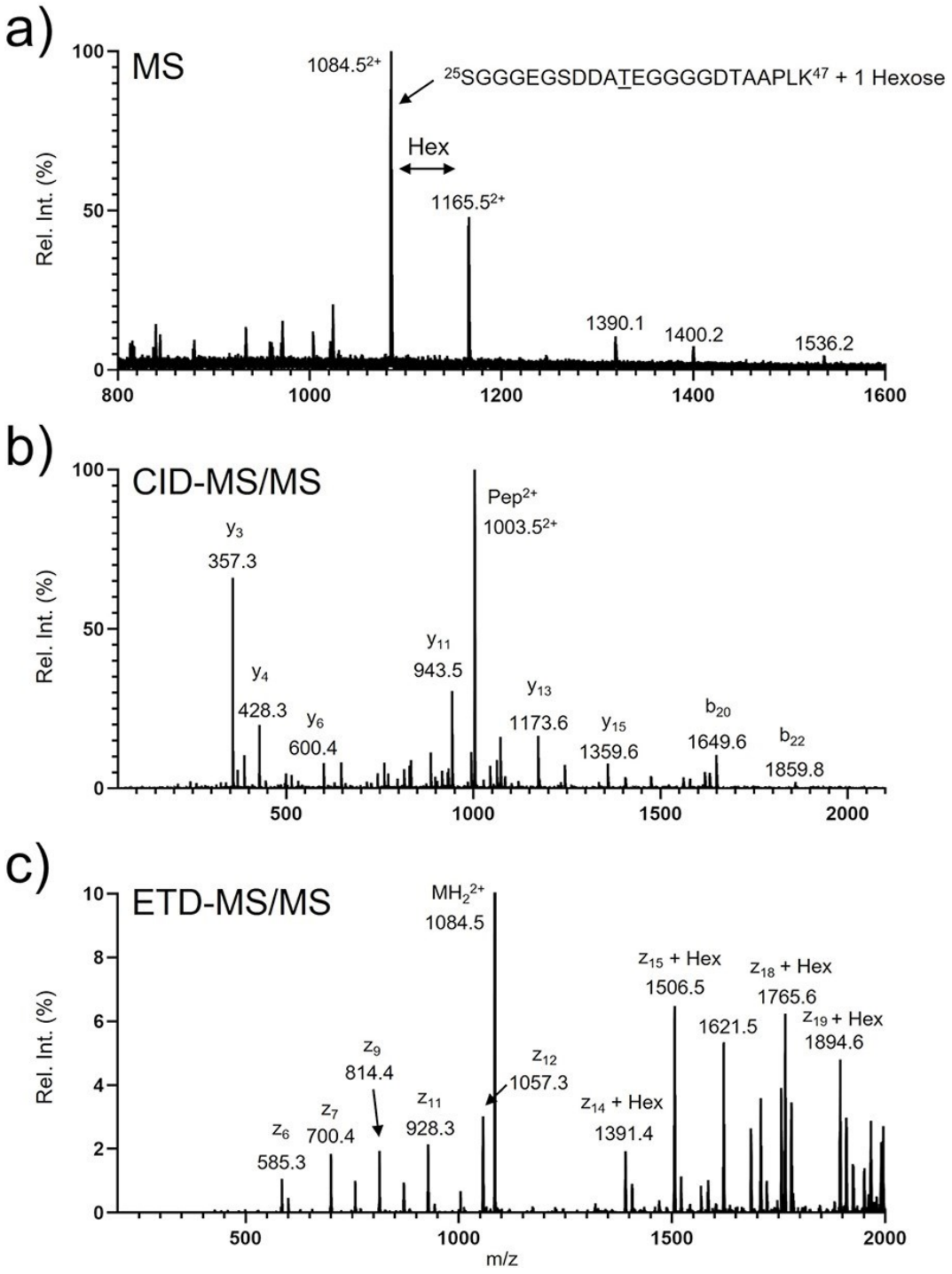


Figure 2. 2. NanoLC-MS and -MS/MS spectra from the Celf_1830 glycopeptide.

(A) NanoLC-MS spectrum acquired for this glycopeptide indication that it is modified by either one or two hexoses. (B) CID-MS/MS spectrum for the doubly protonated glycopeptide ion at m/z 1084.5. (C) ETD-MS/MS data for the same glycopeptide ion.

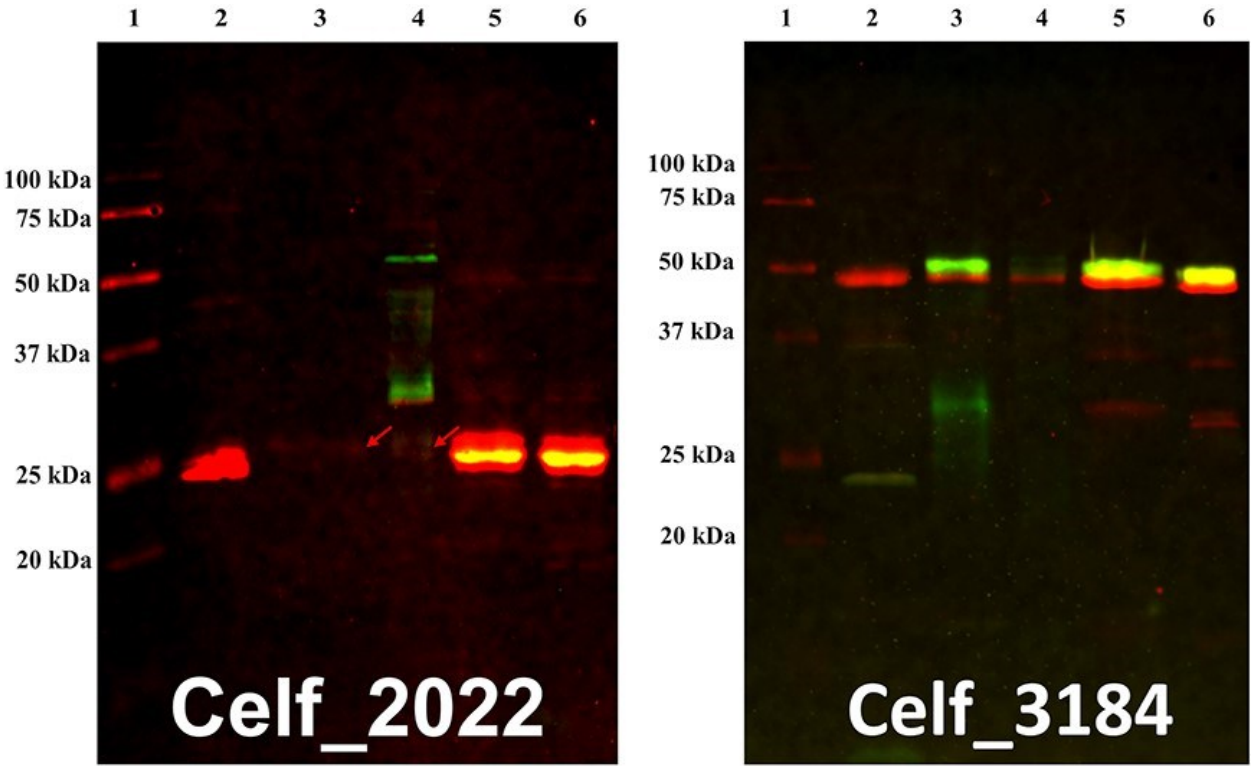


Figure 2. 3. ConA-FITC (green) and Alexa Fluor 647 anti-HIS6 (red) fluorescent blot of C. fimi glycoproteins recombinantly produced in and secreted from *C. glutamicum*.

Proteins Celf_2022 (left) and Celf_3184 (right) were expressed and purified from both *E. coli* and *C. glutamicum*. Lane 1, ladder, lane 2 in each blot is the respective protein purified from *E. coli*, lane 3 is the un-concentrated culture supernatant from *C. glutamicum*, lane 4 is *C. glutamicum* lysate producing the recombinant protein, lane 5 is the purified protein from the *C. glutamicum* culture supernatant, and lane 6 is the protein treated with α-mannosidase. Recombinant proteins purified from *C. glutamicum* exhibit a decrease in molecular weight when treated with α-mannosidase but are still ConA-reactive due to the first mannose residue that is still linked to the peptide.

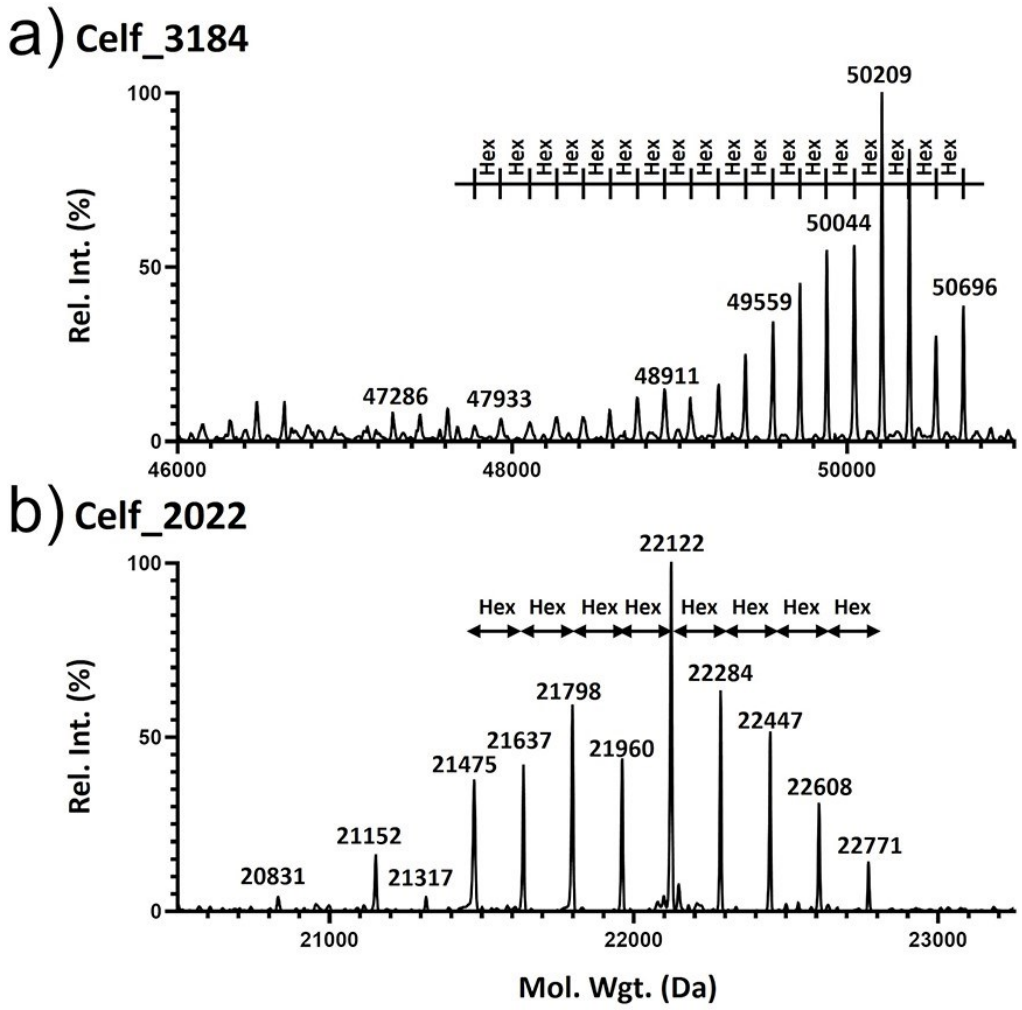


Figure 2. 4. Intact mass LC-MS analysis of (A) Celf_3184 and (B) Celf_2022.

Both proteins were expressed *C. glutamicum* and recovered from the culture supernatant. The calculated mass (signal peptide removed and with no modifications) for Celf-3184 is 44,639 Da. The observed mass profile (panel A) suggests that this protein is modified with up to 37 hexoses. The calculated mass for Celf-2022 is 21,475 Da. The mass profile for the *C. glutamicum*-expressed protein (panel B) indicates that it is modified with 0-8 hexoses. The peaks at 20,831, 21,152 and 21,317 Da are likely to be hexose glycoforms of a truncated form of Celf_2022 in which the N-terminal residue is L63 rather than L54. The presence of this truncated N-terminal sequence was confirmed by nanoLC-MS/MS analysis of the tryptic digest of Celf_2022.

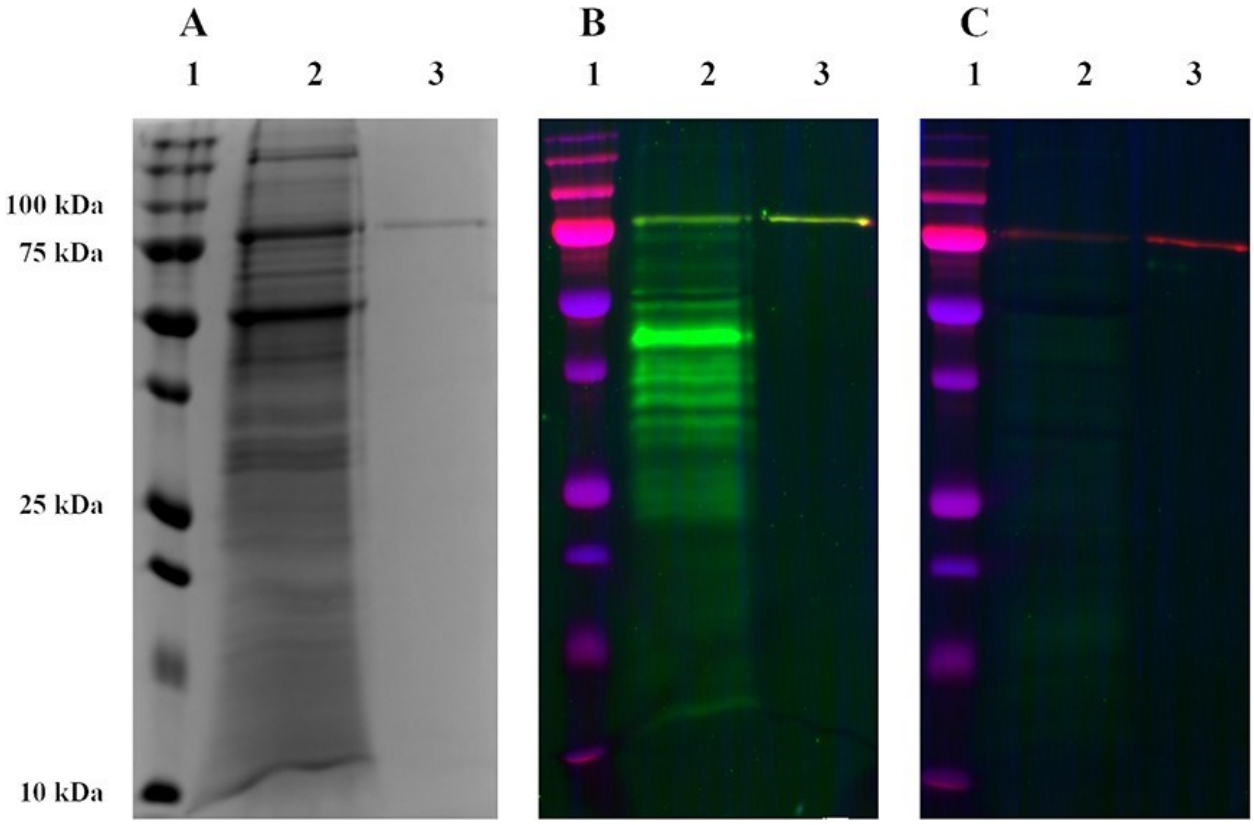


Figure 2. 5. Expression of mannosylated Celf_0189 in *C. glutamicum*.

The membrane fraction of *C. glutamicum* recombinantly producing CElf_0189 was separated on a 12 % SDS-PAGE gel. (A) Coomassie staining and (B) Western-blot image using anti-histidine-Alexa 647 (red), ConA-FITC conjugate (green), and colorimetric showing protein ladder (blue, C). Western blot with 500 mM α-methyl-glucopyranoside. Lane 1, ladder, lane 2, membrane fraction of *C. glutamicum* producing Celf_0189, lane 3 membrane purified Celf_0189. Molecular weight markers are the Bio-Rad All Blue ladder.

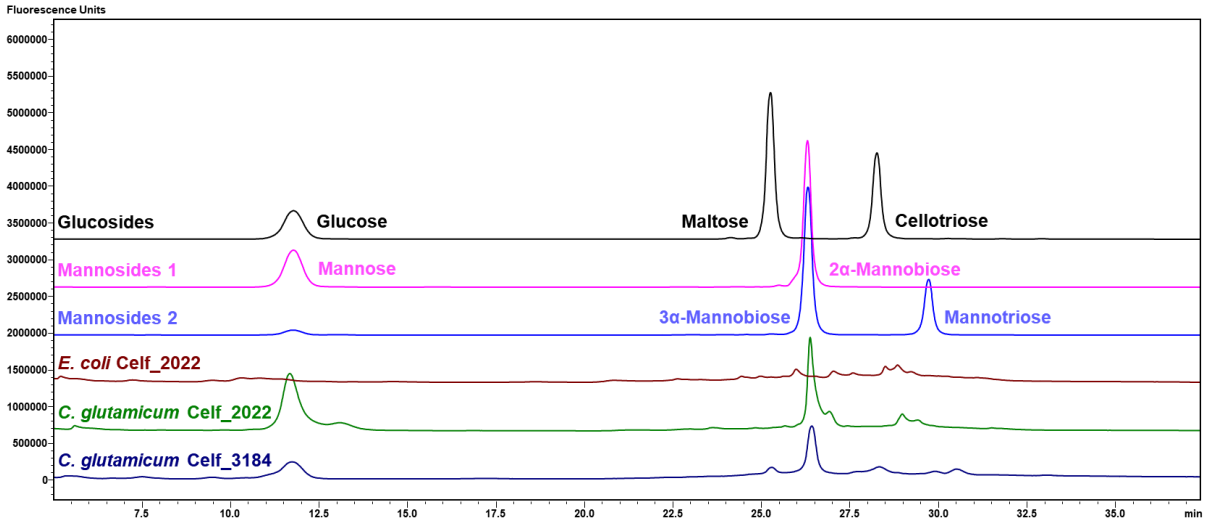


Figure 2. 6. HPLC analysis of 2AB labelled glycans released from *C. glutamicum* expressed Celf_2022 and Celf_3184.

Glucosides (black trace) show monosaccharide (glucose), disaccharide (maltose), and trisaccharide (cellotriose) reference compounds while mannoside standards (magenta and blue) show monosaccharide (mannose), disaccharide (α-1,2-mannobiose, α-1,3-mannobiose), and trisaccharide (α-1,3-mannotriose) reference compounds. *E. coli* expressed Celf_2022 (brown trace). *C. glutamicum* expressed Celf_2022, (green trace) shows essentially only glycans composed of 2 mannose residues. *C. glutamicum* expressed Celf_3184 (dark blue trace) shows a mixture of single and di -mannose glycans.

Table 2. 1. Glycoproteins identified from *Cellulomonas fimi.*

Gene	Annotation	Glycopeptide	Mass (Da)	# Hex Residues
Celf_0029	Stk1 family PASTA domain-containing Ser/Thr kinase	^{652}GPGGNNNGGATPNPTSKP^{669}	70,030	4
Celf_0189	Penicillin-binding protein transpeptidase	^{557}LVDVPAEESAETTATPAAPSADATARPQVPLTTEEAQALR^{596}	68,580	8
Celf_0553	NLP/P6 protein (C40 non peptidase)	^{367}TQPTTPPATGGAYGLGTGR^{385}	50,990	1
Celf_0777	Hypothetical protein	^{41}STPTPTPEVAELTTEDFVAR^{60}	28,730	6
Celf_0922	Hypothetical protein	^{39}GDGSQQPSPTSVTQR^{53}	38,040	5
Celf_1045	Hypothetical protein	^{203}VEPTAPSGR^{211}	21,411	2
Celf_1347	Extracellular solute binding protein family 1	^{19}SSGDDTGSDDATDAGTPAPATIK^{41}	46,700	3
Celf_1421	Hypothetical protein	^{29}SDSGAEDEPTAVETTTESSTEDEAASGQVEVIDQTHLLLATAQDTLEAR^{77}	13,290	3
Celf_1453	Periplasmic binding protein	^{19}ALGASACAADAQAGGATSATPTTGSGFTPVELDNCGTPVTVEAPPER^{65}	34,920	3
Celf_1573	Extracellular solute-binding protein family 3	^{21}SGGGDDDTDAGAGGDATETSGGAEGTIR^{49}	29,760	3
Celf_2022	Peptidyl-prolyl cis-trans isomerase cyclophilin type	^{63}LPSPVVSDQPQATAQPAR^{80}	26,700	7
Celf_2833	Hypothetical protein	^{43}DAATAAQDAVGDLR^{56}	17,100	2
Celf_2865	Hypothetical protein (BMP family ABC transporter substrate-binding protein)	^{22}GSAPEETEETAGTGDAAASDFK^{43}	37,840	1
Celf_3229	Periplasmic binding protein	^{24}SGSTDDDATAEPGSTASDGTTFPVTIESALGTAVVEEKPER^{64}	35,820	5
Celf_3336	Extracellular solute-binding family 1	^{31}AGSGEPAAEATSSGPPEPVEIR^{52}	47,730	6
Celf_3491	periplasmic binding protein/LacI transcriptional regulator (sugar-binding protein)	^{22}SGGGAGSTDGASEGSGGGGEGGLVGVAPTQTSER^{56}	40,632	2
Celf_3669	Periplasmic binding protein	^{24}GTTEEAGAAEPSTSADVAGGPVTITDDR^{51}	34,060	4
Celf_3689	Peptidylprolyl isomerase FKBP-type	^{27}SGDGGSDDSPSADASASVSAAQAADQAALAK^{57}	31,970	6

Table 2. 2. *Cellulomonas* glycoproteins that are potential lipoproteins.

Gene	Annotation	Predicted signal peptidase II cleavage site: LX[A/G]C and glycopeptide
Celf_3229	Periplasmic binding protein	MRHLRRTAVAASVAVLALA**LAAC**[a]**SGSTDDDATAEPGSTASDGTTFPVTIESALGTAVVEEKPER**[b]//[c]
Celf_0591	Periplasmic binding protein	MKLSLHRRAGAVALTGALALA**LAAC**GSDDPTGSGDTTPGSDESVSTL**SGELNGAGASSQEK**//
Celf_3669	Periplasmic binding protein	MRNLRALPAAALAGTAALL**LAACGTTEEAGAAEPSTSADVAGGPVTITDDR**//
Celf_3336	Extracellular solute binding family 1	MRHRPALRGRTAVLRSAALLAVGALA**LTACAGSGEPAAEATSSGPPEPVEIR**//
Celf_1573	Extracellular solute binding family 3	MRARRLVALAAAATLG**LAACSSGGGDDDTDAGAGGDATETSGGAEGTIR**//
Celf_1830	Extracellular ligand binding receptor	MIRSTHAVRAAALAGAAALI**LAACSGGGEGSDDATEGGGGDTAAPLK**//
Celf_1347	Extracellular solute binding protein family 1	MLRFAAVGVAAALT**LTACSSGDDTGSDDATDAGTPAPATIK**//
Celf_0777	Hypothetical protein	MLRSRLFALPAAVLLTAT**LAAC**GGGSDSPSAEKTSASSEKS**TPTPTPEVAELTTEDFVAR**//
Celf_0922	Hypothetical protein	MVRVTGTTRATAGRHGRARAATRWAVPLLGAVLA**LAACGDGSQQPSPTSVTQR**//
Celf_1045	Hypothetical protein	MAARTRHGRPPDRADGSPRPRAAAALVVAVVGT**LAGC**AGAAPEMT//TDYPVER**VEPTAPSGR***
Celf_1421	Hypothetical protein	MKLPATRTRLAVGAATAALAATAF**LAGCSDSGAEDEPTAVETTTESSTEDEAASGQVEVIDQTHLLLATAQDTLEAR**
Celf_3689	Peptidylprolyl isomerase FKBP-type	MRRTTTARAIAAATTALVLSLS**LAAC**SGDGGSDDSPSADASASVSAAQAAD//
Celf_0189	Penicillin-binding protein transpeptidase	MTLDSARDAARRDAADPTTPRPVGGSGTWARVATSAVVLALVGAG**TAAC**//LVDVPAEESAETTATPSADATARPQVPLTTEEAQALR

[a] The putative lipoprotein site is in bold and underlined. [b] The peptide sequence in blue is the one identified in the MS spectra as carrying the glycan. [c] The // indicates a break in the sequence. Names of genes selected for recombinant expression in *C. glutamicum* shown in bold.

Table 2. 3. Primers used to amplify genes from *C. fimi* genome.

Gene	*5'*	*3'*
celf_0189	5'-ATT AGC ATA TGA CGC TCG ACA GCG CGC GCG A-3'	5'-ATT AGG AAT TCT TAA TGG TGA TGG TGA TGG TGG CCC GCG GCG TAG CGG AGG AAG TC-3'
celf_3184	5'-GGG GTA TTC CAT ATG TCC ACC CGC AGA ACC GCC GCA GCG-3'	5'-GGG GAA TTC TCA CCA CCT GGC GTT GCG CGC CAT C-3'

2. 5. Discussion

2. 5. 1. Validation of *C. glutamicum* as a mannosylation host

Our previous proteomic analysis of *C. fimi* (111) revealed several mannoproteins contained within the cell-associated fraction. Here, we report on the identification of 19 cell-associated glycoproteins enriched from *C. fimi* and confirmed as being modified with *O*-linked hexoses by IP-NPLC enrichment and bottom-up NanoLC-MS/MS analysis. Interestingly, many of them are putative lipoproteins (Table 2.2) with orthologous to those previously identified in *M. tuberculosis* (93), where mannosylation leads to increased virulence (49). In addition, *C. fimi* is not the only organism sharing homologous mannoproteins with *M. tuberculosis* (98). Investigation of these mannosylated cell associated proteins and lipoproteins in a mannosylation capable host may allow narrowing down of a more defined cellular purpose to the *O*-linked glycosylation of these proteins. While we did not definitively demonstrate that the hexose is D-mannose, previous literature has shown the sugars to be mannobiose/mannotriose on glycoproteins in *Mycobacterium* (47). The presence of conserved mannosylated orthologues involved in various cellular metabolic pathways makes it intriguing to decipher what glycosylation brings to protein function. The link to cell wall integrity and mannosylated penicillin binding proteins in *Streptomyces*, suggests that this may be common in several actinobacterial species. We are following that up through investigating the presence of other glycosylated penicillin binding proteins in *Cellulomonas* and *Corynebacterium*.

We are assessing a system to investigate mannoprotein orthologues to probe the function of this post-translational modification. We started by looking at three proteins from the soil actinobacterium *C. fimi* ATCC484, a well-known plant cell wall degrading organism. Our target *C. fimi* proteins were produced recombinantly in *C. glutamicum* and shown to be substrates for

the host protein mannosyltransferase, producing native-like mannosylation. Mannosylation of our target proteins was confirmed via ConA lectin blotting, intact/peptide MS and HPLC of liberated *O*-glycans. While we did not identify the modified glycopeptide of natively produced Celf_3184 in this work, the identified peptide in the recombinant protein from *C. glutamicum* (Figure S2.4A) matches well with literature estimates for the level of glycosylation on this protein (114). The modified glycopeptides and hexose distribution of Celf_2022 (Figure S2.4B) and Celf_0189 (Figure S2.6B) produced in *C. glutamicum* closely match those seen in native proteins from *C. fimi* (Figure S2.1, panels K and B). HPLC of the liberated mannoglycans from all three recombinantly produced proteins shows that *C. glutamicum* is capable of mannosylating *C. fimi* proteins to a similar density with mostly disaccharides based on our HPLC data (Figure 2.6). Mannose oligosaccharides were clearly separated from glucose oligosaccharides, and mannosidase treatment of the liberated glycans suggests their composition to be entirely mannose. However, it must be noted that as C2 epimers of each other D-glucose and D-mannose are difficult to differentiate with this methodology. There are other protein mannosyltransferases from the GT105 family, distributed in both eukaryotes and prokaryotes, however actinobacteria do not appear to harbour the required gene, again supporting our assertion that the GT39 orthologue is the initiating transferase.

As *C. glutamicum* does not contain any annotated protein-*O*-glucosyltransferases (GT90) within its genome, we are confident that the initial hexose added is mannose. Lastly, these experiments show that the *C. glutamicum* PMT is capable of recognizing *O*-mannosylation substrates from closely related organisms and performs the transfer of mannose to proteins, with subsequent elongation likely performed by one of the 7 GT4 members annotated in the *C.*

glutamicum genome – a family known to have proteins with α-1,2-, α-1,3-, and/or α-1,6-mannosyltransferase activity.

2. 5. 2. Localization and activity of actinobacterial proteins produced in *C. glutamicum*

Of the three *C. fimi* proteins expressed in *C. glutamicum*, Celf_3184 and Celf_0189 were both localized as they are in the native organism (extracellularly and membrane-bound, respectively). Recombinantly produced Celf_3184 was secreted into the culture supernatant as predicted by its TAT leader sequence and as it was originally found in its native host (96), while Celf_0189 with its (putative) lipoprotein leader was detected predominately in the membrane fraction as predicted. Celf_2022 was originally detected as a protein in the secretome during our initial proteomics analysis of *C. fimi* (111) and when produced recombinantly in *C. glutamicum*, however the SignalP5.0 algorithm predicted this protein to be cytoplasmic as it had neither a strong Sec nor TAT leader sequence. The size discrepancy seen with recombinantly produced Celf_2022 is likely due to the generation of at least 2 major species of truncated products formed in both *E. coli* and *C. glutamicum* in addition to the differential mannosylation of these truncations in *C. glutamicum*. At present we do not know which export system is used in *C. glutamicum* to export the Celf_2022 protein.

C. glutamicum could produce enzymatically active Celf_3184 with similar cellulase activity to the enzyme produced in *E. coli* (unpublished data). The glycoside hydrolase activity of Celf_3184 produced in both *C. glutamicum* and *E. coli* showed no difference on the preferred substrate barley-β-glucan, but the mannosylated product is slightly more thermostable than its unmodified counterpart (unpublished data). We were unable to demonstrate a functional assay for assessing the putative prolyl-isomerase enzymatic activities of Celf_2022, and there is not a

strong orthologue in *C. glutamicum* which will limit our ability to determine the function of the glycan on this protein.

Celf_0189 is annotated as a penicillin binding protein, and indeed reacts with BOCILLIN-FL when produced in either *E. coli* or *C. glutamicum* (unpublished data), so the glycosylation is not required for this kind of functional assay. However, it is possible that the *in vivo* activity may use this modification in interactions with other proteins involved in peptidoglycan remodelling. We did not verify if the protein is a lipoprotein, as the yield and solubility of this protein made that analysis more difficult. We are looking for glycosylated orthologues in *Corynebacterium* as well as other glycosylated penicillin binding proteins in *Cellulomonas* so that we can design assays to determine what role the glycan plays in the function of these critical cellular proteins. The ability of *C. glutamicum* to achieve native-like mannosylation and export of glycoproteins originating from *Cellulomonas* suggests that its use would aid in a more detailed study of the process when a suitable biological assay can be established for the heterologous proteins being examined.

2. 5. 3. Conclusion

Our comparison of select mannoproteins produced in *C. fimi* and *C. glutamicum* has shown that the latter is a suitable host for studying mannosylation for these heterologous proteins. However, we realize that a subtle effect of glycosylation may be difficult to identify with these orthologous proteins. To address this deficiency, we plan to generate a ΔPMT strain of *C. glutamicum* that can be complemented with PMTs from other organisms, starting with *C. fimi*. This approach will allow us to produce mannoproteins which are mannosylated by their cognate PMT enzyme in *C. glutamicum*. Because of the low mannoprotein background in *C. glutamicum*,

a simple readout of the substrate specificities of the various PMTs may be possible and will give us insight on the beginning of this modification process, like a minimum sequence requirement for modification. We are also currently looking for *Corynebacterium* orthologues of the *C. fimi* glycoproteins that may be more suitable for this kind of study. In addition, we will follow up the investigation of the role of export in mannosylation. As we and others (98) have shown, there are TAT secreted proteins and possibly cytoplasmic proteins in these actinobacteria that are also mannosylated outside of the suggested SEC dependent secretion described for *Mycobacteria* (92). Increasing the number of mannoproteins and bacterial PMTs that can be investigated in the *Corynebacterium* host will develop a better understanding of the protein substrates and identify routes to dissect the function protein mannosylation plays in actinobacteria.

2. 6. Materials and Methods

2. 6. 1. Enrichment of CMC grown *C. fimi* mannosylated proteins

C. fimi ATCC 484 was grown for 48 h at 30°C and 180 RPM in LB broth supplemented with 0.5% CMC. Cells were harvested by centrifugation at 5,000 x *g* for 30 mins at 4°C and frozen at -20°C. Cells (2.4 g wet weight) were resuspended in 20 ml of ConA buffer A (20 mM Tris, 0.5 M NaCl, 1.0 mM $CaCl_2$ and 1.0 mM $MnCl_2$, pH 7.4) with Sigma P-2714 protease inhibitor cocktail. The cells were lysed using the Emulsiflex-C5 (Avestin) at ≥20 000 psi. After centrifugation at 6,000 x *g* for 5 min and 20,000 x *g* for 20 min (4°C), the final clear, yellow supernatant was used to load the ConA Sepharose column (Cytiva). ConA enrichment was carried out at 4°C.

The 20,000 x *g* pellet fraction was resuspended in 10 mL of ConA buffer A with 0.5% Triton X-100 and incubated at room temperature, using a tube roller, for 1 hour. It was then

centrifuged at 100,000 x *g* for 60 min at 4°C. The 100,000 x *g* supernatant was used for a second ConA Sepharose enrichment and the 100,000 x *g* pellet was resuspended in 10 mL ConA buffer A and kept for analysis.

In both runs, the column was equilibrated in ConA buffer A and washed, after sample loading, with 5 column volumes (CV) of the same buffer. Mannosylated proteins were eluted over a 5 CV linear gradient of 0% – 50% 0.3 M α-methylglucoside in ConA buffer A, while collecting 2.0-mL fractions. The yield of eluted proteins was increased by the inclusion of 0.2% Triton X-100 in the elution buffer.

2. 6. 2. Western and lectin blot protocol

Lectin blots were performed as described in the manufacturers (Sigma) data sheets. Briefly, the proteins of interest were separated on a 12% SDS PAGE gel using a MiniProtean system (Bio-Rad), which was then rinsed three times for 5 minutes, each, in excess Tris buffered saline pH 7.6 (TBS, 50 mM Tris, 150 mM NaCl, pH 7.6) before blotting to PVDF membrane using a Trans-Blot Turbo transfer system (Bio-Rad). The transfer was performed in 48 mM Tris, 39 mM glycine using 2.5 A, 25 V for 8 mins. The protein-bearing PVDF membrane was then rinsed three times for 5 minutes each, in TBS and blocked for one hour in 5% BSA in TBS at room temperature (22°C). The blocked membrane was washed three times for 5 minutes in TBS at room temperature, then incubated overnight at 4ºC in 0.05% Tween 20, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM $MnCl_2$; 0.5 µg/mL ConA-FITC or Rhodamine conjugated lectin (Millipore-Sigma); and 1:10,000 AlexaFluor 647 anti-HIS_6 (Bio-Rad) in TBS. The membrane then underwent three 10-minute washes in TBS pH 7.6 at room temperature and was visualized on a Bio-Rad ChemiDoc. Molecular weight markers were Bio-Rad All Blue standards.

2. 6. 3. BLAST Search Protocol

To identify orthologues of *C. fimi* glycoproteins, BLAST searches were performed using the BLASTp algorithm in the NCBI non-redundant protein sequences database. Comparisons were conducted with actinobacterial species *C. flavigena* (NCBI taxid:1711), *C. glutamicum* (NCBI taxid:1718), *S. coelicolor (*NCBI txid:100226) and *M. tuberculosis* (NCBI taxid:1773). Matches above an E value of 10^{-25} were omitted.

2. 6. 4. Cloning and recombinant production of Celf_2022, Celf_3184, and Celf_0189

The genes *celf_0189* and *celf_3184* were amplified from extracted genomic DNA using specific primers (Table 3), while *celf_2022* was ordered as a synthetic gene (Figure S2.7) from BioBasic (Markham, Ontario, Canada) to remove incompatible internal restriction sites. All genes were restriction cloned into the MCS of an *E. coli*/*C. glutamicum* shuttle vector pTGR-5 (115) using NdeI and EcoRI (New England Biolabs) and transformed into electrocompetent NEB10β (New England Biolabs) *E. coli*. Each construct was verified via sequencing prior to transformation into respective bacterial strains for expression.

2. 6. 5. Expression and purification

Constructs with a plasmid borne kanamycin resistance gene for selection in *E. coli* expression strains (BL21 for Celf_0189 and Celf_2022; SHuffle Express T7 for Celf_3184) were grown overnight in 2YT (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl) containing 50 μg/mL kanamycin at 37°C (BL21) or 30°C (SHuffle Express T7) at 180 RPM. Expression cultures (2YT containing 50 μg/mL kanamycin) were inoculated with the overnight culture to an OD_{600} of ≈ 0.1 then incubated at 37°C (BL21) or 30°C (SHuffle Express T7) at 180 RPM. When the expression

cultures reached an $OD_{600} \approx 0.5– .6$ they were induced with 0.5 mM IPTG and incubated overnight at 180 RPM and at 20°C, 25°C, and 30°C for Celf_3184, Celf_2022, and Celf_0189 respectively. Cells were harvested by centrifugation at 5,000 x *g* for 30 minutes at 4°C and frozen at -20°C.

C. glutamicum containing expression constructs were grown overnight at 30°C and 200 RPM in brain-heart infusion (BHI) media containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid, then induced for 24 (Celf_0189) or 48 hours (Celf_2022 and Celf_3184) under the same conditions using 0.5 mM IPTG. Cells were harvested by centrifugation at 5,000 x *g* for 30 minutes at 4°C and frozen at -20°C.

For Celf_2022 and Celf_3184 (which possess Sec and TAT leaders, respectively), the spent medium was brought to a final concentration of 100 mM HEPES/Na+ (pH 8.0), 0.3 M NaCl using a 10x concentrated stock solution and immediately loaded onto a 5-mL IMAC column. For Celf_0189, cell pellets were lysed mechanically using a mortar and pestle, and membrane fractions were harvested via ultracentrifugation at 100,000 x *g* for 1 hour at 4°C. The resulting pellet was solubilized in 50 mM sodium phosphate (pH 7.8), 50 mM NaCl, 10% glycerol, and 1% dodecylmaltoside (DDM) overnight at room temperature on a shaking table.

Recombinantly expressed proteins from *C. fimi* were recovered via IMAC using cOmplete His-Tag purification resin (Millipore-Sigma). Samples were loaded at a flow rate of 2 mL/min using 100 mM HEPES/Na+ (pH 8.0), 0.3 M NaCl and the column was washed with 5 CV of the same buffer. His-tagged proteins were eluted over 5 CV using a gradient of 0 to 100%, 0.5 M imidazole in 100 mM HEPES/Na+ (pH 8.0), 0.3 M NaCl.

2. 6. 6. In-gel tryptic digestion and Nano LC-MS/MS analysis of selected protein bands from ConA fractions

A second SDS-PAGE gel was prepared using the same loading and separation conditions used for the lectin blot analysis described above. The gel was stained with Colloidal Coomassie Blue and bands corresponding to those ConA reactive bands in the lectin blot were cut out and destained by immersion in 100 mM ammonium bicarbonate (ABC), 30% acetonitrile (ACN). The destaining process was repeated until all the stain was removed. The gel bands were then subjected to in-gel tryptic digestion and nanoLC-MS/MS using established protocols (116). The gel bands were shrunk with ACN, reswollen in 10 mM dithiothreitol in 50 mM ABC and incubated at 54°C for 1 hour. The excess liquid was removed, the gel pieces were covered with 55 mM iodoacetamide in 50 mM ABC and incubated in the dark at room temperature for 1 hour. The excess liquid was removed, and the gel pieces were shrunk with ACN and rehydrated in 50 mM ABC. This process was repeated once more to remove all excess reagent for the gel pieces. The gel pieces were shrunk and finally incubated in 50 mM ABC containing 20 ng/µL of sequencing grade trypsin (Sigma-Aldrich). Sufficient trypsin solution was added to just cover the gel pieces after rehydration. The samples were incubated overnight at 37°C. The excess liquid was then transferred to new Eppendorf tubes and stored at 4°C before analysis. Nano LC-MS/MS analysis was performed using a capLC capillary chromatography system (Waters, Milford, MA) coupled with a Q-TOF Ultima hybrid quadrupole time-of-flight mass spectrometer (Waters). Approximately one fifth of each digest was injected onto a 5 mm x 300 µm i.d. Acclaim PepMap100 C18 µ-precolumn (Dionex/Thermo Scientific, Sunnyvale, CA) and resolved on a 100 µm x 100 mm i.d. 1.7 µm BEH130 C18 column (Waters) using the following gradient conditions: 1% to 45% acetonitrile (ACN) in 0.1% formic acid in 19 min and 45% to

85% ACN in 2 min. The flow rate was 400 nL/min. MS/MS spectra were acquired in Data Dependant Acquisition (DDA) mode on doubly, triply and quadruply charged ions. The resulting MS/MS spectra were searched against a customized *C. fimi* protein sequence database (*C. fimi* ATCC484) using the Mascot™ search engine (Matrix Science Ltd., London, UK). The search parameters were as follows: enzyme = trypsin, number of missed cleavages = 1, peptide mass tolerance = 0.5 Da, fragment mass tolerance = 0.5 Da, fixed medication = carbamidomethyl (cysteine), variable modifications = oxidation (methionine) and deamidation (asparagine, glutamine). Peptides hits with scores less than 20 were discarded and those with scores between 20 and 40 were confirmed manually.

2. 6. 7. Glycopeptide enrichment and Nano LC-MS analysis

Selected fractions from the ConA column were digested with trypsin. A volume of each fraction equivalent to 50 µg of protein was concentrated on a vacuum concentrator (Speedvac) to approximately 120 µL and brought to 200 µL with 1M Tris-HCl, 6 M guanidine, 1 M Tris HCl, pH 7.5. 20 µL of 100 mM DTT was added to each sample and incubated at 56°C for 15 minutes. 22 µL of 250 mM iodoacetamide in 50 mM ABC was added to each sample and incubated at RT in the dark for 20 minutes. The non-detergent samples were buffer exchanged into approximately 50 µL of 50 mM ABC by centrifugal filtration using 3 kDa MWCO filter (Amicon). The detergent-containing fractions were cleaned up using the Detergent Out kit which removes non-ionic detergents (Millipore) and exchanged into 50 mM ABC. Each sample was brought to approximately 100 µL with 50 mM ABC, 5% acetonitrile. Trypsin (5 µg) was added, and the samples were incubated at 37°C overnight.

Glycopeptides from the tryptic digests were enriched by Ion-Pairing Normal Phase Liquid Chromatography (IP-NPLC) using conditions similar to those described previously (112, 117). Approximately half of each digest was evaporated to near dryness (approximately 2 µL remaining) on a vacuum concentrator (Speedvac) and reconstituted to 85% acetonitrile, 1.0% TFA. 8 µL of each sample were injected on a polyhydroxyethyl A™ Javelin® guard column (1 cm x 1 mm i.d., 5 µm, NEST group, MA) on a MDLC chromatography system (GE Healthcare). The guard column had been equilibrated with 85% acetonitrile (ACN). The following mobile phase gradient was used to elute the peptides and glycopeptides from the column: time 0, the gradient goes from 85% to 70% ACN in 10 minutes followed by a further decrease in CAN from 70% to 50% ACN in 15 minutes. The flow rate was 12 µL/min. The column was equilibrated in 85% ACN for 4 minutes prior to the next injection. The column eluate was collected in 4 fractions: (1) 0-5 minutes, (2) 5-8 minutes, (3) 8-14 minutes and (4) 14-27 minutes.

The fractions were evaporated to approximately 12 % of their starting volume to remove ACN and reconstituted to 50 µL with 0.1% formic acid. 10 µL of each fraction was analyzed by Nano LC-MS/MS on the Q-TOF Ultima using the same column setup as described above but using the following gradient conditions: 1% to 45% ACN in 0.1% formic acid in 37 min and 45% to 95% ACN in 2 min. The flow rate was 400 nL/min. MS/MS spectra were acquired in Data Dependant Acquisition (DDA) mode on doubly, triply and quadruply charged ions. Most glycopeptides eluted in fraction 4 though a handful were observed in fraction 3. Most of the non-glycosylated peptides eluted in fractions 1 and 2. Glycopeptide MS/MS spectra were identified manually. A partial amino acid sequence was derived from the glycopeptide MS/MS spectrum which was matched against the *C. fimi* protein sequences in the NCBInr database using blastp

(https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2. 6. 8. Determining the sites of glycosylation using Electron-Transfer Dissociation (ETD)

Selected *O*-linked glycopeptides observed in the IP-NPLC fractions were analyzed by targeted Nano LC-ETD-MS/MS on a LTQ XL linear ion trap mass spectrometer (Thermo Fischer Scientific) equipped to perform ETD (118–120). Fluoranthene was used as the anionic reagent and optimal reaction times varied from 400-800 milliseconds depending on the glycopeptide. Supplemental activation was applied when necessary.

2. 6. 9. Intact mass LC-MS analysis of rCelf-2022 and rCelf-3184

Intact mass analysis was performed using an 1100 chromatography system (Agilent) linked to an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific). 5 µg of each protein was injected on to a 2.1 x 30 mm Poros R2 column (Thermo Fisher Scientific) and resolved using the following rapid gradient: hold at 20% mobile phase A for 3 minutes, 20% - 90% mobile phase B in 3 minutes, hold at 90% mobile phase B for 1 minute. Mobile phase A was 0.1% formic acid in ddH_2O and mobile phase B was acetonitrile. The flow rate was 3 mL/min with 100 µL split to the electrospray ion source. Optimal peak shape was achieved by heating the column and mobile phase to 80^0C. The mass spectrometer was tuned for small protein analysis using myoglobin and the resolution was set to 15,000. Mass spectra were acquired from m/z 400 to 2,000 in the orbitrap at 1 scan per second. The spectra acquired across the protein peak were summed and deconvoluted using MaxEnt 1 (Waters).

2. 6. 10. Analysis of *O*-glycosylation on the *C. fimi* proteins expressed in *C. glutamicum*

Celf_2022 (50 µg) was incubated overnight with trypsin (2.5 µg) at 37°C. Celf_0189 (100 µg, 0.5 µg/µL) and Celf_3184 (50 ug, 0.5 µg/µL) were reduced by heating to 80^0C for 10 minutes with 10 mM dithiothreitol in 100 mM ammonium bicarbonate, followed by alkylation with 37.5 mM iodoacetamide in the dark at RT for 10 minutes. Celf_0189 was then incubated overnight with endoproteinase Glu-C (6.5 µg, Promega). Celf_3184 was digested overnight with chymotrypsin (3 µg) at room temperature. Approximately 2.5 µg of the resulting digests were enriched on a polyhydroxyethyl A spin trap using the following protocol. 5 mg of polyhydroxyethyl A beads (10 µm, 100 Å, PolyLC inc.) was packed into an empty 10-200 µL TopTip disposable pipette tip by centrifugation at 2,000 g for 3 minutes. The tip was washed with 50 µL of H_2O and equilibrated with 50 µL of 80% ACN before use. Protein digest equivalent to approximately 2.5 µg was evaporated to dryness on a Speedvac, redissolved in 50 µL of 80% ACN, 0.1% TFA (loading buffer), loaded in the packed tip and centrifuged at 2,000 g for 3 minutes. The tip was washed 3 times with 50 µL of loading buffer with centrifugation after each application. Finally, the glycopeptides were eluted with 50 µL of ddH_2O, followed by 50 µL of 30% ACN, 0.1% TFA.

The eluates were combined, evaporated to dryness, and redissolved in 20 µL of 0.1% formic acid in ddH_2O. The enriched glycopeptides were analyzed by Nano LC-MS/MS on a NanoAcquity HPLC system (Waters) coupled to an LTQ-Orbitrap XL hybrid mass spectrometer. A quarter of the reconstituted eluate (5 µL) was loaded on a Nano-Acquity Symmetry C18 trap (5µm, 180 µm x 2cm) and washed with 1% mobile phase B at 15 µL/min for 5 minutes. The trap was then switched on-line with a Nano-Acquity BEH C18 nanoflow column (1.7µm, 100µm x 100 mm) and the peptides were separated using the following gradient: 1% - 45% mobile phase

B in 37 minutes, 45% - 95% mobile phase B in 2 minutes. Mobile phase A was 0.1% formic acid in ddH_2O and mobile phase B is ACN, 0.1 % formic acid. The flow rate was 400 µL/min. The instrument was tuned for peptide analysis and the resolution with 30,000. Mass spectra were acquired from m/z 400 to 2,000 in the Orbitrap at 1 scan per second. The instrument was set in data dependent acquisition mode to acquire CAD-MS/MS spectra on multiply charged ions in the ion trap. Glycopeptide MS and MS/MS spectra were examined manually.

2. 6. 11. HILIC-HPLC analysis of *O*-glycans from recombinantly produced Celf_2022 and Celf_3184

The *O*-glycans of recombinantly produced Celf_2022 and Celf_3184 were released and labelled with 2-aminobenzamide (2AB; Ex: 330 nm, Em: 420 nm) using the EZGlyco *O*-Glycan prep kit (Sbio) following the manufacturer's recommendations. The labelled *O*-glycans from 100 µg of recombinant proteins were then analyzed by HILIC on a HPLC system with fluorescent detection (Shimadzu) and a 3µm bead 3 mm x 150 mm TSKgel Amide-80 (Tosoh) using 100% acetonitrile (mobile phase A) and 50 mM ammonium formate pH 4.4 (mobile phase B) with the column temperature set to 45°C. The labelled *O*-glycans were injected at 250 µL/min 10% mobile phase B, which was maintained for 10 minutes. The *O*-glycans were then eluted using a gradient of 10% to 50% mobile phase B over 15 minutes and held at 50% mobile phase B for an additional 5 minutes. The column was then re-equilibrated with 10% mobile phase B for 10 minutes prior to subsequent runs. Mannobiose (α1,2, and α1,3-linked) and α1,2 linked mannotriose standards were obtained from Carbosynth and labelled in the same manner as the liberated glycans. Samples were treated with α-mannosidase (NEB Canada) after labelling and

re-run to examine the presence of mannose in the liberated glycans.

2. 6. 12. Electrocompetent *C. glutamicum* and transformation

C. glutamicum ATCC 13032 was inoculated into 2YT containing 25 µg/mL nalidixic acid and incubated at 30°C overnight at 180 RPM. The next day, 200 mL of 2YT containing 3.5% (w/v%) glycine and 25 µg/mL nalidixic acid was inoculated to an OD_{600} of ≈ 0.1 using the overnight culture and incubated at 30°C and 180 RPM. When the OD_{600} had reached ≈ 0.2 – 0.25, ampicillin was added to a final concentration of 0.5 µg/mL and the culture was incubated under the same conditions for an additional 1.5 hours. Cells were harvested by centrifugation at 5,000 x *g* for 10 minutes at 4°C. Cell pellets were washed in ice cold 20 mM HEPES/Na+, 5% glycerol, pH 7.2 and centrifuged as before three times. The final cell pellet was resuspended in 1.5 mL ice cold 5 mM HEPES/Na+, 15% glycerol, pH 7.2 and 100 µL aliquots were stored at -80°C.

Aliquots of electrocompetent *C. glutamicum* cells were electroporated at 2.5 kV using a MicroPulser (BioRad), then incubated in 1 mL 2YT at 30°C and 180 RPM for 2 – 4 hours for recovery. These cultures were centrifuged at 5,000 x *g* for 5 minutes and cell pellets were resuspended in ≈ 200 µL fresh 2YT. Resuspended cells were plated onto agar media containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid and incubated at 30°C for 48 – 72 hours, or until colonies were visible.

2. 7. Acknowledgements

This work was partially funded by a Natural Sciences and Engineering Research Council grant to WW.

3. PROTEIN-O-MANNOSYLATION BY NON-SEC/TAT SECRETION TRANSLOCONS IN ACTINOBACTERIA

as submitted to Applied and Environmental Microbiology
August 2023
by Saxena, H., Patel, R., Kelly, J., and Wakarchuk, W.

3. 1. Chapter Overview

This chapter describes the knockout of the GT-39 enzyme in *C. glutamicum* responsible for all cellular POM, in addition to the attempted characterization of the mutant and the identification of a discernable phenotype. As *C. glutamicum* is a widely utilized, industrially relevant microorganism, tools for the required genetic manipulations were adapted to study the process of POM in the organism. While recombinant protein production of actinobacterial GT-39s (the enzymes responsible for the initiation of POM) did not result in isolated enzyme for biochemical characterization, *in vivo* assays of these GT-39s did reveal surprising differences in activity and specificity. Most notably, evidence was obtained that another mannosylation capable translocon exists in *C. glutamicum* that functions adjacent to the traditional SEC and TAT translocons.

3. 2. Abstract

Protein-*O*-mannosylation (POM) is a form of *O*-glycosylation that is ubiquitous throughout all domains of life and has been extensively characterized in eukaryotic systems. However, in prokaryotes this process has only been investigated in relation to pathogenicity (in *M. tuberculosis*) even though there are many non-pathogenic bacteria that are known to regularly carry out POM. To date, there is no consensus on what benefit POM imparts to the non-

pathogenic bacteria that can perform it. Though the generation of a POM deficient mutant of *C. glutamicum* – a widely utilized and known mannosylating actinobacteria – this work shows that even closely related actinobacterial GT-39s can potentially have different activities and substrate specificities for targets of POM. Moreover, presented here is evidence that POM does not only occur in a SEC-dependent manner; POM also occurs with TAT and non-SEC secreted substrates in a specific and likely tightly regulated manner. Together these results highlight the need for further biochemical characterization of POM in these and other bacterial species to help elucidate the true nature of its biological functions.

3. 3. Introduction

3. 3. 1. Actinobacterial protein *O*-mannosylation

POM is an essential and ubiquitous posttranslational modification found throughout all domains of life (22). In the half century since this modification was first described in actinobacteria (90) it has been demonstrated that POM plays a critical role in the virulence of *M. tuberculosis* (49), but no definitive biological context has been attributed to it in the non-pathogenic members of the phylum highlighting a significant lack in the understanding of prokaryotic POM. Most of the information on POM and the enzymes responsible for it – protein-*O*-mannosyltransferases (PMTs) – has been garnered from work done in *S. cerevisiae*, which has contributed significantly to the understanding of this process in eukaryotes. A considerable proportion of the mannosylated eukaryotic proteins identified to date are either secreted or cell membrane associated proteins with well defined functions (14, 22, 52, 121). Conversely, only a few of the mannosylated proteins identified in well-known actinobacteria like *M. tuberculosis* (91–93), *C. glutamicum* (56, 94), *C. fimi* (95, 96), and *S. coelicolor* (98, 122) have had functions

attributed to them, complicating the elucidation of the biological context of this modification. This is despite there being over 3,000 bacterial species with an annotated PMTs – belonging to glycosyltransferase family 39 (GT-39) in the Carbohydrate Active Enzyme (CAZy) database (123).

In eukaryotes, POM is initiated in the lumen of the endoplasmic reticulum (ER) as a protein is being translocated across the membrane in an unfolded SEC-dependent manner. In prokaryotes, POM is thought to occur similarly during extracellular translocation across the plasma membrane; however, many TAT-exported proteins, which are folded prior to their export (98, 124, 125) , have also been identified to be *O*-mannosylated. Recently, numerous mannosylated cytosolic and non-SEC translocon secreted proteins have also been identified in *C. fimi* and *C. flavigena* (55, 126). This contradiction highlights the fact that there is still a significant lack of information on both the mechanisms involved in bacterial protein-*O*-mannosylation and its overall function.

3. 3. 2. POM requirements

Four mandatory components are necessary for any cell to successfully carry out POM: an activated mannose residue, a mannose-carrying lipid donor, a PMT (GT-39) enzyme, and the target protein to be modified containing the required Ser or Thr residue(s). Though there exists some variation in these components depending on their parent organism (especially between prokaryotes and eukaryotes) the general steps of protein-*O*-mannosylation are considered to be the same throughout all domains of life (52, 127). There are four generalized steps to protein-O-mannosylation: firstly, the mannosylated lipid donor is synthesized through the action of a GT-2 family glycosyltransferase (known as Ppm1 in actinobacteria) that transfers mannose from the

activated donor GDP-α-D-mannose to the phosphorylated prenyl or dolichol donor lipid. Next, the phosphomannose lipid is flipped across the membrane by a still unknown protein, where it can interact with the PMT. The mannose residue is then transferred from the phospholipid carrier to the PMT enzyme, which finally transfers the mannose through an inverting mechanism to the hydroxyl group of Ser or Thr residues on the target protein as it is translocated across a biological membrane (88, 128–132). The most evident difference in protein *O*-mannosylation between eukaryotes and prokaryotes is the cellular location of the modification reaction. As prokaryotes lack the cellular compartments of eukaryotes, protein *O*-mannosylation is thought to occur in the periplasm or on the extracellular face of the plasma membrane (depending on Gram status of the organism) instead of the luminal face of the ER (8, 22, 133).The mono-mannosylated glycoprotein will then undergo further modification by a number of distinct enzymes to complete the glycan chain, producing the final glycoprotein (134). While the enzymes responsible for the subsequent elongation of mannoglycans in actinobacteria is not currently known, it is likely performed by GT-4 family members recognized to have α-1,2-, α-1,3-, and/or α-1,6-mannosyltransferase activity (7 of which are annotated in the *C. glutamicum* ATCC 13032 genome).

3. 3. 3. Glycosyltransferase family 39 (GT-39)

All PMT enzymes identified to date are integral membrane proteins containing several transmembrane domains and contain enough sequence homology to be grouped into GT-39, highlighting the conservation of *O*-mannosylation (123). In *S. cerevisiae,* there are several PMT orthologues belonging to three PMT subfamilies, POMT1, POMT2, and POMT4 (135, 136) which were previously thought to be a form of redundancy; however, it is now known that

enzymatic activity requires members of POMT1 and POMT2 to form heterodimers while POMT4 members form homodimers (137, 138). Recently, the first cryo-electron microscopy structure of a GT-39 PMT1-PMT2 heterodimer from *S. cerevisiae* was solved (PDB: 6P2R), while only topology reports and hydropathy profiles of several bacterial PMTs have been published (108, 127, 139).

The related architecture of the *S. cerevisiae* GT-39 to GT-66 oligosaccharyltransferases from both prokaryotes (but not including actinobacteria as there are no annotated GT-66 enzymes in these organisms) and other eukaryotes implies that PMTs across the domains of life maintain at least three characteristics: a cytosolic N-terminal region and C-terminal region on the opposing side of the membrane (endoplasmic reticulum or plasma membrane), multiple hydrophilic loops on either side of the membrane, with the first luminal/periplasmic loop containing at least a part of the PMT catalytic site while carrying the conserved neighbouring residues DE (140–142) aligning with D^{55} and E^{56} in the yeast PMT1. While these two conserved residues are not believed to be a part of the PMT active site – as exchange of either does not completely abolish PMT activity – they may be involved in the stabilization of the active site by the coordinating and positioning of the required divalent cations (143). As many studies have focused on characterizing these enzymes in eukaryotic cells, there is precious little first-hand information about the bacterial members of this family (8, 22, 144).

Across species, GT-39s can show a marked degree of variation. The only prokaryotic GT-39s investigated to date originate from *M. tuberculosis* and *C. glutamicum* and show only 25.7% and 26% global amino acid sequence similarity, respectively, to the PMT1 of *S. cerevisiae*. Interestingly, these two prokaryotic PMTs share a 55.9% similarity (Table S1), which could be partially explained by their close taxonomic relation. Genomic analyses have revealed other

putative bacterial GT-39s in other actinobacteria, like *C. fimi* and *C. flavigena* (25.4% and 23.2% similarity to *S. cerevisiae* PMT1, 46% and 46.9% similarity to the *M. tuberculosis* GT-39, and 42% and 40.5% similarity to the *C. glutamicum* GT-39, respectively), including other prokaryotes.

3. 3. 4. Lipid donor and glycosyltransferase family 2 (GT-2)

Glycolipid intermediates provide activated mannose to PMTs for POM. There is one specific mannosyl donor in all eukaryotes, and one specific mannosyl donor in all *O*-mannosylating bacteria. These donors are dolichol phosphate β-D-mannose (Dol-P-Man) and polyprenyl monophosphomannose (PPM), respectively (132, 144, 145). The interaction of PMTs with the mannosyl donor depends on the presence of an activated mannose moiety anchored to the periplasmic side of the membrane (146, 147) in prokaryotes, or luminal side of the ER for eukaryotes (131, 148).

