Molecular, Genetic, and Process Approaches for Improving Secreted Pharmaceutical Protein Quality in *Komagataella phaffii*

by

Yuchen Yang

Submitted to the Department of Chemical Engineering in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Authored by: _____

Yuchen Yang
Department of Chemical Engineering, July 2024

Certified by: _____

J. Christopher Love
Raymond A. (1921) and Helen E. St. Laurent Professor of Chemical Engineering
Thesis Supervisor

Accepted by: _____

Hadley D. Sikes
Willard Henry Dow Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Molecular, Genetic, and Process Approaches for Improving Secreted Pharmaceutical Protein Quality in *Komagataella phaffii*

by Yuchen Yang

Submitted to the Department of Chemical Engineering in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Chemical Engineering

Thesis supervisor:
J. Christopher Love
Raymond A. (1921) and Helen E. St. Laurent Professor of Chemical Engineering

ABSTRACT

Biopharmaceutical products constitute a significant portion of the global bioeconomy. Compared to traditional synthetic small-molecule drugs, recombinant therapeutic proteins offer advantages like enhanced specificity and reduced side effects, and there has been tremendous growth in their innovation thanks to modern DNA technologies and AI-driven algorithms. While mammalian platforms such as Chinese Hamster Ovary (CHO) cells are commonly used for their high production titer and capability for complex post-translational modifications, thier high cost of goods manufactured can greatly constrain biopharmaceutical global accessibility. The yeast *Komagataella phaffii* is the prime candidate for next-generation biomanufacturing for reasons including simpler host biology, reduced time to market, and better sustainability. Nevertheless, product quality, such as size/charge variants and non-human glycosylation, can be of major concern for proteins secreted from this host organism.

This thesis explores three different engineering approaches aimed at improving the quality of both aglycosylated and glycosylated proteins, with a particular focus on monoclonal antibodies, the leading class of protein biopharmaceuticals by both sales and innovation.

Firstly, we demonstrated significant quality improvements through molecular sequence engineering of aglycosylated monoclonal antibody backbones. By making informed, conservative mutations to two or three amino acid residues, we greatly reduced product-related variants from proteolysis and N-terminal variations. We further showed the comparability between yeast- and CHO-secreted products, providing a framework for rapid product development with this unconventional yeast.

Secondly, we applied CRISPR-Cas9 gene editing technology to humanize the glycosylation pathway of *K. phaffii*. We achieved homogeneous G0 glycosylation on a reporter peptide by resolving a previously unreported synthetic lethality via a transcriptomics-informed approach. Key challenges for monoclonal antibody glycosylation were also identified through further comprehensive pathway engineering.

Lastly, we examined the performance of glycoengineered *K. phaffii* strains under varied process conditions. Employing a machine learning algorithm, we improved the desired glycan abundance on a subunit vaccine candidate. The process-robustness of engineered strains suggests the potential of this host as a viable commercial biomanufacturing host.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of many people. In research and in day-to-day life, I am sincerely grateful for their invaluable contributions.

To my advisor, Professor J. Christopher Love, whose guidance and support have been indispensable throughout my graduate studies. From sharing your vision of future biomanufacturing to inspiring the lab and many others, you have offered me the freedom to explore innovative ideas and pursue ambitious research goals. Thank you for your mentoring and for helping me become a better independent scientist.

To my committee Professor Kristala Prather and Professor Kate Galloway, who have provided invaluable feedback on my research and supported both my personal and professional growth.

To Danielle Camp, who has been a key mentor in developing my soft skills and the biggest cheerleader for all of my accomplishments, no matter how small or inconsequential they might seem.

To my lab mates, both past and present, who have inspired and supported me day in and day out. Neil Dalvie for welcoming me to the lab and for many engaging discussions about synthetic biology. Chris Naranjo for being a technical wizard in protein analytics and for passing down his wisdom. Carmen Elenberger, Shuting Shi, Hayley Ford, and Sergio Rodriguez for always lending an ear and offering their support whenever I needed it. Joe Brady, Ryan Johnston, Tim Logeree, Josh Hinckley, Harini Narayanan, Jason Zhang, Raghav Acharya, Brittney Sunday, Anastasiya Grebin, and Shervin Tabrizi for their camaraderie and collaboration.

To the friends made throughout my time at MIT, who have kept me going over the past six years. The "9 hours per week" crew, for enduring the first-year experience together, then practice school in Houston and Dubai, and all the memories since. N/A Dance, Michael Mandanas, Xochitl Luna, and many others, for satisfying my appetite for the performing arts and for being amazing people both on and off-stage. Ellison Scheuller and Tess Engst-Mansilla for being a consistent source of support and joy.

To the friends supporting me since Berkeley, especially Laura Strong and Jennifer Zhang, who have been a constant through both challenges and triumphs. Thank you for listening to my vents, for lending me confidence, for supporting my goals, for keeping me motivated, for celebrating my victories, and, most importantly, for having been doing so since before my thesis work began.

To my family – my day-one supporters – thank you for your unconditional love and unwavering support from thousands of miles away.

TABLE OF CONTENT

LIST OF FIGURES AND TABLES

1. INTRODUCTION

1.1. Recombinant proteins as a biopharmaceutical modality

Since the early 1980s, the global biopharmaceutical market has grown tremendously and accounts for more than US $343 billion (2021 estimate).[1] Compared to traditional synthetic small-molecule drugs, biopharmaceuticals can have enhanced specificity and activity, thus reducing adverse side effects.[2] Multiple therapeutic modalities are included under the biopharmaceutical umbrella. Recombinant proteins, expressed in non-native host organisms – dominate the market with more than 95% of the market sales and comprise the majority of newly approved products.[1] New modalities such as nucleic acid-based and engineered cell-based products are emerging, but they are still undergoing continued technological development, and their market shares remain relatively low.[3,4]

Types of pharmaceutical proteins also can vary from recombinant hormones, blood factors, to vaccines, but the predominant product class is monoclonal antibodies (mAbs), which have accounted for more than half of the global protein-based biologics sales since 2012 and steadily grown over the years.[1] mAbs have been in clinical use since the 1980s for treatment of a wide range of diseases and conditions, especially cancer and autoimmune diseases.[5]

1.2. *Komagataella phaffii* as a workhorse for biomanufacturing

Mammalian systems, such as Chinese Hamster Ovary (CHO) cells and Human Embryonic Kidney (HEK) cells, are the most frequently used expression system for pharmaceutical recombinant proteins, especially for genuinely novel active pharmaceutical ingredients (APIs). These systems can support high production titer, as mAb titers of more than 10 g/L have been reported from recombinant CHO cell lines,[6] and are capable of complex post-translational modifications (PTMs), most notably glycosylation. The high production titer

upstream, combined with the development of mature downstream recovery and purification,[7] makes mammalian platforms, especially CHO, the preferred host for manufacturing new biologics.

Nevertheless, next-generation biomanufacturing strategies are still being explored for both societal and commercial reasons. Motivating factors including improving access to these medicines for global populations,[8] increasing demand for sustainable bioprocessing,[9] reducing time to market as a key business advantage, and responding rapidly to global events like pandemics.[10] For mAbs production in CHO, the complex culture media, stringent operating conditions to reduce risks of contamination, and viral inactivation/filtration of secreted products all contribute to the high cost of goods manufactured (COGm), which today can range from $30-100 per gram.[11] This greatly constrains access to these medicines in low- and middle-income countries (LMICs) – indeed, 80% of all doses for registered products are administered only in North America and Europe.[12] Corporate commitments to reduce carbon footprints and geopolitical interests for enabling a circular bioeconomy also underscore the importance for continued developments to intensify, consolidate, and reduce energy, water, and raw materials used in bioprocessing.[13,14] Finally, accelerated development demonstrated during the CoVID-19 pandemic showed the potential to reduce the time from discovery to first-in-human clinical trials to four to six months, albeit with significant corporate and regulatory cooperation to accomplish.[15] These examples emphasize the possibility for more timely development, clinical assessment and commercialization of innovative new biopharmaceuticals. Together, these factors motivate the need for continued innovations in the approaches to manufacture mAbs and other recombinant proteins with shorter development timelines, lower production costs, and improved sustainability.

Other expression platforms used for recombinant protein production include bacteria and fungi. Compared to mammalian cells, they can grow to much higher density on much cheaper media.[16,17] Although *Escherichia coli* and other prokaryotic systems can achieve high volumetric productivity of recombinant proteins, their usage in producing complex pharmaceutical proteins is limited due to the lack of native cellular machinery for PTMs.[17–19] Engineering efforts have been reported to facilitate the implementation of some PTMs in *E. coli*,[20–22] but yeasts are commonly regarded as the non-mammalian host of interest for producing proteins with more complex modification. The biology of these single-cell eukaryotic organisms can potentially directly increase the volumetric output of production (through faster doubling times and reduced process requirements),[10,23,24] and boasts other advantages including proteolytic maturation, glycosylation, and formation of disulfide bonds.[25] *Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha*, *Kluvomyces lactis*, and *Yarrowia lipolytica* are among yeast expression systems that are already designated Generally Regarded As Safe (GRAS) by regulatory agencies.[26] They contain an advanced secretory pathway with a stacked Golgi apparatus for producing PTMs in a manner similar to higher eukaryotes and secrete fewer native host cell proteins, simplifying downstream processing and facility design.[27] In the case of pharmaceutical protein production, however, *Komagataella phaffii* is the alternative host of choice because of its high productivity, simple process, and past industry experience – indeed, it is routinely used to manufacture therapeutic proteins like insulin and subunit vaccines in both high income countries and LMICs.[27,28]

1.3. Critical product qualities of pharmaceutical proteins

Compared to synthetic small-molecule drugs, pharmaceutical proteins are much larger and more complex. For example, common over-the-counter painkiller acetaminophen (Tylenol®)

only contains 20 atoms, while insulin, one of the smallest pharmaceutical proteins, contains 788 atoms,[29] and trastuzumab (Herceptin®), a common monoclonal antibody for cancer therapy, is comprised of more than 20,000 atoms.[30] Because these proteins are produced from biological processes, molecular heterogeneity, imperfect cellular processing, and chemical and enzymatic alterations can naturally occur during production.[31] Common product-related impurities include aggregation, fragmentation, C- and N-terminal modifications, oxidation, deamidation, N- and O-linked glycosylation, glycation, conformation, and disulfide bond modifications.[32] The complex structures of these proteins mean that these impurities are not easily characterized and can significantly impact biological activity, pharmacokinetic/pharmacodynamic (PK/PD) profiles, immunogenicity, and overall safety/toxicity if altered.[32,33] Thus lie the reason behind why *K. phaffii*, despite its advantages in ease and reduced costs for large-scale manufacturing, fast growth, no risk of viral contamination, remains sidelined in producing complex pharmaceutical proteins.[25] The yeast-like glycosylation imparted on expressed proteins can lead to immunogenicity and rapid clearance, making it unsuitable for many products.[34] Proteolysis of the target sequences can also affect the yield and purity of desired proteins.[35–37] Yeast also has a lower folding capacity compared to mammalian cells, which can result in the aggregation of incorrectly folded proteins.[38]

1.4. Diverging Glycosylation Pathways: Yeast vs. Human

Many of the described impurities and product-related variants can be reduced or removed during downstream purification processes. Due to the purity requirements for pharmaceutical proteins, high-resolution chromatography methods are usually used in purification and can distinguish the desired product from its impurities based on their differences in charge (ion exchange chromatography), hydrophobicity (hydrophobic interaction chromatography),

15

molecular recognition (affinity chromatography), or size/shape (size exclusion chromatography).[39] Yeast-produced glycan variants, however, cannot be removed by these methods, because the glycosylation pathways differ significantly between yeast and higher eukaryotes like human or CHO.

The N-linked glycosylation pathway in the endoplasmic reticulum (ER) is conserved across most eukaryotic organisms.[18,40] Glycan synthesis starts on the cytoplasmic side of ER with the phosphorylation of dolichol, a membrane anchor upon which two N-acetylglucosamines (GlcNAc) and five mannoses (Man) are added by glycosyltransferases, using nucleotide-activated sugars as substrates. The lipid-linked oligosaccharide (LLO) is then flipped from the cytoplasmic side to the lumen side of ER by flippase.[41] Once inside the ER, the LLO is further modified by a series of glucosyltransferases until the oligosaccharide is consisted of three glucoses (Glc), nine mannoses, and two core GlcNAc's (Glc3Man9, the core GlcNAc's are often omitted in writing). The glycan is then transferred onto the appropriate asparagine residue in the nascent peptide by the oligosaccharide transferase (OST) complex.[41] Two glucosidases cleave off the three glucoses, during which the membrane-bound chaperone calnexin selectively binds to GlcMan9 glycan and facilitates protein folding. Finally, an ER-resident mannosidase cleaves off one of the terminal mannoses before the glycoprotein is transported to the Golgi apparatus for further glycosylation.[40]

The glycosylation pathway in the Golgi apparatus is vastly different between yeasts and mammalian cells. Upon entering Golgi, most glycoproteins carry Man8 glycan structure. In yeasts, the protein glycan is further modified by more mannoses into a hypermannosylated structure. The outer chain hypermannosylation starts with the addition of an α-1,6-mannose by Och1p. Depending on the yeast species, more than 50 α-1,2-mannoses can then be added by mannan polymerase complexes and mannosyltransferases, before phosphomannoses and α-1,3-mannoses

cap the hypermannosylated outer chain.[18,42,43] The mannosylation of core mannose structure is much less extensive, and only phosphomannose and α-1,3-mannose addition is observed.[40] Compared to other yeasts, *K. phaffii* exhibits a lesser degree of hypermannosylation due to its inability to perform α-1,3-mannosylation.[40]

In contrast to yeasts, mammalian cells process N-linked glycans to structures with more complex and diverse sugar moieties. Golgi-resident mannosidases I and II remove five mannoses from the glycan, while GlcNAc transferases I and II add two GlcNAc residues onto the core trimannosyl structure. The resulting GlcNAc2Man3 structure is often referred to as the G0 glycoform (for the absence of galactose moiety on the glycan).[44] Common further glycan modifications in higher eukaryotes include adding galactoses (and sometimes sialic acids, specifically N-acetylneuraminic acids) to the biantennary structure, adding core fucoses, and branching further into tri-antennary, tetra-antennary, and penta-antennary structures.[40,45,46]

1.5. Engineering Approaches for improving product quality

In the biomanufacturing of pharmaceutical proteins, downstream processing is resource-intensive and can account for up to 80% of production costs for a mAb process.[47] It is therefore economically attractive to reduce impurities and variants early in production to reduce the burden on downstream purification.

1.5.1. Improving product quality by sequence engineering

Protein sequence is the earliest intervention step in ensuring the success of a therapeutic drug. When producing a recombinant protein in its non-native host, it is very likely that the non-native environment is less than optimal for protein expression, folding, maturation, and secretion (depending on the host). Advances in molecular dynamics simulation,[48] X-ray crystallography,[49] and, more recently, machine learning[50] have equipped researchers with better tools to visualize

protein structures and potentially diagnose causes for product-related variants on a sequence level. Indeed, through understanding the host biology and modifying hotspots in the protein sequence, impurities and variants can be eliminated before they ever enter the process. Our lab has demonstrated the successful molecular engineering of several vaccine candidates to reduce their secreted product-related variants, thereby improving their production yield and manufacturability.[51,52]

In Chapter 2 of this thesis, we discuss the adaptation of aglycosylated monoclonal antibodies for improved production in *K. phaffii*. Monoclonal antibodies are the leading therapeutic protein in both market sales and approval,[1] making it an attractive candidate to demonstrate the possibility of adopting an alternative host for its industrial-scale production. We show that significant improvement in product quality can be achieved by making small, conservative, informed modifications to the IgG1 conserved region. By demonstrating the comparability between an engineered mAb secreted by *K. phaffii* and by CHO, we are able to provide further evidence supporting the adoption of alternative hosts for therapeutic protein production.

1.5.2. Improving product quality by strain engineering

Although molecular sequence engineering is able to eliminate product-related variants from the get-go, it is imperative to ensure that any amino acid modification does not negatively impact the efficacy and safety of recombinant therapeutic proteins. In the biopharmaceutical industry, sequence modifications to existing biologics also preclude the labeling of these engineered molecules as their respective biosimilars and can raise concerns during regulatory reviews.[53,54] Furthermore, not all proteins are created equal, and some are a lot less tolerable to the change of production hosts than aglycosylated monoclonal antibodies. Many pharmaceutical

products, including antibodies, hormones, and growth factors, are glycoproteins that require proper glycosylation for folding, stability, serum half-life, and appropriate PK/PD properties.[17,19,40,55] Since the native yeast glycosylation pathway significantly differs from its counterpart in mammals, amino acid mutations alone are insufficient to support the translation of a glycoprotein process to yeast. Thus, although protein sequence modifications can be an attractive option for improving manufacturability, especially in the adaptation of an existing technology to a new host, it may not always be a viable strategy.

In addition to molecular sequence, the choice of strain (or cell line) also heavily impacts the quality of secreted proteins. Engineering the host cell is thus an alternative option for improving product quality in a minimally invasive manner. Our ability to engineer the host has grown tremendously thanks to advances in genetic engineering techniques. Compared to mammalian systems, yeasts have much smaller genomes and are more amenable to genome engineering.[27] Nevertheless, compared to *E. coli* and *S. cerevisiae*, two model organisms with extensive studies on their basic biology and development on appropriate synthetic biology toolbox,[56,57] *K. phaffii* is still in need of more complex and host-informed engineering tools that allow more efficient pathway engineering, especially regarding therapeutically relevant protein glycosylation.[58,59]

In Chapter 3 of this thesis, we demonstrate the adaptation of a CRISPR-Cas9 system for multiplexed humanization of *K. phaffii* glycosylation pathway. Chapter 4 then builds upon this and describes our pathway engineering to achieve homogeneous G0 glycosylation on a reporter peptide by tuning down the activity of the commitment step enzyme. Chapter 5 details our subsequent efforts in engineering the glycosylation pathway to accommodate more complex proteins, including a vaccine subunit, a colony-stimulating factor, and a monoclonal antibody.

We demonstrate tuning multiple "knobs" to better balance pathway flux and competing kinetics and suggest that accessibility of N-linked glycosylation site is likely the pathway bottleneck.

1.5.3. Improving product quality by process engineering

In bioprocesses, cellular behavior can vary drastically responding to any environmental changes. Hence, in addition to molecular and strain engineering, process conditions also dictate product quality. Indeed, cultivation temperature, carbon source, and media supplementation, among many more bioprocessing independent variables, all have tremendous impacts on intracellular machineries.[60–62] Like sequence modifications, these process factors have their limitations, especially for post-translational modifications like glycosylation, which requires catalysis of non-native enzymes. Furthermore, process condition optimization is a multivariate optimization problem – the media alone can consist of more than ten different components.[63] Standard One-Factor-At-a-Time (OFAT) and Design-Of-Experiment (DOE) approaches can be labor-intensive and may be constrained by local optima.[64] Modern machine learning-based methods can overcome these challenges and have been successfully applied to both media composition and process parameter optimization.[65,66]

In Chapter 6 of this thesis, we investigate the effects of process parameters on the glycan profile of recombinant proteins secreted by *K. phaffii* strains engineered with the humanized, heterologous glycosylation pathway. We show the pathway activity's relative invariability with respect to temperature and carbon source concentration, contrasting with the responses observed in CHO cells under similar perturbations. We also explore machine learning as a tool for optimizing media supplementation and demonstrate the potential of this method for enhancing glycan homogeneity of glycosylated pharmaceutical proteins.

2. ADAPTING AGLYCOSYLATED MONOCLONAL ANTIBODIES FOR PRODUCTION IN *K. PHAFFII*

Protein sequence is the starting point of any biomanufacturing process, and a protein "designed-for-success" can circumvent many quality issues down the line. In this chapter, we use monoclonal antibodies (mAbs) as a case study. Next-generation biomanufacturing of this major class of biopharmaceuticals using alternative hosts like *Komagataella phaffii* could improve the accessibility of these medicines, address broad societal goals for sustainability, and offer financial advantages for accelerated development of new products. Antibodies produced by *K. phaffii*, however, may manifest unique molecular quality attributes that could raise potential concerns for clinical use. We demonstrate here conservative modifications to the amino acid sequence of aglycosylated antibodies based on the human IgG1 isotype that minimize product-related variations when secreted by *K. phaffii*. We further show comparable biophysical properties and molecular variations between the sequence-modified NIST mAb secreted by *K. phaffii* and CHO cells. This suggests a path towards production of high-quality mAbs that could be expressed interchangeably by either yeast or mammalian cells.

2.1. Background and Motivation

Recombinant monoclonal antibodies (mAbs) are one of the most important classes of biopharmaceuticals and among the fastest-growing biologic medicines by the number of new approved products and biosimilars, numbers of patients treated, and total revenue.[67] Advances in single-cell screening and machine learning algorithms to predict binding and manufacturability have accelerated the design, discovery and optimization of mAbs as therapeutics in the past ten years.[68,69] Manufacturing these medicines presently relies on highly-standardized processes using Chinese hamster ovary (CHO) cells for recombinant expression and Protein A (ProA)-based chromatography for recovery, with new advances in continuous manufacturing emerging for clinical and commercial use.[70]

The existing infrastructure to manufacture therapeutic mAbs has evolved through the continuous improvement of now-standardized "platform" processes using CHO cells.[71] The space-time yields for state-of-the-art fed-batch processes can reach ~0.2-0.5 g/L/d with fully continuous processing achieving up to ~2-4 g/L/d.[71,72] With these outputs, the costs of drug substance could approach ~$30 per gram, albeit with limited additional gains expected without new technologies for recovery.[8,72] Ultimately, reducing COGSm (and improving sustainability of bioprocessing) requires maximizing space-time yields from the smallest facilities with reduced labor.[73] Additional gains will result from removing process operations, increasing automation, and reducing biological variations through improved control or host biology.[74,75]

Here we have considered an alternative approach to advance the utility of *K. phaffii* for producing high-quality, full-length aglycosylated human immunoglobulin IgG1. Aglycosylated antibodies have emerged as an important new engineered sub-class for these drugs.[76] Advances in protein engineering make it possible to modulate the engagement of Fc receptors to minimize

immunological functions *in vivo*.[77] In 2020, the FDA-approved eptinezumab (Vyepti®), an aglycosylated mAb manufactured by Alder/Lundbeck in *K. phaffii* for treatment of chronic migraines.[78] Several other aglycosylated mAbs have been approved or are in clinical trials for indications ranging from diabetes to non-small-cell lung cancer.[79] Given this emerging use of engineered sequences, we sought here to adapt the sequence of the common human IgG1 mAb to reduce the typical variations observed in yeast-expressed mAbs. We postulated that conservative strategies for protein engineering (similar to those that we have applied previously to subunit vaccine candidates[51,52,80]) could improve the quality attributes while introducing minimal or no new liabilities. Specifically, we present in this chapter studies on the vector designs, including the signal peptides, and the sequence-related liabilities in the human IgG1 mAb that affect the proteolysis of the heterodimer. We show that minimal variations (2-3 amino acids) can improve the quality attributes of seven aglycosylated variants of mAbs in *K. phaffii*. Finally, we demonstrate that the quality of the engineered mAb secreted by *K. phaffii* is comparable to that produced by industrial-grade CHO cells.

2.2. Results

2.2.1. Characterization of product-related variants of a clinical mAb secreted from *K. phaffii*

We aimed to define a generalizable strategy to express high-quality, aglycosylated mAbs in *K. phaffii*. Towards this goal, we first analyzed the common molecular variations manifest in a monoclonal antibody expressed by *K. phaffii*. We assessed the impact of the expression vector and recombinant protein sequence on mAb expression in *K. phaffii* for a commonly studied IgG1κ mAb – trastuzumab (trade name Herceptin). This molecule is well characterized, and has

been previously expressed in *K. phaffii* by us and others.[81,82] It is also a useful model since most approved mAbs (and in clinical development) contain the IgG1κ constant region.

We used an in-house custom vector to create expression cassettes for both the heavy and light chains of trastuzumab and integrate them into a modified strain of *K. phaffii* (AltHost S-63) to evaluate the expression of each component. We expressed the heavy chain using the canonical methanol-inducible P_{AOX1} as the promoter and integration locus. We also expressed the light chain using another previously identified strong, methanol-inducible promoter, P_{DAS2}, as the promoter and integration locus.[83] We constructed expression vectors for both light chain and heavy chain using the described custom vectors. Previous reports have shown that a hybrid secretion signal comprising the pre region of *Saccharomyces cerevisiae* OST1 signal sequence and the pro region of *α*-mating factor signal peptide (αSP) can significantly improve recombinant protein expression in *K. phaffii* by promoting co-translational translocation.[84] We elected to use this preOST1-proαSP in our initial studies without the Glu-Ala-Glu-Ala (EAEA) sequence found at the end of the native αSP. The Golgi-resident dipeptidyl aminopeptidase *STE13* can remove this EAEA sequence,[85] but the efficiency of cleavage varies by molecule (leading to N-terminal variations on the recombinant protein).[51,86] As αSP processing can occur without the EAEA motif,[87] the four-amino acid sequence is often omitted.[88,89] This dual expression system allowed us to test variations in both overall and relative copy number (Figure 2.1A).

We analyzed the variations manifest in trastuzumab when secreted from *K. phaffii* following cultivation in flasks and purification by Protein A chromatography. Product-related variations in the drug substance of a biopharmaceutical can reduce productivity of the cell culture, increase the complexity and costs for purification, and may introduce risks for patients.[90] Variations of mAbs expressed by CHO cells can include *N*-terminal variations, modulated

glycosylation, sulfation, phosphorylation, hydroxylation, carboxylation, amidation, glycation, and misfolding, among others.[91,92] Apart from variants resulting from native yeast glycans, detailed assessments of other variations of mAbs secreted from *K. phaffii* have not been well-documented, however.

The purified trastuzumab showed many product-related variants (Figure 2.1B). To differentiate these variants, we fractionated the purified protein with size exclusion chromatography (Figure 2.1C) and evaluated the predominant fractions by intact mass spectrometry (Figure 2.1D). There were three major product-related variants. First, we observed high molecular weight (HMW) variants of both the trastuzumab heavy chain and light chain. Observed mass shifts for the HMW variants of light chain were consistent with *N*-terminal extensions resulting from incomplete cleavage of the proαSP (Figure A2.1). The identities of the HMW species of heavy chain were more difficult to determine, likely due to the presence of glycans. To facilitate our analysis, we treated the purified mAb with PNGase to cleave N-linked glycans and confirmed the presence of similar *N*-terminal extensions due to incomplete cleavage of the proαSP on the heavy chain (Figure 2.2A). This portion of the proαSP should be removed in the Golgi apparatus of *K. phaffii* by cleavage at a dibasic motif KR by *KEX2* protease.[85] The proαSP contains both N- and O-linked glycosylation sites and these modifications may also contribute to the formation of HMW species when the *N*-terminal extension is not efficiently removed. This observation is similar to ones we have reported previously for subunit protein vaccine antigens secreted from *K. phaffii*.[51]

Second, we observed a variant with a slightly higher apparent mass than the unmodified heavy chain by SDS-PAGE. Treatment with PNGase did not remove this variant, but incremental mass shifts of 161 Da are consistent with hexose addition. Furthermore, digestion with α1-2,3,6-

mannosidase removed this variant, suggesting that the product variant may contain O-glycosylation (Figure 2.2B). We attempted to identify potential sites of O-glycosylation on this variant by performing an in-gel tryptic digest and LC-MS/MS. We found only one tryptic peptide [STSGGTAALGCLVK] early in the IgG1 heavy chain constant region that appeared differentially O-glycosylated compared to the unmodified heavy chain (Figure A2.2). We tested several sequence variants that removed the Ser and Thr residues in this peptide, but we still observed O-glycosylated variants by SDS-PAGE (Figure A2.3), suggesting there may be other potential sites for O-linked mannosylation. The observed degree of O-mannosylation is low, and similar post-translational modification has been reported in mAbs produced by higher eukaryotes including CHO and COS cells as single-mannose additions on the light chain of IgG2.[93]

Third, we observed a cleaved heavy chain fragment. We performed intact LCMS and identified the variant as a *C*-terminal fragment of the trastuzumab heavy chain (Figure 2.2C). Interestingly, this fragment begins near a dibasic motif (K/KVE…) that may also be a cleavage target for *KEX2* protease. While *KEX2* cleavage canonically occurs at the *C*-terminal side of a dibasic motif, additional mass adducts or losses may confound the intact mass analysis.

A Vectors for genomic integration:

Ori
TT
Geneticin marker
P_{AOX1}
Signal peptide
Heavy chain
TT

Linkers for Gibson assembly

Ori
TT
Zeocin marker
P_{DAS2}
Signal peptide
Light chain
TT

B

◄ Heavy chain   ◁ Light chain

kDa
IgG Std.   Purified trastuzumab

130
100
70
55
40
35
25

High molecular weight species (*N*-terminal extension)
O-mannosylated
N-glycosylated
Aglycosylated
High molecular weight species (*N*-terminal extension)
N-glycosylated cleaved product
Aglycosylated cleaved product
Light chain

C

Fraction 1   Fraction 2

UV absorbance (mAU)

Aggregates
Fragments

Elution volume (mL)

D

Light chain | Cleaved heavy chain | Heavy chain

Full length
O-mannoses
+ *N*-term EEGVSLEKR
+ *N*-term AKEEGVSLEKR
+ *N*-term SIAAKEEGVSLEKR

+ N-glycosylation
Fc cleaved product

Full length
+ O-mannoses
+ putative *N*-term extension
+ N-glycosylation

Relative abundance

Deconvoluted mass (kDa)

Fraction 1
Fraction 2

27

Figure 2.1. Characterization of *K. phaffii*-secreted trastuzumab product-related variants. (A) Vector design for the expression of trastuzumab light chain and heavy chain. (B) SDS-PAGE analysis of purified trastuzumab secreted from *K. phaffii*, as compared to an IgG standard. (C) Size exclusion chromatography trace of purified trastuzumab. Two fractionations of interest are highlighted and analyzed with LC-MS. (D) Mass spectrometry spectra of Fractions 1 and 2. Product-related variants including N-terminal variants, Fc cleavage, and non-human glycosylation are noted.
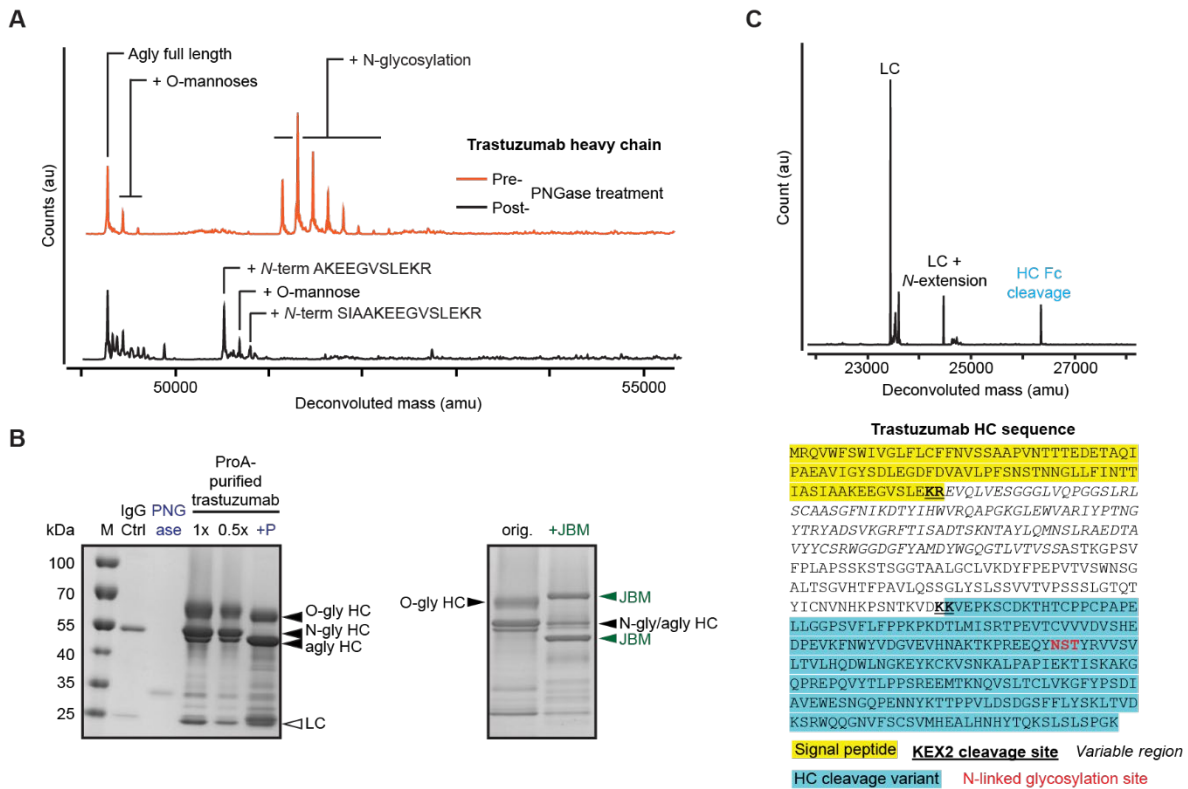
Figure 2.2. Characterization of glycosylation on *K. phaffii*-secreted trastuzumab.
(A) Mass spectra of trastuzumab heavy chain pre- and post-PNGase treatment, which removes
N-linked glycosylation. (B) SDS-PAGE analysis of purified trastuzumab with PNGase treatment
and/or α1-2,3,6-mannosidase (JBM) digestion. O-linked glycosylation is identified with JBM
digestion. (C) Mass spectrum of Fc cleavage, with heavy chain sequence analysis supporting the
existence of a Kex2 cleavage site in the conserved HC region.

2.2.2. Effects of signal peptide on mAb quality

Given our observations on incomplete cleavage of the signal peptide from the expressed heavy and light chains, we decided to test other signal peptides, including ones that direct native secretion of immunoglobins in other organisms, to assess how these could alter the N-terminal extensions observed. The signal peptide, which directs translocation of the polypeptide into the endoplasmic reticulum, can impact both secreted titer and quality, likely due to the difference in timing between translation and translocation and in subsequent processing in the yeast secretory pathway.[94] The most commonly used signal peptide for expression of recombinant proteins in *K. phaffii* is the signal peptide from the *Saccharomyces cerevisiae* alpha mating factor gene (αSP).[85] Previous reports of mAb expression in *K. phaffii* have suggested that alternative signal peptides from the *K. phaffii* genome or from other eukaryotic organisms may yield higher secreted titers of mAbs.[95] We first evaluated the expression of the trastuzumab light chain with seven additional signal peptides, and observed the highest secreted titers with αSP or the preOST1-proαSP. We then tested the expression of the trastuzumab heavy chain vectors with different signal peptides in strains containing expression cassettes for the light chain bearing these two signal peptides (αSP or preOST1-proαSP). Overall, we observed higher secreted titers of both chains when the light chain was expressed with the preOST1-proαSP (Figure 2.3A). The αSP, preOST1-proαSP and the signal peptide from human serum albumin (HSA-SP) yielded the highest secreted titers of the trastuzumab heavy chain (as measured by Protein A biolayer interferometry). These combinations yielded a 79-fold higher specific productivity compared to the murine immunoglobin signal peptide pre- region previously reported for expressing mAbs in *K. phaffii* (Figure 2.3B).[95] There was no significant improvement in product quality evident by SDS-

PAGE, though the titers for many of tested signal peptides were likely not high enough to observe the changes in product-related variants. These data together suggested the best constructs from this set of peptides tested still used the preOST1-proαSP signal sequence, and we chose to use this sequence for further engineering to improve product quality.
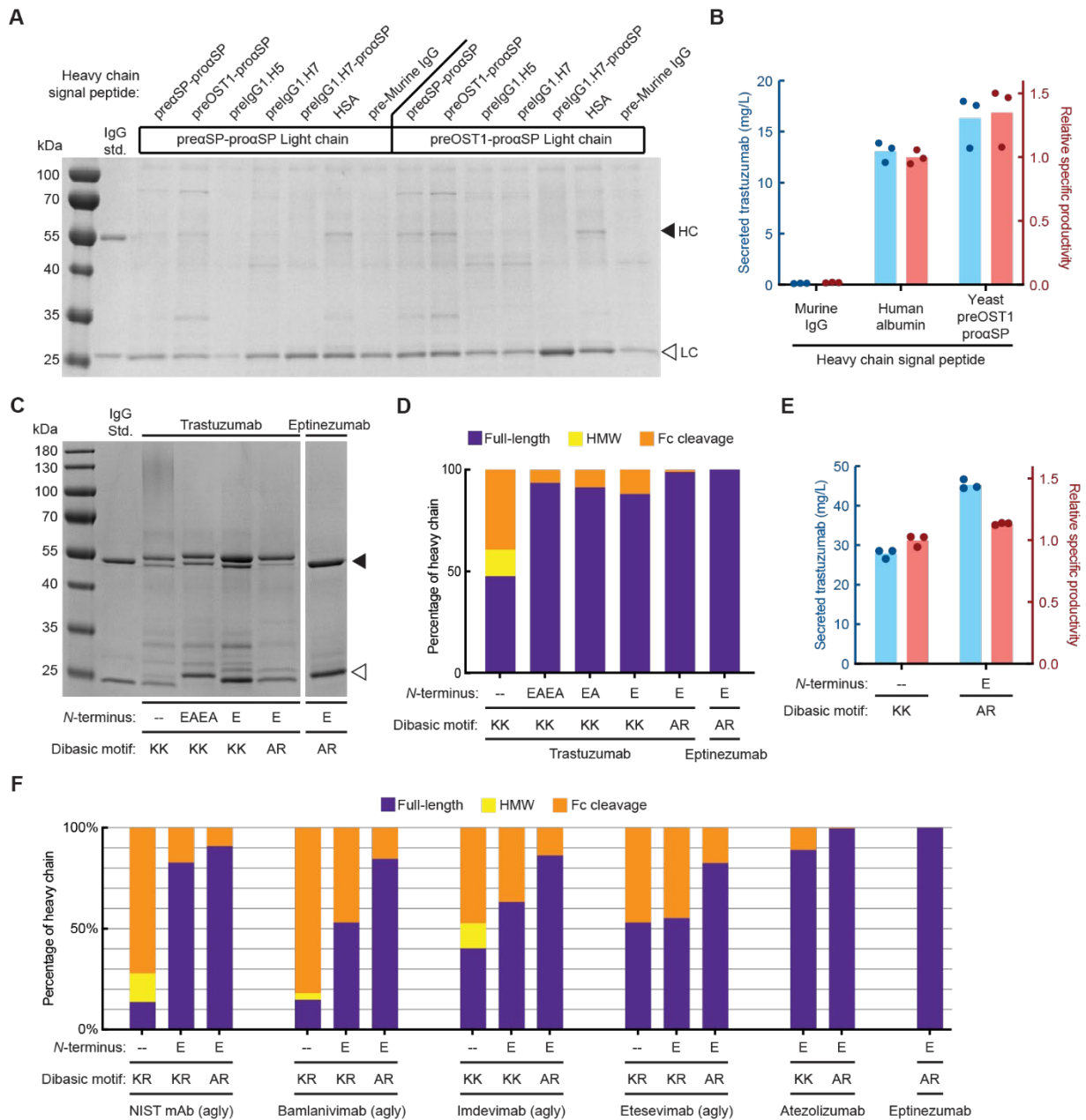
Figure 2.3. Conservation mutations in IgG backbone to improve product quality.
(A) SDS-PAGE analysis of trastuzumab secretion with different signal peptides. (B) Secretion titer of trastuzumab with select three signal peptides, measured by bio-layer interferometry. (C) SDS-PAGE, (D) LC-MS protein variant analysis, and (E) BLI titer measurement of engineered trastuzumab, as compared to eptinezumab. (F) LC-MS-based protein variant analysis of the heavy chains of different monoclonal antibodies.

2.2.3. Effects of integration copy number on mAb quality

We then examined the copy number ratio between the light chain and heavy chain vectors to explore its impact on the secretion of the mAb. It is typical to evaluate the secreted expression of several *K. phaffii* transformants to identify an optimal clone for production of a heterologous recombinant protein. In theory, transformants differ primarily by the copy number of the vector integrated into the host genome. Higher copy numbers can yield higher levels of the recombinant transcript, and in turn improve the secreted titer of simple recombinant proteins.[96] High transcript levels of complex proteins, however, may activate other competitive pathways such as endoplasmic reticulum associated degradation (ERAD) and the unfolded protein response (UPR).[97] Lower copy numbers may be beneficial, therefore, for more complex proteins. Previous studies have suggested that the light chain can properly fold and secrete without the heavy chain, but that the heavy chain has more complex folding requirements and requires the light chain for efficient folding and secretion.[98]

We transformed a vector encoding the heavy chain of trastuzumab modified to remove the site for N-linked glycosylation (N300A) into a base strain of *K. phaffii*. We selected three transformants that exhibited high, medium, or low growth on selection plates, as an approximation for high, medium, and low copy numbers of the heavy chain vector. We then transformed a vector encoding the trastuzumab light chain into each of these transformants and identified 6-8 transformants from each that exhibited a range of growth rates on selections plates, as an approximation for the copy number of the light chain vector. We evaluated the secreted expression of all transformants and observed a positive correlation between secretion of the light chain and secretion of the heavy chain (Figure A2.4). To test if this correlation was universal or specific to trastuzumab, we performed the same series of transformations with a second human

IgG1 mAb that is currently in preclinical development (mAb1). Interestingly, we observed a negative correlation between secretion of the mAb1 light chain and the mAb1 heavy chain. These results suggest that the optimal ratio of light chain copy number to heavy chain copy number depends on the nature of the recombinant mAb and the *K. phaffii* strain background. Although different copy number ratio between light chain and heavy chain yielded different titers, we did not observe any correlation between copy number ratio and product-related variants, based on SDS-PAGE.

2.2.4. Minimal, conservative changes to mAb sequence for quality improvement

We then examined if product-related variants of trastuzumab could be reduced by modifying the sequence. We and others have shown that small, conservative changes to the amino acid sequence of therapeutic proteins can have large impacts on quality and manufacturability.[51,52] Sequence engineering has also improved the quality and manufacturability of mAbs.[99] We previously reported that addition of amino acid residues to the *N*-terminus of a recombinant protein can reduce *N*-terminal extensions, likely by increasing the steric accessibility of the *KEX2* cleavage site.[100] We hypothesized that this strategy may also reduce *N*-terminal extension of the light chain and heavy chain of trastuzumab. We expressed and purified trastuzumab with EAEA, EA, or E residues added to the *N*-terminus of both the heavy chain and light chain (Figure 2.3C). We observed nearly complete elimination of *N*-terminal extension with EAEA addition, and significant reduction thereof with the addition of a singular E residue. Interestingly, we also observed reduction of the O-glycosylated heavy chain variant and the cleaved *C*-terminal heavy chain fragment. We analyzed each purified sample by intact LCMS and observed that residues attached to the *N*-terminus of the heavy chain and light chain

remained on the purified protein (Figure A2.5). This observation is consistent with our previous observations for several subunit vaccine antigens.[100]

Next, we sought to further reduce or eliminate the *C*-terminal heavy chain fragment. We analyzed the sequence of the recently approved mAb for headache treatment, eptinezumab (Vyepti®), that is manufactured in *K. phaffii*.[78] Eptinezumab has two sequence modifications (K207A, K208R) that removed the dibasic motif in the heavy chain constant region that may be the cause of *C*-terminal cleavage product observed for trastuzumab. We expressed eptinezumab in *K. phaffii* using preOST1-proαSP with a single E residue on the *N*-terminus of both the light chain and heavy chain and purified the modified eptinezumab by ProA chromatography. We did not observe any *C*-terminal heavy chain fragment (Figure 2.3C).

To evaluate if this change could confer the same benefit to trastuzumab, we created a variant with the same mutations. We quantified the relative abundance of the two major heavy chain variants by LCMS and observed a large reduction in the abundance of the *C*-terminal heavy chain fragment – the detected fraction of cleaved heavy chain dropped from ~11% to <1% with the KK→AR mutations (Figure 2.3D). We also evaluated the secreted titer of trastuzumab with and without both sequence modifications and observed that the modified trastuzumab molecule was secreted with slightly increased specific productivity (Figure 2.3E). The strain secreting the modified trastuzumab also grew to a higher cell density, which resulted in a ~60% increase in secreted titer. This result suggests that reduction or elimination of product-related variants also improved cellular processing and secretion of the mAb, likely because misfolded or incorrectly modified mAb molecules may trigger degradation pathways or stress responses in the yeast secretory pathway, reducing the overall secreted titer.[101]

We then tested the benefits of these changes for five additional IgG1 sequences, including NIST mAb, bamlanivimab, imdevimab, etesevimab, and atezolizumab. We mutated the conserved N-linked glycosylation site with N-to-A mutation in the heavy chain Fc region in NIST mAb, bamlanivimab, imdevimab, and etesevimab to reduce product quality variability due to glycosylation. Analysis with LCMS showed improvement in the fraction of full-length heavy chain in all mAbs tested (Figure 2.3F). Overall, these results demonstrate that small modifications to the human IgG1 backbone can greatly improve both the quality and titer of mAbs secreted from *K. phaffii*.

2.2.5. Comparability of engineered mAbs secreted from yeast and CHO

One potential use for yeast-based expression of aglycosylated mAbs would be to support rapid production for first-in-human clinical studies.[10] It is feasible to consider changing the host used for further production or development of such mAbs since the analytical characterization of the resulting protein could support comparability of the two sources.[102] To evaluate if these modifications to IgG1 backbone are transferrable to other expression systems, we compared the secreted aglycosylated NIST mAb with and without sequence engineering to those secreted by CHO. We saw a significant amount of N-terminal extension and heavy chain Fc cleavage in the unmodified yeast-secreted product by SDS-PAGE (Figure 2.4A). N-terminal E-addition and the dibasic site mutation essentially eliminated these product-related variants. Aglycosylated NIST mAbs secreted from CHO, both unmodified and engineered, showed no N-terminal extension or heavy chain cleavage, but they did exhibit C-terminal lysine clipping, a known post-translational modification by endogenous carboxypeptidases during CHO cultivation (Figure 2.4B).[103] We then compared the secondary structures of both the yeast- and CHO-secreted products. Far-ultraviolet circular dichroism spectroscopy showed converging confirmational profiles among

both CHO-secreted molecules and the engineered yeast-secreted product. Purified un-engineered yeast product showed a different CD profile, likely due to N-terminal extension and heavy chain cleavage (Figure 2.4C). These data show that the modified IgG1 sequence can be expressed in both *K. phaffii* and CHO cells with highly similar structures and minimal differences.

Figure 2.4. Comparability of engineered mAbs secreted from *K. phaffii* and CHO.
Comparison of secreted aglycosylated NISTmAb from both hosts in (A) SDS-PAGE analysis,
(B) LC-MS, and (C) circular dichroism.
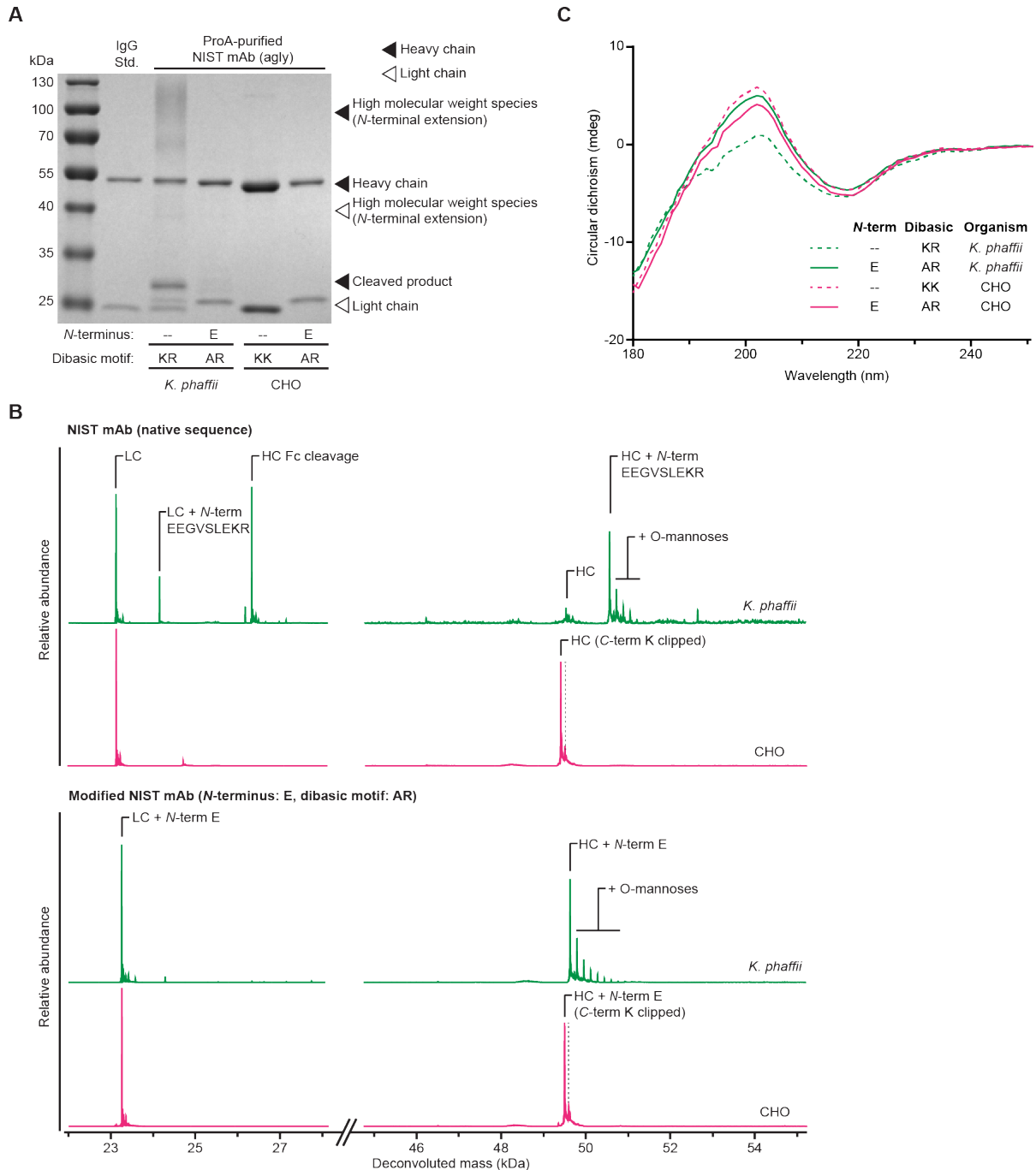
2.3. Discussion

In this chapter, we developed a modular approach to express aglycosylated antibodies in *K. phaffii*. By using a dual integration system with our customized vectors, we secreted full-length trastuzumab, but several apparent product-related variants were evident with the original sequence. After characterizing the variants and identifying them as an N-terminal extension and heavy chain Fc cleavage, we modified the amino acid sequence for the human IgG1 in two places to eliminate most product-related variants after affinity purification with Protein A. These sequence changes resulted in higher titers of trastuzumab in flasks, likely due to the reduction of variants. These engineering changes also improved the quality of five other aglycosylated variants of IgG1's used as biopharmaceuticals when expressed in *K. phaffii*, greatly reducing N-terminal signal peptide extension and Fc cleavage.

These conservative modifications have prior clinical precedents that reduce their potential risks for clinical or commercial use. The modification of the dibasic site motif has been used in Vyepti® (eptinezumab). Across IgG subclasses, this motif presents as KK in IgG1, KR in IgG3 and IgG4, and KT in IgG2.[104] The clinical precedent and the intrinsic variability demonstrate tolerance for mutations at this motif. The beneficial addition of a single N-terminal amino acid is similar to the intrinsic modification found on recombinant proteins produced by bacteria such as *E. coli*. Due to the absence of a secretory pathway (and thus no secretion signal peptide processing) in prokaryotic hosts, an extra N-terminal methionine (which initiates recombinant protein translation) is present in the final drug product.[105] Myalept® (met-leptin), Neupogen® (met-G-CSF), and Protropin® (met-hGH, with second generation product Nutropin®, hGH) are among commercial recombinant methionyl pharmaceutical proteins.[1] The N-terminal methionine from bacteria-produced proteins can be cleaved by peptidases, but residual variants with single

amino acid N-terminal extensions have been present in biologic medicines. We have previously shown that the identity of this single amino acid can also vary – including methionine – to give similar reductions of N-terminal variations resulting from incomplete signal peptide cleavage.[100]

With these combined conservative changes, we showed the quality of secreted aglycosylated mAb products were highly comparable between yeast and CHO, the industry standard. In fact, yeast-secreted NIST mAb did not exhibit any C-terminal lysine clipping, a common product-related variant in CHO cultures.[106] The engineered, aglycosylated NIST mAb showed similar secondary structures as the unmodified NIST mAb expressed by CHO, indicating that these minimal sequence changes did not significantly alter the overall protein conformation. The *K. phaffii*-secreted product did exhibit minor O-linked mannosylation. Similar mannosylated variants have been reported to be present in clinical-stage drug products, with no significant impact to biological activity.[107] The impact of O-linked mannosylation differs by molecule, but single-mannose addition has been shown to not induce immunogenic reaction,[108] as higher eukaryotes could have similar modifications.[93] Nevertheless, further characterization and engineering could be pursued to minimize this post-translational modification common to yeast. The feasibility of O-linked glycoengineering has been demonstrated,[109] and the homogeneity of resulting O-glycome could be advantageous for "biobetter" development.[110]

We demonstrated the ability to adapt aglycosylated monoclonal antibodies for high-quality production in *K. phaffii*. The comparability between yeast- and CHO-secreted products provides additional context to advance alternative hosts for mAb production. A widespread transition in commercial manufacturing of mAbs in yeast remains underdeveloped, but the characterizations and engineering presented here provide a new framework for advancing rapid

development cycles for new biologics and for global expansion of production for low-cost, high-quality mAbs.

2.4. Methods

Yeast vectors and strains

The custom vector was constructed by synthesis of DNA fragments (IDT) and Gibson assembly (New England Biolabs). Genes containing mAb light and heavy chains were codon optimized, synthesized (Integrated DNA Technologies), and cloned into the custom vector. Modification of the vector, including replacement of signal peptides and addition of mutations to the IgG1 sequence were performed using PCR and site-directed mutagenesis (NEB). All yeast strains were derived from wild-type *Komagataella phaffii* (NRRL Y-11430). After initial screening of signal peptides, all strains were derived from a modified base strain (AltHost Research Consortium Strain S-63 (RCR2_D196E, RVB1_K8E)) described previously.[111] *K. phaffii* strains were transformed as described previously.[112]

Yeast cultivations

Strains for initial characterization and titer measurement were grown in 3 mL culture in 24-well deep well plates (25°C, 600 rpm), and strains for protein purification were grown in 100 mL culture in 500 mL shake flasks (25°C, 300 rpm). Cells were cultivated in complex media (potassium phosphate buffer pH 6.5, 1.34% nitrogen base w/o amino acids, 1% yeast extract, 2% peptone). Cells were inoculated at 0.1 OD600, outgrown for 24 h with 4% glycerol feed, pelleted, and resuspended in fresh media with 1% methanol and 40 g/L sorbitol to induce recombinant gene expression. Supernatant samples were collected after 24 h of production, filtered, and analyzed.

CHO vectors and strains

The heavy and light chain coding sequences for NISTmAb (with or without the described mutations) were codon optimized[113] for expression in *C. griseus* and inserted via restriction digestion into two otherwise identical expression plasmids that differ only in their metabolic selection markers. Each respective pair of expression plasmids was co-transfected into Chinese Hamster Ovary host cells, previously described elsewhere.[114] Selective pressure was applied twenty-four hours post-transfection through a complete media exchange into chemically defined medium lacking critical metabolites. Transfectant pools were cultured at 36°C and 5% CO_2 and resuspended into fresh selection medium every three to seven days until the cell pools returned above 90% viability.

CHO fed-batch material generation

CHO-derived NISTmAb was produced in a 14-day fed-batch process. Recovered transfectant pools were seeded at a target density of $1x10^6$ cells/mL into a proprietary chemically defined production medium[115] with an initial working volume of 100 mL in a 500 mL flask. Fed-batch cultures were cultivated under shaking conditions at 5% CO_2 and 35°C until day 7, when the temperature was lowered to 31°C. Beginning on day two, daily fixed volumes of a proprietary chemically defined complex feed were added to sustain the culture duration. On day fourteen, cell culture supernatants were harvested by centrifugation and filtration before final storage at -70°C.

Protein purification

The filtered supernatant samples were diluted, 1:1 with 1x phosphate-buffered saline (PBS, pH 7.4) for 3-mL culture or 10:1 with 10x PBS for 100-mL culture, and purified on the GE ÄKTA pure system with a 1-mL ProA column (HiTrap Protein A HP, Cytiva). After sample

loading, the column was equilibrated with and washed with 1x PBS and eluted with 100mM

citric acid (pH 2.8). Eluted products were pH adjusted with 1M Tris-HCl (pH 9.0).

Analytical assays for protein characterization

Purified protein concentrations were determined by absorbance at A280 nm. SDS-PAGE

was carried out as described previously.[74] Supernatant titers were measured by Protein A biolayer

interferometry. Specific productivity was defined as relative titer normalized by cell density,

measured by OD600. Size exclusion chromatography, intact mass spectrometry, and far-

ultraviolet circular dichroism were performed as described previously.[52]

3. APPLYING CRISPR-CAS9 TECHNOLOGY FOR GENOME EDITING IN *K. PHAFFII*

Protein sequence modifications, albeit an attractive option for improving manufacturability, are not always a viable strategy. Engineering the host cell is thus an alternative strategy for improving product quality in a minimally invasive manner, and advances in genetic manipulation techniques have significantly increased the speed and precision with which this can be achieved. With CRISPR-Cas9 gene editing technology, we demonstrate in this chapter its multiple applications, especially multiplexed humanization of *K. phaffii* glycosylation pathway. We also show tunable regulation of cell growth, based on a process-friendly native promoter identified with RNA sequencing.

3.1. Background and Motivation

Model microorganisms like *E. coli* and *S. cerevisiae* are commonly used for manufacturing complex molecules. Recently, interest has risen in the development of non-model microorganisms, including bacteria, yeast, and filamentous fungi as hosts for biomanufacturing and chemical processing, owing to specialized phenotypes like unique metabolic chemistries or high capacity for pathway engineering.[116] The lack of broad gene editing tools, however, impedes the agile development of new potential hosts for producing high-value molecules.[117] *K. phaffii* falls under the category of these less developed, alternative hosts. To date, efforts to elucidate and engineer complex phenotypes have required excessive trial and error using slow, inefficient methods of genetic disruption. Despite the time- and resource-intensive process, genetic engineering of *K. phaffii* has proven to be a widely utilized strategy for improving the quality of secreted proteins. Multiple auxotrophic strains have been developed by knocking out genes encoding crucial metabolic function,[118] and they have seen utilization in the humanization of glycosylation pathway.[119] Flippase recombination is also used to recycle auxotrophic or antibiotic markers, and protease-deficient strains can be created to reduce unwanted proteolytic activity on target recombinant proteins.[120]

Homologous recombination, however, is not efficient in *K. phaffii* – unlike *S. cerevisiae*, in which recombination could be accomplished with only 40 base-pair homology, allowing for the construction of genetic perturbation cassettes by simple polymerase chain reaction (PCR) – and occurs at a much lower frequency than non-homologous end joining (NHEJ).[120] To enhance the efficiency of recombination, CRISPR-Cas9 system can be utilized to induce targeted double-stranded break, putting additional selection pressure on cells to adopt homology-directed DNA repair.[121] This genome editing system has thus become the technique of choice for both

functional genetics and genome engineering in microbial hosts, enabling parallel, targeted genomic disruptions without genomic integration of selection markers or additional genome scarring.[122] In the CHO space, examples include the construction of targeted integration sites for heterologous protein expression to reduce drifts in product quality compared to random integration and glycosyltransferase knockout screen for increased glycan homogeneity.[123] Similar systems have been adapted to *K. phaffii*, but their use cases have largely been limited to simple knockouts and fluorescent reporter protein expression.[124] More recent literature have reported the engineering of metabolic pathways in *K. phaffii* for secondary metabolite production.[125,126]

We demonstrate in this chapter a CRISPR-Cas9 genome editing strategy in *K. phaffii* and the utilization thereof to start addressing product quality issues identified in Chapter 2. More specifically, we describe the engineering of *K. phaffii* native glycosylation pathway through multiplexed genome editing. This paves the way to further explore glycosylation humanization in Chapter 4 and that of more complex, therapeutically relevant proteins in Chapter 5. We also discuss our search for tunable native promoters that can be controlled in a process-friendly manner and show the potential of this system to control heterologous pathways. Lastly, CRISPR-Cas9 technology is employed to upregulate a peptidase to examine its impact on protein processing.

3.2. Results

3.2.1. Multiplexed editing of *K. phaffii* glycosylation pathway

Our lab has developed a simple, highly efficient CRISPR-Cas9 system that, unlike prior work, which has relied on RNA polymerase II for the expression of sgRNA,[124] uses native RNA polymerase III system and the associated tRNA expression. The system can generate knockouts of *K. phaffii* genes with up to 100% efficiency.[112] Because several tRNA's have proven to be

equally highly effective in generating knockouts, we theorized that these tRNA-sgRNA cassettes could function orthogonally and enable multiplexed genome engineering. The yeast glycosylation pathway is a prime candidate for such multiplexed engineering, because it is a multi-step pathway that requires as many as 18 different genetic modifications for terminal sialylation.[54] This was largely achieved via recycling auxotrophic markers, which leaves genome scarring that could lead to altered metabolic activity and genome instability. CRISPR-Cas9 technology circumvents this issue because it does not depend on the genomic integration of selection markers for editing and thus leaves no scars. It has also been reported that the expression of these heterologous glycosylation pathway enzymes and the knockouts of native mannosyltransferases created growth defects in engineered strains,[127] so speed is an additional advantage of multiplexed editing.

We set our engineering target to be the production of GlcNAcMan5 glycan on the reporter peptide K3, a highly soluble peptide derived from the Kringle 3 domain of human plasminogen with one single highly accessible N-linked glycosylation site. This peptide has also been used in previous glycoengineering efforts as the reporter molecule.[128] Three heterologous enzymes are needed to produce the desired glycan – mannosidase I (from *Caenorhabditis elegans*), UDP-GlcNAc transporter (from *Kluyveromyces lactis*), and UDP-GlcNAc transferase (from *Homo sapiens*), denoted as *MNS1*, *MNN2* (also known as *YEA4*), and *GNT1*, respectively. We elected to express these heterologous genes under methanol-inducible promoter P_{AOX1} to minimize the impact of their expression on biomass accumulation.

We first knocked out *K. phaffii* native *OCH1* and observed slow growth. We then constructed an RNA Pol-III cassette to express three sgRNAs for the simultaneous integration of *MNS1*, *MNNS2*, and *GNT1*. The feasibility for multiplex gRNA cassettes to be expressed under a

single RNA Pol-III has been demonstrated in *S. cerevisiae*.[129] We thus constructed the multiplexed plasmid and cotransformed it with linear DNA integration cassettes for the heterologous genes into wildtype, *Δoch1*, and *Δku70* cells (Figure 3.1A). *KU70* is a conserved gene across species responsible for NHEJ, and its knockout has been reported to enhance the efficiency of homology-directed repair (HDR).[130] We observed that complete multiplexing was in achieved in all three strains with efficiency of up to 40%. Among all loci screened, nearly 50% of possible edits were successful, with only small frequencies of indels, off-target integrations, or mixed colonies (Figure 3.1B). Interestingly, *Δku70* genotype did not perform better than wildtype in integration efficiency. We thus decided to not move forward with characterizing the *Δku70* strain because it should not perform any differently from strains with intact *KU70*.

We therefore transformed K3 peptide into a wildtype strain, a Δ*och1* strain, and the glycoengineered strain, cultivated clones in batch culture, and evaluated the glycosylation pattern on secreted K3 by intact protein mass spectrometry. The glycoengineered strain produced K3 with uniform glycosylation, markedly different from the yeast's native, heterogeneous hypermannosylation (Figure 3.1C). Analysis by LC-MS and treatment with PNGase revealed a nearly uniform, modified glycan structure (Figure 3.1D). These data show the successful engineering of a complex phenotype of interest in a specialized microorganism in only two steps. Equivalent modifications with traditional tools would require four transformations, using either four orthogonal antibiotic or auxotrophic markers, or through recycling of a subset of these markers, which would even more prolong the engineering timeline.
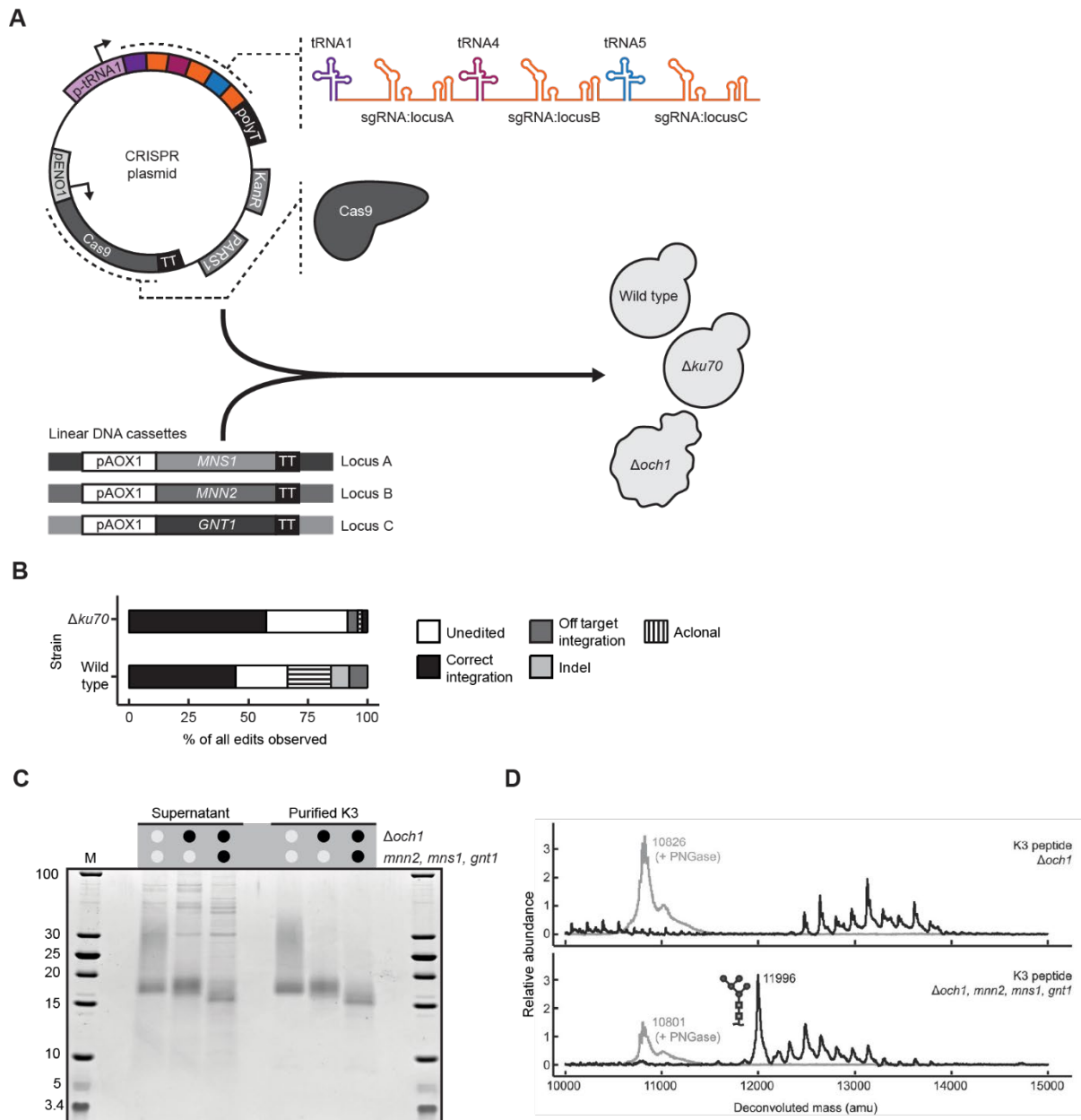
Figure 3.1. Multiplexed engineering of *K. phaffii* glycosylation pathway.
(A) Schematic for co-transformation of a CRISPR-Cas9 plasmid carrying three single-stranded guide RNAs and linear DNA fragments for simultaneous genome editing. (B) Screening of multiplexed engineering in both Δ*ku70* and wildtype strains. (C) SDS-PAGE analysis of secreted K3 peptide in different strain backgrounds. (D) LC-MS analysis of K3 peptide in Δ*och1* or glycoengineered strain with and without PNGase digestion. Man5 glycan was the predominant glycosylation in glycoengineered strain.

3.2.2. Identification of *K. phaffii* native, process-friendly promoters

Impaired growth rate, a direct result from Δ*och1*, is an undesirable phenotype for a host like *P. pastoris*, which prides itself on its ability to rapidly accumulate cell weight during cultivations.[131] To circumvent the detrimental effect of *OCH1* knockout, a genetic switch that induces *OCH1* expression during biomass accumulation phase but represses it during protein production phase would be highly advantageous. The most common strategy for tuning gene expression is through altering transcript levels, typically achieved by promoter engineering.[132] Unlike in *S. cerevisiae*, synthetic biology toolkit development in *P. pastoris* has been slow. This is partially because *S. cerevisiae* is the fungal host of choice for producing chemicals via metabolic engineering, a field that has seen dramatic advancement over the past two decades.[133,134] Genetic elements including promoters, terminators, and riboswitches are developed, characterized, and utilized to optimize metabolic flux for the production of desired products, while genome engineering tools such as CRISPR-Cas9 have been widely employed to facilitate this process.[135–139] In comparison, *P. pastoris* has been used mainly for protein production, and its synthetic biology toolbox, as a result, is largely limited to and tailored for this purpose only.[140,141]

As a methylotrophic host, *P. pastoris* is capable of oxidizing methanol for energy production and cell growth.[140] *AOX1* gene in *P. pastoris* codes for alcohol oxidase I, which catalyzes the first step of methanol metabolism. Its gene expression is repressed in glucose or glycerol media and strongly induced in methanol media. For this reason, the tightly regulated P_{AOX1} has gained wide popularity for recombinant protein production.[142] A library of P_{AOX1} variants of different strengths has been generated for fine-tuning gene expression to improve yield and quality of heterologous proteins.[143] Other methanol inducible promoters, including

dihydroxyacetone synthase II promoter (P_{DAS2}) and formaldehyde dehydrogenase I promoter (P_{FLD1}), are also used in heterologous protein production.[144] Because of the safety concerns with methanol flammability in large-scale fermentation, novel promoters and promoter systems have been developed, either using other inducing chemicals such as ethanol or through transcriptional factor engineering.[145–148] These native carbon source dependent promoters are favorable because they allow for the separation of biomass accumulation from protein production and because they utilize the native regulation mechanisms, which have evolved to support high cell viability.[149]

Inducible promoters such as P_{AOX1} are suitable for protein production, but successful *OCH1* control requires an opposite regulation scheme. To date, only a handful of repressible promoters are characterized in *P. pastoris*. Delic *et al.* reported five repressible promoters in *P. pastoris* and used P_{THI11}, a promoter in thiamine metabolism and the promoter with the largest dynamic range out of the five characterized, to achieve a ten-fold repression with the addition of thiamine into the culture media.[150] This promoter was later used for dynamic control of native ergosterol biosynthesis for secondary metabolite production, demonstrating a practical application of conditional promoters in *P. pastoris*.[151]

The native conditional promoter discovery in *P. pastoris* is typically achieved either through the understanding of native metabolism or through gene homology to another organism, in this case usually *S. cerevisiae*.[145,150,152] Even with high-throughput methods such as DNA microarray, as reported in more recent publications, the number of conditional promoters is still limited.[140,150,153] With the decreasing cost of high-throughput sequencing, RNA-sequencing, or RNA-Seq, has become the preferred method to characterize the host transcriptome. Compared to other methods such as DNA microarray, RNA-Seq has higher sensitivity and larger dynamic

range.[154] This technology has been used in *P. pastoris* to define the transcriptomic landscapes in common media conditions.[83,155]

Our group had previously collected RNA-Seq transcriptomic data of wild-type *P. pastoris* in more than 40 different media conditions by using different carbon sources and doping various additives in glycerol media. The selected additives were largely limited to sugars and common vitamins to keep discovered promoters process-friendly and not significantly raise the cost of media. Glycerol was chosen to be the basis of comparison to other carbon sources and additives, because it has high utilization rate and biomass yield, making it particularly useful for culture outgrowth.[140] Differential expression analyses were carried out on this dataset to identify promoter candidates for further characterization (Figure 3.2A). Through this process, we were able to identify promoters whose downstream genes show significantly up or down regulation in different media conditions compared to glycerol media.

We elected to study one particular promoter P_{INO1} for its potential in regulating gene expression. This promoter was selected because the *INO1* gene showed an over 30-fold down-regulation in glycerol media containing inositol, the strongest down-regulation observed in all media conditions with additives. With CRISPR/Cas9, we replaced the promoter for *GUT1*, encoding glycerol kinase, which catalyzes the first step of glycerol metabolism,[156] with P_{INO1} in a *K. phaffii* Δ*ku70* strain. We theorized that its repression would significantly reduce the cell growth rate on glycerol. By plating serial dilutions of Δ*ku70* strains, both unedited and ΔP_{GUT1}::P_{INO1}, on minimal glycerol agar plates, we observed decreasing growth rate of ΔP_{GUT1}::P_{INO1} strain with increasing inositol concentration, while the presence of inositol had no noticeable effect on the unedited strain. Full repression of *GUT1* expression was achieved at 0.5 mM inositol concentration (Figure 3.2B).

Motivated by the successful repression of *GUT1* by P_{INO1}, we attempted to demonstrate glycosylation control with the same promoter. We replaced the promoter region for *OCH1* with P_{INO1}, and plated serial dilution of this strain on minimal glycerol agar plates with different inositol concentration. We, however, were not able to see any significant change in growth after incubating the plates at 30°C for multiple days. A closer look at the expression level of *GUT1*, *INO1*, and *OCH1* in wild-type *K. phaffii* in glycerol media with and without inositol revealed the cause of the lack of response to inositol. The level of *GUT1* transcripts is on the same order of magnitude as that of *INO1* transcripts when *INO1* expression is induced, while *OCH1* is expressed at a much lower level on par with *INO1* expression level when *INO1* is repressed (Figure 3.2C). Thus, even when *OCH1* expression in the $\Delta P_{OCH1}::P_{INO1}$ strain is fully repressed, its expression level is still comparable to that in wild-type, resulting the absence of significant change in growth on agar plates supplemented with inositol.

A possible solution remains to be explored – engineering the Kozak sequence of P_{INO1}. A Kozak sequence is the stretch of DNA nucleotides that extends approximately from position -6 to +6, where position +1 is the start of gene coding sequence. In *S. cerevisiae*, varying Kozak sequence has been shown to change translational efficiency.[157] Our lab has demonstrated a more than ten-fold change in yeast enhanced green fluorescent protein (yEGFP) expression via varying the promoter Kozak sequence.[158] Varying Kozak sequence is unlikely to affect the regulatory elements of the *INO1* promoter, because Kozak sequence is only the few base pairs around the start codon, while regulatory elements are usually further upstream.
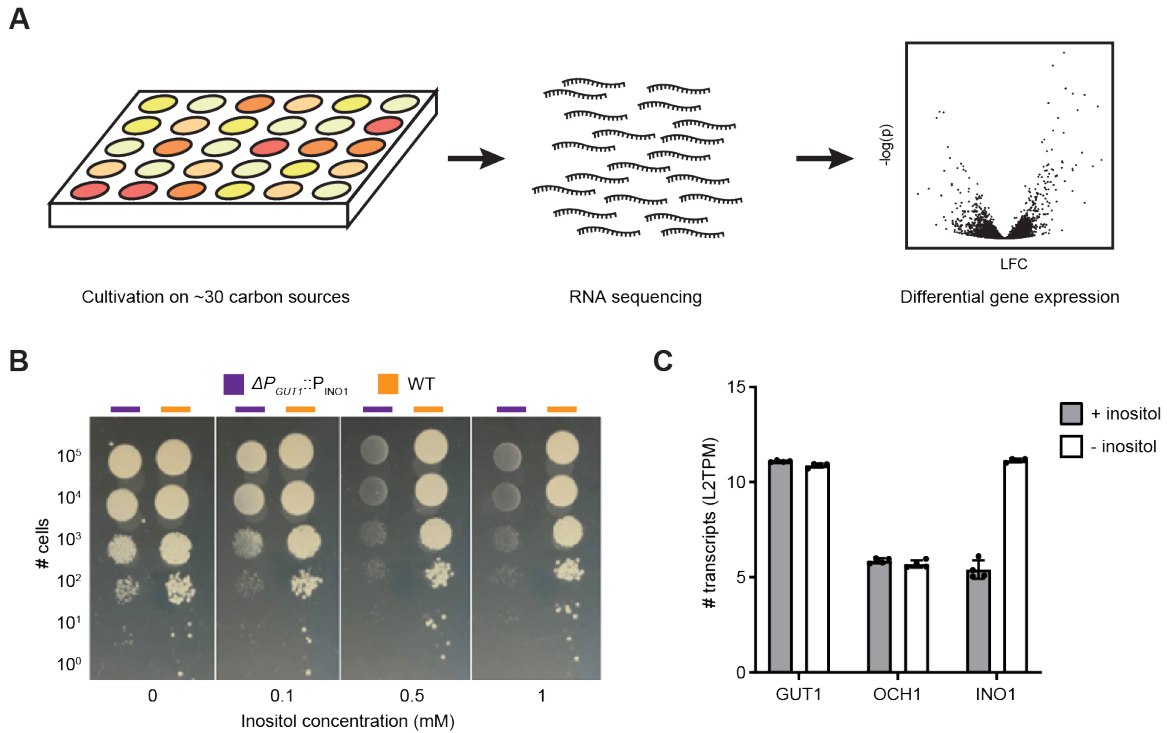
Figure 3.2. Control of cell growth on minimal glycerol media with native tunable promoter.
(A) Process schematic for the identification of native differentially regulated genes in different
media conditions. (B) Suppression of cell growth on minimal glycerol media with inositol by
replacing native *GUT1* promoter with P_{INO1}. Full suppression is achieved at 0.5 mM inositol
concentration. (C) Expression levels of several genes of interest, including *GUT1*, *OCH1*, and
INO1. P_{INO1} is a good candidate for regulating *GUT1* expression, but a poor one for *OCH1*
regulation.

3.2.3. Upregulation of a peptidase for protein processing

In Chapter 2.2.1, we discussed the issue of N-terminal extension from the deletion of Glu-Ala-Glu-Ala (EAEA) sequence found at the end of the native αSP. In our dual expression system for aglycosylated mAbs, we used a hybrid secretion signal comprising pre-OST1 and pro-αSP (without this EAEA motif). The Golgi-resident dipeptidyl aminopeptidase *STE13* is responsible for the removal of this sequence,[85] but the efficiency of cleavage varies by molecule (leading to N-terminal variations on the recombinant protein).[51,86] Although we were able to demonstrate N-terminal glutamate addition could significantly improve signal peptide processing, an alternative strategy could be upregulation of *STE13*, which, in theory, improves the processing of EAEA at the C-terminus of αSP.