Prior to the POM reaction, a separate synthase enzyme which is anchored into the membrane charges these phosphorylated lipid donors with mannose using GDP-Man on their respective cytoplasmic extension (Maeda & Kinoshita, 2008). These enzymes are polyprenyl monophosphomannose synthases (Ppm1) and belong to the GT-2 family (123). The newly synthesized Dol-P-Man or PPM is then flipped to the opposing side of the membrane, by an unknown protein (52, 149–151) to interact with the PMT and other mannosyltransferase enzymes.

3. 3. 5. Transmembrane and tetratricopeptide repeat-containing (TMTC) mannosyltransferases

Recently a novel class of *O*-Man glycosyltransferases was identified, selectively serving to mono-mannosylate cadherins and protocadherins (152). The *O*-mannosylglyans on these proteins were reported to not be elongated, suggesting a novel type of *O*-mannosylation in higher eukaryotes (153–156). This enzyme family of transmembrane (TM) and tetratricopeptide (TPR) repeat-containing (TMTC) proteins is composed of four paralogues, TMTC1–4, with potentially differing roles (157, 158) and belonging to GT-105. While this family is distributed in both eukaryotes and (to a limited degree) prokaryotes, no putative GT-105s are currently annotated in actinobacteria suggesting that GT-39s are the sole initiating mannosyltransferase in the phylum.

3. 3. 6. Investigating actinobacterial POM in *C. glutamicum*

C. glutamicum is a Gram-positive non-pathogenic, non-sporulating, non-motile rod-shaped bacteria that has been widely utilized in industrial applications such as the production of amino acids, nucleotides, and enzymes (159–161). *C. glutamicum*, *M. tuberculosis*, *C. fimi*, and *C. flavigena* are all members of the *Actinomycetales* order and have genomes that are high in GC% content. The wide utilization of *C. glutamicum* means there exists an established molecular toolbox for genetic engineering of this organism. In addition, the close genetic resemblance of *C. glutamicum* to *C. fimi* and their common mannosylation machinery suggests that this organism serves as an ideal candidate for a Gram-positive protein expression/secretion system to investigate actinobacterial mannosylation. This was demonstrated recently, as the *C. glutamicum* platform was shown to accurately mannosylate and localize recombinantly produced heterologous proteins originating from *C. fimi* (126).

A GT-39 deficient mutant of *C. glutamicum* recombinantly producing selected actinobacterial PMTs will allow for the *in vivo* assay of these multipass transmembrane proteins, which are traditionally difficult to biochemically characterize and therefore lack reports in the literature. Using both the native *C. glutamicum* mannoproteome and a natively mannosylated actinobacterial mannoprotein – Celf-3184, expressed alongside each actinobacterial PMT – will show differences in substrate preference between the related actinobacterial GT-39s, allowing for the continued development of the *C. glutamicum* platform for the investigation of POM in actinobacteria. Most importantly, the impact of different secretion pathways on POM will be investigated using actinobacterial targets, some of which are known to be secreted mannoproteins.

3. 4. Results

3. 4. 1. Cg_1014 knockout and complementation

To adequately assess *in vivo* POM by each recombinantly expressed actinobacterial GT-39, the native *C. glutamicum* GT-39 was inactivated. Homologous recombination was used to knockout the PMT gene in *C. glutamicum* (Cg_1014), taking advantage of the wide host range of the pK18mobsacB (162) suicide vector. Predictive software tools revealed the possibility of regulatory transcriptional elements for neighbouring genes to be contained on the non-coding strand of Cg_1014, specifically, the regions that coded for the N- and C- termini of the PMT enzyme. For this reason, a truncated and inactive knockout construct was designed instead of a seamless knockout (Figure S3.1). This truncated construct only consisted of the cytoplasmic N-terminal region, the first transmembrane region, and the extracellular C-terminal region. As the active site and conserved residues D^{65} and E^{66} are contained in the first extracellular loop, it was

predicted that this construct would effectively abolish POM in *C. glutamicum*. Following homologous recombination, clones were screened and confirmed via colony PCR (Figure S3.2) for replacement of the native Cg_1014 gene by the truncated and inactive knockout. An amplicon of 2,539 bps indicated a successful knockout generating the ΔCg_1014 strain compared to the 3,880 bps amplicon in ATCC 13032 with intact GT-39. The final, positive clone was further confirmed for the loss of POM by ConA lectin blotting (Figure 3.2), with any residual ConA reactivity in the POM deficient strain being attributed to mannosylated lipids and/or lipoarabinomannan (LAM) enriched in the membrane fractions due to their resistance to proteinase K digestion (Figure S3.3).

POM in other actinobacteria has been reported to be non-essential and is loss to be lacking a discernable phenotype (49, 56). This was confirmed with the GT-39 deficient strain of *C. glutamicum* as there were no significant differences in growth between ATCC 13032 and the ΔCg_1014 mutant (Figure S3.4). As previous proteomic studies of actinobacteria have identified that many mannoproteins are either membrane-bound or membrane-associated (55, 163), the ΔCg_1014 mutant was further screened for a distinct phenotype with antibiotics targeting either membrane-bound/associated or intracellular components – therefore requiring active or passive transport through the membrane. While no obvious patterns were evident, some differences in antibiotic susceptibility between the two strains were observed (Figure 3.1). The ΔCg_1014 mutant strain was more susceptible to tetracycline (30 µg), chloramphenicol (30 µg), and tobramycin (10 µg), but less susceptible to novobiocin (30 µg), gentamicin (10 µg), and erythromycin (15 µg) compared to the ATCC 13032 strain. Complementation of the ΔCg_1014 strain with pCGE-31 harbouring the native Cg_1014 gene resulted in antibiotic susceptibilities like the ATCC 13032 strain, except for three of the tested antibiotics. Compared to the ATCC

13032 strain, the complemented mutant was still more susceptible to chloramphenicol (30 µg), tobramycin (10 µg), and erythromycin (15 µg).

The traditional biochemical characterization of multipass transmembrane proteins is notoriously difficult (164, 165). Detection and recovery of actinobacterial GT-39s produced recombinantly in both the ATCC 13032 and ΔCg_1014 strain was not possible (data not shown). For this reason, the *O*-mannosylation of the native *C. glutamicum* mannoproteome was initially used to assess the degree of POM in the complemented ΔCg_1014 strains. Only the ΔCg_1014 strain complemented with the *C. glutamicum* GT-39 showed notable reconstitution of mannosylation, with minimal complementation by the *C. fimi* GT-39, Celf_3080 (Figure 3.2). Lack of complementation in the ΔCg_1014 strain with site directed mutants (SDM) of the *C. glutamicum* GT-39 – mutating conserved residues $D^{65}N$ and $E^{66}Q$ – suggests that these residues are necessary for enzymatic activity, but not necessarily catalysis (Figure 3.2). Like the *S. cerevisiae* and *S. coelicolor* GT-39s, it is likely that these residues assist in the coordination of an active site divalent $Mn2^{+}$ cation – based on their structural similarity to the oligosaccharyltransferase from *Campylobacter lari*, PglB, where activity is metal dependent (108, 166, 167).

3. 4. 2. *In vivo O*-mannosylation using an actinobacterial target mannoprotein

As the native *C. glutamicum* mannoproteome proved to be a poor target for the *C. fimi* and *C. flavigena* GT-39s, a selected target actinobacterial mannoprotein – Celf_3184, known to be accurately mannosylated in the ATCC 13032 strain (126) – was expressed alongside each actinobacterial GT-39 in a synthetic *O*-mannosylation operon to better assay the *in vivo* mannosylation of these PMTs. Interestingly, differential mannosylation of Celf_3184 was

detected when co-expressed with the different actinobacterial GT-39 complementation constructs (Figure 3.3). Perhaps unsurprisingly, LC-MS results confirmed that the culture supernatant recovered Celf_3184 co-expressed with the *C. fimi* GT-39 in the ΔCg_1014 strain was more uniformly modified by hexoses (Figure S3.4A) than when produced in the ATCC 13032 strain, which has been previously reported. When produced in the ATCC 13032 strain (or the *C. glutamicum* GT-39 complementation construct) the range of modifications to Celf_3184 typically falls between 29 – 37 hexoses (126), but when co-expressed with the *C. fimi* GT-39, this range is much narrower at 31 – 35 hexoses (Figure 3.4A). Celf_3184 co-expressed with the *C. glutamicum* GT-39 in the ΔCg_1014 strain showed a similar degree of modification (data not shown) to Celf_3184 produced in the ATCC 13032 strain (126). Conversely, the *C. flavigena* GT-39 does not appear to accept any of the *C. glutamicum* native mannoproteome or Celf_3184 as POM targets which could suggest a very stringent substrate specificity for this GT-39.

3. 4. 3. Secretion and POM utilizing a translocon other than SEC and TAT

Originally, predictive tools were unable to identify an *N*-terminal secretion signal associated with Celf_2022 even though it was first identified as a mannoprotein lacking a predicted leader sequence (a predicted cyclophilin type peptidyl-prolyl cis-trans isomerase) in spent medium of *C. fimi* cultures during the proteomic analysis of the secretome of *C. fimi* (55). For this reason, Celf_2022 was initially chosen as a possibly cytoplasmic mannoprotein target for heterologous expression in *C. glutamicum*. The previous validation of *C. glutamicum* for the accurate mannosylation of recombinant actinobacterial mannoproteins showed that the recombinant host is capable of producing, exporting, and mannosylating Celf_2022 similarly to when it is produced in its native organism *C. fimi*. Further, it was determined that Celf_2022 is

actually a highly efficiently secreted mannoprotein, with the identified glycopeptide at the N-terminus of the mature polypeptide (126). While bioinformatic tools are often capable of accurately identifying and classifying signal peptides in addition to predicting protein localization, the potential for misclassification must be noted as in the case of Celf_2022. In contrast to the predicted cytoplasmic localization, significant amounts of this protein were found to be secreted into the culture media when produced in *C. glutamicum* (126).

Subsequent iterations of the SignalP algorithm now classify the leader sequence of this protein (likelihood of 0.998 using SignalP 6.0) as likely utilizing a pathway other than SEC or TAT, leading to the Celf_2022 leader sequence being classified as "Other". To confirm this prediction, the "Other" leader sequence of Celf_2022 was exchanged with the SEC leader of Celf_1230 and the TAT leader of Celf_3184. As a functional assay for this enzyme has yet to be developed, POM was used to assess which translocon could possibly be utilized by this protein during native secretion. These two translocons are both involved in the process of actinobacterial POM but differ mechanistically. Proteins that are secreted through the SEC translocon are exported in an unstructured or linear manner, while the TAT translocon secretes proteins that are in a fully folded state. A matching POM profile between one, or both leader swapped Celf_2022 constructs and the protein with its native "Other" leader would then be a strong indicator of which translocon was being utilized by the "Other" leader sequence.

3. 4. 4. Celf_2022

Celf_2022 is mannosylated (and exported extracellularly) with its native "Other" leader sequence in ATCC 13032, but replacement by either a classical SEC leader (from Celf_1230) or a classical TAT leader (from Celf_3184) completely abolished mannosylation in the secreted

recombinant protein that was recovered from the spent culture media (Figure 3.5A and D). The low levels of recombinant Celf_2022 within the cells (Figure S3.5A and B) confirmed that the protein observed in the unconcentrated culture medium (Figure S3.5C and D) was not due to cell lysis. To confirm POM of native Celf_2022 was carried out by the *C. glutamicum* GT-39 alone, all Celf_2022 constructs were expressed in the ΔCg_1014 strain where no mannosylation was detected on any protein regardless of the secretion leader utilized (Figures S3.6A and D, S3.7). These results support the predictive analysis of SignalP 6.0, suggesting that the native Celf_2022 is likely secreted by a translocon independent of both SEC and TAT as only the "Other" leader results in mature Celf_2022 with a POM profile matching that of the original protein identified from the secretomic analysis of *C. fimi* (55).

3. 4. 5. Choice of secretion pathway impacts *O*-mannosylation profile

As the "Other" leader still results in efficient mannosylation of Celf_2022 and was shown to be adjacent to both the SEC and TAT translocons, the effects of this leader on the SEC secreted Celf_1230 and TAT secreted Celf_3184 were investigated. While both proteins were also identified during the secretomic analysis of *C. fimi* (55), only Celf_3184 is known to be a mannoprotein. Like the previous experiment, the effects of exchanging the native leader sequences of these proteins with the "Other" leader sequence of Celf_2022 were assessed by comparing the POM profiles of the leader-swapped constructs to the proteins with their native leader. Again, the relative abundance of recombinant proteins in the cytoplasm (Figure S3.5A and B) compared to the unconcentrated spent culture medium (Figure S3.5C and D) confirmed their presence was not due to cell lysis and constructs were expressed in the ΔCg_1014 strain to confirm POM was carried out by the *C. glutamicum* GT-39 (Figures S3.6B, C, and D, S3.7).

3. 4. 6. Celf_1230

Celf_1230 is a uniquely thermostable glycoside hydrolase family 6 (GH-6) enzyme from *C. fimi* not identified as a mannoprotein (55, 168) with the lack of ConA lectin blotting of the recombinantly produced enzyme in *C. glutamicum* confirming this (Figure 3.5B). Swapping the native SEC leader of this enzyme with the "Other" leader did not change the mannosylation status of the secreted material (Figures 3.5B and D, S3.5C and D). However, the relative abundance of the "Other" swapped Celf_1230 was greater in both the cytoplasm and unconcentrated spent medium fractions (Figure S3.5), possibly indicative of greater stability/proteolytic resistance of the non-native fusion.

3. 4. 7. Celf_3184

Celf_3184 is another *C. fimi* GH-6 enzyme identified previously (55), different from Celf_1230 in that the former is a highly mannosylated and TAT secreted enzyme. As this protein is an excellent example of a mannoprotein outside of the classical dogma of SEC-dependent POM in actinobacteria, it was originally selected for assessing *C. glutamicum* for its use in producing accurately mannosylated recombinant actinobacterial proteins. When produced in *C. glutamicum*, Celf_3184 – with its classical TAT type leader sequence – is mannosylated with between 29 – 37 hexoses (126) which is congruent to observations seen from the same protein produced natively in *C. fimi* (55, 126). However, replacement by the "Other" leader sequence from Celf_2022 resulted in less secreted protein with a more uniform POM profile (Figure S3.5C and D). Material recovered from the spent culture medium (Figure 3.5C and D) was determined by LC-MS/MS to have 24 – 30 hexoses (Figures 3.4B). The cytoplasmic fraction also showed a

lower abundance of Celf_3184 with the “Other” leader sequence (Figure S3.5A and B) possibly indicative of decreased stability/proteolytic resistance of the non-native fusion. This is the first report of differential POM on the same target protein in actinobacteria based on the secretion leader sequence utilized.

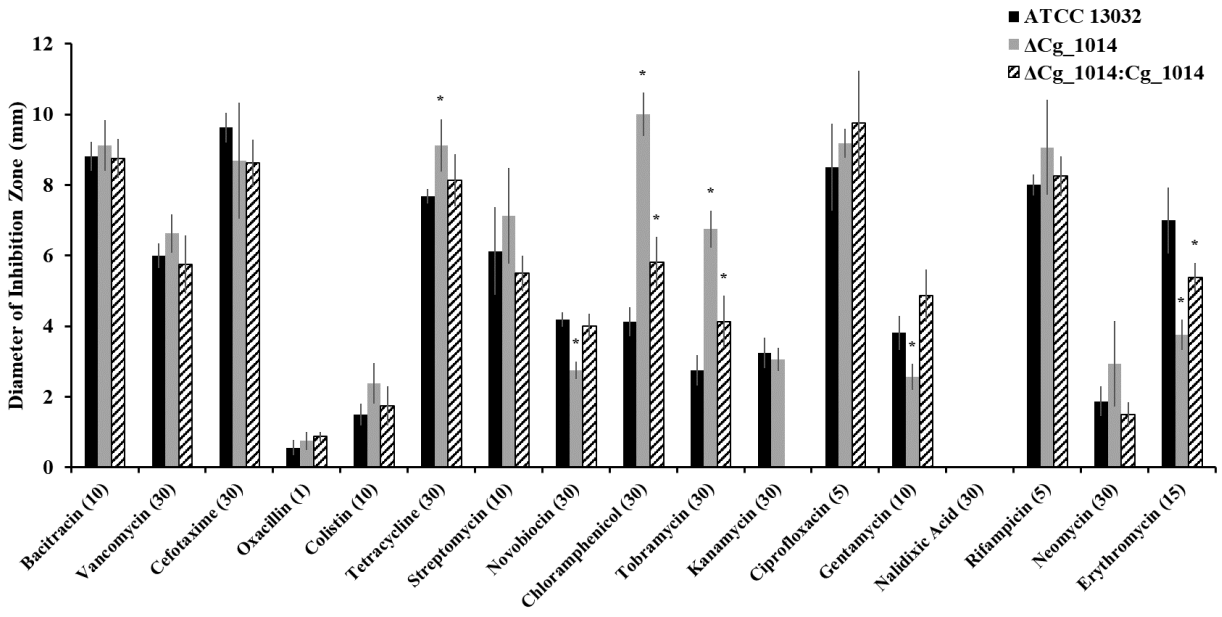


Figure 3. 1. Sensitivity of *C. glutamicum* ΔCg_1014 and ΔCg_1014:Cg_1014 to a range of antibiotics by disc diffusion.

Changes to antibiotic sensitivity in the ΔCg_1014 strain were assessed by diameter (in mm) of inhibition zone. The antibiotics having the largest effect on the mutant on zones of inhibition (ZOI, mm) were tetracycline (30 µg), chloramphenicol (30 µg), tobramycin (10 µg), and erythromycin (15 µg). These antibiotics all inhibit intracellular targets. Complementation of the mutant by Cg_1014 restored sensitivity or resistance to most antibiotics except for chloramphenicol, tobramycin, and erythromycin. These antibiotics target various ribosomal subunits (23S, 30S/50S, and 50S, respectively). Kanamycin resistance of the complemented mutant is conferred by the expression plasmid for antibiotic selection.

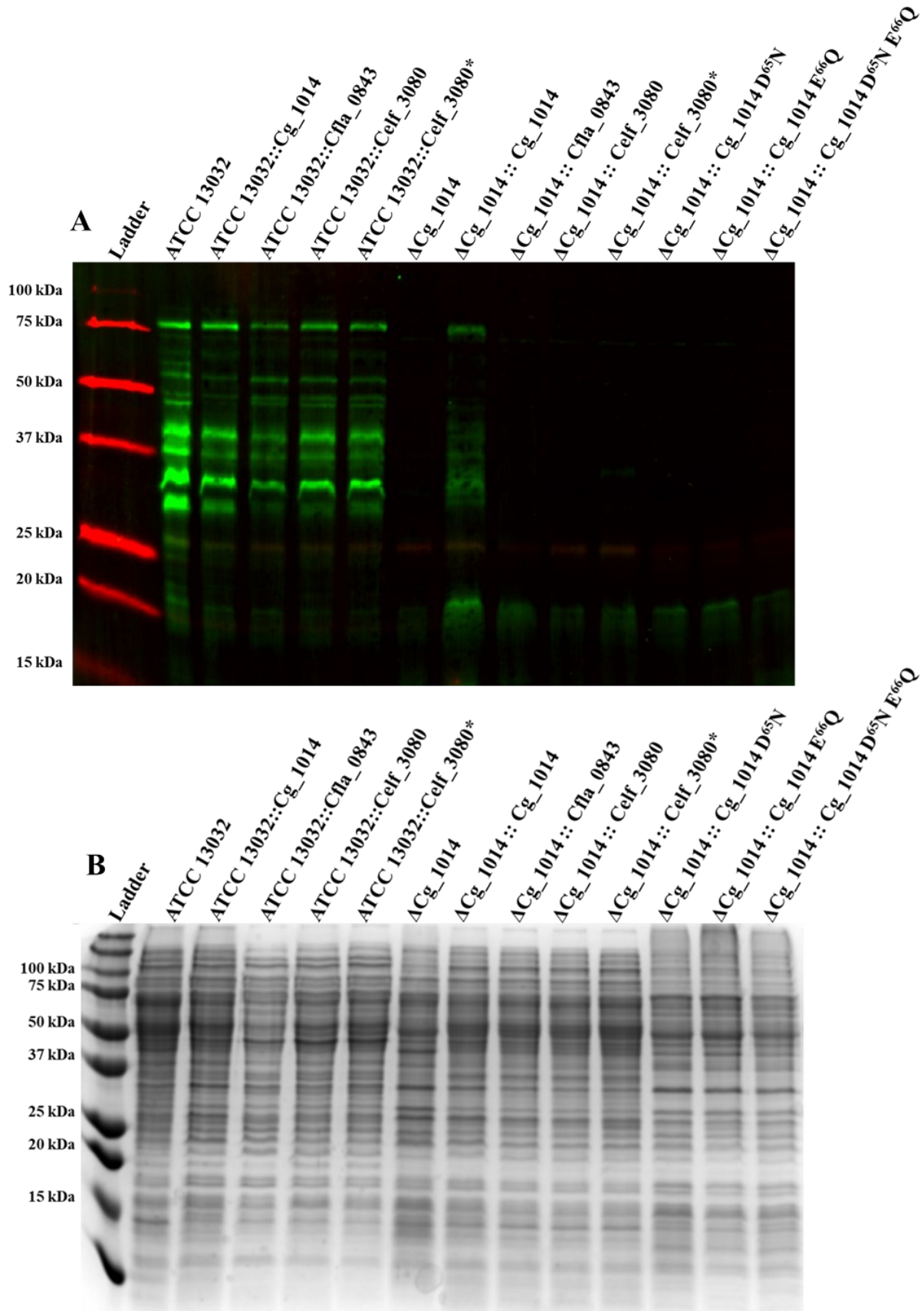

A
Ladder
ATCC 13032
ATCC 13032::Cg_1014
ATCC 13032::Cfla_0843
ATCC 13032::Celf_3080
ATCC 13032::Celf_3080*
ΔCg_1014
ΔCg_1014:: Cg_1014
ΔCg_1014:: Cfla_0843
ΔCg_1014:: Celf_3080
ΔCg_1014:: Celf_3080*
ΔCg_1014:: Cg_1014 D65N
ΔCg_1014:: Cg_1014 E66Q
ΔCg_1014:: Cg_1014 D65N E66Q
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
B
Ladder
ATCC 13032
ATCC 13032::Cg_1014
ATCC 13032::Cfla_0843
ATCC 13032::Celf_3080
ATCC 13032::Celf_3080*
ΔCg_1014
ΔCg_1014:: Cg_1014
ΔCg_1014:: Cfla_0843
ΔCg_1014:: Celf_3080
ΔCg_1014:: Celf_3080*
ΔCg_1014:: Cg_1014 D65N
ΔCg_1014:: Cg_1014 E66Q
ΔCg_1014:: Cg_1014 D65N E66Q
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa

Figure 3. 2. ConA-FITC (green) lectin blot (A) and Coomassie stained 15% SDS-PAGE (B) of *C. glutamicum* ATCC 13032 and ΔCg_1014 membrane fractions expressing recombinant actinobacterial GT-39s and Cg_1014 SDM constructs.

Overexpression of the actinobacterial GT-39s in *C. glutamicum* ATCC 13032 resulted in no changes to the native membrane protein mannoproteome as detected by ConA-FITC lectin conjugate (A, green). Western-blotting using anti-HIS_6 antibodies and fluorescent Ni-NTA conjugates did not detect any proteins indicative of recombinant GT-39s. Lack of POM in the ΔCg_1014 strain was complemented by expression of the *C. glutamicum* GT-39 Cg_1014, with minimal complementation by the *C. fimi* GT-39 Celf_3080, and no complementation was evident with the *C. flavigena* GT-39 Cfla_0843. Substitution of the conserved residues D^{65} and E^{66} to N^{65} and Q^{66}, respectively, confirmed their requirement for catalytic activity. An asterisk (*) denotes the *C. fimi* GT-39 (Celf_3080) codon optimized for expression in *C. glutamicum*. Coomassie stained 15% SDS-PAGE as loading control (B). Molecular weight standards are the Bio-Rad All Blue ladder.

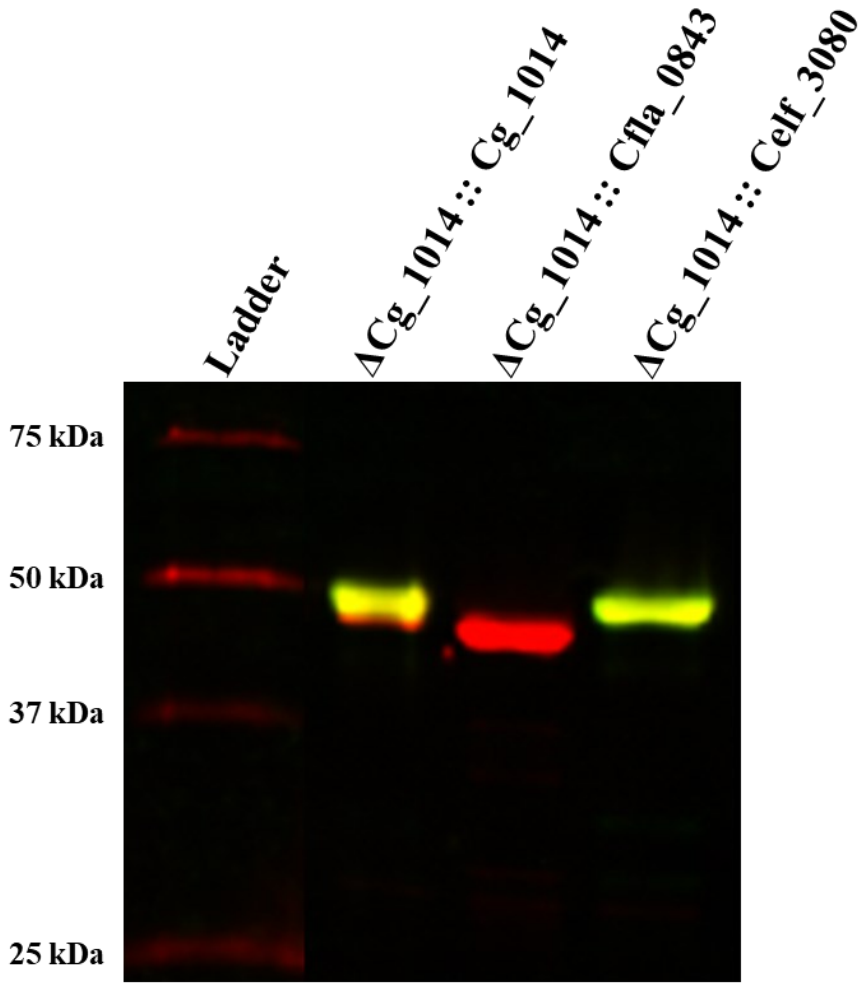


Figure 3. 3. ConA-FITC (green) and Anti-HIS-Alexa647 (red) blot of recombinant Celf_3184 recovered from spent media of *C. glutamicum* ΔCg_1014 expressing actinobacterial GT-39s.

As direct detection of recombinant GT-39s was not possible, Celf_3184 was used as an *in vivo* readout of POM activity. When co-expressed with each actinobacterial GT-39 in the ΔCg_1014 strain both the *C. glutamicum* and *C. fimi,* GT-39s accepted the known mannoprotein as a substrate, with the *C. fimi* GT-39 producing a more homogenously mannosylated product. The GT-39 of *C. flavigena* did not glycosylate this substrate, but it was correctly exported out of the cell. Molecular weight standards are the Bio-Rad All Blue ladder.

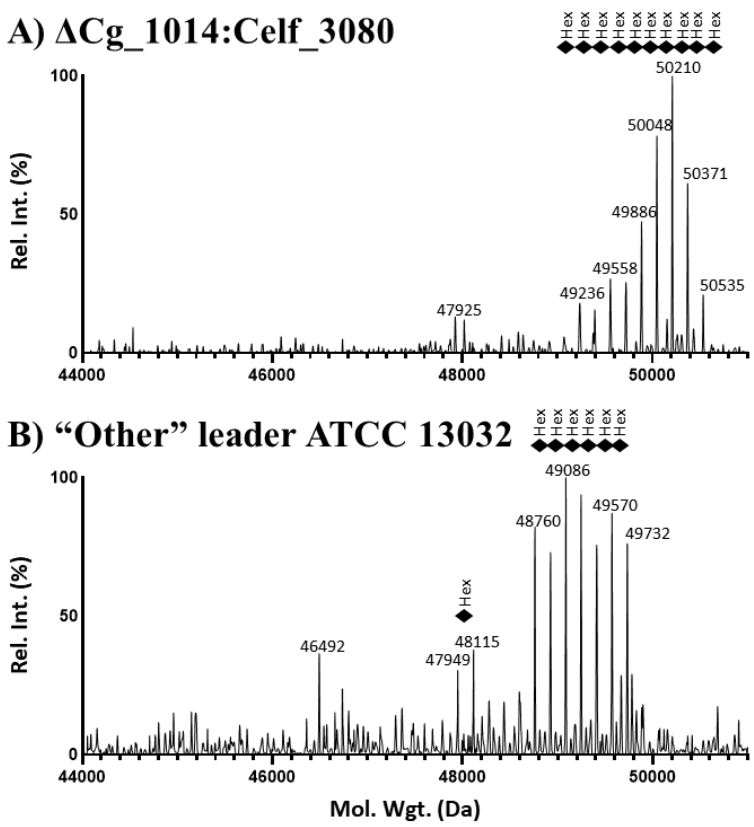


Figure 3. 4. Intact mass LC-MS analysis of Celf_3184 expressed in (A) ΔCg_1014:Celf_3080 and (B) ATCC 13032 using "Other" leader sequence from Celf_2022.

Celf_3184 was recovered from the spent culture medium. The calculated mass (with signal peptide removed and no hexose modifications) of Celf_3184 is 44,880 Da. When co-expressed with the *C. fimi* GT-39 Celf_3080 in the ΔCg_1014 strain, the observed mass profile (A) suggests this protein is modified with between 31– 5 hexoses, which is much more homogenous than the modification seen on native Celf_3184 when produced in ATCC 13032. When expressed in ATCC 13032 and with its native TAT leader replaced by the "Other" leader sequence of Celf_2022, the produced Celf_3184 has an observed mass profile (B) suggesting the addition of 24–30 hexoses; fewer hexose modifications than when native Celf_3184 is produced in ATCC 13032.

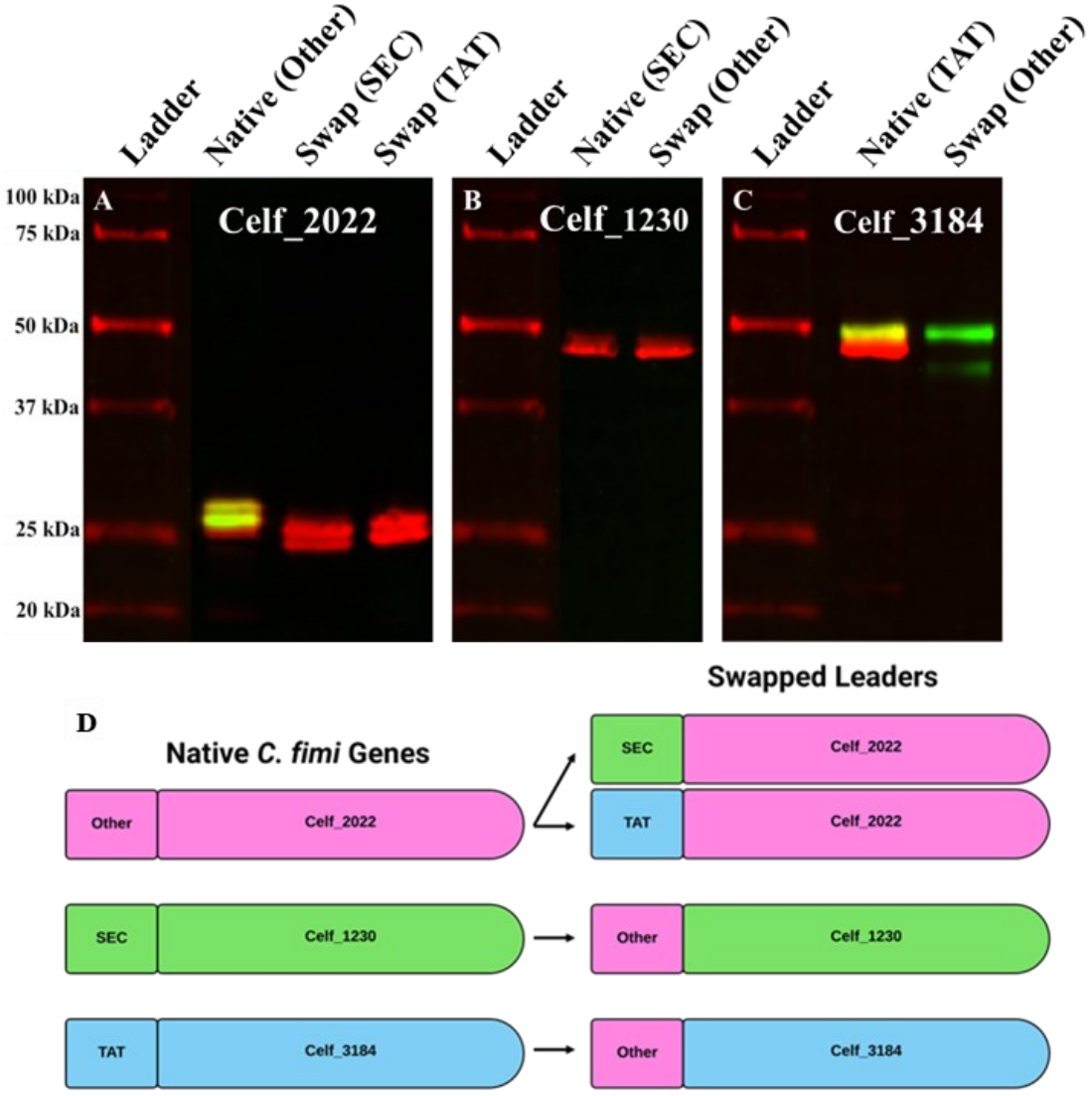


Figure 3. 5. ConA-FITC (green) and Anti-HIS-Alexa647 (red) blot of spent culture media enriched recombinant Celf_2022 expressed with "Other", SEC, and TAT leaders (A), Celf_1230 expressed with SEC and "Other" leaders (B), and Celf_3184 expressed with TAT and "Other" leaders (C) produced in *C. glutamicum* ATCC 13032. Schematic of swapped leader constructs (D).

When lead by its native "Other" leader sequence, Celf_2022 is both secreted and mannosylated in *C. glutamicum* ATCC 13032 (A). Replacement of this "Other" leader sequence by a SEC or

TAT signal peptide retains secretion but abolishes mannosylation. Replacement of the SEC signal peptide of Celf_1230 (a non-mannosylated protein) by the “Other” leader sequence results in no change to secretion or mannosylation pattern (B). Replacement of the TAT signal peptide of Celf_3184 by the “Other” leader sequence results in a more homogenous glycoform of protein enriched from the culture media (C). Schematic diagram showing swapped leader sequences between recombinantly produced *C. fimi* test proteins (D). Molecular weight standards are the Bio-Rad All Blue ladder.

3. 5. Discussion

3. 5. 1. *C. glutamicum* GT-39 is dispensable

Inactivation and complementation of the gene encoding the GT-39 of *C. glutamicum* has been reported previously (56), however only the presence or absence of mannosylation on secreted proteins was assayed; the more abundant membrane associated or even cytoplasmic mannoproteins were not considered. In addition, no phenotype or growth characteristics of this knockout have been reported to date. For this purpose, the inactivated truncation mutant which maintains up- and downstream effectors (Figure S3.1B) for Cg_1013 – a hypothetical protein with unknown function – and Cg_1015 – a uroporphyrin-III C/tetrapyrrole (corrin/porphyrin) methyltransferase (169) – was designed to be a more refined approach to generating the mutant required for the investigation of POM in *C. glutamicum*.

Initially, the presence of some ConA reactive bands and smears on lectin blots (Figure 3.2) led to the assumption of incomplete inactivation of Cg_1014, but this reactivity can be attributed to mannosylated lipids and/or lipoarbinomannan (LAM) enriched during the isolation of membrane proteins (Figure S3.3). It has been reported that POM deficient *M. tuberculosis* exhibits enhanced release of LAM (170), which was also observed with the ΔCg_1014 mutant of *C. glutamicum*. While the ConA reactive smears in Figure S3.3B are indicative of LAM in the samples, it should be noted that there are still some weak but distinct ConA reactive bands in ATCC 13032 following proteinase K digestion. If these bands are in fact proteinaceous in nature, they likely correspond to mannoproteins (or mannose containing glycopeptides) not digested by proteinase K, as POM is known to confer proteolytic resistance to some proteins (96, 171). In *M. tuberculosis,* LAM is mannosylated by a mannosyltransferase known as PimB (Rv0557), belonging to GT-4 (172, 173) – a family with no annotated protein-*O*-mannosylation

functionality (EC 2.4.1.109). Orthologues of the *M. tuberculosis* GT-4 exist in *C. glutamicum* and will be discussed later. Much like in *M. tuberculosis* (49), inactivation of the *C. glutamicum* GT-39 results in no discernable growth phenotype (Figure S3.4) exemplifying the seemingly dispensable nature of this modification in non-pathogenic actinobacteria, at least under laboratory conditions.

While microscopy showed no obvious morphological differences between the strains (data not shown), differences in antibiotic susceptibility between ATCC 13032 and ΔCg_1014 were seen with tetracycline, novobiocin, chloramphenicol, tobramycin, gentamicin, and erythromycin (Figure 3.1) – antibiotics primarily targeting Gram-positives and all with intracellular targets. These antibiotics all require transport (passive or active) across the cell membrane, so these results suggest that POM does affect either permeability or specific uptake mechanisms at least in the cases of tetracycline, chloramphenicol, and tobramycin. Surprisingly, lack of POM increased the resistance of the ΔCg_1014 strain to novobiocin, gentamicin, and erythromycin. These antibiotics can be taken up by active transport – using ABC transporters, efflux pumps, or even porins (174) – which have been characterized in *C. glutamicum* (175) – suggesting the possibility that key enzymes along these uptake routes may require POM for functionality. Hexose-modified ABC transporters have been reported in *M. tuberculosis* (176). In analogy with the non-functional PMT mutants made in *S. coelicolor* (177), only mild increases to antibiotic susceptibility are seen with vancomycin and β-lactams, which were not observed with the ΔCg_1014 strain. This suggests that in *C. glutamicum*, major cell wall effects are not a phenomenon associated with a lack of protein mannosylation.

Complementation of the mutant by Cg_1014 mostly restored antibiotic susceptibilities to that of the ATCC 13032 strain, with three exceptions. Compared to the wild-type strain, partial

complementation of antibiotic susceptibility to chloramphenicol, tobramycin, and erythromycin was observed when the ΔCg_1014 strain was complemented with Cg_1014 – these antibiotics all targeting various ribosomal subunits: 23S, 30S/50S, and 50S, respectively (178–180). This is likely related to POM in the complemented mutant not returning to levels observed in the ATCC 13032 strain (Figure 3.2). Taken together, these data implicate POM in affecting the interaction between some ribosomal subunits and the antibiotics targeting them. While investigations into the cytoplasmic mannoproteome of actinobacteria is lacking, proteomic evidence of ribosomal subunits in *M. tuberculosis* carrying hexose modifications has been reported (176).

The wide reach of POM – in terms of the biological functions of the proteins carrying the modification – makes it difficult to clearly assign any singular cellular role to it. The abundant nature of POM further complicates the answer to the fundamental question: what biological function does protein-*O*-mannosylation have in these non-pathogenic actinobacterial species?

3. 5. 2. Differential complementation of ΔCg_1014 by actinobacterial GT-39s

Verification of the actinobacterial GT-39 expression constructs was first attempted using the ATCC 13032 strain. However, no anti-His_6 reactive bands of corresponding molecular weight were identified in the membrane or intermediate fractions of any construct (data not shown) and all *C. glutamicum* expression strains were subsequently confirmed to harbour the correct vector and insert by plasmid rescue into *E. coli*. As a proxy for direct detection, changes to the ATCC 13032 native mannoproteome during the expression of each actinobacterial GT-39 was also assessed (Figure 3.2); however, no noticeable changes were observed. Integral membrane proteins can be notoriously difficult to detect, enrich, and biochemically characterize (164, 165) as they are often presented at very low levels and lose their native activity and the conformation

required for enzymatic activity when removed from biological membranes (181–183). As the *C. glutamicum* vector pCGE-31 typically produces recombinant proteins to a high (Figure S3.9), it was thought that the majority of the recombinant GT-39s produced in each construct were misfolded and then rapidly degraded. In *E. coli*, the ATP-dependent zinc metalloprotease FtsH degrades misfolded membrane proteins (184); the *C. glutamicum* FtsH homologue (Uniprot A0A1Q6BR68) likely performing a similar function and explaining the lack of detection of even insoluble, misfolded recombinant GT-39s. In *S. coelicolor*, evidence has been presented that even very conservative single amino acid substitutions to the native GT-39 sequence may cause misfolding and rapid degradation leading to both lack activity and detection by Western blotting (166). It is therefore not unreasonable to think that either the N- or C- terminal His_6 tagging of the actinobacterial GT-39s utilized here could have a similar effect.

Following this reasoning, we sought to observe, if enzymatic activity was observable using the ΔCg_1014 strain instead, where any low-level recombinant GT-39 activity would not have to compete with the natively produced enzyme for mannoprotein substrates. In the ΔCg_1014 strain, the *C. glutamicum* GT-39 complemented POM to a degree, but mannosylation levels did not return to that of the ATCC 13032 strain (Figure 3.2). As enzymatic activity is present in this complemented strain with no recombinant protein able to be detected, it is likely that at least a minimal amount of the *C. glutamicum* GT-39 enzyme is produced in an active form and localized correctly. Knowing this, we attempted to confirm the necessity of the conserved DE motif for enzymatic activity in the first extracellular loop of the *C. glutamicum* GT-39 with site directed mutants. Mutation of either $D^{65}N$ or $E^{66}Q$ totally abolishes the enzymatic activity of the *C. glutamicum* GT-39. This finding does not prove that the residues are required for catalysis, instead highlighting their importance as conserved residues within the GT-39 family as the DE

motif has been implicated in the stabilization of the GT-39 active site by coordinating the required divalent cations (143). Minimal complementation of select mannoproteins was observed with the *C. fimi* GT-39 (native and codon optimized for *C. glutamicum* expression), again suggesting that perhaps a minimal amount of recombinant enzyme was functional. However, no complementation by the *C. flavigena* GT-39 was evident. While this could suggest that these closely related GT-39s may display high specificity for their native targets of POM, the possibility of no active *C. flavigena* GT-39 being produced could not be ruled out. For this reason, we decided to utilize the highly mannosylated protein Celf_3184 (Cel6A) from *C. fimi* as a recombinant target for the *in vivo* readout of POM activity by heterologous GT-39s in *C. glutamicum*.

3. 5. 3. *In vivo* mannosylation of Celf_3184 by actinobacterial GT-39s

The *C. fimi* GH-6, Celf_3184, was chosen as the preliminary target protein for the *O*-mannosylation operons as it has previously been shown to be a suitable target for recombinant expression and POM in *C. glutamicum* ATCC 13032. As this protein is heavily mannosylated in both its native organism, *C. fimi*, and the recombinant host *C. glutamicum* (126), this target protein was utilized to assess and compare the *in vivo* activities of the different actinobacterial GT-39s expressed in *C. glutamicum* (21). When expressed in conjunction with the different actinobacterial GT-39s, differences in mannosylation can be seen, but also possible substrate preferences between these closely related enzymes (Figure 3.3). Logically, the *C. fimi* GT-39 (with native codon usage) readily accepts the *C. fimi* GH-6 and mannosylates it much more uniformly than the *C. glutamicum* GT-39 (Figure 3.4A), where different glycoforms were secreted into the culture medium. GT-39s only catalyze the initial attachment of mannose to S/T

residues and do not contribute to their elongation and as such, the increased degree of modification observed with the *C. fimi* GT-39 cannot be solely attributed to its presence. However, as previously reported most of the *O*-mannose glycans on Celf_3184 (produced recombinantly in ATCC 13032) are disaccharides and not further polymerized (126). As the initial attachment of mannose to S/T residues is likely the limiting step in POM, these results show that the *C. fimi* GT-39 recognizes and acts on a greater number of potential *O*-mannosylation sites contained in Celf_3184 than the *C. glutamicum* GT-39. However, the *C. flavigena* GT-39 does not appear to accept this substrate as a target for POM. Again, without a known *C. flavigena* target protein, the lack of active GT-39 production cannot yet be ruled out for this strain. However, *C. flavigena* harbours its own orthologue of the *C. fimi* TAT secreted GH-6 (Cfla_2913) and interestingly, the only significant structural difference between these two related enzymes is in the linker region between the CBM2a and GH6 domains which contains the identified glycopeptide of Celf_3184 (126). The linker region in Cfla_2913 is significantly larger than that in Celf_3184, potentially moving the region homologous to the Celf_3184 glycopeptide (Figure S3.8). As the N-terminus of Cfla_2913 contains TAT motifs similar to those found in Celf_3184, it is likely also secreted via the TAT translocon where respective GT-39s should be interacting with these proteins in a fully folded state. While Cfla_2913 has yet to be identified as a mannoprotein, if it is natively mannosylated similarly to its *C. fimi* counterpart then these preliminary results implicate that there are notable differences in both the structure and spatial functionality of closely related actinobacterial GT-39s. The further dissection of this observation and actinobacterial POM in general requires additional mannoprotein targets from each species be incorporated into the mannosylation operons. Comparing these mannosylation

profiles could potentially allow for the identification of a minimum sequence requirement for actinobacterial POM.

3. 5. 4. "Other" leader of Celf_2022 does not utilize SEC or TAT translocons

To further investigate the potential of Celf_2022 utilizing a translocon other than SEC or TAT, the "Other" leader sequence of the protein was exchanged with the leader sequences from both Celf_1230 (a non-mannosylated SEC secreted protein) and Celf_3184 (a heavily mannosylated TAT secreted protein). While both Celf_2022 constructs were secreted from *C. glutamicum* ATCC 13032 and enriched from spent culture medium, only the native Celf_2022 with "Other" leader sequence was ConA reactive (Figure 3.5A and D). As the SEC and TAT translocons are known to secrete proteins in unfolded and folded states, respectively, the loss of mannosylation by directing Celf_2022 through these two translocons provides evidence that the context of the glycosylation site is determined by the direction provided by the leader sequence – especially as the identified glycopeptide is totally separate from the native leader sequence which is processed following secretion. Celf_2022 is still an uncharacterized protein with a positively charged *N*-terminus followed by a transmembrane helix that spans residues 33–54, the same location where previous iterations of the SignalP algorithm predicted processing of the signal peptide and where the N-terminus of the secreted protein was determined to be (126). While these features make the protein appear similar to a single-pass transmembrane protein that likely follows the positive-inside rule, proteomic evidence suggests that this protein is fully secreted instead of membrane-associated (55). Regardless of the true location of the mature protein outside of the cell, membrane insertion and secretion both require a manner of translocation through the bilayer and the evidence presented here suggests that in the case of the "Other"

leader of Celf_2022 the translocon that is used is neither SEC nor TAT. Finally, to confirm that mannosylation of native Celf_2022 was not carried out by a separate class of mannosyltransferases (like in the case of LAM), these constructs were also expressed in the ΔCg_1014 mutant where POM was not evident on any recombinantly produced protein (Figure S3.6A, B, and C). The observed mass differences in both mannosylated and non-mannosylated Celf_2022 is the result of differential N-terminal truncation products (126).

Definitive confirmation that Celf_2022 is secreted in a TAT independent manner could be obtained with the knockout of the *C. glutamicum* TATABC translocase as previously reported (185), but showing the protein is secreted in a totally SEC-independent manner may not be possible as the knockout of the SEC translocase is known to be lethal in some bacterial species (186) and efforts to do so have yet to be reported in *C. glutamicum*.

3. 5. 5. Secretion translocons can modulate POM

To investigate the effects on POM by the "Other" leader and the currently unidentified translocon utilized by it, both Celf_1230 and Celf_3184 had their native leader sequences replaced with the "Other" leader sequence of Celf_2022. Celf_1230 was not detected as a mannoprotein in the proteomic analysis of secreted *C. fimi* proteins (55); its mannosylation profile does not change when the native SEC leader is replaced by the leader sequence of Celf_2022 (Figure 3.5B and D). This shows that utilizing the "Other" leader and translocon does not result in the aberrant mannosylation of natively non-mannosylated proteins and further reinforces the sole GT-39 of *C. glutamicum* being responsible for all POM.

When the native TAT leader of Celf_3184 was replaced by the Celf_2022 leader sequence a narrower mannosylation pattern was observed on the secreted and enriched GH-6

(Figure 3.5C and D) resulting in 24–30 hexoses (Figure 3.4B) as opposed to 29– 7 hexoses (126). This implies that the unidentified translocon utilized by Celf_2022 may exhibit a modulating effect or strong relationship between mannosylation and secretion, whether it be a form of quality control or a greater degree of interaction with GT-39s. In addition, the conformation of the mannoprotein being exported cannot yet be ruled out as we have yet to identify other actinobacterial mannoproteins like Celf_2022 that are only POM targets when exported by their native translocons.

The lack of Celf_3184 secretion in the ΔCg_1014 strain initially implied some elements of the TAT translocon require mannosylation for functionality, however, the production of Celf_2022 with a TAT leader discredited this hypothesis (Figure S3.6C). While some components of both the SEC and TAT secretion pathways have been reported to be glycosylated in *M. tuberculosis* – with glycans composed of more than just mannose (176) – mannosylation does not appear to influence the functionality of either translocon in *C. glutamicum*. It is more likely that the abolishment of mannosylation in *C. glutamicum* results in the loss of functionality of one or more chaperones – a modification also observed on chaperones from *M. tuberculosis* (176) – specifically required for recombinantly produced Celf_3184 to be folded correctly, leading to its rapid degradation (187). Currently, the translocon utilized by Celf_2022 is unknown and to truly dissect its relationship with POM it must first be identified. Of note is the recently identified class of non-classically secreted proteins lacking a traditionally identifiable N-terminal signal peptide. Currently, it cannot be entirely ruled out that Celf_2022 may be a member of this class of proteins. While the study of non-classical secretion in bacteria is steadily growing – with reports in *Bacillus*, *Listeria*, *Staphylococcus*, *Streptococcus*, *Mycobacterium*, and

others (163, 188–192) – the relationship between this novel secretion system and protein glycosylation in bacteria is under-reported.

3. 5. 6. Conclusion

While POM has been known to occur in various actinobacterial species, a significant lack of information still exists about how this modification benefits them outside of the context of virulence. Even though they are closely related, two of the actinobacterial GT-39s investigated here showed very different substrate specificities for and activities on known actinobacterial mannoprotein targets. Moreover, the mannosylation deficient strain of *C. glutamicum* showed an increased susceptibility to tetracycline, chloramphenicol, and tobramycin suggesting POM may affect cellular permeability in some manner. Most importantly, there is further evidence that POM does not only occur in a solely SEC dependent manner in actinobacteria, as proteins can be exported and mannosylated via the TAT pathway. Moreover, here we have presented evidence for a currently unidentified secretion pathway in actinobacteria that can mannosylate and export proteins without using either the SEC or TAT translocons with different levels and profiles of the modification based on the pathway used. These findings showcase how poorly understood actinobacterial POM is, how it is performed and what impact it has on the organisms that are capable of it.

Currently, *C. glutamicum* is a naïve recombinant host with the success of high-yield recombinant production occurring on a protein-to-protein basis for a variety of reasons. However, a recently developed molecular chaperone system has been shown to improve consistent recombinant protein production in the organism (187); a similar methodology will be utilized to facilitate the production and detection of heterologous GT-39s. While heterologous

protein production and secretion in *C. glutamicum* benefits from the organism's low abundance of secreted proteases (193) intracellular proteases can dramatically affect the yields of heterologous proteins (194), justifying the need for additional engineering of this strain.

3. 6. Methods

3. 6. 1. Media, strains, and expression conditions

All strains were grown in 2YT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, BioShop Canada). NEB® Stable *E. coli* (NEB) was used for routine cloning and plasmid production. BL21 (DE3) *E. coli* and *C. glutamicum* ATCC13032 were used for recombinant protein production. Electrocompetent *C. glutamicum* were cultured in MBGT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 35 g/L glycine, 0.1% Tween-80) and outgrowths were performed in 2YT + 91 g/L sorbitol.

Single colonies of *C. glutamicum* expression constructs were inoculated into 25 mL 2YT containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid, then incubated overnight at 30°C with shaking at 180 RPM. The following day, overnight cultures were diluted into 250 mL 2YT expression cultures (containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid) to an $OD_{600} \approx 0.1$ and incubated at 30°C, 180 RPM. Cultures were induced with 0.5 mM IPTG once they reached an $OD_{600} \approx 0.6$ and allowed to express overnight at 30°C. Following induction, cultures were harvested by centrifugation at 5,000 x *g* for 10 mins at 4°C.

3. 6. 2. *C. glutamicum* GT-39 knockout generation and characterization

The gene encoding the sole GT-39 in *C. glutamicum* ATCC 13032 (Cg_1014) was replaced with a truncated and inactive mutant via homologous recombination assisted by the pK18mobsacB suicide vector. A synthetic gene composed of the truncated Cg_1014 gene containing only the cytoplasmic N-terminal domain (1 – 35 aa), the first transmembrane region (36 – 58 aa), and the extracellular C-terminal domain (506 – 520 aa) flanked by upstream and downstream regions of 1001 bps was synthesized (IDT), restriction cloned into pK18mobsacB using XbaI and SalI, and subsequently cloned into electrocompetent NEB® Stable *E. coli* (NEB). The sequenced knockout construct was transformed directly into electrocompetent *C. glutamicum* ATCC 13032 and selection for gene replacement was carried out using the previously established *sac*B methodology (162) with counterselection in the presence of 20% sucrose. Replacement of Cg_1014 with the truncated gene was confirmed via PCR with specific flanking primers (Table S2).

Growth curves were recorded in triplicate in 2YT medium (containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid) at 30°C and 180 RPM throughout and OD_{600} was measured spectrophotometrically. Antibiotic susceptibility was assayed with the Kirby-Bauer (KB) methodology at 30°C using the following antibiotic discs in quadruplicate: bacitracin (BAC) 10 U, vancomycin (VAN) 30 µg, cefotaxime (CTX) 30 µg, oxacillin (OXA) 1 µg, colistin (COL) 10 µg, tetracycline (TET) 30 µg, streptomycin (STR) 10 µg, novobiocin (NOV) 30 µg, chloramphenicol (CHL) 30 µg, tobramycin (TOB) 10 µg, kanamycin (KAN) 30 µg, ciprofloxacin (CIP) 5 µg, gentamicin (GEN) 10 µg, nalidixic acid (NAL) 30 µg, rifampicin (RIF) 5 µg, neomycin (NEO) 30 µg, and erythromycin (ERY) 15 µg (Fisher).

3. 6. 3. Vector and mannosylation operon design and construction

The *E. coli/C. glutamicum* shuttle vector pTGR-5 (195) was received as a generous gift from Dr. Pablo Ravasi (Instituto de Biología Molecular y Celular de Rosario, Rosario, Santa Fe, Argentina). To generate the high-level expression plasmid, pCGE-31, the Ptac region in pTGR-5 (XbaI – NheI fragment) was replaced with the lac UV5 + tandem Plac system from the expression vector pCW (85) utilizing synthetic primers (Table S2) and maintaining the *sod* RBS and spacing already present in pTGR-5. The improved expression levels of pCGE-31 are shown in Figure S3.9.

GT-39 genes from *C. glutamicum* ATCC 13032, *C. fimi* ATCC 484, and *C. flavigena* ATCC 482 were amplified from genomic DNA using specific primers containing NdeI (5') and HindIII (3') restriction sites (Table S2). The triple lac operator from pCW-MalET (85) was used to replace the single lac operator in pTGR-5 (195) using synthetic primers containing BamHI (5') and NheI (3') restriction sites (Table S2) while maintaining the RBS_{sod} and nucleotide spacing, generating the pCGE-31 shuttle vector (Figure S3.10A) for the recombinant expression of actinobacterial GT-39s in *C. glutamicum*. A synthetic operon used to assay POM *in vivo* using a known mannoprotein (Figure S3.10B) was designed and inserted upstream of the rrnB T1 terminator using NdeI (5') and AvrII (3') containing actinobacterial GT-39s followed by a single lac operator, RBS_{sod}, and actinobacterial mannoprotein Celf_3184. Constructs were confirmed via restriction digest and sequencing, transformation of constructs into *C. glutamicum* was confirmed via plasmid rescue into NEB® Stable *E. coli* (NEB).

3. 6. 4. Production of electrocompetent *C. glutamicum* cells

A single colony of *C. glutamicum* ATCC13032 or ΔCg_1014 was inoculated into 50 mL 2YT containing 25 µg/mL nalidixic acid and incubated overnight at 30°C with shaking at 180 RPM. The following day, 1 L MBGT containing 25 µg/mL nalidixic acid was inoculated to an $OD_{600} \approx 0.1$ using the overnight culture and incubated at 30°C, 180 RPM. When the $OD_{600} \approx 0.25 – 0.25$ (about 2 hours) 0.5 µg/mL ampicillin was added and the culture was allowed to continue incubating at 30°C, 180 RPM for an additional 1.5 hours.

Following incubation, cells were harvested at 5,000 x *g* for 10 mins at 4°C. Cells were resuspended in 150 mL 10% glycerol and centrifuged at 5,000 x *g* for 10 mins at 4°C a total of 3 times. Final cell pellets were resuspended in 10% glycerol to a final $OD_{600} \approx 200$ and aliquots of 100 µL were stored at -80°C.

Electrocompetent *C. glutamicum* ATCC 13032 and ΔCg_1014 were transformed with an adapted protocol (196, 197). Aliquots of competent cell were thawed on ice and allowed to incubate with 750 ng DNA for 10 minutes prior to transformation. Cells and DNA were electroporated in 0.2-cm cuvettes using a BioRad Gene Pulser Mini at 2.5 kV for 4.80 – 5.20 ms. Immediately following, 1 mL 2YT + 91 g/L sorbitol was added to cells and outgrowths were placed at 46°C for 6 mins to inactivate the host restriction system and increase transformation efficiency. Cells were allowed to recover at 30°C, 180 RPM for 2 hours. Cells were harvested by centrifugation at 5,000 x *g* for 1 min and resuspended in 200 µL fresh outgrowth medium prior to plating on agar containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid. Plates were incubated at 30°C for 48 – 72 hours until colonies appeared.

All constructs in *C. glutamicum* were confirmed via plasmid rescue in *E. coli*. Miniprepped plasmid DNA from *C. glutamicum* constructs was transformed into electrocompetent *E. coli* for propagation, then confirmed by restriction digest and sequencing.

3. 6. 5. Purification of secreted His_6-tagged proteins

Cell pellets were harvested by centrifugation at 5,000 x *g* for 10 minutes at 4°C and stored at -20°C. The spent culture medium was clarified by centrifugation at 20,000 x *g* for 30 minutes at 4°C and particulates were removed with a 0.45 µm PES bottle top filter (supp). A 10x stock of HisTrap A Buffer (1 M HEPES, 3 M NaCl, pH 8.0) was diluted to 1X with the clarified spent media, bringing the mixture to a final concentration of 100 mM HEPES, 300 mM NaCl, pH 8.0 prior to affinity chromatography.

All recombinantly produced proteins were enriched via affinity chromatography using Roche cOmplete™ His-Tag Purification Resin (Millipore-Sigma) and chromatography on an AKTA Start (Cytiva). Clarified spent media were loaded onto an equilibrated XK-16 column (Cytiva) containing 15 mL cOmplete™ resin with HisTrap A buffer at a flowrate of 2 mL/min. The column was washed with 3 CV of HisTrap A buffer prior to elution along a linear gradient (0 – 100%) of HisTrap B (100 mM HEPES, 300 mM NaCl, 500 mM imidazole, pH 8.0) over 5 CV. Fractions containing recombinant proteins were pooled, concentrated, and buffer exchanged into 50 mM HEPES, 150 mM NaCl, pH 7.4 using 20 mL VivaSpin concentrators with 10,000 MWCO (Cytiva).