We constructed three different *STE13* expression cassettes with methanol-inducible promoters, upstream of genes *GQ67_02654*, *GQ67_03437*, and *DAS2*. We chose these promoters based on their native gene expression in media with glycerol versus methanol as the sole carbon source (Figure A3.1). All three promoters have similar basal gene expression in glycerol media but different induced expression in methanol media. We then used CRISPR-Cas9 system to integrate these cassettes into the *K. phaffii* genome as an extra copy of *STE13* to minimize any phenotypic disruption during biomass accumulation phase in glycerol. For molecular demonstration, we chose three different molecules: (1) bovine serum albumin (BSA), (2) a SARS-CoV2 receptor binding domain (RBD) variant, and (3) the same SARS-CoV2 RBD variant with a SpyTag[159] at its N-terminal. These proteins were chosen because we have previously seen the quality difference before and after appending EAEA at the N-termini (BSA, internal unpublished data, and RBD[100]); on the other hand, because of the flexibility of SpyTag, the EAEA sequence on αSP was properly processed by *STE13*.[100] However, when we

transformed these recombinant proteins with and without EAEA addition at the N-termini, we did not observe any secretion of recombinant proteins by SDS-PAGE, despite strong secretion of these proteins in unmodified strains. We hence hypothesize that the overexpression of *STE13* negatively impacts protein secretion, possibly due to the disruption in late endosome trafficking, but the exact molecular mechanism remains to be studied.

3.3. Discussion

In this chapter, we reported the application of a simplified, multiplexed method for expression of sgRNAs in *K. phaffii*. The rapid engineering of glycosylation pathway can also be applied to a broad range of applications, including introduction of mammalian chaperones to enhance folding of complex molecules, generation of protease-deficient strains to improve yields of full-length product,[23] reduction and redirection of vacuolar and endoplasmic reticulum-associated protein degradation pathways,[160] enhancement of lipid synthesis and vesicular machinery,[161,162] and introduction of novel metabolic pathways.[163] It is worth noting that since we achieved the integration of three heterologous glycosylation genes, *MNS1*, *MNN2*, and *GNT1*, the expected predominant glycan structure on the secreted K3 peptide should be GlcNAcMan5, but we were only able to achieve Man5 glycosylation, and a significant amount of high-mannose structure persists on the secreted peptide. One possible reason is due to the induced co-expression of this humanized pathway and recombinant protein. As a result, there is no pre-existing machineries in place to properly glycosylate K3 peptide once its expression is induced by methanol. Chapter 4 will explore the constitutive expression of the heterologous glycosylation pathway for the homogeneous glycosylation of K3.

We also applied the CRISPR-Cas9 technology to two different engineering targets. First, RNA sequencing helped guide the identification of differentially expressed genes under process

friendly and economically viable control strategies. We achieved inositol-controlled growth rate modulation with a such promoter, and this promoter can be further engineered (through Kozak sequence mutations, for example) to regulate a wide range of gene expression. Second, we aimed to target a peptidase of interest for upregulation to examine its impact on protein processing, in hopes that some of the protein quality improvements achieved in Chapter 2 (via sequence engineering) can be attained by genetic engineering. The upregulation of this peptidase *STE13*, however, hindered protein expression. To our knowledge, the upregulation of *STE13* is poorly studied – the only phenotypic response being from a large-scale survey in *S. cerevisiae*, where it slows down vegetative growth[164] – and its impact on secretion even less so. This highlights the need for better mechanistic understanding of protein secretion in this alternative host and how it interacts with other cellular systems, possibly by adopting similar genome-wide gene perturbation survey coupled with high-throughput readout for protein secretion titer and quality.

3.4. Methods

CRISPR Plasmid construction

Guide RNAs targeting specific genome loci were designed using the ATUM gRNA Designer tool (www.atum.bio). 60-bp single stranded DNA fragments were synthesized (Azenta), containing 20-bp gRNA sequence and 20-bp homology on either side, and cloned into a Cas9-expressing plasmid via NEB Hifi Assembly. The propagation of plasmids was carried out in DH5α *E. coli* cells (NEB).

Gene knockout

Competent cells were prepared as described elsewhere[165] and transformed with 250 ng of circular plasmid DNA. Cells were allowed to recover in 1:1 YPD:1M sorbitol (Sigma-Aldrich) for 3h at 30°C without shaking before plating on YPD agar (BD Difco) plates supplemented with

200 µg/mL G418 Geneticin (Thermo Fisher Scientific). Plates were incubated for 2 days at 30°C before colonies were picked at random for genomic DNA extraction. DNA was extracted as previously described[166] and purified using the MagJET gDNA Kit (Thermo Fisher Scientific). The knockout locus of interest was then PCR amplified and Sanger sequenced (Azenta or Quintara Biosciences) to confirm successful knockout.

Multiplexed engineering

Glycosylation genes *MNS1*, *MNN2*, and *GNT1* were targeted to intragenic regions of the genome adjacent to genes *GQ67_04576*, *PFK1*, and *ROX1*, respectively using 500 bp flanking sequences for homologous recombination. Genes were codon optimized for *K. phaffii* (*Pichia pastoris*), synthesized (IDT), and subcloned for storage. All genes were placed between the *K. phaffii AOX1* promoter, and the *S. cerevisiae CYC1* terminator. Linear inserts were amplified using PCR and column purified. The multiplexed guide RNA cassette was synthesized in three sections separated by tRNA to avoid self-homology of the structural component of sgRNAs. Fragments were assembled into a multiplexed RNA Pol-III cassette using Golden Gate assembly (NEB). The multiplexed RNA Pol-III cassette was then cloned into the previously constructed Cas9 expression vector using Gibson assembly.

Wild type, Δ*och1*, and Δ*ku70* cells were transformed in duplicate with a mixture of 250 ng of plasmid pND413 and 2.5 µg of each linear insert encoding each glycosylation gene. Sixteen colonies were randomly selected for genomic DNA extraction. All three targeted loci were amplified from each genomic DNA sample and Sanger sequencing was performed to confirm integration of linear inserts at the correct loci. Off-target integrations were identified by amplification using heterologous gene specific primers in samples with an otherwise unedited target locus.

K3 glycan profile characterization

Wild type, Δ*och1*, and glycoengineered cells were transformed[165] with a vector for multicopy expression of human K3 plasminogen peptide. Strains were grown in 24-well deep well plates (25°C, 600 rpm) using glycerol-containing media (BMGY-Buffered Glycerol Complex Medium, Teknova) supplemented to 4% (v/v) glycerol. After 24h of biomass accumulation, cells were pelleted and resuspended in BMMY (Buffered Methanol Complex Medium, Teknova) containing 3% (v/v) methanol. After 24 hours of production, supernatant was harvested for analysis by 16% Tricine SDS-PAGE (Thermo Fisher Scientific). K3 peptide was purified using HisTrap columns (GE), deglycosylated with PNGase F (NEB), and glycosylation verified by LCMS (Agilent QTOF-6530).

Media screening and transcriptome analysis

A rich-defined media[167] was used as the base media for screening carbon source or additives. Table B3.1 details the list of carbon source and additives.

Cells were inoculated at 0.1 OD600 at 3 mL plate scale and harvested after 18 h of biomass accumulation in the described rich-defined media. RNA was extracted and purified according to the Qiagen RNeasy 96 kit. RNA quality was analyzed on an Agilent BioAnalyzer to ensure RNA Quality Number >6.5. RNA was reverse transcribed with Superscript III (ThermoFisher) and amplified with KAPA HiFi HotStart ReadyMix (Roche). RNA libraries were prepared using the Nextera XT DNA Library Preparation Kit with the Illumina DNA/RNA UD Indexes Set A. sequenced on an Illumina Nextseq to generate paired reads of 50 (read 1) and 50 bp (read 2). Sequenced mRNA transcripts were demultiplexed using sample barcodes, aligned to the wildtype *Komagataella phaffii* genome (strain Y11430) and exogenous transgenes, and

quantified using Salmon version 1.6.0.[168] Gene level summaries were prepared using tximport

version 1.24.0[169] running under R version 4.2.1.

Inositol-repressible promoter functional test

Cells were grown overnight in YPD at 30°C and diluted with sterile water (Invitrogen) to

0.6 OD600 the following day. The diluted liquid culture was then serial-diluted 10, 10^2, 10^3, 10^4,

and 10^5 times with sterile water. 5 μL of the original dilution and each serial dilution

(corresponding to approximately 10^5, 10^4, 10^3, 10^2, 10^1, and 1 cell(s)) were stamped onto minimal

glycerol media (6.7 g/L BD Yeast Nitrogen Base w/o Amino Acids, 10 mg/L L-histidine, 20

mg/L L-methionine, 20 mg/L L-tryptophan, 10 mL/L glycerol) plates with 0, 0.2, 0.5, or 1 mM

inositol without selection. Plates were incubated for at least 2 days at 30°C before examination.

4. ENGINEERING G0 PATHWAY FOR HOMOGENEOUS GLYCOSYLATION ON K3 PEPTIDE

Glycosylation is a common post-translational modification and is vital for the safety and efficacy of many therapeutic proteins. Although *K. phaffii* has dedicated machineries for performing glycosylation, its native high-mannose structures can negatively impact recombinant protein folding and stability and raise immunogenicity concerns. In the previous chapter, we demonstrated the feasibility of using CRISPR-Cas9 to humanize the yeast glycosylation pathway in a multiplexed fashion, but additional engineering is needed to increase glycan profile homogeneity. In this chapter, we demonstrate constructing a constitutively expressed heterologous glycosylation pathway in *K. phaffii*. We achieved homogeneous G0 glycosylation on a reporter peptide by using a spontaneously occurring enzyme mutant or tuning down the pathway expression to a host-appropriate level. Informed by RNA sequencing, we further show that growth defects from glycoengineering can be partially rescued by gene knockouts in the MAPK cascade.

4.1. Background and Motivation

The recombinant biopharmaceutical market is currently comprised of over 300 products.[1] Although new kinds of therapeutics such as nucleic acid-based products have made their entry, the market is still dominated by recombinant proteins, many of which require glycosylation for their folding, stability, and efficacy.[79,170] Mammalian cells such as Chinese Hamster Ovarian (CHO) cells are typically the preferred platform for the expression of monoclonal antibodies and other glycoproteins, because their glycosylation machinery can produce human-like glycan profiles, ensuring the safety and efficacy of these protein therapeutics. Nevertheless, high production costs, the requirement for complex media, and risk of viral contamination in mammalian expression systems have spurred the interest in utilizing alternative expression hosts such as yeasts.[10,15]

Unlike CHO cells, which produce human-like glycans, yeasts produce high-mannose glycan structures that can lead to target molecule's fast clearance and poor efficacy.[171] Thus, the emergence of glycoengineered *Komagataella phaffii* is promising for the possibility of using this fast-growing methylotrophic yeast for pharmaceutical glycoprotein production. Humanization of the *K. phaffii* glycosylation machinery involves the inactivation of the native glycosylation pathway and the integration of heterologous glycosylation genes.[172] The selection of these heterologous genes has notably come from screening combinatorial libraries of subcellular localization sequences and catalytic domains, often driven by strong constitutive promoters like *GAPDH* promoter to maximize their effectiveness.[128] However, glycoengineered strains suffer fitness defects, including slower growth,[127] decreased stationary phase cell density,[173] and increased propensity for lysis.[174] Nevertheless, the cellular effect of glycoengineering, to the best of our knowledge, remains poorly characterized on a molecular level.

In this chapter, we describe the reconstruction of humanized glycosylation pathway to GlcNAc2Man3 (G0) glycan structure. We discovered a previously unreported synthetic lethality – the overexpression of *C. elegans* α-1,2-mannosidase I *MNS1* is incompatible with the deletion of native α-1,6-mannosyltransferase *OCH1*. We demonstrate that viable glycoengineering can be achieved via utilizing a spontaneous mutant of *MNS1* or expressing the wild-type *MNS1* with a less active promoter and that other glycosidases and glycosyltransferases in the humanized glycosylation pathway can be driven by weaker, native biology-informed promoters without compromising glycan homogeneity. Through RNA sequencing, we further show that the decreased growth rate stems from less active metabolism and upregulated MAPK cascade. Gene knockouts in the yeast MAPK cascade could partially rescue the growth defects phenotype.

4.2. Results

4.2.1. Discovery of a synthetic lethality during pathway humanization

Previous G0 glycoengineering strategies used combinatorial libraries to screen fusion constructs of different cytosolic, transmembrane, and stem (CTS) domains and catalytic domains for highly effective glycosidases and glycosyltransferases. Most of these fusion protein sequences were constructed using the cDNA libraries of the host organisms, and often expressed under the control of strong constitutive or methanol-inducible promoters such as *GAPDH* and *AOX1* promoters.[128] As different organisms have different codon usage preference, the native sequences of glycosylation enzymes from higher eukaryotic organisms (including *Homo sapiens*, *Rattus norvegicus*, and more) could contain codons rarely used in *K. phaffii*. Codon optimization and gene synthesis technologies have seen vast improvement in recent years and can improve recombinant protein and heterologous pathway expression.[175] Furthermore, existing engineering approaches often depend on the repeated recycling of multiple auxotrophic or antibiotic

markers,[59] which leaves genomic scars and potentially alters cellular amino acid metabolism. To circumvent this problem, the CRISPR-Cas9 system has been adapted and optimized for *K. phaffii*, allowing for fast and markerless genome editing.[112,124]

Unlike previous glycoengineering strategies, which started with the knock-out of the native alpha-1,6-mannosyltransferase OCH1 (the gene responsible for the initiation of outer chain hypermannosylation), we decided to leave the knock-out as the last step of engineering efforts. Through our previous experience with working with *Δoch1* cells, strains with this genotype have significantly reduced growth rate, which slows down the sequential genome editing required for glycosylation pathway humanization.[112] The first step in our glycoengineering strategy involves the cleavage of *K. phaffii* native Man8 to Man5 glycan structure via a heterologous α-1,2-mannosidase. The fusion protein consisting of *S. cerevisiae MNS1* localization sequence and *C. elegans mans-1* catalytic domain has been reported to be among the best performing fusion constructs and was thus selected to initiate the humanized pathway.[128] We codon-optimized the mannosidase fusion construct and a subsequent *K. lactis* uridine diphosphate (UDP)-GlcNAc transporter *YEA4* for expression in *K. phaffii* and integrated these two genes under the control of strong constitutive promoters P_{GAPDH} and P_{TEF1}, respectively, using a previously reported CRISPR-Cas9 system.[112] The subsequent attempt at *GNT1* integration yielded no successful transformant with either promoter. We assayed the transcriptome of wild type *K. phaffii* in common laboratory conditions, and RNA sequencing showed that genes like GAPDH and TEF1 are natively expressed at tens to hundreds of folds stronger than native glycosylation genes (Figure 4.1A). We thus theorized that the overexpression of *GNT1* under the control of P_{GAPDH} or P_{TEF1} was growth inhibiting and switched the promoter for *GNT1* expression to either *ENO1* promoter or *MNN4* promoter. The

native expression of *ENO1* is comparable to that of *GAPDH* and *TEF1*, so P$_{ENO1}$ serves as another test case for strong constitutive expression of the heterologous glycosylation genes. *MNN4* is a native putative regulator of phosphomannosylation of N-linked glycans,[176] so using its promoter to drive *GNT1* expression could allow us to mimic native glycosylation pathway expression. We were able to obtain successful *GNT1* integrants with both P$_{MNN4}$ and, surprisingly, P$_{ENO1}$. This could be due to the differential expression of native *ENO1* upon glycoengineering, which changes the strength of the promoter compared to that in a wildtype background.

Following the integration of *MNS1*, *YEA4*, and *GNT1*, we decided to benchmark the generated glycoengineered strains using the K3 peptide. We transformed K3 into the glycoengineered strains and targeted *OCH1* for knockout. *OCH1* knockout was achieved through replacing its open reading frame with that of blasticidin-S deaminase, which confers *Δoch1* cells resistance to blasticidin. We then evaluated the glycan profile of K3 peptide secreted from blasticidin-resistant transformants with mass spectrometry. However, we were unable to detect the presence of GlcNAcMan5 glycan. Since the glycan profile was mainly consisted of high mannose Man8 to Man11 structures (Figure A4.1), we suspected that the heterologous Mns1p, responsible for cleaving Man8 to Man5, lost its enzymatic activity. By sequencing the integrated *MNS1* construct of the cultured clones, we discovered that there were spontaneous mutations in its coding sequence. After confirming that no *MNS1* mutations existed in the parent strains (with intact *OCH1*), we repeated the *OCH1* knockout in glycoengineered strains expressing all three heterologous genes (with *GNT1* controlled by either P$_{ENO1}$ or P$_{MNN4}$), both with and without K3 integration. In total, we sequenced the *MNS1* integration of 82 clones and discovered mutations at 23 unique amino acid residues (Figure A4.2). The spontaneous occurrence of unlinked

mutations in the *MNS1* coding sequence could mean the incompatibility between OCH1 knockout and overexpression of *MNS1* with P_{GAPDH}. It is possible that its knockout, when combined with the integration and overexpression of codon-optimized α-1,2-mannosidase *MNS1*, alters the glycosylation landscape of native glycoproteins, which include structurally integral cell wall mannoproteins, to a lethal extent, and only cells that had rendered *MNS1* inactive through spontaneous mutations could survive.

Because this incompatibility, to the best of our knowledge, has never been reported before, we investigated if further humanization, namely integration of *MNS2* and *GNT2* for G0 glycosylation, could resolve the incompatibility. We followed the same engineering strategy as we did with *GNT1* and constructed two lineages of *OCH1*-intact glycoengineered strains, with *GNT1*, *MNS2*, and *GNT2* expressed under P_{ENO1} or P_{MNN4}, in addition to the already integrated P_{GAPDH}-*MNS1* and P_{TEF1}-*YEA4*. We attempted *OCH1* knockout in these strains and assessed the *MNS1* coding sequence. Unfortunately, unlinked mutations were still detected in all successful knockout transformants. Combining all *OCH1* knockouts across different glycoengineered strains, we screened a total of 158 colonies, 121 of which were frameshift mutations, and the remaining 37 were amino acid substitutions. These mutations occurred throughout *MNS1* ORF at 47 unique amino acid residues (Figure 4.1B) and showed that further glycoengineering till G0 glycan did not resolve the synthetic lethality between *MNS1* and *Δoch1*.
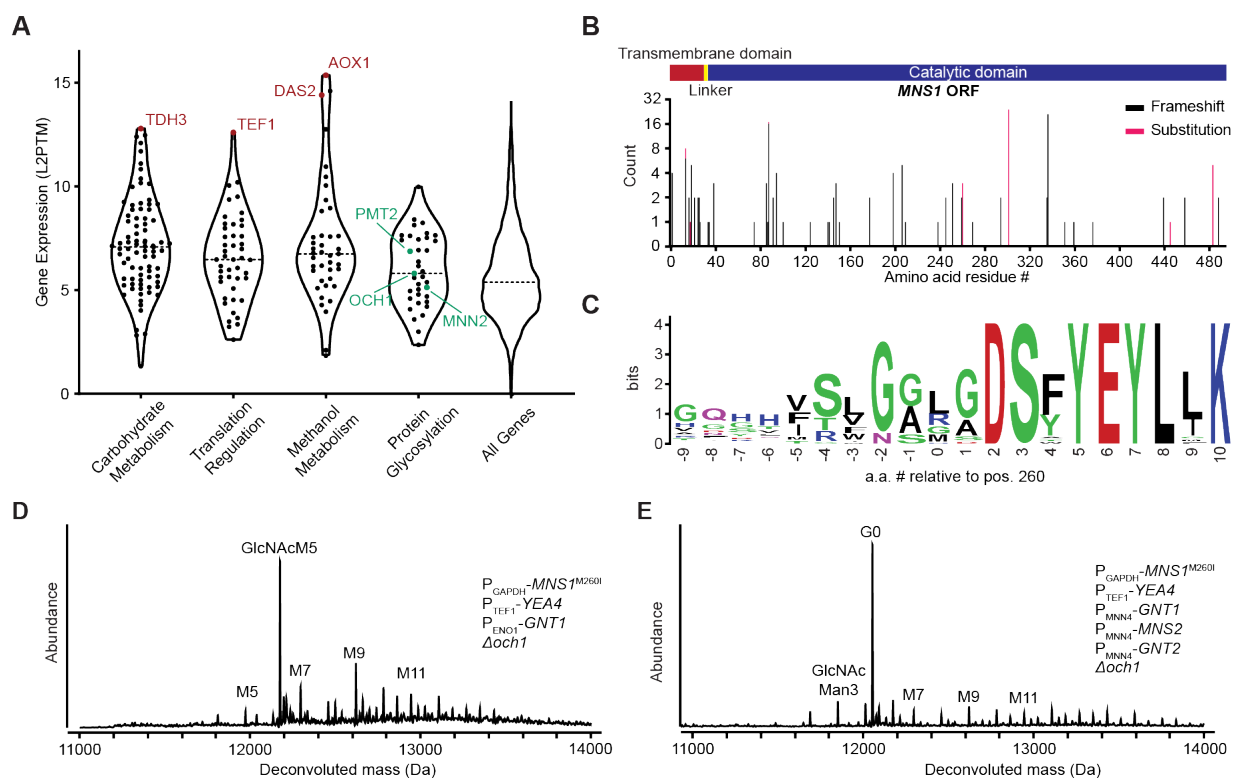
Figure 4.1. Resolution of a synthetic lethality during glycosylation pathway construction. (A) Expression levels of select representative native gene sets. Note that *K. phaffii* native protein glycosylation pathway genes (with some highlighted in green) are expressed at much lower levels than those of commonly used promoters (highlighted in red) for heterologous gene expression. (B) Spontaneous mutations in the *MNS1* open reading frame after *OCH1* knockout. (C) Logo plot of amino acid residues surrounding M260 across different organisms. (D) Mass spectrum of K3 peptide secreted from a GlcNAcMan5-glycoengineered strain. (E) Mass spectrum of K3 peptide secreted from a G0-glycoengineered strain.

4.2.2. Identification of a spontaneous *MNS1* mutant

We examined the mutation sites within *MNS1* more closely. Through the sequence homology between the *C. elegans mans-1* and its homologs in other eukaryotic organisms, we identified in the fusion protein construct E373 and T484 as the proton donor active site and calcium ion cofactor binding site, respectively. Frameshift mutations result in a completely different translation from the original, abolishing the enzymatic function, so the frameshift mutants were not investigated further. Most point mutations are substitutions of hydrophobic residues to charged residues, or vice versa, in conversed regions across species. However, we identified one mutant with methionine residue at position 260 mutated to isoleucine that could be of interest for further study. By comparing mannosidase homologs across different eukaryotic organisms, including *C. elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Penicillium citrinum*, *Mus musculus*, and more, we found that the methionine residue is not highly conserved across species, and leucine is the most frequent amino acid, followed by arginine and glycine (Figure 4.1C). Given the structural similarity between leucine and isoleucine, we decided to investigate this *MNS1* M260I mutant further because it could potentially retain its enzymatic activity.

The *MNS1* M260I mutant was identified in the *Δoch1* glycoengineered strain expressing *YEA4* and *GNT1* under the control of P_{TEF1} and P_{MNN4}, respectively. We cultivated this mutant clone in a batch culture, purified the secreted 6His-tagged K3 peptide using affinity-based chromatography, and analyzed its glycan profile via intact protein LC-MS (Figure 4.1D). This glycoengineered strain with mutated *MNS1* produced predominantly GlcNAcMan5 glycan, with some high-mannose glycan structures ranging from Man7 to Man11. The presence of high mannose glycans indicated insufficient cleavage of mannoses by Mns1p, and this decreased enzymatic activity could explain why the M260I mutant was able to survive even when

expressed under the control of strong constitutive P_{GAPDH}. To test if this spontaneously occurring

mutant would enable G0 glycoengineering, we reverted the defective *MNS1* in previously

constructed *Δoch1* G0 strains and re-integrated mutant *MNS1* under the control of P_{GAPDH}. Post-

transformation confirmation of *MNS1* mutant did not show any new mutations, and the glycan

profile of its secreted K3 peptide was predominantly G0 (Figure 4.1E).

4.2.3. Generation of viable G0 strains with attenuated *MNS1* expression

Seeing that mutant *MNS1* enabled G0 glycoengineering, likely because of its reduced

enzymatic activity, we investigated whether viable transformants could be obtained by

expressing wildtype *MNS1* with weaker promoters. Less active promoters can allow for similar

level of overall mannosidase activity, but they could have the advantage of using less cellular

resources. To test if less active promoters can support the expression of wildtype *MNS1* without

inducing cytotoxicity, we built wildtype *MNS1* constructs expressed under the control of

additional seven promoters, selected based on their native gene expression – P_{ENO1}, P_{PPA2}, P_{CDA2},

P_{MNN4}, P_{MNN10}, P_{BCK1}, and P_{1500} (Figure 4.2A). These constructs were then integrated into *Δoch1*

G0-glycoengineered strains without *MNS1*, whose original integration was reverted after

detecting mutations in the ORF. Table B4.1 shows successful integration of all attempted *MNS1*

integration. Wildtype *MNS1* expressed under strong *ENO1* promoter cannot be integrated into

either G0 strain, consistent with our previous experience with the synthetic lethality between

Δoch1 and strong constitutive expression of *MNS1*. G0 strain with *GNT1*, *MNS2*, and *GNT2*

under the control of P_{MNN4} can tolerate a higher expression level of wildtype *MNS1*, indicating a

possible inverse connection between highest tolerable *MNS1* expression and cellular stress from

heterologous pathway overexpression. Surprisingly, P_{MNN4} and P_{MNN10}, two promoters from the

K. phaffii native glycosylation pathway, did not generate viable transformants when used to

express *MNS1*, especially in the G0 strain with high pathway expression. One possible explanation is that the native glycosylation pathway could be upregulated in the glycoengineered strains, so the expression level of *MNS1* was in fact higher than what we had expected based on MNN4 and MNN10 gene expression in wildtype *K. phaffii* strain. We then cultivated the G0 strains with successful *MNS1* integration and assayed the glycan profile of their secreted K3 peptide. All viable *MNS1* constructs were able to generate homogeneous G0 glycan, and the presence of high-mannose structures was reduced in most G0 strains with wildtype *MNS1* compared to with mutant *MNS1*, despite that a strong constitutive P_{GAPDH} was used to drive the expression of the mutant (Figure 4.2B). This further confirms that M260I mutation reduces the enzymatic activity of *MNS1*, and that this reduction in activity can be achieved via using less active promoters with wildtype sequence.
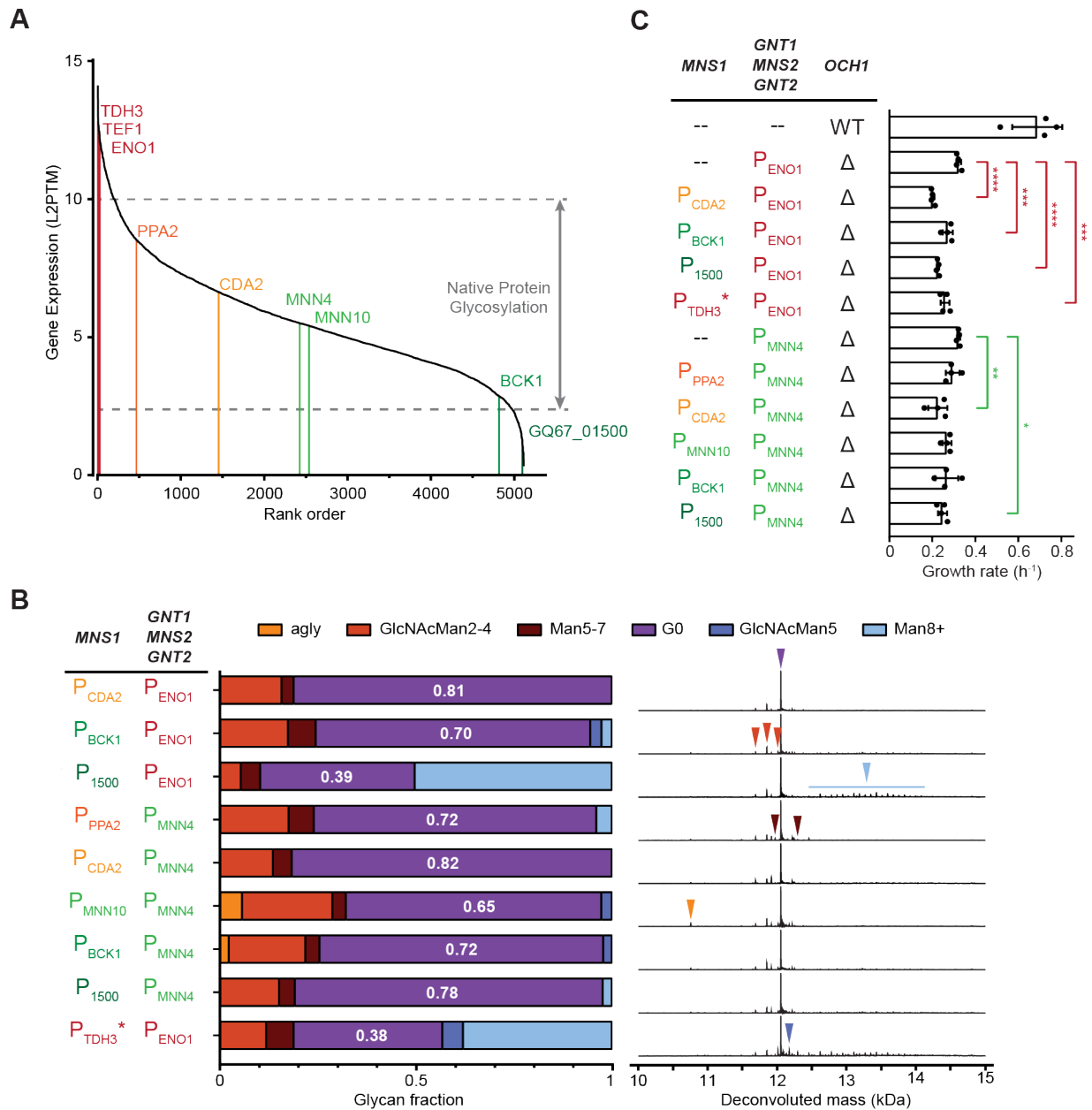
Figure 4.2. Attenuating *MNS1* expression enables G0 glycoengineering.
(A) Activity of promoters for wildtype *MNS1* expression. (B) Glycan profiles based on intact protein LC-MS of secreted K3 peptide from different G0 strains. (C) Growth rates of different G0 strains, compared to wildtype and "partially glycoengineered" strains. Wildtype growth rate is significantly different from all other strains ($p < 0.0001$, one-way ANOVA). Red brackets: growth comparisons among P_{ENO1}-*GNT1/MNS2/GNT2* strains. Green brackets: growth comparisons among P_{MNN4}-*GNT1/MNS2/GNT2* strains. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.0001$, one way ANOVA.

4.2.4. Further characterization of G0 glycoengineered strains

Because glycoengineered strains have been reported to exhibit growth defects,[127] we examined the growth rate of our strains and compared them to wildtype *K. phaffii* strain. We included G0 strains of both P_{ENO1} and P_{MNN4} engineering lineages and their parent strains (with the original *MNS1* integration reverted). At room temperature, the growth rate of glycoengineered strains is less than half of wildtype growth rate, with the slowest growing glycoengineered strain at approximately 30% of wildtype (Figure 4.2C). Because we saw the incompatibility between *Δoch1* and highly expressed *MNS1*, we compared G0 strains before and after *MNS1* re-integration. Compared to their respective parent strains, six of nine strains showed significant growth rate decrease upon *MNS1* knock-in, but there is no apparent correlation between growth rate and *MNS1* promoter strength. Furthermore, comparing strains with the same *MNS1* constructs but differentially expressed *GNT1*, *MNS2*, and *GNT2* (using either P_{ENO1} or P_{MNN4}), we did not observe any significant difference in their growth rate. To better understand the major engineering changes that impact growth, we performed additional growth rate assays (Figure A4.3). We observed that the integration of heterologous pathway genes resulted in more than 40% decrease in growth rate in wildtype *K. phaffii*. In *Δoch1* strain, however, pathway integration had no negative impact on growth, and G0-engineered strains had higher growth rates than *Δoch1* strain. This partial recovery of growth defects is contrary to previous reports of serious cellular burden upon *MNS2* and *GNT2* integration[127] and likely stems from using attenuated expression cassettes for the heterologous pathway. We also examined K3 peptide secretion from select glycoengineered strains – different *MNS1* constructs or expression levels of other heterologous pathway enzymes did not seem to greatly impact K3 peptide secretion (Figure A4.4).

We then characterized the cell walls of these glycoengineered strains against wildtype strain. First, different sugar moieties on glycoproteins would interact differently with Alcian blue, a polyvalent cationic dye that binds to negative charges, such as those from phosphomannoses, at the cell wall.[176] We thus compared the Alcian blue staining of different strains and observed that wildtype strain had intense staining, because of the phosphomannoses present on the mannan chain of cell-surface proteins (Figure 4.3A). The knockout of *OCH1* inhibits outer chain extension, where most of the phosphomannoses reside, and significantly decreases Alcian blue staining. Integration of *MNS1* cleaves down mannose sugars, further precluding phosphomannose addition and decreasing staining. However, after the subsequent introduction of *GNT1*, *MNS2*, and *GNT2*, the resulting glycoengineered strain showed significantly more Alcian blue staining compared to both Δ*och1* and *MNS1*-integrated strains. This was unexpected because Alcian blue should not have high binding affinity to the terminal GlcNAc on G0 glycan due to its neutral charge. One possible explanation is that although the secreted K3 peptide did not exhibit any phosphomannosylation on its glycan profile, some degree of phosphates could still be present on the cell-surface proteins. The integration of a complete G0-humanized glycosylation pathway could have triggered the upregulation of native phosphomannosylation pathway, leading to increased negative charge on the cell surface and increased Alcian blue staining. The glycan profile of the cell surface proteins thus remains to be further studied to directly confirm this hypothesis.

The resistance of different strains to Congo red and Calcofluor white was also examined. These two chemicals interfere with the construction and stress response of fungal cell wall, and their susceptibility can thus serve as good representation of cell wall integrity.[177] We notably observed that Δ*och1* and *MNS1* integration greatly decreased the wildtype strain's tolerance

against Calcofluor white, but the integration of additional genes to G0 helped partially restore this phenotype (Figure 4.3B). This observation could serve as one piece of evidence for the upregulation of phosphomannosylation pathway upon further engineering of the *K. phaffii* native glycosylation pathway, since the decoration of additional phosphomannose moieties on cell-wall proteins should confer the cells increased resistance to Calcofluor white. Congo red susceptibility assay showed similar results, Δ*och1* and *MNS1* integrated strains showed zero growth, while G0-glycoengineered strains showed minimal growth.
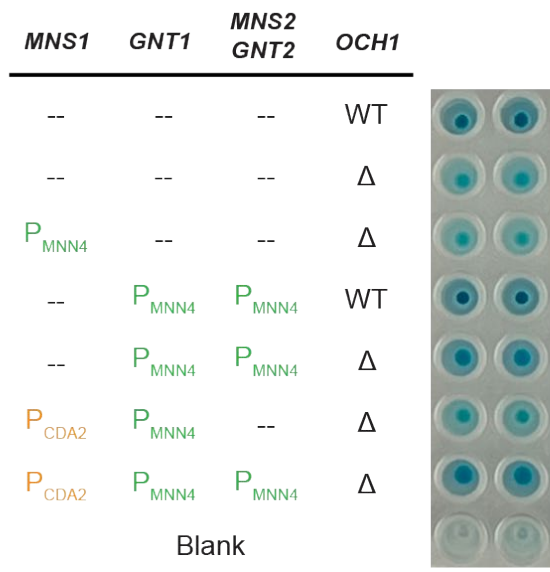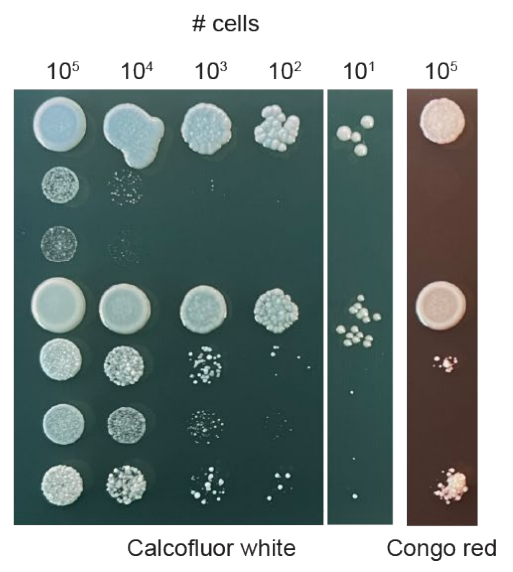
Figure 4.3. Cell wall characterization of glycoengineered strains.
(A) Alcian blue binds to negative charges on the cell wall (such as mannosylphosphates), resulting in the blue coloring of *OCH1*-intact strains. Surprisingly, increased staining is also observed in strains expressing GlcNAc transferases. (B) Cell wall integrity assay of the same strains by either Calcofluor white (with Evans blue) or Congo red.

4.2.5. Transcriptomic analyses of G0 glycoengineered strains

To better understand the impact of glycoengineering, we collected the transcriptomes of several G0-glycoengineered strains and compared them to those of partially glycoengineered strains (with all genome edits required for G0 except for MNS1 integration) and of wildtype *K. phaffii*. By principal component analysis, we see that the first component, which accounts for 51% of the variance in the data, separates glycoengineered and wildtype *K. phaffii* strains (Figure A4.5). It also worth noting that although we have observed cellular toxicity from the overexpression of *MNS1*, its expression level in viable glycoengineered strains does not seem be correlated with either principal component in this two-dimensional projection of transcriptomic dataset, even in the strains with the complete humanized glycosylation pathway except for *MNS1*. We then performed pathway enrichment analysis of the best performing glycoengineered strain (with P_{CDA2}-*MNS1* integration) compared to wildtype. We observed that in wildtype strain, pathways related to electron transport train, metabolic processes, and redox activity are upregulated, while those related to signal transduction are upregulated in the G0 glycoengineered strain (Figure 4.4A). This would explain the difference between growth rate since the wildtype strain exhibits high energy generation and metabolism, while signal transduction (the mitogen-activated protein kinase (MAPK) cascade) mediates cell growth and stress response.[178] Similar pathway enrichment comparisons were carried out in other assayed glycoengineered strains against wildtype strain, and they showed similar pathway expression profiles.

Based on this finding, we targeted select heavily upregulated but relatively non-essential genes in the MAPK pathway for knockout to examine if such genetic perturbations could partially restore cell growth (Table B4.2). Three of seven genomic perturbations had positive impact on cell growth, with Δ*msb2* yielding a 15% increase in growth rate (Figure 4.4B). Future

studies could include assaying K3 peptide secreted from these knockout strains to ensure the proper function of their glycosylation pathway.

Because principal components from PCA are linear combinations of all variables, we aimed to identify the truly differentially expressed genes in the dataset using sparse PCA (sPCA). This method employs additional regularization techniques to introduce sparsity in the loading of principal components, thus isolating high contributing variables to each PC.[179] By applying sPCA, we were able to construct five modules with 110 genes (as opposed to more than 3,200 genes in the complete transcriptome, Table B4.3) that best characterize the difference in the dataset. Through functional enrichment analysis,[180] modules 1 and 2 and each sample's respective scores confirm the results from pathway enrichment analysis above: module 1 comprises of essential pathways in the central metabolism, while module 2 is enriched with MAPK cascade (Figure 4.4C). Additional metabolic pathways are enriched in modules 3 and 4, but there is no apparent correlation between their scoring and sample type (Figure A4.6).

We then examined the transgene expression level of these glycoengineered strains, especially that of *MNS1*, the first gene in the humanized glycosylation pathway. All tested *MNS1* cassettes were integrated at the same intergenic locus with different promoters, but we observed an unexpected "baseline" expression level of *MNS1* that was above the expected values for weaker promoters (Figure 4.5A). In the G0 strain with P_{CDA2}-driven *MNS1*, the expression level of *MNS1* match up with *CDA2* in wildtype or in glycoengineered strains. However, in P_{GAPDH}-*MNS1mut* G0 strain, *MNS1* had a lower expression level, while in all other glycoengineered strains, *MNS1* was more strongly expressed than expected. We thus hypothesized that *MNS1* expression is, to some degree, influenced by neighboring gene expression. The gene upstream of *MNS1* integration is expressed at, on average, approximately 8.7 L2TPM in both wildtype and

glycoengineered strains, and although there is a predicted terminator, based on RNA hairpin secondary structure and a poly-T sequence,[181] read-throughs of RNA polymerase could happen and impact the downstream gene expression.[138,182] Although unexpected, the altered expression of *MNS1* does not change the fact that P$_{GAPDH}$-driven wildtype *MNS1* was cytotoxic, evidenced by the fact that mutant *MNS1*, driven by the same promoter, was expressed at least 2.5 fold stronger than the strongest viable wildtype *MNS1* construct and almost 10 fold stronger than the median.

To address this issue, we selected new integration loci for the expression of *MNS1* with less active promoters. These new loci had upstream/downstream gene expression closely matching with the promoter strength to allow for better examination of the impact of *MNS1* promoter on glycan profile (Figure A4.7). After expressing K3 peptide in these new G0 strains, we analyzed its glycan profiles with intact protein LC-MS. We observed that integrating *MNS1* at these new loci does not seem to strongly impact G0 glycan homogeneity (Figure 4.5B). Nevertheless, future studies should include additional characterization of these strains, including growth rate, cell wall integrity, transcriptomics *etc*. to confirm the further downregulation of *MNS1* and fully assess the impact of promoter strength. This also shows that in adopting *K. phaffii* as a host for routine genome engineering, we still need to further explore and identify optimal landing pads for the knock-ins of heterologous genes with consistent predictability.
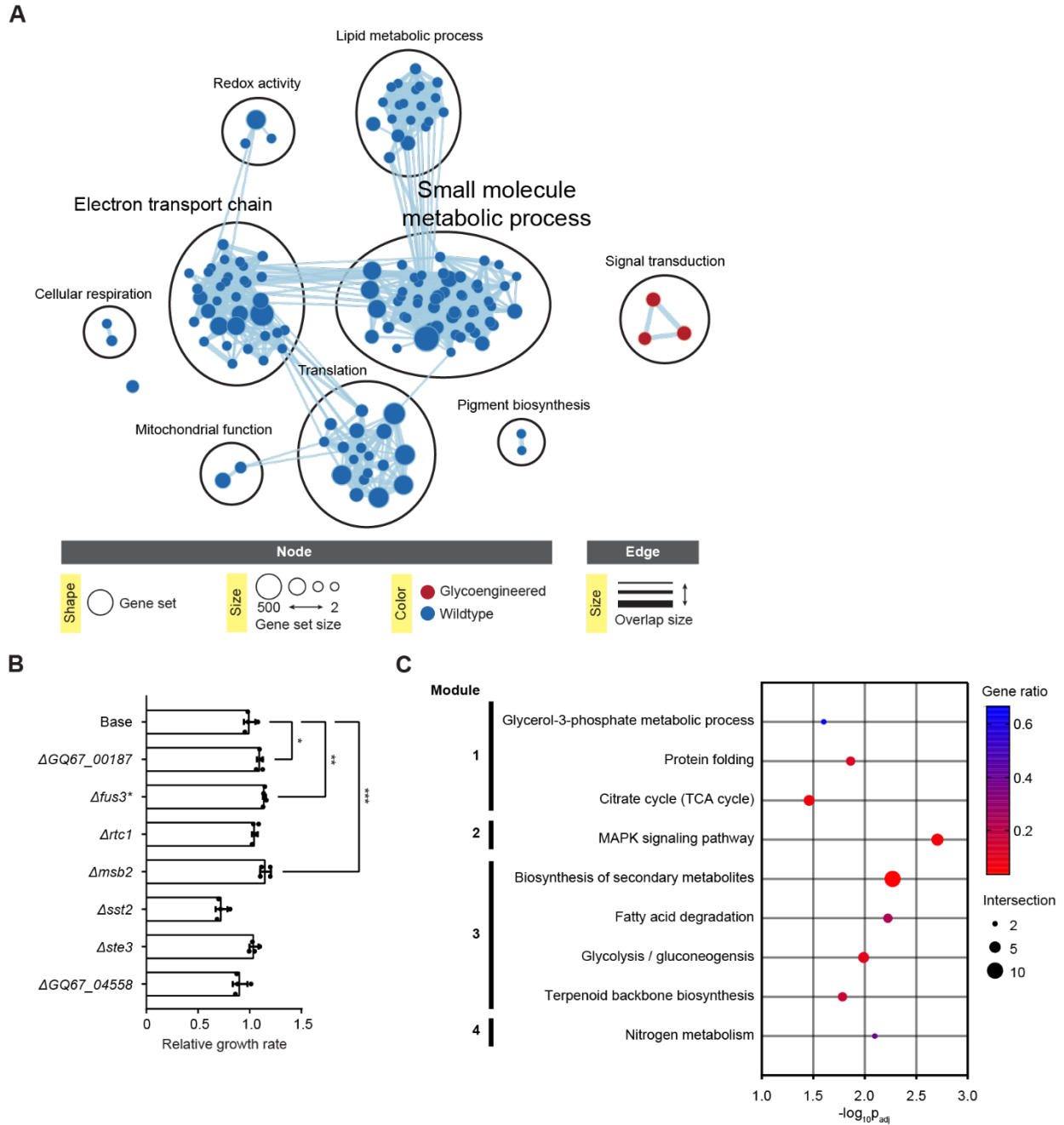
Figure 4.4. Transcriptomic analysis of glycoengineered strains.
(A) Pathway gene set enrichment map of wildtype vs. glycoengineered strains. (B) Restoration of growth rate defects of a glycoengineered strain through gene knockouts in the MAPK cascade signal transduction pathway. (C) Gene set enrichment analysis dotplot of modules 1-4 generated from sparse PCA analysis. Genes in module 5 did not show any significant enrichment of gene sets.
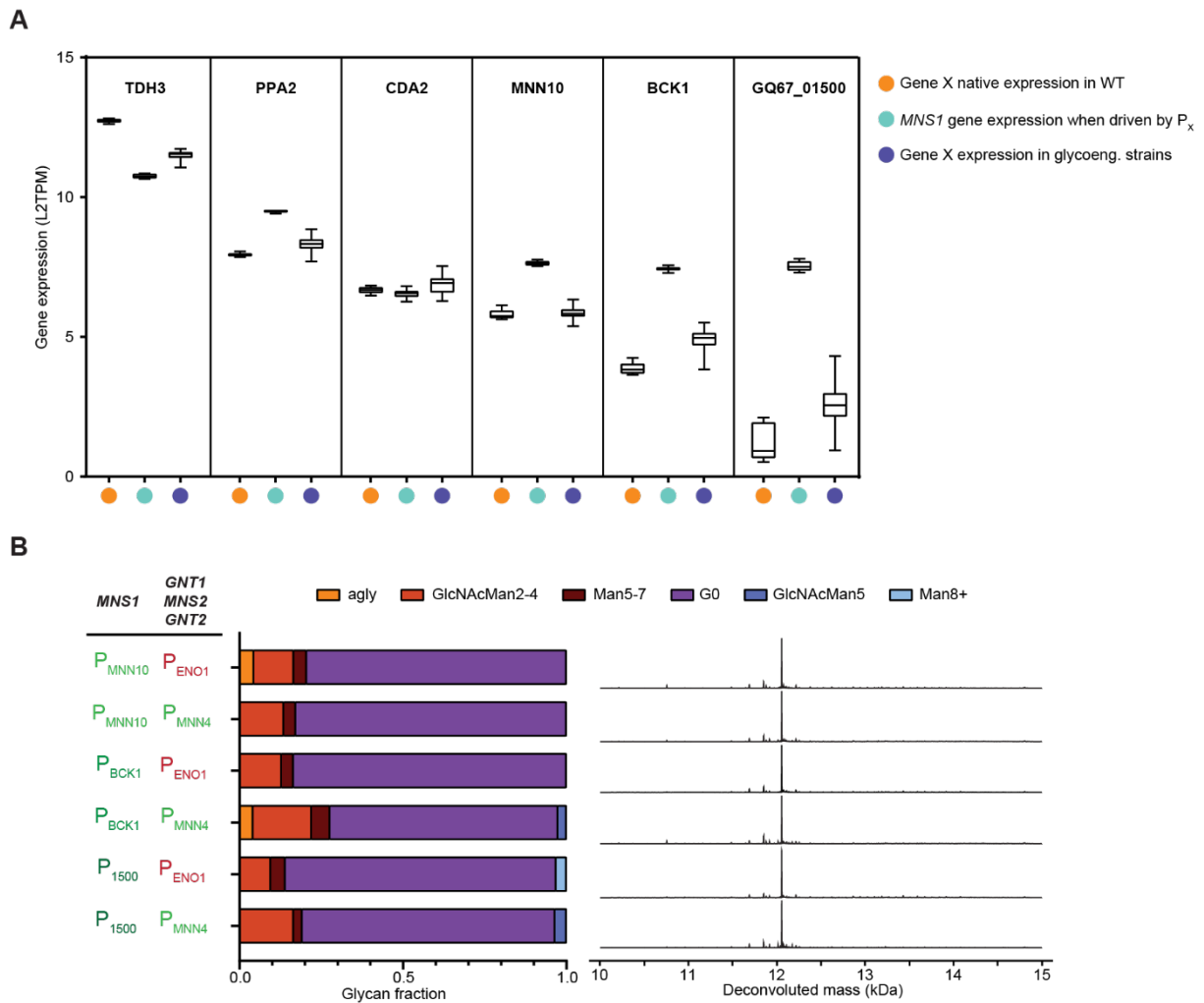
Figure 4.5. Selection of alternative integration loci for *MNS1* integration.
(A) While the expression of native genes in wildtype vs. glycoengineered strains is largely consistent, the expression of *MNS1* exhibits a mismatch with the expected promoter activity level. (B) Glycan profiles based on intact protein LC-MS of secreted K3 peptide from different G0 strains with *MNS1* integrated at new loci.

4.2.6. Extension of humanized pathway for galactosylation

Beyond G0, there are two major glycoforms – Gal2GlcNAc2Man3 (commonly referred to as G2) and Sia2Gal2GlcNAc2Man3 (or Sia2G2), where Gal refers to galactose, and Sia refers to sialic acid or, more specifically, N-acetylneuraminic acid (Neu5Ac). These two glycoforms have important therapeutic values - for example, terminal galactose in G2 glycoform enhances complement dependent cytotoxicity in monoclonal antibodies, and terminal sialic acid greatly prolongs protein half-life for erythropoietin.[54,91] Since native *P. pastoris* glycans do not undergo galactosylation or sialylation, the precursors of these modifications, UDP-galactose and CMP-sialic acid, must be *de novo* synthesized.

G2 glycan requires the additional of two terminal galactoses by a galactosyltransferase with UDP-galactose as the reaction substrate. UDP-galactose can be generated from the epimerization of UDP-glucose, and this can be achieved by *Schizosaccharomyces pombe* UDP-Gal 4-epimerase *UGE1*. UDP-galactose is then transported to the Golgi lumen through *Drosophila melanogaster* UDP-galactose transporter *senju* and added onto the G0 glycan via *H. sapiens* galactosyltransferase *GALT*. Previous glycoengineering literature reported a tripartite fusion protein consisting of a Golgi single-pass membrane tether from *S. cerevisiae MNN2*, the Gal/Glu epimerase *UGE1*, and the transferase *GALT*.[127] This approach relies on the diffusion of UDP-glucose, a byproduct of ER-resident glycosylation pathway, from ER to Golgi, where it is isomerized to UDP-galactose and transferred onto G0 glycan. In addition to this tripartite protein construct, we decided to include another fusion protein construct which integrates the epimerase on the cytosolic tail, UDP-galactose transporter for subcellular localization, and the transferase for galactose addition. For this protein construct, we attached preαSP at the N-terminus to facilitate entry into ER. We integrated these constructs, with either P_{GAPDH} or P_{MNN4}, into a G0-

engineered strain (with P$_{GAPDH}$-MNS1mut) and assessed if any galactosylation can be detected on secreted K3 peptide via intact LC-MS. From all constructs tested, only P$_{GAP}$-driven Golgi-resident construct resulted in mass spectrometry peaks that could be galactosylation (Figure 4.6A, Figure A4.8A). However, due to the identical mass between galactose and mannose, additional tests would be needed to confirm the identity of glycan.

Glycan confirmation can be done via digesting the glycoprotein with different exoglycosidases to release sugar moieties at the reducing end of glycans. To differentiate terminal galactose from mannose, we digested the secreted K3 peptide with either β1-4 galactosidase, α1-2,3- and α1-6-mannosidases, or a combination or all three. After digestion with galactosidase, two of the predominant peaks disappeared, whereas the same peaks persisted after mannosidase digestion (Figure 4.6B). We thus were able to confirm the presence of terminal galactose on the K3 peptide. The degree of galactosylation, however, is low, likely because the reserve of UDP-glucose diffused from ER to Golgi was not enough to support the flux of UDP-galactose needed. Additional genetic perturbation is likely needed to enhance galactosylation, such as the introduction of an additional transporter to increase concentration of sugar precursor in the Golgi.

Additionally, the other proposed galactosylation construct that combines epimerase, transporter, and transferase did not produce any galactosylated product (Figure A4.8B). We reasoned that this could be due to either the misfolding of the protein or incorrect subcellular localization. Based on previous reports, UDP-galactose transporter *senju* is capable of Golgi localization, but, in our fusion protein, the N-terminal addition of epimerase likely disrupted the transporter's proper localization. The removal of preαSP did not generate any galactosylated K3 molecules, either (Figure A4.8C, D). To mediate this issue, future work could include utilizing

synthetic protein scaffold to colocalize isomerase and transferase to the transporter. This strategy

has been used in *E. coli* for metabolic engineering, where increasing local substrate

concentration can increase enzyme turnover and enhance productivity.[183] In the case of protein

galactosylation, similar colocalization can increase enzyme efficiency and, thus, degree of

galactosylation.
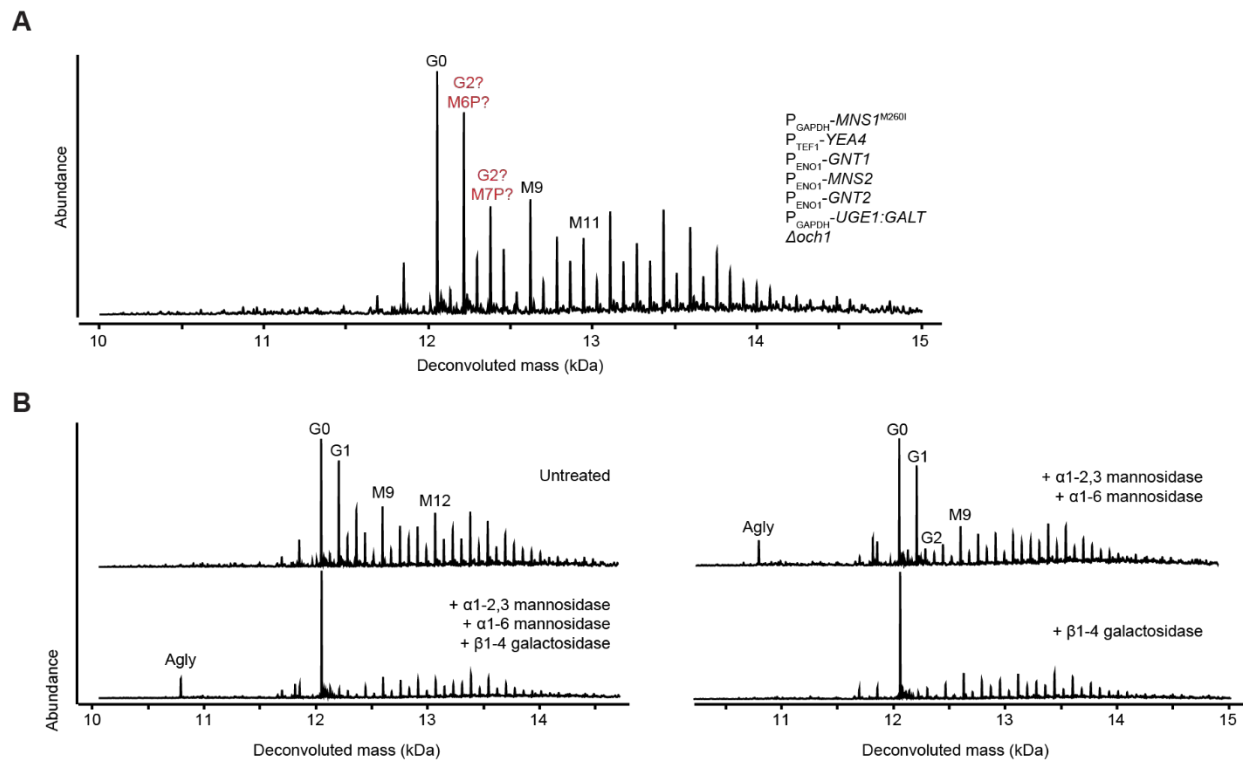
Figure 4.6. Extension of G0 glycosylation pathway for galactosylation.
(A) Mass spectrum of secreted K3 peptide in a galactosylation-engineered strain. Note the identical mass shift between galactose and mannose. (B) Exoglycosidase digestion of purified K3 sample. Digestion by galactosidase, and not by mannosidase, resulted in a mass shift of the suspected galactosylation peak, thus confirming active galactosylation machinery in the glycoengineered strain.

4.3. Discussion

In this chapter, we demonstrated the engineering of *K. phaffii* native glycosylation pathway via integrating constitutively-expressed heterologous genes. By switching from methanol-induced pathway expression to constitutive expression, all enzymes within the heterologous pathway are active and we achieved homogeneous G0 glycosylation on human plasminogen K3 peptide. Furthermore, in constructing the heterologous pathway, we identified a previously unreported synthetic lethality between Δ*och1* and *MNS1*, evident by the spontaneous, unlinked mutations in *MNS1* open reading frame after *OCH1* knockout. Previous literature did not report such mutations in their glycoengineering efforts, but during the screening of combinatorial libraries of transmembrane localization sequences and catalytic domains, the authors reported different glycosylation efficiency from different combinations.[128] Some of the variations observed could be explained by the catalytic activity of mannosidases from different organisms or the effectiveness of subcellular localization, but it is also possible that some genetic combinations were mutated *in vivo* after cassette integration - confirmation by Sanger sequencing of each individual clone within combinatorial libraries is resource-intensive and was very likely cost prohibitive in the early 2010s.

In our glycoengineering efforts, we isolated a spontaneously occurring *MNS1* mutant that enabled G0 pathway integration. This mutant had reduced enzymatic activity, evidenced by the tolerance of its strong constitutive expression. Similar reduction in *MNS1* activity was achieved by switching P_{GAPDH} for less active promoters, such as P_{CDA2}, P_{BCK1}, and P_{1500}, all generating predominantly G0 glycan on secreted K3 peptide, by downregulating the expression level of *MNS1* by 2.5- to almost 10-fold by transcriptomic analysis. It is worth noting that the expression

of *MNS1* under these weaker promoters was likely influenced by its upstream gene expression, which imposed a "floor" for the transgene expression level, resulting in a high-than-expected gene expression for *MNS1*. We selected additional integration loci with weaker neighboring gene expression for *MNS1* expression, and, similarly, all generated strains yielded homogeneous G0 glycosylation. Although our lab has started to identify intergenic loci for integration,[184] there is still great need for characterizing safe landing pads for genomic integration cassettes that enable predictable and consistent transgene expression under different promoters. Further characterization of glycoengineered strains showed that glycoengineered strains exhibited decreased cell wall integrity and slower growth, but the growth defects mainly came from *OCH1* knockout, and the integration of additional genes did not significantly impact growth rate. Pathway enrichment analysis of transcriptomic data supported this observation – wildtype strain was metabolically active, while the glycoengineered strains had upregulated MAPK cascade, regulating stress response and cell cycle.

Finally, we demonstrated further conversion of G0 to its galactosylated product by integrating a Golgi-resident tripartite fusion protein. The degree of galactosylation is low, however, and an alternative fusion protein combining isomerase, transporter, and transferase did not show any *in vivo* enzymatic activity, likely due to enzyme misfolding or incorrect subcellular targeting.

In conclusion, we were able to humanize the native glycosylation pathway with new gene manipulation techniques to homogeneously glycosylate K3 peptide, and advances in sequencing technology supported our diagnostic and engineering strategy. Proper glycosylation of K3 peptide serves as a start to achieving better glycan profile, and thus quality, control over a wide range of therapeutic proteins, which will be further explored in Chapter 5.

4.4. Methods

Yeast strains

All strains were derived from wildtype *K. phaffii* (NRRL Y-14430). The DNA fragment containing 6His-tagged K3 peptide was codon optimized, synthesized (Integrated DNA Technologies), and cloned into a custom vector for roll-in integration. Integration cassettes containing *MNS1*, *MNN2*, *GNT1*, *MNS2*, and *GNT2* were similarly codon optimized, synthesized, and cloned into a custom vector. They were integrated at the intergenic loci near *PFK1*, *GQ67_04576*, *GQ67_03878*, *GQ67_04704*, and *GQ67_05172*, respectively, following a previously described protocol through co-electroporation of 250ng of CRISPR plasmid and 2.5μg of linear DNA.

Cultivation

Strains for glycan profile characterization were grown in 3 mL culture in 24-well deep well plates (25°C, 600 rpm). Cells were cultivated in BMGY (buffered glycerol complex media, Teknova) with 4% v/v glycerol. For glycoengineered strains, after 40h of biomass accumulation, cells were pelleted and resuspended in BMMY (buffered methanol complex media, Teknova) with 3% v/v methanol. After 48 hours of production, supernatant was harvested for analysis.

Protein purification and glycan characterization

Supernatant was filtered and diluted 1:1 with Ni-IMAC (nickel-immobilized metal affinity chromatography) binding buffer (25mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4). Purification was carried out on the GE ÄKTA pure system with 1-mL HisTrap columns (Cytiva). After sample loading, the column was equilibrated with and washed with the described binding buffer before eluting with elution buffer (500mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4).

Intact protein LC-MS was performed as described previously.[52] Mass spectra were processed using MassHunter Bioconfirm software (Agilent Technologies) with a deconvolution range of 10-15 kDa, using a mass step of 0.5 Dalton.

Growth rate assay

Strains were grown in YPD (overnight for wildtype, and over two nights for glycoengineered strains) at 30°C and 250 rpm. They were then inoculated in 100μL BMGY supplemented with 4% v/v glycerol to an initial OD of 0.05 (glycoengineered strains) or 0.01 (wildtype) in Nunclon® Delta surface 96-well plate. Cultures were subsequently grown in a microplate reader (Tecan) with shaking at 1000rpm and ambient temperature for 24 hours, with OD600 measurements taken every 30 minutes. Growth data during exponential phase was fit to an exponential function to obtain growth rate.

Transcriptome analysis

Cells were inoculated at 0.1 OD600 at 3 mL plate scale and harvested after 18 h of biomass accumulation in BMGY supplemented with 4% v/v glycerol. RNA was extracted and purified according to the Qiagen RNeasy 96 kit. RNA quality was analyzed on an Agilent BioAnalyzer to ensure RNA Quality Number >6.5. RNA was reverse transcribed with Superscript III (ThermoFisher) and amplified with KAPA HiFi HotStart ReadyMix (Roche). RNA libraries were prepared using the Nextera XT DNA Library Preparation Kit with the Illumina DNA/RNA UD Indexes Set A. sequenced on an Illumina Nextseq to generate paired reads of 50 (read 1) and 50 bp (read 2). Sequenced mRNA transcripts were demultiplexed using sample barcodes, aligned to the wildtype *Komagataella phaffii* genome (strain Y11430) and exogenous transgenes, and quantified using Salmon version 1.6.0.[168] Gene level summaries were

prepared using tximport version 1.24.0[169] running under R version 4.2.1. Pathway enrichment analysis was carried as previously described.[185]

Sparse PCA analysis was carried out as previously described[186] with the appropriate R package. Sparsity index of 3 and number of factors of 5 were chosen.

Alcian blue, Congo red, and Calcofluor white assays

Alcian blue assay was carried out as described previously[187] with minor modifications. Briefly, strains were growth in YPD (overnight for wildtype, and over two nights for glycoengineered strains) at 30°C and 250 rpm. 1 OD600 equivalent of cells were then pelleted and washed once with and resuspended in 100μL 0.02N HCl. The resuspensions were transferred into a 96-well V-bottomed PCR plate (Eppendorf), and 100μL of Alcian blue solution was added. After incubation at room temperature for 15 minutes, the plate was centrifuged at 3,100 g for 15 minutes.

Congo red and Calcofluor white assays were carried out similarly[187] but with final concentrations of Congo red and Calcofluor white at 25 mg/L and 15 mg/L, respectively.

MAPK signaling pathway knockout

Essentiality score was based on a previously described dataset which used a CRISPR-Cas9 genome-wide knockout library to investigate gRNA drop-out after transformation.[188] Twenty genes in the signaling pathway with the highest normalized enrichment score were chosen, and 0.75 non-essentiality score was selected to be the cut-off, generating eight genes for knockout target. Non-essentiality score was defined as the adjusted *p*-value of gRNA drop-out, so that closer the *p*-value is to 1, the less essential the gene is. Gene knockouts were carried out as previously described.[189]

Genomic integration of galactosylation machinery

Fusion protein constructs were codon optimized, synthesized (IDT), and cloned into a custom vector for integration at intergenic locus near *GQ67_4099*. Genome editing was carried out similar to a previous method[112] with 250ng of CRISPR plasmid and 2.5μg of linear DNA.

Exoglycosidase digestion

ÄKTA-purified K3 peptide samples were buffered exchanged, using Amicon® Ultra (3,000 molecular weight cutoff) centrifuge columns (Millipore), and concentrated in water. Protein concentration was determined by absorbance at A280 nm (DeNovix). Glycan digestion by α1-2,3-mannosidase, α1-6-mannosidase, and β1-4-galactosidase S (New England Biolabs) was carried out according to manufacturer protocol with minor changes. Briefly, 8 μg of protein was digested in 1x buffer with 4 μL of each exoglycosidase of interest, in total volume of 30 μL. After overnight incubation at 37°C, 120 μL cold acetone was added to each reaction. Following 1-hour incubation at -20°C, proteins were precipitated by centrifugation at 15,000g for 15 minutes at -10°C. The supernatant was decanted, and the air-dried protein pellets were resuspended in 5% acetone in water with 0.1% formic acid for intact protein LC-MS analysis.

5. ENGINEERING STRAINS FOR GLYCOSYLATING PROTEINS OF INCREASED COMPLEXITIES

In the previous chapter, we demonstrated homogeneous G0 glycosylation on a simple peptide. Therapeutically relevant proteins are however much more complex, and proper glycosylation of these molecules can be challenging. In this chapter, we discuss the interplay between humanized glycosylation pathway and yeast native protein folding and glycosylation machineries. We observed a negative correlation between protein complexity and desired G0 glycan abundance on the secreted product. Through genetically probing many aspects of the native and heterologous pathways, we identified glycosylation site accessibility as the main contributing factor to diminished humanized glycosylation activity. Promiscuous activity by Golgi-resident mannosyltransferases likely also has detrimental effects, but our finding suggests that more extensive re-engineering of the native machineries would be required to meaningfully impact glycosylation landscape.

5.1. Background and Motivation

Glycosylation is a vital post-translational modification due to its large impact on pharmaceutical protein efficacy and safety.[190] Yeasts such as *K. phaffii* have emerged as a class of favorable alternative hosts, but their use in glycoprotein production is hindered by their native hypermannosylation, which, in addition to possible structural disruptions, can lead to unwanted immunogenicity and fast clearance.[119] Human plasminogen Kringle 3 domain-derived peptide (K3) has served as a good reporter molecule for past glycoengineering efforts,[59,112,128] but therapeutically relevant proteins are usually much bigger and more complex than simple peptides. A prominent example is monoclonal antibodies (mAbs) – the dominant pharmaceutical protein by innovation and by sales.[1] A mAb molecule is a heterotetramer, with two light chain and two heavy chain molecules linked together by disulfide bridges, and requires glycosylation in the constant Fc (crystallizable fragment) for protein stabilization and therapeutic functions.[191] Even in mammalian hosts such as CHO, the rate of mAb folding in the endoplasmic reticulum is relatively low, and folding/assembly machineries in the ER have been the target for engineering.[192] Disulfide bridge formation, in particular, can be rate-limiting and lead to the accumulation of incorrectly folded/assembled molecules and ER-associated degradation.[193] In yeasts like *S. cerevisiae* and *K. phaffii*, secretory pathway engineering, including overexpression of soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs), heat shock transcription factors (Hsf), and protein chaperones, and knockout of native proteases.[191]

Not only is protein folding important in the context of product quality, it also has inextricable ties with protein glycosylation. In the ER, oligosaccharide transferase (OST) complex transfers the Glc3Man9 glycan onto the newly synthesized peptide, and after processing by glycosidases, the resulting GlcMan9 glycan is recognized by two ER-resident lectin

chaperones calreticulin and calnexin to facilitate protein folding. Additional ER mannosidase also plays a role in protein quality control by diverting unfolded proteins (with prolonged residence time in the ER) to ER-associated protein degradation.[194]

In addition to the ER, the humanized glycosylation pathway also involves the engineering of Golgi apparatus. There is notably redundancy among yeast Golgi glycosylation enzymes, suggesting that many of these proteins perform similar or overlapping functions.[195] As a result, the exact functionality of each enzyme is not well-characterized in yeasts, and even less so in *K. phaffii*. Knockouts of a subset of native enzymes have been combined with heterologous pathway integration to clean up high-mannose structures, often a combination of α-mannose, β-mannose, and phosphomannose transferases.[54]

Thus lies the challenge for glycosylation pathway engineering, which spans both ER and Golgi apparatus – in the ER, disruption of the native pathway could interfere with protein folding and quality control machinery, while in the Golgi, the organelle is primed for hypermannosylation by many native mannosyltransferases. In this chapter, we show the difference in the performance of humanized glycosylation pathway for expressing more complex proteins than K3 peptide. We detail our engineering strategies to improve pathway processivity, including altering enzyme localization and expression strength, increasing gene dosage, introducing helper chaperones, and knocking out native mannosyltransferases. Despite the limited impact of most of these changes on the glycan profile of a commercial monoclonal antibody, but we identified key areas that likely require further engineering interventions.

5.2. Results

5.2.1. Construction of a dual-expression vector for monoclonal antibodies

In Chapter 4, we achieved homogeneous G0 glycosylation on K3 peptide by either using a spontaneously occurring *MNS1* mutant or attenuating the expression of wildtype *MNS1*. To assess how viable G0 strains would perform with more complex proteins, we used trastuzumab, a commercial monoclonal antibody with one N-linked glycosylation site on its heavy chain, for molecular demonstration. The expression and secretion of trastuzumab in *K. phaffii* required the sequential integrations of its light chain and heavy chain (Chapter 2), which was challenging and time-consuming in glycoengineered strains due to their slow growth and decreased transformation efficiency. We were ultimately able to accomplish the double transformation in the strain expressing *GNT1, MNS2,* and *GNT2* with P_{ENO1} and *MNS1* M260I mutant with P_{GAPDH}. After cultivating this strain and analyzing the glycan profile of the secreted monoclonal antibody by LC-MS, we observed that the amount of G0 glycan was very limited compared to aglycosylated and hypermannosylated fractions (Figure 5.1).

To facilitate the utilization of monoclonal antibodies as a readout for glycosylation pathway engineering, we constructed a dual-expression vector, whose single transformation enables co-expression of light chain and heavy chain (Figure 5.1B). We included Glu-Ala-Glu-Ala (EAEA) sequence at the N-termini of trastuzumab light chain and heavy chain due to the significant amount of N-terminal extension observed. We assessed the glycan profile of trastuzumab heavy chain secreted from the same glycoengineered strain and did not observe any significant changes from altering protein expression vector (Figure A5.1).
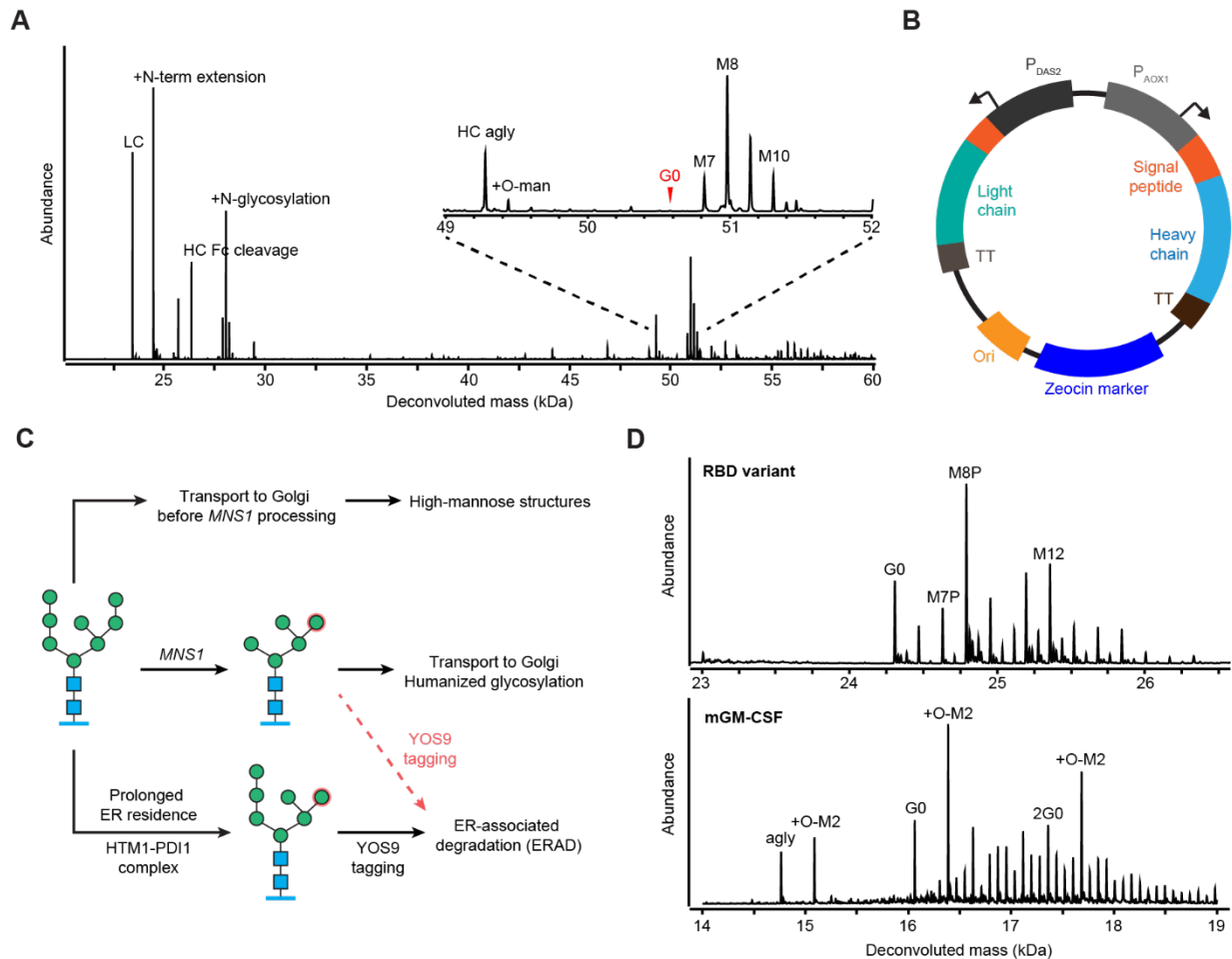
Figure 5.1. First-pass expression of complex proteins in glycoengineered strains.
(A) Expression of trastuzumab (with intact N-linked glycosylation site in its heavy chain Fc
domain) in a glycoengineered strain. LC-MS analysis showed minimal G0 glycosylation. (B)
Design of a dual expression vector for trastuzumab (and other IgG's) expression. (C) Proposed
mechanism of the interaction between heterologous glycosylation pathway and yeast native
ERAD. (D) Expression of a SARS-CoV2 receptor binding domain (RBD) variant and murine
granulocyte macrophage colony-stimulating factor (mGM-CSF) in a glycoengineered strain.

5.2.2. Decrease in G0 abundance with increasing protein complexity

The most significant differences between reporter molecules K3 and trastuzumab include their sizes and consequently their folding requirements within the ER. K3 is a simple soluble peptide, essentially without any need for folding machinery recruitment in the secretory pathway. In contrast, trastuzumab is a much bigger and more complex molecule, whose production in higher eukaryotes requires chaperones and therefore much longer residence time in the ER.[193] While most of the heterologous glycosylation enzymes (*GNT1*, *MNS2*, and *GNT2*) are targeted to the Golgi apparatus, *MNS1* was localized to the ER with *S. cerevisiae* Mns1 localization sequence. This choice is supported by prior reports because Man8, the substrate for *MNS1* processing, is the diverging point between yeast native glycosylation and human complex glycosylation.[59] By localizing the commitment step to the humanized glycosylation pathway to the ER, we could preemptively disrupt yeast native hypermannosylation, which largely resides in the Golgi apparatus. However, recent work in *S. cerevisiae* has reported an ER quality control mechanism for glycosylated proteins, which would suggest this localization strategy be less than ideal (Figure 5.1C).[194] Normally, in the ER of a yeast cell with intact glycosylation pathway, folded glycosylated proteins with Man8 glycan would be transported to the Golgi apparatus for mannosylation. If a protein is misfolded or forming aggregates, however, it would be detected by the Htm1-Pdi1 complex, which cleaves off another mannose, forming Man7 structure and exposing the terminal α1,6-mannose. This terminal mannose is recognized by Yos9p, which tags the glycosylated protein for ER-associated degradation (ERAD). In the engineering glycosylation pathway, *MNS1* is localized to ER and trims the Man8 structure to Man5, which has a similarly exposed terminal α1,6-mannose that can be tagged by Yos9p. In the case of K3 peptide, its minimal folding requirement means short residence time in the ER, and the

anterograde transport to Golgi likely dominates over Yos9 tagging for ERAD. However, monoclonal antibodies such as trastuzumab would prove a much bigger challenge for yeast folding machinery. The propensity for misfolding and aggregation due to the lack of evolved specialized chaperones is compounded by the trimming of mannoses and the exposure of α1,6-mannose by the heterologous Mns1p. The export of correctly folded and properly glycosylated antibody molecules could be outcompeted by the hyperactive ERAD system, greatly decreasing the amount of Man5-glycosylated trastuzumab for further humanized glycosylation processing in the Golgi. Antibody molecules that are aglycosylated or not modified by Mns1p have relatively lower chance of degradation, resulting in the enrichment of undesired glycan structures present in the glycan profile.

To test this hypothesis, we first tried expressing two other proteins in the glycoengineered strains – a SARS-CoV-2 receptor binding domain (RBD) variant[52] (with one N-linked glycosylation site) and murine granulocyte macrophage colony-stimulating factor (mGM-CSF, with two N-linked glycosylation sites).[127] The complexity of these two molecules fall between κ3 peptide and trastuzumab, with well-defined tertiary structure but without interpeptide interactions or multiple disulfide bridges. These two proteins were expressed in the same glycoengineered strain with *MNS1* mutant, and LC-MS analysis of the secreted product showed that mGM-CSF exhibited largely homogeneous G0 glycosylation, while RBD, the larger of the two additional molecular demonstrations, had significantly more hypermannosylated structures (Figure 5.1D). This serves as evidence for our hypothesis that the extended ER residence time required by more complex molecules did increase the fraction of hypermannosylated product, likely due to the degradation of Man5 glycoproteins through ERAD.