3. 6. 6. Isolation of *C. glutamicum* membrane proteins

Frozen cell pellets were resuspended (1 g/ 10 mL) in ConA buffer A (20 mM Tris, 0.5 M NaCl, 1.0 mM CaCl2 and 1.0 mM $MnCl_2$, pH 7.4) with Sigma P-2714 protease inhibitor cocktail. The cells were lysed using an Emulsiflex-C5 (Avestin) at ≥ 20,000 psi. Following centrifugation at 6,000 x g for 5 minutes at 4°C to remove debris, the supernatant was centrifuged again at 20,000 x g for 20 minutes at 4°C. The 20,000 x g pellet fraction was resuspended in 0.5 mL of ConA buffer A with 0.1% Triton X-100 and incubated at room temperature, using a tube roller, overnight at 4°C. Solubilized membranes were centrifuged at 100,000 x g for 1 hour at 4°C and the resultant supernatants were used for Western and lectin blotting.

3. 6. 7. Western and lectin blot protocol

Lectin blots were performed as described in the manufacturer's (Sigma) data sheets. Briefly, the proteins of interest were separated on a 12% SDS PAGE gel using a MiniProtean system (Bio-Rad), which was then rinsed three times for 5 minutes each in excess Tris buffered saline pH 7.6 (TBS, 50 mM Tris, 150 mM NaCl, pH 7.6) before being blotting to PVDF membrane using a Trans-Blot Turbo transfer system (Bio-Rad). The transfer was performed in 48 mM Tris, 39 mM glycine using 2.5 A, 25 V for 8 minutes. The protein-bearing PVDF membrane was then rinsed three times for 5 minutes, each, in TBS and blocked for one hour in 5% BSA in TBS at room temperature. The blocked membrane was washed three times for 5 minutes in TBS at room temp, then incubated overnight at 4ºC in 0.05% Tween 20; 1mM $CaCl_2$; 1mM $MgCl_2$; 1mM $MnCl_2$; 0.5 µg/mL ConA-FITC conjugated lectin (Millipore-Sigma); and 1:10, 000 AlexaFluor 647 anti-HIS_6 (Bio-Rad) in TBS. The membrane then underwent three 10-minute

washes in TBS pH 7.6 at room temp and was visualized on a Bio-Rad ChemiDoc. Molecular weight markers were Bio-Rad All Blue standards.

3. 6. 8. LC-MS analysis of intact mass from Celf-3184 mannosylated by actinobacterial GT-39s

Intact mass analysis was performed using an Ultimate 3000 (Dionex/Thermo Fisher Scientific) linked to an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific). 5 µg of each protein was injected on to a 2.1 x 30 mm Poros R2 column (Thermo Fisher Scientific) and resolved using the following rapid gradient: hold at 20% mobile phase A for 3 minutes, 20% - 90% mobile phase B in 3 minutes, hold at 90% mobile phase B for 1 minute. Mobile phase A was 0.1% formic acid in ddH2O and mobile phase B was acetonitrile. The flow rate was 3 mL/min with 100 µL split to the electrospray ion source. Optimal peak shape was achieved by heating the column and mobile phase to 80°C. The mass spectrometer was tuned for small protein analysis using myoglobin and the resolution was set to 15,000. Mass spectra were acquired from m/z 400 to 2,000 in the Orbitrap at 1 scan per second. The spectra acquired across the protein peak were summed and deconvoluted using MaxEnt 1 (Waters).

3. 7. Acknowledgements

This work was partially funded by a Natural Sciences and Engineering Research Council grant to WW.

4. *CORYNEBACTERIUM GLUTAMICUM* AS A RECOMBINANT PLATFORM FOR THE EXPRESSION OF BACTERIAL AND EUKARYOTIC SIALYLTRANSFERASES

as submitted to Microbial Cell Factories

September 2023

by Saxena, H., Siapuatco, D., Thompson, N., Leclaire, L., Kirby, M., and Wakarchuk W.

4. 1. Chapter Overview

This chapter describes the development process of adapting *Corynebacterium glutamicum* into a more robust heterologous protein production platform. The engineering undertaken here dramatically improved the recombinant yield from the organism and highlighted areas that the strain requires additional improvements. To showcase the capabilities of the organism, a co-expression strategy was developed to produce hST6GalI – a eukaryotic sialyltransferase containing disulfide bonds – in what is essentially a wild-type strain of *C. glutamicum*. The results presented here show that *C. glutamicum* could produce the enzyme in comparable yield and of comparable activity to the same enzyme produced in a strain of *E. coli* more tailored towards protein production. In addition, this chapter compares the production capabilities of *C. glutamicum* to *E. coli* using prokaryotic sialyltransferases and shows that these enzymes can be utilized for the *in vitro* glycosylation of therapeutic proteins.

4. 2. Abstract

The accurate glycosylation of protein therapeutics is currently one of the various stratagems to increase both their efficacy and serum half-life. It has been repeatedly demonstrated

that the capping of *N*-linked glycans by terminal sialic (*N*-acetylneuraminic) acids significantly increases the *in vivo* half-life of a given glycoprotein. Terminal glycosylation has proven to be difficult to reproduce in mammalian cell culture systems due to a large variation in the efficiency of native sialyltransferases in various production strains, in addition to non-human cell lines incorporating a non-human analogue of sialic acid - *N*-glycolylneuraminic acid (Neu5Gc) - which is highly antigenic to humans. Thus, the incorporation of these antigenic epitopes into recombinant therapeutics must be avoided. In this regard, it is currently our goal to perform the final glycosylation step of *N*-linked glycans *in vitro*, with sialyltransferases (STs) produced in prokaryotic expression systems. However, enzymes made in *E. coli* are contaminated with endotoxin which can be costly and challenging to remove. The Gram-positive soil micro-organism *C. glutamicum* has been widely utilized in industrial applications such as the production of amino acids, nucleotides and even enzymes – with one major benefit of the final product being essentially endotoxin free. *C. glutamicum* is currently a naïve recombinant host, but with minimal engineering and the co-expression of folding chaperones we have produced both prokaryotic and eukaryotic STs in the organism with comparable activities and yields to the traditional recombinant host *E. coli*.

4. 3. Introduction

4. 3. 1. Sialic acids

Sialic acids are family of acidic nine-carbon monosaccharides (Figure S4.1), which are derivatives of neuraminic acid (Neu), 5-amino-3,5-dideoxy-2-nonulosonic acid (198). These monosaccharides are widely found in nature, with the most common form being 5-*N*-acetyl-neuraminic acid, Neu5Ac. In mammalian cells, these are usually presented as the terminal sugar

residue(s) of cell-surface or secreted glycoproteins and glycolipids (199, 200), where they play important roles in various biological processes – especially in cell recognition and signalling. There are many nonulosonic acids with differing physiological relevance such as 2-keto-3-deoxy-nononic acid (Kdn), or those limited to bacterial species such as legionaminic acid (Leg), and pseudaminic acid (Pse) (201–203). In non-human animals, another variant is present, 5-*N*-glycolylneuraminic acid (Neu5Gc) which is made in most non-human cell lines and can even be incorporated into human glycoproteins through recycling the Neu5Gc from animal milk and meat products (204, 205). As 5-*N*-acetlyneraminic acid (Neu5Ac) is the most common sialic acid presented in human tissues (206), viruses and pathogenic bacteria have also evolved to decorate their own cell surfaces with Neu5Ac in an effort evade the host's innate immune response and facilitate their specific interaction with different host-cell surfaces (207–209).

It is not only the various substituents of the sialic acids that affect their function; the nature of the underlying glycosidic linkage that attaches sialic acids to neighbouring sugar residues further defines the role of a given sialic acid. Through the specificity of sialyltransferases (STs), Neu5Ac can form different linkages through its C-2 hydroxyl group with the C-3, C-6, and the C-8/9 hydroxyl of other sugars – the latter in the case of other sialic acids (206). α2-3 linkages are commonly found in glycoconjugates present in the respiratory tract and gastrointestinal tract; they are also the preferred linkage for the binding of certain influenza A and B viruses to host cells, facilitating viral infection (210, 211). α2-6 linkages are abundant in glycans found on the surface of mammalian cells – including those in the respiratory tract and certain tissues – and is also exploited by certain influenza A viruses as a receptor for cell entry (212, 213). α2-8 linkages are less common than the α2-3 and α2-6 linkages but are found on specific glycoconjugates, particularly in specific gangliosides like GD3 and GT1a, or in

polysialic acids (PolySia). PolySia, consisting of repeating units of α2-8-linked Neu5Ac, plays important roles in neural development, plasticity, and cell migration (206, 214).

In addition to their roles in cell recognition, sialic acids like Neu5Ac also have protective functions. When presented as the terminal monosaccharide or oligosaccharides, like in PolySia, sialic acids act as critical factors that determine the circulating half-life of glycoproteins (68, 215). As such, when sialic acids are missing or removed from the termini of glycans the underlying monosaccharides (such as galactose) are recognized by asialoglycoprotein receptors in the liver and other organs, resulting in the rapid clearance of the glycoprotein (216, 217). As many biotherapeutics (cytokines, antibodies, and hormones) are glycoproteins, this phenomenon is significant as inadequate sialylation results in the rapid clearance of the therapeutic molecule. Sialylation of therapeutic proteins can also alter their bioactivity and specificity, improving their efficacy (74, 218). Moreover, the sialylation of therapeutic proteins can reduce their immunogenicity thereby decreasing the likelihood of an immune response and reducing the risk of adverse effects (219). Currently, there are numerous efforts to enhance the sialylation in mammalian cell lines (75–77, 220) or to produce sialylated glycoproteins in plants and plant cell lines (78, 79, 221), yeast (222, 223), insect (80, 224, 225), and even bacterial cells (81, 82).

4. 3. 2. Sialyltransferases (STs)

STs are glycosyltransferases that are involved in the transfer of Neu5Ac to glycoconjugates, *N*- and *O*-linked glycoproteins, and glycolipids (206, 226). STs are responsible for the transfer of sialic acid residues from a nucleotide (cytidine 5'-monophosphate, CMP) - activated sugar donor to a variety of specific acceptor molecules such as a glycoprotein or a glycolipid, often terminated with a galactose (Gal), *N*-acetylgalactosamine (GalNAc), or another

Sia residue (227). As Sialic acids (and therefore STs) are important virulence factors for several bacterial pathogens – including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* (207, 228, 229) – there have been significant efforts in developing drugs that target STs, with the goal of inhibiting their activity (230–232). Currently, STs are of substantial interest because of their application to the "capping" of therapeutic glycoprotein-based biologics (both *in vivo* and *in vitro*), improving both their efficacy and circulating half-lives (73, 84, 216, 233).

STs have been classified based on the glycosidic linkages they form, commonly α2,3 and α2,6 linking to galactose (Gal), and α2,8 linking to other sialic acids (234) – though other linkages like α2,9 to sialic acids have been identified in bacteria like *Escherichia coli* K92 (235) and *N. meningitidis* group C (236). All STs identified to date have been classified based on their protein sequence homology into five unique glycosyltransferase (GT) families in the Carbohydrate Active enZymes (CAZy) database (123) available at www.cazy.org. Currently, STs and polysialyltransferases (PSTs) originating from eukaryotes and viruses are categorized into GT family 29 (GT-29), while the comparatively less conserved bacterial STs encompass GT-38, GT-42, GT-52, and GT-80. Because of the diversity in their respective sialomes (defined as the total complement of sialic acid types, linkages, and their mode of presentation) and the variety of STs produced in individual species, mammalian STs are further subcategorized based on the type of linkage formed and sugar acceptor molecule (237).

While mammalian STs generally exhibit rather strict substrate specificities, there can be some overlap – the donor substrate specificities of mammalian STs can be somewhat more relaxed; capable of transferring CMP-Neu5Ac, CMP-Neu5Gc, CMP-Neu5,9Ac_2, CMP-Kdn or even fluorophore-conjugated CMP-Neu5Ac (238–243). While this phenomenon is intriguing,

relaxed donor preferences are highly undesirable in the context of producing sialylated biologics in non-human cell lines like Chinese Hamster Ovary cells (CHO), which produce CMP-Neu5Gc that is highly antigenic in humans (244). As all mammalian STs are type II membrane proteins with a single transmembrane and stem domain (as Golgi localized proteins) and typically contain three disulfide bonds (227, 245), recombinant bacterial expression producing soluble and active protein in adequate amounts for *in vitro* capping of glycoproteins is a more complicated matter. In the case of eukaryotic STs, mammalian cell culture can sometimes outperform bacterial expression systems, however this benefit comes with significantly increased production costs.

Differing from their typically monofunctional eukaryotic counterparts, bacterial STs are much less conserved and can have strong back reactions (sialidase like) which can limit their use in *in vitro* reactions and some enzymes in GT-42 can even form two different linkages (227). While bacterial STs have been primarily studied for their relation to pathogenesis – like in *C. jejuni, H. influenza, N. gonorrheae,* and *N. meningitidis* (246–249) – non-pathogenic bacteria like *Photobacterium* spp. also harbour STs which have been characterized (250, 251). As the promiscuous nature of bacterial STs can allow for both the production of differentially linked sialosides and a wider range of sugar acceptor substrates, these enzymes show considerable promise for utilization in therapeutic glycoprotein capping strategies. In addition, bacterial STs are typically much simpler to produce recombinantly than their eukaryotic counterparts. However, bacterial STs can be less active on glycoprotein substrates than on their natural oligosaccharide or small molecule structures (85). As GT-29, GT-42, and GT-80 enzymes are all capable of producing human-like sialosides (Neu5Ac), enzymes from these families show significant potential in producing uniformly sialylated therapeutic proteins using both *in vivo* and *in vitro* approaches.

4. 3. 3. *In vivo* versus *in vitro* sialylation of therapeutics

There exist several platforms for the *in vivo* production of sialylated therapeutic glycoproteins. However, these platforms commonly exhibit marked variation in the Neu5Ac content of their final products due to intrinsic culture and expression conditions (72, 244, 252). Additionally, protein production in mammalian cell culture is expensive; frequently utilizing non-human cell lines (like CHO cells) which requires strict quality control to ensure the final product is free from antigenic non-human sialic acids (253), ultimately impacting final yields and further increasing costs. While overall a cheaper platform, production of sialosides in bacteria faces similar issues, specifically production in *E. coli* including the battle against LPS/endotoxin contamination (254, 255). Common to all platforms is the sheer amount of development required to steer production towards a single, homogenous glycoform. In the case of eukaryotic cell lines, this has been accomplished by knocking out or knocking down numerous genes responsible for (but not limited to) generating branched mannose structures, transferring antigenic sugars, creating antigenic linkages, and shunting precursors down alternative pathways to ensure adequate sugar donor levels. In addition, both eukaryotic and bacterial expression systems require the stable or transient addition of many GT genes, often of human origin, to accurately "humanize" the produced glycan structures. Finally, whole biosynthetic pathways sometimes need to be incorporated as many expression platforms do not produce Neu5Ac (74–76, 80, 82–84, 220, 222, 224).

As the final enzymatic step in sialoside production, the implementation and optimization of Neu5Ac and STs production can be a lengthy process. In contrast, *in vitro* sialylation methodologies utilizing recombinantly produced STs allows for the rapid production of highly

sialylated products that can be brought to market during the development and/or optimization of incorporating sialylation into an expression platform. Additionally, *in vitro* sialylation strategies are inherently more flexible than their *in vivo* counterparts. Not only can Neu5Ac linkages be changed on demand, more complex sialic acid structures like PolySia can easily be enzymatically installed over multistep reactions with well characterized enzymes like CST-I and CST-II from *C. jejuni* (256, 257).

4. 3. 4. *Corynebacterium glutamicum*

C. glutamicum is a Gram-positive, facultatively anaerobic soil bacterium that is generally regarded as safe (GRAS) and was originally identified in the 1950s for its ability to produce L-glutamate (258–260). Since then, *C. glutamicum* has been widely utilized in industrial scale fermentation processes to produce other amino acids, biofuels, plant secondary metabolites, compounds for health, cosmetic, food and animal feed industries, and even heterologous proteins like single-chain variable fragment (scFv) antibodies (261–264). The development of various genetic engineering tools allowed for *C. glutamicum* to become the industrial powerhouse it is now (265–269). As *C. glutamicum* is endotoxin-free, can achieve true secretion (with minimal amounts of endogenously secreted proteins), and exhibit very low extracellular protease abundance (270–272) , the organism possesses significant potential in its application as a heterologous bacterial expression platform.

4. 3. 5. Eukaryotic ST production in *C. glutamicum*

Typically, strains of *E. coli* are not capable of forming cytoplasmic disulfide bonds, primarily due to reductive pathways involving enzymes like glutaredoxin and thioredoxin

reductase protecting the cell from oxidative stress (273). To circumvent this issue and successfully produce disulfide bond-containing proteins cytoplasmically, a variety of engineered strains of *E. coli* – like the SHuffle® strains from New England Biolabs (NEB) – have been designed with partial or total disruption to the interfering reductive pathways in addition to taking advantage of the disulfide bond "proof-reading" capabilities of disulfide bond isomerase C (DsbC) overexpressed cytoplasmically (274–276). Furthermore, co-expression of the human protein disulfide isomerase (hPDI) and the yeast sulfhydryl oxidase (Erv1p) facilitates the high-level expression of complex disulfide bonded eukaryotic proteins in the cytoplasm of *E. coli* (277). Co-expression of the human α2-6 ST (hST6GalI) containing three disulfide bonds with hPDI results in the production of soluble and active enzyme in high yield. The co-expression of a hPDI and human quiescin/sulfhydryl oxidase 1b (hQSOX1b) fusion dramatically boosts the specific activity of a maltose binding protein (MBP)-hST6GalI fusion protein produced in *E. coli* compared to hPDI alone (85).

As high-yield mammalian enzyme production in *C. glutamicum* is currently in its infancy, the platform suffers from many of growing pains experienced with *E. coli*. Utilizing existing molecular biology tools for genetic manipulations in the organism (268, 269) and folding chaperone co-expression strategies (85, 277), this work serves as a proof of concept highlighting the capabilities of *C. glutamicum* as a recombinant expression host for the high-yield production of both bacterial and eukaryotic STs by adapting methodologies originally developed for use in *E. coli*.

4. 4. Results

4. 4. 1. pCGE-31 and pDual for co-expression of folding chaperones in *C. glutamicum*

To increase overall protein expression level, the *E. coli*/*C. glutamicum* shuttle vector pCGE-31 (Figure S4.2A) was generated by replacing the singular P_{tac} of pTGR-5 (Figure S4.3A) with the P_{lacUV5} and tandem P_{tac} promoter system from pCW-MBPT (Figure S4.3B). The T4 phage lysozyme RBS of pCW-MBPT is suitable for high-yield heterologous expression in *E. coli* (data not shown), while in *C. glutamicum* the *sod* RBS was shown to provide the highest expression levels when compared to the *lacZ* (*E. coli*) and *cspB* (*C. glutamicum*) RBS sequences (268). A comparison between the *sod* RBS and T4 phage lysozyme RBS sequences showed the necessity of the former for recombinant production in *C. glutamicum*, where the improved expression of HIS_6-tagged eGFP by pCGE-31 was confirmed by anti-HIS_6 Western blotting (Figure S4.4).

The co-expression of folding chaperones has been widely utilized in prokaryotic hosts to facilitate the production of "difficult" recombinant proteins, improving solubility, functionality, and yield (278). Recently, the compatible two-plasmid and differentially inducible pRG_Duet system for the simultaneous expression of multiple genes or operons in *C. glutamicum* was developed (269). Incorporating the $P_{tetR/tetA}$ of this system into pCGE-31 would allow for the same functionality but circumvent the increased burden maintaining two plasmids would have on bacterial physiology (279) and therefore heterologous production. In addition, the bicistronic pCGE-31 would be capable of higher level isopropylthio-β-galactoside (IPTG) induced expression levels in *C. glutamicum* than the similarly designed bicistronic pOGOduet vector (280).

As such, the addition of the secondary expression cassette to pCGE-31 for the co-expression of various folding chaperones (Figures S2C and D) ultimately resulted in the generation of pDual_4 (Figure S4.2B). This expression cassette contained the small ubiquitin related modifier peptide (SUMO) as an N-terminal solubility enhancer linked to the hPDI-hQSOX1b fusion driven by the anhydrotetracycline (ATc) inducible $P_{tetR/tetA}$. A synthetic gene containing the repressor, promoters, and mRuby3 as a stuffer sequence was inserted to pCGE-31 using the PstI restriction site, with the mRuby3 sequence being replaced by target chaperone sequences (Figure S4.5A and B). The relative reductase activity of *C. glutamicum* lysates expressing the different folding chaperone constructs was assessed using a PDI activity assay (Figure 4.1). The construct harbouring both the hPDI-hQSOX1b and DsbC chaperones (pDual_4) increased the reductase activity of *C. glutamicum* lysates approximately 1.6-fold compared to the ATCC 13032 strain.

4. 4. 2. Bacterial ST production in *C. glutamicum*

The accurate sialylation of therapeutic glycoproteins requires a library of competent ST enzymes capable of producing the necessary α2,3, α2,6, and α2,8 sialic acid linkages. As prokaryotic STs are generally simpler to produce in prokaryotic hosts than eukaryotic STs, the previously reported α2,3 sialyltransferase from *Bibersteinia trehalosi* (BST) and a rationally engineered mutant of the *Helicobacter cetorum* α2,6 ST (HcST) were selected to test the recombinant production capabilities of *C. glutamicum* (85, 281). The multifunctional α2,3/α2,8 ST from *Campylobacter jejuni* (CSTII) was also selected for *C. glutamicum* production for its α2,8 activity as this enzyme generates the oligosaccharide substrate required by PST enzymes (282).

Soluble HIS_6-tagged BST and HcST were expressed to a slightly greater degree in *C. glutamicum* lysates, compared to *E. coli* BL21 lysates – all with clear induction bands (Figures 4.2A, S4.6A). However, expression levels of CSTII were much higher in *E. coli* BL21 lysates than in *C. glutamicum*. Overall, the expression of CSTII in *C. glutamicum* displayed far more variability and did not approach the expression levels of the same construct produced in *E. coli* BL21. All STs of bacterial origin were enriched via IMAC (Figure S4.8) from lysates of both *E. coli* and *C. glutamicum* in comparable quantity and purity (Figures 4.3A, S4.6A), except for CSTII where the low cytoplasmic abundance likely contributed to the non-specific binding of intrinsic *C. glutamicum* proteins.

4. 4. 3. hST6GalI made in *C. glutamicum*

While much more difficult to produce in prokaryotic systems, the human α2,6 ST (hST6GalI) acts on glycoprotein substrates much more readily than its prokaryotic counterparts (85). To assess the effects of co-expressing folding chaperones on both the production and activity of this enzyme in *C. glutamicum*, each hST6GalI construct was first produced in both *E. coli* BL21 and *C. glutamicum* ATCC 13032 and recombinant fusion proteins were detected in the soluble fractions of lysates by Western-blotting with an anti-MBP-HRP conjugated antibody (Figures 4.2B and C, S4.6B and C). In both strains, the MBP-hST6GalI fusion was not produced in a soluble form and found entirely in the insoluble fraction (Figure S4.9). The addition of an N-terminal SUMO tag or the co-expression of the hPDI-hQSOX1b fusion – either alone or in concert with DsbC –improved the yield of soluble hST6GalI in both strains to differing degrees. All constructs enriched from *C. glutamicum* soluble lysates showed significantly less degradation than their *E. coli* counterparts (Figures 4.3B and C, S4.7 B and C). In both strains, the MBP-

hST6GalI fusion co-expressed with the hPDI-hQSOX1b fusion and DsbC displayed the greatest yield of soluble protein per litre culture (Figure S4.10).

4. 4. 4. Small molecule (BDP) assays

To assess and compare the activities of each ST construct, both *E. coli* and *C. glutamicum* produced enzymes were assayed on fluorescently tagged oligosaccharide substrates. BODIPY-*N*-acetyllactosamine (BDP-LacNAc) was utilized for assessing the α2,3 transfer or α2,6 transfer of Neu5Ac to Gal by BST and HcST respectively, while BODIPY-sialyl-LacNAc was used to assess the first α2,8 transfer of Neu5Ac by CSTII; these synthetic analogues have previously been utilized for the characterization of STs of both pro- and eukaryotic origin (85, 283). The determined specific activities (nmol/min/mg) of each bacterial ST produced in both prokaryotic hosts are summarized in Table 1 and reverse phase HPLC traces showing the facile separation of each substrate and product are shown in Figure S4.11. BST and HcST were produced in both organisms with comparable specific activities; however, the α2,8 activity of *C. glutamicum* produced CSTII was slightly lower than the *E. coli* produced material. This trend was also observed with BDP- α2,6-sialyl-LacNAc, as CSTII is also capable of accepting this substrate, though with a lower activity than its preferred substrate (Figure S4.12A).

As the MBP-hST6GalI construct co-expressed with the hPDI-hQSOX1b fusion and DsbC (referred to as HUST-166 from here) displayed the greatest relative activity on BDP-LacNAc in both strains (Figure S4.13), this construct was selected for further comparisons. The specific activity of *E. coli* produced HUST-166 utilizing BDP-LacNAc was determined to be 59.9 ± 11.6 nmol/min/mg and the *C. glutamicum* produced enzyme was comparable with a specific activity of 45.6 ± 17.3 nmol/min/mg (Table 1).

4. 4. 5. Ethyl-*p*-aminobenzoic acid (EPAB) labelled *N*-glycan assays

While small oligosaccharide substrates do allow for the facile assessment of ST activity, they often do not closely approximate the natural glycan/glycoprotein substrates of the enzymes they are used to assay. Primarily, small molecule substrates are not able to show the nuanced differences in activities or preferences of any particular ST enzyme. For example, an ST enzyme may preferentially sialylate one arm of a multi-antennary glycan prior to the modification of the remaining sites – if the enzyme is even capable of interacting with and modifying all potential sites on any particular glycan substrate. For this reason, the relatively simple bi-antennary G2 *N*-glycan – composed of the typical *N*-glycan core, with each arm carrying a terminal LacNAc moiety – was used to better asses the activities of STs on a more natural substrate.

E. coli and *C. glutamicum* produced BST and HcST displayed similar activities when assayed on the free G2-EPAB labelled *N*-glycan substrates. Overall, HcST showed a preference for producing biantennary *N*-glycans with one sialylated arm (Figure 4.4A and B), while BST was more agnostic in regard to substrate arm preference (Figure 4.4C and D). CSTII was not assayed with this substrate as the weak anion exchange (WAX) column used for separation of the different sialylated products was not capable of resolving two α2,8 Neu5Ac additions on the same arm from one α2,8 addition to each arm (Figure S4.14). In addition, determining any potential arm preference of the CSTII enzyme is not critical for its ability to fully modify the *N*-glycans of glycoproteins (283). When assayed on free *N*-glycans, the activities of HUST-166 produced in either strain were comparable and displayed the well known (284) strong preference for first modifying one arm of the biantennary *N*-glycan before sialylating the remaining arm (Figure 4.4E and F).

4. 4. 6. Asialo-glycoprotein modification

Synthetic substrates are indispensable tools allowing for the kinetic characterization of STs; however, the purposes of this work require the assessment and comparison of STs acting on glycoprotein substrates. As glycoproteins are inherently much more complex than the synthetic substrates mentioned and utilized previously, here they serve as a much better substrate for the characterization of ST activity. Moreover, as the underlying protein influences both the orientation and accessibility of the attached glycans, one ST can display significantly different activities on various glycoproteins (238). For the purposes of therapeutic glycoprotein sialylation, this necessitates the need for a library of competent ST enzymes. The human glycoprotein α1-antitrypsin (A1AT) with 3 bi-antennary complex-type *N*-glycans (285) was used as a readily modifiable glycoprotein substrate, while the (recombinantly produced *in plantae*) human glycoprotein butyrylcholinesterase (BuChE) with up to nine complex and hybrid-type *N*-glycans (286) was utilized as an example of recombinant protein with high mannose, hybrid and incomplete *N*-glycans (85, 283).

Bacterial STs can be much less active on glycoprotein substrates than their eukaryotic counterparts. This can be observed with HcST acting on both the terminally desialylated A1AT (aisaloA1AT) and terminally desialylated BuChE (asialoBuChE) substrates (Figure 4.5A and B, Figure S4.15A and B) when detected by lectin blotting using an mNeonGreen-tagged diCBM40 – previously shown to bind both α2,6 and α2,3 linked Neu5Ac (287). These asialoglycoprotein substrates were generated by treatment with sialidase to remove terminal sialic acids. HcST produced in both *E. coli* and *C. glutamicum* are comparably active on these asialoglycoproteins. Conversely, BST is a prokaryotic enzyme that appears to readily modify the target

asialoglycoproteins with an α2,3 Neu5Ac (Figure 4.5C and D, Figure S4.15C and D). Most importantly, both the *E. coli* and *C. glutamicum* enzymes are comparable in their ability to modify these asialoglycoproteins.

The GM3 type oligosaccharide on a bacterial glycolipid is the natural substrate of CSTII, but the enzyme will readily accept (α2,3/α2,6) sialyl-LacNAc. As such, these asialoglycoprotein substrates were "primed' with both an α2,3- and an α2,6-Neu5Ac to serve as substrates for the CSTII enzymes. As the primary intent was to show that *C. glutamicum* is a comparable recombinant host to *E. coli* BL21and not the enzymatic characterization of the CSTII enzyme, the utilization of a slightly lower preference substrate was deemed acceptable. CBM40 has been reported to bind α2,8 linked Neu5Ac but weaker than α-2,3/2,6 (288), so these additions catalyzed by CSTII were detected by both a loss of fluorescence from the primed glycoprotein substrates and an increase in molecular weight as evident by Coomassie stained SDS-PAGEs. Again, a similar specific activity of the *E. coli* and *C. glutamicum* enzymes was observed with α2,3-Neu5Ac primed glycoproteins (Figure 4.5E and F, Figure S4.15E and F), with the trend extending to α2,6-Neu5Ac "primed" glycoproteins (Figure S4.12C – F). The asialoA1AT and asialoBuChE substrates were also comparably sialylated by HUST-166 produced in both strains (Figures 4.5G and H, S4.15G and H) showing no loss in functionality of the enzyme when produced in *C. glutamicum*. Taken together, these results show that *C. glutamicum* holds a great potential for recombinant protein production with minimal engineering.

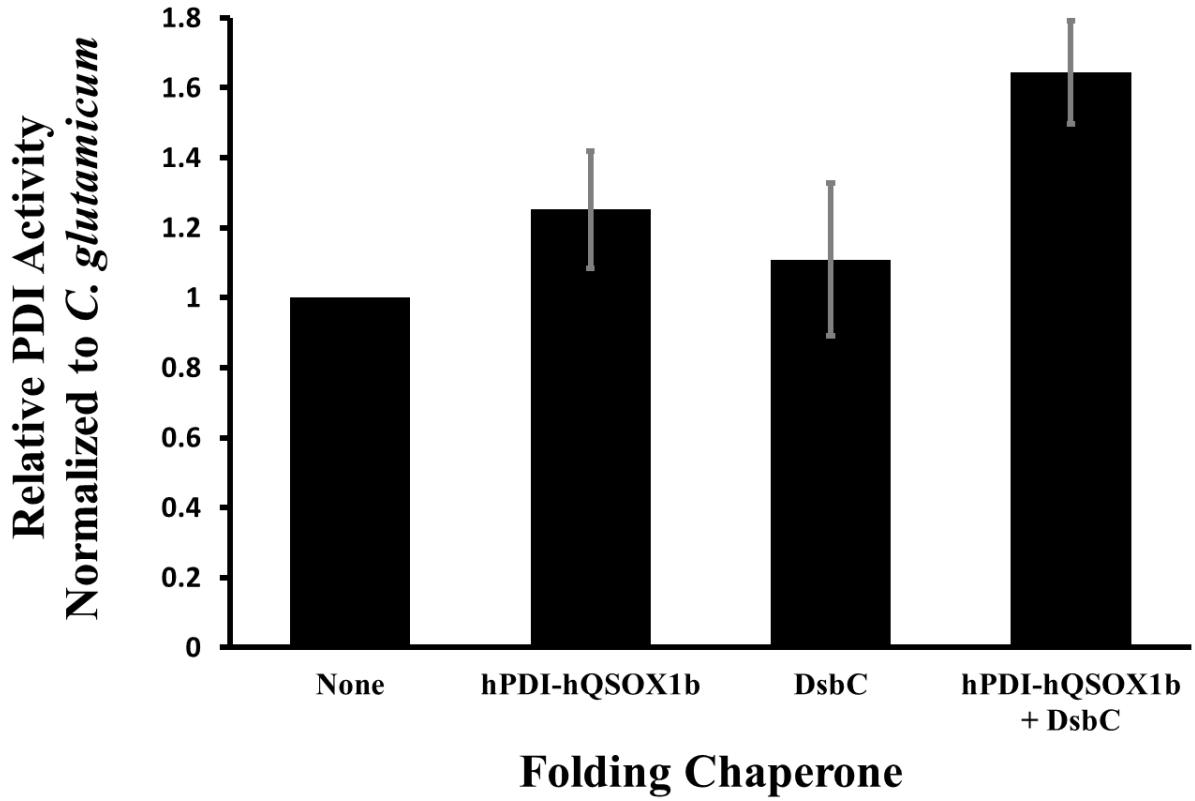


Figure 4. 1. Relative PDI activity of *C. glutamicum* lysates expressing folding chaperones via pDual. constructs.

Compared to *C. glutamicum* ATCC 13032 lysates, co-expression of the hPDI-hQSOX1b fusion and DsbC (via the pDual_4 construct) improved the cytoplasmic PDI activity of the strain by ≈1.6-fold as determined by protein disulfide isomerase (PDI) activity assay kit (Abcam).

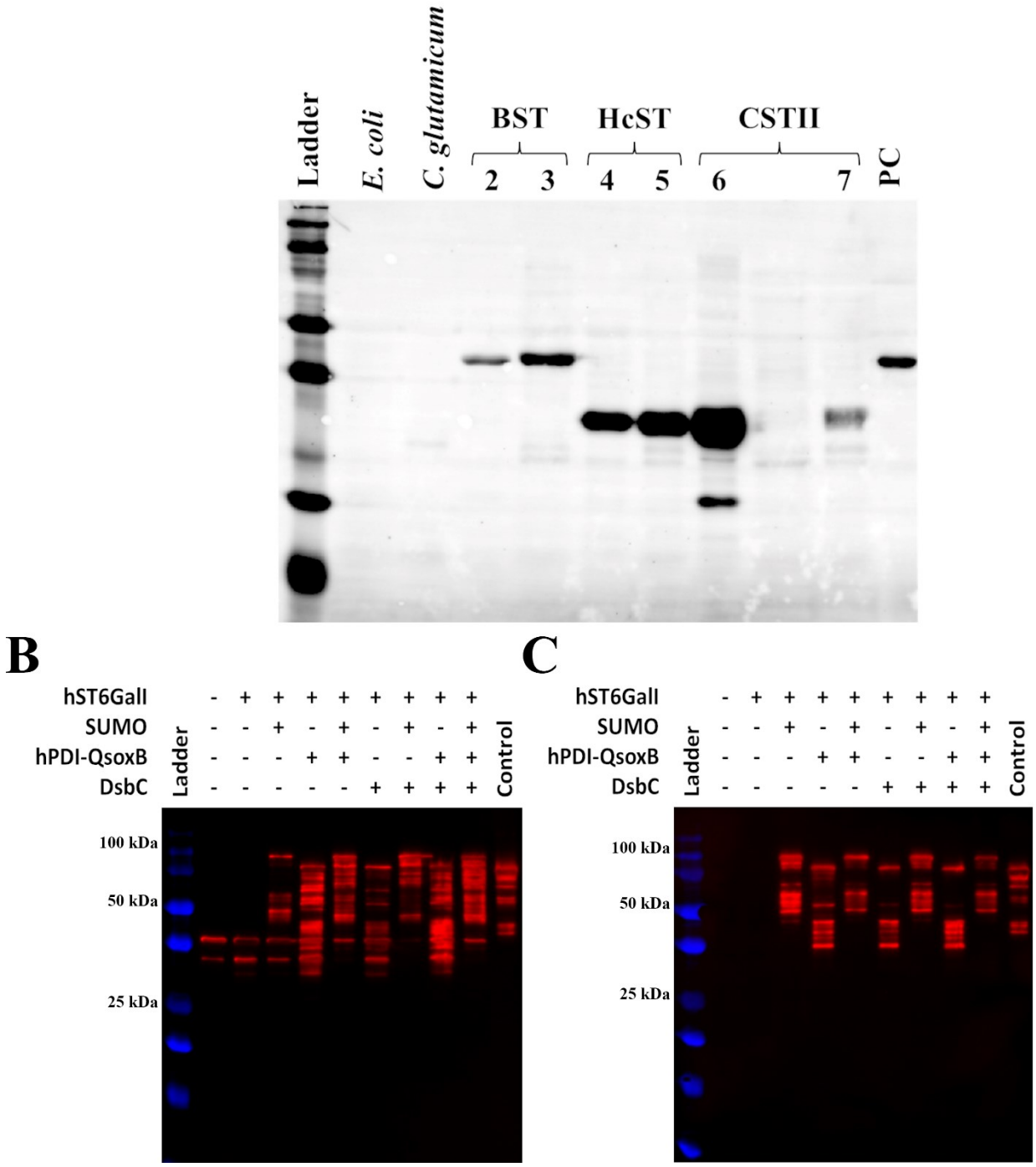


Figure 4. 2. Anti-HIS$_6$-Alexa647 Western blots showing soluble production of each bacterial ST (BST, HcST, CSTII) in BL21 *E. coli* and *C. glutamicum* (A), with anti-MBP-HRP Western blots (red) showing soluble production of each hST6GalI co-expression construct in *E. coli* BL21 (B) and *C. glutamicum* (C).

Production of each bacterial ST in *E. coli* is shown in lanes 1, 3, and 5. *C. glutamicum* production of each bacterial ST is shown in lanes 2, 4, and 6 (A). *C. glutamicum* is not as proficient as *E. coli* at expressing CSTII. The sole expression of the MBP-hST6GalI in both *E. coli* (C, lane 3) and *C. glutamicum* (D, lane 3) produces no soluble ST enzyme. Addition of an N-terminal SUMO tag, or the co-expression of folding chaperones (hPDI-QSOX1B or DsbC) improves the solubility of the produced MBP-hST6GalI in both strains; however, significant degradation is observed in *E. coli* (C). The HUST-166 construct (C and D, lane 9) co-expressing both hPDI-QSOX1B and DsbC and lacking an N-terminal SUMO tag produced the most full-length and active ST enzyme. Control protein is MBP-hST6GalI produced in SHuffle® Express *E. coli*. Expected molecular weights: HcST 32.8 kDa, BST 49.1 kDa, CSTII 31.7 kDa, MBP-hST6GalI 79.6 kDa, SUMO-MBP-hST6GalI 90.6 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.

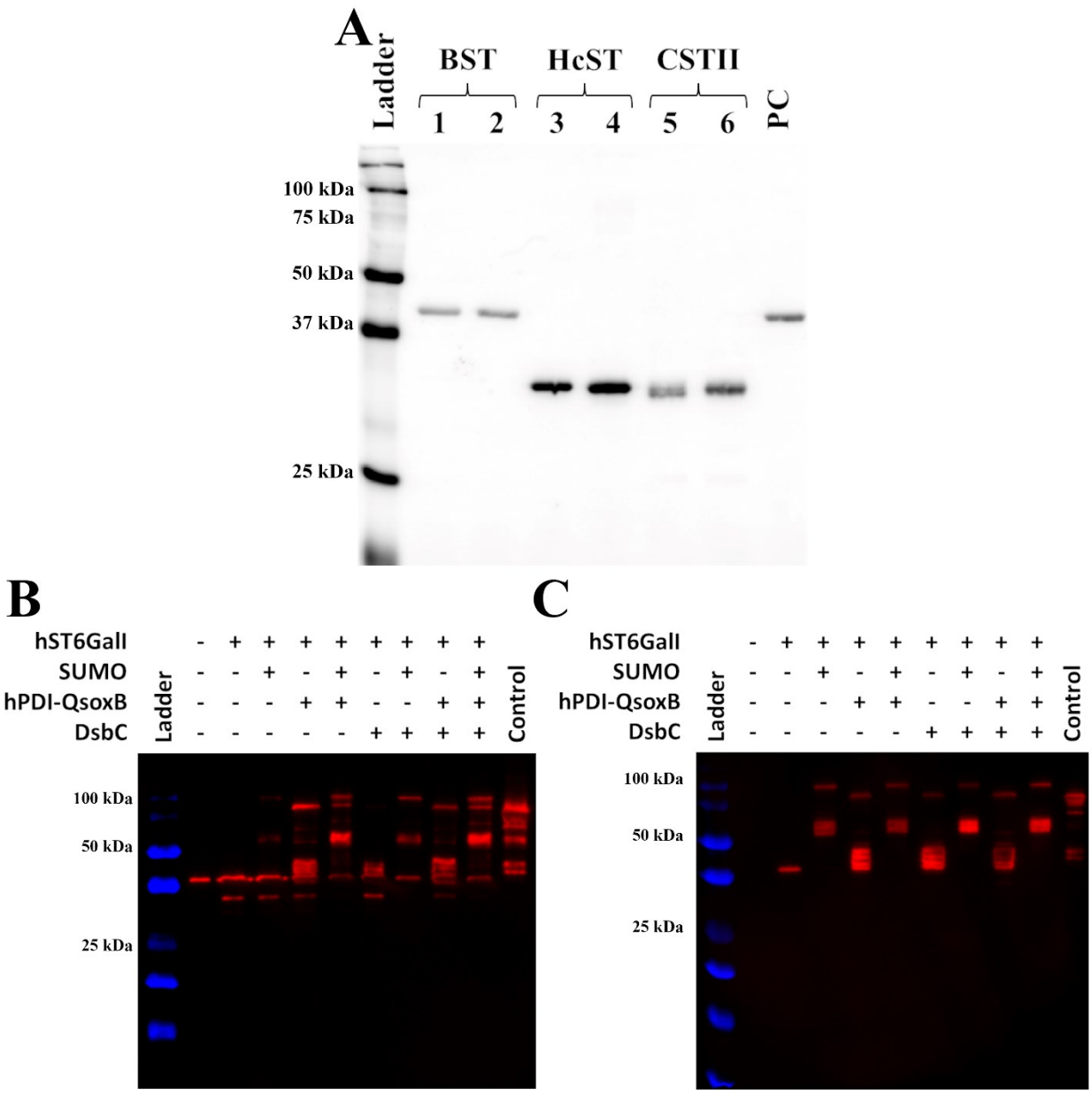


Figure 4. 3. Anti-HIS$_6$-Alexa647 Western blots showing IMAC purified bacterial STs (BST, HcST, CSTII) from BL21 *E. coli* and *C. glutamicum* (A), with anti-MBP-HRP Western blots (red) showing affinity purified hST6GalI co-expression constructs from BL21 *E. coli* (B) and *C. glutamicum* (C).

Each bacterial ST enriched from *E. coli* is shown in lanes 1, 3, and 5. Each bacterial ST enriched from *C. glutamicum* is shown in lanes 2, 4, and 6 (A). All bacterial STs were comparably

enriched from both strains. The sole expression of the MBP-hST6GalI in both *E. coli* (C, lane 3) and *C. glutamicum* (D, lane 3) produces no soluble ST enzyme, only free MBP was recovered. In *E. coli* (C), the addition of an *N*-terminal SUMO tag, or the co-expression of folding chaperones (hPDI-QSOX1B or DsbC) allows for the recovery of full-length MBP-hST6GalI fusion – though significant and variable degradation is observed. In *C. glutamicum* (D,) the degradation patterns are much more consistent between all constructs. The HUST-166 construct (C and D, lane 9) co-expressing both hPDI-QSOX1B and DsbC and lacking an *N*-terminal SUMO tag produced the most full-length and active ST enzyme. Control protein is MBP-hST6GalI produced in SHuffle® Express *E. coli*. Expected molecular weights: HcST 32.8 kDa, BST 49.1 kDa, CSTII 31.7 kDa, MBP-hST6GalI 79.6 kDa, SUMO-MBP-hST6GalI 90.6 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.

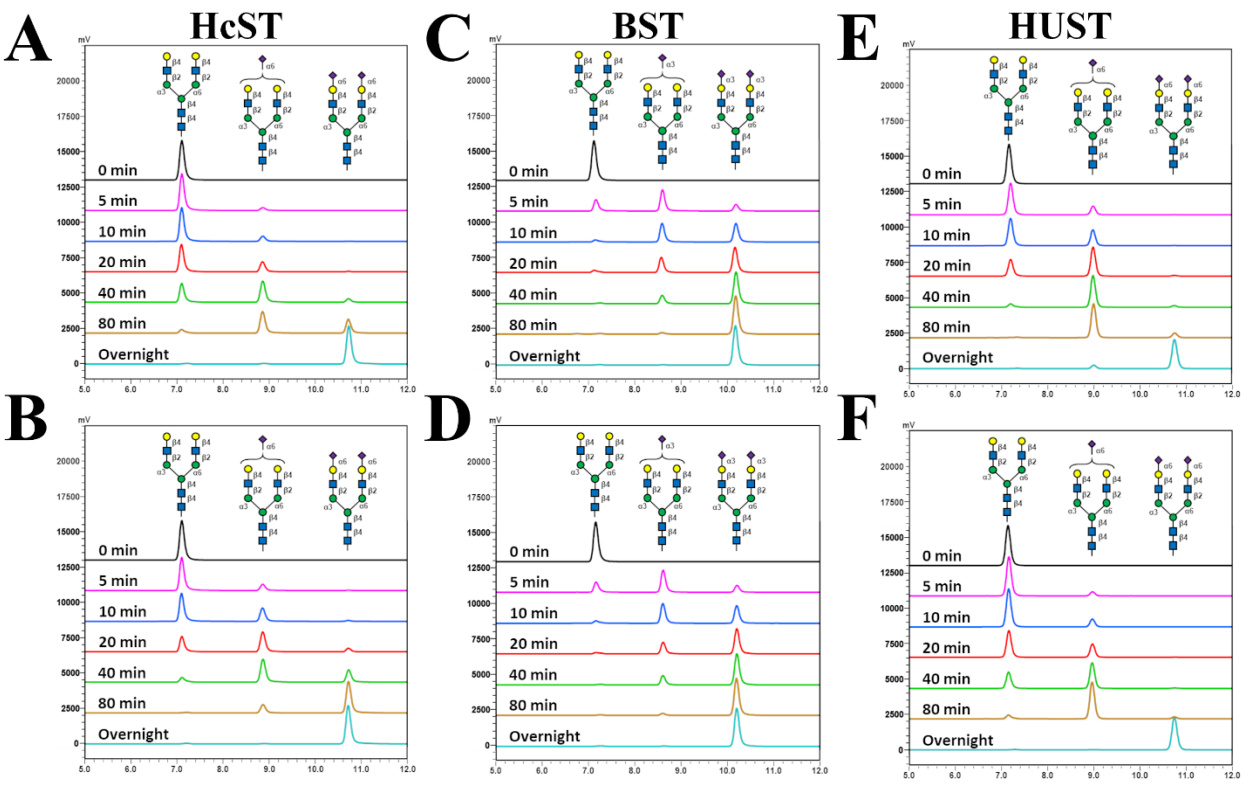


Figure 4. 4. HPLC traces showing G2 *N*-glycan modification by HcST (A and B), BST (C and D), and HUST-166 (E and F) in *E. coli* (top) and *C. glutamicum* (bottom).

No strain dependent changes were observed in enzyme activity or arm preference. The HcST enzyme (A and B) displays a preference for first producing biantennary glycans with one α2,6 sialylated arm. The BST enzyme (C and D) also displays this preference, but to a lesser degree. HUST-166 construct displays a very strong preference for first producing one α2,6 sialylated arm before modifying the remaining arm (E and F).

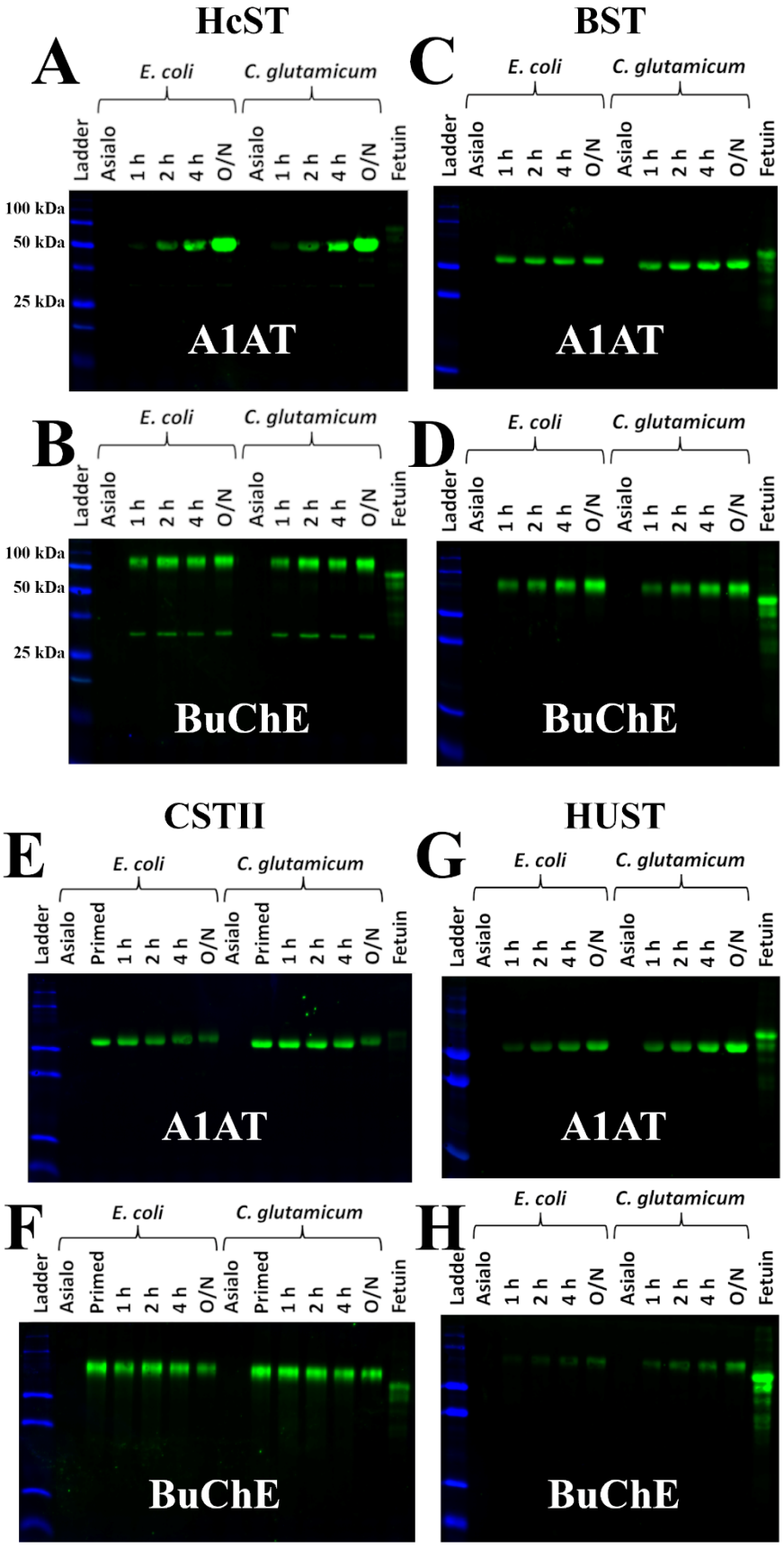

HcST
BST
A
C
B
D
E. coli
C. glutamicum
Ladder
Asialo
1 h
2 h
4 h
O/N
Fetuin
100 kDa
50 kDa
25 kDa
A1AT
BuChE
CSTII
HUST
E
G
F
H
Primed

Figure 4. 5. mNeonGreen-diCBM40 lectin blots comparing *in vitro* sialylation of asialoA1AT (A, C, E, G) and asialoBuChE (B, D, F, H) by STs purified from *E. coli* and *C. glutamicum*.

The HcST enzyme produced in both stains was highly active on both glycoprotein substrates (A and B). The BST enzyme was readily able to modify the *N*-glycans of both asialoA1AT (C) and asialoBuChE (D) with a terminal α2,3 Neu5Ac. The CSTII enzyme requires a terminal α2,3 Neu5Ac, to which it adds another α2,3 Neu5Ac. This addition cannot be monitored via the mNeonGreen-diCBM40 probe, but the subsequent α2,8 Neu5Ac addition can be monitored by molecular weight shift and loss of signal as the probe does not bind α2,8 Neu5Ac. The CSTII enzyme is capable of modifying both target asialoglycoproteins (E and F); molecular weight shifts are more evident with A1AT as it is a discrete band, the heterogenous glycan profile of BuChE results in a more diffuse band where molecular weight shifts are more difficult to discern. The HUST-166 construct (G and H) readily modifies both targets with α2,6 Neu5Ac as glycoproteins are the native substrate of this enzyme. Expected molecular weight of asialoA1AT is 46.7 kDa, asialoBuChE is 68.4 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.

Table 4. 1. Specific activities (nmol/min/mg) of STs produced in *E. coli* and *C. glutamicum* assayed on fluorescent small molecule acceptors.

HUST, like other GT29 enzymes, typically displays poor activity on synthetic substrates as these enzymes prefer protein substrates.

	Specific Activity (nmol/min/mg)	
ST Enzyme (Neu5Ac Linkage)	***E. coli* Produced**	***C. glutamicum* Produced**
HcST (α2,6)	123.6 ± 7.8	121.6 ± 8.7
BST (α2,3)	764.1 ± 77.2	829.1 ± 125.4
CSTII (α2,8)	183.5 ± 25.5	150.7 ± 18.4
HUST (α2,6)	59.9 ± 11.6	45.6 ± 17.3

4. 5. Discussion

4. 5. 1. pCGE-31 production of bacterial STs

Overall, pCGE-31 succeeded in improving the expression levels of recombinant proteins in *C. glutamicum* compared to pTGR-5 by a factor of 3–4x. Utilization of the *sod* RBS was previously shown to drive higher expression of a reporter gene than either the *lacZ* (from *E. coli*) or the *cspB* (from *C. glutamicum*) RBS regions (268). We also verified that the *sod* RBS is well suited to driving recombinant expression in *C. glutamicum* as the T4 page lysozyme RBS native to pCW-MBPT does not appear to drive expression as well in this host.

The *C. glutamicum* expression of both HcST and BST by pCGE-31 showed slightly higher expression levels than in *E. coli* BL21. Unfortunately, the CSTII enzyme does not appear to express very reproducibly in our current *C. glutamicum* expression system. As the other STs in this work express well enough in *C. glutamicum* to produce induction bands visible by Coomassie staining, it is possible that the unique structure of CSTII complicates its production and folding in *C. glutamicum*. Both HcST and CSTII are members of GT42, sharing a modified GT-A type fold (289), while BST is a member of GT80 with a GT-B type fold (281). While CSTII is a tetrameric enzyme whose GT-A fold is characterized by a single α/β Rossmann fold and a flexible lid domain which is composed of a coil and two helices, we do not yet know if HcST shares this quaternary structure.[109]. Though the cytoplasmic chaperones assisting in each GT fold type are currently unknown, *C. glutamicum* does not contain any annotated GT42 enzymes according to the CAZy database. Due to the unique structural nature CSTII may not be folded correctly in *C. glutamicum*, which has been reported to result in rapid degradation in other actinobacteria (166). Interestingly, the improved expression and activity of BST produced in *C.*

glutamicum suggests that the unmodified strain may be slightly more proficient at generating active STs (and GTs) with GT-B type folds.

4. 5. 2. Activity of bacterial STs produced in *C. glutamicum*

BL21 was selected as a "minimally" engineered strain of *E. coli* to better assess the ATCC 13032 strain of *C. glutamicum* for its ability to produce active prokaryotic and eukaryotic STs. For this reason, kinetic data was only utilized as a metric for the performance of *C. glutamicum* as a recombinant host and not to rank the utility of the different STs enzymes for the purpose of *in vitro* sialylation. Moreover, bacterial STs are typically more active on small molecule substrates than they are on glycoprotein substrates as the former more closely resemble their native oligosaccharide substrates. As these STs are intended to be utilized for the sialyation of glycoproteins, this kinetic data only serves as a comparison between production strains. Overall, *C. glutamicum*-produced bacterial STs were comparable to their *E. coli*-produced counterparts, except for CSTII (Table 1). The ability of CSTII to transfer to Neu5Ac-α2,6-LacNac remained intact though, and comparable between both strains (Figure S4.12).

HcST produced in both organisms prefers the unsialylated G2 core *N*-glycan structure as a substrate and will only progress to sialylating the remaining arm once most of this preferred substrate is exhausted. This is likely caused by the first Neu5Ac addition slightly crowding the interaction of the possibly multimeric HcST enzyme with the remaining, unsialylated arm and thereby affecting the kinetics of the reaction. As a GT80 enzyme, BST displays significantly less arm preference than the other monofunctional STs assayed – likely due to its monomeric GT-B type fold. Though both strains produced all bacterial STs with comparable activities on asialoglycoproteins (or α2,3/α2,6 primed glycoproteins for CSTII), BST was surprisingly active

on glycoprotein substrates (Figure 4.5). Previous investigations of GT80 members – like the α2,6 ST of *Photobacterium damsela* (85) – showed very weak activity on glycoprotein substrates, making BST a very interesting bacterial α2,3 ST for the potential application of producing sialylated therapeutics.

4. 5. 3. pDual production of hST6GalI in *C. glutamicum*

Following generation of the intermediate pDual_1 construct (Figure S4.5C), functionality of the secondary expression cassette was verified using mRuby3 fluorescence in response to the addition of ATc to the culture (data not shown). Expression of the hPDI-QSOX1b fusion in *C. glutamicum* increased the PDI activity (measured *via* the reductase activity of hPDI) within lysates by ≈1.2-fold in comparison to the ATCC 13032 strain (Figure 4.1). The expression of DsbC however, did not significantly alter the PDI activity found in *C. glutamicum* lysates as this prokaryotic enzyme functions as a proofreader – breaking non-native, improperly paired disulfide bonds (275, 276). Interestingly, expression of both folding chaperones displayed a synergistic effect on the PDI activity present within the *C. glutamicum* lysate, increasing the PDI activity by ≈1.6-fold compared to the ATCC 13032 strain. As anti-SUMO Western blotting of pDual_2 and pDual_4 showed no significant changes to the expression levels or degradation profiles of the hPDI-QSOX1b fusion (data not shown), it is likely that the proofreading capability of DsbC in pDual_4 increases the proportion of active hPDI-QSOX1b with correctly paired disulfide bonds.

While not the ideal choice for protein folding and disulfide formation, the minimally engineered *E. coli* BL21 strain was chosen as a reasonable comparison to the more naïve expression host *C. glutamicum*. Many commercially available strains of *E. coli* – like SHuffle®

from NEB – commonly utilized for the expression of multiple disulfide bond containing proteins - have had interfering reductive pathways removed to boost the oxidative potential within the cytoplasm of the cell (in addition to other engineering) and these strains have been successfully utilized to produce hST6GalI as an MBP fusion at the cost of endotoxin contamination (85). The *E. coli* genes encoding glutaredoxin (*gor*) and thioredoxin reductase (*trxB*) are common targets for deletion to this end (273, 274). *C. glutamicum* harbours homologous proteins that play a role in protecting the cell from oxidative stress, mycoredoxin (WP_011013923) and the thioredoxin-disulfide reductase (WP_003855300) – both of which serve as potential targets to improve the oxidative potential and folding of disulfides in the cytoplasm of *C. glutamicum.*

In *C. glutamicum*, the simple addition of an *N*-terminal SUMO tag dramatically increases the amount of active MBP-ST6GalI that is found in the soluble fraction of lysates (Figure 4.2C, lane 4); this effect has been reported previously (85, 290). Notably, in *E. coli* the N-terminal SUMO tag seems prone to cleavage leaving the MBP-hST6GalI fusion mostly intact. In both strains, utilization of one or both folding chaperones improves the proportion of soluble enzyme produced; less dramatic effects were observed by DsbC alone. In terms of soluble protein expression, the effects of the SUMO solubility tag and folding chaperones are much more consistent across the *C. glutamicum* constructs than in *E. coli*. Moreover, in *C. glutamicum*, the recombinant MBP-hST6GalI construct appears to be significantly less susceptible to proteolysis. This is likely because of a limited number of cytoplasmic proteases identified to date in *C. glutamicum* (291–294) – some of which could be targets for the continued engineering of the strain. These results show that while it is a naïve protein expression platform, *C. glutamicum* harbours great potential for recombinant protein expression.