In mGM-CSF, we also observed O-linked glycosylation, presented as a mass shift consistent with two O-linked mannoses. To better visualize the degree of N-linked hypermannosylation and humanized glycosylation, we digested the purified secreted proteins with α1-2,3,6-mannosidase, which has catalytic activity against all terminal α-mannoses. This enzymatic digestion allows us to distinguish between O-linked mannosylation, N-linked hypermannosylation, and N-linked humanized glycosylation. All yeast native O-linked mannoses can be cleaved off by this enzyme, while N-linked humanized glycans are unaffected because G0 glycans should have terminal N-acetylglucosamines, and N-linked hypermannosylation are cleaved to M1GlcNAc2 structure because other than the final mannose, which is linked to the GlcNAc2 core by a β1-4 linkage, all other mannoses in yeast native glycosylation are α-mannoses. The resulting mass spectrum (Figure A5.2) showed predominant G0 glycosylation, and like trastuzumab, glycan occupancy of mGM-CSF proved to be an issue as less than 50% of the secreted molecules had both N-linked glycosylation sites occupied. RBD, however, did not exhibit any occupancy issues, as the vast majority of the secreted molecules were glycosylated. The difference in glycan occupancies can likely be explained by the accessibility of the glycosylate site, and future studies could include closer examination of protein crystal structure to confirm this theory.

We also examined if *YOS9* knockout can improve glycan processing by reducing the tagging of glycoproteins with the exposed α-1,6-mannose for degradation. Prior work in *S. cerevisiae* demonstrated the viability of this genomic change and its impact on glycoengineering, especially disruption in the *ALG* pathway.[196] We identified a possible *YOS9* homolog in *K. phaffii* based on homology, but, to the best of our knowledge, functional studies of this gene do not exist. After its knockout in G0-glycoengineered trastuzumab-secreting strain, the secreted mAb

titer did not improve, and no significant change in the glycosylation profile was observed. We also attempted *YOS9* knockout in G0 strain secreting RBD variant but were not successful in obtaining a positive transformant. It is possible that *K. phaffii* has an alternative, unreported system for tagging protein for ER-associated degradation, because the identified homolog has no significant homology to Yos9 dimerization domain (Figure A5.3). It thus remains to be explored in future studies the exact ERAD recognition system employed in *K. phaffii*.

5.2.3. Different subcellular localization of *MNS1*

We aimed to resolve the unexpected degradation of desired product by targeting *MNS1* to later in the secretory pathway. Based on previous literature, we selected two additional localization sequences – *S. cerevisiae* Sec12, involved in the formation of COPII vesicle for ER to Golgi transport, and *S. cerevisiae* Van1, a component of mannan polymerase I.[128] These two localization sequences likely target *MNS1* to late ER/early Golgi and Golgi, respectively, based on their protein functions. Changing the subcellular localization would logically alter not only the required expression level for sufficient glycan processing but also the tolerance level of maximum expression. To reflect this change, we decided to combine three different promoters (P_{MNN4}, P_{MNN10}, and P_{CDA2}, although the difference in their activities is likely smaller than expected, per Chapter 4.2.5), two different catalytic regions (wild type and the M-to-I mutant), and the two localization sequences, resulting in twelve new *MNS1* constructs. Through MS analysis of K3 peptide secreted by these strains, we quickly ruled out the M-to-I mutants, because the majority of K3 molecules were hypermannosylated, likely due to the reduced enzymatic activity (Figure A5.4). Wildtype *MNS1* constructs, however, when targeted to later in the secretory pathway (especially late ER/early Golgi) and expressed under weaker constitutive

promoters, still had enough enzymatic activity to efficiently cleave Man8 to Man5 for further downstream humanized glycosylation to G0 (Figure 5.2A).

We then expressed trastuzumab in these G0 strains with differently targeted *mns1* to examine if decreasing mannosidase presence in ER would increase the fraction of correctly glycosylated heavy chain. Through intact protein LC-MS, while these strains secreted homogeneously G0-glycosylated K3 peptide, the mass spectra of trastuzumab heavy chain revealed that there was little to no detectable G0 decoration (Figure 5.2B). Whether or not these differently targeted *mns1* constructs have alleviated ERAD remains to be studied, but moving *MNS1* to later in the secretory pathway could have opened up possibilities for native promiscuous mannosyltransferases to modify the intermediate glycan structures in the humanized pathway. We also investigated an alternative *MNS1* construct using the catalytic domain from *M. musculus* homolog with *S. cerevisiae SEC12* localization sequence.[128] The G0-glycoengineered strain expressing this *MNS1* construct under strong constitutive promoter P_{GAPDH} produced predominantly G0 glycan on K3 peptide, but did not produce significant amount of desired G0 glycan on the heavy chain of trastuzumab.
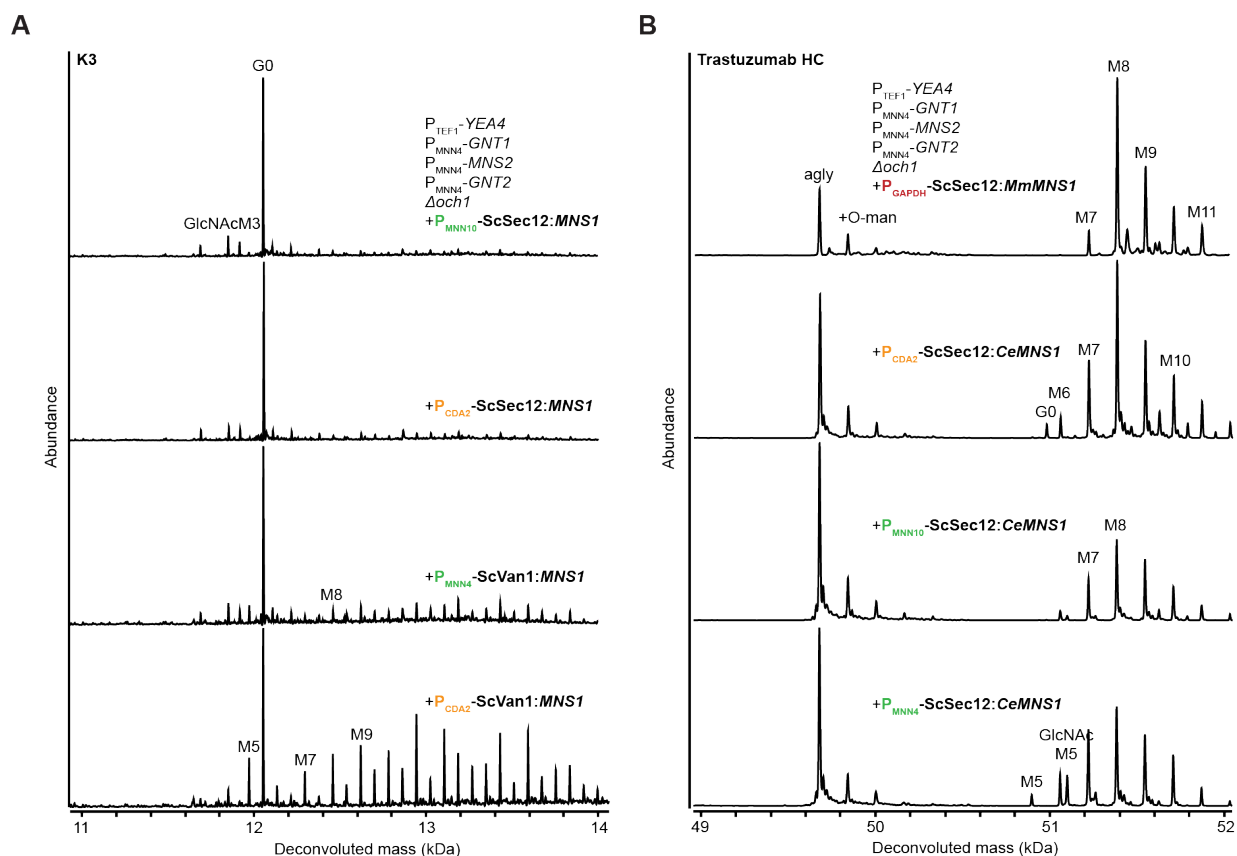
Figure 5.2. Effects of subcellular localization of *MNS1* on secreted protein glycan profile. (A) K3 peptide and (B) trastuzumab heavy chain intact protein LC-MS analysis. Targeting *MNS1* to later in the secretory pathway increases high-mannose structures on K3 peptide. We only observed incremental improvements in mAb HC G0 abundance with different *MNS1* subcellular localization.

5.2.4. Introduction of protein folding chaperones

An alternative approach to facilitate faster anterograde transport from ER to Golgi is to help accelerate protein folding so that Man5-decorated mAb molecules exit ER before interacting with the ERAD system. We thus integrated human calnexin and human calreticulin, two ER-resident lectin-based chaperones that can selectively bind to the glucose residue in GlcMan9 structure,[194] under the control of *K. phaffii* native *CNE1* promoter, the homolog of human calnexin. We chose this promoter to mimic yeast native gene regulation for protein chaperones. We transformed the two constructs encoding these two chaperones into the G0 strain expressing the *M. musculus MNS1* construct and analyzed the secreted K3 peptide and trastuzumab via LC-MS (Figure 5.3A). The G0 fraction on trastuzumab heavy chain minimally improved after the integration of calnexin and calreticulin. For K3 peptide, the amount of hypermannosylated glycans increased after chaperone integration. The integration of these two human lectin chaperones could have introduced additional cellular burden (resources for both their expression and their folding), and the output G0 glycan abundance is likely determined by the balance between additional ER stress from chaperone expression and the reduced protein degradation from the benefits conferred by the same chaperones. K3 is a simple peptide with little folding requirement, so cost of chaperone expression could outweigh the benefits. On contrary, trastuzumab requires much more support for its folding, so the integration of calnexin and calreticulin likely promotes the interaction between recombinant protein and yeast native chaperones, speeding up its processing and reducing clearance by ERAD.

The improvement of G0 glycosylation on trastuzumab heavy chain was very limited with the introduction of calnexin and calreticulin, possibly because yeast native chaperones are not

fully equipped to help fold complex proteins like mAbs. In human ER, there are other

chaperones such as protein disulfide isomerase (*PDI*), peptidyl-prolyl cis-trans isomerase (*PPI*),

and a number of oxidoreductases that act synergistically with calnexin and calreticulin.[197] We

thus built constructs of human disulfide isomerase *PDI1* and peptidyl-prolyl isomerase *CypB*

under the control of promoters of their respective *K. phaffii* homologs *PDI1* and *CPR1*. After

integrating these constructs into G0 strains (with Sec12-targeted *MNS1*) with and without human

calnexin and calreticulin and analyzing the glycan profiles of secreted trastuzumab, we observed

minimal improvement in G0 abundance only with the introduction of *H. sapiens CypB* in the G0

strain without human calnexin and calreticulin integration, while the integration of either

chaperone did not elicit measurable response in lectin chaperone-engineered strain (Figure 5.3B).

We thus concluded that the synergy between calnexin and calreticulin and the additional

chaperones was insufficient in changing the glycan profile. Future studies could extend

chaperone engineering to include additional helper proteins such as the oxidoreductase co-

chaperones to facilitate redox homeostasis in the ER.
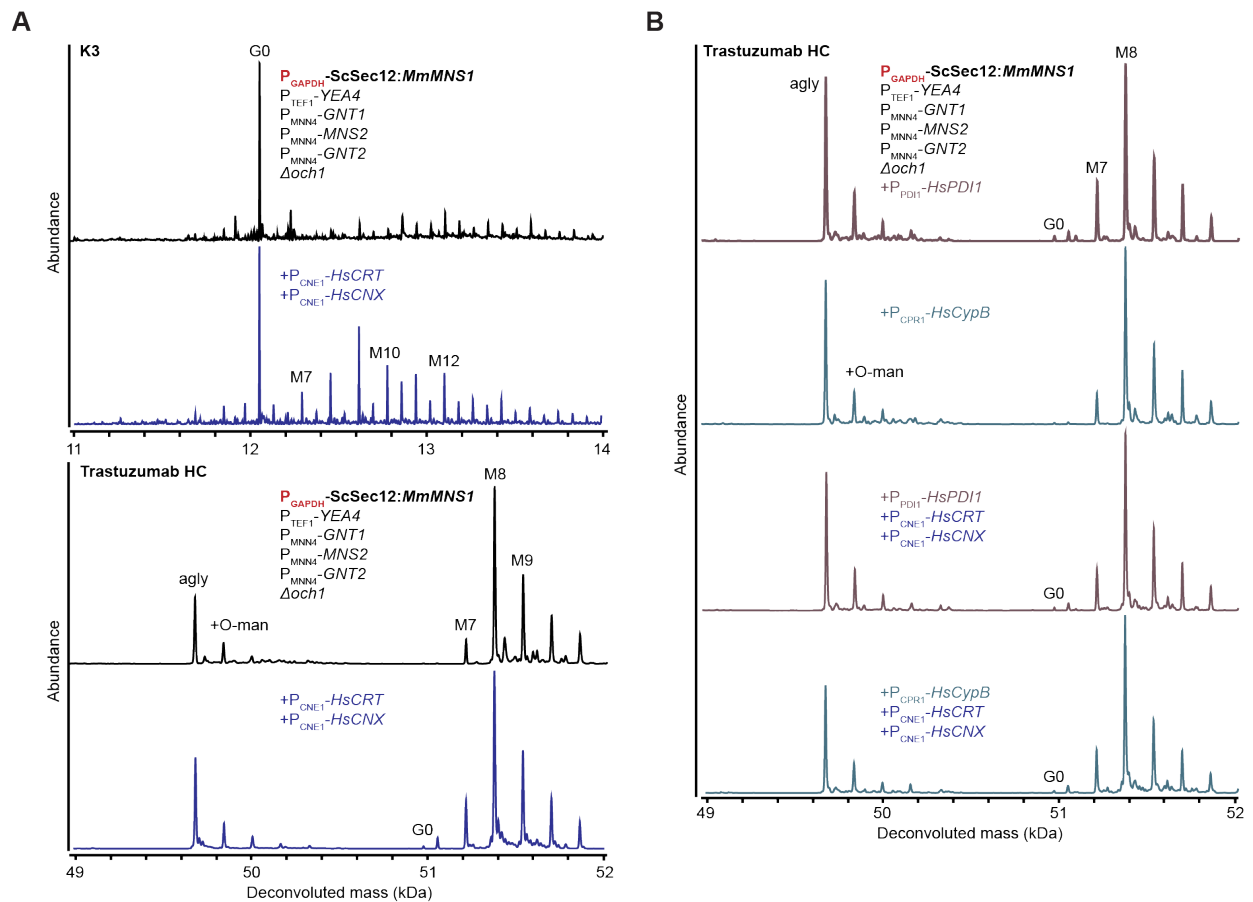
Figure 5.3. Effects of chaperone engineering in glycoengineered strains.
(A) Mass spectra of K3 and trastuzumab intact protein LC-MS analysis. Calnexin and calreticulin integration had detrimental effects on K3 glycosylation but yielded a small improvement with trastuzumab. (B) Mass spectra of trastuzumab HC with the integration of two additional chaperones – protein disulfide-isomerase and peptidyl-prolyl cis-trans isomerase.

5.2.5. Gene dosage of heterologous *MNS1*

Roll-in integration is a commonly used technique in *K. phaffii* for recombinant protein expression, because multiple copies of the same construct can be integrated at the same time to potentially maximize transcript expression level. This strategy is rarely used in heterologous gene expression, however, due to the inability to precisely control copy number. We decided to explore this strategy for upregulating *MNS1*. The cleavage of Man8 to Man5 glycan is often regarded as the commitment step in upper eukaryotes glycosylation, so increasing *MNS1* abundance in the secretory pathway could help divert more flux through the heterologous pathway. We thus integrated four different *MNS1* roll-in constructs – wildtype *C. elegans MNS1* with either *S. cerevisiae* Mns1 or Sec12 localization sequence expressed under P_{CDA2}, mutant *C. elegans MNS1* with *S. cerevisiae* Mns1 expressed under P_{GAPDH}, and *M. musculus MNS1* with *S. cerevisiae* Sec12 expressed under strong constitutive promoter P_{GAPDH}. Out of the four constructs attempted, we were unable to obtain any positive transformants for the two wildtype *C. elegans MNS1* constructs. This is unsurprising given the observed cytotoxicity of strongly expressed wildtype *CeMNS1* (Chapter 4.2.1). LC-MS analysis of secreted trastuzumab showed that roll-in integration of *CeMNS1* mutant or *MmMNS1* constructs generated strains that showed mannosidase activity, compared to the parent strain with no *MNS1* integration (Figure A5.5). However, the benefits of additional copies of either *mns1* construct were not apparent, as there was no significant improvement of G0 glycan fraction over strains with only single *MNS1* construct integrated into the genome.

5.2.6. Accessibility of N-linked glycans for mannosidase modification

The observed inverse relationship between G0 glycan abundance and protein complexity can also be explained by possible steric hindrance that limits the accessibility of *MNS1* to the

105

glycosylation site. All aforementioned *MNS1* constructs use various transmembrane domains as their localization sequence, making them membrane-bound proteins. Although the transmembrane domains contain linker sequence and, in some cases, flexible tails on the ER lumenal side, membrane bound *MNS1* could still possibly be physically constrained from accessing the glycosylation sites on more complex proteins. This hinders the proper initiation of humanized pathway, resulting in the observed decrease in desired G0 glycan abundance.

In yeast, the short peptide sequence HDEL is a well-known ER-retention tag,[198] so we extended the C-termini of existing *MNS1* constructs to include this sequence for ER-localization. The N-terminal localization sequences, composed of a short cytosolic tail and a single-pass transmembrane domain, of *MNS1* constructs were replaced with the pre-region of αSP, a 19-amino acid sequence that directs translocation into the ER. Four different HDEL-tagged *MNS1* constructs were transformed into a trastuzumab-expressing G0 strain as an additional copy. Two out of the four tested constructs (P_{MNN4}-*CeMNS1* and P_{BCK1}-*CeMNS1*) showed a shift in the glycan profile of trastuzumab, evident by the significant increase of Man5 glycan fraction (Figure 5.4A). This enhanced mannosidase activity may result from the combined effect of the additional gene copy and less steric hindrance of ER-lumenal Mns1p. To examine if only using a single copy of HDEL-tagged *MNS1* would achieve the same result, we transformed the two *mns1* constructs that had tested positive into G0Δ*mns1* strain and analyzed the glycan profile of the expressed trastuzumab. Only one of the two constructs tested, P_{MNN4}-*CeMNS1* with the stronger constitutive promoter, yielded a significant amount of Man5 glycan on trastuzumab heavy chain (Figure 5.4B). This suggests that the previously observed increase in mannosidase activity indeed was the result of the combined effect of an extra copy of *mns1* and the likely decrease in steric hindrance.
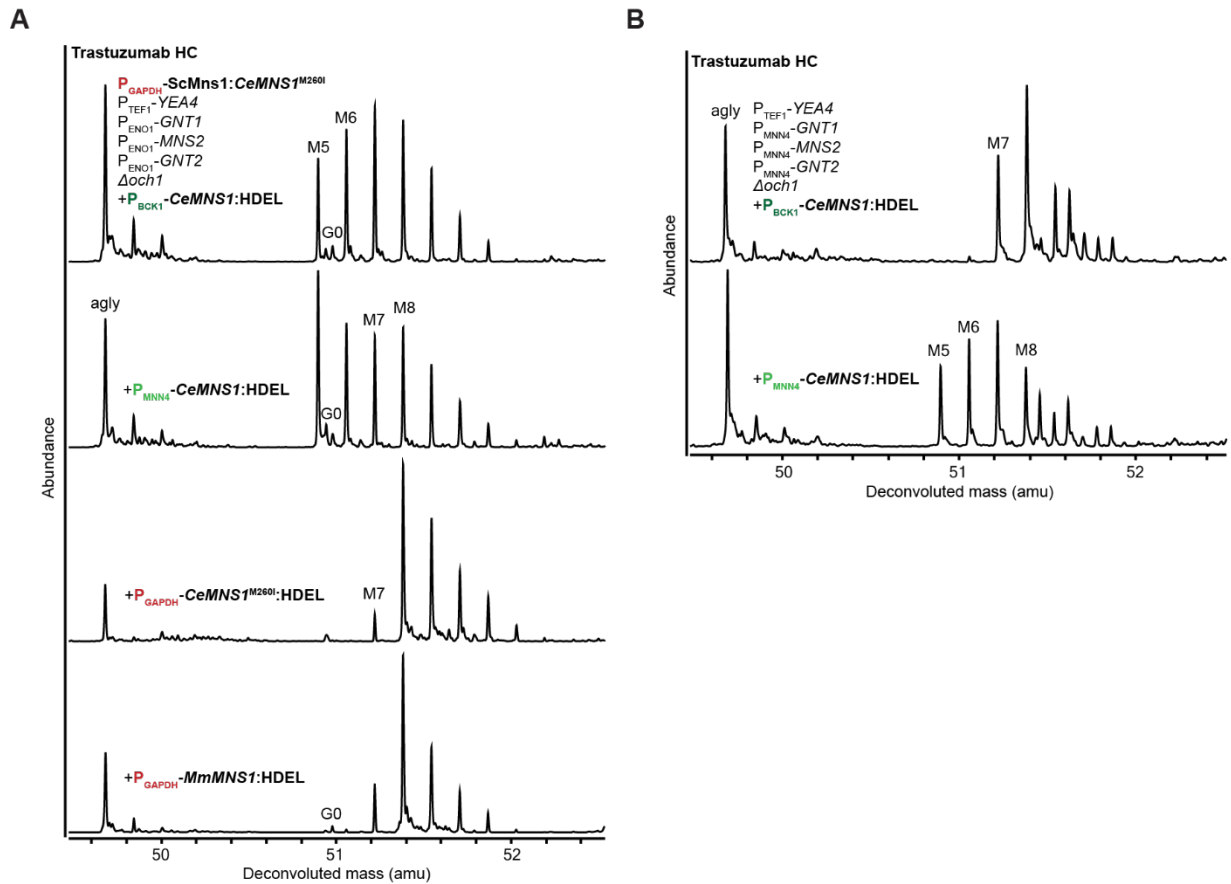
Figure 5.4. Effects of ER lumen-targeted *MNS1* constructs.
(A) An additional copy of HDEL-tagged *MNS1* resulted in modulated glycan structure on trastuzumab HC, as analyzed by intact protein LC-MS. (B) Decoupling the effects of copy number and ER lumen targeting. P_{MNN4}, the stronger promoter of the two, showed mannosidase activity against high mannose structures on trastuzumab HC, as analyzed by LC-MS.

5.2.7. Different subcellular localization of *GNT1*

In humanized glycosylation pathway, *MNS1* processing is often regarded as the key engineering step because it is the commitment step in higher eukaryotes.[40] Nevertheless, unlike higher eukaryotes, yeasts have many mannosyltransferases, some forming mannan polymerase complexes, residing in the Golgi apparatus for hypermannosylation of secretory proteins.[199] Even after proteins with Man8 glycan have been successfully trimmed down to Man5 by heterologous Mns1p and transported to Golgi apparatus, they could still act as the target substrate of promiscuous Golgi-resident mannosyltransferases. The kinetic competition between these promiscuous activities and the desired GlcNAc addition by *GNT1* dictates the subsequent glycosylation reactions and ultimately the glycan profiles of secreted proteins. One approach of diverting protein flux towards *GNT1* is by altering its subcellular localization, giving the heterologous enzyme an advantage by having the opportunity to act on Man5 glycan before any Golgi mannosyltransferases. The existing *GNT1* constructs uses *S. cerevisiae* Mnn2 transmembrane domain to localize the enzyme to Golgi. Thus, similar to the strategy we employed for *MNS1*, we tested two additional localization sequences for *GNT1* – *S. cerevisiae* Mns1 transmembrane domain (for ER localization) and *S. cerevisiae* Sec12 transmembrane domain (for late ER/early Golgi localization). We transformed *GNT1* constructs with these alternative localization sequences into G0Δ*gnt1* strains with different *MNS1* constructs. Any combination of *MNS1* and *GNT1* where *GNT1* localization precedes that of *MNS1* was excluded due to resulting incorrect order of biochemical reactions in the glycosylation pathway. We then transformed trastuzumab in the generated strains and analyzed the glycan profile of secreted proteins. All *GNT1* constructs integrated into the G0Δ*gnt1* strain using the mutant version of ER-localized *CeMNS1* yielded heavy chain G0 peak that was below detection limit by mass

spectrometry, regardless of *GNT1* localization (Figure A5.6A). In G0 strain expressing co-localized *MNS1* and *GNT1* in late ER/early Golgi, we did not observe a significant increase in G0-decorated heavy chain (Figure A5.6B). Additional chaperone engineering (the integration of two human lectin chaperones, calnexin and calreticulin) also did not meaningfully change the glycan profile. The lack of significant improvement with differently localized *GNT1* shows that native promiscuous mannosyltransferases remain kinetically dominant over the native pathway, so modulation of these competitive pathways is warranted.

5.2.8. Knockout of Golgi-resident mannosyltransferases

The knockout of MNN4 and MNN6, two genes involved in *K. phaffii* native phosphomannosylation, has been used to reduce the degree of hypermannosylation in glycoengineered *K. phaffii* strains.[54,176] To investigate if similar knockouts in our G0 glycoengineered strains would reduce the degree of hypermannosylation, we knocked out MNN4 and MNN6 in RBD- and mGM-CSF-expressing strains. The resulting strains secreted these recombinant proteins with less extensive phosphomannosylation, but the overall fraction of correctly G0-glycosylated molecules was not positively affected by the knockouts (Figure 5.5A).

We then examined if knockouts of Golgi-resident mannan polymerase subunits can modulate glycan profiles of mAbs. Mannan polymerase typically acts on the outer chain of N-linked glycans, whose elongation is disrupted by Δ*och1*, but any promiscuity by this complex could still act on mannose moieties in humanized glycosylation pathway intermediates. We targeted three different genes, *MNN9*, *ANP1*, and *VAN1*, for knockout, based on their relative low essentiality based on a previously described library screen,[188] in G0 strains expressing trastuzumab or RBD. By LC-MS analysis, however, all of these single knockouts did not significantly impact glycan profiles of these two proteins (Figure 5.5B, Figure A5.7). This shows

that more extensive engineering of Golgi-resident glycosylation machinery is likely needed to properly modulate the activity of native promiscuous enzymes. Past work has demonstrated viable additional knockouts of native β-mannosyltransferases, and future work could include CRISPR-Cas9-aided multiplexed knockouts for faster strain generation and testing.
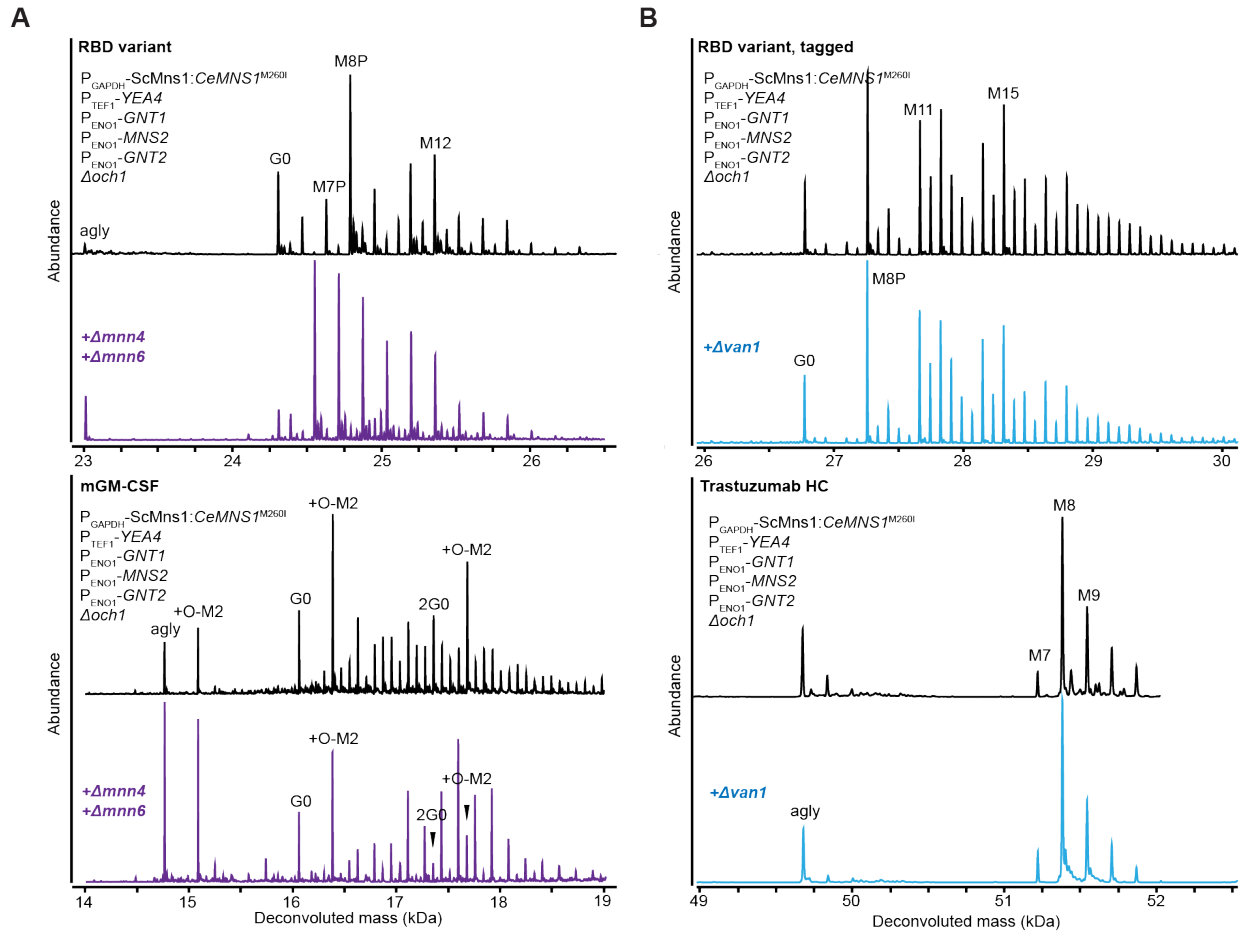
Figure 5.5. Effects of Golgi-resident mannosyltransferase knockout.
(A) Mass spectra of RBD and mGM-CSF in G0 glycoengineered strains before and after Δ*mnn4* and Δ*mnn6*. Note the decrease in phosphomannosylation after gene knockouts. (B) Mass spectra of RBD and trastuzumab HC before and after Δ*van1*. Degree of hypermannosylation on RBD shows a slight decrease, but similar effects are minimal with trastuzumab.

5.2.9. Combination library of *MNS1* and *GNT1*

Most of the engineering strategies described so far are one-off approaches, and despite the potential benefits of diagnosing the exact bottleneck, single changes are likely not sufficient in generating meaningful changes in complex pathways such as glycosylation. To address this issue, we decided to co-modulate the first two genes, *MNS1* and *GNT1*, in a library format to examine its impacts on the glycan profile of secreted trastuzumab.

We included many of the representative constructs tested thus far (Table B5.1) in this library and ligated different *MNS1* constructs with *GNT1* constructs for the single-locus integration of both genes (Figure 5.6A). We then assessed the library diversity by transforming the ligation reaction into *E. coli* and picking single colonies for Sanger sequencing confirmation. We observed that 5 (out of 11) *MNS1* and 2 (out of 5) *GNT1* constructs were under-represented in the combined library (Figure A5.8A). We thus performed a second round of ligation reactions to enrich these under-represented constructs. In the 40 single *E. coli* colonies assayed in the second round of library construction, the success rate of integration cassette PCR amplification was much lower than the first round (approximately 25%), which could negatively impact our ability to infer library diversity. Combining the results of first and second rounds, there are two *MNS1* constructs that were completely absent in the library – *CeMNS1* targeted to late ER/early Golgi with ScMns1 localization sequence driven by P_{MNN4} and P_{MNN10} (Figure A5.8C). The dropout of these constructs likely would not affect the effective library coverage, as the difference in promoter strength could be smaller than expected (Chapter 4.2.5).

We decided to transform this combined library into a trastuzumab-expressing G0 strain, because prior results have showed that increasing gene dosage can have beneficial effects in

glycan modulation (Chapter 5.2.6). The knock-in efficiency, determined by PCR amplifying the integration locus, was low, likely due to the size of integration cassette. After cultivating positive clones and analyzing the glycan profiles of secreted trastuzumab, we identified several possible hits, whose integration cassettes were then verified via Sanger sequencing. By integrating another copy of *MNS1* and *GNT1*, we were able to make incremental improvements in G0 glycan abundance (Figure 5.6B). Most significant changes were observed in strains with the extra copy of *MNS1* being retained in the ER with C-terminal HDEL tag, but despite the enrichment in trimmed-down Man5 glycan, the co-integration of an additional copy of *GNT1* did not generate more desired G0 glycan. In some positive hits, we only detected integration of *MNS1* – this observation, combined with low integration frequency due to cassette size, demonstrates the need for a plasmid-based system for library screens like this. Such systems have been reported, using either a *Pichia*-specific autonomously replicating sequence[200] or centromeric DNA[201] for propagation in *K. phaffii*, but are in practice sub-optimal due to challenges in plasmid upkeep or plasmid size.
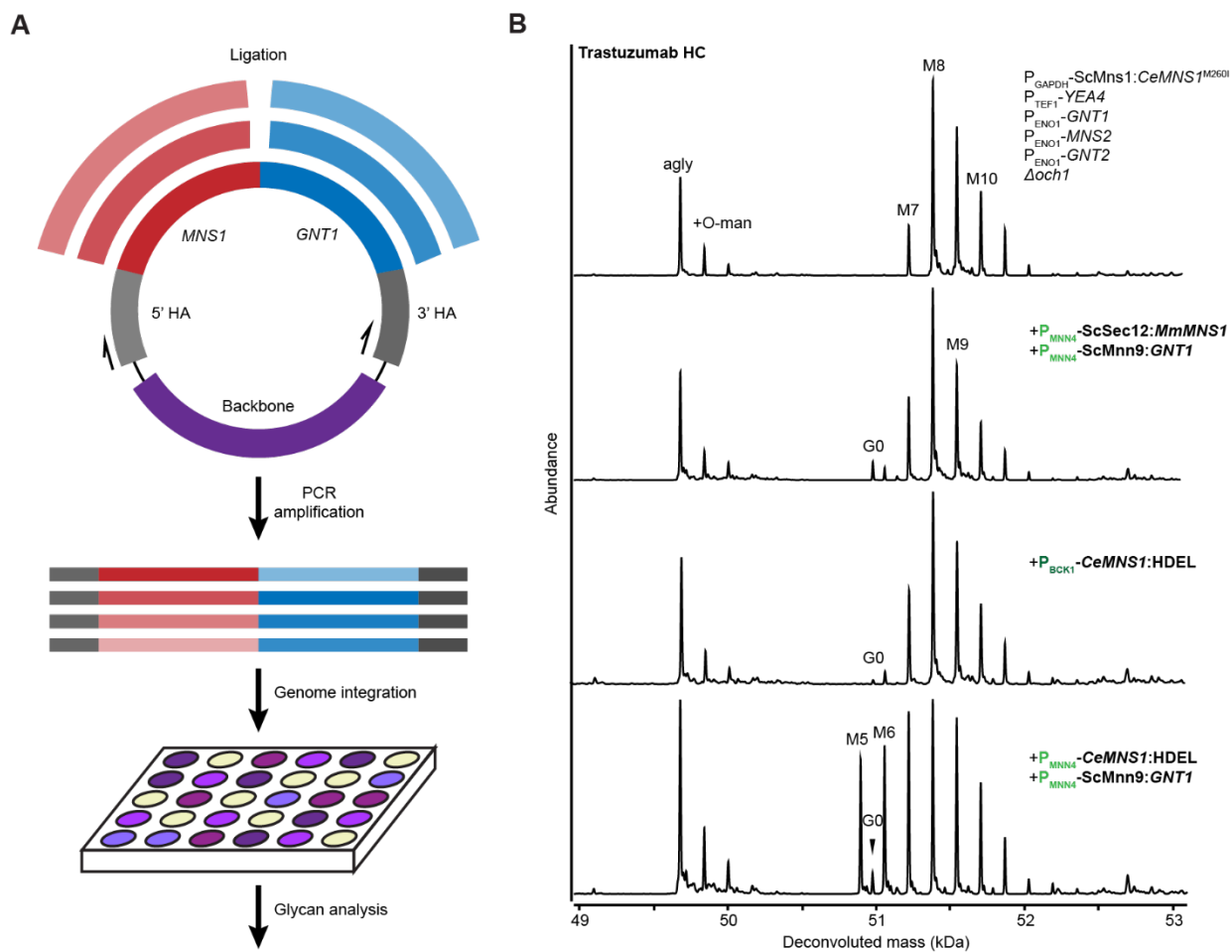
Figure 5.6. Combined *MNS1-GNT1* library for glycan modulation.
(A) Process schematic for library construction. (B) Positive hits identified in the library screen. The identity of integration constructs was verified through Sanger sequencing after differences in intact protein LC-MS were observed.

5.3. Discussion

In this chapter, we detailed our engineering strategies in adapting the glycoengineered strains for the production of more complex proteins, especially a commercial monoclonal antibody. We demonstrate the inverse relationship between desired G0 glycan abundance and secreted protein complexity, using human plasminogen K3 peptide, murine GM-CSF, SARS-CoV2 RBD variant, and trastuzumab as molecular demonstrations. We focused most of our engineering efforts on *MNS1*, the first gene in the humanized glycosylation pathway. Through altering its subcellular localization, catalytic domain, and gene dosage, we demonstrated that for modifying complex enzymes, membrane-bound mannosidase likely encounters steric hindrance, which limits its accessibility to buried N-linked glycans, like those in the trastuzumab Fc region. Instead of a membrane anchor for its ER-localization, *MNS1* showed enhanced activity with C-terminal HDEL tag, a yeast-specific ER retention tag. Altering the localization of *GNT1* with different transmembrane domains had limited impact on glycan profile, however, against the hope that capping intermediate Man5 glycan to form GlcNAcMan5 would preclude further modification by native promiscuous mannosyltransferase activity. We also attempted re-engineering the secretory pathway by introducing ER-resident chaperones, reducing protein recognition by ERAD, and knocking out select Golgi-resident (phospho)mannosyltransferases, but these changes did not significantly improve trastuzumab G0 abundance.

Through modulating different aspects of the humanized glycosylation pathway, we were able to gain valuable insights on its key challenges and crucial bottlenecks. While the existing pathway is capable of glycosylating simple, soluble peptide, extensive additional engineering is required to achieve homogeneous humanized glycosylation on complex proteins. The accessibility issue of N-linked glycosylation sites needs to be resolved, as many of the

membrane-bound enzymes in the humanized pathway likely have reduced activity due to steric hindrance. Longer linker sequences between the catalytic and the transmembrane domains can be considered, but prior reports, using K3 peptide as the reporter molecule, suggest that they underperformed compared to their short counterparts.[128] Additionally, although HDEL retention tag outperforms other ER membrane tethers, similar retention tags for Golgi, to the best of our knowledge, have not been reported. Thus, strategies to localize enzymes to Golgi apparatus while providing them maximum flexibility to carry out their biochemical reactions should be investigated in future studies.

Furthermore, re-engineering of the *K. phaffii* secretory pathway likely requires a combined approach with targets on multiple fronts, ranging from protein folding and redox balance in the ER to mannosyltransferase knockouts and membrane homeostasis in the Golgi. Although the one-off diagnostic strategy employed here was only able to achieve incremental improvements, combining them could create significant impact on therapeutic protein glycosylation.

5.4. Methods

Yeast strains

All strains were derived from wildtype *Komagataella phaffii* (NRRL Y-11430). Genes containing human plasminogen K3 peptide, murine granulocyte macrophage colony-stimulating factor, SARS-CoV2 receptor binding domain variant, and trastuzumab were codon optimized, synthesized (Integrated DNA Technologies), and cloned into a custom vector. *K. phaffii* strains for protein secretion were transformed as described previously.[112] CRISPR-Cas9 genetic perturbation was carried out as described previously[112] with some modifications – 250 ng CRISPR plasmid and 2.5 μg linear DNA were used for integration.

Cultivation

Strains for glycan profile characterization were grown in 3 mL culture in 24-well deep well plates (25°C, 600 rpm). Cells were cultivated in BMGY (buffered glycerol complex media, Teknova) with 4% v/v glycerol. For glycoengineered strains, after 40h of biomass accumulation, cells were pelleted and resuspended in BMMY (buffered methanol complex media, Teknova) with 3% v/v methanol. After 48 hours of production, supernatant was harvested for analysis.