Soluble MBP-hST6GalI could be enriched from all constructs in both strains (except when expressed on its own), and the material enriched from *C. glutamicum* was less degraded overall and of higher purity (Figures 4.3B and C, S4.7B and C). While C-terminal degradation is evident across all constructs in both strains, it should be noted that a similar pattern is evident when the MBP-hST6GalI is produced in highly engineered protein expression strains of *E. coli* like SHuffle® and SHuffle® Express (NEB). While the addition of a SUMO tag generally improved the total yield of MBP-hST6GalI produced in *C. glutamicum* constructs (Figure S4.10), total yields of SUMO-MBP-hST6GalI were lower when co-expressed with DsbC. This could indicate that the small, 11-kDa SUMO tag may occlude some of the interactions between DsbC and the MBP-hST6GalI required for full solubility and/or proper folding. Overall, co-expression of the hPDI-QSOX1b fusion and DsbC displayed the greatest yield of recoverable MBP-hST6GalI (HUST-166) in both strains. Comparison of activities between the different hST6GalI constructs on small molecule substrates showed that in both strains, the HUST-166 construct produced enzyme of the highest specific activity (Figure S4.13). Moreover, HUST-166 produced in both organisms showed no significant difference in its strong preference for first modifying one arm of biantennary *N*-glycans (Figure 4.4E and F) or its ability to modify asialoglycoproteins (Figure 4.5G and H). This is the first report of *C. glutamicum* producing an active ST of eukaryotic origin. As *C. glutamicum* is regarded as endotoxin-free (272), glycosyltransferases produced in the strain would be ideally suited for the *in-vitro* glycan modification of therapeutic glycoproteins – improving their efficacy and circulating half-lives at significantly reduced costs (74, 215, 218).

The synergistic effects of hPDI-QSOX1b and DsbC on the solubility and activity of hST6GalI produced in these prokaryotic systems is likely due to disulfide bond proofreading

functionality on both the hST6GalI and the folding chaperones themselves. Eukaryotic proteins often contain multiple disulfide bonds and as prokaryotes typically do not produce disulfide bonds in their cytoplasm, heterologous production can often result in improperly folded and/or totally insoluble proteins (295). As DsbC is of prokaryotic origin, the prokaryotic hosts are capable of producing this chaperone (Figure S4.6B, lanes 7 and 8), however, both components of the hPDI-QSOX1b fusion are of eukaryotic origin. The expression of hPDI-QSOX1b alone in both hosts produces a nearly equivalent ratio of both soluble and insoluble fusion protein; the addition of DsbC pushes this ratio to favour the soluble form (data not shown). This suggests that DsbC not only proofreads disulfide bonds in hST6GalI, but also in hPDI-QSOX1b, producing more soluble and active chaperone which is congruent with the PDI activity of *C. glutamicum* lysates expressing both chaperones (Figure 4.1). The greater proportion of active hPDI-QSOX1b is then also capable of assisting in the correct formation and proofreading of disulfide bonds in the hST6GalI fusion, increasing soluble yield and specific activity. Moving forward, this platform will be applied to the production of other complex mammalian enzymes and potentially therapeutics protein targets as well.

4. 5. 4. Conclusion

As previously stated, *C. glutamicum* is currently a naïve host for recombinant enzyme production. The data presented here shows promise for the strain as a recombinant host, as, with minimal engineering, it was able to produce the human ST6GalI – a mammalian enzyme normally difficult to produce in prokaryotic systems – with both comparable yield and activity to the traditional *E. coli* host. In concert with its endotoxin-free status (270) and the capability of true secretion from the cell into the culture media (271) facilitated by both SEC and TAT

secretion pathways, *C. glutamicum* could be adapted into an industrial recombinant protein production platform with additional engineering to tailor the strain for the application.

4. 6. Materials and Methods

4. 6. 1. Media, strains, and expression conditions

All strains were grown in 2YT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, BioShop Canada). NEB® Stable *E. coli* (NEB) was used for routine cloning and plasmid production. BL21 (DE3) *E. coli* and *C. glutamicum* ATCC13032 were used for recombinant protein production. Electrocompetent *C. glutamicum* were cultured in MBGT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 35 g/L glycine, 0.1% Tween-80) and outgrowths were performed in 2YT + 91 g/L sorbitol.

Single colonies of expression constructs (in both *E. coli* and *C. glutamicum*) were inoculated into 25 mL 2YT containing 50 µg/mL kanamycin (*C. glutamicum* cultures also contained 25 µg/mL nalidixic acid) and incubated overnight at 37°C (*E. coli*) or 30°C (*C. glutamicum*) with shaking at 180 RPM. On the following day, overnight cultures were diluted into 250 mL 2YT expression cultures (containing the appropriate antibiotics) to an $OD_{600} \approx 0.1$ and incubated at 30°C, 180 RPM. Cultures were induced with 0.5 mM IPTG (and 0.5 µg/mL ATc for pDual constructs) once they reached an $OD_{600} \approx 0.6$ ($OD_{600} \approx 0.45$ for pDual constructs) and allowed to express for 16h at 30°C (25°C for pDual constructs). Following induction, cultures were harvested by centrifugation at 5,000 x *g* for 10 mins at 4°C.

4. 6. 2. Vector design and ST construct cloning

The *E. coli/C. glutamicum* shuttle vector pTGR-5 (268) was received as a generous gift from Dr. Pablo Ravasi. To generate the high-level expression plasmid, pCGE-31, the P_{tac} region in pTGR-5 (XbaI – NheI fragment) was replaced with the lac UV5 + tandem P_{lac} promoter system from pCW-MBPT utilizing synthetic primers and maintaining the *sod* RBS and spacing already present in pTGR-5. *E. coli* codon optimized bacterial ST synthetic genes (Bleu Heron, BioBasic, IDT) were cloned into the MCS of pCGE-31 replacing the eGFP gene utilizing NdeI and HindIII (NEB).

The pDual co-expression constructs were constructed using the desired $P_{tetR/tetA}$ expression cassette of pRG_Duet1 and pRG_Duet2 (269), which were generous gifts from Dr. Bernhard Eikmanns. pDual_1 was constructed using a synthetic gene composed of $P_{tetR/tetA}$ upstream of the *sod* RBS, mRuby3 (flanked by NsiI and SacI restriction sites), rrnB T1 terminator, and the T7Te terminator was generated (IDT) flanked by PstI restriction sites for insertion into pCGE-31. The mRuby3 gene was replaced by a SUMO tagged hPDI (18 aa – 508 aa) hQSOX1b (30 aa – 604 aa) fusion, the *E. coli* DsbC, or an operon containing both to generate pDual_2, pDual_3, and pDual_4, respectively. MBP- and SUMO-MBP-hST6GalI fusions were independently inserted into each expression vector (including pCGE-31) utilizing NheI and AvrII restriction sites. All expression constructs and intermediates were confirmed via restriction digest, sequencing, and confirmation of sialyltransferase activity where appropriate.

4. 6. 3. Electrocompetent *C. glutamicum* cells

A single colony of *C. glutamicum* ATCC13032 was inoculated into 50 mL 2YT containing 25 µg/mL nalidixic acid and incubated overnight at 30°C with shaking at 180 RPM.

The following day, 1 L MBGT containing 25 µg/mL nalidixic acid was inoculated to an $OD_{600} \approx 0.1$ using the overnight culture and incubated at 30°C, 180 RPM. When the $OD_{600} \approx 0.25 – 0.25$ (about 2 hours) 0.5 µg/mL ampicillin was added and the culture was allowed to continue incubating at 30°C, 180 RPM for an additional 1.5 hours.

Following incubation, cells were harvested at 5,000 x *g* for 10 minutes at 4°C. Cells were resuspended in 150 mL 10% glycerol and centrifuged at 5,000 x *g* for 10 minutes at 4°C a total of 3 times. Final cell pellets were resuspended in 10% glycerol to a final $OD_{600} \approx 200$ and aliquots of 100 µL were stored at -80°C.

Aliquots of competent cells were thawed on ice and allowed to incubate with 750 ng DNA for 10 mins prior to transformation. Cells and DNA were electroporated in 0.2 cm cuvettes using a BioRad Gene Pulser Mini at 2.5 kV for 4.80–5.20 ms. Immediately following, 1 mL 2YT + 91 g/L sorbitol was added to cells and outgrowths were placed at 46°C for 6 minutes to inactivate the host restriction system and increase transformation efficiency. Recovery continued at 30°C, 180 RPM for 2 hours. Cells were harvested via centrifugation at 5,000 x *g* for 1 min and resuspended in 200 µL fresh outgrowth medium prior to plating on agar containing 50 µg/mL kanamycin and 25 µg/mL nalidixic acid. Plates were incubated at 30°C for 48–72 hours until colonies appeared.

All constructs in *C. glutamicum* were confirmed via plasmid rescue in *E. coli*. Miniprepped plasmid DNA from *C. glutamicum* constructs was transformed into electrocompetent *E. coli* for propagation, then confirmed by restriction digest.

4. 6. 4. PDI activity assay of *C. glutamicum* expressing folding chaperone constructs

Relative PDI activity in lysates of *C. glutamicum* expressing the hPDI-QSOX1b chimera, DsbC, or both were assessed using the Fluorometric Protein Disulfide Isomerase (PDI) Activity Assay Kit (Abcam) and compared to the ATCC 13032 strain. 5 µg of each lysate were assayed following the manufacturer recommended protocol and read on a (BioTek) plate reader, measuring fluorescence intensity (Ex/Em=490/580 nm) every 5 mins for 1 hour.

4. 6. 5. Purification of recombinantly produced STs

Cell pellets were resuspended in HisTrap A Buffer (100 mM HEPES, 300 mM NaCl, pH 8.0) for bacterial STs or MBPTrap A Buffer (100 mM HEPES, 150 mM NaCl, pH 7.4) for hST6GalI constructs with 1 mM PMSF to a final concentration of 1 g cell paste/10 mL buffer. Resuspended cells were lysed mechanically via a single pass through an Avestin C3 homogenizer at ≈ 20,000 PSI. Insoluble material and cell debris were removed from the lysate by centrifugation at 20,000 x *g* for 30 minutes at 4°C and soluble aggregates were removed by ultracentrifugation at 100,000 x *g* for 1 hour at 4°C. Prior to purification, clarified lysates were passed through a 0.45 µm syringe filter.

All recombinant STs were enriched via affinity chromatography using either Roche cOmplete™ His-Tag Purification Resin (Millipore-Sigma) for bacterial STs or Dextrin Sepharose® High Performance (Cytiva) for hST6GalI constructs and AKTA Start (Cytiva) FPLC. Clarified lysates were loaded onto an equilibrated XK-16 column (Cytiva) containing 15 mL of the appropriate resin with either HisTrap A or MBPTrap A at a flowrate of 2 mL/min. The column was washed with 3 CV of HisTrapA or MBPTrap A prior to elution along a linear gradient (0–100%) of HisTrap B (100 mM HEPES, 300 mM NaCl, 500 mM imidazole, pH 8.0)

or MBPTrap B (100 mM HEPES, 150 mM NaCl, pH 7.4) over 5 CV. Fractions containing recombinant STs were pooled, concentrated, and buffer exchanged into 50 mM HEPES, 150 mM NaCl, pH 7.4 using 20 mL VivaSpin concentrators (Cytiva).

4. 6. 6. ST activity on small molecule substrates and glycoproteins

All recombinant STs were assayed for activity using *N*-acetyllactosamine (LacNAc) or Neu5Ac-α2,3/α2,6-*N*-acetyllactosamine (α2,3/α2,6-sialyl-LacNAc) coupled to the BODIPY (BDP) fluorophore (283). Small molecule reactions contained 0.05 mg/mL enzyme, 0.5 mM BDP-LacNAc or BDP-α2,3/α2,6-sialyl-LacNAc, 50 mM HEPES pH 7.4, 10 mM $MgCl_2$, 50 mU/mL alkaline phosphatase, and 2.0 mM CMP-Neu5Ac (with 0.01% Triton X-100 for hST6GalI constructs). Reactions were incubated at 30°C and stopped either by spotting 1 µL onto silica TLC plates or by the addition of 1:99 (v/v) stop solution (100% ACN). TLC plates were developed in a solvent mixture of 4:2:1:0.1 (EtOAc:MeOH:H_2O:HOAc), while 5 pmol of stopped small molecule reactions were separated on an Accucore™ C18 (Thermo Scientific) column in 50 mM ammonium acetate (pH 4.5) at 0.6 mL/min over an increasing gradient of ACN (20 – 40%) with fluorescence monitored at 503/514 nm Ex/Em. Assay conditions using glycoprotein substrates terminally de-sialylated (asialo) and sialylated A1AT and BuChE were performed as previously described (85).

4. 6. 7. Bacterial ST activity on free *N*-glycans

Recombinantly produced bacterial STs were assayed for activity using G2-EPAB (NatGlycan) as a substrate. Reactions contained 0.1 mg/mL enzyme, 0.5 mM G2-EPAB, 50 mM HEPES pH 7.4, 10 mM $MgCl_2$, and 2.0 mM CMP-Neu5Ac (with 0.01% Triton X-100 for

hST6GalI constructs). Reactions were incubated at 30 °C and stopped by the addition of 1:39 (v/v) stop solution (50% ACN). 50 pmol of modified and unmodified *N*-glycan analogues were separated on a GlycanPac™ (Thermo Scientific) column in 50 mM ammonium formate (pH 4.5) at 1.2 mL/min over a decreasing gradient of ACN (75 – 25%) with fluorescence monitored at 289/375 nm Ex/Em.

4. 6. 8. Generation of mNeonGreen-diCBM40 for detecting terminal α-2,3 and α-2,6 Neu5Ac

An *E. coli* codon optimized sequence of the *C. perfringens* diCBM40_NanI described by Ribeiro et al. (2016) was obtained as a synthetic gene from IDT and inserted into pCW using EcoRI and SalI restriction sites downstream of a 6xHis-mNeonGreen coding sequence. The mNeonGreen-diCBM40 was expressed in *E. coli* AD202 incubated with shaking in 2YT medium containing 150 µg/mL ampicillin at 37°C, 180 RPM. Cells were induced at $OD_{600} \approx 0.6$ with 0.5 mM IPTG and further incubated overnight at 25°C, 180 RPM. Harvested cell pellets were lysed and purified as described for the recombinant STs.

4. 6. 9. Western and lectin blotting

Western and lectin blots were performed as described in the manufacturer's data sheets. Briefly, the proteins of interest were separated on a 12% SDS PAGE gel using a MiniProtean system (Bio-Rad), which was then rinsed three times for 5 minutes each in excess Tris buffered saline pH 7.6 (TBS, 50 mM Tris, 150 mM NaCl, pH 7.6) before being blotted to PVDF membrane using a Trans-Blot Turbo transfer system (Bio-Rad). The transfer was performed in 48 mM Tris, 39 mM glycine using 2.5 A, 25 V for 8 mins. The protein-bearing PVDF membrane was then rinsed three times for 5 minutes each, in TBS and blocked for one hour in 5% BSA in

TBS at room temp. The blocked membrane was washed three times for 5 minutes in TBS at room temp, then incubated overnight at 4°C with respective antibody or lectin: 0.05 µg/mL Anti-MBP-HRP (NEB) in TBS, 1:10, 000 AlexaFluor 647 anti-HIS6 (Bio-Rad) in TBS, or 15 µg/mL mNeonGreen-diCBM40. The membrane then underwent three 10-minute washes in TBS at room temp and was visualized on a Bio-Rad ChemiDoc. Molecular weight markers were Bio-Rad Precision Plus All Blue standards.

4. 7. Acknowledgments

Funding NSERC alliance grant ALLRP 550399-20. The recombinant BuChE protein was a gift of PlantForm Corporation, Toronto Ontario.

5. FUTURE DIRECTIONS

5. 1. Dissecting actinobacterial POM

Attempting to assign a singular function to POM in actinobacteria may prove to be a futile endeavor as this modification has been identified on proteins with a wide variety of functions not easily grouped into an all-encompassing category. At the crux of the issue is the fact that the reported actinobacterial GT-39 deficient mutants display no discernable phenotype when grown under laboratory conditions, and only the virulence of pathogenic strains seems to be affected by the loss of POM (49, 140, 176). While interesting and important in their own right, these results and conclusions only create more questions – if the primary function of POM is to facilitate the virulence and immune evasion, what function does this modification serve to the non-pathogenic actinobacteria that perform it?

Firsthand biochemical data from the characterization of an actinobacterial GT-39 would be of great use towards understanding this process in prokaryotes. Utilizing this data and comparing it to the breadth of knowledge that has been reported for eukaryotic GT-39s would finally provide insight into how related these enzymes truly are across domains. However, the difficulty lies in both the production and recovery of active enzyme produced in a recombinant host (164, 165). This work has taken the first steps in this regard, by at least achieving detectable POM from recombinantly produced GT-39s in a POM deficient strain of *C. glutamicum*. As no recombinant PMTs were able to be directly detected, a primary focus should involve the further optimization of *C. glutamicum* for use as a recombinant host – potential avenues towards this goal will be discussed later. Following this, experiments elucidating the actual membrane topology of these enzymes (such as cystine or alanine scanning mutagenesis) would be able to distinguish the orientation of GT-39s in the actinobacterial membrane (296, 297); whether they

are inserted into the membrane with the active site on the extracellular face, luminal face, or a mixture of the two. These results would then logically lead to dissecting the process that leads to POM on cytoplasmic proteins – is there a mechanism than exports unmodified proteins before importing them, or is there a cytoplasmic POM pathway that utilizes GT-39s imbedded in the membrane so that their active sites are luminal? Recently, nuclear mannoproteins were identified in *S. cerevisiae,* with *O*-man glycosites overlapping known *O*-GlcNac glycosites and implicating that both forms of *O*-glycosylation may serve similar or overlapping biological functions in the organism (298). Investigating the presence and biosynthesis of cytoplasmic mannoproteins in actinobacteria may provide similar insights towards the possible functions of actinobacterial POM.

The identification of different activities and substrate specificities of actinobacterial GT-39s assayed has reinforced that amino acid sequence similarity does not necessarily tell the whole story, especially in the case of enzyme function. The GT-39s investigated here are all more closely related to each other than to their closest eukaryotic counterparts – if structural or functional differences exist between the actinobacterial enzymes, how significant might their differences be to the eukaryotic ones? Additional actinobacterial mannoproteins must first be identified then subsequently expressed alongside the various GT-39s in *C. glutamicum* to gain further insight into the different specificities of these enzymes. High throughput protein expression methodologies may potentially derive enough data to determine individual consensus sequences for the GT-39 enzymes, if not actinobacterial POM in general.

5. 2. Non-SEC/TAT secretion in actinobacteria

There is a great need for the continued investigation of prokaryotic POM, as these organisms have provided evidence in exception to the dogma that POM must occur in a SEC dependent manner (51, 52). Numerous non-SEC secreted mannoproteins have been identified in actinobacteria (111, 163, 188, 192), including the novel class of non-classically secreted proteins – a class of proteins lacking a traditional signal peptide or known secretion motifs that are predicted to be found in the cytoplasm but are found extracellularly (167). Though the study of non-classical secretion in bacteria is steadily growing, with reports in *Bacillus*, *Listeria*, *Staphylococcus*, *Streptococcus*, *Mycobacterium*, and others (163, 188–192), the relationship between this novel secretion system and protein glycosylation in bacteria is under-reported.

This work has shown that actinobacteria are capable of POM that is independent of SEC secretion, as is the case with the TAT secreted GH6 Celf_3184 and several cytoplasmic proteins in *C. glutamicum*, *C. fimi*, and *C. flavigena* (Figure 5.1). The identification of cytoplasmic mannoproteins has also been reported in *M. tuberculosis* and *M. smegmatis* (93). Most interestingly, the current work has provided evidence that the secreted mannoprotein Celf_2022 – which lacks a strong SEC or TAT leader sequence – is not secreted *via* either the SEC or TAT translocon when produced recombinantly in *C. glutamicum*. This finding does not necessarily suggest the presence of a novel secretion system unique to some actinobacteria functioning in concert with POM; it rather serves as evidence that one or more of the identified secretion pathways in *C. glutamicum* (other than SEC or TAT) are intimately involved in the POM of the mannosylated secretome. Moreover, utilizing this unknown "Other" leader sequence to secrete recombinant Celf_3184 produced in the organism altered its mannosylation profile, making it more uniform. This data implicates at least some interaction between the unknown translocon

and the *C. glutamicum* GT-39 – potentially indicating another secretion pathway involved with POM outside of the traditional SEC dependent dogma. As the SEC translocon is an essential pathway required by many bacteria (186), proving that Celf_2022 is truly exported in a SEC independent manner would be difficult via the traditional knockout methodology. However, proximity labelling experiments using the novel leader sequence would be simple to attempt and may reveal chaperones or other proteins that would assist in the identification of this currently unknown translocon. In addition, CRISPR interference (Figure 5. 2) systems have recently been used for the targeted knockdown of specific genes in *C. glutamicum*; utilizing this methodology would definitively rule out the SEC and TAT pathways and could facilitate the identification of this currently unknown secretion pathway (299–301).

5. 3. Engineering *C. glutamicum* as a recombinant host

The ATCC 13032 strain of *C. glutamicum* has proven to be remarkably adept at producing heterologous enzymes in high quantities. However, reliability is currently the platform's greatest hindrance. Though widely utilized industrially, *C. glutamicum* has only begun to be explored for its use as a recombinant host. There are many pitfalls associated with recombinant protein expression and as such, many reports of failed expression in the organism go unexplained (302). Codon optimization, the G+C content of recombinant genes, promoter strength, native proteases, folding chaperones or cofactors, and even mRNA decay all dramatically impact protein expression in even highly engineered strains of *E. coli* – strains tailored for protein expression and ease of use. As this work has shown, considerations like these are even more critical when attempting to produce recombinant enzymes in a naïve host system.

To further develop *C. glutamicum* as a recombinant expression platform and generally improve its reliability, steps similar to those taken with *E. coli* strains tailored for protein production can be applied. To improve the production of disulfide bond-containing proteins, the *C. glutamicum* enzymes mycoredoxin (WP_011013923) and thioredoxin-disulfide reductase (WP_003855300) in addition to genes encoding other anti-oxidative enzymes or gene products responsible for their expression can be knocked out (274, 303). In addition, the non-essential alternative sigma factor of *C. glutamicum* SigB has been shown to be involved in the transition from exponential growth to stationary phase – knockout of this gene could further decrease the expression of genes encoding anti-oxidative enzymes which hinder cytoplasmic disulfide bond formation (304). PBPs responsible for the crosslinking and remodelling of peptidoglycan can also be knocked out to facilitate secretion of proteins from the cell, such as PBP1a (305). Most critically, native *C. glutamicum* proteases like those belonging to the Clp family can be knocked out in addition to the overexpression of native folding chaperones to aid in the solubility of recombinantly produced proteins (194, 291, 302). Finally, cell-free protein synthesis (CFPS) methods (Figure 5. 3) have rapidly advanced with current methodologies allowing for rapid and high-yield protein production from cell extracts (306). This system has recently been applied to and optimized in *C. glutamicum* (307); used in concert with liposomes, CFPS has been shown to improve the production and yield of some membrane proteins (308).

C. glutamicum can achieve true secretion of proteins into the culture media by both SEC and TAT secretion pathways. As the TAT translocon secretes proteins in a fully folded state and accommodates proteins of highly varied molecular weights (125, 185); the utilization of cytoplasmic folding chaperones and TAT secretion in concert could allow for the production and simple recovery of enzymes containing multiple disulfide bonds. In addition to the heterologous

expression of folding chaperones from other organisms, the utilization of chaperones native to *C. glutamicum* has been shown to improve the overall yield of recombinant protein production and the overall reliability of the platform (302, 309).

5. 4. *In vitro* sialylation of therapeutic proteins

One of the greatest assets to enzyme production in *C. glutamicum* is its status as low-endotoxin or endotoxin-free (193, 271, 272). Utilizing STs produced in the organism for the purpose of sialylating therapeutic glycoproteins translates to decreased costs overall. This is because the final product will be comparatively free of the endotoxin contamination that is all too common with proteins and enzymes produced in *E. coli*. The Wakarchuk lab already has an enzyme immobilization strategy under development which would be easily scalable in an industrial setting, where ST enzymes are produced as fusions with carbohydrate binding modules (CBMs) – typically binding cellulose (Figure 5. 4). Utilized in concert with *C. glutamicum* based expression, these CBM-ST fusions can then be immobilized onto spherical particles of cellulose which can be packed into a chromatography column, much like an affinity chromatography resin. Target glycoproteins to be sialylated (or glycosylated by any other GT produced as a CBM fusion) can then be passed through the column with the required donor sugars. The sialylated glycoproteins eluted from the column would only require minor polishing and be free of endotoxin contamination, further decreasing costs and displaying improved and more consistent sialylation levels compared to many therapeutic glycoproteins produced in mammalian cell lines.

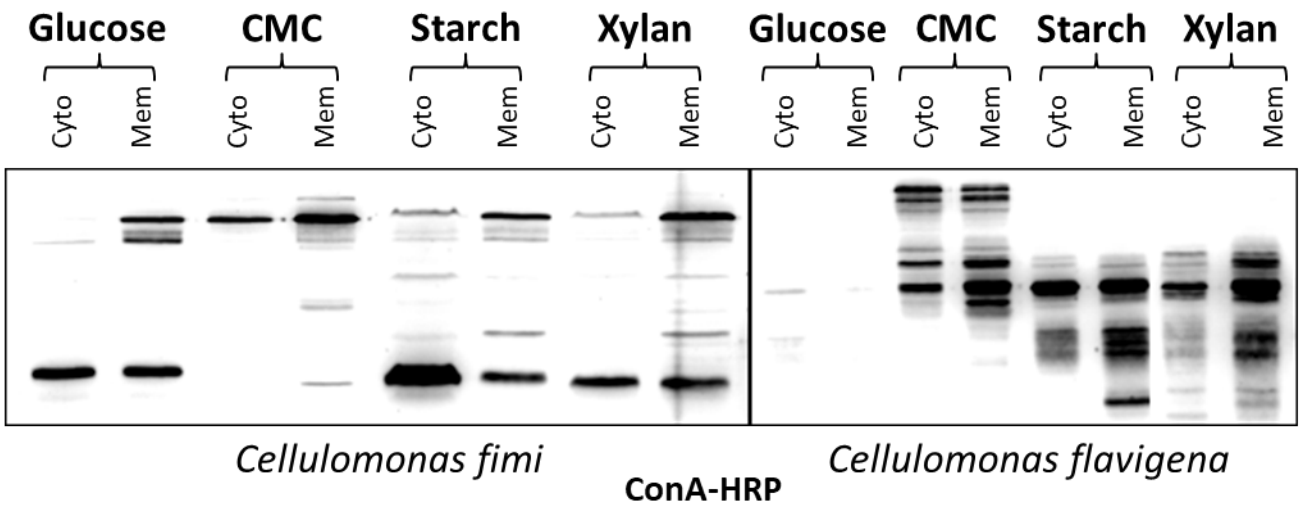


Figure 5. 1. ConA-HRP lectin blot of *C. fimi* and *C. flavigena* mannoproteins based on cellular localization and carbon source.

In *C. fimi* and *C. flavigena*, differential mannoprotein profiles are observed in both cytoplasmic and membrane fractions based on the carbon source utilized. The identification of numerous cytoplasmic mannoproteins in these bacteria is incongruent with the current understanding of bacterial *O*-mannosylation, which is believed to be Sec-dependent. These findings provided the initial rationale towards dissecting the process in actinobacteria. Carbon sources: glucose, carboxymethylcellulose (CMC), starch, soluble beechwood xylan.

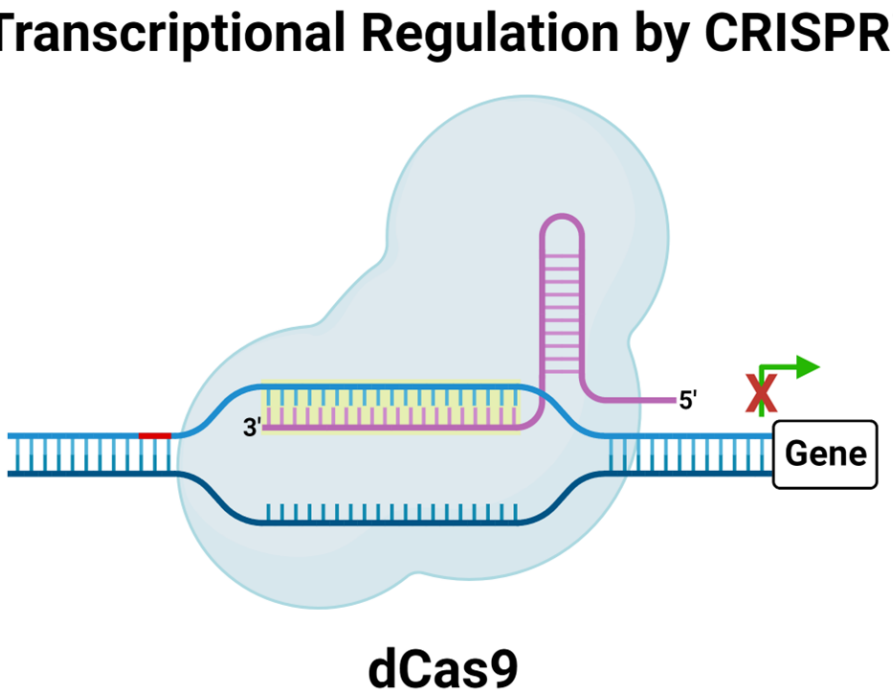


Figure 5. 2. Overview of CRISPR interference (CRISPRi) mediated transcriptional repression.

A catalytically inactivated Cas9 (dCas9) is targeted to a specific genomic DNA sequence by a designed sgRNA (purple). The complex binds the DNA sequence complementary to the sgRNA (yellow region) with high affinity, but the dCas9 is not capable of catalyzing cleavage. The dCas9-sgRNA complex physically blocks target gene transcription and is typically targeted within the promoter of the gene of interest or immediately following the initiation codon. Potential target regions require a protospacer adjacent motif (PAM, shown in red) immediately upstream, composed of the sequence NGG for the *Streptomyces pyogenes* Cas9 protein. Cas9 proteins from different organisms have different PAM sequence requirements.

Adapted from: (310)

Lo, A., and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR–Cas systems. *F1000Res*. **6**, 747.

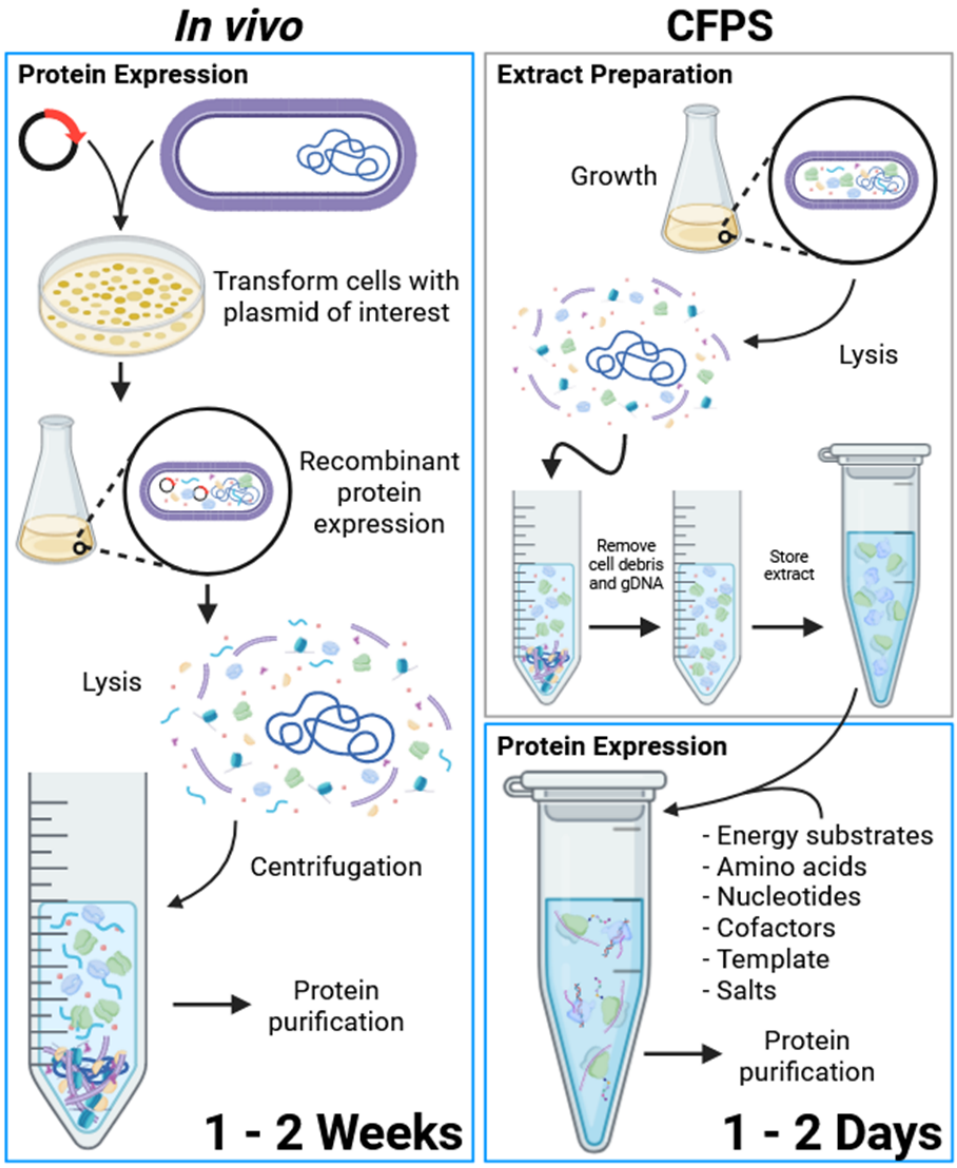


Figure 5. 3. Comparison of traditional *in vivo* protein expression and cell free protein synthesis (CFPS).

As CFPS systems do not require the transformation of cells with expression plasmids carrying genes of interest, they can produce recombinant proteins on an accelerated timeline compared to traditional *in vivo* protein expression systems. Protein expression strains are typically grown to mid- or late-exponential phase, harvested, and lysed. Processing of extracts includes removal of cell debris and genomic DNA prior to long term storage at -80°C. The addition of liposomes to

the extracts prior to protein expression can even allow for the facile recovery of some membrane proteins from CFPS systems. This methodology has been optimized in *C. glutamicum*.

Adapted from: (306)

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

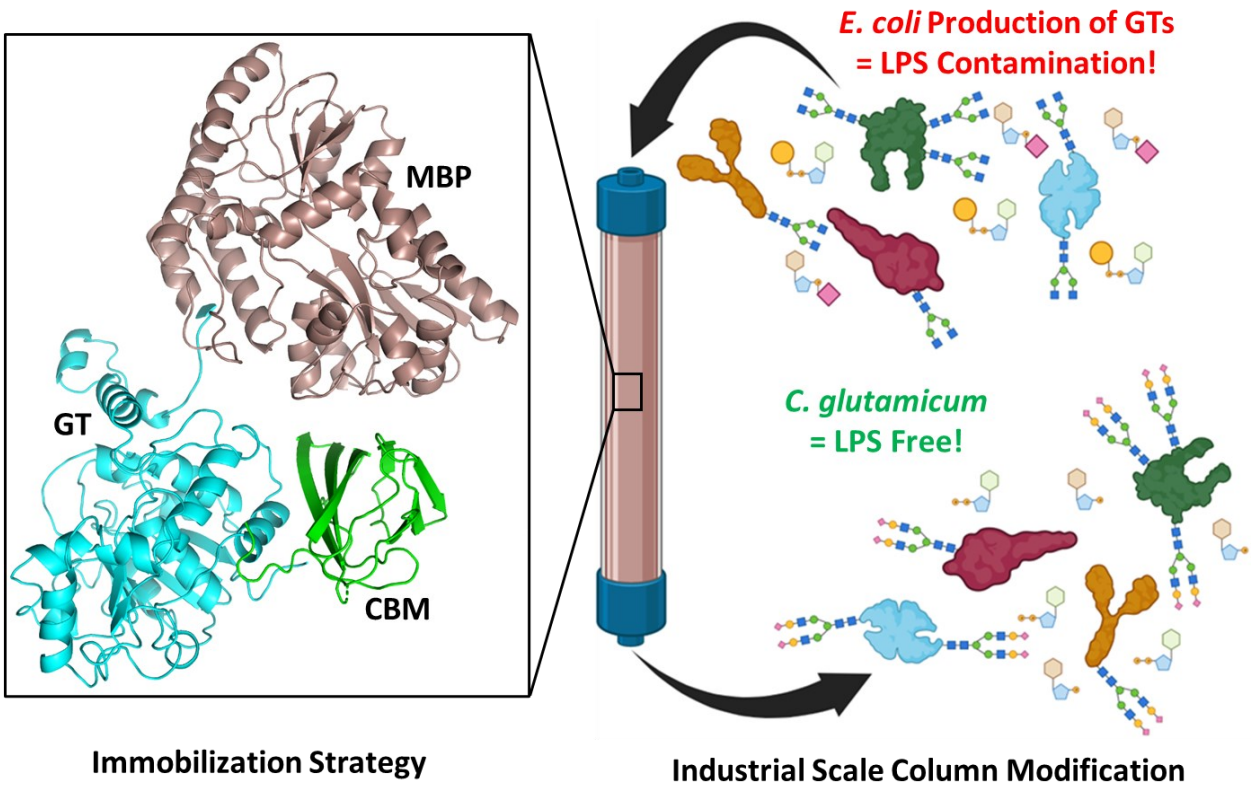


Figure 5. 4. Immobilization strategy for the *in vitro* sialylation of therapeutic glycoproteins using STs produced in *C. glutamicum*.

As *C. glutamicum* is a low LPS/LPS-free organism, STs (or GTs) produced in the organism could be used for the glycosylation of therapeutic glycoproteins without the worry of endotoxin/LPS contamination or the additional costs associated with removal. An enzyme immobilization strategy using carbohydrate binding modules (CBM) as fusions to GT enzymes could be applied as an industrially scalable methodology for the *in vitro* glycosylation of therapeutic proteins.

References

1. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (2017) Essentials of Glycobiology. *Cold Spring Harbor (NY)*

2. Montreuil, J. (1980) Primary Structure of Glycoprotein Glycans Basis for the Molecular Biology of Glycoproteins. *Adv Carbohydr Chem Biochem*. **37**, 157–223

3. Spiro, R. G. (1973) Glycoproteins. *Adv Protein Chem*. **27**, 349–467

4. Beyer, T. A., Rearick, J. I., Paulson, J. C., Prieels, J.-P., Sadler, J. E., and Hill, R. L. (1979) Biosynthesis of mammalian glycoproteins. Glycosylation pathways in the synthesis of the nonreducing terminal sequences. 10.1016/S0021-9258(19)86347-0

5. de Beer, T., Vliegenthart, J. F. G., Löffler, A., and Hofsteenge, J. (1995) The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase Us is alpha-mannopyranose. *Biochemistry*. **34**, 11785–11789

6. Ferguson, M. A. J. (1999) The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. *J Cell Sci*. **112**, 2799–2809

7. Haynes, P. A. (1998) Phosphoglycosylation: A new structural class of glycosylation? *Glycobiology*. **8**, 1–5

8. Dell, A., Galadari, A., Sastre, F., and Hitchen, P. (2010) Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *Int J Microbiol*. 10.1155/2010/148178

9. Eichler, J. (2019) Protein glycosylation. *Current Biology*. **29**, R229–R231

10. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*. **126**, 855–867

11. Nothaft, H., and Szymanski, C. M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology 2010 8:11*. **8**, 765–778

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

13. Cummings, R. D. (2009) The repertoire of glycan determinants in the human glycome. *Mol Biosyst*. **5**, 1087

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

15. Van Den Steen, P., Rudd, P. M., Dwek, R. A., and Opdenakker, G. (2008) Concepts and Principles of O-Linked Glycosylation. *http://dx.doi.org/10.1080/10409239891204198*. **33**, 151–208

16. Magalhães, A., Duarte, H. O., and Reis, C. A. (2021) The role of O-glycosylation in human disease. *Mol Aspects Med*. **79**, 100964

17. Chauhan, J. S., Rao, A., and Raghava, G. P. S. (2013) In silico Platform for Prediction of N-, O- and C-Glycosites in Eukaryotic Protein Sequences. *PLoS One*. **8**, e67008

18. Oman, T. J., Boettcher, J. M., Wang, H., Okalibe, X. N., and Van Der Donk, W. A. (2011) Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol*. **7**, 78

19. Stepper, J., Shastri, S., Loo, T. S., Preston, J. C., Novak, P., Man, P., Moore, C. H., Havlíček, V., Patchett, M. L., and Norris, G. E. (2011) Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett*. **585**, 645–650

20. Ahmed, N., and Furth, A. J. (1992) Failure of Common Glycation Assays to Detect Glycation by Fructose. *Clin Chem*. **38**, 1301–1303

21. Stadler, R. H., Blank, I., Varga, N., Robert, F., Hau, J., Guy, P. A., Robert, M. C., and Riediker, S. (2002) Acrylamide from Maillard reaction products. *Nature 2002 419:6906*. **419**, 449–450

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

23. Rodriguez, I. R., and Whelan, W. J. (1985) A novel glycosyl-amino acid linkage: Rabbit-muscle glycogen is covalently linked to a protein via tyrosine. *Biochem Biophys Res Commun*. **132**, 829–836

24. Chen, P. S., Mitchell, H. K., and Neuweg, M. (1978) Tyrosine glucoside in Drosophila busckii. *Insect Biochem*. **8**, 279–286

25. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glycosylation – putting the pieces together. *FEBS J*. **277**, 81–94

26. Tran, D. T., and Ten Hagen, K. G. (2013) Mucin-type o-glycosylation during development. *Journal of Biological Chemistry*. **288**, 6921–6929

27. Bansil, R., and Turner, B. S. (2006) Mucin structure, aggregation, physiological functions and biomedical applications. *Curr Opin Colloid Interface Sci*. **11**, 164–170

28. Sharon, N., and Lis, H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*. 10.1093/GLYCOB/CWH122

29. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J*. **32**, 1478–1488

30. Belický, Š., Katrlík, J., and Tkáč, J. (2016) Glycan and lectin biosensors. *Essays Biochem*. **60**, 37–47

31. Takeuchi, H., Schneider, M., Williamson, D. B., Ito, A., Takeuchi, M., Handford, P. A., and Haltiwanger, R. S. (2018) Two novel protein *O* -glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. *Proceedings of the National Academy of Sciences*. 10.1073/pnas.1804005115

32. Hart, G. W. (2019) Nutrient regulation of signaling and transcription. *Journal of Biological Chemistry*. **294**, 2211–2231

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

34. Vasudevan, D., Takeuchi, H., Johar, S. S., Majerus, E., and Haltiwanger, R. S. (2015) Peters Plus Syndrome Mutations Disrupt a Noncanonical ER Quality-Control Mechanism. *Current Biology*. **25**, 286–295

35. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998) NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J*. **15**, 115–130

36. Chen, Y. Z., Tang, Y. R., Sheng, Z. Y., and Zhang, Z. (2008) Prediction of mucin-type O-glycosylation sites in mammalian proteins using the composition of k-spaced amino acid pairs. *BMC Bioinformatics*. **9**, 101

37. Dwek, R. A., Edge, C. J., Harvey, D. J., Wormald, M. R., and Parekh, R. B. (2003) ANALYSIS OF GLYCOPROTEIN-ASSOCIATED OLIGOSACCHARIDES. *https://doi.org/10.1146/annurev.bi.62.070193.000433*. **62**, 65–100

38. Iwase, H., and Hotta, K. (1993) Release of O-linked glycoprotein glycans by endo-alpha-N-acetylgalactosaminidase. *Methods Mol Biol*. **14**, 151–159

39. Stimson, E., Virji, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. R., Jennings, M. P., Barker, S., Panico, M., Blench, I., and Moxon, E. R. (1995) Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol Microbiol*. **17**, 1201–1214

40. Power, P. M., Seib, K. L., and Jennings, M. P. (2006) 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

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

42. Børud, B., Aas, F. E., Vik, Å., Winther-Larsen, H. C., Egge-Jacobsen, W., and Koomey, M. (2010) Genetic, structural, and antigenic analyses of glycan diversity in the O-linked protein glycosylation systems of human Neisseria species. *J Bacteriol*. **192**, 2816–2829

43. Børud, B., Viburiene, R., Hartley, M. D., Paulsen, B. S., Egge-Jacobsen, W., Imperiali, B., and Koomey, M. (2011) Genetic and molecular analyses reveal an evolutionary trajectory for glycan synthesis in a bacterial protein glycosylation system. *Proc Natl Acad Sci U S A*. **108**, 9643–9648

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

45. Qutyan, M., Paliotti, M., and Castric, P. (2007) PilO of Pseudomonas aeruginosa 1244: subcellular location and domain assignment. *Mol Microbiol*. **66**, 1444–1458

46. Smedley, J. G., Jewell, E., Roguskie, J., Horzempa, J., Syboldt, A., Stolz, D. B., and Castric, P. (2005) Influence of pilin glycosylation on Pseudomonas aeruginosa 1244 pilus function. *Infect Immun*. **73**, 7922–7931

47. Dobos, K. M., Khoo, K. H., Swiderek, K. M., Brennan, P. J., and Belisle, J. T. (1996) Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of Mycobacterium tuberculosis. *J Bacteriol*. **178**, 2498–2506

48. Sartain, M. J., and Belisle, J. T. (2009) N-Terminal clustering of the O-glycosylation sites in the Mycobacterium tuberculosis lipoprotein SodC. *Glycobiology*. **19**, 38

49. Liu, C. F., Tonini, L., Malaga, W., Beau, M., Stella, A., Bouyssié, D., Jackson, M. C., Nigou, J., Puzo, G., Guilhot, C., Burlet-Schiltz, O., and Rivière, M. (2013) Bacterial protein-O-mannosylating enzyme is crucial for virulence of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*. **110**, 6560–6565

50. Facciuolo, A., and Mutharia, L. M. (2014) Mycobacterial glycoproteins: a novel subset of vaccine candidates. *Front Cell Infect Microbiol*. 10.3389/FCIMB.2014.00133

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

52. Loibl, M., and Strahl, S. (2013) Protein O-mannosylation: What we have learned from baker's yeast. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. **1833**, 2438–2446

53. Wehmeier, S., Varghese, A. S., Gurcha, S. S., Tissot, B., Panico, M., Hitchen, P., Morris, H. R., Besra, G. S., Dell, A., and Smith, M. C. M. (2009) Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolor by a pathway that resembles protein O-mannosylation in eukaryotes. *Mol Microbiol*. **71**, 421–433

54. Jia, L., Sha, S., Yang, S., Taj, A., and Ma, Y. (2021) Effect of Protein O-Mannosyltransferase (MSMEG_5447) on M. smegmatis and Its Survival in Macrophages. *Front Microbiol*. 10.3389/FMICB.2021.657726/FULL

55. Wakarchuk, W. W., Brochu, D., Foote, S., Robotham, A., Saxena, H., Erak, T., and Kelly, J. (2016) Proteomic Analysis of the Secretome of Cellulomonas fimi ATCC 484 and Cellulomonas flavigena ATCC 482. *PLoS One*. 10.1371/JOURNAL.PONE.0151186

56. Mahne, M., Tauch, A., Pühler, A., and Kalinowski, J. (2006) The Corynebacterium glutamicum gene pmt encoding a glycosyltransferase related to eukaryotic protein-O-mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins. *FEMS Microbiol Lett*. **259**, 226–233

57. Beals, J. M., and Shanafelt, A. B. (2006) Enhancing exposure of protein therapeutics. *Drug Discov Today Technol*. **3**, 87–94

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

59. Sola, R. J., and Griebenow, K. (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci*. **98**, 1223–1245

60. Walsh, G. (2014) Biopharmaceutical benchmarks 2014. *Nature Biotechnology 2014 32:10*. **32**, 992–1000

61. Oshima, Y., Takahashi, S., Tani, K., and Tojo, A. (2019) Granulocyte colony-stimulating factor-associated aortitis in the Japanese Adverse Drug Event Report database. *Cytokine*. **119**, 47–51

62. Saxena, A., Raveh, L., Ashani, Y., and Doctor, B. P. (1997) Structure of Glycan Moieties Responsible for the Extended Circulatory Life Time of Fetal Bovine Serum Acetylcholinesterase and Equine Serum Butyrylcholinesterase. *Biochemistry*. **36**, 7481–7489

63. Perlmutter, D. H. (1991) The cellular basis for liver injury in α1-antitrypsin deficiency. *Hepatology*. **13**, 172–185

64. Gan, J. C. (1979) Catabolism of desialylated human plasma α1-antitrypsin and its trypsin complex in the rat. *Arch Biochem Biophys*. **194**, 149–156

65. Byrne, B., Donohoe, G. G., and O'Kennedy, R. (2007) Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells. *Drug Discov Today*. **12**, 319–326

66. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., and Egrie, J. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nature Biotechnology 2003 21:4*. **21**, 414–421

67. Beck, A., Wagner-Rousset, E., Bussat, M.-C., Lokteff, M., Klinguer-Hamour, C., Haeuw, J.-F., Goetsch, L., Wurch, T., Dorsselaer, A., and Corvaia, N. (2008) Trends in

Glycosylation, Glycoanalysis and Glycoengineering of Therapeutic Antibodies and Fc-Fusion Proteins. *Curr Pharm Biotechnol*. **9**, 482–501

68. Ngantung, F. A., Miller, P. G., Brushett, F. R., Tang, G. L., and Wang, D. I. C. (2006) RNA interference of sialidase improves glycoprotein sialic acid content consistency. *Biotechnol Bioeng*. **95**, 106–119

69. Mahmood, I., and Green, M. D. (2005) Pharmacokinetic and Pharmacodynamic Considerations in the Development of Therapeutic Proteins. *Clin Pharmacokinet*. **44**, 331–347

70. Dammen-Brower, K., Epler, P., Zhu, S., Bernstein, Z. J., Stabach, P. R., Braddock, D. T., Spangler, J. B., and Yarema, K. J. (2022) Strategies for Glycoengineering Therapeutic Proteins. *Front Chem*. 10.3389/fchem.2022.863118

71. Ashwell, G., and Harford, J. (1982) Carbohydrate-Specific Receptors of the Liver. *Annu Rev Biochem*. **51**, 531–554

72. Lewis, A. M., Croughan, W. D., Aranibar, N., Lee, A. G., Warrack, B., Abu-Absi, N. R., Patel, R., Drew, B., Borys, M. C., Reily, M. D., and Li, Z. J. (2016) Understanding and Controlling Sialylation in a CHO Fc-Fusion Process. *PLoS One*. 10.1371/JOURNAL.PONE.0157111

73. Jeong, Y. T., Choi, O., Son, Y. D., Park, S. Y., and Kim, J. H. (2009) Enhanced sialylation of recombinant erythropoietin in genetically engineered Chinese-hamster ovary cells. *Biotechnol Appl Biochem*. **52**, 283

74. Kwak, C. Y., Park, S. Y., Lee, C. G., Okino, N., Ito, M., and Kim, J. H. (2017) Enhancing the sialylation of recombinant EPO produced in CHO cells via the inhibition of glycosphingolipid biosynthesis. *Scientific Reports 2017 7:1*. **7**, 1–9

75. Zhong, X., Ma, W., Meade, C. L., Tam, A. S., Llewellyn, E., Cornell, R., Cote, K., Scarcelli, J. J., Marshall, J. K., Tzvetkova, B., Figueroa, B., DiNino, D., Sievers, A., Lee, C., Guo, J., Mahan, E., Francis, C., Lam, K., D'Antona, A. M., Zollner, R., Zhu, H. L., Kriz, R., Somers, W., and Lin, L. (2019) Transient CHO expression platform for robust antibody production and its enhanced N-glycan sialylation on therapeutic glycoproteins. *Biotechnol Prog*. **35**, e2724

76. Zhong, X., Schwab, A., Ma, W., Meade, C. L., Zhou, J., D'Antona, A. M., Somers, W., and Lin, L. (2022) Large-Scale Transient Production in ExpiCHO-S™ with Enhanced N-Galactosylation-Sialylation and PEI-Based Transfection. *Methods in Molecular Biology*. **2313**, 143–150

77. Lee, J. H., Jeong, Y. R., Kim, Y. G., and Lee, G. M. (2017) Understanding of decreased sialylation of Fc-fusion protein in hyperosmotic recombinant Chinese hamster ovary cell culture: N-glycosylation gene expression and N-linked glycan antennary profile. *Biotechnol Bioeng*. **114**, 1721–1732

78. Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., Galuska, C. E., Zlatina, K., Galuska, S. P., Werner, S., Thiesler, H., Werneburg, S., Hildebrandt, H., Gerardy-Schahn, R., and Steinkellner, H. (2016) Engineering of complex protein sialylation in plants. *Proc Natl Acad Sci U S A*. **113**, 9498–9503

79. Kajiura, H., Misaki, R., Fujiyama, K., and Seki, T. (2011) Stable coexpression of two human sialylation enzymes in plant suspension-cultured tobacco cells. *J Biosci Bioeng*. **111**, 471–477

80. Aumiller, J. J., Mabashi-Asazuma, H., Hillar, A., Shi, X., and Jarvis, D. L. (2012) A new glycoengineered insect cell line with an inducibly mammalianized protein N-glycosylation pathway. *Glycobiology*. **22**, 417

81. Zhu, J., Ruan, Y., Fu, X., Zhang, L., Ge, G., Wall, J. G., Zou, T., Zheng, Y., Ding, N., and Hu, X. (2020) An Engineered Pathway for Production of Terminally Sialylated N-glycoproteins in the Periplasm of Escherichia coli. *Front Bioeng Biotechnol*. **8**, 526474

82. Sim, L., Thompson, N., Geissner, A., Withers, S. G., and Wakarchuk, W. W. (2022) Mammalian sialyltransferases allow efficient Escherichia coli-based production of mucin-type O-glycoproteins but can also transfer Kdo. *Glycobiology*. **32**, 429–440

83. McLean, M. D. (2017) Trastuzumab Made in Plants Using vivoXPRESS ® Platform Technology

84. Shantha Raju, T., Briggs, J. B., Chamow, S. M., Winkler, M. E., and Jones, A. J. S. (2001) Glycoengineering of Therapeutic Glycoproteins: In Vitro Galactosylation and Sialylation of Glycoproteins with Terminal N-Acetylglucosamine and Galactose Residues. *Biochemistry*. **40**, 8868–8876

85. Janesch, B., Saxena, H., Sim, L., and Wakarchuk, W. W. (2019) Comparison of α2,6-sialyltransferases for sialylation of therapeutic proteins. *Glycobiology*. **29**, 735–747

86. González-Morelo, K. J., Vega-Sagardía, M., and Garrido, D. (2020) Molecular Insights Into O-Linked Glycan Utilization by Gut Microbes. *Front Microbiol*. 10.3389/fmicb.2020.591568

87. VanderVen, B. C., Harder, J. D., Crick, D. C., and Belisle, J. T. (2005) Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science*. **309**, 166–168

88. Girrbach, V., Zeller, T., Priesmeier, M., and Strahl-Bolsinger, S. (2000) Structure-function analysis of the dolichyl phosphate-mannose: Protein O-mannosyltransferase ScPmt1p. *Journal of Biological Chemistry*. **275**, 19288–19296

89. Maverakis, E., Kim, K., Shimoda, M., Gershwin, M. E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L. R., and Lebrilla, C. B. (2015) Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J Autoimmun*. **57**, 1–13

90. BÉGUIN, P., and EISEN, H. (1978) Purification and Partial Characterization of Three Extracellular from Cellulomunas sp. *Eur J Biochem*. **87**, 525–531

91. Espitia, C., and Mancilla, R. (1989) Identification, isolation and partial characterization of Mycobacterium tuberculosis glycoprotein antigens. *Clin Exp Immunol*. **77**, 378

92. VanderVen, B. C., Harder, J. D., Crick, D. C., and Belisle, J. T. (2005) Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science*. **309**, 166–168

93. González-Zamorano, M., Hernández, G. M., Xolalpa, W., Parada, C., Vallecillo, A. J., Bigi, F., and Espitia, C. (2009) Mycobacterium tuberculosis glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins. *J Proteome Res*. **8**, 721–733

94. Hartmann, M., Barsch, A., Niehaus, K., Pühler, A., Tauch, A., and Kalinowski, J. (2004) The glycosylated cell surface protein Rpf2, containing a resuscitation-promoting factor motif, is involved in intercellular communication of Corynebacterium glutamicum. *Arch Microbiol*. **182**, 299–312

95. Langsford, M. L., Gilkes, N. R., and Wakarchuk, W. W. (1984) The cellulase system of Cellulomonas fimi. *J Gen Microbiol*. **130**, 1367–1376

96. Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1987) Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Lett*. **225**, 163–167

97. Wehmeier, S., Varghese, A. S., Gurcha, S. S., Tissot, B., Panico, M., Hitchen, P., Morris, H. R., Besra, G. S., Dell, A., and Smith, M. C. M. (2009) Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolor by a pathway that resembles protein O-mannosylation in eukaryotes. *Mol Microbiol*. **71**, 421–433

98. Keenan, T., Dowle, A., Bates, R., and Smith, M. C. M. (2019) Characterization of the Streptomyces coelicolor Glycoproteome Reveals Glycoproteins Important for Cell Wall Biogenesis. *mBio*. 10.1128/MBIO.01092-19

99. Cowlishaw, D. A., and Smith, M. C. M. (2001) Glycosylation of a Streptomyces coelicolor A3(2) cell envelope protein is required for infection by bacteriophage phi C31. *Mol Microbiol*. **41**, 601–610

100. Howlett, R., Anttonen, K., Read, N., and Smith, M. C. M. (2018) Disruption of the GDP-mannose synthesis pathway in Streptomyces coelicolor results in antibiotic hyper-susceptible phenotypes. *Microbiology (Reading)*. **164**, 614–624

101. Michell, S. L., Whelan, A. O., Wheeler, P. R., Panico, M., Easton, R. L., Etienne, A. T., Haslam, S. M., Dell, A., Morri, H. R., Reason, A. J., Herrmann, J. L., Young, D. B., and Hewinson, R. G. (2003) The MPB83 antigen from Mycobacterium bovis contains O-linked mannose and (1 → 3)-mannobiose moieties. *Journal of Biological Chemistry*. **278**, 16423–16432

102. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 10.1093/NAR/GKT1178

103. Strahl-Bolsinger, S., and Scheinost, A. (1999) Transmembrane Topology of Pmt1p, a Member of an Evolutionarily Conserved Family of Protein O-Mannosyltransferases. *Journal of Biological Chemistry*. **274**, 9068–9075

104. Adamany, A. M., and Spiro, R. G. (1975) Glycoprotein biosynthesis: studies on thyroid mannosyltransferases. I. Action on glycopeptides and simple glycosides. *Journal of Biological Chemistry*. **250**, 830–841

105. BABCZINSKI, P., HASELBECK, A., and TANNER, W. (1980) Yeast Mannosyl Transferases Requiring Dolichyl Phosphate and Dolichyl Phosphate Mannose as Substrate. *Eur J Biochem*. **105**, 509–515

106. Cooper, H. N., Gurcha, S. S., Nigou, J., Brennan, P. J., Belisle, J. T., Besra, G. S., and Young, D. (2002) Characterization of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay. *Glycobiology*. **12**, 427–434

107. Gurcha, S. S., Baulard, A. R., Kremer, L., Locht, C., Moody, D. B., Muhlecker, W., Costello, C. E., Crick, D. C., Brennan, P. J., and Besra, G. S. (2002) Ppm1, a novel polyprenol monophosphomannose synthase from Mycobacterium tuberculosis. *Biochem J*. **365**, 441–450

108. Bai, L., Kovach, A., You, Q., Kenny, A., and Li, H. (2019) Structure of the eukaryotic protein O-mannosyltransferase Pmt1-Pmt2 complex. *Nat Struct Mol Biol*. **26**, 704–711

109. Khurana, S., Sanli, G., Powers, D. B., Anderson, S., and Blaber, M. (2000) Molecular modeling of substrate binding in wild-type and mutant Corynebacteria 2,5-diketo-D-gluconate reductases. *Proteins*. **39**, 68–75

110. Tsuchidate, T., Tateno, T., Okai, N., Tanaka, T., Ogino, C., and Kondo, A. (2011) Glutamate production from β-glucan using endoglucanase-secreting Corynebacterium glutamicum. *Appl Microbiol Biotechnol*. **90**, 895–901

111. Wakarchuk, W. W., Brochu, D., Foote, S., Robotham, A., Saxena, H., Erak, T., and Kelly, J. (2016) Proteomic Analysis of the Secretome of Cellulomonas fimi ATCC 484 and Cellulomonas flavigena ATCC 482. *PLoS One*. 10.1371/JOURNAL.PONE.0151186

112. Ding, W., Hill, J. J., and Kelly, J. (2007) Selective enrichment of glycopeptides from glycoprotein digests using ion-pairing normal-phase liquid chromatography. *Anal Chem*. **79**, 8891–8899

113. Paradis, F. W. (1990) The expression of cellulomomas fimi cellulase genes in Brevibacterium lactofermentum and characterization of recombinant C. fimi B-glucosidase A from E coli. 10.14288/1.0098751