Protein purification and glycan characterization

Untagged RBD was purified as previously described.[202]

For His-tagged proteins (K3 peptide, mGM-CSF, and RBD), supernatant was filtered and diluted 1:1 with Ni-IMAC (nickel-immobilized metal affinity chromatography) binding buffer (25mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4). Purification was carried out on the GE ÄKTA pure system with 1-mL HisTrap columns (Cytiva). After sample loading, the column was equilibrated with and washed with the described binding buffer before eluting with elution buffer (500mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4).

For trastuzumab, the filtered supernatant samples were diluted, 1:1 with 1x phosphate-buffered saline (PBS, pH 7.4) and purified on the GE ÄKTA pure system with a 1-mL ProA column (HiTrap Protein A HP, Cytiva). After sample loading, the column was equilibrated with and washed with 1x PBS and eluted with 100mM citric acid (pH 2.8). Eluted products were pH adjusted with 1M Tris-HCl (pH 9.0).

Intact protein LC-MS was performed as described previously.[52] Mass spectra were processed using MassHunter Bioconfirm software (Agilent Technologies) with a deconvolution

range of 10-15 kDa (for K3), 10-20 kDa (for mGM-CSF), 20-35 kDa (for RBD), or 10-60 kDa (for trastuzumab) using a mass step of 0.5 or 1 Dalton.

Exoglycosidase digestion

Purified protein samples were buffered exchanged, using Amicon® Ultra centrifuge columns (Millipore), and concentrated in water. Protein concentration was determined by absorbance at A280 nm (DeNovix). Glycan digestion by α1-2,3,6-mannosidase (New England Biolabs) was carried out according to manufacturer protocol. After overnight incubation at 37°C, four times sample volume cold acetone was added to each reaction. Following 1-hour incubation at -20°C, proteins were precipitated by centrifugation at 15,000g for 15 minutes at -10°C. The supernatant was decanted, and the air-dried protein pellets were resuspended in 5% acetone in water with 0.1% formic acid for intact protein LC-MS analysis.

Combined library construction and verification

Different *MNS1* and *GNT1* constructs (Table B5.1) were PCR amplified with overhangs for ligation. Equimolar of *MNS1* constructs and *GNT1* constructs were pooled for ligation with Hifi assembly (New England Biolabs). Ligation mixture was transformed into DH5α (NEB) for library diversity assessment. Individual *E. coli* colonies were randomly picked, and their colony PCR product was sent to Sanger sequencing (Azenta) to confirm the identity of *MNS1* and *GNT1* pairing.

6. OPTIMIZING PROCESS CONDITIONS FOR INCREASED GLYCAN HOMOGENEITY

In addition to host engineering, process engineering has also been proven successful for modulating product quality, especially in CHO processes. In this chapter, we examined how humanized glycosylation pathway in glycoengineered yeast strains respond to process changes, especially cultivation temperature and media composition. We used a machine learning-guided approach to optimize media additives and enriched the G0 abundance on a subunit protein vaccine candidate. We show the relative invariance of secreted protein glycan profile to cultivation environment, which hints at the robustness of glycoengineered *K. phaffii* fermentation.

6.1. Background and Motivation

Komagataella phaffii is one of the common microbial hosts for recombinant protein expression, and there is growing interest in wider utilization of this host to produce pharmaceutical proteins. It shares similar advantages with bacterial hosts, including inexpensive media, fast growth, and facile genetic manipulation, while also reaping the advantage of having a secretory pathway that allows for protein folding and post-translational modifications.[25] Nevertheless, compared to mammalian hosts such as Chinese Hamster Ovary (CHO) cells or common murine myeloma cell lines, proteins secreted from *K. phaffii* often face quality challenges such as proteolysis and incorrect glycosylation.[203] While these issues can be addressed through molecular engineering[51,80,100] and cell line engineering,[54,109,120] modulating the production process is another strategy to improve protein quality.

Indeed, process development and optimization has been extensively studied in CHO cultivations due to the industry's preference for the host. This can include both tuning process parameters and changing media composition because of their influence on host metabolism and post-translational modifications.[204] Charge heterogeneities, from asparagine deamidation and aspartate isomerization, have been reported to be correlated to pH and carbon dioxide tension.[205] The media osmolarity and trace metal presence have also been linked to protein aggregation and/or fragmentation.[204] Monoclonal antibody glycosylation, in particular, has been the target of many studies,[205–207] because mAb has become a commercially successful platform molecule that can be adapted for many disease indications.[19]

Similar studies have been conducted in *K. phaffii*, albeit to a lesser extent. Temperature and pH has been shown to impact host proteolytic activity,[208] and different modes of operation (fed-batch versus continuous) and media development have also been investigated.[74,167] Strain

engineering can also be employed to facilitate changes in bioprocesses – our lab has reported inducible protein production without methanol.[209] For controlling product quality that involves non-native PTMs like glycosylation, genetic engineering is required before any process development can be employed to improve process output.

In this chapter, we assessed G0 strains expressing different proteins for their performance under different cultivation conditions. We observed that carbon source concentration and temperature impact protein secretion titer, but these factors do not significantly change glycan profiles. Furthermore, we examined how different media supplementations, many of whose impacts have been demonstrated in CHO cultures, affected protein glycosylation. By employing a machine learning-guided algorithm, we were able to efficiently search the design space and improve the desire glycan abundance on a SARS-CoV2 receptor binding domain variant.

6.2. Results

6.2.1. Effects of culture temperature

Creating an optimal environment for heterologous enzyme functionality can be beneficial to efficient pathway engineering. In the humanized glycosylation pathway, the catalytic domain of many enzymes come from warm-blooded upper eukaryotes – *MNS2* comes from *R. norvegicus*, *GNT2* comes from human, and a subset of *MNS1* constructs come from *M. musculus*, in addition to the human chaperones that were included in several G0 strains. *K. phaffii* cultures are typically carried out at ambient temperature (approximately 25°C), which could be kinetically suboptimal for directing flux through the humanized pathway. We thus tested 30°C as an alternative cultivation temperature to examine if glycan composition would be affected. Conversely, reducing cultivation temperature could facilitate the folding of large heterologous enzymes and recombinant proteins.[210] We thus also included cultivation at 20°C to explore how

slower growth rate would affect recombinant protein glycan profile. We tested trastuzumab-expressing strains with multi-copy *MNS1* integration and strains with chaperone engineering and observed that cells grew much faster at 30°C than at 20°C. Thus, to avoid flocculation due to overgrowth, the outgrowth and production phases at 30°C were limited to 24 hours each, while those at 20°C were allowed 48 hours each. SDS-PAGE analysis showed that cultivation at 20°C yielded higher titer, but this could be confounded by the difference in cultivation length (Figure 6.1A). Cultivations at 30°C also did not yield enough material for purification in many cases. Furthermore, although we had seen no significant differences in the trastuzumab heavy chain glycan profiles between chaperone engineered G0 strain, the addition of calnexin and calreticulin seemed to boost the secreted protein titer, an effect more noticeable at 20°C. We then analyzed the glycan profile of heavy chain via LC-MS. Many of the strains, post cultivation at 30°C, did not produce enough materials for purification. Compared to a typical cultivation at 25°C, the mass spectra of trastuzumab heavy chain showed no significant differences in glycan profile when the strains were cultivated at 30°C or 20°C (Figure 6.1B). Thus, changing the enzymatic activity on a global scale through varying cultivation temperature was not beneficial to generating more G0 glycans, likely because the enzyme kinetics for both native and heterologous pathways were likely affected by temperature. For producing monoclonal antibodies, and possibly other complex proteins as well, cultivation at 30°C could be largely infeasible due to the significant drop in productivity.
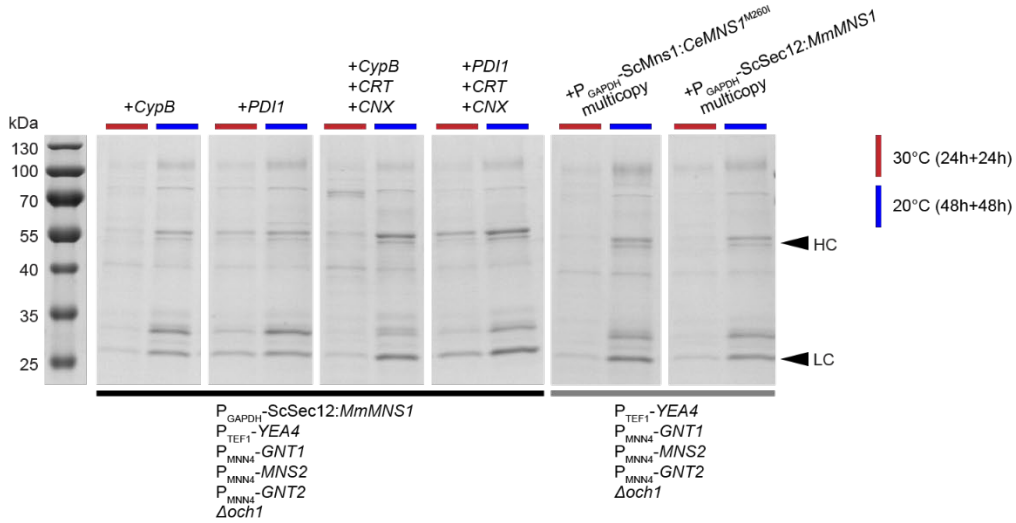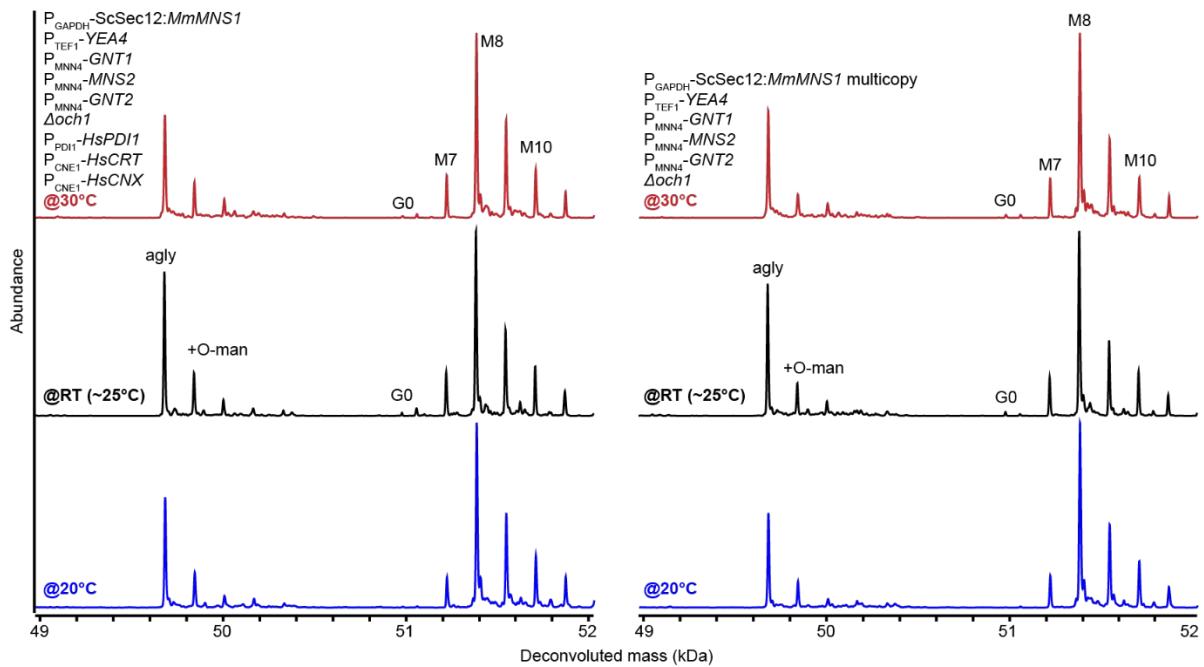
Figure 6.1. Temperature effects on secreted protein glycan profiles.
(A) SDS-PAGE analysis of trastuzumab in different glycoengineered strains in 30°C vs. 20°C cultivation. (B) Mass spectra of trastuzumab HC when cultivated at different temperatures. Cultivation temperature does not have a significant impact on trastuzumab glycan profile.

6.2.2. Effects of carbon source concentration

Recombinant protein production in *K. phaffii* is typically driven by P_{AOX1}, a tightly controlled promoter repressed by glycerol and glucose and induced by methanol. Methanol concentration has been linked to the expression level of recombinant proteins,[52] and reducing the flux of recombinant proteins could be beneficial to mediating ER stress response and prolonging the interaction time between them and heterologous glycosylation enzymes.

We analyzed the glycan profiles of K3 peptide, RBD, and trastuzumab in several G0-glycoengineered strains under four different production phase conditions – 0.5% or 3% v/v methanol with and without 4% w/v sorbitol. Sorbitol is common cofeed carbon source in *K. phaffii* cultivations and serves as a non-repressive carbon source for protein production.[211] By SDS-PAGE analysis, the additional of sorbitol had different effects on different conditions (Figure 6.2A). For K3 peptide and RBD, cofeeding the production phase with sorbitol greatly decreased the titer of secreted proteins, in some cases below purification limit. However, for trastuzumab, the addition of sorbitol showed similar trend with 0.5% v/v methanol but the reverse with 3% v/v methanol. By comparing the glycosylation patterns of secreted K3 peptide and trastuzumab between different conditions via LC-MS, we observed that the inclusion of sorbitol in production media reduced the degree of phosphomannosylation on K3 peptide but had no noticeable effect on trastuzumab (Figure 6.2B). Future studies could include examining whether adding sorbitol to media could downregulate the expression of phosphomannosyltransferases in *K. phaffii*. We were unable to obtain a sufficient amount of RBD-J6 through purification to perform similar comparative analysis because its titer was too low under cofed cultivation conditions.
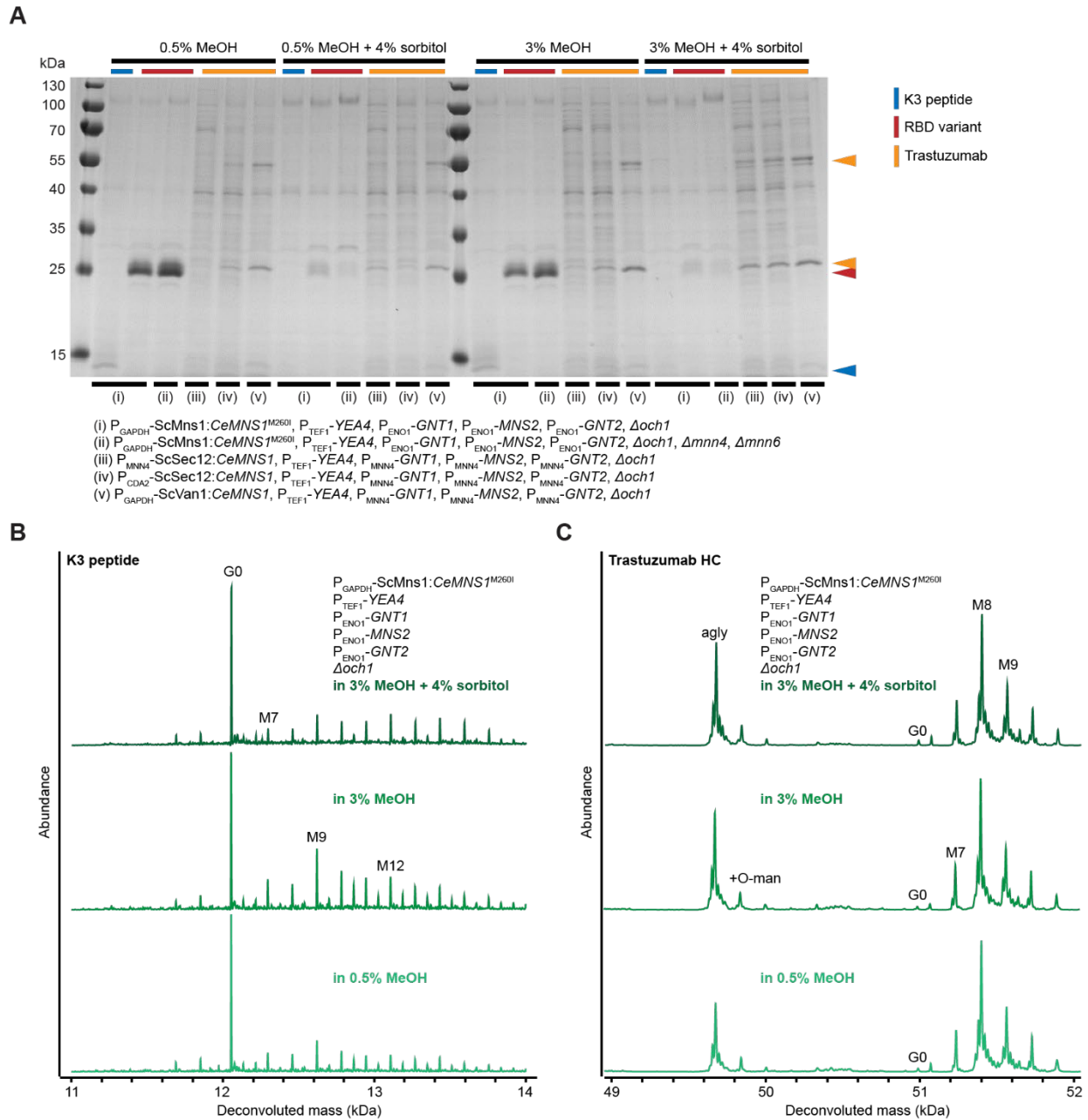
Figure 6.2. Carbon source concentration effects on secreted protein glycan profiles.
(A) SDS-PAGE analysis of K3 peptide, RBD, and trastuzumab in G0 strains with production phase media containing different concentrations of carbon source. (B) Mass spectra of K3 peptide and trastuzumab HC in different production media. Sorbitol cofeed reduces the amount of phosphomannoses on K3 peptide, but no significant impact of carbon source concentration was observed for trastuzumab HC glycosylation.

6.2.3. Effects of media supplementation

Glycosylation control of monoclonal antibodies in CHO fermentation has been extensively studied, of which media supplementation is a widely used strategy for modulating their glycan profiles. Divalent ions (especially Mn^{2+}) supplementation, as the cofactor ion of oligosaccharide transferase (OST), have been reported to help enhance its activity and improve glycan occupancy.[212] Glycosylation reaction precursors like galactose, GlcNAc, and N-acetylmannosamine (ManNAc) help push the biochemical reaction equilibrium towards a more favorable direction and reduce the metabolic burden required for glycosylating the recombinant protein molecules.[206,213] Bases like ammonia and uridine are also key supplementations in CHO culture and impact N-linked glycosylation.[207,213] Proper exploration of the design space for the optimization of media supplements, however, is a multi-variate optimization problem. Standard One-Factor-At-a-Time (OFAT) and Design-Of-Experiment (DOE) approaches can be labor intensive, possibly constrained by local optima, and in practice infeasible due to the lack of easy, high-throughput screening method for glycan profiles of recombinant proteins. Modern machine learning-based methods can overcome these challenges and have been successfully applied in both media composition and process parameter optimization.[65,66]

We here chose a Bayesian optimization algorithm to accelerate the screening of media supplements. It is well-suited for optimizing the "black box" function where the underlying relationship between the input and output is unknown. In short, Bayesian optimization suggests experiments sequentially using a surrogate model that mimics the system of interest based on the experiments observed previously. Due to this property of adaptively sampling locations, the algorithm is capable of achieving optimal conditions faster and with reduced number of overall experiments.[214] We selected the G0 strain with mutant *MNS1* as the base glycoengineered strain

126

and examined how different media supplementations would change the glycan profiles of three different glycoproteins – K3 peptide, RBD, and trastuzumab. We chose the testing range for each supplement based on prior CHO literature (Table B6.1) and generated an initial set of 12 conditions (Table B6.2) to sample uniformly from the design space for round one study. After cultivation, the secreted proteins were purified, and their glycan structures were analyzed by LC-MS. For K3 peptide, in all 12 media supplementation conditions, G0 remained the predominant glycan structure, and different conditions did not significantly change the abundance of high mannose structures (Figure A6.1A). K3 peptide had already been an easy-to-glycosylate molecule, so it is not surprising that adding supplements did not drastically further improve the glycan profile. For trastuzumab, G0 abundance remained minimal – between 0.5% and 1% of trastuzumab light chain abundance – and fell beneath the threshold of the peak calling algorithm in a subset of conditions (Figure A6.1B). Because a significant fraction of trastuzumab heavy chain was aglycosylated, we also examined the ratio between aglycosylated and Man8-glycosylated heavy chain and surprisingly observed a decrease in occupancy in many media supplementation conditions. We observed the biggest effect of media supplementation with RBD – the ratio of G0 to Man11, a common high-mannose structure observed via LC-MS, improved from 0.56 to as high as 1.03 with media supplementation (Figure A6.1C).

Based on the results of the round one study, we decided that K3 peptide was only suitable for base-line measurements and would not meaningfully further contribute to media supplementation optimization. We thus excluded K3 peptide from the study and included mGM-CSF instead. mGM-CSF has two N-linked glycosylation sites, and we have previously observed glycan occupancy issues on this molecule (Chapter 5.2.2). Because RBD did not present any occupancy problem, mGM-CSF would be a great candidate to study how this issue can be

improved with media supplementation. In addition, we chose to increase the exploration range of Mn^{2+} in the round two study, because CHO-informed testing range did not take into account of high cell densities of yeast cultures.[215]

We then carried out two sequential rounds of media supplementation optimization to maximize G0 glycan abundance in RBD, each with 11 different conditions (Table B6.2), and analyzed the resulting glycan profiles of RBD, mGM-CSF, and trastuzumab. In Round 3, calcium ion, the enzyme cofactor for *MNS1*, was added into the model to examine its impact on the heterologous pathway activity. After three rounds of media supplementation testing (Figure 6.3A), we were able to increase the abundance of G0 glycosylation on RBD. Supplementation condition 18 performed the best, resulting in noticeable increase in G0 abundance and decrease in hypermannosylated structures (Figure 6.3B). Simple linear regression is insufficient in relating normalized G0 score to different media additives, as only galactose showed a statistically significant, albeit weak, positive correlation (Figure 6.3C). This shows that the Bayesian optimization was able to capture non-linear relationships among the variables. It is worth noting, however, that media osmolarity could be another inherent variable with different additives. For rigorousness, future studies should include investigating how media supplementation changes osmolarity and its impact on secreted protein glycan profiles.

Because G0 glycan abundance on RBD was set as the optimization target, we did not observe improvements in trastuzumab or mGM-CSF occupancy in the resulting media supplementation conditions (Figure A6.2). The optimized media supplements are thus not applicable for other glycoproteins, and new optimization workflow would be required for new molecules.
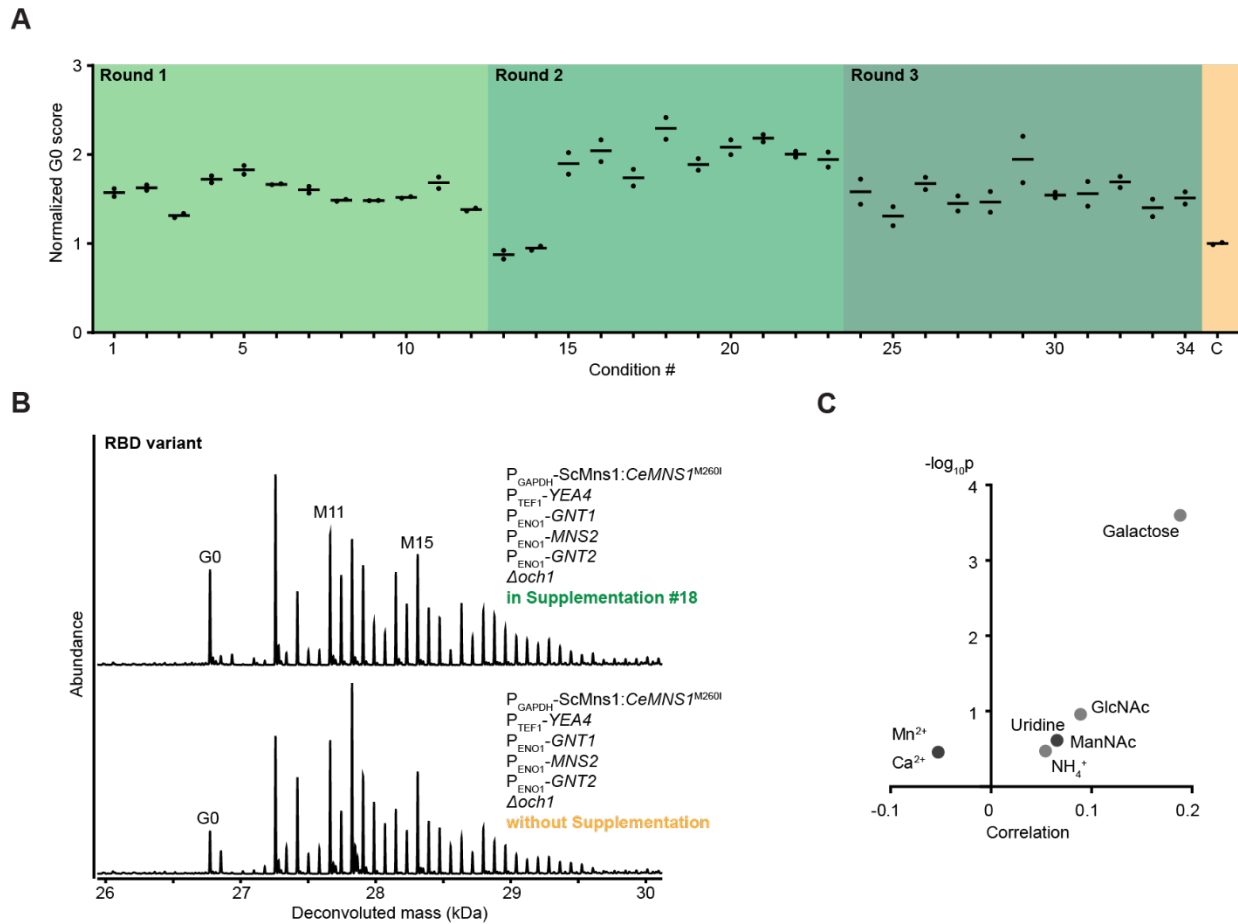
Figure 6.3. Machine learning-guided algorithm to improve G0 abundance on secreted RBD.
(A) Normalized G0 score (ratio of G0 to Man11 abundance) across 34 tested conditions, as compared to control condition without any media additives. (B) Mass spectra of RBD in the best-performing condition and in control. Note the enrichment of G0 glycan and the decrease in hypermannosylated structures. (C) Correlation-*p*-value volcano plot of normalized G0 score to different media additives. Correlation was estimated based on a linear regression model.

6.3. Discussion

In this chapter, we demonstrated the modulation of glycosylation profiles on secreted proteins through varying culture conditions. By combining host engineering and process engineering, we interrogated how the multi-step humanized glycosylation pathway responds to environmental changes. We showed that temperature, despite being a common process variable with great impact on glycan profile in CHO cultures, did not significantly affect the performance of glycoengineered strains. Altering media composition, both carbon source and additives, changed the glycosylation pattern on several molecular demonstrations, including K3 peptide, RBD variant, and murine GM-CSF. We employed a machine learning-guided algorithm to better investigate the design space for media supplementation and significantly improved the G0 abundance on RBD.

One major observation we made in this chapter is the relative invariance of humanized glycosylation pathway towards changes in process parameters, and, to a lesser degree, media composition. This observation is also corroborated by literature as well.[216] The robustness of *K. phaffii* fermentation is a direct contrast from CHO, where small changes in process parameters or media micronutrients can significantly change glycosylation landscape.[206,217,218] It thus follows that strains exhibiting desired glycan profiles are, to some extent, process agnostic, making potential scale-up from bench-scale to pilot- and commercial scale simpler. However, any undesired glycans would also likely persist, and process changes have relatively small impacts. In our case, although we were able to decrease their abundance, high-mannose structures at N-linked glycosylation sites, a result from the combined effect of limited glycan accessibility and Golgi mannosyltransferase promiscuity, still plague the best-performing media supplementation condition. This points to the need for additional genome engineering to better address this issue.

Lastly, the development of better and faster screening technologies would greatly benefit the engineering cycle. Currently, commonly used methods include released glycan analysis (with high-performance liquid chromatography and/or matrix-assisted laser desorption ionization-time of flight mass spectrometry)[219] or intact protein mass spectrometry. Released glycan analysis often cannot accurately measure glycan occupancy, while intact protein mass spectrometry can be confounded by other PTMs (including O-linked glycosylation and phosphorylation), and additional interventions such as enzymatic digestion are needed to decouple mass shifts from different sources. High-throughput methods for glycan identity and abundance characterization would aid in screening more conditions and strains, and, with more data, predictive models could be more accurate in optimizing host engineering and process tuning for a wide range of glycoproteins.

6.4. Methods

Yeast vectors and strains

All strains were derived from wildtype *K. phaffii* (NRRL Y-14430). The DNA fragment containing different recombinant proteins was codon optimized, synthesized (Integrated DNA Technologies), and cloned into a custom vector for roll-in integration. All yeast strains were derived from wild-type *Komagataella phaffii* (NRRL Y-11430). *K. phaffii* strains were transformed as described previously.[112]

Cultivation

Strains for glycan profile characterization were grown in 3 mL culture in 24-well deep well plates (25°C, 600 rpm). Cells were cultivated in BMGY (buffered glycerol complex media, Teknova) with 4% v/v glycerol. For glycoengineered strains, after 40h of biomass accumulation, cells were pelleted and resuspended in BMMY (buffered methanol complex media, Teknova)

with 3% v/v methanol. For mAb cultivations and media supplementation, 10 mM glutathione was added to the production media. After 48 hours of production, supernatant was harvested for analysis.

Protein purification and glycan characterization

Untagged RBD was purified as previously described.[202]

For His-tagged proteins (K3 peptide, mGM-CSF, and RBD), supernatant was filtered and diluted 1:1 with Ni-IMAC (nickel-immobilized metal affinity chromatography) binding buffer (25mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4). Purification was carried out on the GE ÄKTA pure system with 1-mL HisTrap columns (Cytiva). After sample loading, the column was equilibrated with and washed with the described binding buffer before eluting with elution buffer (500mM imidazole, 25mM sodium phosphate, 500mM sodium chloride, pH 7.4).

For trastuzumab, the filtered supernatant samples were diluted, 1:1 with 1x phosphate-buffered saline (PBS, pH 7.4) and purified on the GE ÄKTA pure system with a 1-mL ProA column (HiTrap Protein A HP, Cytiva). After sample loading, the column was equilibrated with and washed with 1x PBS and eluted with 100mM citric acid (pH 2.8). Eluted products were pH adjusted with 1M Tris-HCl (pH 9.0).

Intact protein LC-MS was performed as described previously.[52] Mass spectra were processed using MassHunter Bioconfirm software (Agilent Technologies).

Exoglycosidase digestion

Purified protein samples were buffered exchanged, using Amicon® Ultra centrifuge columns (Millipore), and concentrated in water. Protein concentration was determined by absorbance at A280 nm (DeNovix). Glycan digestion by α1-2,3,6-mannosidase (New England

Biolabs) was carried out according to manufacturer protocol. After overnight incubation at 37°C, four times sample volume cold acetone was added to each reaction. Following 1-hour incubation at -20°C, proteins were precipitated by centrifugation at 15,000g for 15 minutes at -10°C. The supernatant was decanted, and the air-dried protein pellets were resuspended in 5% acetone in water with 0.1% formic acid for intact protein LC-MS analysis.

Bayesian optimization of media supplementation

Bayesian optimization, which applies a sequential procedure in which a surrogate Gaussian process is used to suggest the next set of experiments based on already acquired data, optimizes the trade-off between exploration and optimization. The implementation of the model was carried out as previously described,[214] with RBD G0 abundance as the optimization target.

7. DISCUSSION AND FUTURE OUTLOOK

In this thesis, we presented *K. phaffii* as an alternative host organism for producing pharmaceutical proteins to improve the accessibility of these medicines, address broad societal goals for sustainability, and offer financial advantages for accelerated development of new products. We demonstrated the engineering of molecular sequence, host genome, and process variables to improve the quality of secreted recombinant proteins, especially aglycosylated and glycosylated monoclonal antibodies.

7.1. Design-for-Success with Deeper Understanding of Secretory Pathway

Chapter 2 of this thesis discussed the adaptation of aglycosylated monoclonal antibodies for improved production in *K. phaffii*. By making small, conservative, informed modifications to the IgG1 backbone, we significantly reduced product-related variants and achieved product quality comparable to that in CHO, the industry gold standard. The molecular sequence changes introduced in this chapter are broadly applicable to a wide range of recombinant proteins, and our lab have since demonstrated proof-of-concept quality improvement with milk proteins. Admittedly, although we were successful in improving the quality of secreted mAbs, additional functional studies would be needed to ensure protein activity. In fact, risks of reduced or abolished activity are the main deterrent against molecular sequence engineering. In the case of monoclonal antibodies, we can hedge our bets based on approved commercial molecules like eptinezumab, but many other classes of biologics do not have such precedents. It is thus prudent to design molecules with production host and potential quality issues in mind. This necessitates deeper understanding of host biology, especially the secretory pathway. Luckily, *K. phaffii*, compared to common mammalian platforms, has a much smaller genome and simpler, more robust biology.[27] Aided by modern techniques like gene editing, high-throughput sequencing, and

other -omics techniques, a better mechanistic understanding of protein translation, translocation, folding, transport, modifications, and eventual secretion can potentially pinpoint problem areas in the molecular sequence and suggest strategies for their modifications.

7.2. Faster Pathway Tuning with Titratable Promoters

Chapters 3 and 4 of this thesis discussed the implementation of a CRISPR-Cas9 system in *K. phaffii* for improving product quality, especially non-human glycosylation. We demonstrated the engineering of yeast native glycosylation pathway to produce homogeneous G0 glycan on a reporter peptide. Although the feasibility of glycoengineering has been demonstrated before, we intended to leverage CRISPR-Cas9 technology for faster pathway manipulation. The advantage of speed was evident in the multiplexed engineering of glycosylation pathway, achieving three genomic changes at three different loci. We also discovered that, however, engineering a multistep pathway with such intricate ties with other aspects of cellular functions was more complicated than we initial thought – viable glycoengineered strains were only possible with a spontaneous gene mutant or tuning down gene expression level to a host-appropriate level. Heterologous pathway optimization is not a new problem in synthetic biology. In fact, multi-step pathways are very common in metabolic engineering,[220,221] and many tools have been developed in conventional hosts like *E. coli* and *S. cerevisiae* for precise tuning of gene expression levels.[56,222] The utilization of *K. phaffii* as a host for similar metabolic engineering has driven innovation in developing novel promoters for pathway optimization, but the scope has been limited to mutagenizing common promoters P_{AOX1} and P_{GAPDH}.[223] We have started the work of identifying native promoters that are responsive to different carbon sources and additives in the media and demonstrated the regulation of cell growth using an inositol-repressible promoter, but additional assay is needed to confirm their titratability. Development and characterization of such

promoters and other gene regulation systems would accelerate the engineering cycle of glycosylation pathway and help inform gene expression to achieve glycan homogeneity.

7.3. Data-Driven Host Engineering

The importance of host engineering guided by different -omics techniques should not be understated. Chapter 4 of this thesis, in particular, used bulk RNA sequencing data to inform an appropriate expression level for heterologous pathway. After creating viable glycoengineered strains, similar techniques were employed to compare transcriptomes between different strains and identify engineering targets to address growth defects. Including the host cell biology into consideration can and should be employed in strain engineering. In glycoengineering, non-native sugar moieties, such as galactose and sialic acid, need to be *de novo* synthesized for glycans beyond G0, and proper balance of carbon flux could achieve homogeneous glycosylation without compromising cell fitness. In a broader context, genome engineering of the host can also be more effective when guided by model and data.[224,225]

Furthermore, transcriptomic analysis highlighted areas for improvement in the experimental design. Specifically, heterologous gene expression is likely impacted by surrounding genes, and, consequently, the dynamic range with which we were able to modulate expression level is narrower than expected. Our lab has previously reported the interrogation of different intergenic regions for genomic integration via ATAC-seq,[184] but the scope was limited to a single construct of a fluorescent protein driven by a highly active promoter. Based on the results presented in this thesis, further characterization of integration loci is warranted to identify landing pads for heterologous genes that can be regulated consistently and across several orders of magnitude in expression level. This underscores the importance of comprehensive -omics analysis in reliable genome editing.

7.4. High-Throughput Method for Screening Complex Phenotypes

Chapters 5 and 6 discussed our engineering strategies to improve the production of desired glycans on more complex proteins. We observed an inverse correlation between protein complexity and the abundance of G0 glycosylation in secreted product. Chapter 5 aimed to address this issue by engineering the heterologous pathway and different secretory machineries. We identified accessibility as the main challenge in properly glycosylating larger, more complex proteins, especially those with buried N-linked glycosylation sites, as is the case for monoclonal antibodies. Furthermore, although the performance of the humanized pathway remained largely invariant with chaperone engineering, native mannosyltransferase knockouts, and *GNT1* localization, we hypothesize that single perturbations are likely not sufficient for impacting pathway activity and that simultaneous engineering of multiple targets would be required. Chapter 6 demonstrated multiple process engineering strategies for improving glycosylation profile, including adjusting process parameters and media supplementation. We were able to increase G0 abundance on a subunit vaccine candidate with machine learning-guided supplementation, but beneficial effects were minimal for monoclonal antibodies.

With the development of CRISPR-Cas9 genome-wide library screens, powerful artificial intelligence tools, and more synthetic biology parts, we are able to rapidly generate new strains and testing conditions – such as combining targets identified in Chapter 5 or introducing additional parameters or media additives like in Chapter 6. The workflow in improving glycan profile (and other secretion-related phenotypes) is often bottlenecked at characterization. Improvements to the throughput of screening methods can greatly accelerate engineering cycle. Yeast surface display could be a viable strategy in improving glycosylation pathway engineering, since any genomic changes in this pathway would result in a global shift in glycan profile,

including that of surface anchor proteins. Identification of glycan-specific antibodies could benefit future studies using such platforms.

7.5. Conclusion

Protein quality is a complex problem, but it is an indispensable aspect of drug development to ensure safety and efficacy. Alternative hosts like *K. phaffii* are capable of secreting therapeutic proteins and can boast advantages including faster development, lower cost, greener process, and increased accessibility, but the product quality can be of concern compared to industry standard mammalian platforms. With the advancements presented in this thesis and the future directions suggested above, I hope to have conveyed that many quality issues observed in this host can be addressed through the engineering of molecular sequence, host biology, production process, or a combination of them. Nevertheless, there is a need to further understand the fundamental biology with data- and model-driven techniques, to develop better genetic tools to manipulate this biology, and to screen the resulting phenotypes with high-throughput method. The overarching multifaceted framework is applicable to other unconventional hosts as well. *K. phaffii* is not the magical cure-all for biologics production – no organism is, but through understanding and engineering this host and applying such insights to other hosts, we are right on track to providing high-quality pharmaceutical proteins to patients around the world.

8. REFERENCES

(1)   Walsh, G.; Walsh, E. Biopharmaceutical Benchmarks 2022. *Nat. Biotechnol.* **2022**, *40* (12), 1722.

(2)   Kesik-Brodacka, M. Progress in Biopharmaceutical Development. *Biotechnol. Appl. Biochem.* **2018**, *65* (3), 306–322.

(3)   Bulcha, J. T.; Wang, Y.; Ma, H.; Tai, P. W. L.; Gao, G. Viral Vector Platforms within the Gene Therapy Landscape. *Signal Transduct. Target. Ther. 2021 61* **2021**, *6* (1), 1–24.

(4)   Sterner, R. C.; Sterner, R. M. CAR-T Cell Therapy: Current Limitations and Potential Strategies. *Blood Cancer J. 2021 114* **2021**, *11* (4), 1–11.