114. Langsford, M. L. (1988) The purification and characterization of two cellulose-binding, glycosylated cellulases from the bacterium Cellulomonas fimi. 10.14288/1.0098066

115. Ravasi, P., Peiru, S., Gramajo, H., and Menzella, H. G. (2012) Design and testing of a synthetic biology framework for genetic engineering of Corynebacterium glutamicum. *Microb Cell Fact*. **11**, 1–11

116. Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. **1**, 2856–2860

117. Ding, W., Nothaft, H., Szymanski, C. M., and Kelly, J. (2009) Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Mol Cell Proteomics*. **8**, 2170–2185

118. Kus, J. V., Kelly, J., Tessier, L., Harvey, H., Cvitkovitch, D. G., and Burrows, L. L. (2008) Modification of Pseudomonas aeruginosa Pa5196 type IV Pilins at multiple sites with D-Araf by a novel GT-C family Arabinosyltransferase, TfpW. *J Bacteriol*. **190**, 7464–7478

119. Loignon, M., Perret, S., Kelly, J., Boulais, D., Cass, B., Bisson, L., Afkhamizarreh, F., and Durocher, Y. (2008) Stable high volumetric production of glycosylated human recombinant IFNalpha2b in HEK293 cells. *BMC Biotechnol*. 10.1186/1472-6750-8-65

120. Liu, D., Wei, Q., Xia, W., He, C., Zhang, Q., Huang, L., Wang, X., Sun, Y., Ma, Y., Zhang, X., Wang, Y., Shi, X., Liu, C., and Dong, S. (2021) O-Glycosylation Induces Amyloid-β to Form New Fibril Polymorphs Vulnerable for Degradation. *J Am Chem Soc*. **143**, 20216–20223

121. Lehle, L., Strahl, S., and Tanner, W. (2006) Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angewandte Chemie International Edition*. **45**, 6802–6818

122. Wehmeier, S., Varghese, A. S., Gurcha, S. S., Tissot, B., Panico, M., Hitchen, P., Morris, H. R., Besra, G. S., Dell, A., and Smith, M. C. M. (2008) Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolor by a pathway that resembles protein O-mannosylation in eukaryotes. *Mol Microbiol*. **71**, 421–433

123. Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. **50**, D571–D577

124. Xu, C., Wang, S., Thibault, G., and Ng, D. T. W. (2013) Futile protein folding cycles in the ER are terminated by the unfolded protein O-mannosylation pathway. *Science*. **340**, 978–981

125. Posey, J. E., Shinnick, T. M., and Quinn, F. D. (2006) Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis. *J Bacteriol*. **188**, 1332–1340

126. Saxena, H., Buenbrazo, N., Song, W.-Y., Li, C., Brochu, D., Robotham, A., Ding, W., Tessier, L., Chen, R., Kelly, J., and Wakarchuk, W. (2023) Towards an experimental system for the examination of protein mannosylation in Actinobacteria. *Glycobiology*. 10.1093/GLYCOB/CWAD023

127. Lommel, M., Schott, A., Jank, T., Hofmann, V., and Strahl, S. (2011) A Conserved Acidic Motif Is Crucial for Enzymatic Activity of Protein O-Mannosyltransferases. *J Biol Chem*. **286**, 39768

128. Willer, T., Valero, M. C., Tanner, W., Cruces, J., and Strahl, S. (2003) O-mannosyl glycans: From yeast to novel associations with human disease. *Curr Opin Struct Biol*. **13**, 621–630

129. Maeda, Y., and Kinoshita, T. (2008) Dolichol-phosphate mannose synthase: structure, function and regulation. *Biochim Biophys Acta*. **1780**, 861–868

130. Tanner, W. (1969) A lipid intermediate in mannan biosynthesis in yeast. *Biochem Biophys Res Commun*. **35**, 144–150

131. Kruszewska, J. S., Saloheimo, M., Migdalski, A., Orlean, P., Penttilä, M., and Palamarczyk, G. (2000) Dolichol phosphate mannose synthase from the filamentous fungus Trichoderma reesei belongs to the human and Schizosaccharomyces pombe class of the enzyme. *Glycobiology*. **10**, 983–991

132. Orlean, P., Albright, C., and Robbins, P. W. (1988) Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein. *Journal of Biological Chemistry*. **263**, 17499–17507

133. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 10.1093/GLYCOB/12.4.43R

134. Strahl-Bolsinger, S., Gentzsch, M., and Tanner, W. (1999) Protein O-mannosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects*. **1426**, 297–307

135. Endo, T., and Suzuki, K. (2019) Mammalian O-mannosyl glycans: Biochemistry and glycopathology. *Proceedings of the Japan Academy, Series B*. **95**, 39–51

136. Sheikh, M. O., Halmo, S. M., and Wells, L. (2017) Recent advancements in understanding mammalian O-mannosylation. *Glycobiology*. **27**, 806–819

137. Fernández-Álvarez, A., Elias-Villalobos, A., and Ibeas, J. I. (2009) The O-Mannosyltransferase PMT4 Is Essential for Normal Appressorium Formation and Penetration in Ustilago maydis. *Plant Cell*. **21**, 3397

138. Manya, H., Chiba, A., Yoshida, A., Wang, X., Chiba, Y., Jigami, Y., Margolis, R. U., and Endo, T. (2004) Demonstration of mammalian protein O-mannosyltransferase activity: Coexpression of POMT1 and POMT2 required for enzymatic activity. *Proc Natl Acad Sci U S A*. **101**, 500–505

139. Vigerust, D. J. (2011) Protein glycosylation in infectious disease pathobiology and treatment. *Cent Eur J Biol*. **6**, 802–816

140. Lengeler, K. B., Tielker, D., and Ernst, J. F. (2008) Protein-O-mannosyltransferases in virulence and development. *Cellular and Molecular Life Sciences*. **65**, 528–544

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

142. Girrbach, V., and Strahl, S. (2003) Members of the evolutionarily conserved PMT family of protein O-mannosyltransferases form distinct protein complexes among themselves. *Journal of Biological Chemistry*. **278**, 12554–12562

143. Tytgat, H. L. P., and Lebeer, S. (2014) The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol Mol Biol Rev*. **78**, 372–417

144. Gurcha, S. S., Baulard, A. R., Kremer, L., Locht, C., Moody, D. B., Muhlecker, W., Costello, C. E., Crick, D. C., Brennan, P. J., and Besra, G. S. (2002) Ppm1, a novel polyprenol monophosphomannose synthase from Mycobacterium tuberculosis. *Biochemical Journal*. **365**, 441–450

145. Haselbeck, A. (1989) Purification of GDP mannose: dolichyl-phosphate O-β-D-mannosyltransferase from Saccharomyces cerevisiae. *Eur J Biochem*. **181**, 663–668

146. Korduláková, J., Gilleron, M., Mikušová, K., Puzo, G., Brennan, P. J., Gicquel, B., and Jackson, M. (2002) Definition of the First Mannosylation Step in Phosphatidylinositol Mannoside Synthesis. *Journal of Biological Chemistry*. **277**, 31335–31344

147. Rainczuk, A. K., Yamaryo-Botte, Y., Brammananth, R., Stinear, T. P., Seemann, T., Coppel, R. L., McConville, M. J., and Crellin, P. K. (2012) The lipoprotein LpqW is essential for the mannosylation of periplasmic glycolipids in Corynebacteria. *J Biol Chem*. **287**, 42726–42738

148. Haselbeck, A., and Tanner, W. (1983) O-glycosylation in Saccharomyces cerevisiae is initiated at the endoplasmic reticulum. *FEBS Lett*. **158**, 335–338

149. Maeda, Y., and Kinoshita, T. (2008) Dolichol-phosphate mannose synthase: Structure, function and regulation. *Biochimica et Biophysica Acta (BBA) - General Subjects*. **1780**, 861–868

150. Szymanski, C. M., and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology 2005 3:3*. **3**, 225–237

151. Mohiman, N., Argentini, M., Batt, S. M., Cornu, D., Masi, M., Eggeling, L., Besra, G., and Bayan, N. (2012) The ppm operon is essential for acylation and glycosylation of lipoproteins in Corynebacterium glutamicum. *PLoS One*. 10.1371/JOURNAL.PONE.0046225

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

153. Bartels, M. F., Winterhalter, P. R., Yu, J., Liu, Y., Lommel, M., Möhrlen, F., Hu, H., Feizi, T., Westerlind, U., Ruppert, T., and Strahl, S. (2016) Protein O-Mannosylation in the Murine Brain: Occurrence of Mono-O-Mannosyl Glycans and Identification of New Substrates. *PLoS One*. 10.1371/JOURNAL.PONE.0166119

154. Lommel, M., Winterhalter, P. R., Willer, T., Dahlhoff, M., Schneider, M. R., Bartels, M. F., Renner-Müller, I., Ruppert, T., Wolf, E., and Strahl, S. (2013) Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion. *Proc Natl Acad Sci U S A*. **110**, 21024–21029

155. Larsen, I. S. B., Narimatsu, Y., Joshi, H. J., Yang, Z., Harrison, O. J., Brasch, J., Shapiro, L., Honig, B., Vakhrushev, S. Y., Clausen, H., and Halim, A. (2017) Mammalian O-mannosylation of cadherins and plexins is independent of protein O-mannosyltransferases 1 and 2. *J Biol Chem*. **292**, 11586–11598

156. Vester-Christensen, M. B., Halim, A., Joshi, H. J., Steentoft, C., Bennett, E. P., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2013) Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proc Natl Acad Sci U S A*. **110**, 21018–21023

157. Eisenhaber, B., Sinha, S., Jadalanki, C. K., Shitov, V. A., Tan, Q. W., Sirota, F. L., and Eisenhaber, F. (2021) Conserved sequence motifs in human TMTC1, TMTC2, TMTC3, and TMTC4, new O-mannosyltransferases from the GT-C/PMT clan, are rationalized as ligand binding sites. *Biol Direct*. **16**, 1–18

158. Sunryd, J. C., Cheon, B., Graham, J. B., Giorda, K. M., Fissore, R. A., and Hebert, D. N. (2014) TMTC1 and TMTC2 Are Novel Endoplasmic Reticulum Tetratricopeptide Repeat-containing Adapter Proteins Involved in Calcium Homeostasis. *Journal of Biological Chemistry*. **289**, 16085–16099

159. Sanli, G. (2004) Structural alteration of cofactor specificity in Corynebacterium 2,5-diketo-D-gluconic acid reductase. *Protein Science*. **13**, 504–512

160. Tsuchidate, T., Tateno, T., Okai, N., Tanaka, T., Ogino, C., and Kondo, A. (2011) Glutamate production from β-glucan using endoglucanase-secreting Corynebacterium glutamicum. *Appl Microbiol Biotechnol*. **90**, 895–901

161. Mori, H., and Cline, K. (2001) Post-translational protein translocation into thylakoids by the Sec and ΔpH-dependent pathways. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. **1541**, 80–90

162. Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994) Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli

plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene*. **145**, 69–73

163. He, Z., and De Buck, J. (2010) Cell wall proteome analysis of Mycobacterium smegmatis strain MC2 155. *BMC Microbiol*. **10**, 121

164. Lehrer, J., Vigeant, K. A., Tatar, L. D., and Valvano, M. A. (2007) 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

165. Kaur, J., and Bachhawat, A. K. (2009) A modified Western blot protocol for enhanced sensitivity in the detection of a membrane protein. *Anal Biochem*. **384**, 348–349

166. Holman, N. D. M., Wilkinson, A. J., and Smith, M. C. M. (2021) Alanine-scanning mutagenesis of protein mannosyl-transferase from Streptomyces coelicolor reveals strong activity-stability correlation. *Microbiology (N Y)*. 10.1099/mic.0.001103

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

168. Saxena, H., Hsu, B., De Asis, M., Zierke, M., Sim, L., Withers, S. G., and Wakarchuk, W. (2018) Characterization of a thermostable endoglucanase from Cellulomonas fimi ATCC484. *Biochem Cell Biol*. **96**, 68–76

169. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A. C., Meyer, F., Möckel, B., Pfefferle, W., Pühler, A., Rey, D. A., Rückert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegräbe, I., and Tauch, A. (2003) The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*. **104**, 5–25

170. Alonso, H., Parra, J., Malaga, W., Payros, D., Liu, C. F., Berrone, C., Robert, C., Meunier, E., Burlet-Schiltz, O., Rivière, M., and Guilhot, C. (2017) Protein O-mannosylation deficiency increases LprG-associated lipoarabinomannan release by Mycobacterium tuberculosis and enhances the TLR2-associated inflammatory response. *Sci Rep*. 10.1038/S41598-017-08489-7

171. Chen, L., Drake, M. R., Resch, M. G., Greene, E. R., Himmel, M. E., Chaffey, P. K., Beckham, G. T., and Tan, Z. (2014) Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. *Proc Natl Acad Sci U S A*. **111**, 7612–7617

172. Torrelles, J. B., DesJardin, L. E., MacNeil, J., Kaufman, T. M., Kutzbach, B., Knaup, R., McCarthy, T. R., Gurcha, S. S., Besra, G. S., Clegg, S., and Schlesinger, L. S. (2009) Inactivation of Mycobacterium tuberculosis mannosyltransferase pimB reduces the cell

wall lipoarabinomannan and lipomannan content and increases the rate of bacterial-induced human macrophage cell death. *Glycobiology*. **19**, 743

173. Schaeffer, M. L., Khoo, K. H., Besra, G. S., Chatterjee, D., Brennan, P. J., Belisle, J. T., and Inamine, J. M. (1999) The pimB gene of Mycobacterium tuberculosis encodes a mannosyltransferase involved in lipoarabinomannan biosynthesis. *J Biol Chem*. **274**, 31625–31631

174. Delcour, A. H. (2009) Outer Membrane Permeability and Antibiotic Resistance. *Biochim Biophys Acta*. **1794**, 808

175. Lichtinger, T., Burkovski, A., Niederweis, M., Krámer, R., and Benz, R. (1998) Biochemical and Biophysical Characterization of the Cell Wall Porin of Corynebacterium glutamicum: The Channel Is Formed by a Low Molecular Mass Polypeptide†. *Biochemistry*. **37**, 15024–15032

176. Birhanu, A. G., Yimer, S. A., Kalayou, S., Riaz, T., Zegeye, E. D., Holm-Hansen, C., Norheim, G., Aseffa, A., Abebe, M., and Tønjum, T. (2019) Ample glycosylation in membrane and cell envelope proteins may explain the phenotypic diversity and virulence in the Mycobacterium tuberculosis complex. *Scientific Reports 2019 9:1*. **9**, 1–15

177. Howlett, R., Read, N., Varghese, A., Kershaw, C., Hancock, Y., and Smith, M. C. M. (2018) Streptomyces coelicolor strains lacking polyprenol phosphate mannose synthase and protein O-mannosyl transferase are hyper-susceptible to multiple antibiotics. *Microbiology (Reading)*. **164**, 369–382

178. Trevor, A. J., Katzung, B. G., Kruidering-Hall, M. M., and Masters, S. B. (2013) Chapter 44. Chloramphenicol, Tetracyclines, Macrolides, Clindamycin, Streptogramins, & Linezolid. in *Katzung & Trevor's Pharmacology: Examination & Board Review, 10e*, The McGraw-Hill Companies, New York, NY

179. Hahn, F. E., C. L. Wisseman, Jr., and Hopps, H. E. (1955) MODE OF ACTION OF CHLORAMPHENICOL III. : Action of Chloramphenicol on Bacterial Energy Metabolism. *J Bacteriol*. **69**, 215

180. Yang, G., Trylska, J., Tor, Y., and McCammon, J. A. (2006) Binding of aminoglycosidic antibiotics to the oligonucleotide A-site model and 30S ribosomal subunit: Poisson-Boltzmann model, thermal denaturation, and fluorescence studies. *J Med Chem*. **49**, 5478–5490

181. Liu, S., Li, S., Krezel, A. M., and Li, W. (2022) Stabilization and structure determination of integral membrane proteins by termini restraining. *Nature Protocols 2021 17:2*. **17**, 540–565

182. Tiefenauer, L., and Demarche, S. (2012) Challenges in the Development of Functional Assays of Membrane Proteins. *Materials*. **5**, 2205

183. Kongpracha, P., Wiriyasermkul, P., Isozumi, N., Moriyama, S., Kanai, Y., and Nagamori, S. (2022) Simple But Efficacious Enrichment of Integral Membrane Proteins and Their Interactions for In-Depth Membrane Proteomics. *Mol Cell Proteomics*. **21**, 100206

184. Karata, K., Verma, C. S., Wilkinson, A. J., and Ogura, T. (2001) Probing the mechanism of ATP hydrolysis and substrate translocation in the AAA protease FtsH by modelling and mutagenesis. *Mol Microbiol*. **39**, 890–903

185. Oertel, D., Schmitz, S., and Freudl, R. (2015) A TatABC-Type Tat Translocase Is Required for Unimpaired Aerobic Growth of Corynebacterium glutamicum ATCC13032. *PLoS One*. **10**, e0123413

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

187. Wang, Y., Liu, X., Li, Y., Yang, Y., Liu, C., Linhardt, R. J., Zhang, F., and Bai, Z. (2023) Enhanced production of recombinant proteins in Corynebacterium glutamicum using a molecular chaperone. *J Gen Appl Microbiol*. 10.2323/JGAM.2022.10.002

188. Målen, H., Berven, F. S., Fladmark, K. E., and Wiker, H. G. (2007) Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv. *Proteomics*. **7**, 1702–1718

189. Schaumburg, J., Diekmann, O., Hagendorff, P., Bergmann, S., Rohde, M., Hammerschmidt, S., Jänsch, L., Wehland, J., and Kärst, U. (2004) The cell wall subproteome of Listeria monocytogenes. *Proteomics*. **4**, 2991–3006

190. Pasztor, L., Ziebandt, A. K., Nega, M., Schlag, M., Haase, S., Franz-Wachtel, M., Madlung, J., Nordheim, A., Heinrichs, D. E., and Götz, F. (2010) Staphylococcal major autolysin (Atl) is involved in excretion of cytoplasmic proteins. *J Biol Chem*. **285**, 36794–36803

191. Lei, B., Mackie, S., Lukomski, S., and Musser, J. M. (2000) Identification and Immunogenicity of Group A Streptococcus Culture Supernatant Proteins. *Infect Immun*. **68**, 6807

192. Naclerio, G., Baccigalupi, L., Caruso, C., De Felice, M., and Ricca, E. (1995) Bacillus subtilis Vegetative Catalase Is an Extracellular Enzyme. *Appl Environ Microbiol*. **61**, 4471–4473

193. Lee, M. J., and Kim, P. (2018) Recombinant Protein Expression System in Corynebacterium glutamicum and Its Application. *Front Microbiol*. 10.3389/fmicb.2018.02523

194. Liu, X., Meng, L., Wang, X., Yang, Y., and Bai, Z. (2022) Effect of Clp protease from Corynebacterium glutamicum on heterologous protein expression. *Protein Expr Purif*. **189**, 105928

195. Ravasi, P., Peiru, S., Gramajo, H., and Menzella, H. G. (2012) Design and testing of a synthetic biology framework for genetic engineering of Corynebacterium glutamicum. *Microb Cell Fact*. **11**, 147–147

196. Van der Rest, M. E., Lange, C., and Molenaar, D. (1999) A heat shock following electroporation induces highly efficient transformation of Corynebacterium glutamicum with xenogeneic plasmid DNA. *Appl Microbiol Biotechnol*. **52**, 541–545

197. Ruan, Y., Zhu, L., and Li, Q. (2015) Improving the electro-transformation efficiency of Corynebacterium glutamicum by weakening its cell wall and increasing the cytoplasmic membrane fluidity. *Biotechnol Lett*. **37**, 2445–2452

198. Schauer, R., Srinivasan, G. V., Wipfler, D., Kniep, B., and Schwartz-Albiez, R. (2011) O-Acetylated Sialic Acids and Their Role in Immune Defense. *The Molecular Immunology of Complex Carbohydrates-3*. **705**, 525

199. Brinkman-Van Der Linden, E. C. M., and Varki, A. (2000) New aspects of siglec binding specificities, including the significance of fucosylation and of the sialyl-Tn epitope. Sialic acid-binding immunoglobulin superfamily lectins. *J Biol Chem*. **275**, 8625–8632

200. Varki, A. (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature 2007 446:7139*. **446**, 1023–1029

201. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *J Biol Chem*. **276**, 34862–34870

202. Kooistra, O., Lüneberg, E., Knirel, Y. A., Frosch, M., and Zähringer, U. (2002) N-Methylation in polylegionaminic acid is associated with the phase-variable epitope of Legionella pneumophila serogroup 1 lipopolysaccharide. Identification of 5-(N,N-dimethylacetimidoyl)amino and 5-acetimidoyl(N-methyl)amino-7-acetamido-3,5,7,9-tetradeoxynon-2-ulosonic acid in the O-chain polysaccharide. *Eur J Biochem*. **269**, 560–572

203. Gil-Serrano, A. M., Rodríguez-Carvajal, M. A., Tejero-Mateo, P., Espartero, J. L., Menendez, M., Corzo, J., Ruiz-Sainz, J. E., and Buendía-Clavería, A. M. (1999) Structural determination of a 5-acetamido-3,5,7, 9-tetradeoxy-7-(3-hydroxybutyramido)-L-glycero-L-manno-nonulos onic acid-containing homopolysaccharide isolated from Sinorhizobium fredii HH103. *Biochemical Journal*. **342**, 527

204. Martin, M. J., Muotri, A., Gage, F., and Varki, A. (2005) Human embryonic stem cells express an immunogenic nonhuman sialic acid. *Nature Medicine 2005 11:2*. **11**, 228–232

205. Hashii, N., Kawasaki, N., Nakajima, Y., Toyoda, M., Katagiri, Y., Itoh, S., Harazono, A., Umezawa, A., and Yamaguchi, T. (2007) Study on the quality control of cell therapy products: Determination of N-glycolylneuraminic acid incorporated into human cells by nano-flow liquid chromatography/Fourier transformation ion cyclotron mass spectrometry. *J Chromatogr A*. **1160**, 263–269

206. Lewis, A. L., Chen, X., Schnaar, R. L., and Varki, A. (2022) Sialic Acids and Other Nonulosonic Acids. *Essentials of Glycobiology*. 10.1101/GLYCOBIOLOGY.4E.15

207. Severi, E., Hood, D. W., and Thomas, G. H. (2007) Sialic acid utilization by bacterial pathogens. *Microbiology (N Y)*. **153**, 2817–2822

208. Ilver, D., Johansson, P., Miller-Podraza, H., Nyholm, P. G., Teneberg, S., and Karlsson, K. A. (2003) Bacterium-host protein-carbohydrate interactions. *Methods Enzymol*. **363**, 134–157

209. Schauer, R. (2000) Achievements and challenges of sialic acid research. *Glycoconj J*. **17**, 485

210. Suzuki, Y. (2005) Sialobiology of Influenza: Molecular Mechanism of Host Range Variation of Influenza Viruses. *Biol Pharm Bull*. **28**, 399–408

211. Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C., and Wilson, I. A. (2006) Glycan Microarray Analysis of the Hemagglutinins from Modern and Pandemic Influenza Viruses Reveals Different Receptor Specificities. *J Mol Biol*. **355**, 1143–1155

212. Gagneux, P., Cheriyan, M., Hurtado-Ziola, N., Brinkman Van Der Linden, E. C. M., Anderson, D., McClure, H., Varki, A., and Varki, N. M. (2003) Human-specific Regulation of α2-6-linked Sialic Acids. *Journal of Biological Chemistry*. **278**, 48245–48250

213. Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006) Influenza virus receptors in the human airway. *Nature 2006 440:7083*. **440**, 435–436

214. Rutishauser, U. (2008) Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. *Nature Reviews Neuroscience 2007 9:1*. **9**, 26–35

215. Ashwell, G., and Harford, J. (2003) Carbohydrate-Specific Receptors of the Liver. *https://doi.org/10.1146/annurev.bi.51.070182.002531*. **51**, 531–554

216. Morell, A. G., Irvine, R. A., Sternlieb, I., Scheinberg, I. H., and Ashwell, G. (1968) Physical and Chemical Studies on Ceruloplasmin: V. METABOLIC STUDIES ON SIALIC ACID-FREE CERULOPLASMIN IN VIVO. *Journal of Biological Chemistry*. **243**, 155–159

217. Weigel, P. H., and Yik, J. H. N. (2002) Glycans as endocytosis signals: the cases of the asialoglycoprotein and hyaluronan/chondroitin sulfate receptors. *Biochimica et Biophysica Acta (BBA) - General Subjects*. **1572**, 341–363

218. Vattepu, R., Sneed, S. L., and Anthony, R. M. (2022) Sialylation as an Important Regulator of Antibody Function. *Front Immunol*. **13**, 818736

219. Huang, J., Huang, J., and Zhang, G. (2022) Insights into the Role of Sialylation in Cancer Metastasis, Immunity, and Therapeutic Opportunity. *Cancers (Basel)*. 10.3390/CANCERS14235840

220. Wang, Q., Yin, B., Chung, C. Y., and Betenbaugh, M. J. (2017) Glycoengineering of CHO cells to improve product quality. *Methods in Molecular Biology*. **1603**, 25–44

221. Misaki, R., Fujiyama, K., and Seki, T. (2006) Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. *Biochem Biophys Res Commun*. **339**, 1184–1189

222. Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., Devos, S., Rousseau, F., Plets, E., Houthuys, E., Malinge, P., Magistrelli, G., Cons, L., Chatel, L., Devreese, B., and Callewaert, N. (2014) GlycoDelete technology: simplifying mammalian cell N-glycosylation for recombinant protein expression. *Nat Biotechnol*. **32**, 485

223. Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T. A., Li, H., Choi, B. K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., and Gerngross, T. U. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science (1979)*. **313**, 1441–1443

224. Chang, G. D., Chen, C. J., Lin, C. Y., Chen, H. C., and Chen, H. (2003) Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *J Biotechnol*. **102**, 61–71

225. Hollister, J., Grabenhorst, E., Nimtz, M., Conradt, H., and Jarvis, D. L. (2002) Engineering the Protein N-Glycosylation Pathway in Insect Cells for Production of Biantennary, Complex N-Glycans. *Biochemistry*. **41**, 15093

226. Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M. A., Samyn-Petit, B., Julien, S., and Delannoy, P. (2001) The human sialyltransferase family. *Biochimie*. 10.1016/S0300-9084(01)01301-3

227. Chen, X., and Varki, A. (2010) Advances in the Biology and Chemistry of Sialic Acids. *ACS Chem Biol*. **5**, 163

228. Bouchet, V., Hood, D. W., Li, J., Brisson, J. R., Randle, G. A., Martin, A., Li, Z., Goldstein, R., Schweda, E. K. H., Pelton, S. I., Richards, J. C., and Moxon, E. R. (2003) Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc Natl Acad Sci U S A*. **100**, 8898

229. Bhattacharjee, A. K., Jennings, H. J., Kenny, C. P., Martin, A., and Smith, I. C. P. (1975) Structural Determination of the Sialic Acid Polysaccharide Antigens of Neisseria meningitidis Serogroups B and C with Carbon 13 Nuclear Magnetic Resonance*. *Journal of Biological Chemistry*. **250**, 1926–1932

230. Miyazaki, T., Angata, K., Seeberger, P. H., Hindsgaul, O., and Fukuda, M. (2008) CMP substitutions preferentially inhibit polysialic acid synthesis. *Glycobiology*. **18**, 187–194

231. Preidl, J. J., Gnanapragassam, V. S., Lisurek, M., Saupe, J., Horstkorte, R., and Rademann, J. (2014) Fluorescent Mimetics of CMP-Neu5Ac Are Highly Potent, Cell-Permeable Polarization Probes of Eukaryotic and Bacterial Sialyltransferases and Inhibit Cellular Sialylation. *Angewandte Chemie International Edition*. **53**, 5700–5705

232. Ni, L., Chokhawala, H. A., Cao, H., Henning, R., Ng, L., Huang, S., Yu, H., Chen, X., and Fisher, A. J. (2007) Crystal structures of Pasteurella multocida sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry*. **46**, 6288–6298

233. Chia, S., Tay, S. J., Song, Z., Yang, Y., Walsh, I., and Pang, K. T. (2023) Enhancing pharmacokinetic and pharmacodynamic properties of recombinant therapeutic proteins by manipulation of sialic acid content. *Biomedicine & Pharmacotherapy*. **163**, 114757

234. Steenbergen, S. M., and Vimr, E. R. (2003) Functional relationships of the sialyltransferases involved in expression of the polysialic acid capsules of Escherichia coli K1 and K92 and Neisseria meningitidis groups B or C. *Journal of Biological Chemistry*. **278**, 15349–15359

235. Shen, G.-J., Datta, A. K., Izumi, M., Koeller, K. M., and Wong, C.-H. (1999) Expression of α2,8/2,9-Polysialyltransferase fromEscherichia coli K92. *Journal of Biological Chemistry*. **274**, 35139–35146

236. Freiberger, F., Claus, H., Günzel, A., Oltmann-Norden, I., Vionnet, J., Mühlenhoff, M., Vogel, U., Vann, W. F., Gerardy-Schahn, R., and Stummeyer, K. (2007) Biochemical characterization of a Neisseria meningitidis polysialyltransferase reveals novel functional motifs in bacterial sialyltransferases. *Mol Microbiol*. **65**, 1258–1275

237. Varki, A., and Angata, T. (2006) Siglecs—the major subfamily of I-type lectins. *Glycobiology*. **16**, 1R-27R

238. Abukar, T., Rahmani, S., Thompson, N. K., Antonescu, C. N., and Wakarchuk, W. W. (2021) Development of BODIPY labelled sialic acids as sialyltransferase substrates for direct detection of terminal galactose on N- and O-linked glycans. *Carbohydr Res*. **500**, 108249

239. Higa, H. H., and Paulson, J. C. (1985) Sialylation of glycoprotein oligosaccharides with N-acetyl-, N-glycolyl-, and N-O-diacetylneuraminic acids. *Journal of Biological Chemistry*. **260**, 8838–8848

240. Angata, T., Matsuda, T., and Kitajima, K. (1998) Synthesis of neoglycoconjugates containing deaminated neuraminic acid (KDN) using rat liver α2,6-sialyltransferase. *Glycobiology*. **8**, 277–284

241. Datta, A. K., Sinha, A., and Paulson, J. C. (1998) Mutation of the sialyltransferase s-sialylmotif alters the kinetics of the donor and acceptor substrates. *Journal of Biological Chemistry*. **273**, 9608–9614

242. Datta, A. (2009) Comparative Sequence Analysis in the Sialyltransferase Protein Family: Analysis of Motifs. *Curr Drug Targets*. **10**, 483–498

243. Anne Harduin-Lepers (2010) Comprehensive Analysis of Sialyltransferases in Vertebrate Genomes. *Glycobiol Insights*. **2**, 29–61

244. Borys, M. C., Dalal, N. G., Abu-Absi, N. R., Khattak, S. F., Jing, Y., Xing, Z., and Li, Z. J. (2010) Effects of culture conditions on N-glycolylneuraminic acid (Neu5Gc) content of a recombinant fusion protein produced in CHO cells. *Biotechnol Bioeng*. **105**, 1048–1057

245. Datta, A. K., Chammas, R., and Paulson, J. C. (2001) Conserved Cysteines in the Sialyltransferase Sialylmotifs Form an Essential Disulfide Bond. *Journal of Biological Chemistry*. **276**, 15200–15207

246. Romanow, A., Keys, T. G., Stummeyer, K., Freiberger, F., Henrissat, B., and Gerardy-Schahn, R. (2014) Dissection of hexosyl- and sialyltransferase domains in the bifunctional capsule polymerases from Neisseria meningitidis W and Y defines a new sialyltransferase family. *J Biol Chem*. **289**, 33945–33957

247. Jen, F. E. C., Ketterer, M. R., Semchenko, E. A., Day, C. J., Jennings, M. P., Seib, K. L., and Apicella, M. A. (2021) The Lst Sialyltransferase of Neisseria gonorrhoeae Can Transfer Keto-Deoxyoctanoate as the Terminal Sugar of Lipooligosaccharide: a Glyco-Achilles Heel That Provides a New Strategy for Vaccines to Prevent Gonorrhea. *mBio*. 10.1128/MBIO.03666-20

248. Li, N., Qi, Y., Zhang, F. Y., Yu, X. H., Wu, Y. G., Chen, Y., Jiang, C. L., and Kong, W. (2011) Overexpression of α-2,6 sialyltransferase stimulates propagation of human influenza viruses in Vero cells. *Acta Virol*. **55**, 147–153

249. Yuki, N. (2007) Campylobacter sialyltransferase gene polymorphism directs clinical features of Guillain-Barré syndrome. *J Neurochem*. **103 Suppl 1**, 150–158

250. Mertsch, A., He, N., Yi, D., Kickstein, M., and Fessner, W. D. (2020) An α2,3-Sialyltransferase from Photobacterium phosphoreum with Broad Substrate Scope: Controlling Hydrolytic Activity by Directed Evolution. *Chemistry – A European Journal*. **26**, 11614–11624

251. Yamamoto, T., Nakashizuka, M., and Terada, I. (1998) Cloning and Expression of a Marine Bacterial β-Galactoside α2,6-Sialyltransferase Gene from Photobacterium damsela JT0160. *The Journal of Biochemistry*. **123**, 94–100

252. Ivarsson, M., Villiger, T. K., Morbidelli, M., and Soos, M. (2014) Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *J Biotechnol*. **188**, 88–96

253. Ghaderi, D., Taylor, R. E., Padler-Karavani, V., Diaz, S., and Varki, A. (2010) Implications of the presence of N-glycolylneuraminic Acid in Recombinant Therapeutic Glycoproteins. *Nat Biotechnol*. **28**, 863

254. Magalhães, P., Lopes, A., Mazzola, P., Rangel-Yagui, C., Penna, T., and Pessoa, A. (2007) Methods of endotoxin removal from biological preparations: A review. *J Pharm Pharm Sci*. **10**, 388–404

255. Sampath, V. P. (2018) Bacterial endotoxin-lipopolysaccharide; structure, function and its role in immunity in vertebrates and invertebrates. *Agriculture and Natural Resources*. **52**, 115–120

256. Willis, L. M., Gilbert, M., Karwaski, M. F., Blanchard, M. C., and Wakarchuk, W. W. (2008) Characterization of the α-2,8-polysialyltransferase from Neisseria meningitidis with synthetic acceptors, and the development of a self-priming polysialyltransferase fusion enzyme. *Glycobiology*. **18**, 177–186

257. Lindhout, T., Iqbal, U., Willis, L. M., Reid, A. N., Li, J., Liu, X., Moreno, M., and Wakarchuk, W. W. (2011) Site-specific enzymatic polysialylation of therapeutic proteins using bacterial enzymes. *Proc Natl Acad Sci U S A*. **108**, 7397–7402

258. Kranz, A., Polen, T., Kotulla, C., Arndt, A., Bosco, G., Bussmann, M., Chattopadhyay, A., Cramer, A., Davoudi, C. F., Degner, U., Diesveld, R., Freiherr von Boeselager, R., Gärtner, K., Gätgens, C., Georgi, T., Geraths, C., Haas, S., Heyer, A., Hünnefeld, M., Ishige, T., Kabus, A., Kallscheuer, N., Kever, L., Klaffl, S., Kleine, B., Kočan, M., Koch-Koerfges, A., Kraxner, K. J., Krug, A., Krüger, A., Küberl, A., Labib, M., Lange, C., Mack, C., Maeda, T., Mahr, R., Majda, S., Michel, A., Morosov, X., Müller, O., Nanda, A. M., Nickel, J., Pahlke, J., Pfeifer, E., Platzen, L., Ramp, P., Rittmann, D., Schaffer, S., Scheele, S., Spelberg, S., Schulte, J., Schweitzer, J. E., Sindelar, G., Sorger-Herrmann, U., Spelberg, M., Stansen, C., Tharmasothirajan, A., Ooyen, J. van, van Summeren-Wesenhagen, P., Vogt, M., Witthoff, S., Zhu, L., Eikmanns, B. J., Oldiges, M., Schaumann, G., Baumgart, M., Brocker, M., Eggeling, L., Freudl, R., Frunzke, J., Marienhagen, J., Wendisch, V. F., and Bott, M. (2022) A manually curated compendium of expression profiles for the microbial cell factory Corynebacterium glutamicum. *Sci Data*. 10.1038/S41597-022-01706-7

259. Sun, H., Zhao, D., Xiong, B., Zhang, C., and Bi, C. (2016) Engineering Corynebacterium glutamicum for violacein hyper production. *Microb Cell Fact*. **15**, 1–9

260. Kinoshita, S., Udaka, S., and Shimono, M. (1957) STUDIES ON THE AMINO ACID FERMENTATION PART I. PRODUCTION OF L-GLUTAMIC ACID BY VARIOUS MICROORGANISMS. *J Gen Appl Microbiol*. **3**, 193–205

261. Yim, S. S., An, S. J., Choi, J. W., Ryu, A. J., and Jeong, K. J. (2014) High-level secretory production of recombinant single-chain variable fragment (scFv) in Corynebacterium glutamicum. *Appl Microbiol Biotechnol*. **98**, 273–284

262. Wendisch, V. F., Jorge, J. M. P., Pérez-García, F., and Sgobba, E. (2016) Updates on industrial production of amino acids using Corynebacterium glutamicum. *World J Microbiol Biotechnol*. 10.1007/S11274-016-2060-1

263. Ikeda Masato, and Takeno, S. (2013) Amino Acid Production by Corynebacterium glutamicum. in *Corynebacterium glutamicum: Biology and Biotechnology* (Yukawa Hideaki and Inui, M. ed), pp. 107–147, Springer Berlin Heidelberg, Berlin, Heidelberg, 10.1007/978-3-642-29857-8_4

264. Liu, X. X., Li, Y., and Bai, Z. H. (2021) Corynebacterium glutamicum as a robust microbial factory for production of value-added proteins and small molecules: fundamentals and applications. *Microbial Cell Factories Engineering for Production of Biomolecules*. 10.1016/B978-0-12-821477-0.00006-4

265. Li, C., Swofford, C. A., Rückert, C., and Sinskey, A. J. (2021) Optimizing recombineering in Corynebacterium glutamicum. *Biotechnol Bioeng*. **118**, 2255–2264

266. Kirchner, O., and Tauch, A. (2003) Tools for genetic engineering in the amino acid-producing bacterium Corynebacterium glutamicum. *J Biotechnol*. **104**, 287–299

267. Bakkes, P. J., Ramp, P., Bida, A., Dohmen-Olma, D., Bott, M., and Freudl, R. (2020) Improved pEKEx2-derived expression vectors for tightly controlled production of recombinant proteins in Corynebacterium glutamicum. *Plasmid*. 10.1016/J.PLASMID.2020.102540

268. Ravasi, P., Peiru, S., Gramajo, H., and Menzella, H. G. (2012) Design and testing of a synthetic biology framework for genetic engineering of Corynebacterium glutamicum. *Microb Cell Fact*. 10.1186/1475-2859-11-147

269. Gauttam, R., Desiderato, C., Jung, L., Shah, A., and Eikmanns, B. J. (2019) A step forward: Compatible and dual-inducible expression vectors for gene co-expression in Corynebacterium glutamicum. *Plasmid*. **101**, 20–27

270. Lee, M. J., and Kim, P. (2018) Recombinant Protein Expression System in Corynebacterium glutamicum and Its Application. *Front Microbiol*. **9**, 2523

271. Liu, X., Yang, Y., Zhang, W., Sun, Y., Peng, F., Jeffrey, L., Harvey, L., McNeil, B., and Bai, Z. (2015) Expression of recombinant protein using Corynebacterium Glutamicum: progress, challenges and applications. *http://dx.doi.org/10.3109/07388551.2015.1004519*. **36**, 652–664

272. Srivastava, P., and Deb, J. K. (2005) Gene expression systems in corynebacteria. *Protein Expr Purif*. **40**, 221–229

273. Prinz, W. A., Åslund, F., Holmgren, A., and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *Journal of Biological Chemistry*. **272**, 15661–15667

274. Ren, G., Ke, N., and Berkmen, M. (2016) Use of the SHuffle Strains in Production of Proteins. *Curr Protoc Protein Sci*. **85**, 5.26.1-5.26.21

275. Hatahet, F., Nguyen, V. D., Salo, K. E. H., and Ruddock, L. W. (2010) Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microb Cell Fact*. **9**, 67

276. Nakamoto, H., and Bardwell, J. C. A. (2004) Catalysis of disulfide bond formation and isomerization in the Escherichia coli periplasm. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. **1694**, 111–119

277. Nguyen, V. D., Hatahet, F., Salo, K. E. H., Enlund, E., Zhang, C., and Ruddock, L. W. (2011) Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microb Cell Fact*. 10.1186/1475-2859-10-1

278. Fatima, K., Naqvi, F., and Younas, H. (2021) A Review: Molecular Chaperone-mediated Folding, Unfolding and Disaggregation of Expressed Recombinant Proteins. *Cell Biochemistry and Biophysics 2021 79:2*. **79**, 153–174

279. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C., and San Millán, Á. (2021) Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev Microbiol*. **19**, 347–359

280. Goldbeck, O., and Seibold, G. M. (2018) Construction of pOGOduet – An inducible, bicistronic vector for synthesis of recombinant proteins in Corynebacterium glutamicum. *Plasmid*. **95**, 11–15

281. Talafová, K., Hrabárová, E., and Nahálka, J. (2015) A semi-multifunctional sialyltransferase from Bibersteinia trehalosi and its comparison to the Pasteurella multocida ST1 mutants. *J Biotechnol*. **216**, 116–124

282. Gilbert, M., Karwaski, M.-F., Bernatchez, S., Young, N. M., Taboada, E., Michniewicz, J., Cunningham, A.-M., and Wakarchuk, W. W. (2002) The Genetic Bases for the Variation in the Lipo-oligosaccharide of the Mucosal Pathogen, Campylobacter jejuni. *Journal of Biological Chemistry*. **277**, 327–337

283. Abukar, T., Buenbrazo, N., Janesch, B., Kell, L., and Wakarchuk, W. (2019) Assay methods for the glycosyltransferases involved in synthesis of bacterial polysaccharides. *Methods in Molecular Biology*. **1954**, 215–235

284. Barb, A. W., Meng, L., Gao, Z., Johnson, R. W., Moremen, K. W., and Prestegard, J. H. (2012) NMR characterization of Immunoglobulin G Fc glycan motion on enzymatic sialylation. *Biochemistry*. **51**, 4618

285. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H.-P., and Altmann, F. (2006) Comprehensive glyco-proteomic analysis of human α1-antitrypsin and its charge isoforms. *Proteomics*. **6**, 3369–3380

286. Kolarich, D., Weber, A., Pabst, M., Stadlmann, J., Teschner, W., Ehrlich, H., Schwarz, H.-P., and Altmann, F. (2008) Glycoproteomic characterization of butyrylcholinesterase from human plasma. *Proteomics*. **8**, 254–263

287. Ribeiro, J. P., Pau, W., Pifferi, C., Renaudet, O., Varrot, A., Mahal, L. K., and Imberty, A. (2016) Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form. *Biochemical Journal*. **473**, 2109–2118

288. Connaris, H., Crocker, P. R., and Taylor, G. L. (2009) Enhancing the Receptor Affinity of the Sialic Acid-binding Domain of Vibrio cholerae Sialidase through Multivalency. *J Biol Chem*. **284**, 7339

289. Chiu, C. P. C., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. J. (2004) Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. *Nature Structural & Molecular Biology 2004 11:2*. **11**, 163–170

290. Zha, J., Zang, Y., Mattozzi, M., Plassmeier, J., Gupta, M., Wu, X., Clarkson, S., and Koffas, M. A. G. (2018) Metabolic engineering of Corynebacterium glutamicum for anthocyanin production. *Microb Cell Fact*. **17**, 143

291. Hong, E. J., Park, J. S., Kim, Y., and Lee, H. S. (2014) Role of Corynebacterium glutamicum sprA Encoding a Serine Protease in glxR-Mediated Global Gene Regulation. *PLoS One*. **9**, e93587

292. Lüdke, A., Krämer, R., Burkovski, A., Schluesener, D., and Poetsch, A. (2007) A proteomic study of Corynebacterium glutamicum AAA+ protease FtsH. 10.1186/1471-2180-7-6

293. Engels, S., Schweitzer, J. E., Ludwig, C., Bott, M., and Schaffer, S. (2004) clpC and clpP1P2 gene expression in Corynebacterium glutamicum is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor σH. *Mol Microbiol*. **52**, 285–302

294. Peng, F., Liu, · Xiuxia, Wang, · Xinyue, Chen, J., Liu, · Meng, Yang, Y., and Bai, Z. (2019) Triple deletion of clpC, porB, and mepA enhances production of small ubiquitin-like modifier-N-terminal pro-brain natriuretic peptide in Corynebacterium glutamicum. *J Ind Microbiol Biotechnol*. **46**, 67–79

295. Landeta, C., Boyd, D., and Beckwith, J. (2018) Disulfide bond formation in prokaryotes. *Nature Microbiology 2018 3:3*. **3**, 270–280

296. Holman, N. D. M., Wilkinson, A. J., and Smith, M. C. M. (2021) Alanine-scanning mutagenesis of protein mannosyl-transferase from Streptomyces coelicolor reveals strong activity-stability correlation. *Microbiology (N Y)*. **167**, 1103

297. Newell, J. G., and Czajkowski, C. (2007) Cysteine scanning mutagenesis: Mapping binding sites of ligand-gated ion channels. *Handbook of Neurochemistry and Molecular*

Neurobiology: Practical Neurochemistry Methods. 10.1007/978-0-387-30401-4_21/COVER

298. Halim, A., Larsen, I. S. B., Neubert, P., Joshi, H. J., Petersen, B. L., Vakhrushev, S. Y., Strahl, S., and Clausen, H. (2015) Discovery of a nucleocytoplasmic O-mannose glycoproteome in yeast. *Proc Natl Acad Sci U S A*. **112**, 15648–15653

299. Zhang, B., Liu, Z.-Q., Liu, C., and Zheng, Y.-G. (2016) Application of CRISPRi in Corynebacterium glutamicum for shikimic acid production. *Biotechnol Lett*. **38**, 2153–2161

300. Cleto, S., Jensen, J. V., Wendisch, V. F., and Lu, T. K. (2016) *Corynebacterium glutamicum* Metabolic Engineering with CRISPR Interference (CRISPRi). *ACS Synth Biol*. **5**, 375–385

301. Gauttam, R., Seibold, G. M., Mueller, P., Weil, T., Weiß, T., Handrick, R., and Eikmanns, B. J. (2019) A simple dual-inducible CRISPR interference system for multiple gene targeting in Corynebacterium glutamicum. *Plasmid*. **103**, 25–35

302. Wang, Y., Liu, X., Li, Y., Yang, Y., Liu, C., Linhardt, R. J., Zhang, F., and Bai, Z. (2023) Enhanced production of recombinant proteins in Corynebacterium glutamicum using a molecular chaperone. *J Gen Appl Microbiol*. **69**, 2022.10.002

303. Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012) SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact*. **11**, 1–16

304. Larisch, C., Nakunst, D., Hüser, A. T., Tauch, A., and Kalinowski, J. (2007) The alternative sigma factor SigB of Corynebacterium glutamicum modulates global gene expression during transition from exponential growth to stationary phase. *BMC Genomics*. **8**, 4

305. Matsuda, Y., Itaya, H., Kitahara, Y., Theresia, N. M., Kutukova, E. A., Yomantas, Y. A. V., Date, M., Kikuchi, Y., and Wachi, M. (2014) Double mutation of cell wall proteins CspB and PBP1a increases secretion of the antibody Fab fragment from Corynebacterium glutamicum. *Microb Cell Fact*. **13**, 56

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

307. Zhang, L., Lin, X., Wang, T., Guo, W., and Lu, Y. (2021) Development and comparison of cell-free protein synthesis systems derived from typical bacterial chassis. *Bioresour Bioprocess*. **8**, 58

308. Kuruma, Y., and Ueda, T. (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc*. **10**, 1328–1344

309. Huang, M., Zhao, Y., Feng, L., Zhu, L., Zhan, L., and Chen, X. (2020) Role of ClpB From Corynebacterium crenatum in Thermal Stress and Arginine Fermentation. *Front Microbiol*. 10.3389/fmicb.2020.01660

310. Lo, A., and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR–Cas systems. *F1000Res*. **6**, 747

Appendix A

Supplemental to Chapter 1 (S1)

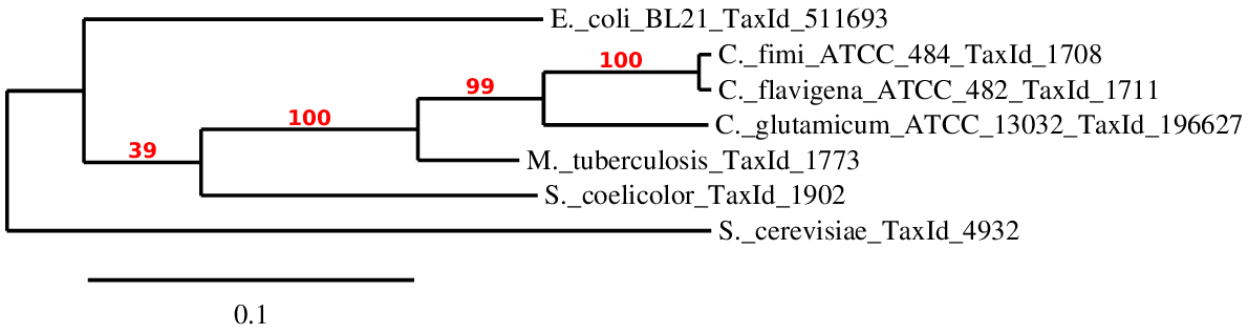
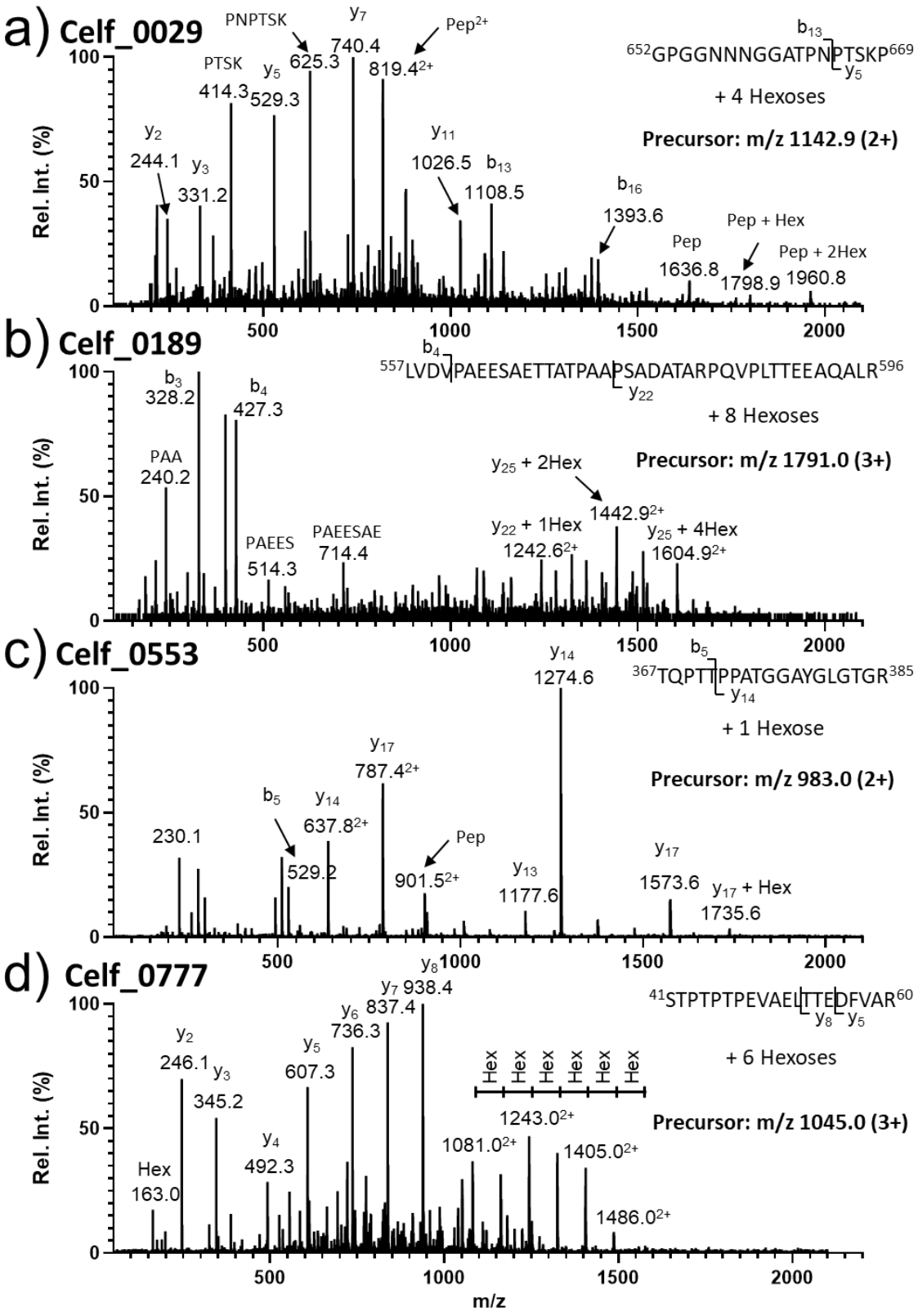
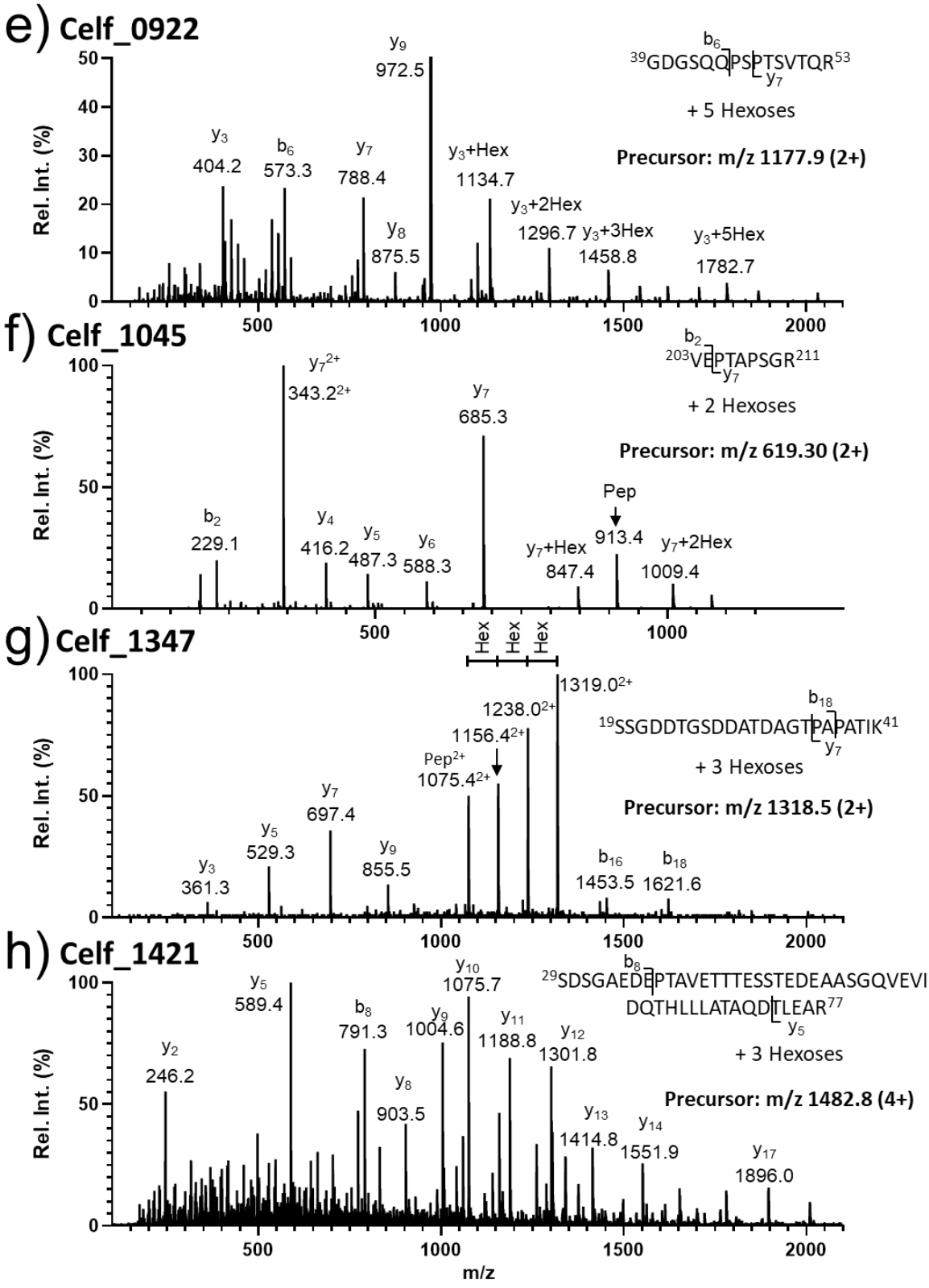


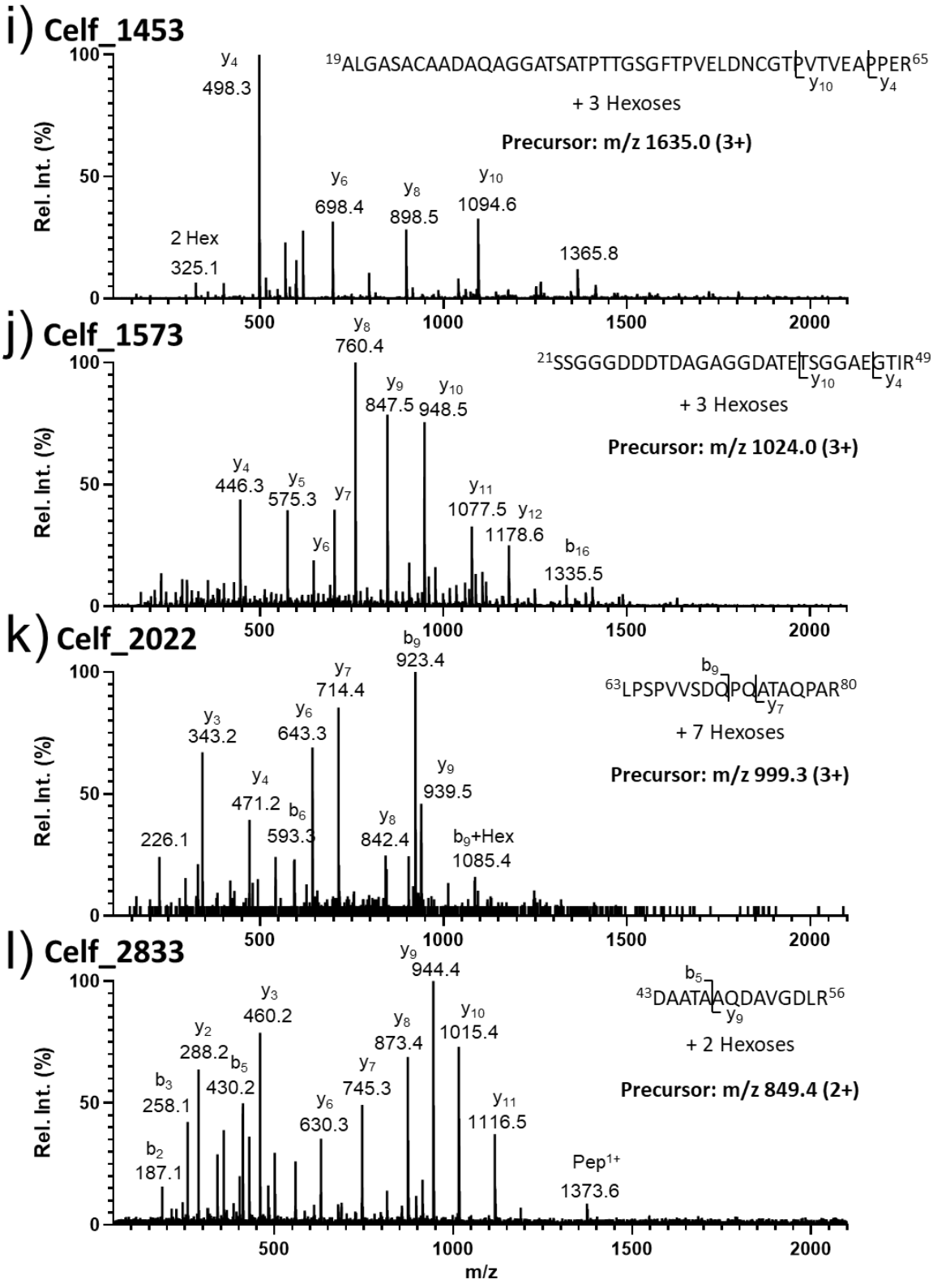
Figure S2. 1. Phylogenetic tree showing the relatedness of mannosylating actinobacteria based on 16S rRNA gene sequences.