(5)   Breedveld, F. C. Therapeutic Monoclonal Antibodies. *Lancet* **2000**, *355* (9205), 735–740.

(6)   Kim, J. Y.; Kim, Y. G.; Lee, G. M. CHO Cells in Biotechnology for Production of Recombinant Proteins: Current State and Further Potential. *Appl. Microbiol. Biotechnol.* **2012**, *93* (3), 917–930.

(7)   Liu, H. F.; Ma, J.; Winter, C.; Bayer, R. Recovery and Purification Process Development for Monoclonal Antibody Production. *MAbs* **2010**, *2* (5), 480–499.

(8)   Kelley, B.; Renshaw, T.; Kamarck, M.; Brian Kelley, C.; Biotechnology, V.; Francisco, S. Process and Operations Strategies to Enable Global Access to Antibody Therapies. *Biotechnol. Prog.* **2021**, *37* (3), e3139.

(9)   Lalor, F.; Fitzpatrick, J.; Sage, C.; Byrne, E. Sustainability in the Biopharmaceutical Industry: Seeking a Holistic Perspective. *Biotechnol. Adv.* **2019**, *37* (5), 698–707.

(10)   Brady, J. R.; Love, J. C. Alternative Hosts as the Missing Link for Equitable Therapeutic Protein Production. *Nat. Biotechnol. 2021 394* **2021**, *39* (4), 404–407.

(11)   Shukla, A. A.; Thömmes, J. Recent Advances in Large-Scale Production of Monoclonal Antibodies and Related Proteins. *Trends in Biotechnology*. Elsevier Current Trends May 2010, pp 253–261.

(12)   Wellcome Trust. *Expanding Access to Monoclonal Antibody-Based Products: A Global Call to Action*; 2020.

(13)   Bunnak, P.; Allmendinger, R.; Ramasamy, S. V.; Lettieri, P.; Titchener-Hooker, N. J. Life-Cycle and Cost of Goods Assessment of Fed-Batch and Perfusion-Based Manufacturing Processes for MAbs. *Biotechnol. Prog.* **2016**, *32* (5), 1324–1335.

(14)   Pietrzykowski, M.; Flanagan, W.; Pizzi, V.; Brown, A.; Sinclair, A.; Monge, M. An Environmental Life Cycle Assessment Comparison of Single-Use and Conventional Process Technology for the Production of Monoclonal Antibodies. *J. Clean. Prod.* **2013**, *41*, 150–162.

(15)   Kelley, B. *Developing Therapeutic Monoclonal Antibodies at Pandemic Pace*; 2020; Vol. 38, pp 540–545.

(16)   Demain, A. L.; Vaishnav, P. Production of Recombinant Proteins by Microbes and Higher Organisms. *Biotechnology Advances*. Elsevier May 1, 2009, pp 297–306.

(17)   Martínez, J. L.; Liu, L.; Petranovic, D.; Nielsen, J. Pharmaceutical Protein Production by Yeast: Towards Production of Human Blood Proteins by Microbial Fermentation. *Current Opinion in Biotechnology*. Elsevier Current Trends December 1, 2012, pp 965–971.

(18)   De Pourcq, K.; De Schutter, K.; Callewaert, N. Engineering of Glycosylation in Yeast and Other Fungi: Current State and Perspectives. *Appl. Microbiol. Biotechnol.* **2010**, *87* (5), 1617–1631.

(19)   Walsh, G. Biopharmaceutical Benchmarks 2018. *Nat. Biotechnol.* **2018**, *36* (12), 1136–

1145.

(20)   Valderrama-Rincon, J. D.; Fisher, A. C.; Merritt, J. H.; Fan, Y. Y.; Reading, C. A.; Chhiba, K.; Heiss, C.; Azadi, P.; Aebi, M.; DeLisa, M. P. An Engineered Eukaryotic Protein Glycosylation Pathway in Escherichia Coli. *Nat. Chem. Biol. 2012 85* **2012**, *8* (5), 434–436.

(21)   Mueller, P.; Gauttam, R.; Raab, N.; Handrick, R.; Wahl, C.; Leptihn, S.; Zorn, M.; Kussmaul, M.; Scheffold, M.; Eikmanns, B.; Elling, L.; Gaisser, S. High Level in Vivo Mucin-Type Glycosylation in Escherichia Coli. *Microb. Cell Fact.* **2018**, *17* (1), 1–15.

(22)   de Marco, A. Strategies for Successful Recombinant Expression of Disulfide Bond-Dependent Proteins in Escherichia Coli. *Microb. Cell Fact.* **2009**, *8* (1), 1–18.

(23)   Jiang, H.; Horwitz, A. A.; Wright, C.; Tai, A.; Znameroski, E. A.; Tsegaye, Y.; Warbington, H.; Bower, B. S.; Alves, C.; Co, C.; Jonnalagadda, K.; Platt, D.; Walter, J. M.; Natarajan, V.; Ubersax, J. A.; Cherry, J. R.; Love, J. C. Challenging the Workhorse: Comparative Analysis of Eukaryotic Micro-organisms for Expressing Monoclonal Antibodies. *Biotechnol. Bioeng.* **2019**, *116* (6), 1449–1462.

(24)   Coleman, E. M. Establishment of a Novel Pichia Pastoris Host Production Platform, Massachusetts Institute of Technology, 2020.

(25)   Hejnaes, K. R.; Ransohoff, T. C. Chemistry, Manufacture and Control. *Biopharm. Process. Dev. Des. Implement. Manuf. Process.* **2018**, 1105–1136.

(26)   Çelik, E.; Çalik, P. Production of Recombinant Proteins by Yeast Cells. *Biotechnol. Adv.* **2012**, *30* (5), 1108–1118.

(27)   Love, K. R.; Dalvie, N. C.; Love, J. C. The Yeast Stands Alone: The Future of Protein Biologic Production. *Current Opinion in Biotechnology*. Elsevier Ltd October 1, 2018, pp 50–58.

(28)   Shekhar, C. Pichia Power: India's Biotech Industry Puts Unconventional Yeast to Work. *Chem. Biol.* **2008**, *15* (3), 201–202.

(29)   Weiss, M.; Steiner, D. F.; Philipson, L. H. Insulin Biosynthesis, Secretion, Structure, and Structure-Activity Relationships. *Endotext* **2015**.

(30)   Trastuzumab - PubChem https://pubchem.ncbi.nlm.nih.gov/compound/Trastuzumab (accessed Jun 1, 2024).

(31)   Goetze, A. M.; Schenauer, M. R.; Flynn, G. C. Assessing Monoclonal Antibody Product Quality Attribute Criticality through Clinical Studies. *MAbs* **2010**, *2* (5), 500–507.

(32)   Eon-Duval, A.; Broly, H.; Gleixner, R. Quality Attributes of Recombinant Therapeutic Proteins: An Assessment of Impact on Safety and Efficacy as Part of a Quality by Design Development Approach. *Biotechnol. Prog.* **2012**, *28* (3), 608–622.

(33)   Tebbey, P. W.; Varga, A.; Naill, M.; Clewell, J.; Venema, J. Consistency of Quality Attributes for the Glycosylated Monoclonal Antibody Humira® (Adalimumab). *MAbs* **2015**, *7* (5), 805–811.

(34)   Li, H.; Sethuraman, N.; Stadheim, T. A.; Zha, D.; Prinz, B.; Ballew, N.; Bobrowicz, P.; Choi, B. K.; Cook, W. J.; Cukan, M.; Houston-Cummings, N. R.; Davidson, R.; Gong, B.; Hamilton, S. R.; Hoopes, J. P.; Jiang, Y.; Kim, N.; Mansfield, R.; Nett, J. H.; Rios, S.; Strawbridge, R.; Wildt, S.; Gerngross, T. U. Optimization of Humanized IgGs in Glycoengineered Pichia Pastoris. *Nat. Biotechnol.* **2006**, *24* (2), 210–215.

(35)   Sinha, J.; Plantz, B. A.; Inan, M.; Meagher, M. M. Causes of Proteolytic Degradation of Secreted Recombinant Proteins Produced in Methylotrophic Yeast Pichia Pastoris: Case Study with Recombinant Ovine Interferon-τ. *Biotechnol. Bioeng.* **2005**, *89* (1), 102–112.

(36)   Kerry-Williams, S. M.; Gilbert, S. C.; Evans, L. R.; Ballance, D. J. Disruption of the Saccharomyces Cerevisiae YAP3 Gene Reduces the Proteolytic Degradation of Secreted Recombinant Human Albumin. **1998**, *14*, 161–169.

(37)   Gil, D. F.; García-Fernández, R.; Alonso-del-Rivero, M.; Lamazares, E.; Pérez, M.; Varas, L.; Díaz, J.; Chávez, M. A.; González-González, Y.; Mansur, M. Recombinant Expression of ShPI-1A, a Non-Specific BPTI-Kunitz-Type Inhibitor, and Its Protection Effect on Proteolytic Degradation of Recombinant Human Miniproinsulin Expressed in Pichia Pastoris. *FEMS Yeast Res.* **2011**, *11* (7), 575–586.

(38)   Kohno, K. Stress-Sensing Mechanisms in the Unfolded Protein Response: Similarities and Differences between Yeast and Mammals. *J. Biochem.* **2010**, *147* (1), 27–33.

(39)   Labrou, N. E. Protein Purifi Cation: An Overview. *Methods Mol. Biol.* **2014**, *1129*, 3–10.

(40)   Laukens, B.; De Visscher, C.; Callewaert, N. Engineering Yeast for Producing Human Glycoproteins: Where Are We Now? *Future Microbiol.* **2015**, *10* (1), 21–34.

(41)   Aebi, M. N-Linked Protein Glycosylation in the ER. *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier November 1, 2013, pp 2430–2437.

(42)   Tanner, W.; Lehle, L. Protein Glycosylation in Yeast. *BBA - Reviews on Biomembranes*. Elsevier April 27, 1987, pp 81–99.

(43)   Janik, A.; Juchimiuk, M.; Kruszewska, J.; Orowski, J.; Pasikowska, M.; Palamarczyk, G. Impact of Yeast Glycosylation Pathway on Cell Integrity and Morphology. In *Glycosylation*; InTech, 2012.

(44)   Sha, S.; Agarabi, C.; Brorson, K.; Lee, D. Y.; Yoon, S. N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends in Biotechnology*. Elsevier Ltd October 1, 2016, pp 835–846.

(45)   CAMPION, B.; LEGER, D.; WIERUSZESKI, J.-M.; MONTREUIL, J.; SPIK, G. Presence of Fucosylated Triantennary, Tetraantennary and Pentaantennary Glycans in Transferrin Synthesized by the Human Hepatocarcinoma Cell Line Hep G2. *Eur. J. Biochem.* **1989**, *184* (2), 405–413.

(46)   Takahashi, M.; Kuroki, Y.; Ohtsubo, K.; Taniguchi, N. Core Fucose and Bisecting GlcNAc, the Direct Modifiers of the N-Glycan Core: Their Functions and Target Proteins. *Carbohydrate Research*. Elsevier August 17, 2009, pp 1387–1390.

(47)   McNulty, M. J.; Berliner, A. J.; Negulescu, P. G.; McKee, L.; Hart, O.; Yates, K.; Arkin, A. P.; Nandi, S.; McDonald, K. A. Evaluating the Cost of Pharmaceutical Purification for a Long-Duration Space Exploration Medical Foundry. *Front. Microbiol.* **2021**, *12*, 700863.

(48)   Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nat. 1990 3476294* **1990**, *347* (6294), 631–639.

(49)   Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. *Methods Mol. Biol.* **2008**, *452*, 63–87.

(50)   AlQuraishi, M. Machine Learning in Protein Structure Prediction. *Curr. Opin. Chem. Biol.* **2021**, *65*, 1–8.

(51)   Dalvie, N. C.; Brady, J. R.; Crowell, L. E.; Tracey, M. K.; Biedermann, A. M.; Kaur, K.; Hickey, J. M.; Kristensen, D. L.; Bonnyman, A. D.; Rodriguez-Aponte, S. A.; Whittaker, C. A.; Bok, M.; Vega, C.; Mukhopadhyay, T. K.; Joshi, S. B.; Volkin, D. B.; Parreño, V.; Love, K. R.; Love, J. C. Molecular Engineering Improves Antigen Quality and Enables Integrated Manufacturing of a Trivalent Subunit Vaccine Candidate for Rotavirus. *Microb. Cell Fact.* **2021**, *20* (1), 1–14.

(52)   Dalvie, N. C.; Rodriguez-Aponte, S. A.; Hartwell, B. L.; Tostanoski, L. H.; Biedermann,

A. M.; Crowell, L. E.; Kaur, K.; Kumru, O. S.; Carter, L.; Yu, J.; Chang, A.; McMahan, K.; Courant, T.; Lebas, C.; Lemnios, A. A.; Rodrigues, K. A.; Silva, M.; Johnston, R. S.; Naranjo, C. A.; Tracey, M. K.; Brady, J. R.; Whittaker, C. A.; Yun, D.; Brunette, N.; Wang, J. Y.; Walkey, C.; Fiala, B.; Kar, S.; Porto, M.; Lok, M.; Andersen, H.; Lewis, M. G.; Love, K. R.; Camp, D. L.; Silverman, J. M.; Kleanthous, H.; Joshi, S. B.; Volkin, D. B.; Dubois, P. M.; Collin, N.; King, N. P.; Barouch, D. H.; Irvine, D. J.; Love, J. C. Engineered SARS-CoV-2 Receptor Binding Domain Improves Manufacturability in Yeast and Immunogenicity in Mice. **2021**, *118* (38), e2106845118.

(53)   McCamish, M.; Woollett, G. The State of the Art in the Development of Biosimilars. *Clin. Pharmacol. Ther.* **2012**, *91* (3), 405–417.

(54)   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.; Gerngross, T. U. Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins. *Science (80-.).* **2006**.

(55)   Sola, R. J.; Griebenow, K. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Journal of Pharmaceutical Sciences*. John Wiley and Sons Inc. April 1, 2009, pp 1223–1245.

(56)   Meyer, A. J.; Segall-Shapiro, T. H.; Glassey, E.; Zhang, J.; Voigt, C. A. Escherichia Coli "Marionette" Strains with 12 Highly Optimized Small-Molecule Sensors. *Nat. Chem. Biol.* **2019**, *15* (2), 196–204.

(57)   Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol. 2016 143* **2016**, *14* (3), 135–149.

(58)   Gao, J.; Jiang, L.; Lian, J. Development of Synthetic Biology Tools to Engineer Pichia Pastoris as a Chassis for the Production of Natural Products. *Synth. Syst. Biotechnol.* **2021**, *6* (2), 110–119.

(59)   Choi, B.-K.; Bobrowicz, P.; Davidson, R. C.; Hamilton, S. R.; Kung, D. H.; Li, H.; Miele, R. G.; Nett, J. H.; Wildt, S.; Gerngross, T. U. Use of Combinatorial Genetic Libraries to Humanize N-Linked Glycosylation in the Yeast Pichia Pastoris. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (9), 5022–5027.

(60)   de Groot, N. S.; Ventura, S. Effect of Temperature on Protein Quality in Bacterial Inclusion Bodies. *FEBS Lett.* **2006**, *580* (27), 6471–6476.

(61)   Kaufmann, H.; Mazur, X.; Fussenegger, M.; Bailey, J. E. Influence of Low Temperature on Productivity, Proteome and Protein Phosphorylation of CHO Cells. *Biotechnol Bioeng* **1999**, *63*, 573–582.

(62)   Görgens, J. F.; Van Zyl, W. H.; Knoetze, J. H.; Hahn-Hägerdal, B. Amino Acid Supplementation Improves Heterologous Protein Production by Saccharomyces Cerevisiae in Defined Medium. *Appl. Microbiol. Biotechnol.* **2005**, *67* (5), 684–691.

(63)   Ehret, J.; Zimmermann, M.; Eichhorn, T.; Zimmer, A. Impact of Cell Culture Media Additives on IgG Glycosylation Produced in Chinese Hamster Ovary Cells. *Biotechnol. Bioeng.* **2019**, *116* (4), 816–830.

(64)   Du, J.; Yuan, Y.; Si, T.; Lian, J.; Zhao, H. Customized Optimization of Metabolic Pathways by Combinatorial Transcriptional Engineering. *Nucleic Acids Res.* **2012**.

(65)   Hashizume, T.; Ozawa, Y.; Ying, B. W. Employing Active Learning in the Optimization of Culture Medium for Mammalian Cells. *npj Syst. Biol. Appl. 2023 91* **2023**, *9* (1), 1–10.

(66)   Bader, J.; Narayanan, H.; Arosio, P.; Leroux, J. C. Improving Extracellular Vesicles

Production through a Bayesian Optimization-Based Experimental Design. *Eur. J. Pharm. Biopharm.* **2023**, *182*, 103–114.

(67) Ecker, D. M.; Jones, S. D.; Levine, H. L. The Therapeutic Monoclonal Antibody Market. *MAbs* **2015**, *7* (1), 9–14.

(68) Makowski, E. K.; Kinnunen, P. C.; Huang, J.; Wu, L.; Smith, M. D.; Wang, T.; Desai, A. A.; Streu, C. N.; Zhang, Y.; Zupancic, J. M.; Schardt, J. S.; Linderman, J. J.; Tessier, P. M. Co-Optimization of Therapeutic Antibody Affinity and Specificity Using Machine Learning Models That Generalize to Novel Mutational Space. *Nat. Commun. 2022 131* **2022**, *13* (1), 1–14.

(69) Williams, K. L.; Guerrero, S.; Flores-Garcia, Y.; Kim, D.; Williamson, K. S.; Siska, C.; Smidt, P.; Jepson, S. Z.; Li, K.; Dennison, S. M.; Mathis-Torres, S.; Chen, X.; Wille-Reece, U.; MacGill, R. S.; Walker, M.; Jongert, E.; King, C. R.; Ockenhouse, C.; Glanville, J.; Moon, J. E.; Regules, J. A.; Tan, Y. C.; Cavet, G.; Lippow, S. M.; Robinson, W. H.; Dutta, S.; Tomaras, G. D.; Zavala, F.; Ketchem, R. R.; Emerling, D. E. A Candidate Antibody Drug for Prevention of Malaria. *Nat. Med. 2024 301* **2024**, *30* (1), 117–129.

(70) Coffman, J.; Brower, M.; Connell-Crowley, L.; Deldari, S.; Farid, S. S.; Horowski, B.; Patil, U.; Pollard, D.; Qadan, M.; Rose, S.; Schaefer, E.; Shultz, J. A Common Framework for Integrated and Continuous Biomanufacturing. *Biotechnol. Bioeng.* **2021**, *118* (4), 1735–1749.

(71) Kelley, B.; Kiss, R.; Laird, M. A Different Perspective: How Much Innovation Is Really Needed for Monoclonal Antibody Production Using Mammalian Cell Technology? *Adv. Biochem. Eng. Biotechnol.* **2018**, *165*, 443–462.

(72) Mahal, H.; Branton, H.; Farid, S. S. End-to-End Continuous Bioprocessing: Impact on Facility Design, Cost of Goods, and Cost of Development for Monoclonal Antibodies. *Biotechnol. Bioeng.* **2021**, *118* (9), 3468–3485.

(73) Pollock, J.; Coffman, J.; Ho, S. V.; Farid, S. S. Integrated Continuous Bioprocessing: Economic, Operational, and Environmental Feasibility for Clinical and Commercial Antibody Manufacture. *Biotechnol. Prog.* **2017**, *33* (4), 854–866.

(74) Crowell, L. E.; Lu, A. E.; Love, K. R.; Stockdale, A.; Timmick, S. M.; Wu, D.; Wang, Y. A.; Doherty, W.; Bonnyman, A.; Vecchiarello, N.; Goodwine, C.; Bradbury, L.; Brady, J. R.; Clark, J. J.; Colant, N. A.; Cvetkovic, A.; Dalvie, N. C.; Liu, D.; Liu, Y.; Mascarenhas, C. A.; Matthews, C. B.; Mozdzierz, N. J.; Shah, K. A.; Wu, S. L.; Hancock, W. S.; Braatz, R. D.; Cramer, S. M.; Love, J. C. On-Demand Manufacturing of Clinical-Quality Biopharmaceuticals. *Nature Biotechnology*. Nature Publishing Group November 2018, p 988.

(75) Crowell, L. E.; Rodriguez, S. A.; Love, K. R.; Cramer, S. M.; Love, J. C. Rapid Optimization of Processes for the Integrated Purification of Biopharmaceuticals. *Biotechnol. Bioeng.* **2021**, *118* (9), 3435–3446.

(76) Jung, S. T.; Kang, T. H.; Kelton, W.; Georgiou, G. Bypassing Glycosylation: Engineering Aglycosylated Full-Length IgG Antibodies for Human Therapy. *Curr. Opin. Biotechnol.* **2011**, *22* (6), 858–867.

(77) Ju, M. S.; Jung, S. T. Aglycosylated Full-Length IgG Antibodies: Steps toward next-Generation Immunotherapeutics. *Curr. Opin. Biotechnol.* **2014**, *30*, 128–139.

(78) Dhillon, S. Eptinezumab: First Approval. *Drugs* **2020**, *80* (7), 733–739.

(79) Mimura, Y.; Katoh, T.; Saldova, R.; O'Flaherty, R.; Izumi, T.; Mimura-Kimura, Y.; Utsunomiya, T.; Mizukami, Y.; Yamamoto, K.; Matsumoto, T.; Rudd, P. M. Glycosylation

Engineering of Therapeutic IgG Antibodies: Challenges for the Safety, Functionality and Efficacy. *Protein Cell 2017 91* **2017**, *9* (1), 47–62.

(80)   Rodriguez-Aponte, S. A.; Dalvie, N. C.; Wong, T. Y.; Johnston, R. S.; Naranjo, C. A.; Bajoria, S.; Kumru, O. S.; Kaur, K.; Russ, B. P.; Lee, K. S.; Cyphert, H. A.; Barbier, M.; Rao, H. D.; Rajurkar, M. P.; Lothe, R. R.; Shaligram, U. S.; Batwal, S.; Chandrasekaran, R.; Nagar, G.; Kleanthous, H.; Biswas, S.; Bevere, J. R.; Joshi, S. B.; Volkin, D. B.; Damron, F. H.; Love, J. C. Molecular Engineering of a Cryptic Epitope in Spike RBD Improves Manufacturability and Neutralizing Breadth against SARS-CoV-2 Variants. *Vaccine* **2023**, *41* (5), 1108.

(81)   Liu, L.; Stadheim, A.; Hamuro, L.; Pittman, T.; Wang, W.; Zha, D.; Hochman, J.; Prueksaritanont, T. Pharmacokinetics of IgG1 Monoclonal Antibodies Produced in Humanized Pichia Pastoris with Specific Glycoforms: A Comparative Study with CHO Produced Materials. *Biologicals* **2011**, *39* (4), 205–210.

(82)   Zhang, N.; Liu, L.; Dan Dumitru, C.; Cummings, N. R. H.; Cukan, M.; Jiang, Y.; Li, Y.; Li, F.; Mitchell, T.; Mallem, M. R.; Ou, Y.; Patel, R. N.; Vo, K.; Wang, H.; Burnina, I.; Choi, B. K.; Huber, H.; Stadheim, T. A.; Zha, D. Glycoengineered Pichia Produced Anti-HER2 Is Comparable to Trastuzumab in Preclinical Study. *MAbs* **2011**, *3* (3), 289–298.

(83)   Love, K. R.; Shah, K. A.; Whittaker, C. A.; Wu, J.; Bartlett, M. C.; Ma, D.; Leeson, R. L.; Priest, M.; Borowsky, J.; Young, S. K.; Love, J. C. Comparative Genomics and Transcriptomics of Pichia Pastoris. *BMC Genomics* **2016**.

(84)   Barrero, J. J.; Casler, J. C.; Valero, F.; Ferrer, P.; Glick, B. S. An Improved Secretion Signal Enhances the Secretion of Model Proteins from Pichia Pastoris. *Microb. Cell Fact.* **2018**, *17* (1), 1–13.

(85)   Lin-Cereghino, G. P.; Stark, C. M.; Kim, D.; Chang, J. W. J.; Shaheen, N.; Poerwanto, H.; Agari, K.; Moua, P.; Low, L. K.; Tran, N.; Huang, A. D.; Nattestad, M.; Oshiro, K. T.; Chang, J. W. J.; Chavan, A.; Tsai, J. W.; Lin-Cereghino, J. The Effect of α-Mating Factor Secretion Signal Mutations on Recombinant Protein Expression in Pichia Pastoris. *Gene* **2013**, *519* (2), 311–317.

(86)   Kozlov, D. G.; Yagudin, T. A. Antibody Fragments May Be Incorrectly Processed in the Yeast Pichia Pastoris. *Biotechnol. Lett.* **2008**, *30* (9), 1661–1663.

(87)   Ghosalkar, A.; Sahai, V.; Srivastava, A. Secretory Expression of Interferon-Alpha 2b in Recombinant Pichia Pastoris Using Three Different Secretion Signals. *Protein Expr. Purif.* **2008**, *60* (2), 103–109.

(88)   Wang, X.; Zhu, M.; Zhang, A.; Yang, F.; Chen, P. Synthesis and Secretory Expression of Hybrid Antimicrobial Peptide CecA-Mag and Its Mutants in Pichia Pastoris. *Exp. Biol. Med.* **2012**, *237* (3), 312–317.

(89)   Neiers, F.; Belloir, C.; Poirier, N.; Naumer, C.; Krohn, M.; Briand, L. Comparison of Different Signal Peptides for the Efficient Secretion of the Sweet-Tasting Plant Protein Brazzein in Pichia Pastoris. *Life 2021, Vol. 11, Page 46* **2021**, *11* (1), 46.

(90)   Gramer, M. J. Product Quality Considerations for Mammalian Cell Culture Process Development and Manufacturing. *Adv. Biochem. Eng. Biotechnol.* **2014**, *139*, 123–166.

(91)   Beck, A.; Liu, H. Macro- and Micro-Heterogeneity of Natural and Recombinant IgG Antibodies. *Antibodies 2019, Vol. 8, Page 18* **2019**, *8* (1), 18.

(92)   Walsh, G.; Jefferis, R. Post-Translational Modifications in the Context of Therapeutic Proteins. *Nat. Biotechnol. 2006 2410* **2006**, *24* (10), 1241–1252.

(93)   Martinez, T.; Pace, D.; Brady, L.; Gerhart, M.; Balland, A. Characterization of a Novel

Modification on IgG2 Light Chain: Evidence for the Presence of O-Linked Mannosylation. *J. Chromatogr. A* **2007**, *1156* (1–2), 183–187.

(94) Hegde, R. S.; Bernstein, H. D. The Surprising Complexity of Signal Sequences. *Trends Biochem. Sci.* **2006**, *31* (10), 563–571.

(95) Aw, R.; McKay, P. F.; Shattock, R. J.; Polizzi, K. M. A Systematic Analysis of the Expression of the Anti-HIV VRC01 Antibody in Pichia Pastoris through Signal Peptide Optimization. *Protein Expr. Purif.* **2018**, *149*, 43–50.

(96) Yang, J.; Lu, Z.; Chen, J.; Chu, P.; Cheng, Q.; Liu, J.; Ming, F.; Huang, C.; Xiao, A.; Cai, H.; Zhang, L. Effect of Cooperation of Chaperones and Gene Dosage on the Expression of Porcine PGLYRP-1 in Pichia Pastoris. *Appl. Microbiol. Biotechnol.* **2016**, *100* (12), 5453–5465.

(97) Delic, M.; Göngrich, R.; Mattanovich, D.; Gasser, B. Engineering of Protein Folding and Secretion—Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins. *Antioxid. Redox Signal.* **2014**, *21* (3), 414–437.

(98) Haryadi, R.; Ho, S.; Kok, Y. J.; Pu, H. X.; Zheng, L.; Pereira, N. A.; Li, B.; Bi, X.; Goh, L. T.; Yang, Y.; Song, Z. Optimization of Heavy Chain and Light Chain Signal Peptides for High Level Expression of Therapeutic Antibodies in CHO Cells. *PLoS One* **2015**, *10* (2), e0116878.

(99) Pettit, D. K.; Rogers, R. S.; Arthur, K.; Brodsky, Y.; Clark, R. H.; Crowell, C.; Ennis, J.; Gillespie, A.; Gillespie, R.; Livingston, B.; Nalbandian, E.; Pace, D.; Smidt, P.; Pauly, M.; Timmons, K.; Trentalange, M.; Whaley, K. J.; Zeitlin, L.; Thomas, J. N. CHO Cell Production and Sequence Improvement in the 13C6FR1 Anti-Ebola Antibody. *MAbs* **2016**, *8* (2), 347–357.

(100) Dalvie, N. C.; Naranjo, C. A.; Rodriguez-Aponte, S. A.; Johnston, R. S.; Christopher Love, J. Steric Accessibility of the N-Terminus Improves the Titer and Quality of Recombinant Proteins Secreted from Komagataella Phaffii. *Microb. Cell Fact.* **2022**, *21* (1), 1–11.

(101) Delic, M.; Valli, M.; Graf, A. B.; Pfeffer, M.; Mattanovich, D.; Gasser, B. The Secretory Pathway: Exploring Yeast Diversity. **2013**.

(102) Ratih, R.; Asmari, M.; Abdel-Megied, A. M.; Elbarbry, F.; El Deeb, S. Biosimilars: Review of Regulatory, Manufacturing, Analytical Aspects and Beyond. *Microchem. J.* **2021**, *165*, 106143.

(103) Brorson, K.; Jia, A. Y. Therapeutic Monoclonal Antibodies and Consistent Ends: Terminal Heterogeneity, Detection, and Impact on Quality. *Curr. Opin. Biotechnol.* **2014**, *30*, 140–146.

(104) Alan Lazar, G.; Llp, W. Immunoglobulin Variants Outside the Fc Region. **2005**, *353* (60).

(105) Xiao, Q.; Zhang, F.; Nacev, B. A.; Liu, J. O.; Pei, D. Protein N-Terminal Processing: Substrate Specificity of Escherichia Coli and Human Methionine Aminopeptidases. *Biochemistry* **2010**, *49* (26), 5588.

(106) Dick, L. W.; Qiu, D.; Mahon, D.; Adamo, M.; Cheng, K. C. C-Terminal Lysine Variants in Fully Human Monoclonal Antibodies: Investigation of Test Methods and Possible Causes. *Biotechnol. Bioeng.* **2008**, *100* (6), 1132–1143.

(107) Bernstein, J. A.; Qazi, M. Ecallantide: Its Pharmacology, Pharmacokinetics, Clinical Efficacy and Tolerability. *Expert Rev. Clin. Immunol.* **2010**, *6* (1), 29–39.

(108) Cukan, M. C.; Hopkins, D.; Burnina, I.; Button, M.; Giaccone, E.; Houston-Cummings, N. R.; Jiang, Y.; Li, F.; Mallem, M.; Mitchell, T.; Moore, R.; Nylen, A.; Prinz, B.; Rios, S.;

Sharkey, N.; Zha, D.; Hamilton, S.; Li, H.; Stadheim, T. A. Binding of DC-SIGN to Glycoproteins Expressed in Glycoengineered Pichia Pastoris. *J. Immunol. Methods* **2012**, *386* (1–2), 34–42.

(109) Hamilton, S. R.; Cook, W. J.; Gomathinayagam, S.; Burnina, I.; Bukowski, J.; Hopkins, D.; Schwartz, S.; Du, M.; Sharkey, N. J.; Bobrowicz, P.; Wildt, S.; Li, H.; Stadheim, T. A.; Nett, J. H. Production of Sialylated O-Linked Glycans in Pichia Pastoris. *Glycobiology* **2013**, *23* (10), 1192–1203.

(110) Torres-Obreque, K. M.; Meneguetti, G. P.; Muso-Cachumba, J. J.; Feitosa, V. A.; Santos, J. H. P. M.; Ventura, S. P. M.; Rangel-Yagui, C. O. Building Better Biobetters: From Fundamentals to Industrial Application. *Drug Discov. Today* **2022**, *27* (1), 65–81.

(111) Brady, J. R.; Whittaker, C. A.; Tan, M. C.; Kristensen, D. L.; Ma, D.; Dalvie, N. C.; Love, K. R.; Love, J. C. Comparative Genome-Scale Analysis of Pichia Pastoris Variants Informs Selection of an Optimal Base Strain. *Biotechnol. Bioeng.* **2020**.

(112) Dalvie, N. C.; Leal, J.; Whittaker, C. A.; Yang, Y.; Brady, J. R.; Love, K. R.; Love, J. C.; Christopher Love, J. Host-Informed Expression of CRISPR Guide RNA for Genomic Engineering in Komagataella Phaffii. *ACS Synth. Biol.* **2020**, *9* (1), 26–35.

(113) Raab, D.; Graf, M.; Notka, F.; Schödl, T.; Wagner, R. The GeneOptimizer Algorithm: Using a Sliding Window Approach to Cope with the Vast Sequence Space in Multiparameter DNA Sequence Optimization. *Syst. Synth. Biol.* **2010**, *4* (3), 215.

(114) Wright, C.; Alves, C.; Kshirsagar, R.; Pieracci, J.; Estes, S. Leveraging a CHO Cell Line Toolkit to Accelerate Biotherapeutics into the Clinic. *Biotechnol. Prog.* **2017**, *33* (6), 1468–1475.

(115) Huang, Y. M.; Hu, W. W.; Rustandi, E.; Chang, K.; Yusuf-Makagiansar, H.; Ryll, T. Maximizing Productivity of CHO Cell-Based Fed-Batch Culture Using Chemically Defined Media Conditions and Typical Manufacturing Equipment. *Biotechnol. Prog.* **2010**, *26* (5), 1400–1410.

(116) Matthews, C. B.; Wright, C.; Kuo, A.; Colant, N.; Westoby, M.; Love, J. C. Reexamining Opportunities for Therapeutic Protein Production in Eukaryotic Microorganisms. *Biotechnol. Bioeng.* **2017**, *114* (11), 2432–2444.

(117) Shapiro, R. S.; Chavez, A.; Collins, J. J. CRISPR-Based Genomic Tools for the Manipulation of Genetically Intractable Microorganisms. *Nat. Rev. Microbiol.* **2018**, *16* (6), 333–339.

(118) Nett, J. H.; Hodel, N.; Rausch, S.; Wildt, S. Cloning and Disruption of ThePichia Pastoris ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6 Genes and Their Use as Auxotrophic Markers. *Yeast* **2005**, *22* (4), 295–304.

(119) Hamilton, S. R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R. C.; Li, H.; Mitchell, T.; Nett, J. H.; Rausch, S.; Stadheim, T. A.; Wischnewski, H.; Wildt, S.; Gerngross, T. U. Production of Complex Human Glycoproteins in Yeast. *Science (80-.).* **2003**, *301* (5637), 1244–1246.

(120) Ahmad, M.; Winkler, C. M.; Kolmbauer, M.; Pichler, H.; Schwab, H.; Emmerstorfer-Augustin, A. Pichia Pastoris Protease-Deficient and Auxotrophic Strains Generated by a Novel, User-Friendly Vector Toolbox for Gene Deletion. *Yeast* **2019**, *36* (9), 557–570.

(121) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157* (6), 1262–1278.

(122) Raschmanová, H.; Weninger, A.; Glieder, A.; Kovar, K.; Vogl, T. Implementing CRISPR-Cas Technologies in Conventional and Non-Conventional Yeasts: Current State and Future

Prospects. *Biotechnol. Adv.* **2018**.

(123) Tihanyi, B.; Nyitray, L. Recent Advances in CHO Cell Line Development for Recombinant Protein Production. *Drug Discov. Today Technol.* **2020**, *38*, 25–34.

(124) Weninger, A.; Hatzl, A. M.; Schmid, C.; Vogl, T.; Glieder, A. Combinatorial Optimization of CRISPR/Cas9 Expression Enables Precision Genome Engineering in the Methylotrophic Yeast Pichia Pastoris. *J. Biotechnol.* **2016**, *235*, 139–149.

(125) Gao, J.; Xu, J.; Zuo, Y.; Ye, C.; Jiang, L.; Feng, L.; Huang, L.; Xu, Z.; Lian, J. Synthetic Biology Toolkit for Marker-Less Integration of Multigene Pathways into Pichia Pastoris via CRISPR/Cas9. *ACS Synth. Biol.* **2022**, *11* (2), 623–633.

(126) Liu, Q.; Shi, X.; Song, L.; Liu, H.; Zhou, X.; Wang, Q.; Zhang, Y.; Cai, M. CRISPR-Cas9-Mediated Genomic Multiloci Integration in Pichia Pastoris. *Microb. Cell Fact.* **2019**, *18* (1), 1–11.

(127) Jacobs, P. P.; Geysens, S.; Vervecken, W.; Contreras, R.; Callewaert, N. Engineering Complex-Type N-Glycosylation in Pichia Pastoris Using GlycoSwitch Technology. *Nat. Protoc.* **2009**, *4* (1), 58–70.

(128) Nett, J. H.; Terrance, A. S.; Li, H.; Bobrowicz, P.; Hamilton, S. R.; Davidson, R. C.; Choi, B.-K.; Mitchell, T.; Bobrowicz, B.; Rittenhour, A.; Wildt, S.; Gerngross, T. U. A Combinatorial Genetic Library Approach to Target Heterologous Glycosylation Enzymes to the Endoplasmic Reticulum or the Golgi Apparatus of Pichia Pastoris. *Yeast* **2010**, *27* (2), 67–76.

(129) Zhang, Y.; Wang, J.; Wang, Z.; Zhang, Y.; Shi, S.; Nielsen, J.; Liu, Z. A GRNA-TRNA Array for CRISPR-Cas9 Based Rapid Multiplexed Genome Editing in Saccharomyces Cerevisiae. *Nat. Commun. 2019 101* **2019**, *10* (1), 1–10.

(130) Näätsaari, L.; Mistlberger, B.; Ruth, C.; Hajek, T.; Hartner, F. S.; Glieder, A. Deletion of the Pichia Pastoris KU70 Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology. *PLoS One* **2012**, *7* (6), e39720.

(131) Kunert, R.; Reinhart, D. Advances in Recombinant Antibody Manufacturing. *Applied Microbiology and Biotechnology*. Springer Verlag April 1, 2016, pp 3451–3461.

(132) Jeschek, M.; Gerngross, D.; Panke, S. Combinatorial Pathway Optimization for Streamlined Metabolic Engineering. *Current Opinion in Biotechnology*. Elsevier Ltd October 1, 2017, pp 142–151.

(133) Jensen, M. K.; Keasling, J. D. Recent Applications of Synthetic Biology Tools for Yeast Metabolic Engineering. *FEMS Yeast Res.* **2015**, *15*, 1–10.

(134) Khalil, A. S.; Collins, J. J. Synthetic Biology: Applications Come of Age. *Nat. Publ. Gr.* **2010**, *11*, 367.

(135) Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E. Expression-Level Optimization of a Multi-Enzyme Pathway in the Absence of a High-Throughput Assay. *Nucleic Acids Res.* **2013**.

(136) Blazeck, J.; Garg, R.; Reed, B.; Alper, H. S. Controlling Promoter Strength and Regulation in Saccharomyces Cerevisiae Using Synthetic Hybrid Promoters. *Biotechnol. Bioeng.* **2012**, *109* (11), 2884–2895.

(137) Hubmann, G.; Thevelein, J. M.; Nevoigt, E. Natural and Modifi Ed Promoters for Tailored Metabolic Engineering of the Yeast Saccharomyces Cerevisiae. *Methods Mol. Biol.* **2014**, *1152*, 17–42.

(138) Curran, K. A.; Karim, A. S.; Gupta, A.; Alper, H. S. Use of Expression-Enhancing Terminators in Saccharomyces Cerevisiae to Increase MRNA Half-Life and Improve Gene

Expression Control for Metabolic Engineering Applications. *Metab. Eng.* **2013**, *19*, 88–97.

(139) Babiskin, A. H.; Smolke, C. D. A Synthetic Library of RNA Control Modules for Predictable Tuning of Gene Expression in Yeast. *Mol. Syst. Biol.* **2011**, *7*, 471.