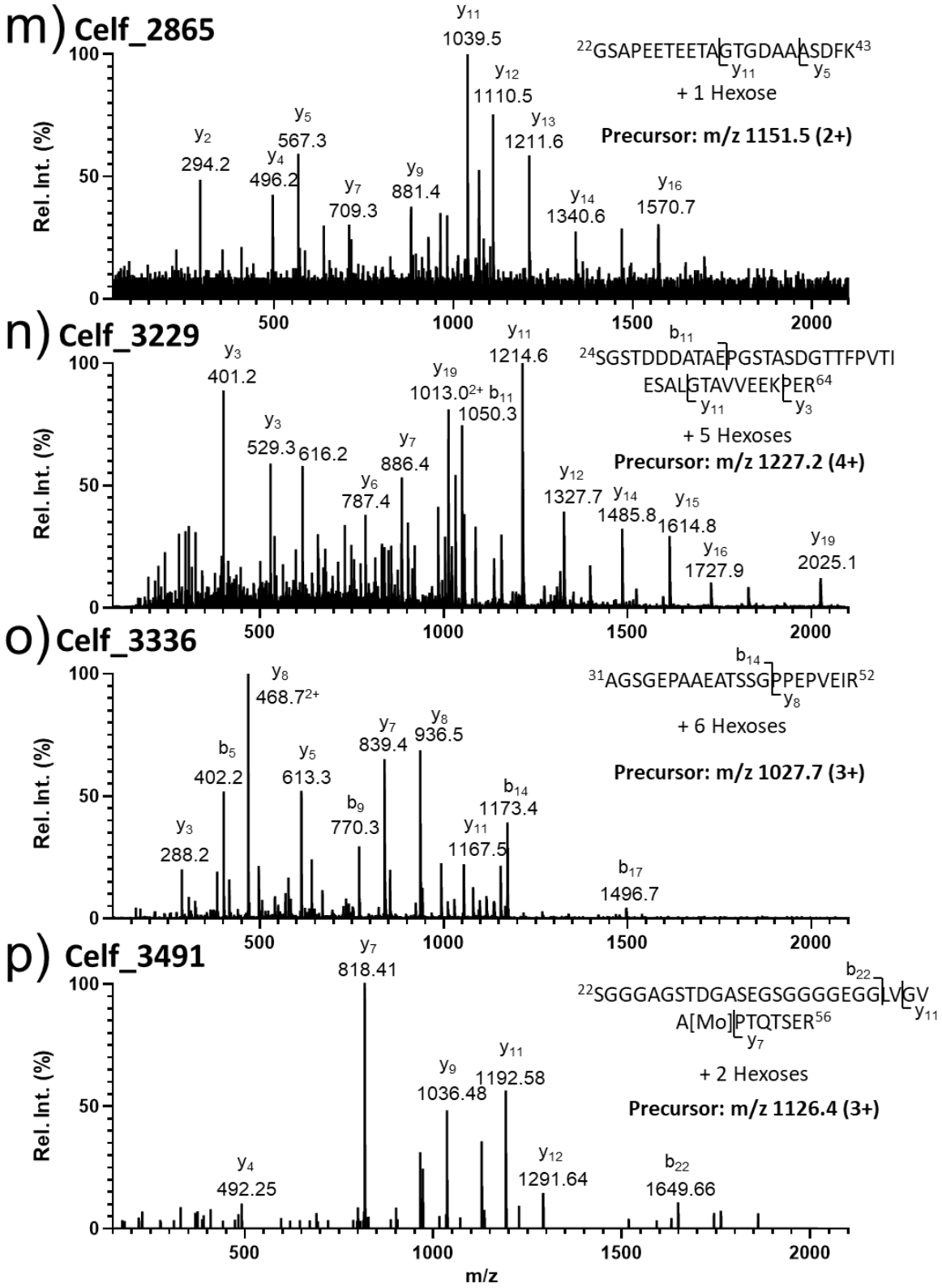
C. glutamicum is more closely related to other mannosylating actinobacteria than the traditional recombinant host *E. coli* and would therefore serve as an ideal host (genetically) for investigating actinobacterial protein-*O*-mannosylation. PMTs all belong to the GT-39 family, but there is significant genetic distance separating the prokaryotic and eukaryotic members of this family. *E. coli* was used as the outgroup. Branch support values (%) are shown in red. Bar, 0.1% sequence divergence, created using TreeDyn (Dereeper, 2008).

Spectra for the MS/MS data for the *Cellulomonas* derived glycopeptides.




e) Celf_0922
Rel. Int. (%)
y9 972.5
y3 404.2
b6 573.3
y7 788.4
y8 875.5
y3+Hex 1134.7
y3+2Hex 1296.7
y3+3Hex 1458.8
y3+5Hex 1782.7
b6 y7
39GDGSQQPSPTSVTQR53
+ 5 Hexoses
Precursor: m/z 1177.9 (2+)
f) Celf_1045
Rel. Int. (%)
y7 2+ 343.2 2+
y7 685.3
b2 229.1
y4 416.2
y5 487.3
y6 588.3
y7+Hex 847.4
Pep 913.4
y7+2Hex 1009.4
b2 y7
203VEPTAPSGR211
+ 2 Hexoses
Precursor: m/z 619.30 (2+)
g) Celf_1347
Rel. Int. (%)
Hex Hex Hex
1319.0 2+
1238.0 2+
1156.4 2+
Pep 2+ 1075.4 2+
y7 697.4
y5 529.3
y3 361.3
y9 855.5
b16 1453.5
b18 1621.6
b18 y7
19SSGDDTGSDDATDAGTPAPATIK41
+ 3 Hexoses
Precursor: m/z 1318.5 (2+)
h) Celf_1421
Rel. Int. (%)
y10 1075.7
y5 589.4
b8 791.3
y9 1004.6
y11 1188.8
y12 1301.8
y2 246.2
y8 903.5
y13 1414.8
y14 1551.9
y17 1896.0
b8 y5
29SDSGAEDEPTAVETTTESSTEDEAASGQVEVI
DQTHLLLATAQDTLEAR77
+ 3 Hexoses
Precursor: m/z 1482.8 (4+)
m/z


i) Celf_1453
19ALGASACAADAQAGGATSATPTTGSGFTPVELDNCGTPVTVEAPPER65
y4
498.3
y10
y4
+ 3 Hexoses
Precursor: m/z 1635.0 (3+)
y6
698.4
y8
898.5
y10
1094.6
2 Hex
325.1
1365.8
Rel. Int. (%)
j) Celf_1573
21SSGGGDDDTDAGAGGDATETSGGAEGTIR49
y10
y4
+ 3 Hexoses
Precursor: m/z 1024.0 (3+)
y8
760.4
y9
847.5
y10
948.5
y4
446.3
y5
575.3
y7
y6
y11
1077.5
y12
1178.6
b16
1335.5
k) Celf_2022
63LPSPVVSDQPQATAQPAR80
b9
y7
+ 7 Hexoses
Precursor: m/z 999.3 (3+)
b9
923.4
y7
714.4
y3
343.2
y6
643.3
y4
471.2
b6
593.3
y9
939.5
y8
842.4
b9+Hex
1085.4
226.1
l) Celf_2833
43DAATAAQDAVGDLR56
b5
y9
+ 2 Hexoses
Precursor: m/z 849.4 (2+)
y9
944.4
y3
460.2
y2
288.2
b5
430.2
y8
873.4
y10
1015.4
y7
745.3
b3
258.1
y6
630.3
y11
1116.5
b2
187.1
Pep1+
1373.6
100
50
0
500
1000
1500
2000
m/z


m) Celf_2865
^{22}GSAPEETEETAGTGDAAASDFK^{43}
+ 1 Hexose
Precursor: m/z 1151.5 (2+)
n) Celf_3229
^{24}SGSTDDDATAEPGSTASDGTTFPVTI
ESALGTAVVEEKPER^{64}
+ 5 Hexoses
Precursor: m/z 1227.2 (4+)
o) Celf_3336
^{31}AGSGEPAAEATSSGPPEPVEIR^{52}
+ 6 Hexoses
Precursor: m/z 1027.7 (3+)
p) Celf_3491
^{22}SGGGAGSTDGASEGSGGGGEGGLVGV
A[Mo]PTQTSER^{56}
+ 2 Hexoses
Precursor: m/z 1126.4 (3+)
Rel. Int. (%)
m/z

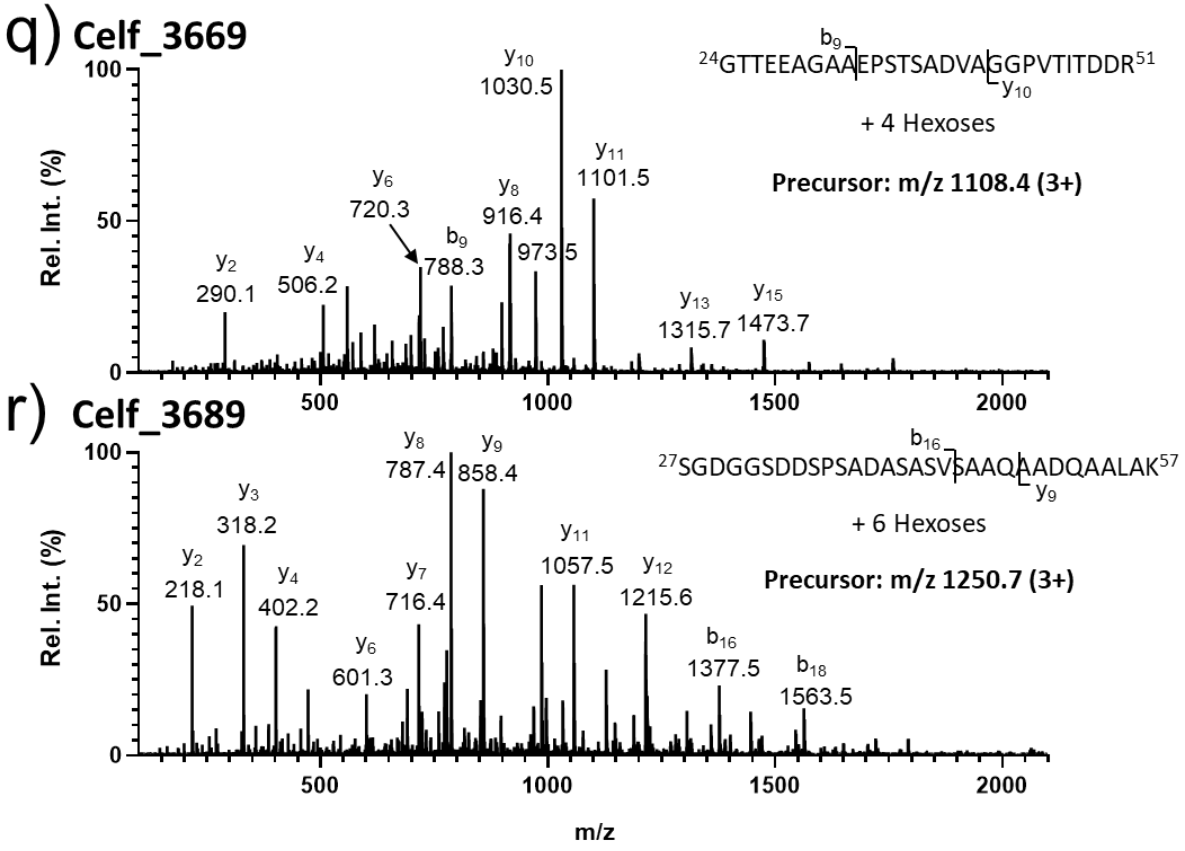


Figure S2. 2. NanoLC-CID-MS/MS spectra of tryptic glycopeptides identified in the enriched fractions from *C. fimi*.

MS/MS spectra of 19 *Cellulomonas* derived glycopeptides. Identified glycopeptides were observed to be modified by up to 8 hexoses**.** The proteins from which they originate are listed in Table 2. 1.

Figure S2. 3. Features of *C. fimi* test proteins expressed in *C. glutamicum*.

A)

>Celf_3184 GH6 Endoglucanase with CBM2a

MSTRRTAAALLAAAAVAVGGLTALTTTAAQAAPGCRVDYAVTNQWPGGFGANVTITNLG

DPVSSWKLDWTYTAGQRIQQLWNGTASTNGGQVSVTSLPWNGSIPTGGTASFGFNGSW

AGSNPTPASFSLNGTTCTGTVPTTS**PTPTPTPTTPTPTPTPTPTPTPT**VTPQPTSGFYVDP

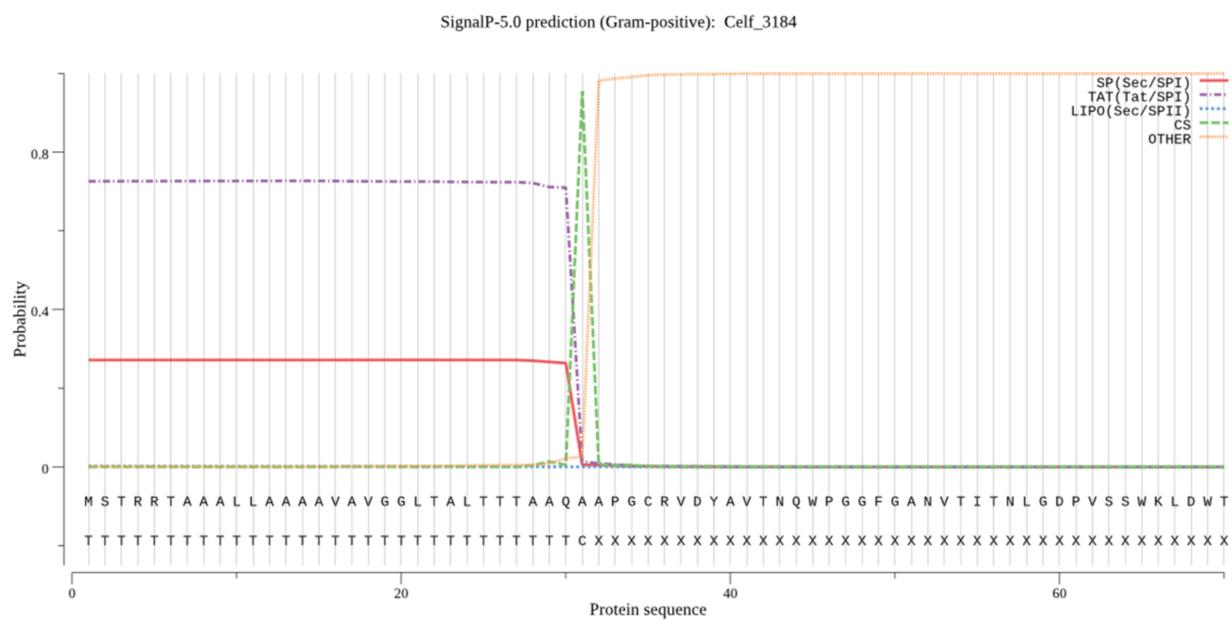
TTQGYRAWQAASGTDKALLEKIALTPQAYWVGNWADASHAQAEVADYTGRAVAAGKT

PMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNPIVILEPDALAQLGDCSGQ

GDRVGFLKYAAKSLTLKGARVYIDAGHAKWLSVDTAVNRLNQVGFEYAVGFALNTSNY

QTTADSKAYGQQISQRLGGKKFVIDTSRNGNGSNGEWCNPRGRALGERPVAVNDGSGL

DALLWVKLPGESDGACNGGPAAGQWWQEIALEMARNARWHHHHHH*



Features of Celf_3184: TAT leader from aa 1-31 (0.7266 probability from SignalP5.0), CBM2a from aa 31 to 165 including the “PT” linker shown in bold and the GH6 catalytic domain from aa 165 – 449. Previously, the “PT” linker had been shown to be the site of glycosylation in *C. fimi*. A PhD thesis from the University of British Columbia where the original work on *Cellulomonas*

glycosylation was performed shows Celf_3184 made by *C. fimi* (then called EngA or CenA) had 10% by weight mannose, so 51 kDa / 10 = 5100 Da, this in turn is 5100 Da/162 Da per hexose = 31.5 hexoses - so around 30 hexoses per protein molecule. (Langsford 1988, PhD thesis UBC)

B)

>Celf_2022 Annotation as peptidyl-prolyl cis-trans isomerase cyclophilin type

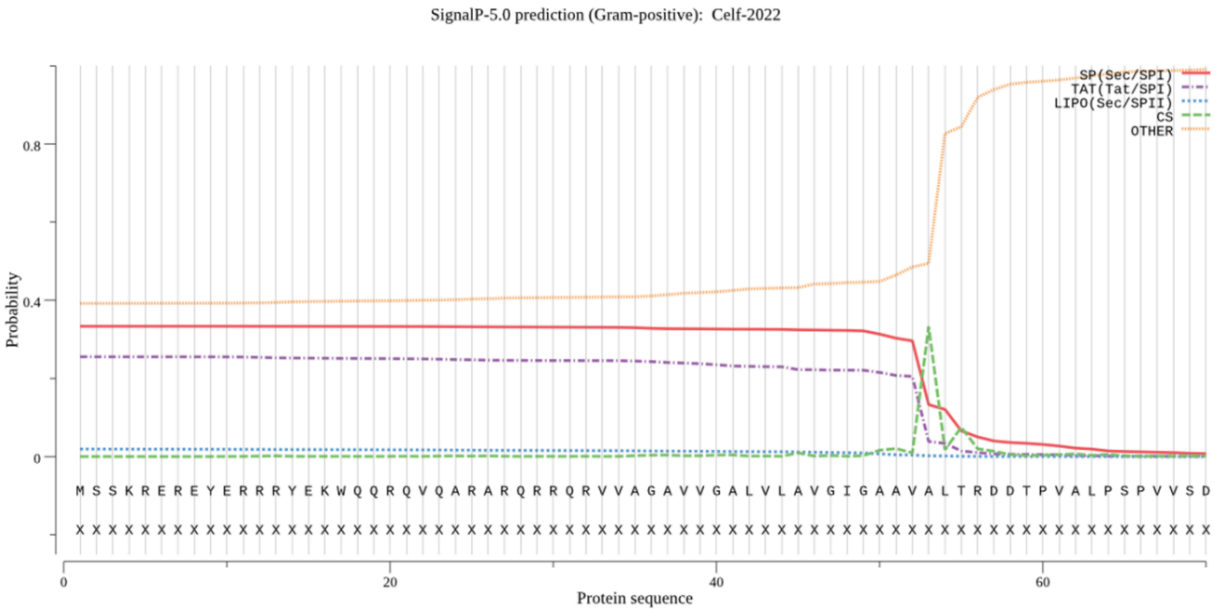
MSSKREREYERRRYEKWQQRQVQARARQRRQRVVAGAVVGALVLAVGIGAAVALTR**D**

DTPVALPSPVVSDQPQATAQPARTGNGGVVPEAALAEGRTWTGTLALSQGDVAIELDG

AAAPQAVANFVTLAREGYFDGTSCHRLVTGGIHVLQCGDPTATGQGGPGYSWGPVENA

PSDDVYPAGTIAMARVGGDGSSMGSQFFLVYEDSQIPSDSAGGYTVFGRITSGLDVVQAI

ADAGTVEGSETPVDDVIIEGVETQHHHHHH*



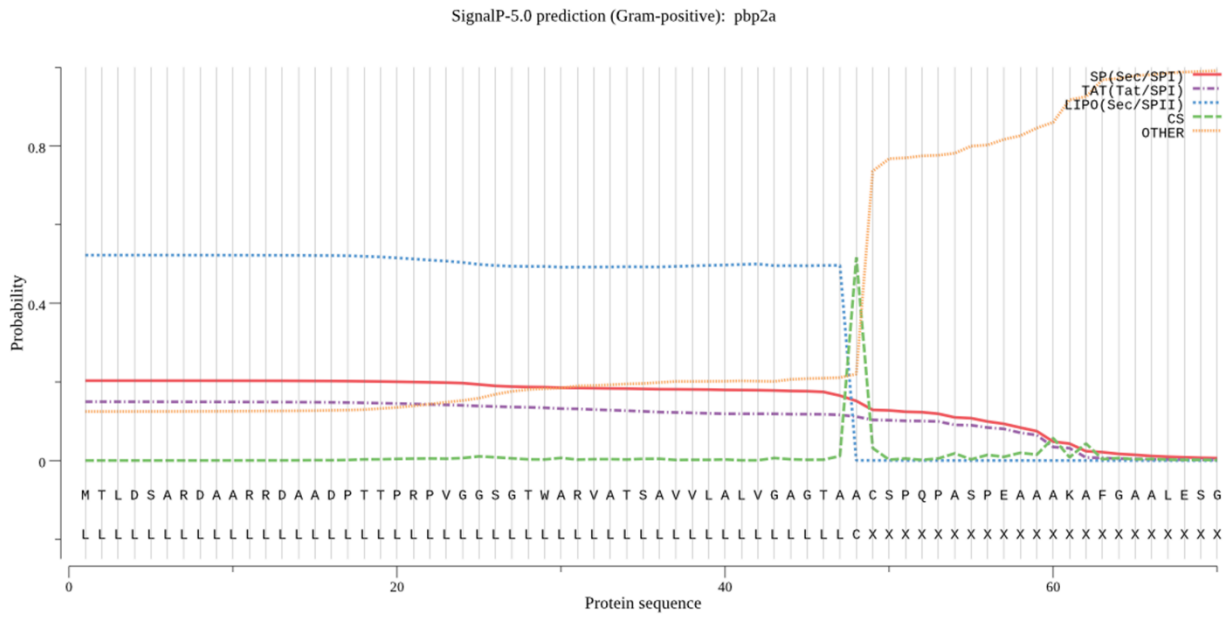
Features of Celf_2022: This protein from *C. fimi*, lacks a strong signal peptide according to the SignalP5.0 algorithm. Despite that bioinformatic clue, we find Celf_2022 almost exclusively in the culture supernatant when expressed in *C. glutamicum*. The glycopeptide is shown in bold; the apparent leader sequence is underlined. When expressed and purified from *E. coli* there is

essentially one species of protein at 21474 Da which corresponds to Δ53aa (see Figure S4 below). There is one weak orthologue from *C. glutamicum,* however the sequence identity is only 35% over the whole protein.

C)

>Celf_0189 Annotation as putative penicillin binding transpeptidase PBP2a

MTLDSARDAARRDAADPTTPRPVGGSGTWARVATSAVVLALVGAG**TAAC**SPQPASPEA
AAKAFGAALESGDFSGVELAAGTDAAAVVESRATAFEGLTPWKPDVAFDGVTSPDDDP
DVATASYTFTWDVDDRDADWTYTTHAALARDEKDPDVWRVTWKRSALVPDLTDAEIL
RVTREQAPRGRVLGAGGAVIVEPRNVSRVGVDKTLTDAAGQGAAARELAAQLAMDAE
AYAQRVAGAGEKAFVEAIVVREGDPAYNLATLTAIPGVRAVPDTLPLAPTRRFARPILGTV
GAATAEIVEKSQGQIAAGDLTGLSGLQRQYDALLRGNPALTVEAVSPDGAAVRPLFQTEP
TPGADLVTTLDPRLQDGAEEILAGVGPASAIVAIRPSTGDVLAAASGPGGEGMSTATLGQ
YAPGSTFKVASALALLRSGATPATTFSCPPTVAVDGREFSNYPDYPSSALGDVPLSTAFAQS
CNTAFISARDSAPQTALVDAAGSLGLVPDADLAFAAFLGAVPSESTGTDHAASMIGQGRV
LASPLGMATVAASVAAGQTVTPR**LVDVPAEESAETTATPAAPSADATARPQVPLTTEEA**
QALRQLMRGVVTEGGGTFLQDVPGGEVAAKTGTAQFGDAANLQNHVWMIAIQGDLAV
AVFVDVGEYGSTTAGPLLEDFLRYAAGHHHHHH*



Features of Celf_0189: This protein appears to have a lipoprotein leader (probability 0.52 from SignalP 5.0, underlined) and the putative lipoprotein consensus site is shown in bold. The identified glycopeptide is also shown in bold.

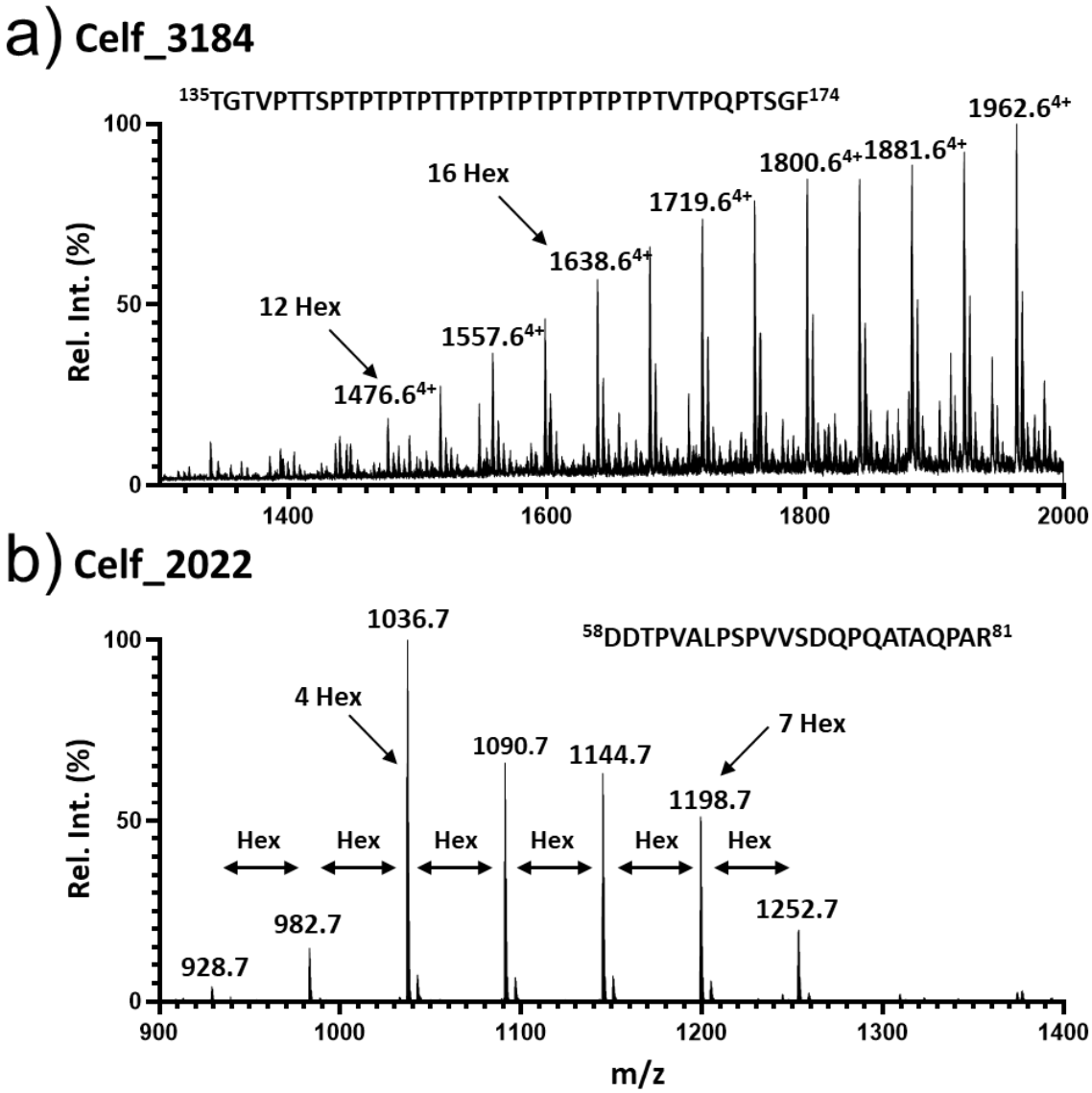


Figure S2. 4. NanoLC-MS analysis of the glycosylation of Celf_3184 and Celf_2022, both of which were expressed in *C. glutamicum*.

NanoLC-MS spectra of HILIC-enriched glycopeptides from (a) the chymotryptic digest of Celf-3184 and (b) the tryptic digests of Celf_2022. Their peptide sequences are presented in the panels and were confirmed by MS/MS analysis (data not shown).

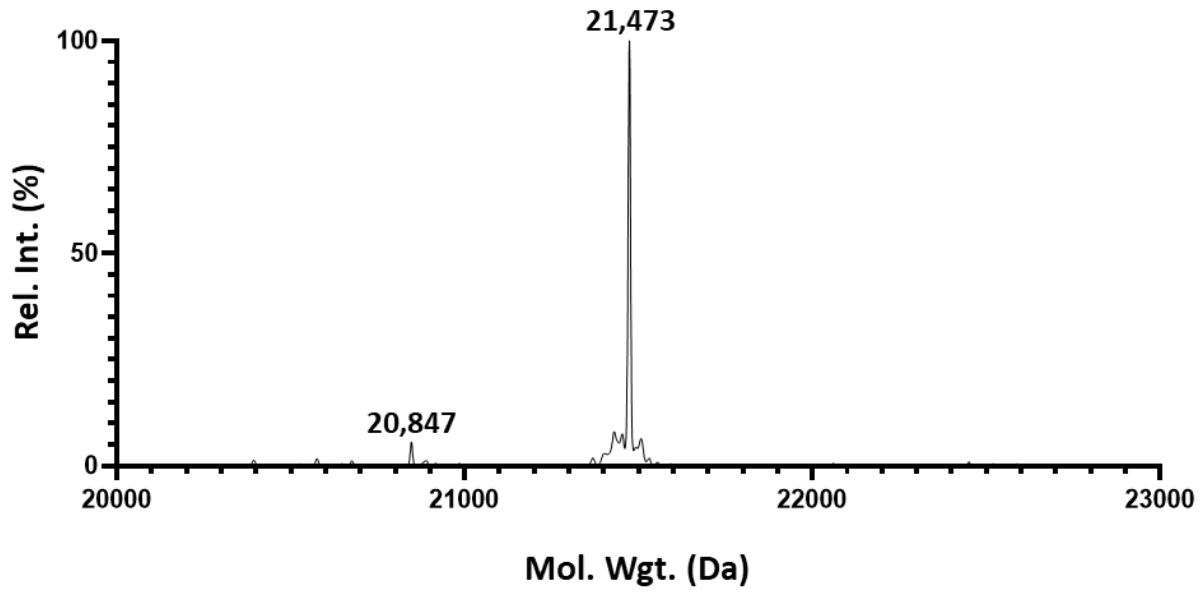


Figure S2. 5. Intact Mass LC-MS of Celf_2022 expressed in *E. coli.*

The calculated mass (signal peptide removed) for this protein is 21,474 Da which agrees well with the observed mass.

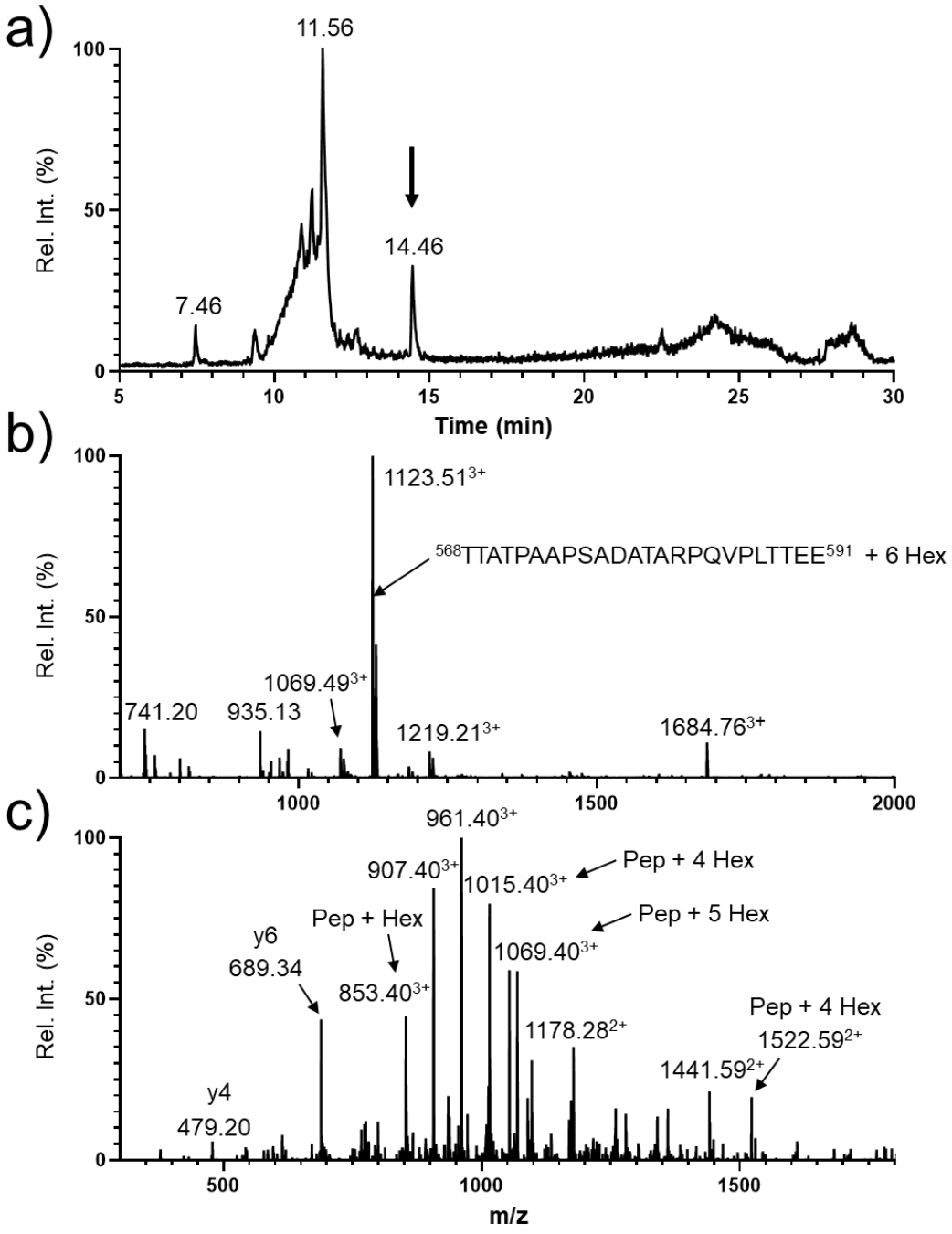


Figure S2. 6. NanoLC-MS/MS analysis of glycopeptides enriched from the endoproteinase Glu-C digest of Celf_0189 using Polyhydroxyethyl A media.

(A) The total ion chromatogram (m/z 400-2000). (B) Mass spectrum for the peak at 14.46 minutes which is dominated by the triply charged ion at m/z 1123.51. Panel C: CID-MS/MS

spectrum for m/z 1123.51 confirming that it corresponds to the Celf_0189 peptide, ^{568}TTATPAAPSADATARPQVPLTTEE^{591}, modified with 6 hexoses. In fact, all the earlier peaks observed in the TIC (panel A) are glycopeptides with shorter peptides derived from this same region of Celf_0189.

5’**CATATG**TCCAGCAAGCGCGAACGTGAGTACGAGCGCCGTCGTTACGAGAAA

TGGCAGCAGCGCCAGGTCCAGGCACGTGCACGTCAACGTCGTCAACGTGTCG

TCGCAGGTGCAGTAGTCGGTGCACTGGTACTGGCAGTGGGTATCGGTGCAGCT

GTGGCACTGACTCGTGACGACACTCCTGTGGCTCTGCCTAGCCCAGTGGTGAG

CGATCAACCACAGGCTACCGCTCAGCCGGCTCGTACTGGTAACGGTGGTGTGG

TGCCGGAAGCTGCTCTGGCTGAAGGTCGTACTTGGACTGGTACGCTGGCTCTG

AGCCAGGGTGATGTGGCGATCGAACTGGATGGTGCGGCCGCCCCGCAGGCGG

TTGCGAACTTCGTTACCCTGGCGCGTGAAGGTTACTTCGATGGTACGTCTTGTC

ATCGCCTGGTTACCGGTGGTATCCACGTTCTGCAGTGCGGTGATCCGACCGCG

ACCGGTCAGGGTGGCCCGGGCTACTCTTGGGGCCCGGTTGAAAATGCGCCGT

CTGATGATGTATACCCGGCGGGCACGATCGCGATGGCCCGCGTAGGCGGCGAC

GGCTCTTCTATGGGCTCTCAGTTCTTCCTGGTTTATGAAGACAGCCAGATCCCG

TCCGACTCCGCCGGCGGCTATACCGTTTTTGGCCGCATCACCTCCGGCCTGGA

CGTAGTTCAGGCGATTGCCGACGCGGGCACCGTTGAAGGCTCCGAAACCCCG

GTTGACGACGTAATTATTGAAGGCGTTGAAACCCAG**CACCATCACCATCACC**

ATTGATAA**GAATTC** 3’

Figure S2. 7. Synthetic *celf_2022* gene sequence.

The synthetic gene sequence of synthesized *celf_2022.* Show in bold are restriction sites (5’ NdeI, 3’ EcoRI), tandem stop codons are underlined, and C-terminal HIS^6 tag is shown in blue.

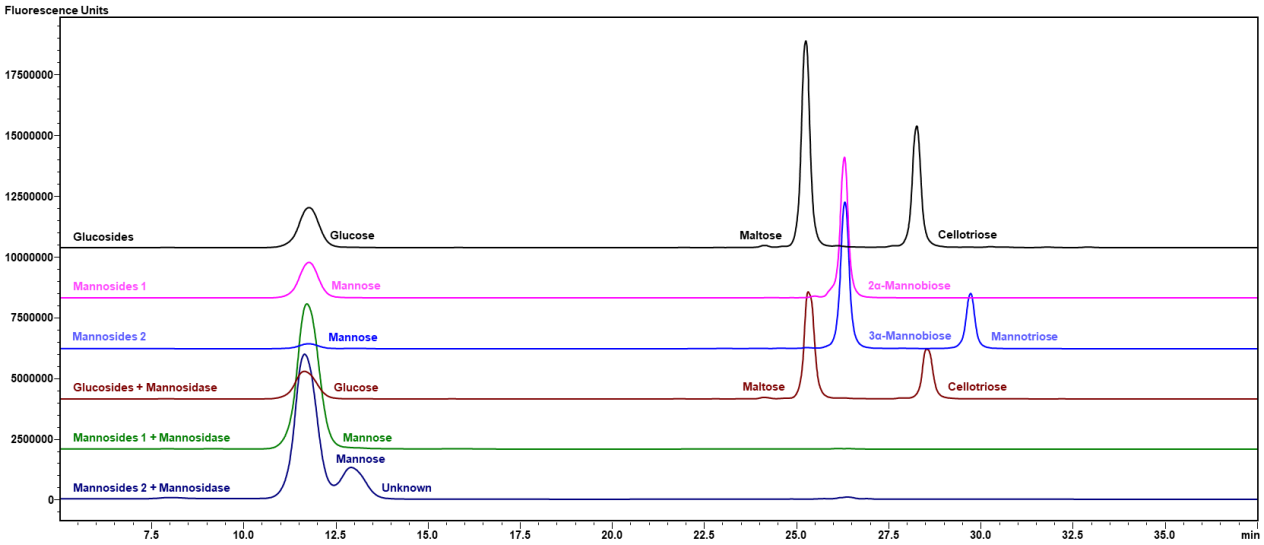


Figure S2. 8. O-glycan standards as 2-aminobenzamide labeled sugars, analyzed by HILIC HPLC.

Mixtures of mono-, di- and trisaccharides were labelled with 2-AB. These were subsequently treated with α-mannosidase to show where the monosaccharide runs, and that the enzyme is specific. The order of the analysis for HPLC traces from top to bottom is: glucoside standards; mannose and α-1,2-mannobiose; mannose, α1,3-mannobiose, and α-1,3-mannotriose. The bottom three traces are the same samples treated with jackbean mannosidase.

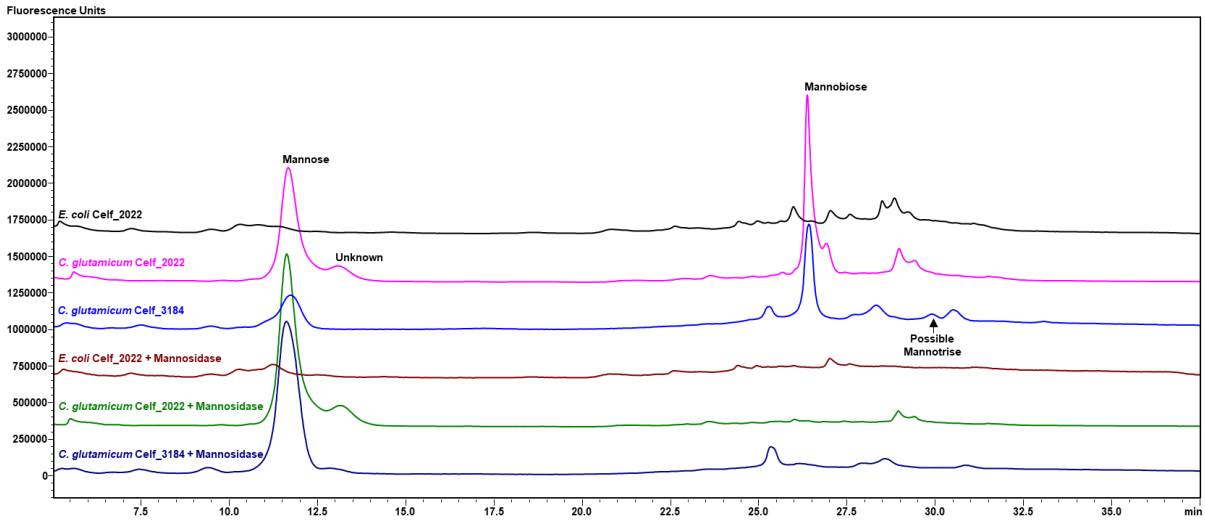


Figure S2. 9. HPLC analysis of 2AB labelled glycans released from *C. glutamicum* expressed Celf_2022 and Celf_3184.

Cleaved glycans were treated with α-mannosidase to show that the disaccharides were in fact mannose terminated. The top trace is glycans from Celf_2022 isolated from an *E. coli* lysate (non-glycosylated), The second trace is glycans from culture supernatant purified Celf_2022 produced in *C. glutamicum*. The third trace is glycans from culture supernatant purified Celf_3184 produced in *C. glutamicum*. The bottom three traces are the same samples treated with jack bean mannosidase.

Table S2. 1. Blast analysis of *Cellulomonas* glycoproteins across related actinobacteria.

Gene/ Protein	Annotation	*C. flavigena taxid:1711*	*C. glutamicum taxid:1718*	*S. coelicolor taxid:100226*	*S. coelicolor taxid:1902*	*M. tuberculosis taxid:1773*
Celf_0029	**Stk1 family PASTA domain-containing Ser/Thr kinase**	8E-134 WP_048772469.1	5E-95 WP_220313280.1	5E-126 WP_011029269.1	4E-128 NUV56788.1	2E-121 CNE63184.1
Celf_0189	**Penicillin-binding protein transpeptidase**	0.00E+00 WP_013118662.1	2.00E-40 HJE09866.1	4.00E-34 WP_011029224.1	1.00E-36 NUV54027.1	4.00E-61 CNG04032.1
Celf_0553	**C40 family (non-peptidase)**	4E-54 ADG73525.1	**2E-18 WP_065366685.1**	**3E-22 WP_011029498.1**	**6E-22 NSL83314.1**	**1E-18 CNG09295.1**
Celf_0777	**Hypothetical protein**	1E-37 WP_013118564.1	No Match	No Match	No Match	No Match
Celf_0922	**Hypothetical protein**	6E-144 WP_013117944.1	No Match	5E-30 WP_011031120.1	2E-34 NUV53951.1	**9E-22 CNG01146.1**
Celf_1045	**Hypothetical protein**	**2E-15 WP_148234365.1**	No Match	No Match	No Match	No Match
Celf_1347	**Extracellular solute binding protein family 1**	7E-79 WP_239659478.1	**1E-11 WP_038583938.1**	5E-55 WP_011027184.1	8E-55 WP_202492664.1	8E-21 CNE20358.1
Celf_1421	**Hypothetical protein**	No Match	No Match	No Match	No Match	No Match
Celf_1453	**Periplasmic binding protein, putative F420-0 ABC transporter substrate-binding protein**	3E-135 WP_013116660.1	No Match	1E-54 WP_011027505.1	3E-54 WP_202492748.1	3E-35 SGC68319.1
Celf_1573	**Extracellular solute-binding protein family 3**	2E-58 WP_013116782.1	2E-60 WP_040967531.1	7E-81 WP_011030447.1	9E-81 BDD75250.1	9E-66 CNE19607.1

Celf_1830	**Extracellular ligand binding receptor**	0E+00 ADG74871.1	No Match	No Match	No Match	No Match
Celf_2022	**Peptidyl-prolyl cis-trans isomerase cyclophilin type**	5E-109 WP_013117035.1	1E-32 WP_208396298.1	6E-40 WP_003977316.1	1E-39 WP_202492880.1	2E-39 CNF85300.1
Celf_2833	**Hypothetical protein**	No Match	No Match	No Match	No Match	No Match
Celf_2865	**Hypothetical protein (BMP family ABC transporter substrate-binding protein)**	9E-171 WP_013116260.1	No Match	No Match	No Match	No Match
Celf_3229	**Periplasmic binding protein, iron-siderophore ABC transporter substrate-binding protein**	8E-58 WP_013117801.1	3E-30 QYO72903.1	**2E-24** **WP_003978370.1**	**5E-24** **WP_003978370.1**	6E-28 CNG07028.1
Celf_3336	**Extracellular solute-binding family 1**	No Match	No Match	7E-39 WP_011027838.1	2E-38 WP_011027838.1	1E-96 CNF54055.1
Celf_3491	**periplasmic binding protein/LacI transcriptional regulator (sugar-binding protein)**	2E-159 WP_013117736.1	No Match	3E-103 WP_003976402.1	8E-103 WP_003976402.1	7E-130 CNF28068.1
Celf_3669	**Periplasmic binding protein – Iron uptake**	No Match	No Match	5E-89 WP_003971744.1	2E-90 NUV51519.1	**5E-15** **MBR7503274.1**
Celf_3689	**Peptidylprolyl isomerase FKBP-type**	2E-70 WP_013118439.1	**2E-13** **WP_060564232.1**	**1E-22** **WP_003977188.1**	**2E-22** **WP_202492429.1**	4E-29 CNE51518.1

There are two distinct sequences for *S. coelicolor,* so we included both in the search.
Taxid:100226 and Taxid:1902

Appendix B

Supplemental to Chapter 2 (S2)

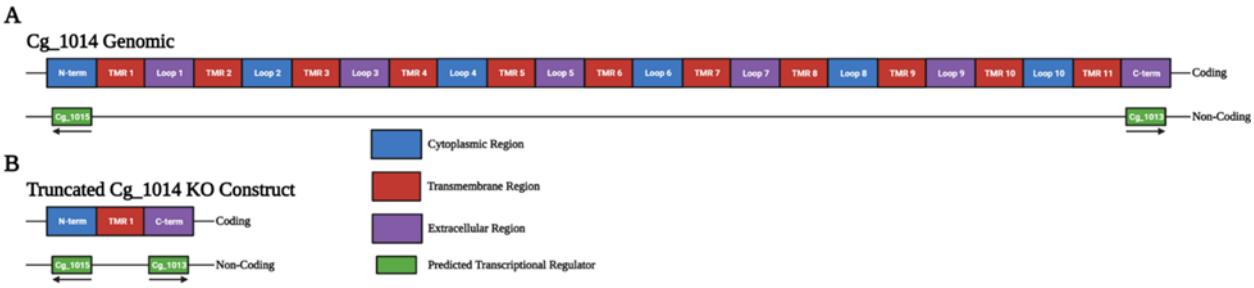


Figure S3. 1. Schematic and predicted topology of genomic Cg_1014 and predicted transcriptional regulators on the non-coding strand (A) maintained in the inactivated knockout mutant (B).

The *C. glutamicum* GT-39 Cg_1014 contains (as predicted by TMHMM-2.0) 11 transmembrane regions (TMR, red), 5 cytoplasmic loops (blue), and 6 extracellular loops (purple). The predicted cytoplasmic N-terminal and extracellular C-terminal domains also contain predicted transcriptional regulators for the neighbouring genes Cg_1013 and Cg_1015 (A). As catalytic activity is thought to be harboured on extracellular Loop 1 a truncated and inactive mutant was designed (B) containing only the N- and C-terminal domains connected by TMR 1, maintaining down- and upstream transcriptional effectors.

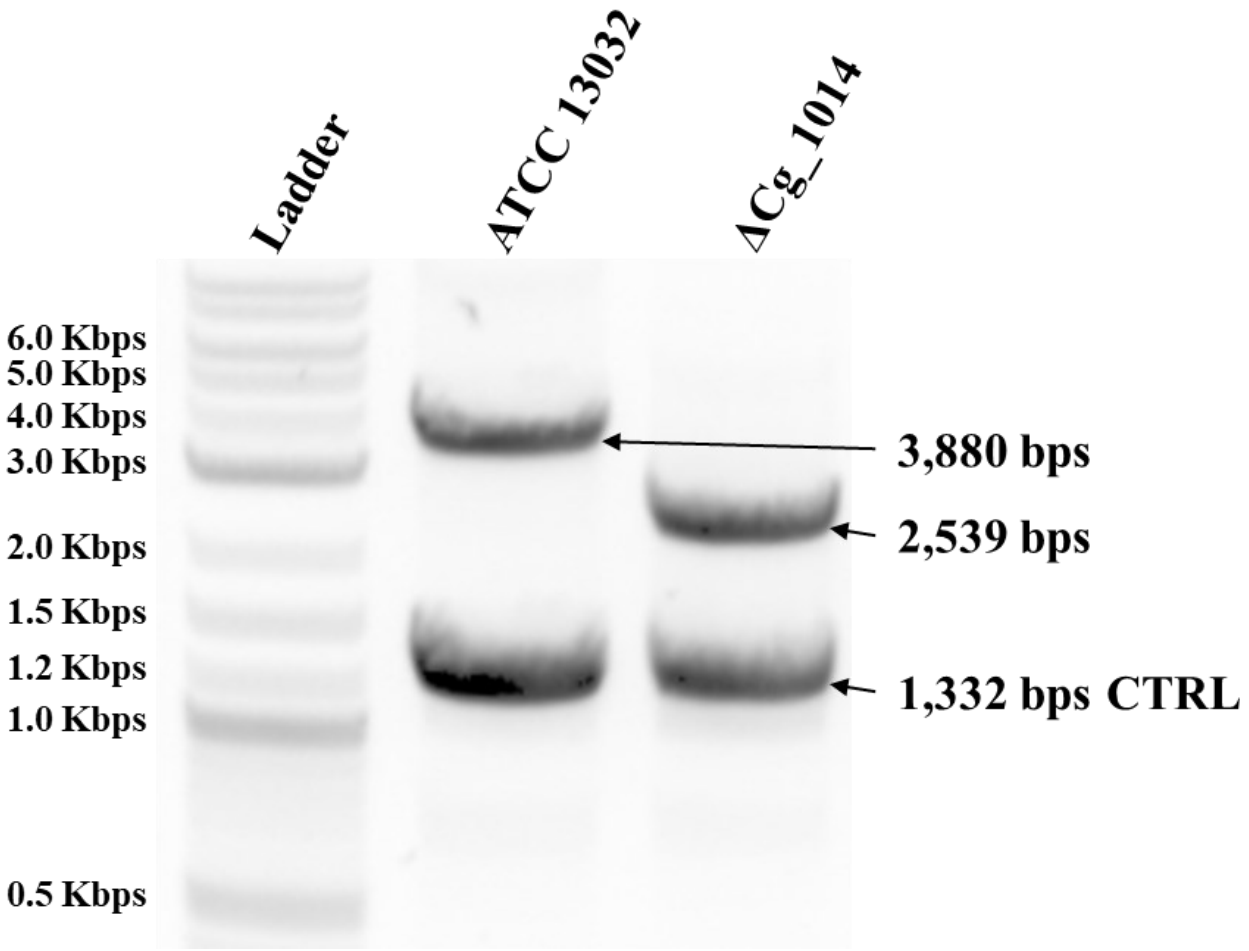


Figure S3. 2. 0.8% agarose gel showing genomic knockout of GT-39 in *C. glutamicum* ATCC 13032 and ΔCg_1014.

Primers specific to the flanking regions ≈ 1,000 bps up- and downstream of Cg_1014 were used to identify successful homologous recombination events. In ATCC 13032 this amplicon is 3,880 bps and when Cg_1014 is replaced by the truncated inactive sequence the amplicon is 2,539 bps. This decrease exactly corresponds to the 1,341 bps removed from Cg_1014 (Δ59 – 505 aa) to generate the mutant gene. An additional amplicon of 1,332 was also produced using primers specific for the *C. glutamicum* homologue of the maltose binding protein (MBP) as a control.

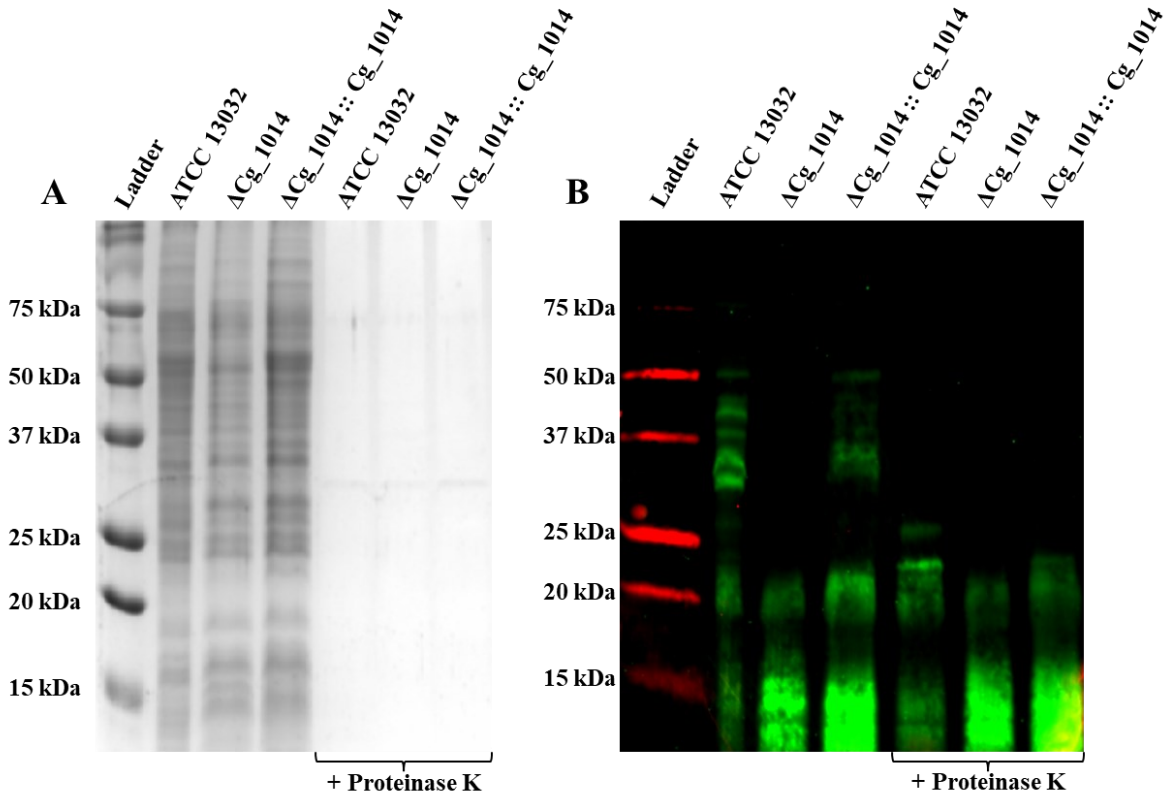


Figure S3. 3. Coomassie stained 15% SDS-PAGE (A) and ConA-FITC (green) lectin blot (B) of *C. glutamicum* ATCC 13032, ΔCg_1014, and ΔCg_1014:Cg_1014 membrane fractions following digestion with proteinase K.

Coomassie stained gel (A) shows membrane fractions before (Lanes 2 – 4) and after proteinase K digestion (Lanes 5 – 7). Digestion of proteins in each membrane fraction is evident by lack of Coomassie strained bands in proteinase K-treated samples (A). ConA reactive smears in lectin blot (B) of membrane fractions from *C. glutamicum* ATCC 13032, ΔCg_1014, and complemented strains are attributed to the presence of LAM in the samples (low molecular weight smears present in ConA blot). Distinct bands that remain in ATCC 13032 and ΔCg_1014:Cg_1014 membrane fractions following proteinase K digest (Lanes 5- 7) are

attributed to protease resistant mannoproteins as POM is known to confer proteolytic resistance. Molecular weight standards are the Bio-Rad All Blue ladder.

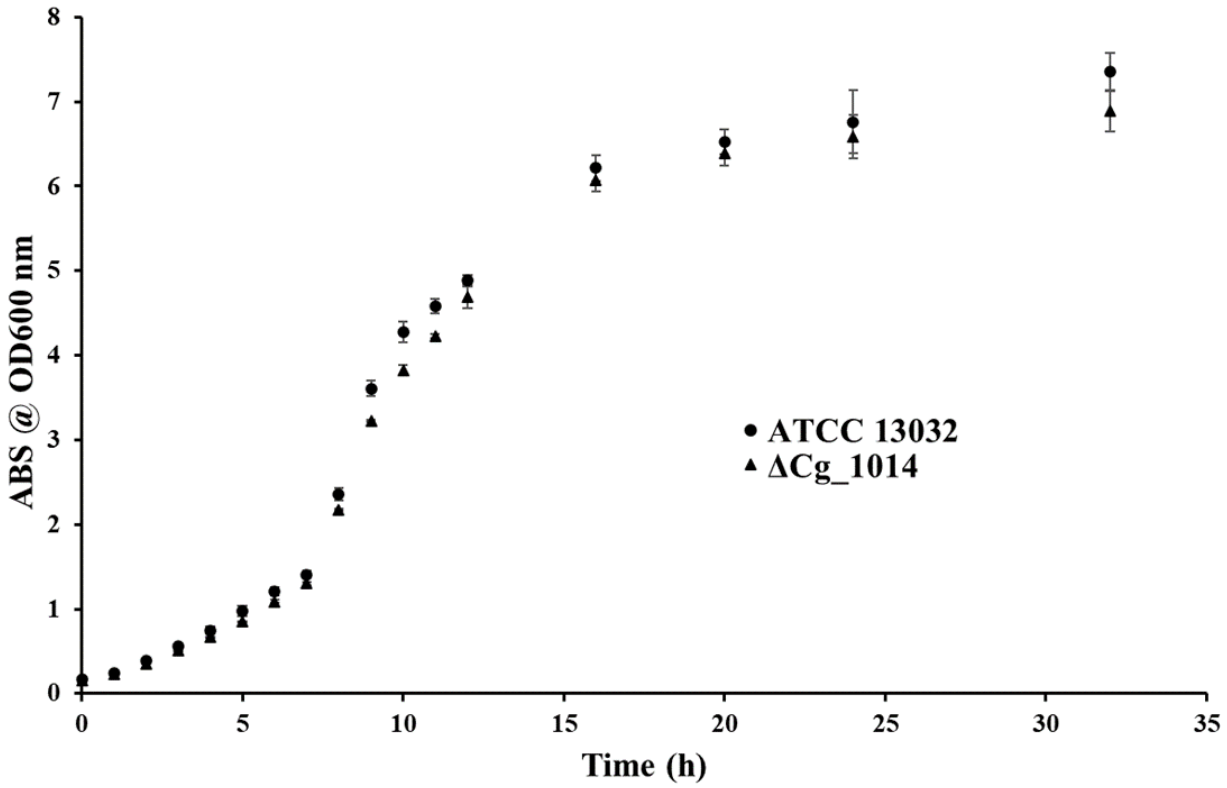


Figure S3. 4. Effects of GT-39 inactivation on the growth of *C. glutamicum* ΔCg_1014.

No significant differences to the ATCC 13032 strain were evident at any stage of growth in the ΔCg_1014 strain.

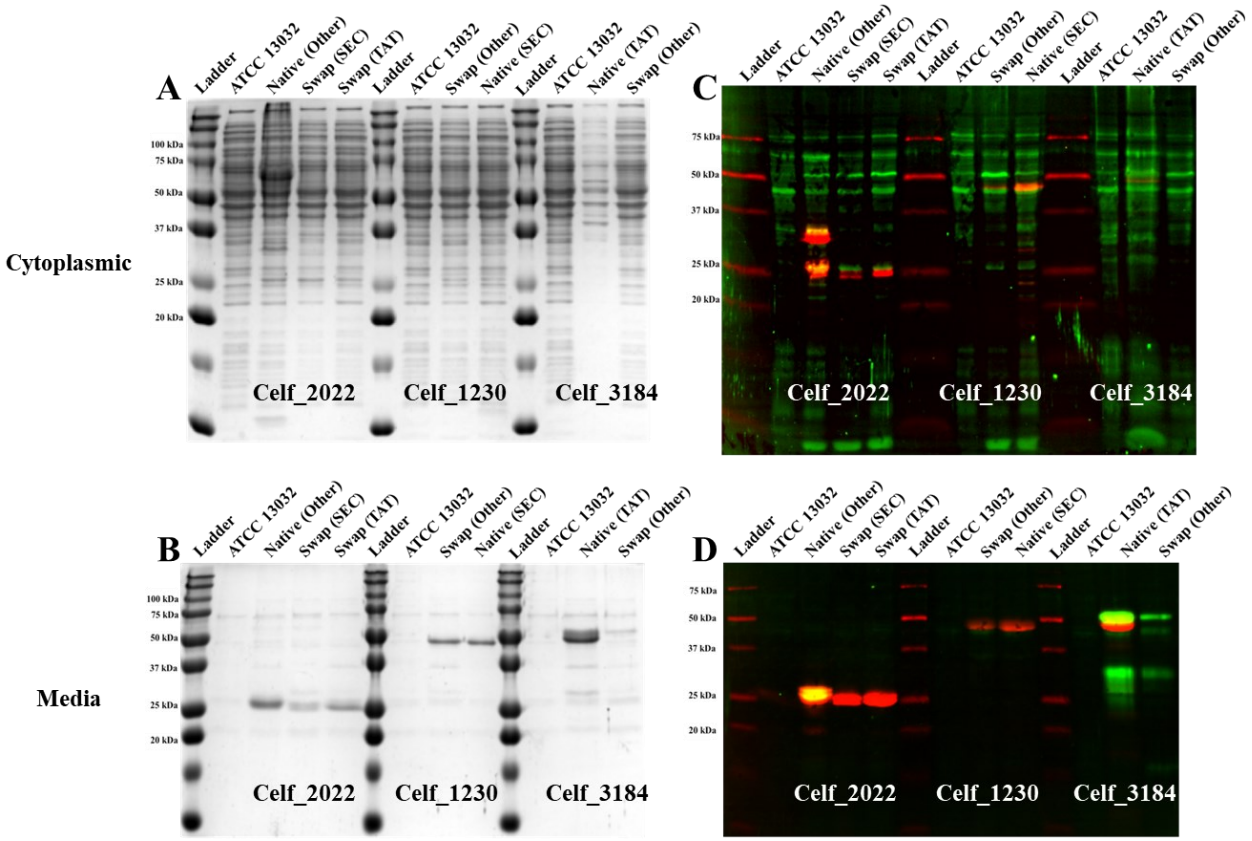


Figure S3. 5. Coomassie stained 15% SDS-PAGE (A and C) and ConA-FITC (green) lectin blot (B and D) of *C. glutamicum ATCC 13032* cytoplasmic (A and B) and undiluted spent culture media fractions (C and D) producing leader swap constructs.

Celf_2022 was expressed with its native "Other" and swapped SEC, and TAT leaders, Celf_1230 was expressed with its native SEC and swapped "Other" leader, and Celf_3184 was expressed with its native TAT and swapped "Other" leader. Less overall protein is observed in the cytoplasm when Celf_2022 has its native "Other" leader swapped to either a SEC or TAT leader with no mannosylation evident on the secreted material, but secretion of the protein itself is not affected. Secreted Celf_1230 levels and its mannosylation status do not change when exchanging the native SEC leader to the "Other" leader, but cytoplasmic levels do decrease. Both secreted and cytoplasmic levels of Celf_3184 decrease when its native TAT leader is replaced by the "Other" leader, however the secreted material is much more homogenously mannosylated.

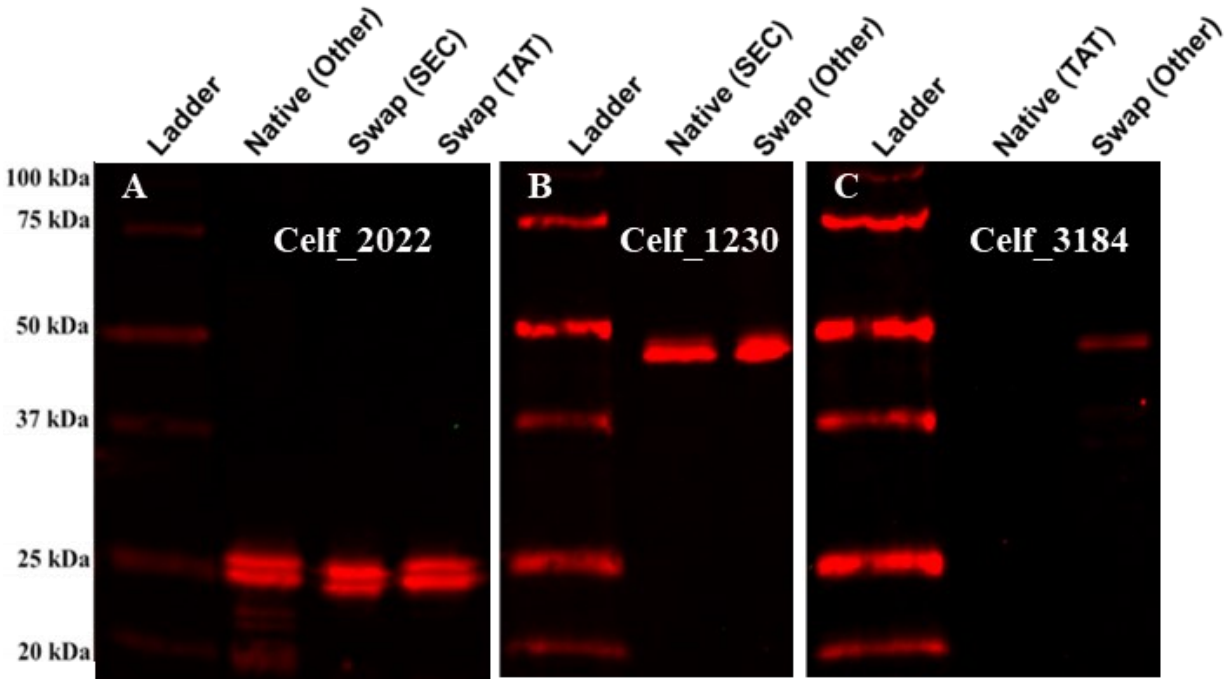


Figure S3. 6. ConA-FITC (green) and Anti-HIS-Alexa647 (red) blot of spent culture media enriched recombinant Celf_2022 expressed with "other", SEC, and TAT leaders (A), Celf_1230 expressed with SEC and "other" leaders (B), and Celf_3184 expressed with TAT and "other leaders (C) produced in *C. glutamicum* ΔCg_1014.

All proteins were produced in the POM deficient strain of *C. glutamicum*, indicated by the lack of ConA lectin reactivity (green). When lead by its native "other" leader sequence, Celf_2022 is secreted but not mannosylated in ΔCg_1014 (A). Secretion and mannosylation status are retained following the replacement of this "other" leader sequence by a SEC or TAT signal peptide. Replacement of the SEC signal peptide of Celf_1230 (a non-mannosylated protein) by the "other" leader sequence results in no change to secretion or mannosylation pattern (B). The native TAT led Celf_3184 is not produced or secreted in the ΔCg_1014 strain (C), but replacement of the TAT signal peptide by the "other" leader sequence results production and secretion of non-mannosylated Celf_3184. Molecular weight standards are the Bio-Rad All Blue ladder.

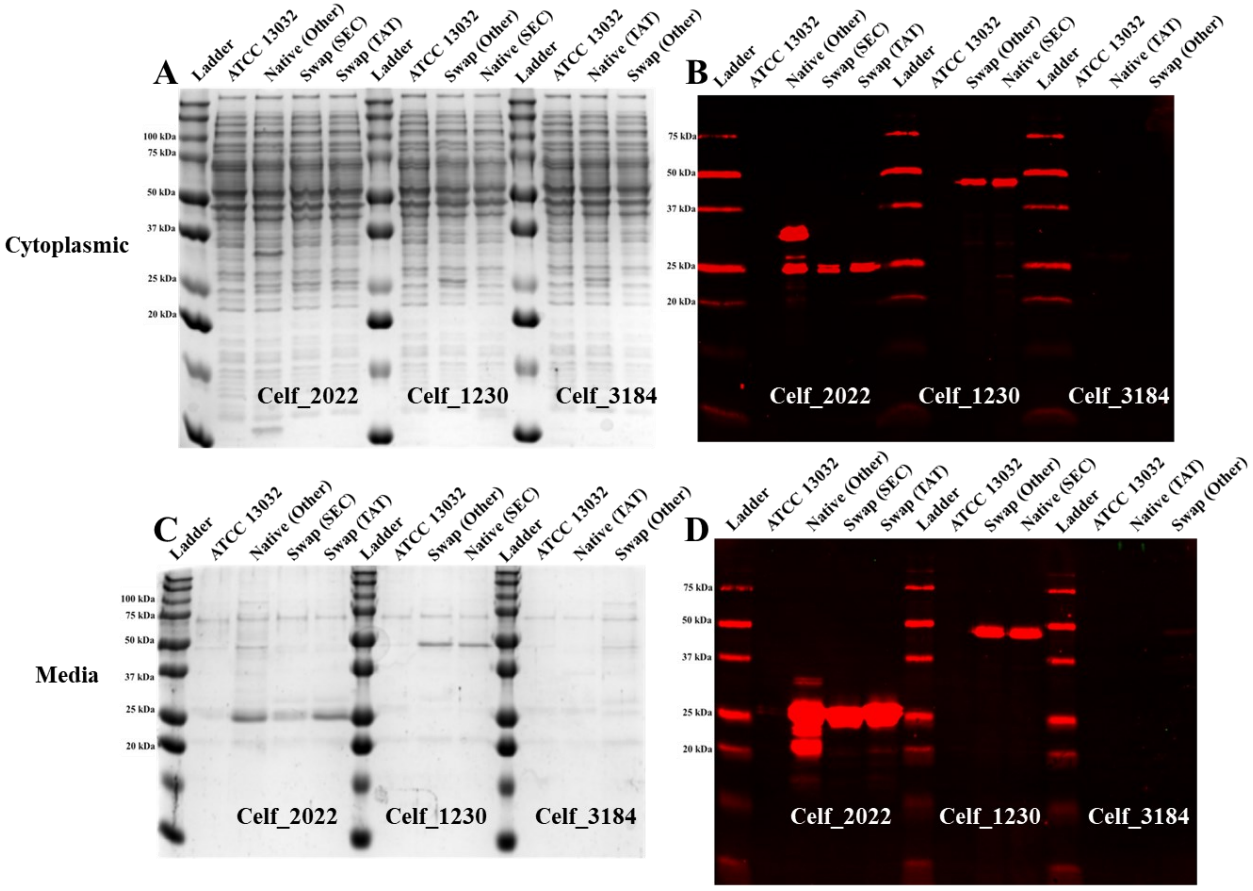
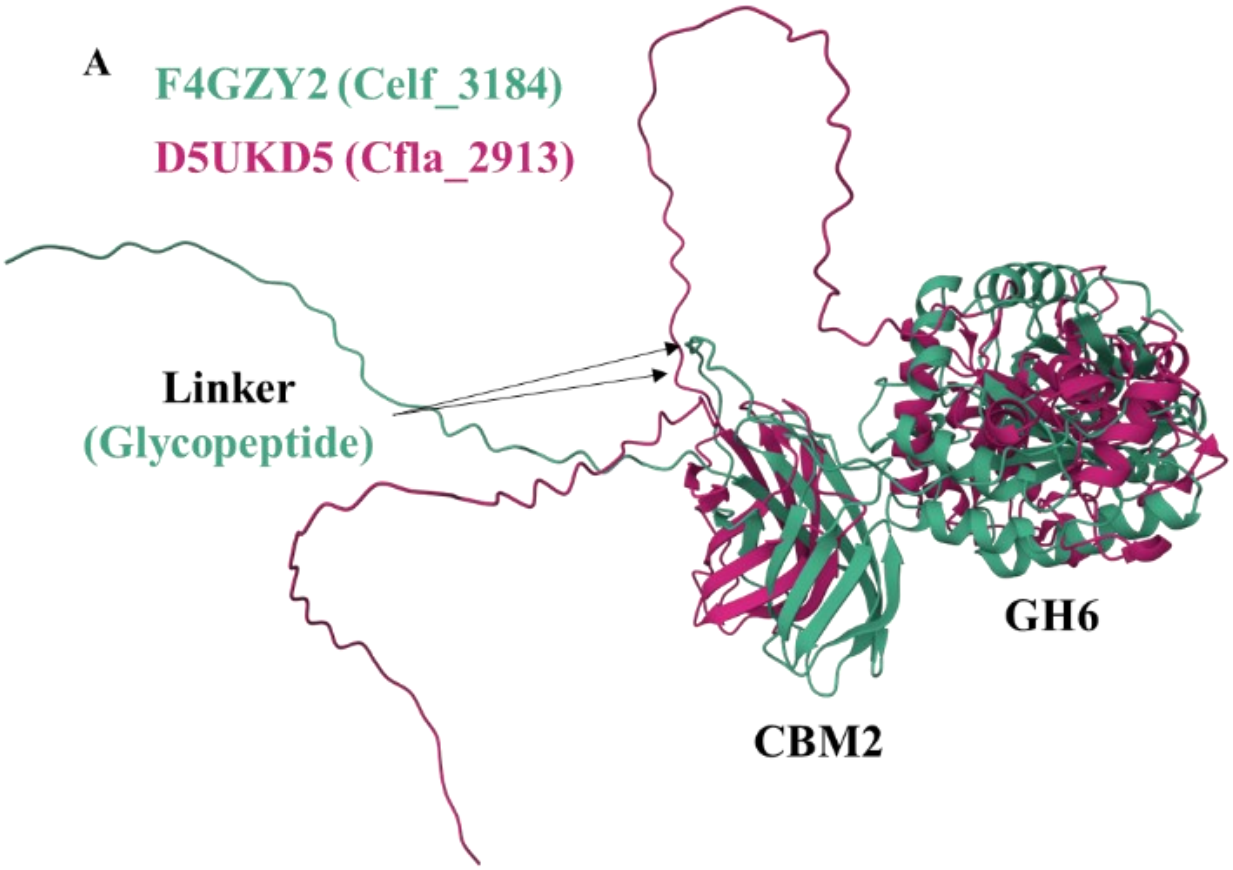


Figure S3. 7. Coomassie stained 15% SDS-PAGE (A and C) and ConA-FITC (green) lectin blot (B and D) of ΔCg_1014 *C. glutamicum* cytoplasmic (A and B) and undiluted spent culture media fractions (C and D) producing leader swap constructs.

Celf_2022 was expressed with its native "Other" and swapped SEC, and TAT leaders, Celf_1230 was expressed with its native SEC and swapped "Other" leader, and Celf_3184 was expressed with its native TAT and swapped "Other" leader. No mannosylation is evident on any of the constructs as they were expressed in the ΔCg_1014 mutant. Cytoplasmic and secreted protein abundance is very similar to these constructs expressed in the ATCC 13032 strain: however, Celf_3184 with its native TAT leader is not detectable (via Western blot and activity assay) in either the cytoplasmic or spent media fractions. Replacement of the TAT leader with the "Other" leader appears to restore the production and export of this protein.



B

```
Celf_3184       MSTR--------RTAAALLAAAAVAVGGLTALTTTAAQAAPGCRVDYAVTNQWPGGFGAN
Cfla_2913       MSPRPKPALAAGRKAVAVLAAGAVLAAGTTVLASTAANAAAGCRVDYAVTNEWPGGFGAS
                **.*        *.*.*:***.** ..* *.*.:.***:**.**********.*******.

Celf_3184       VTITNLGDPVSSWKLDWTYTAGQRIQQLWNGTASTNGGQVSVTSLPWNGSIPTGGTASFG
Cfla_2913       INVTNLGDPLSSWDLRWTFPSGQSIQQLWNGAASSSGSQVTVVNSPWNGSVGTNGVISLG
                :.:******:***.* **:.:** *******.**:.*.**:*.. *****: *.*. *:*

Celf_3184       FNGSWAGSNPTPASFSLNGTTCTGTV--------PTTSPT--PTPTPT-TPTPTPTPTPT
Cfla_2913       FNGSWNGSNAKPTSFTLNGTACTGSVSGGQPTQQPTQQPTQQPTQQPTQQPTQQPTQQPT
                ***** ***.:*:**:****:***:*         **.**  **  **  **  **  **

Celf_3184       PTPTVTPQPTSGFYVDPTTQGYRAWQAASGTDKALLEKIALTPQAYWVGNWADASHAQAE
Cfla_2913       QQPTQQPQPSGDFYVDPETAAYAAWQAASGNDKVLLAKIAQTPQALWIGDWSSASVIQQQ
                  **  ***.:.***** * .* *******.**.** *** **** *:*:*:.**  * :

Celf_3184       VADYTGRAVAAGKTPMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNPIVILE
Cfla_2913       VRDYTGKARSAGKIAQIVVYAIPGRDCGNYSGGGVATSEYARWVDTVAQGVQGNPWVILE
                * ****.* :*** . :********** .:*****: ************.:*** ****

Celf_3184       PDALAQLGDCSGQGDRVGFLKYAAKSLTLKGARVYIDAGHAKWLSVDTAVNRLNQVGFEY
Cfla_2913       PDALAQLGDCQGQGDRVGFLQYAAKAFAAKGARVYIDAGNSAWLSASEAANRLNRVGWDG
                **********.*********:****::: **********:: ***.. *.****.**::

Celf_3184       AVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSRNGNGSNG-EWCNPRGRALGER
Cfla_2913       AVGFSLNVSNYRTTAEAKAYGQEISRLTGGKKFVIDTSRNGNGASGSEWCNPSGRALGDR
                ****.**.***.***.:*****:**.  ***************.:* ***** *****:*

Celf_3184       PVAVNDGSGLDALLWVKLPGESDGACNGGPAAGQWWQEIALEMARNARW
Cfla_2913       PTRVNDGSGLDALLWIKRPGESDGTCNGGPAAGAWWQSMALELARNAKW
                *. ************:* ******:******** ***.:***:****.*
```

Figure S3. 8. Alphafold predicted structures (A) comparing Celf_3184 (F4GZY2, green) and Clfa_2913 (D5UKD5, magenta) and MUSCLE 3.8 sequence alignment (B).

Cfla_2913 – the *C. flavigena* homologue of the *C. fimi* GH6, Celf_3184 – contains a number of insertions in the linker region spanning the CBM2a and GH6 domains. In Celf_3184, this linker has been identified as the glycopeptide containing *O*-mannosyl glycans. The longer linker domain of Cfla_2913 promotes a significantly different orientation to this domain, potentially moving the glycosylation sites away from the active site of the *C. glutamicum* GT-39 (A). The *C. flavigena* and *C. glutamicum* GT-39s only share 40.5% global amino acid sequence homology. The twin arginine (RR) motif of Cfla_2913 contains an 8 amino acid insertion (B) – compared to the same motif in Celf_3184 – causing predictive tools like SignalP5.0 to misrepresent their identity. The full length Cfla_2913 protein also contains the hydrophobic and C-terminal regions indicative of TAT proteins. The C-terminal region often ends with a short (A-X-A) motif specifying cleavage by the signal peptidase. Predicted signal peptide cleavage of both proteins indicated by black arrow and twin arginine motif indicated by magenta arrows.

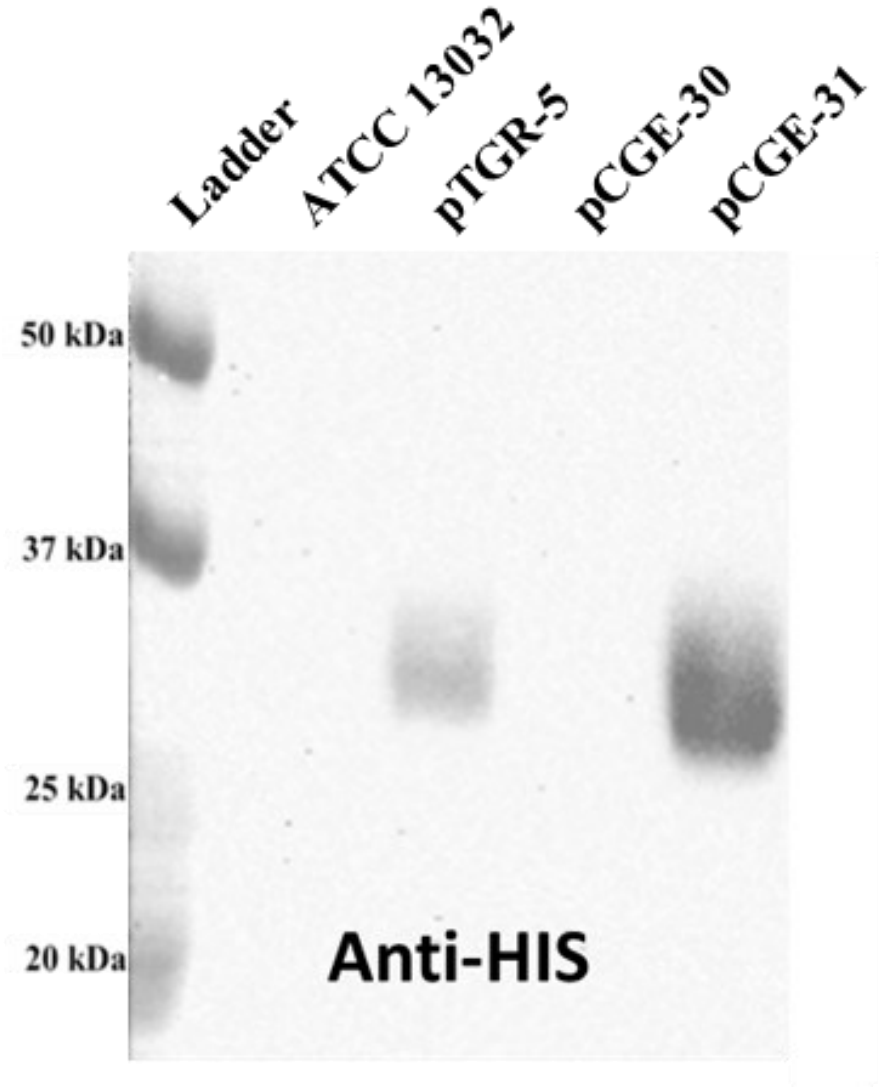
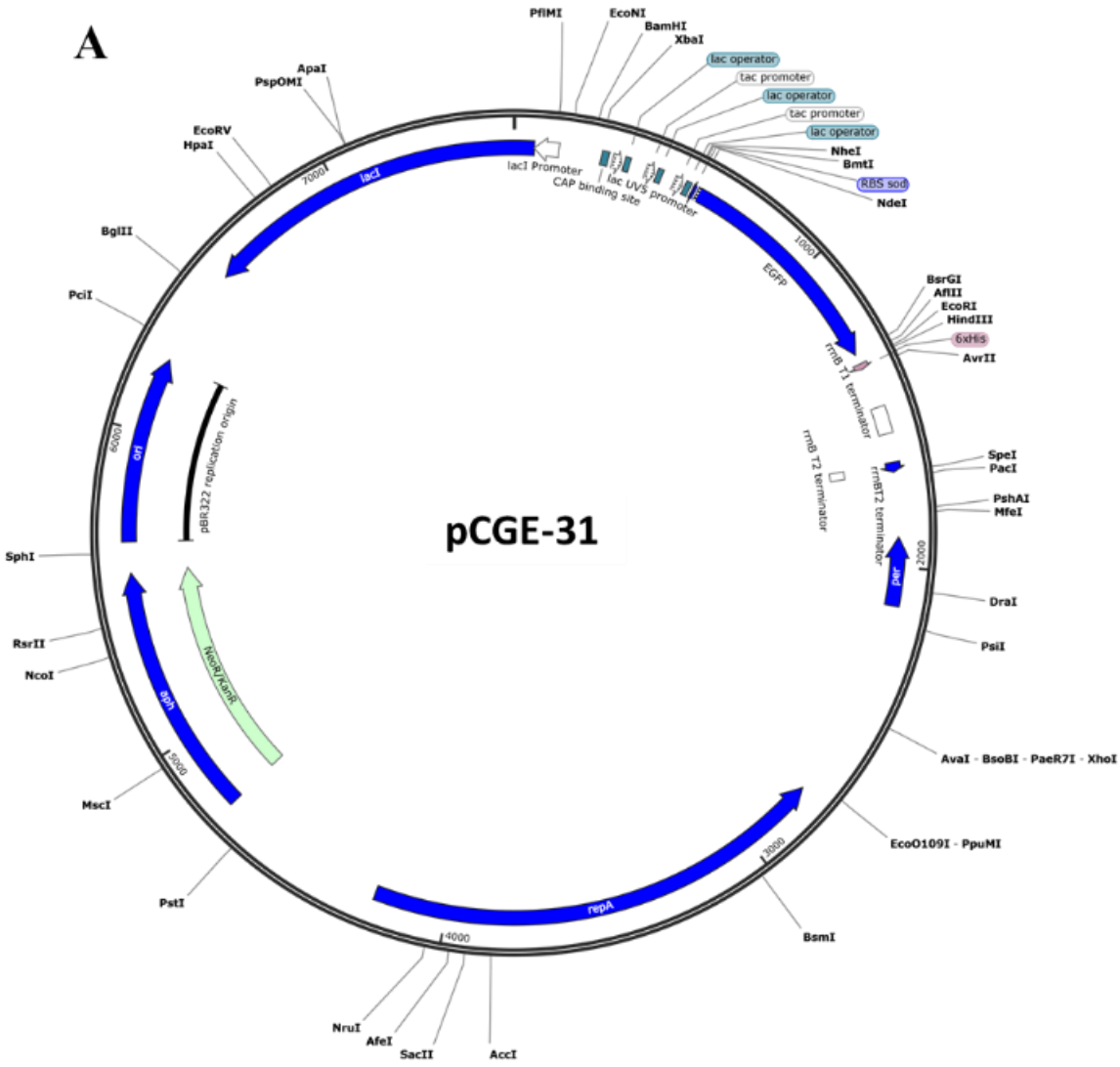


Figure S3. 9. Anti-HIS-HRP Western blot showing expression level differences of eGFP produced in *C. glutamicum* by pTGR-5 and pCGE-31.

The *E. coli*/*C. glutamicum* shuttle vector pCGE-31 can drive higher expression levels of recombinant proteins in *C. glutamicum* than its parent pTGR-5. The inclusion of the sod RBS is critical to its functionality in *C. glutamicum*, as pCGE-30 uses the triple promoter system and RBS from pCW and is incapable of producing as much eGFP as either sod RBS containing vector.


A
pCGE-31
PflMI
EcoNI
BamHI
XbaI
lac operator
tac promoter
lac operator
tac promoter
lac operator
NheI
BmtI
RBS sod
NdeI
ApaI
PspOMI
EcoRV
HpaI
BglII
PciI
SphI
RsrII
NcoI
MscI
PstI
NruI
AfeI
SacII
AccI
BsmI
EcoO109I - PpuMI
AvaI - BsoBI - PaeR7I - XhoI
PsiI
DraI
PshAI
MfeI
SpeI
PacI
BsrGI
AflII
EcoRI
HindIII
6xHis
AvrII
lacI
lacI Promoter
CAP binding site
lac UV5 promoter
EGFP
rrnB T1 terminator
rrnB T2 terminator
rrnBT2 terminator
pBR322 replication origin
ori
aph
HoxB/KanR
repA
pair
1000
2000
3000
4000
5000
6000
7000

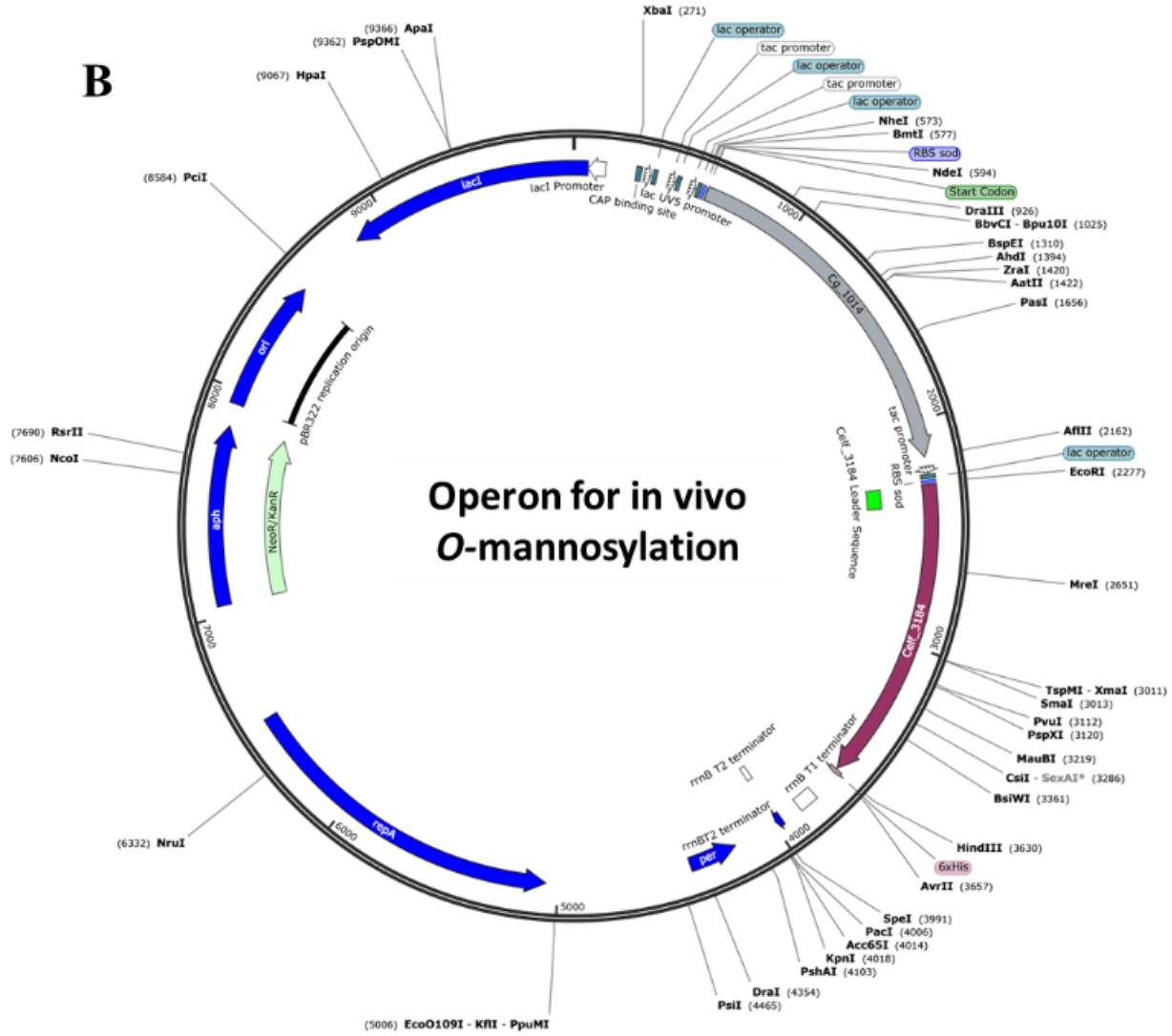


Figure S3. 10. Plasmid maps of pCGE-31 used for recombinant expression of GT-39s (A) and *O*-mannosylation operons used for co-expression of GT-39s with target actinobacterial mannoproteins (B).

Features of pCGE-31 *E. coli*/*C. glutamicum* shuttle vector (A). The pCGE-31 vector originates from pTGR-5 and received the triple promoter system from pCW via restriction cloning. Actinobacterial GT-39 genes amplified from genomic DNA replaced eGFP using NdeI and HindIII restriction sites. Features of the *O*-mannosylation operon constructs (B). A synthetic

secondary ORF containing the Celf_3184 gene was added via restriction cloning (NdeI – AvrII) under the control of Ptac and also utilizing the sod RBS.

Table S3. 1. Global sequence similarity of actinobacterial GT-39s compared to *S. cerevisiae* PMT1. Global sequence similarity was determined using EMBOSS Needle pairwise sequence alignment.

	***S. cerevisiae* (PMT1)**	***M. tuberculosis***	***M. smegmatis***	***C. glutamicum***	***C. fimi***	***C. flavigena***
***S. cerevisiae* (PMT1)**	**100.0%**	25.7%	25.1%	26.0%	25.4%	23.2%
M. tuberculosis	25.7%	**100.0%**	83.9%	55.9%	46%	46.9%
M. smegmatis	25.1%	83.9%	**100.0%**	57.7%	46%	48.7%
C. glutamicum	26.0%	55.9%	57.7%	**100.0%**	42%	40.5%
C. fimi	25.4%	46.0%	46%	42.0%	**100.0%**	50.5%
C. flavigena	23.2%	46.9%	48.7%	40.5%	50.5%	**100.0%**

Table S3. 2. Selected primers used in this study.

Target Amplicon	*5'*	*3'*
Cg_1014	5'-TTG ATT CAT ATG GTG AGC CAA GCC CTA CCT GTT CG-3'	5'-TTG ATT CTT AAG TTA GCG CCA GCT TGG GAA CCA CAT CAA GG-3'
Celf_3080	5'-TTG ATT CAT ATG GTG CCG CCC ACG CGA GAC GAC-3'	5'-TTG ATT CTT AAG TCA GAT CCA GCT CGT CAG CCA CAT GTG GCT GTG C-3'
Cfla_0843	5'-TTG ATT CAT ATG GTG CCG ACC GAC GGA GAC GAC ACC GAG-3'	5'-TTG ATT CTT AAG TCA GAT CCA GGT CGG CAG CCA CAT CCG GAT GTG C-3'
pCW triple promoter	5'-ATT AGT CTA GAT AAT GTG AGT TAG CTC ACT CAT TAG G-3'	5'-ATT AGG CTA GCA AAT TGT TAT CCG CTC ACA ATT CCA-3'
Cg_1014 upstream	5'-GGG GAT CCT TCT TCG GTT GCG GTA ATT TGC TCT GGC T-3'	5'-AGC CAT CTC TCA CTC GGT TGA TTG TAG AGC CTT GGC-3'
Cg_1014 downstream	5'-TAG ATC GCC CTC CCC TTT TAC CGC ACC AGG TGA CC-3'	5'-CGA CTC TAG AGT CGA TGT CAT GAA CCA CTG GCT CGA C-3'
Cg_1014 deletion	5'-GGG GAT CCT TCT TCG GTT GCG GTA ATT TGC TCT GGC T-3'	5'-CGA CTC TAG AGT CGA TGT CAT GAA CCA CTG GCT CGA C-3'
C. glutamicum MBP	5'-TTG AAG CGT CTT ACT CGC ATC GCA TCC ATC-3'	5'-TTA GCC CCA GTT GGA TTC CTT CTC AGC AG-3'
celf_3184	5'-GGG GTA TTC CAT ATG TCC ACC CGC AGA ACC GCC GCA GCG-3'	5'-GGG GAA TTC TCA CCA CCT GGC GTT GCG CGC CAT C-3'
pCGE-31 MCS	5'-GCA TGA TAT GGA TCC ATA TAT GCG GCC GCA TAT TCT AGA-3'	5'-GCG CTA CGG CGT TTC ACT TCT GAG-3'
SEC – Celf_2022 N-term	5'-ATT AGC ATA TGG TGG CCC GAC CCT TCC G-3'	5'-ACG AGT CAG CGC GGC CGC GG-3'
SEC – Celf_2022 C-term	5'-CGG CCG CGC TGA CTC GTG ACG ACA CTC CTG TGG-3'	5'-ATT AGG AAT TCT TAT CAA TGG TGA TGG TGA TGG TGC TGG GT-3'
TAT – Celf_2022 N-term	5'-CAT ATG TCC ACC CGC AGA ACC GCC-3'	5'-CAC GAG TCA GCG CCT GCG CGG CGG TG-3'
TAT – Celf_2022 C-term	5'-GCG CAG GCG CTG ACT CGT GAC GAC ACT CCT-3'	5'-AAG CTT AGA ATT CTT ATC AAT GGT GAT GGT GAT GGT GC-3'
"Other" – Celf_1230 N-term	5'-ATT AGC ATA TGT CCA GCA AGC GCG AAC GTG AG-3'	5'-GCC GCC GTG TGC CAC AGC TGC ACC GAT ACC C-3'

"Other" – Celf_1230 C-term	5'-GCT GTG GCA CAC GGC GGC CCC CC-3'	5'-ATT AGA AGC TTG CTG CGC GGG C-3'
"Other" – Celf_3184 N-term	5'-ATT AGC ATA TGT CCA GCA AGC GCG AAC GTG AG-3'	5'-AGC CGG GAG CCG CCA CAG CTG CAC CGA TAC CC-3'
"Other" – Celf_3184 C-term	5'-GCT GTG GCG GCT CCC GGC TGC-3'	5'-ATT AGA AGC TTC CAC CTG GCG TTG CG-3'

Appendix C

Supplemental to Chapter 2 (S2)

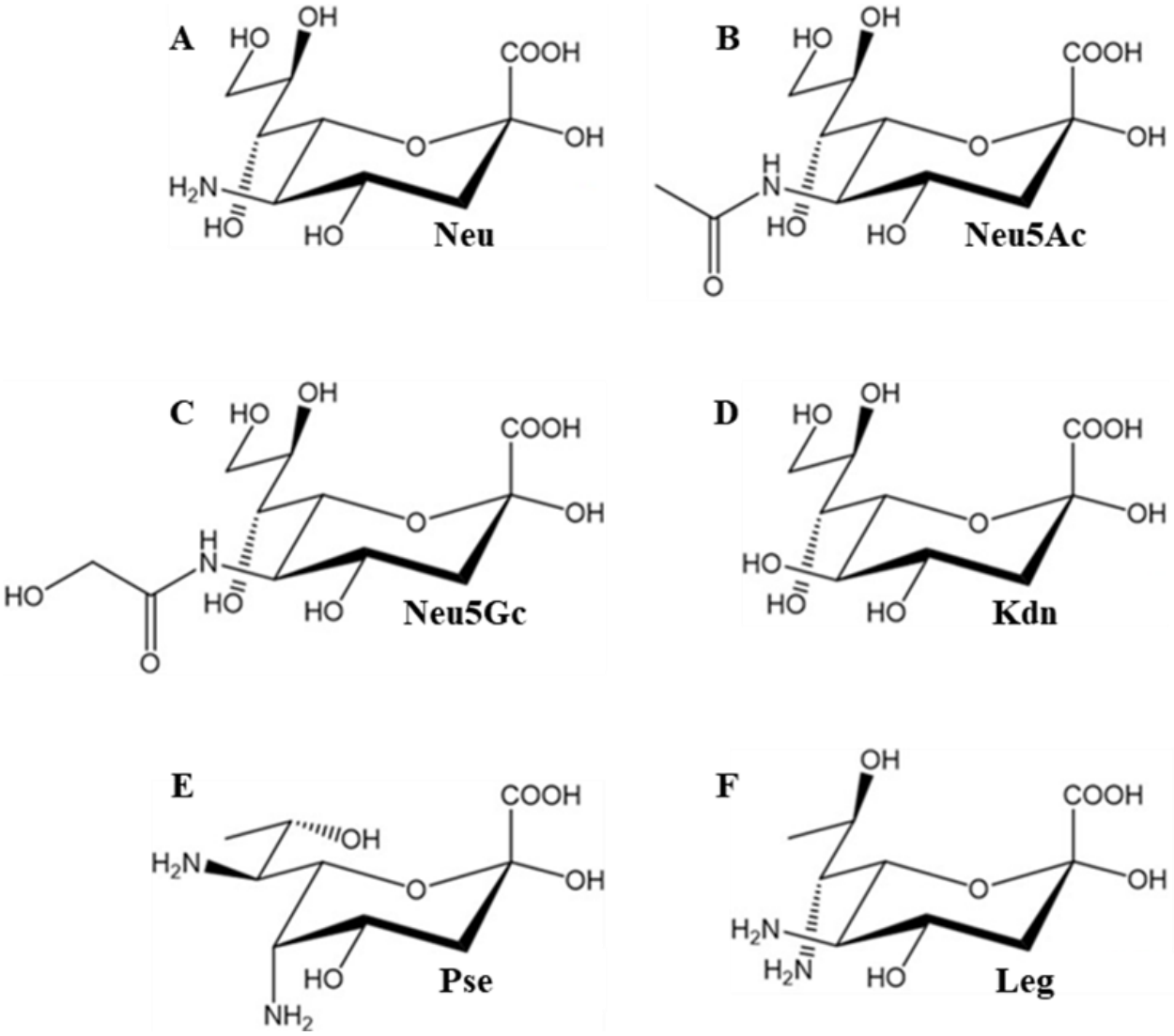
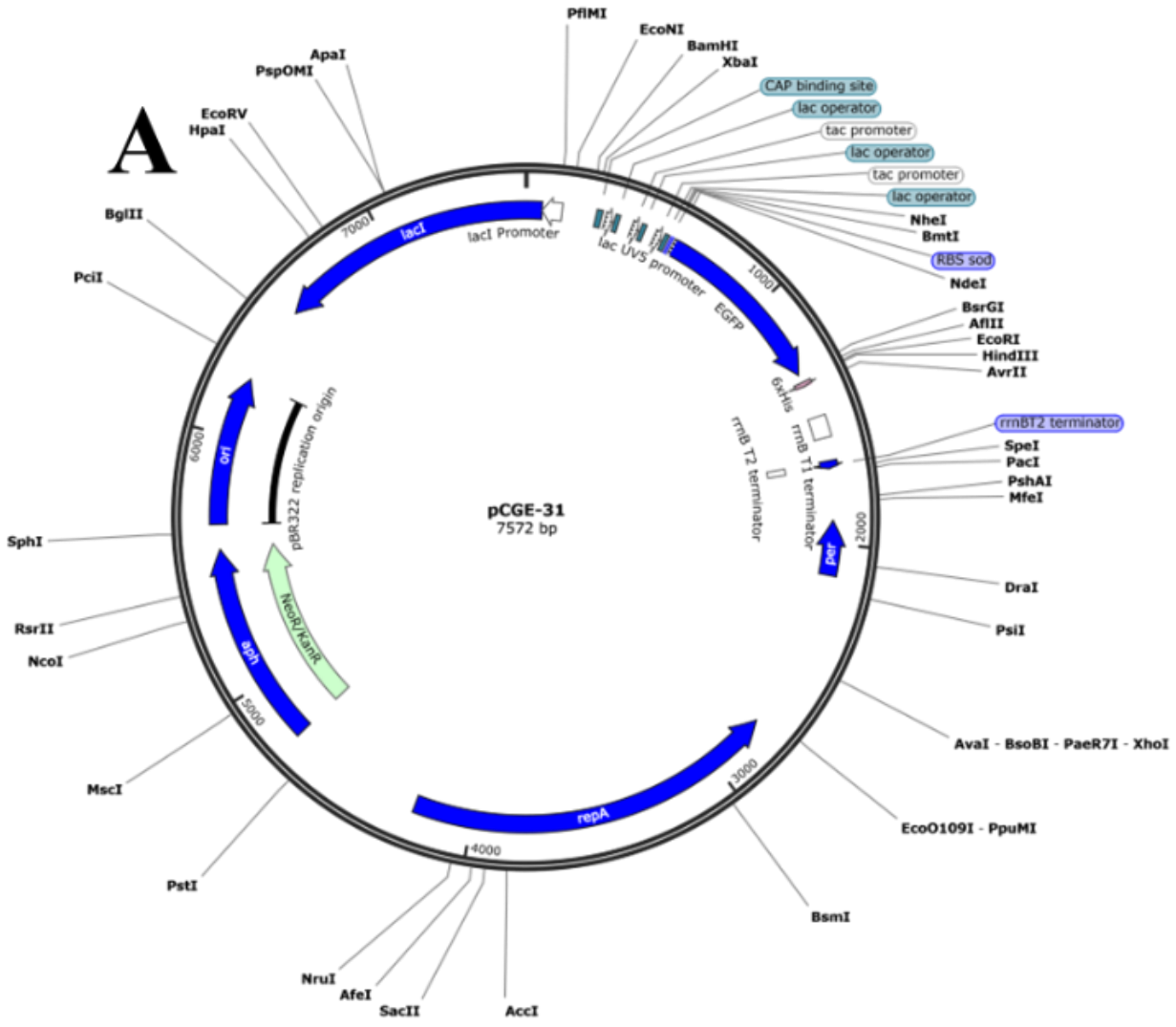


Figure S4. 1. Structural comparison of selected nonulosonic acids.

The parent compound in this lineage is (A) 5-amino-3,5-dideoxy-2-nonulosonic acid (neuraminic acid). Acetamido substituents at C5 results in (B) 5-acetamido-3,5-dideoxy-2-nonulosonic acid (*N*-acetylneuraminic acid, Neu5Ac). Hydroxyacetamido substitution at C5 results in (C) 3,5-dideoxy-5-hydroxyacetamido-2-nonulosonic acid (*N*-glycolylneuraminic acid, Neu5Gc). The C5 amine replacement by a hydroxyl group results in (D) 3-deoxy-2-nonulosonic acid

(ketodeoxynononic acid, Kdn). Also shown are: (E) 5,7-diamino-3,5,7,9-tetradeoxy-2-nonulosonic acid (pseudaminic acid, Pse), and (F) 5,7-diamino-3,5,7,9-tetradeoxy-2-nonulosonic acid (legionaminic acid, Leg).


A
pCGE-31
7572 bp
PflMI
EcoNI
BamHI
XbaI
CAP binding site
lac operator
tac promoter
lac operator
tac promoter
lac operator
NheI
BmtI
RBS sod
NdeI
BsrGI
AflII
EcoRI
HindIII
AvrII
rrnBT2 terminator
SpeI
PacI
PshAI
MfeI
DraI
PsiI
AvaI - BsoBI - PaeR7I - XhoI
EcoO109I - PpuMI
BsmI
AccI
SacII
AfeI
NruI
PstI
MscI
SphI
RsrII
NcoI
BglII
PciI
ApaI
PspOMI
EcoRV
HpaI
lacI
lacI Promoter
lac UV5 promoter
EGFP
6xHis
rrnB T1 terminator
rrnB T2 terminator
per
repA
aph
NeoR/KanR
ori
pBR322 replication origin
1000
2000
3000
4000
5000
6000
7000

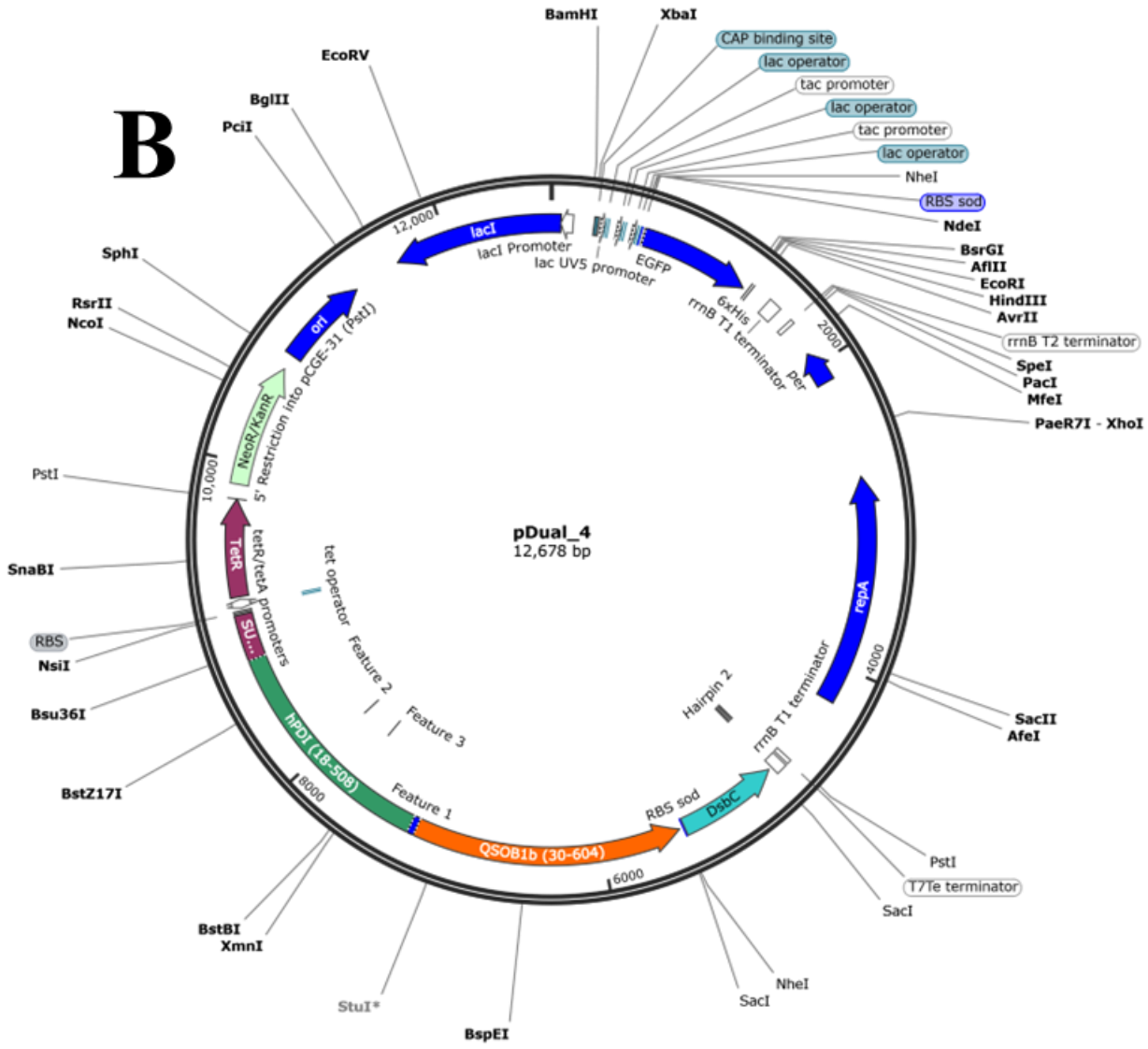
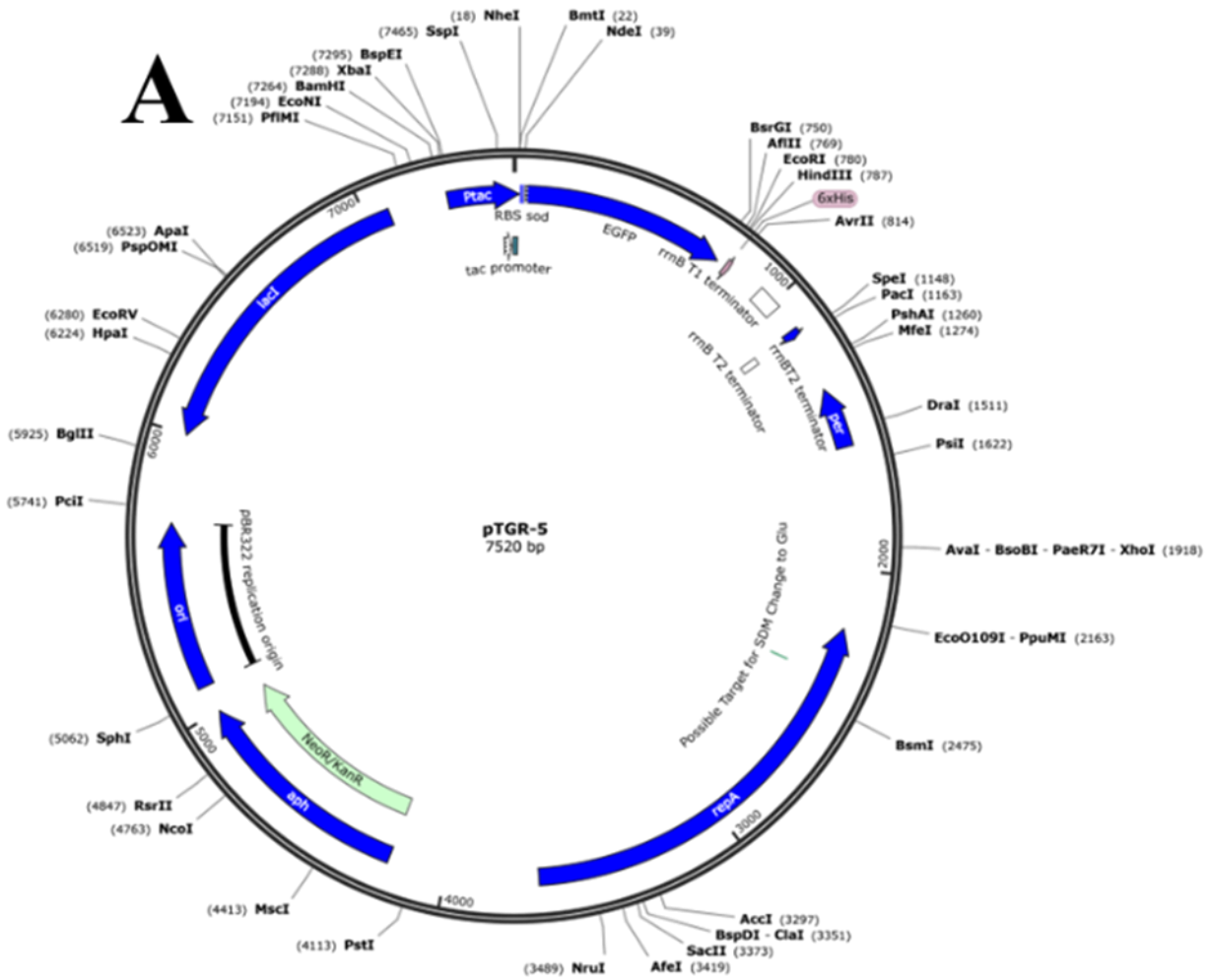
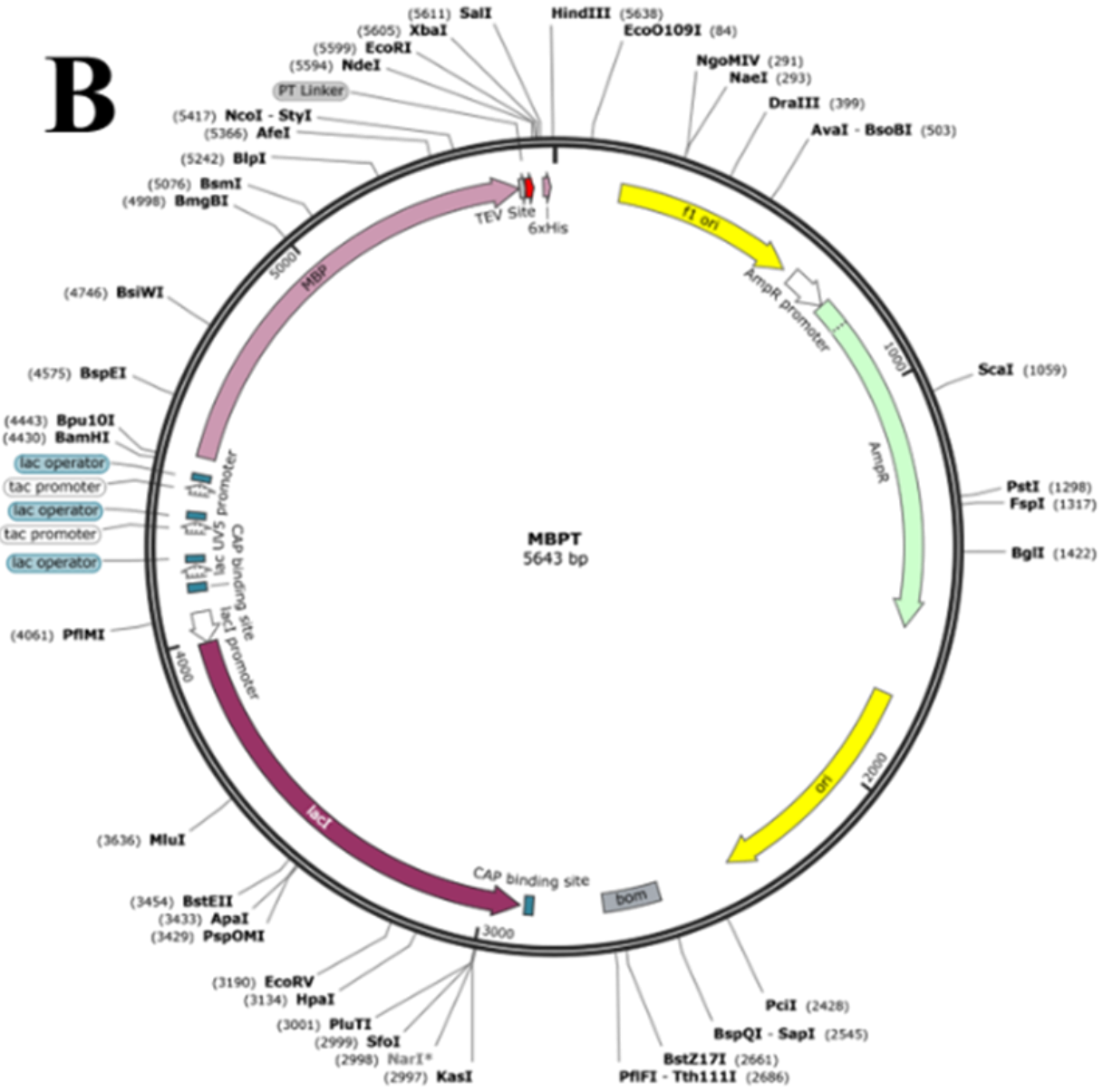


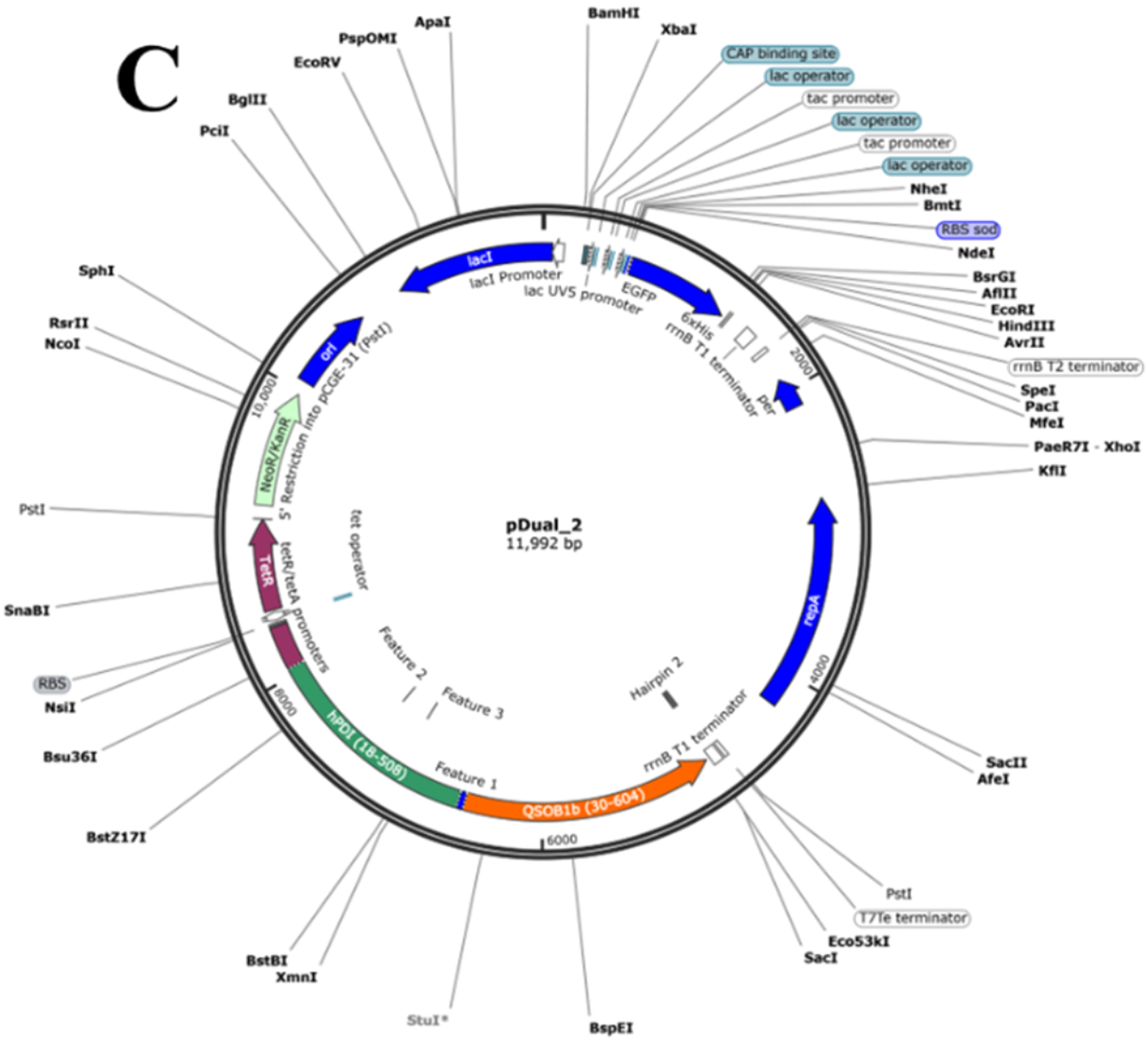
Figure S4. 2. Schematics of pCGE-31 (A) and pDual_4 (B) used for recombinant expression of STs in *C. glutamicum*.

pCGE-31 (A) and its derivative pDual_4 (B) contain the IPTG inducible P_{lacUV5} and tandem P_{tac} promoter system to drive high-yield expression of heterologous proteins in both *E. coli* and *C. glutamicum*. pDual_4 (B) also contains the ATc inducible $P_{tetR/tetA}$ promoter for the co-expression of folding chaperones in both organisms – both the hPDI-QSOX1b fusion and DsbC.


A
pTGR-5
7520 bp
(18) NheI
BmtI (22)
NdeI (39)
(7465) SspI
(7295) BspEI
(7288) XbaI
(7264) BamHI
(7194) EcoNI
(7151) PflMI
BsrGI (750)
AflII (769)
EcoRI (780)
HindIII (787)
6xHis
AvrII (814)
SpeI (1148)
PacI (1163)
PshAI (1260)
MfeI (1274)
DraI (1511)
PsiI (1622)
AvaI - BsoBI - PaeR7I - XhoI (1918)
EcoO109I - PpuMI (2163)
BsmI (2475)
AccI (3297)
BspDI - ClaI (3351)
SacII (3373)
AfeI (3419)
(3489) NruI
(4113) PstI
(4413) MscI
(4763) NcoI
(4847) RsrII
(5062) SphI
(5741) PciI
(5925) BglII
(6224) HpaI
(6280) EcoRV
(6519) PspOMI
(6523) ApaI
Ptac
RBS sod
tac promoter
EGFP
rrnB T1 terminator
rrnB T2 terminator
rrnBT2 terminator
per
repA
aph
KeoR/KanR
ori
lacI
p8A322 replication origin
Possible Target for SDM Change to Glu
1000
2000
3000
4000
5000
6000
7000


B
MBPT
5643 bp
(5611) SalI
HindIII (5638)
(5605) XbaI
EcoO109I (84)
(5599) EcoRI
(5594) NdeI
NgoMIV (291)
NaeI (293)
PT Linker
DraIII (399)
(5417) NcoI - StyI
AvaI - BsoBI (503)
(5366) AfeI
(5242) BlpI
(5076) BsmI
(4998) BmgBI
TEV Site
6xHis
f1 ori
MBP
5000
AmpR promoter
1000
(4746) BsiWI
ScaI (1059)
(4575) BspEI
(4443) Bpu10I
(4430) BamHI
lac operator
tac promoter
lac operator
tac promoter
lac operator
lac UV5 promoter
CAP binding site
lacI promoter
AmpR
PstI (1298)
FspI (1317)
BglII (1422)
(4061) PflMI
4000
ori
2000
lacI
(3636) MluI
CAP binding site
bom
(3454) BstEII
(3433) ApaI
(3429) PspOMI
3000
(3190) EcoRV
(3134) HpaI
PciI (2428)
(3001) PluTI
(2999) SfoI
BspQI - SapI (2545)
(2998) NarI*
BstZ17I (2661)
(2997) KasI
PflFI - Tth111I (2686)

C


pDual_2
11,992 bp
lacI
lacI Promoter
lac UV5 promoter
EGFP
6xHis
rrnB T1 terminator
ori
pBR322 ori
2000
4000
repA
rrnB T1 terminator
Hairpin 2
6000
QSOB1b (30-604)
Feature 1
hPDI (18-508)
Feature 3
Feature 2
8000
tetR/tetA promoters
tet operator
TetR
5' Restriction into pCGE-31 (PstI)
NeoR/KanR
10,000
CAP binding site
lac operator
tac promoter
lac operator
tac promoter
lac operator
NheI
BmtI
RBS sod
NdeI
BsrGI
AflII
EcoRI
HindIII
AvrII
rrnB T2 terminator
SpeI
PacI
MfeI
PaeR7I - XhoI
KflI
SacII
AfeI
PstI
T7Te terminator
Eco53kI
SacI
BspEI
StuI*
BstBI
XmnI
BstZ17I
Bsu36I
NsiI
RBS
SnaBI
PstI
SphI
RsrII
NcoI
PciI
BglII
EcoRV
PspOMI
ApaI
BamHI
XbaI

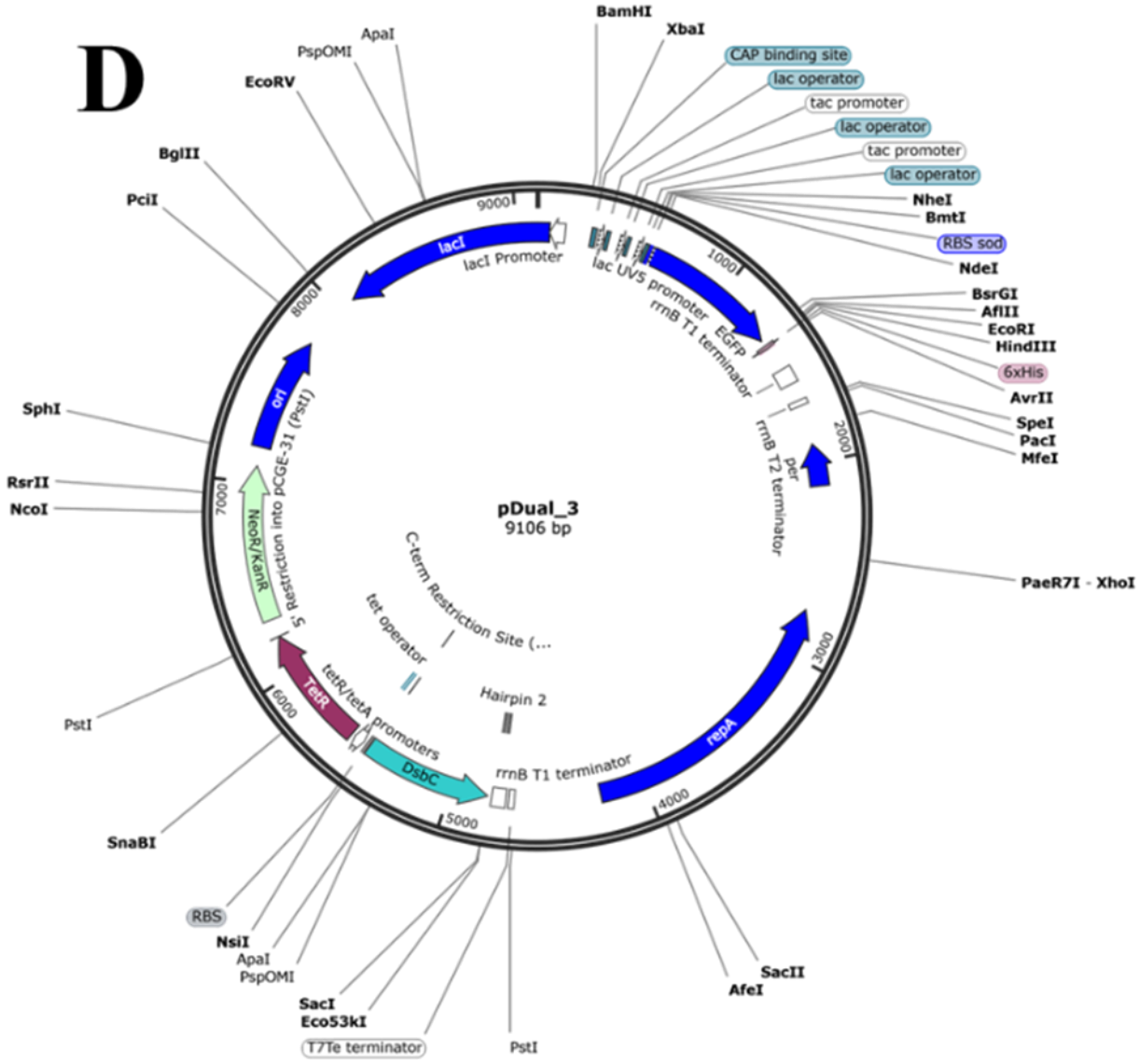


Figure S4. 3. Schematics of pTGR-5 (A), pCW-MBPT (B), pDual_2 (C), and pDual_3 (D).

The *E.* coli/*C. glutamicum* shuttle vector pTGR-5 (A) was utilized as the basis for the pDual series plasmids (C and D) allowing for the co-expression of the folding chaperones hPDI-QSOX1b and DsbC, respectively. The triple promoter system (P_{lacUV5} and tandem P_{tac}) from pCW-MBPT (B) was used to replace the P_{tac} region of pTGR-5, generating pCGE-31.

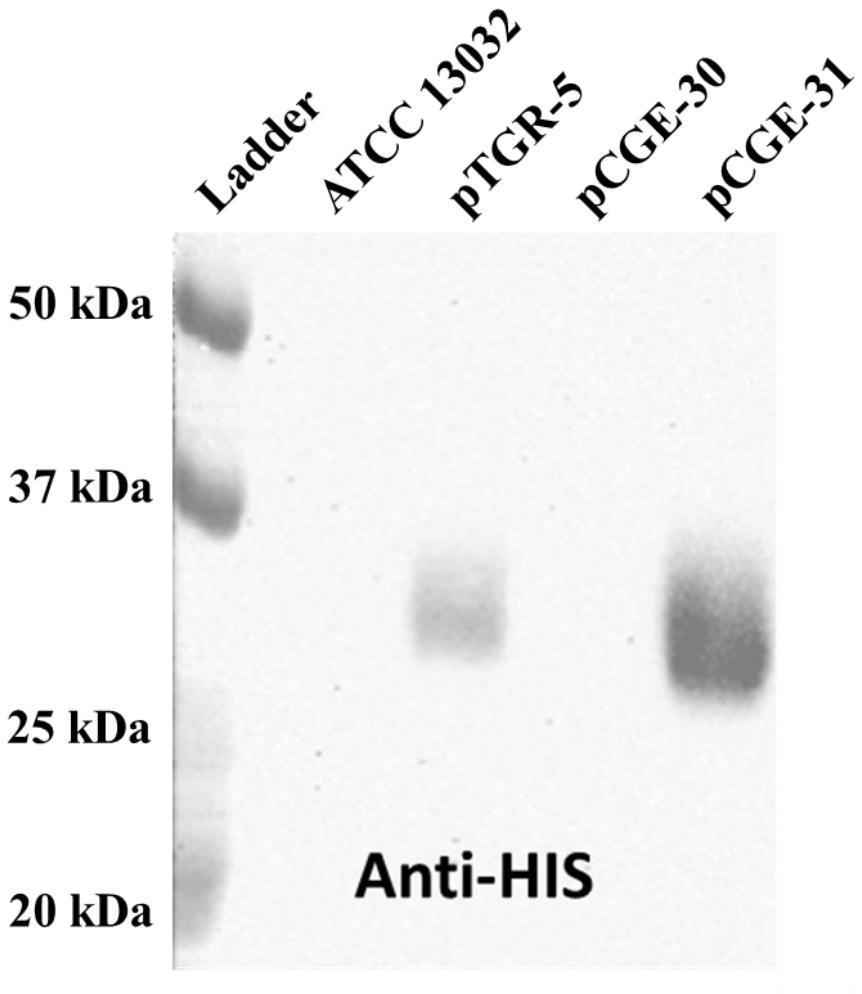


Figure S4. 4. Anti-HIS$_6$-HRP Western blot showing expression level differences of eGFP produced in *C. glutamicum* by pTGR-5 and pCGE-31.

The *E. coli*/*C. glutamicum* shuttle vector pCGE-31 can drive higher expression levels of recombinant proteins in *C. glutamicum* than its parent pTGR-5. The inclusion of the sod RBS is critical to its functionality in *C. glutamicum*. pCGE-30 uses the triple promoter system and T4 phage lysozyme RBS from pCW and is incapable of driving expression of eGFP in *C. glutamicum*.

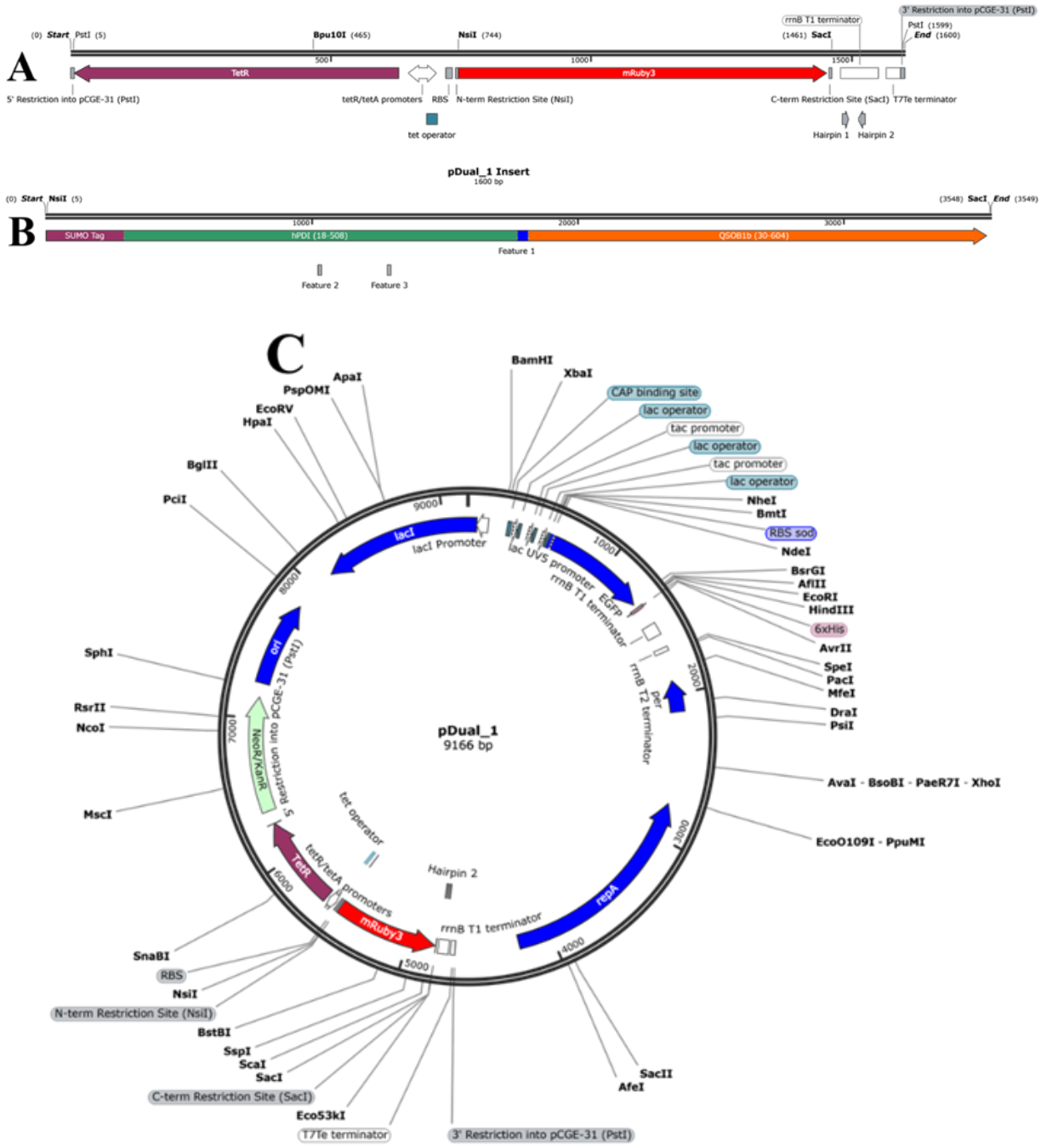


Figure S4. 5. Schematic of PtetR/tetA synthetic gene used to generate pDual vectors (A), synthetic gene of hPDI-QSOX1b fusion (B), and intermediate pDual_1 construct with mRuby3 to assess functionality of PtetR/tetA (C).

Synthetic gene of the ATc inducible $P_{tetR/tetA}$ promoter (A) cloned into pCGE-31 via PstI restriction sites to generate the intermediate construct pDual_1 (C). The synthetic gene contains

an mRuby3 stuffer sequence in-frame with P_{tetA} to allow for the facile verification of functionality and simple replacement by desired folding chaperones by restriction cloning. The synthetic gene of the hDPI-QSOX1b fusion with an *N*-terminal SUMO tag (B) was cloned into pDual_1 utilizing NsiI and SacI restriction sites.

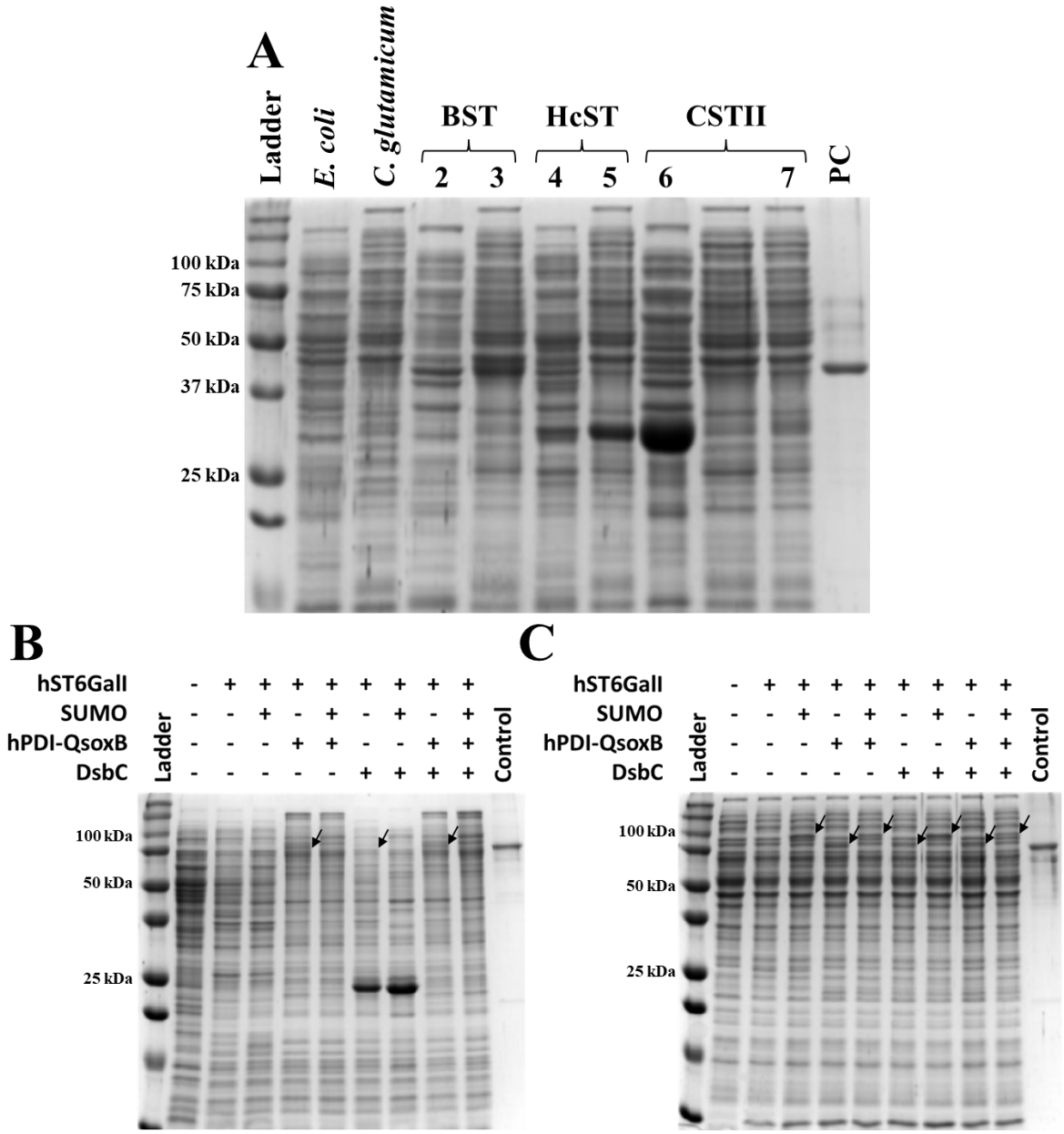


Figure S4. 6. Coomassie stained 15% SDS-PAGEs of lysates producing bacterial STs (BST, HcST, CSTII) in BL21 E. coli (A) and C. glutamicum (B), and lysates producing hST6GalI co-expression constructs in BL21 *E. coli* (C) and *C. glutamicum* (D).

No evident induction band of either MBP-hST6GalI or SUMO-MBP-hST6GalI is evident in *E. coli* lysates (C). Cytoplasmic DsbC overexpression is evident in pDual_3 constructs produced in *E. colii* (C, lanes 7 and 8). The sole expression of the MBP-hST6GalI in *C. glutamicum* (D, lane

3) produces no soluble ST enzyme, but the addition of an *N*-terminal SUMO tag, or the co-expression of folding chaperones (hPDI-QSOX1B and/or DsbC) produces a distinct induction band of corresponding molecular weight (black arrows). Control protein is MBP-hST6GalI produced in SHuffle® Express *E. coli*. Expected molecular weight of MBP-hST6GalI is 79.6 kDa, SUMO-MBP-hST6GalI is 90.6 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.

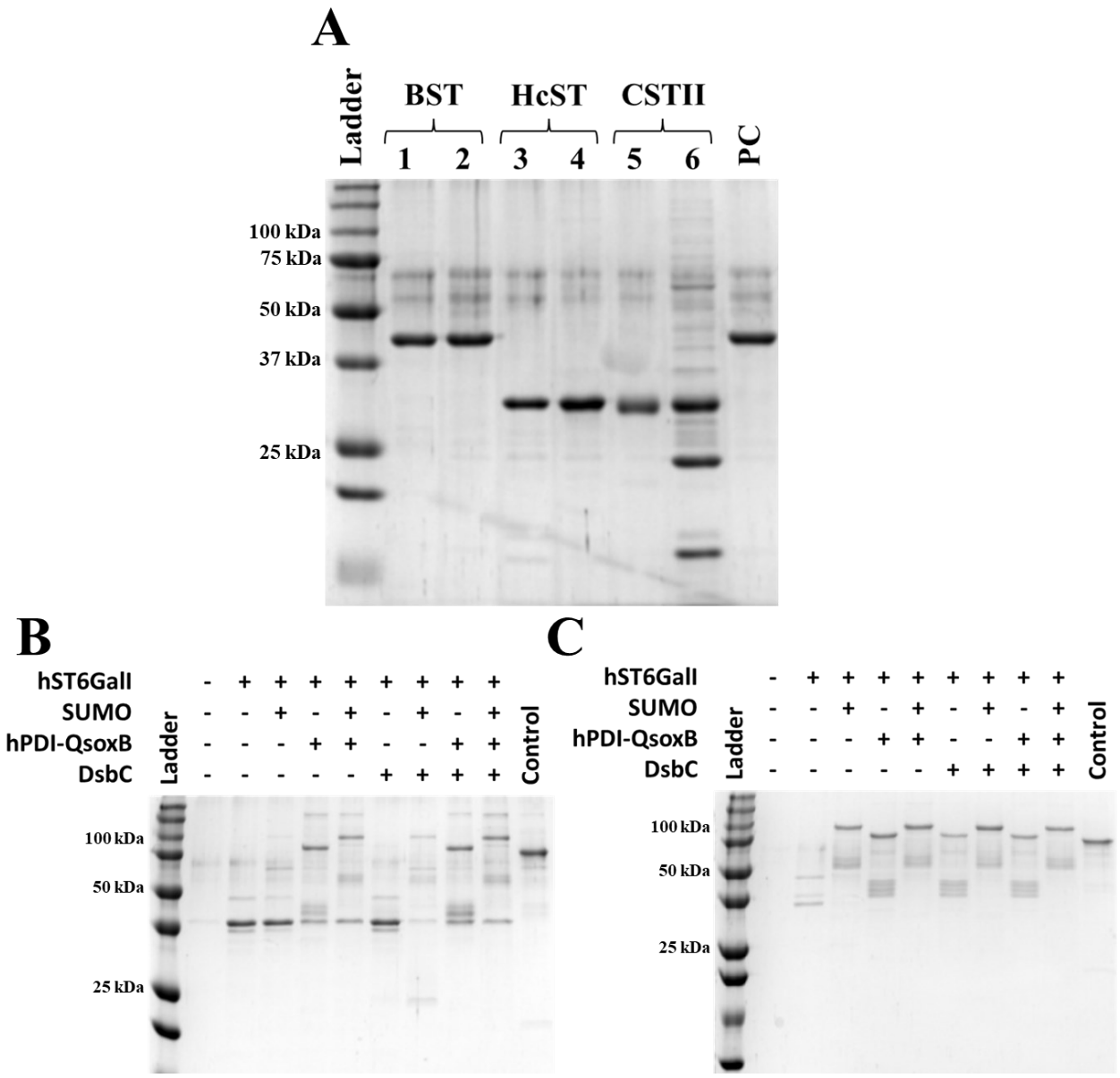


Figure S4. 7. Coomassie stained 15% SDS-PAGEs of IMAC purified bacterial STs (BST, HcST, CSTII)from BL21 E. coli (A) and C. glutamicum (B), and affinity purified hST6GalI co-expression constructs from BL21 *E. coli* (C) and *C. glutamicum* (D).

The sole expression of the MBP-hST6GalI in both *E. coli* (C, lane 3) and *C. glutamicum* (D, lane 3) produces no soluble ST enzyme, only free MBP is recovered. In *E. coli* (C) the co-expression of the hPDI-QSOX1B fusion had the greatest impact on the recovery of full-length and soluble MBP-hST6GalI. In *C. glutamicum* (D) the degradation patterns are much more consistent

between all constructs. The HUST-166 construct (C and D, lane 9) with both hPDI-QSOX1B and DsbC co-expressed and lacking an *N*-terminal SUMO tag produced the most full-length and active ST enzyme. Control protein is MBP-hST6GalI produced in SHuffle® Express *E. coli*. Expected molecular weight of MBP-hST6GalI is 79.6 kDa, SUMO-MBP-hST6GalI is 90.6 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.

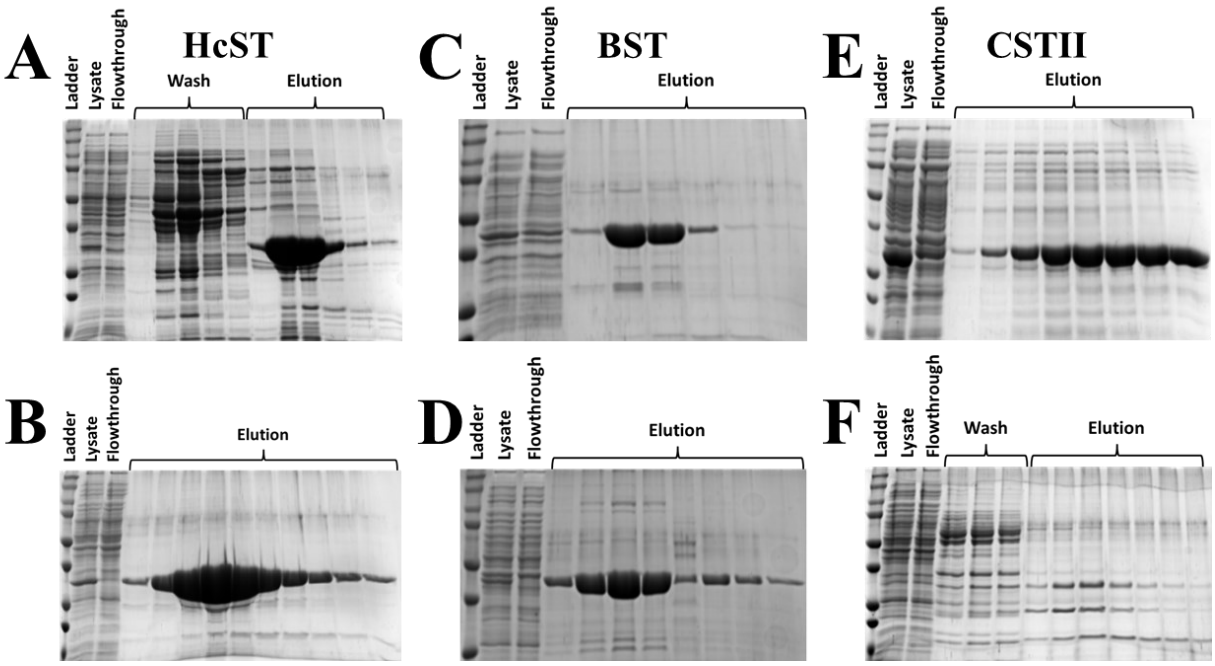


Figure S4. 8. Coomassie stained 15% SDS-PAGEs of fraction from IMAC purification of HcST (A and B), BST (C and D), and CSTII (E and F) produced in *E. coli* (top) and *C. glutamicum* (bottom).

All bacterial STs were recovered in similar quantities and of similar purity, except for CSTII produced in *C. glutamicum*. This enzyme either does not express well in the organism as only a minimal induction band is observed (F, lane 2), or is not folded correctly leading to rapid degradation.

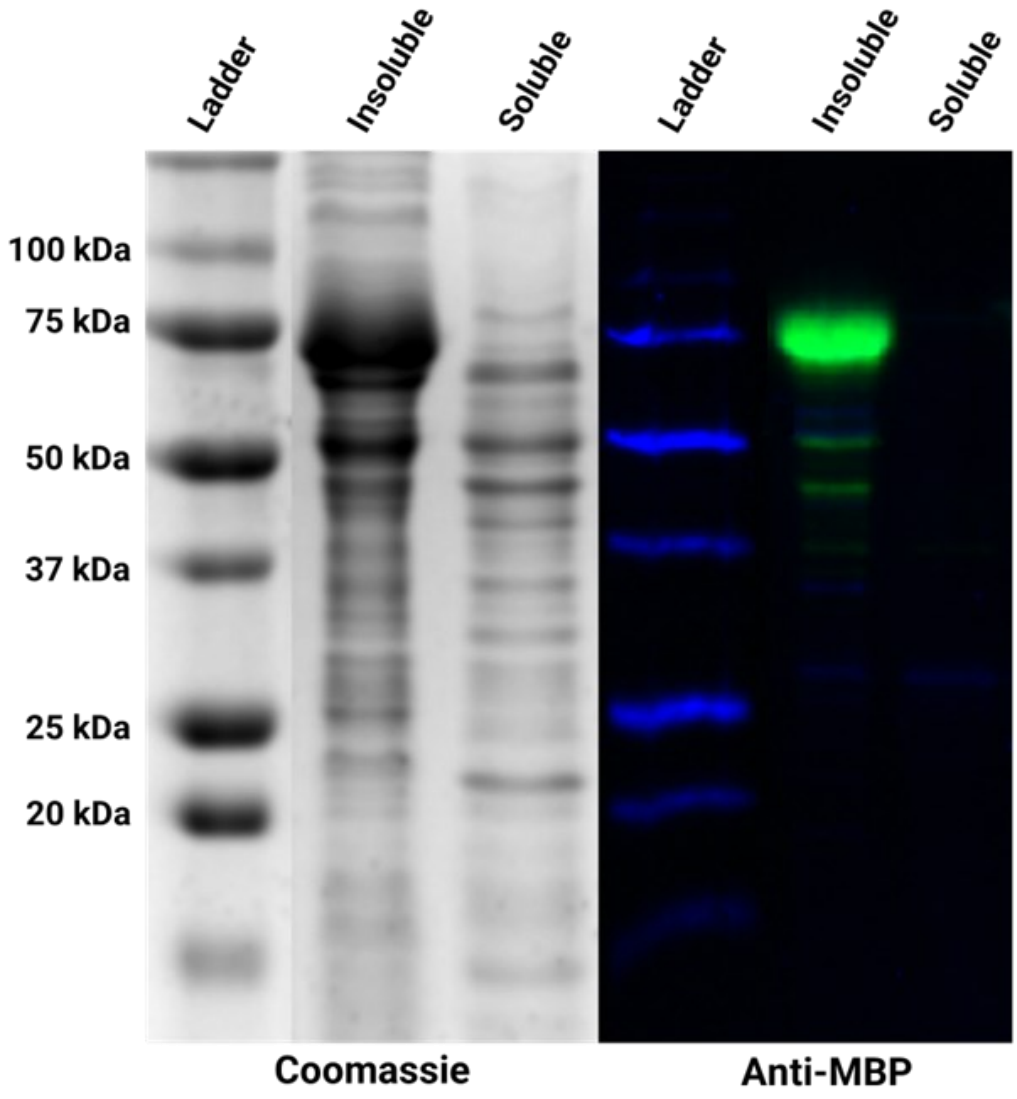


Figure S4. 9. Coomassie stained 15% SDS-PAGE (left) and anti-MBP-HRP Western blot (right) showing insoluble production of MBP-hST6GalI fusion protein in *C. glutamicum*.

When produced in *E. coli* (not shown) or *C. glutamicum* via pCGE-31 expression, the MBP-hST6GalI is found entirely in the insoluble fraction of 100,000 x g clarified supernatants. Coomassie staining shows a strong induction band (A) in the insoluble fraction and Western blots probed by anti-MBP-HRP (B) confirm that no fraction of the produced fusion is soluble.

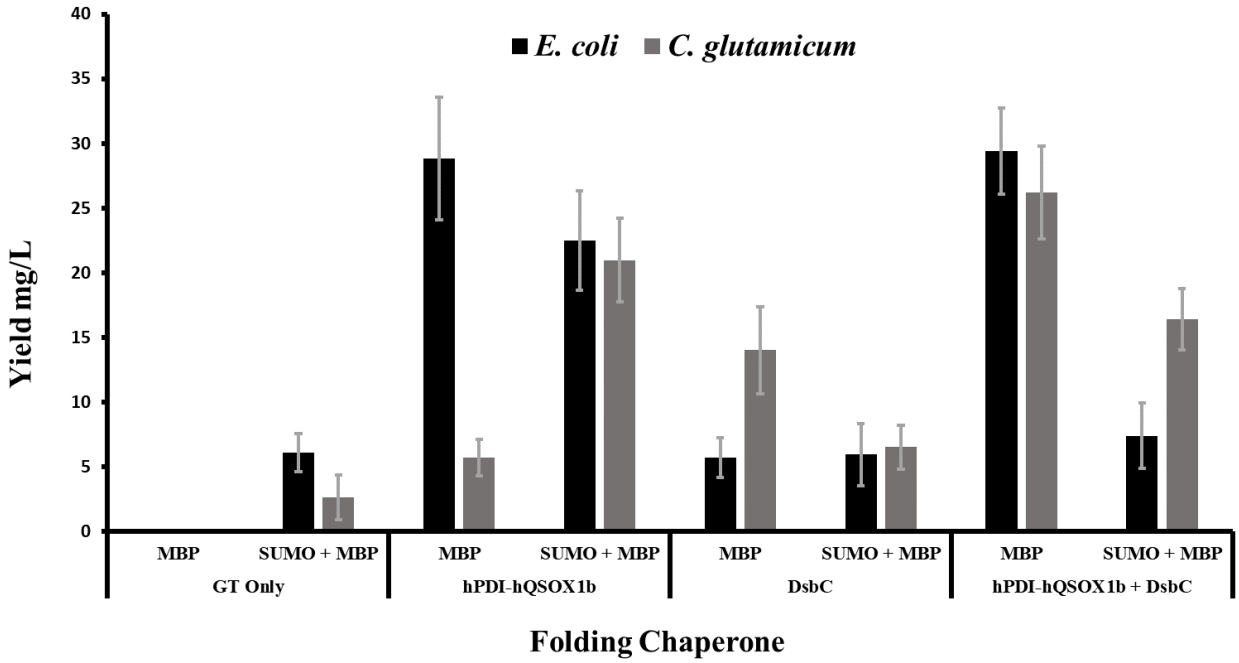


Figure S4. 10. Total yield (mg) of each recovered hST6GalI construct per litre of culture.

Both an *N*-terminal SUMO tag and all folding chaperones tested allowed for the recovery of some full-length MBP-hST6GalI fusion, whereas the MBP-hST6GalI fusion expressed on its own in both organisms did not produce any soluble or recoverable full-length fusion. The MBP-hST6GalI co-expression construct utilizing both folding chaperones (hPDI-QSOX1b and DsbC) had the greatest total yield (mg/L culture) in *C. glutamicum* of approximately 25 mg/L culture. The MBP-hST6GalI construct co-expressing either hPDI-QSOX1b or hPDI-QSOX1b and DsbC in *E. coli* had comparably high total yields of approximately 30 mg/L culture.

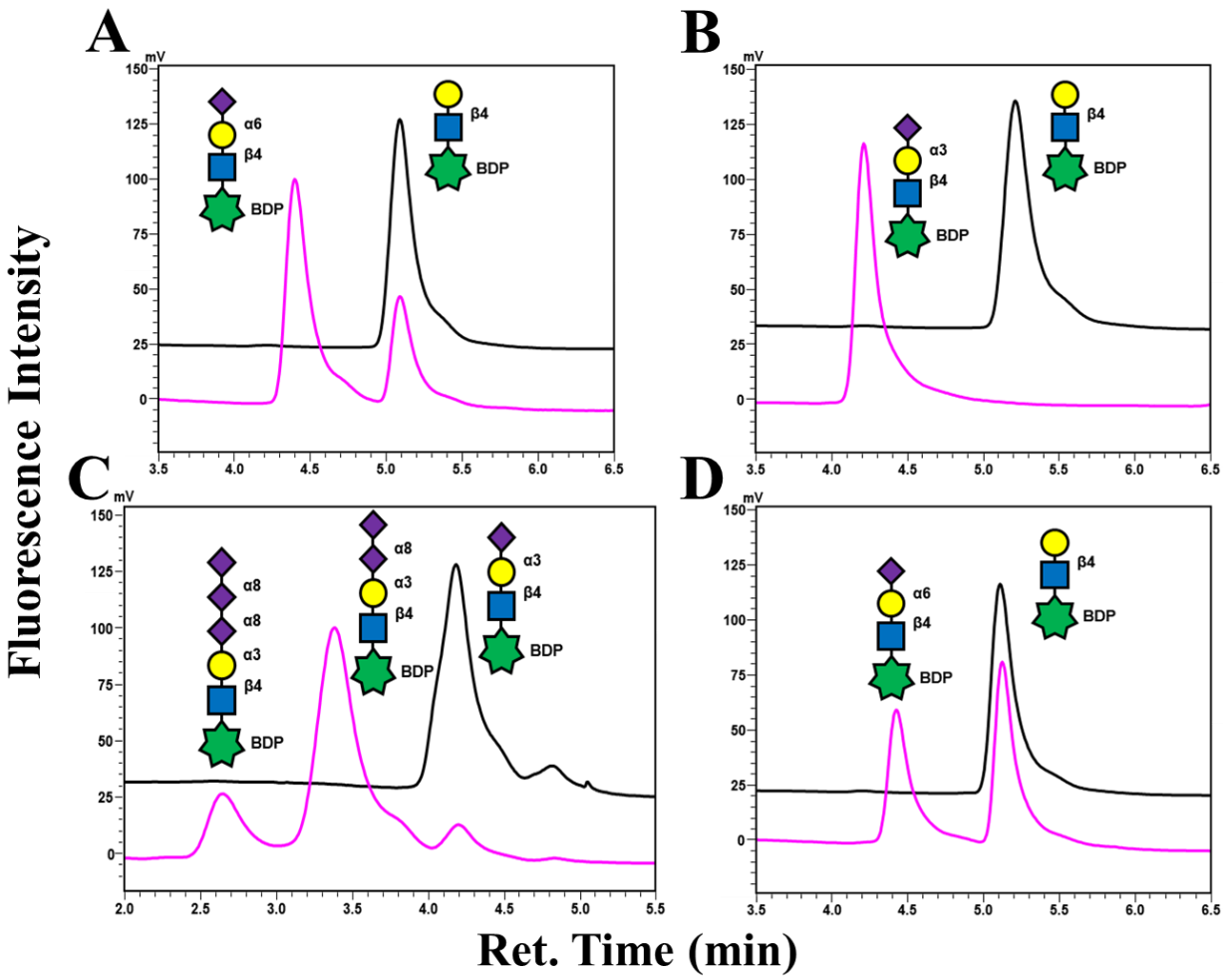
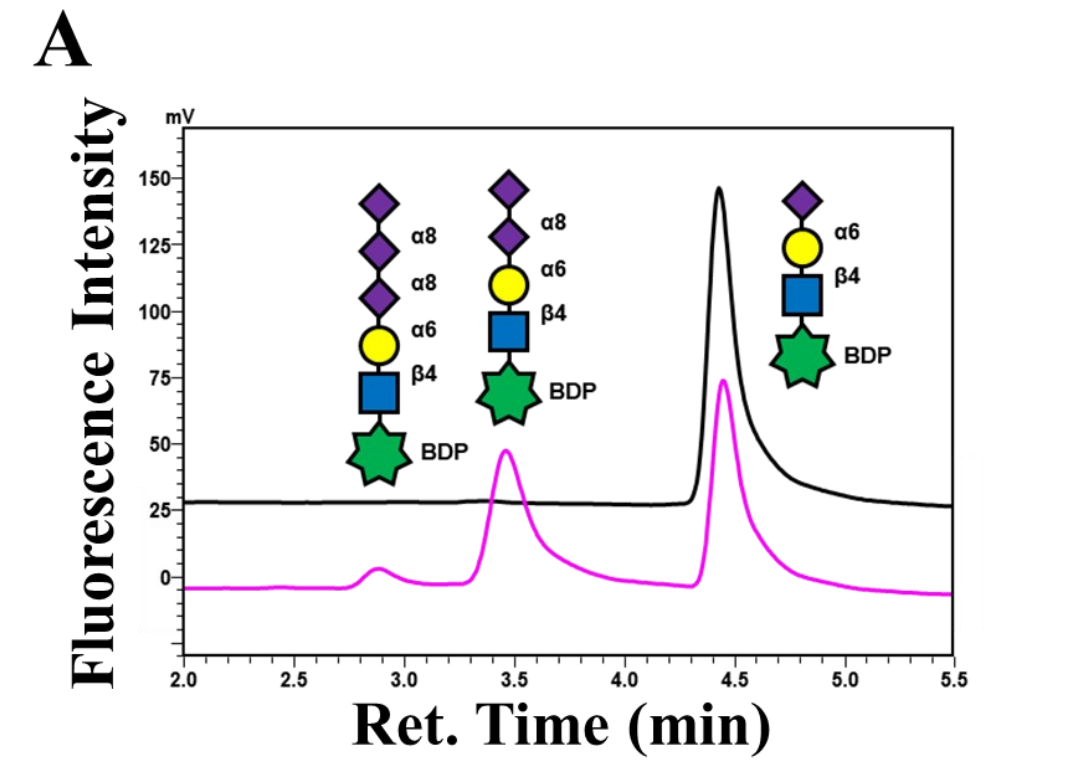


Figure S4. 11. Example HILIC HPLC traces from small molecule reactions by HcST (A), BST (B), CSTII (C), and HUST (D) showing separation by Neu5Ac addition and differentiating between α2,3 and α2,6 linkages.

Products from small molecule (BDP-LacNAC/BDP-GM3) reactions were separated using an Accucore™ C18 (Thermo Scientific) column. Distinct separation can be seen in retention times between oligosaccharide structures of the same charge and containing differentially linked Neu5Ac residues: α2,6-Neu5Ac for HcST and HUST (A and D), α2,3-Neu5Ac for BST (B). Separation of differentially charged produced is also evident for α2,8-Neu5Ac additions CSTII (C).


A
Fluorescence Intensity
mV
150
125
100
75
50
25
0
α8
α8
α6
β4
BDP
α8
α6
β4
BDP
α6
β4
BDP
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
Ret. Time (min)

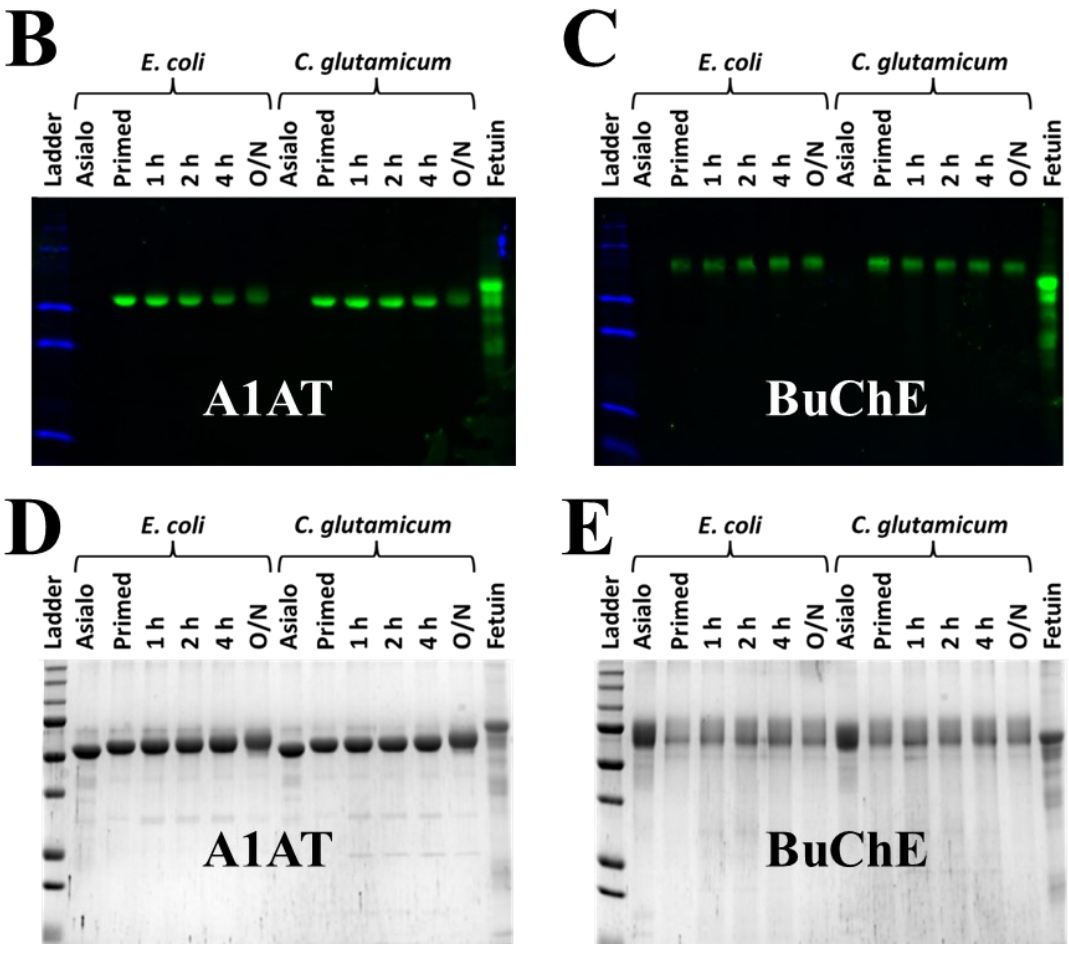

B
E. coli
C. glutamicum
Ladder
Asialo
Primed
1 h
2 h
4 h
O/N
Asialo
Primed
1 h
2 h
4 h
O/N
Fetuin
A1AT
C
E. coli
C. glutamicum
Ladder
Asialo
Primed
1 h
2 h
4 h
O/N
Asialo
Primed
1 h
2 h
4 h
O/N
Fetuin
BuChE
D
E. coli
C. glutamicum
Ladder
Asialo
Primed
1 h
2 h
4 h
O/N
Asialo
Primed
1 h
2 h
4 h
O/N
Fetuin
A1AT
E
E. coli
C. glutamicum
Ladder
Asialo
Primed
1 h
2 h
4 h
O/N
Asialo
Primed
1 h
2 h
4 h
O/N
Fetuin
BuChE

Figure S4. 12. CSTII assays using α2,6 primed substrates: small molecule reactions (A), free N-glycans (B), and mNeonGreen-diCBM40 lectin blots showing *in vitro* sialylation of A1AT (C) and BuChE (D) with corresponding Coomassie-stained 15% SDS-PAGEs (E and F).

CSTII also accepts α2,6-linked Neu5Ac as a substrate, though with lower efficiency. To confirm the *C. glutamicum* expression of the enzyme was not altering this functionality, assays using α2,6-Neu5Ac primed substrates were performed. An example HILIC HPLC trace of small molecule reactions performed with CSTII using Neu5Ac-α2,6-lactosamine (A) showing how activity was calculated and compared between the enzyme made in the different hosts – the α2,8-Neu5Ac transfer to a α2,6-Neu5Ac substrate was comparable. The *in vitro* sialylation of α2,6 primed A1AT (left) and BuChE (right) shows enzyme made in both hosts had comparable activity on these substrates detected by loss of mNeonGreen-diCBM40 fluorescence (B and C) and increase in molecular weight as seen by Coomassie staining (D and E).

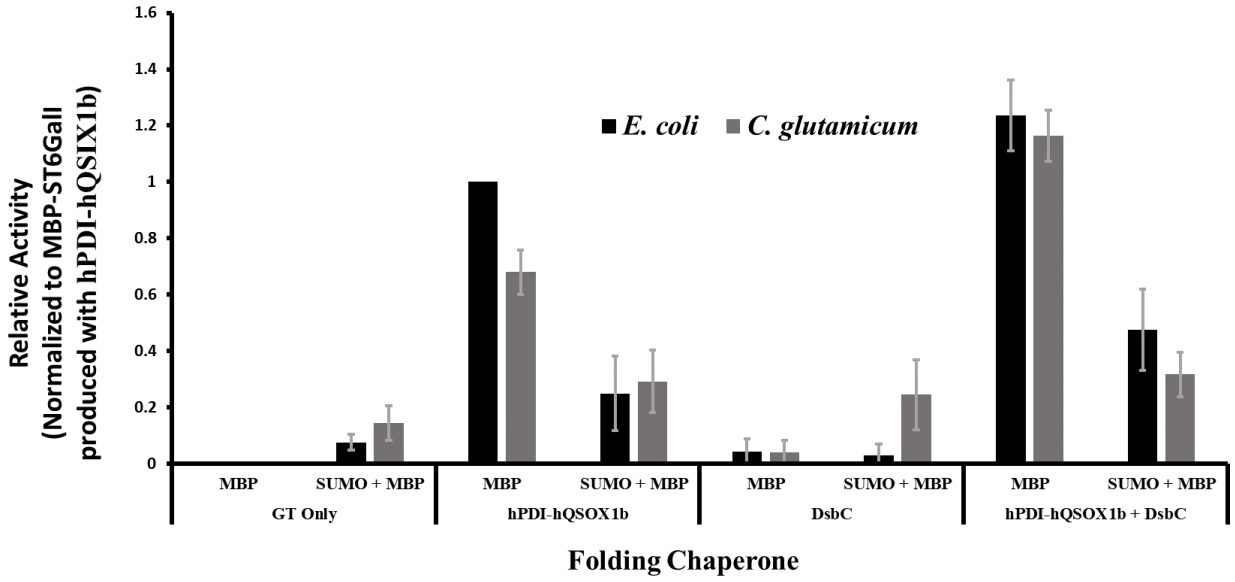


Figure S4. 13. Activities of each MBP-hST6GalI pDual construct as determined by small molecule assays and compared relative to MBP-hST6GalI co-expressed with the hPDI-QSOX1b fusion in BL21 *E. coli.*

In both strains, the recovered MBP-hST6GalI fusion co-expressed with both folding chaperones (hPDI-QSOX1b and DsbC) displayed the greatest specific activity on the small molecule substrate BDP-LacNAc, approximately 1.2-fold greater than the MBP-hST6GalI co-expressed with hPDI-QSOX1b in *E. coli*. The construct (HUST-166) was therefore selected for further comparison to assess the potential of *C. glutamicum* for the purposes of producing active eukaryotic STs.

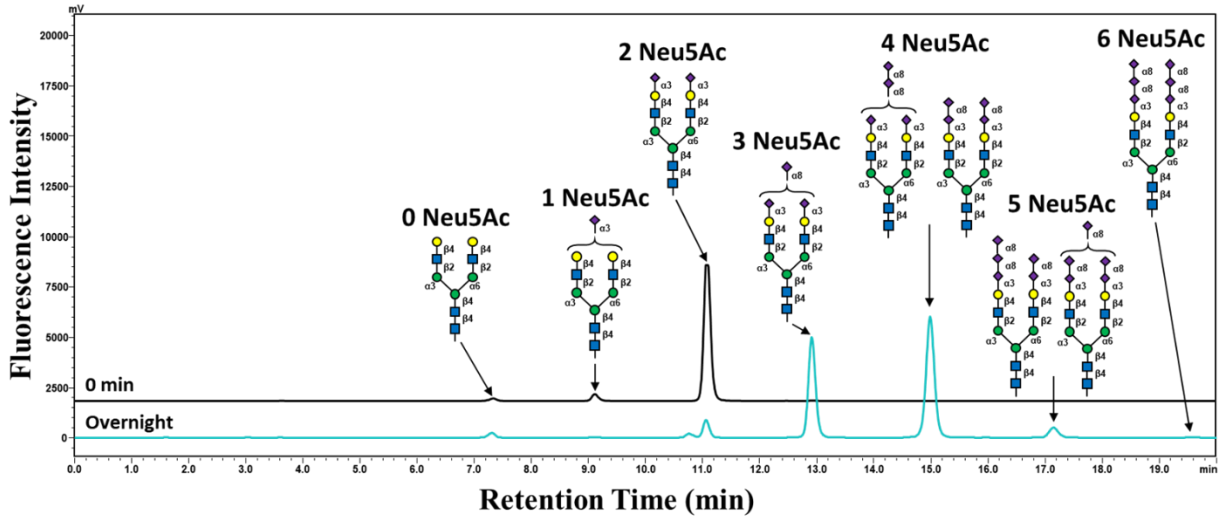


Figure S4. 14. HPLC traces showing α2,6-Neu5Ac primed G2 N-glycan modification by CSTII and lack of resolution of higher sialylated glycoforms.

CSTII from both strains was comparably active on both α2,3- and α2,6-Neu5Ac primed free *N*-glycans; however, the multiple additions of α2,8-linked Neu5Ac by the enzyme in addition to the lack of separation of glycans of similar charge by the column did not allow for the accurate determination of an branch preference for the enzyme. While an interesting addendum to the biochemical characterization of this enzyme, the information was not required to assess the production capability of *C. glutamicum*. The *in vitro* sialylation of glycoproteins (which has been previously reported with this enzyme) was therefore used to better evaluate the proposed heterologous host, *C. glutamicum*.

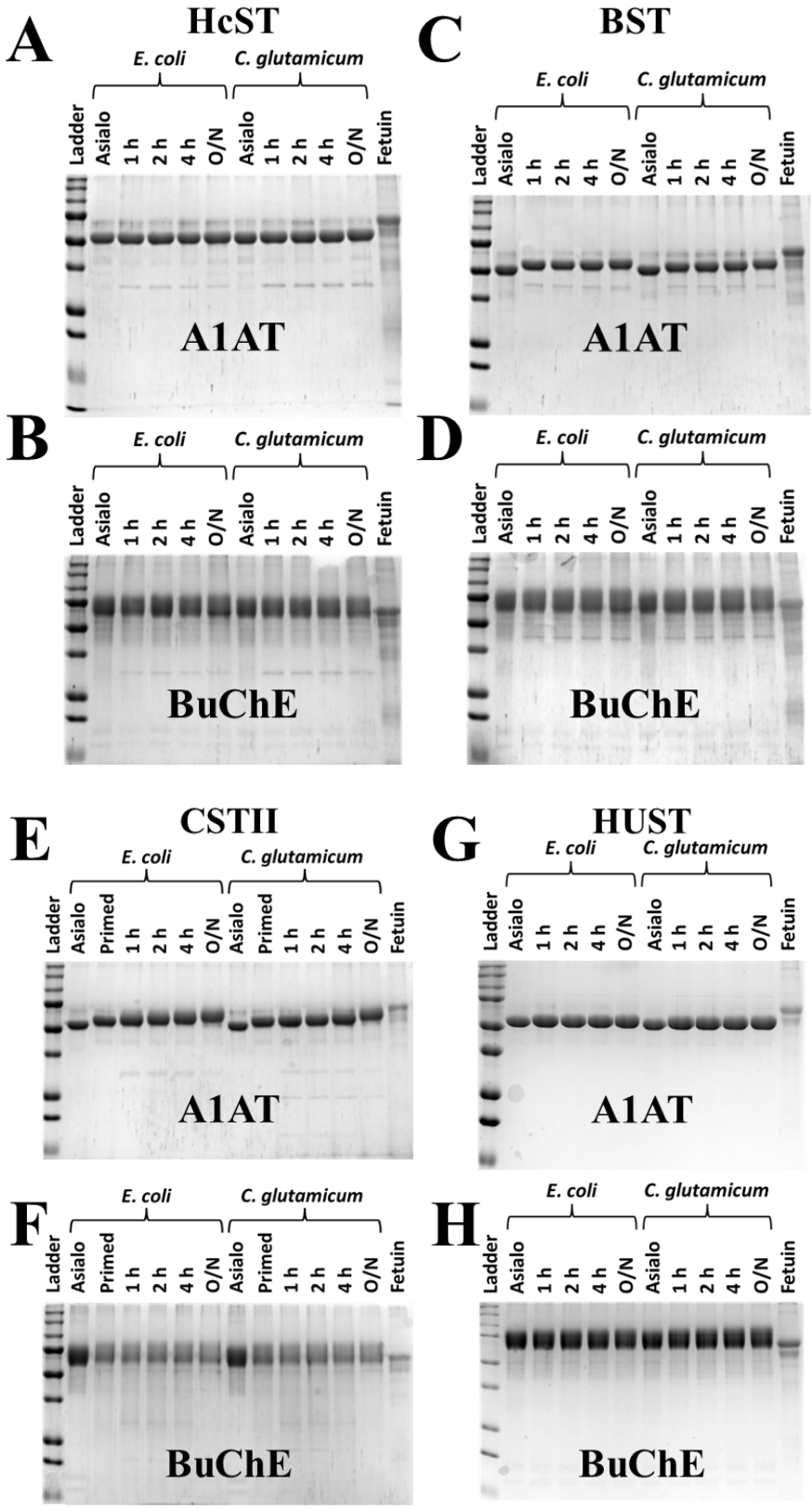

A
HcST
E. coli
C. glutamicum
Ladder
Asialo
1 h
2 h
4 h
O/N
Asialo
1 h
2 h
4 h
O/N
Fetuin
A1AT
B
BuChE
C
BST
A1AT
D
BuChE
E
CSTII
E. coli
C. glutamicum
Ladder
Asialo
Primed
1 h
2 h
4 h
O/N
Asialo
Primed
1 h
2 h
4 h
O/N
Fetuin
A1AT
F
BuChE
G
HUST
A1AT
H
BuChE

Figure S4. 15. Coomassie stained 15% SDS-PAGEs comparing in vitro sialylation of asialoA1AT (A, C, E, G) and asialoBuChE (B, D, F, H) by STs purified from *E. coli* and *C. glutamicum*.

The HcST enzyme produced in both stains was only minimally active on both glycoprotein substrates (A and B). The BST enzyme was readily able to modify the *N*-glycans of both asialoA1AT (C) and asialoBuChE (D) with a terminal α2,3 Neu5Ac. The CSTII enzyme requires a terminal α2,3 Neu5Ac, to which it adds another α2,3 Neu5Ac. Both this addition and the subsequent α2,8 Neu5Ac addition can be monitored by molecular weight shift and diffusion of the band corresponding to the target protein. The CSTII enzyme is capable of modifying both target asialoglycoproteins (E and F); molecular weight shifts are more evident with A1AT as it is a discreet band, the heterogenous glycan profile of BuChE results in a more diffuse band where molecular weight shifts are more difficult to discern. The HUST-166 construct (G and H) readily modifies both targets with α2,6 Neu5Ac as glycoproteins are the native substrate of this enzyme. Expected molecular weight of asialoA1AT is 46.7 kDa, asialoBuChE is 68.4 kDa. Molecular weight markers are the Bio-Rad All Blue ladder.