(140) Peña, D. A.; Gasser, B.; Zanghellini, J.; Steiger, M. G.; Mattanovich, D. Metabolic Engineering of Pichia Pastoris. *Metabolic Engineering*. Academic Press Inc. November 1, 2018, pp 2–15.

(141) Obst, U.; Lu, T. K.; Sieber, V. A Modular Toolkit for Generating Pichia Pastoris Secretion Libraries. *ACS Synth. Biol.* **2017**, *6* (6), 1016–1025.

(142) Inan, M.; Meagher, M. M. Non-Repressing Carbon Sources for Alcohol Oxidase (AOX1) Promoter of Pichia Pastoris. *J. Biosci. Bioeng.* **2001**, *92* (6), 585–589.

(143) Hartner, F. S.; Ruth, C.; Langenegger, D.; Johnson, S. N.; Hyka, P.; Lin-Cereghino, G. P.; Lin-Cereghino, J.; Kovar, K.; Cregg, J. M.; Glieder, A. Promoter Library Designed for Fine-Tuned Gene Expression in Pichia Pastoris. *Nucleic Acids Res.* **2008**, *36* (12), e76–e76.

(144) Cos, O.; Ramón, R.; Montesinos, J. L.; Valero, F. Operational Strategies, Monitoring and Control of Heterologous Protein Production in the Methylotrophic Yeast Pichia Pastoris under Different Promoters: A Review. *Microbial Cell Factories*. 2006.

(145) Menendez, J.; Valdes, I.; Cabrera, N. The ICLI Gene of Pichia Pastoris, Transcriptional Regulation and Use of Its Promoter. *Yeast* **2003**.

(146) Shen, W.; Xue, Y.; Liu, Y.; Kong, C.; Wang, X.; Huang, M.; Cai, M.; Zhou, X.; Zhang, Y.; Zhou, M. A Novel Methanol-Free Pichia Pastoris System for Recombinant Protein Expression. *Microb. Cell Fact.* **2016**.

(147) Wang, J.; Wang, X.; Shi, L.; Qi, F.; Zhang, P.; Zhang, Y.; Zhou, X.; Song, Z.; Cai, M. Methanol-Independent Protein Expression by AOX1 Promoter with Trans-Acting Elements Engineering and Glucose-Glycerol-Shift Induction in Pichia Pastoris. *Sci. Rep.* **2017**, *7*.

(148) Perez-Pinera, P.; Han, N.; Cleto, S.; Cao, J.; Purcell, O.; Shah, K. A.; Lee, K.; Ram, R.; Lu, T. K. Synthetic Biology and Microbioreactor Platforms for Programmable Production of Biologics at the Point-of-Care. *Nat. Commun.* **2016**, *7* (1), 12211.

(149) Weinhandl, K.; Winkler, M.; Glieder, A.; Camattari, A. Carbon Source Dependent Promoters in Yeasts. *Microbial Cell Factories*. 2014.

(150) Delic, M.; Mattanovich, D.; Gasser, B. Repressible Promoters – A Novel Tool to Generate Conditional Mutants in Pichia Pastoris. *Microb. Cell Fact.* **2013**, *12* (1), 6.

(151) Liu, X. Bin; Liu, M.; Tao, X. Y.; Zhang, Z. X.; Wang, F. Q.; Wei, D. Z. Metabolic Engineering of Pichia Pastoris for the Production of Dammarenediol-II. *J. Biotechnol.* **2015**.

(152) Stadlmayr, G.; Mecklenbräuker, A.; Rothmüller, M.; Maurer, M.; Sauer, M.; Mattanovich, D.; Gasser, B. Identification and Characterisation of Novel Pichia Pastoris Promoters for Heterologous Protein Production. *J. Biotechnol.* **2010**.

(153) Prielhofer, R.; Maurer, M.; Klein, J.; Wenger, J.; Kiziak, C.; Gasser, B.; Mattanovich, D. Induction without Methanol: Novel Regulated Promoters Enable High-Level Expression in Pichia Pastoris. *Microb. Cell Fact.* **2013**, *12*, 1.

(154) Waern, K.; Nagalakshmi, U.; Snyder, M. RNA Sequencing. *Methods Mol. Biol.* **2011**.

(155) Liang, S.; Wang, B.; Pan, L.; Ye, Y.; He, M.; Han, S.; Zheng, S.; Wang, X.; Lin, Y. Comprehensive Structural Annotation of Pichia Pastoris Transcriptome and the Response to Various Carbon Sources Using Deep Paired-End RNA Sequencing. *BMC Genomics*

2012.

(156) Klein, M.; Swinnen, S.; Thevelein, J. M.; Nevoigt, E. Glycerol Metabolism and Transport in Yeast and Fungi: Established Knowledge and Ambiguities. *Environmental Microbiology*. 2017.

(157) Li, J.; Liang, Q.; Song, W.; Marchisio, M. A. Nucleotides Upstream of the Kozak Sequence Strongly Influence Gene Expression in the Yeast S. Cerevisiae. *J. Biol. Eng.* **2017**.

(158) Brady, J. R. A Multi-Omics Approach to Improving Productivity of Therapeutic Proteins in Pichia Pastoris (Komagataella Phaffii), Massachusetts Institute of Technology, 2019.

(159) Reddington, S. C.; Howarth, M. Secrets of a Covalent Interaction for Biomaterials and Biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* **2015**, *29*, 94–99.

(160) Marsalek, L.; Puxbaum, V.; Buchetics, M.; Mattanovich, D.; Gasser, B. Disruption of Vacuolar Protein Sorting Components of the HOPS Complex Leads to Enhanced Secretion of Recombinant Proteins in Pichia Pastoris. *Microb. Cell Fact.* **2019**, *18* (1).

(161) Huang, M.; Wang, G.; Qin, J.; Petranovic, D.; Nielsen, J. Engineering the Protein Secretory Pathway of Saccharomyces Cerevisiae Enables Improved Protein Production. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (47), E11025–E11032.

(162) de Ruijter, J. C.; Koskela, E. V.; Frey, A. D. Enhancing Antibody Folding and Secretion by Tailoring the Saccharomyces Cerevisiae Endoplasmic Reticulum. *Microb. Cell Fact.* **2016**, *15* (1), 87.

(163) Li, P.; Sun, H.; Chen, Z.; Li, Y.; Zhu, T. Construction of Efficient Xylose Utilizing Pichia Pastoris for Industrial Enzyme Production. *Microb. Cell Fact.* **2015**, *14* (1), 22.

(164) Yoshikawa, K.; Tanaka, T.; Ida, Y.; Furusawa, C.; Hirasawa, T.; Shimizu, H. Comprehensive Phenotypic Analysis of Single-Gene Deletion and Overexpression Strains of Saccharomyces Cerevisiae. *Yeast* **2011**, *28* (5), 349–361.

(165) Lin-Cereghino, J.; Wong, W. W.; Xiong, S.; Giang, W.; Luong, L. T.; Vu, J.; Johnson, S. D.; Lin-Cereghino, G. P. Condensed Protocol for Competent Cell Preparation and Transformation of the Methylotrophic Yeast Pichia Pastoris. *Biotechniques* **2005**, *38* (1), 44–48.

(166) Lõoke, M.; Kristjuhan, K.; Kristjuhan, A. Extraction of Genomic DNA from Yeasts for PCR-Based Applications. *Biotechniques* **2011**, *50* (5), 325–328.

(167) Matthews, C. B.; Kuo, A.; Love, K. R.; Love, J. C. Development of a General Defined Medium for Pichia Pastoris. *Biotechnol. Bioeng.* **2018**, *115* (1), 103–113.

(168) Patro, R.; Duggal, G.; Love, M. I.; Irizarry, R. A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods 2017 144* **2017**, *14* (4), 417–419.

(169) Soneson, C.; Love, M. I.; Robinson, M. D. Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences. *F1000Research* **2016**, *4*.

(170) Li, H.; d'Anjou, M. Pharmacological Significance of Glycosylation in Therapeutic Proteins. *Curr. Opin. Biotechnol.* **2009**, *20* (6), 678–684.

(171) Gmeiner, C.; Spadiut, O. Protein Production with a Pichia Pastoris OCH1 Knockout Strain in Fed-Batch Mode. In *Glyco-Engineering: Methods and Protocols*; Springer New York, 2015; pp 91–101.

(172) Hamilton, S. R.; Gerngross, T. U. Glycosylation Engineering in Yeast: The Advent of Fully Humanized Yeast. *Curr. Opin. Biotechnol.* **2007**, *18* (5), 387–392.

(173) Pekarsky, A.; Veiter, L.; Rajamanickam, V.; Herwig, C.; Grünwald-Gruber, C.; Altmann,

F.; Spadiut, O. Production of a Recombinant Peroxidase in Different Glyco-Engineered Pichia Pastoris Strains: A Morphological and Physiological Comparison. *Microb. Cell Fact.* **2018**, *17* (1), 1–15.

(174) Jiang, B.; Argyros, R.; Bukowski, J.; Nelson, S.; Sharkey, N.; Kim, S.; Copeland, V.; Davidson, R. C.; Chen, R.; Zhuang, J.; Sethuraman, N.; Stadheim, T. A. Inactivation of a GAL4-like Transcription Factor Improves Cell Fitness and Product Yield in Glycoengineered Pichia Pastoris Strains. *Appl. Environ. Microbiol.* **2015**, *81* (1), 260–271.

(175) Elena, C.; Ravasi, P.; Castelli, M. E.; Peirú, S.; Menzella, H. G. Expression of Codon Optimized Genes in Microbial Systems: Current Industrial Applications and Perspectives. *Front. Microbiol.* **2014**, *5* (FEB), 78262.

(176) Miura, M.; Hirose, M.; Miwa, T.; Kuwae, S.; Ohi, H. Cloning and Characterization in Pichia Pastoris of PNO1 Gene Required for Phosphomannosylation of N-Linked Oligosaccharides. *Gene* **2004**, *324* (1–2), 129–137.

(177) Ram, A. F. J.; Klis, F. M. Identification of Fungal Cell Wall Mutants Using Susceptibility Assays Based on Calcofluor White and Congo Red. *Nat. Protoc. 2006 15* **2006**, *1* (5), 2253–2256.

(178) Gustin, M. C.; Albertyn, J.; Alexander, M.; Davenport, K. MAP Kinase Pathways in the Yeast Saccharomyces Cerevisiae. *Microbiol. Mol. Biol. Rev.* **1998**, *62* (4), 1264.

(179) Zou, H.; Hastie, T.; Tibshirani, R. Sparse Principal Component Analysis. *J. Comput. Graph. Stat.* **2006**, *15* (2), 265–286.

(180) Kolberg, L.; Raudvere, U.; Kuzmin, I.; Adler, P.; Vilo, J.; Peterson, H. G:Profiler—Interoperable Web Service for Functional Enrichment Analysis and Gene Identifier Mapping (2023 Update). *Nucleic Acids Res.* **2023**, *51* (W1), W207–W212.

(181) Naville, M.; Ghuillot-Gaudeffroy, A.; Marchais, A.; Gautheret, D. ARNold: A Web Tool for the Prediction of Rho-Independent Transcription Terminators. *RNA Biol.* **2011**, *8* (1), 11–13.

(182) Curran, K. A.; Morse, N. J.; Markham, K. A.; Wagman, A. M.; Gupta, A.; Alper, H. S. Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol.* **2015**, *4* (7), 824–832.

(183) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27* (8), 753–759.

(184) Brady, J. R.; Tan, M. C.; Whittaker, C. A.; Colant, N. A.; Dalvie, N. C.; Love, K. R.; Love, J. C. Identifying Improved Sites for Heterologous Gene Integration Using ATAC-Seq. *ACS Synth. Biol.* **2020**, *9* (9), 2515–2524.

(185) Reimand, J.; Isserlin, R.; Voisin, V.; Kucera, M.; Tannus-Lopes, C.; Rostamianfar, A.; Wadi, L.; Meyer, M.; Wong, J.; Xu, C.; Merico, D.; Bader, G. D. Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc. 2019 142* **2019**, *14* (2), 482–517.

(186) Witten, D. M.; Tibshirani, R.; Hastie, T. A Penalized Matrix Decomposition, with Applications to Sparse Principal Components and Canonical Correlation Analysis. *Biostatistics* **2009**, *10* (3), 515–534.

(187) Claes, K.; Van Herpe, D.; Vanluchene, R.; Roels, C.; Van Moer, B.; Wyseure, E.; Vandewalle, K.; Eeckhaut, H.; Yilmaz, S.; Vanmarcke, S.; Çıtak, E.; Fijalkowska, D.; Grootaert, H.; Lonigro, C.; Meuris, L.; Michielsen, G.; Naessens, J.; van Schie, L.; De Rycke, R.; De Bruyne, M.; Borghgraef, P.; Callewaert, N. OPENPichia: Licence-Free

Komagataella Phaffii Chassis Strains and Toolkit for Protein Expression. *Nat. Microbiol. 2024 93* **2024**, *9* (3), 864–876.

(188) Dalvie, N. C. Product and Host Engineering for Low-Cost Manufacturing of Therapeutic Proteins in the Yeast Komagataella Phaffii, Massachusetts Institute of Technology, 2022.

(189) Dalvie, N. C.; Lorgeree, T.; Biedermann, A. M.; Love, K. R.; Love, J. C. Simplified Gene Knockout by CRISPR-Cas9-Induced Homologous Recombination. *ACS Synth. Biol.* **2022**, *11* (1), 497–501.

(190) Lingg, N.; Zhang, P.; Song, Z.; Bardor, M. The Sweet Tooth of Biopharmaceuticals: Importance of Recombinant Protein Glycosylation Analysis. *Biotechnol. J.* **2012**, *7* (12), 1462–1472.

(191) Spadiut, O.; Capone, S.; Krainer, F.; Glieder, A.; Herwig, C. Microbials for the Production of Monoclonal Antibodies and Antibody Fragments. *Trends Biotechnol.* **2014**, *32* (1), 54–60.

(192) Dinnis, D. M.; James, D. C. Engineering Mammalian Cell Factories for Improved Recombinant Monoclonal Antibody Production: Lessons from Nature? *Biotechnol. Bioeng.* **2005**, *91* (2), 180–189.

(193) Mathias, S.; Wippermann, A.; Raab, N.; Zeh, N.; Handrick, R.; Gorr, I.; Schulz, P.; Fischer, S.; Gamer, M.; Otte, K. Unraveling What Makes a Monoclonal Antibody Difficult-to-Express: From Intracellular Accumulation to Incomplete Folding and Degradation via ERAD. *Biotechnol. Bioeng.* **2020**, *117* (1), 5–16.

(194) Xu, C.; Ng, D. T. W. Glycosylation-Directed Quality Control of Protein Folding. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group December 1, 2015, pp 742–752.

(195) Dean, N. Asparagine-Linked Glycosylation in the Yeast Golgi. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1426* (2), 309–322.

(196) Quan, E. M.; Kamiya, Y.; Kamiya, D.; Denic, V.; Weibezahn, J.; Kato, K.; Weissman, J. S. Defining the Glycan Destruction Signal for Endoplasmic Reticulum-Associated Degradation. *Mol. Cell* **2008**, *32* (6), 870–877.

(197) Adams, B. M.; Canniff, N. P.; Guay, K. P.; Hebert, D. N. The Role of Endoplasmic Reticulum Chaperones in Protein Folding and Quality Control. *Prog. Mol. Subcell. Biol.* **2021**, *59*, 27–50.

(198) Pelham, H. R.; Hardwick, K. G.; Lewis, M. J. Sorting of Soluble ER Proteins in Yeast. *EMBO J.* **1988**, *7* (6), 1757–1762.

(199) Stolz, J.; Munro, S. The Components of the Saccharomyces Cerevisiae Mannosyltransferase Complex M-Pol I Have Distinct Functions in Mannan Synthesis. *J. Biol. Chem.* **2002**, *277* (47), 44801–44808.

(200) Lueking, A.; Holz, C.; Gotthold, C.; Lehrach, H.; Cahill, D. A System for Dual Protein Expression in Pichia Pastoris and Escherichia Coli. *Protein Expr. Purif.* **2000**, *20* (3), 372–378.

(201) Nakamura, Y.; Nishi, T.; Noguchi, R.; Ito, Y.; Watanabe, T.; Nishiyama, T.; Aikawa, S.; Hasunuma, T.; Ishii, J.; Okubo, Y.; Kondo, A. A Stable, Autonomously Replicating Plasmid Vector Containing Pichia Pastoris Centromeric DNA. *Appl. Environ. Microbiol.* **2018**, *84* (15), 2882–2899.

(202) Rodriguez-Aponte, S. A.; Naranjo, C. A.; Johnston, R. S.; Dalvie, N. C.; Crowell, L. E.; Bajoria, S.; Kumru, O. S.; Joshi, S. B.; Volkin, D. B.; Love, J. C. Minimal Purification Method Enables Developability Assessment of Recombinant Proteins. *Biotechnol. Bioeng.*

2023.

(203) Kim, H.; Yoo, S. J.; Kang, H. A. Yeast Synthetic Biology for the Production of Recombinant Therapeutic Proteins. *FEMS Yeast Research*. Narnia August 1, 2015, p n/a-n/a.

(204) Brühlmann, D.; Jordan, M.; Hemberger, J.; Sauer, M.; Stettler, M.; Broly, H. Tailoring Recombinant Protein Quality by Rational Media Design. *Biotechnol. Prog.* **2015**, *31* (3), 615–629.

(205) Brunner, M.; Fricke, J.; Kroll, P.; Herwig, C. Investigation of the Interactions of Critical Scale-up Parameters (PH, PO2 and PCO2) on CHO Batch Performance and Critical Quality Attributes. *Bioprocess Biosyst. Eng.* **2017**, *40* (2), 251–263.

(206) Blondeel, E. J. M.; Braasch, K.; McGill, T.; Chang, D.; Engel, C.; Spearman, M.; Butler, M.; Aucoin, M. G. Tuning a MAb Glycan Profile in Cell Culture: Supplementing N-Acetylglucosamine to Favour G0 Glycans without Compromising Productivity and Cell Growth. *J. Biotechnol.* **2015**, *214*, 105–112.

(207) Gramer, M. J.; Eckblad, J. J.; Donahue, R.; Brown, J.; Shultz, C.; Vickerman, K.; Priem, P.; van den Bremer, E. T. J.; Gerritsen, J.; van Berkel, P. H. C. Modulation of Antibody Galactosylation through Feeding of Uridine, Manganese Chloride, and Galactose. *Biotechnol. Bioeng.* **2011**, *108* (7), 1591–1602.

(208) Potvin, G.; Ahmad, A.; Zhang, Z. Bioprocess Engineering Aspects of Heterologous Protein Production in Pichia Pastoris: A Review. *Biochem. Eng. J.* **2012**, *64*, 91–105.

(209) Dalvie, N. C.; Biedermann, A. M.; Rodriguez-Aponte, S. A.; Naranjo, C. A.; Rao, H. D.; Rajurkar, M. P.; Lothe, R. R.; Shaligram, U. S.; Johnston, R. S.; Crowell, L. E.; Castelino, S.; Tracey, M. K.; Whittaker, C. A.; Love, J. C. Scalable, Methanol-free Manufacturing of the SARS-CoV-2 Receptor-binding Domain in Engineered Komagataella Phaffii. *Biotechnol. Bioeng.* **2022**, *119* (2), 657.

(210) Ongley, S. E.; Bian, X.; Neilan, B. A.; Müller, R. Recent Advances in the Heterologous Expression of Microbial Natural Product Biosynthetic Pathways. *Nat. Prod. Rep.* **2013**, *30* (8), 1121–1138.

(211) Çelik, E.; Çalik, P.; Oliver, S. G. Fed-Batch Methanol Feeding Strategy for Recombinant Protein Production by Pichia Pastoris in the Presence of Co-Substrate Sorbitol. *Yeast* **2009**, *26* (9), 473–484.

(212) Gawlitzek, M.; Estacio, M.; Fürch, T.; Kiss, R. Identification of Cell Culture Conditions to Control N-Glycosylation Site-Occupancy of Recombinant Glycoproteins Expressed in CHO Cells. *Biotechnol. Bioeng.* **2009**, *103* (6), 1164–1175.

(213) St. Amand, M. M.; Radhakrishnan, D.; Robinson, A. S.; Ogunnaike, B. A. Identification of Manipulated Variables for a Glycosylation Control Strategy. *Biotechnol. Bioeng.* **2014**, *111* (10), 1957–1970.

(214) Narayanan, H.; Dingfelder, F.; Condado Morales, I.; Patel, B.; Heding, K. E.; Bjelke, J. R.; Egebjerg, T.; Butté, A.; Sokolov, M.; Lorenzen, N.; Arosio, P. Design of Biopharmaceutical Formulations Accelerated by Machine Learning. *Mol. Pharm.* **2021**, *18* (10), 3843–3853.

(215) OKOROKOV, L. A.; LICHKO, L. P.; KADOMTSEVA, V. M.; KHOLODENKO, V. P.; TITOVSKY, V. T.; KULAEV, I. S. Energy-Dependent Transport of Manganese into Yeast Cells and Distribution of Accumulated Ions. *Eur. J. Biochem.* **1977**, *75* (2), 373–377.

(216) Ye, J.; Ly, J.; Watts, K.; Hsu, A.; Walker, A.; Mclaughlin, K.; Berdichevsky, M.; Prinz, B.; Sean Kersey, D.; d'Anjou, M.; Pollard, D.; Potgieter, T. Optimization of a

Glycoengineered Pichia Pastoris Cultivation Process for Commercial Antibody Production. *Biotechnol. Prog.* **2011**, *27* (6), 1744–1750.

(217) Prabhu, A.; Gadgil, M. Trace Metals in Cellular Metabolism and Their Impact on Recombinant Protein Production. *Process Biochem.* **2021**, *110*, 251–262.

(218) Grainger, R. K.; James, D. C. CHO Cell Line Specific Prediction and Control of Recombinant Monoclonal Antibody N-Glycosylation. *Biotechnol. Bioeng.* **2013**, *110* (11), 2970–2983.

(219) Royle, L.; Campbell, M. P.; Radcliffe, C. M.; White, D. M.; Harvey, D. J.; Abrahams, J. L.; Kim, Y. G.; Henry, G. W.; Shadick, N. A.; Weinblatt, M. E.; Lee, D. M.; Rudd, P. M.; Dwek, R. A. HPLC-Based Analysis of Serum N-Glycans on a 96-Well Plate Platform with Dedicated Database Software. *Anal. Biochem.* **2008**, *376* (1), 1–12.

(220) Martin, V. J. J.; Piteral, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. Engineering a Mevalonate Pathway in Escherichia Coli for Production of Terpenoids. *Nat. Biotechnol. 2003 217* **2003**, *21* (7), 796–802.

(221) Liu, Y.; Zhao, X.; Gan, F.; Chen, X.; Deng, K.; Crowe, S. A.; Hudson, G. A.; Belcher, M. S.; Schmidt, M.; Astolfi, M. C. T.; Kosina, S. M.; Pang, B.; Shao, M.; Yin, J.; Sirirungruang, S.; Iavarone, A. T.; Reed, J.; Martin, L. B. B.; El-Demerdash, A.; Kikuchi, S.; Misra, R. C.; Liang, X.; Cronce, M. J.; Chen, X.; Zhan, C.; Kakumanu, R.; Baidoo, E. E. K.; Chen, Y.; Petzold, C. J.; Northen, T. R.; Osbourn, A.; Scheller, H.; Keasling, J. D. Complete Biosynthesis of QS-21 in Engineered Yeast. *Nat. 2024 6298013* **2024**, *629* (8013), 937–944.

(222) Park, J. H.; Bassalo, M. C.; Lin, G. M.; Chen, Y.; Doosthosseini, H.; Schmitz, J.; Roubos, J. A.; Voigt, C. A. Design of Four Small-Molecule-Inducible Systems in the Yeast Chromosome, Applied to Optimize Terpene Biosynthesis. *ACS Synth. Biol.* **2023**, *12* (4), 1119–1132.

(223) Wu, X.; Cai, P.; Yao, L.; Zhou, Y. J. Genetic Tools for Metabolic Engineering of Pichia Pastoris. *Eng. Microbiol.* **2023**, *3* (4), 100094.

(224) Bezjak, L.; Erklavec Zajec, V.; Baebler, Š.; Stare, T.; Gruden, K.; Pohar, A.; Novak, U.; Likozar, B. Incorporating RNA-Seq Transcriptomics into Glycosylation-Integrating Metabolic Network Modelling Kinetics: Multiomic Chinese Hamster Ovary (CHO) Cell Bioreactors. *Biotechnol. Bioeng.* **2021**, *118* (4), 1476–1490.

(225) Presnell, K. V.; Alper, H. S. Systems Metabolic Engineering Meets Machine Learning: A New Era for Data-Driven Metabolic Engineering. *Biotechnol. J.* **2019**, *14* (9), 1800416.

Appendix A
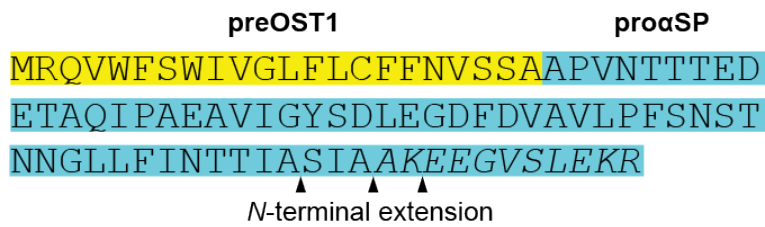
Chapter 2 supplemental figure(s)

preOST1 **proαSP**

MRQVWFSWIVGLFLCFFNVSSA APVNTTTED
ETAQIPAEAVIGYSDLEGDFDVAVLPFSNST
NNGLLFINTTIASIA*AKEEGVSLEKR*

N-terminal extension

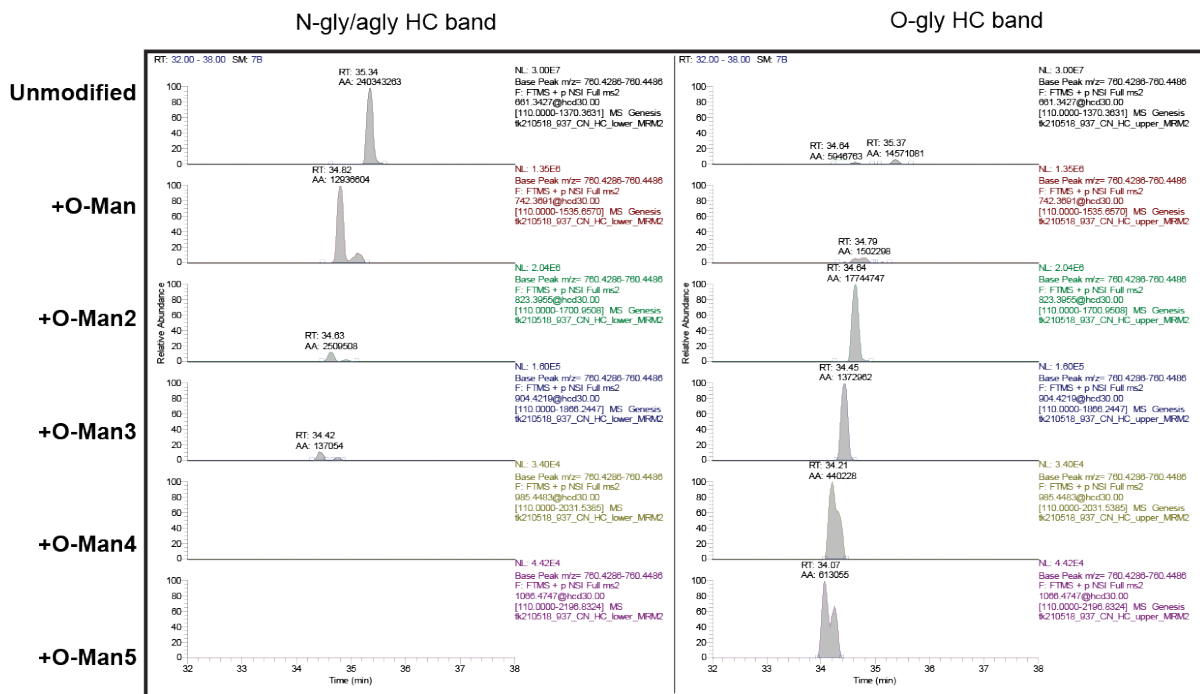**Figure A2.1. N-terminal extension of trastuzumab light chain.**

Figure A2.2. Tandem MS of a differentially mannosylated tryptic peptide in HC.
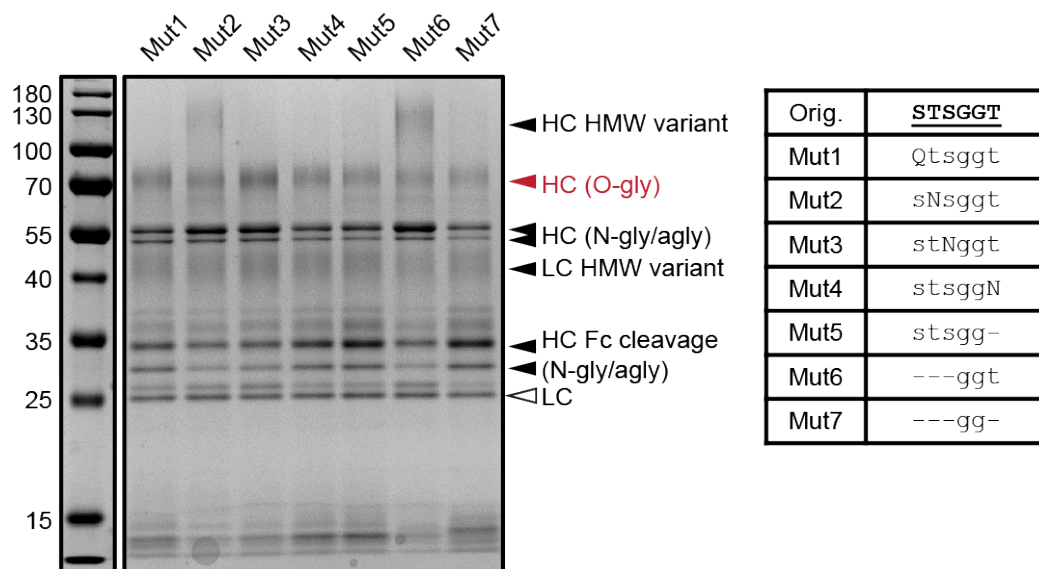
155

Figure A2.3. Mutations in the identified tryptic peptide to reduce O-mannosylation.
No significant reductions in O-mannosylation were observed with different mutations.

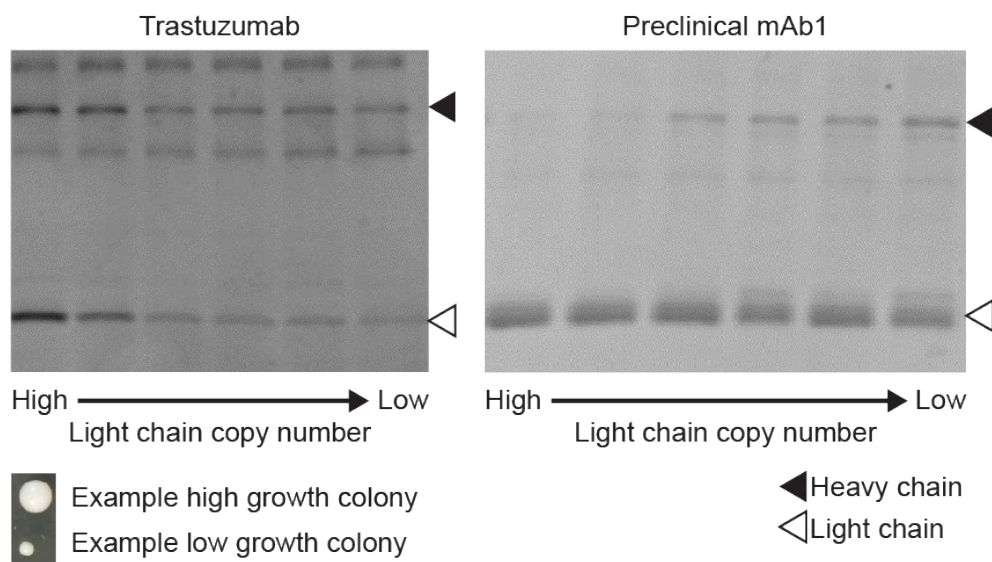| Orig.	**STSGGT**
Mut1	Qtsggt
Mut2	sNsggt
Mut3	stNggt
Mut4	stsggN
Mut5	stsgg–
Mut6	---ggt
Mut7	---gg–

Figure A2.4. Secretion of trastuzumab and a preclinical mAb as a function of gene copy number.
Colony growth on selection plate was treated as a proxy for gene copy number.
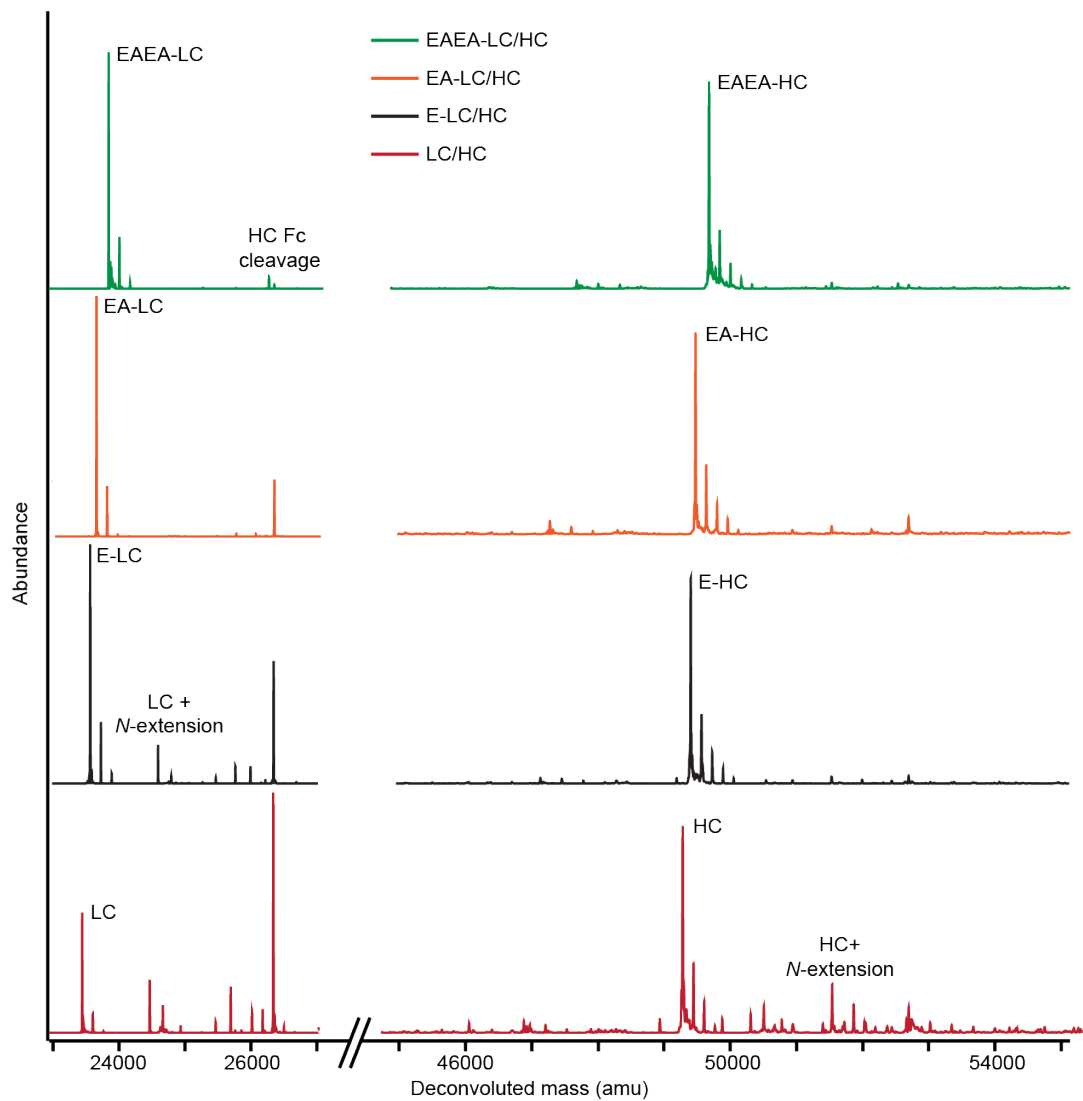
Figure A2.5. Product-related variants of engineered trastuzumab, LC-MS analysis.
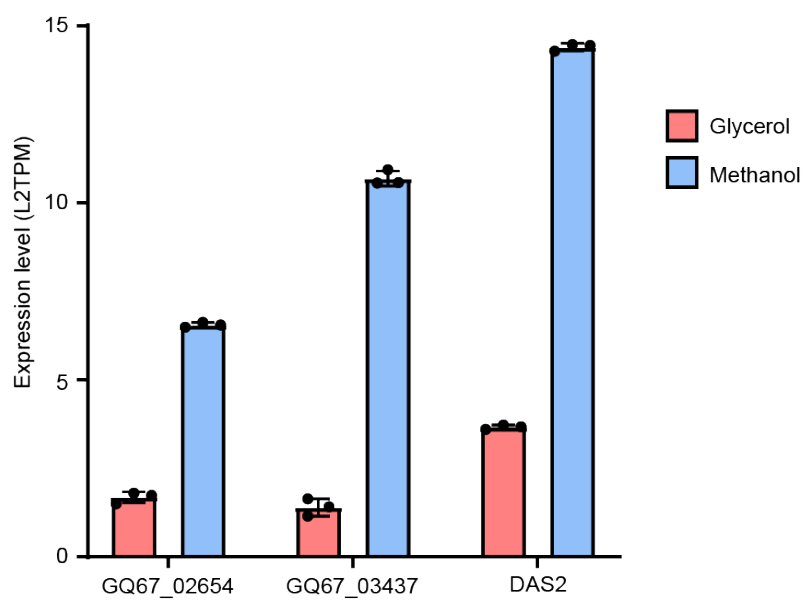
Chapter 3 supplemental figure(s)

Figure A3.1. Expression of methanol-inducible genes of interest.
The native promoters of these genes can be reappropriated for methanol-inducible heterologous pathways.
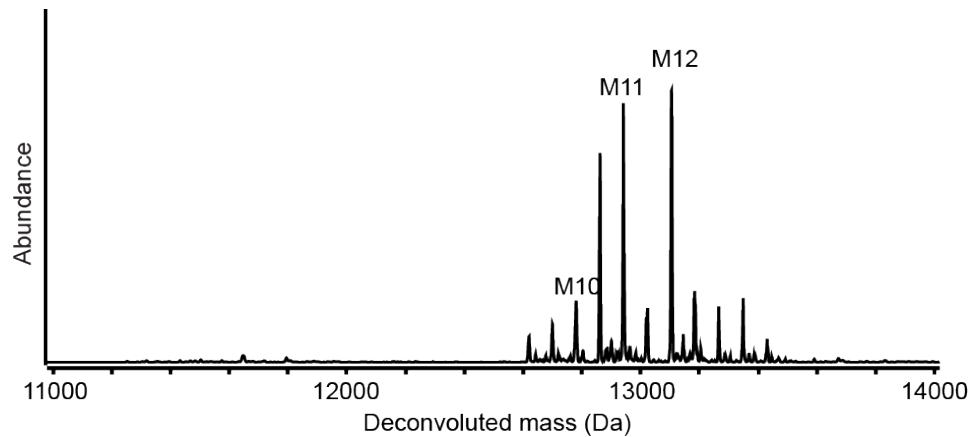
Chapter 4 supplemental figure(s)

Figure A4.1. Example K3 peptide mass spectrum in a G0 strain with mutated, dysfunctional *MNS1*.
Note the absence of any glycan structures with lower numbers of mannose moieties. Mutated *MNS1* is likely incapable of cleaving off mannose(s) from hypermannosylated structures.
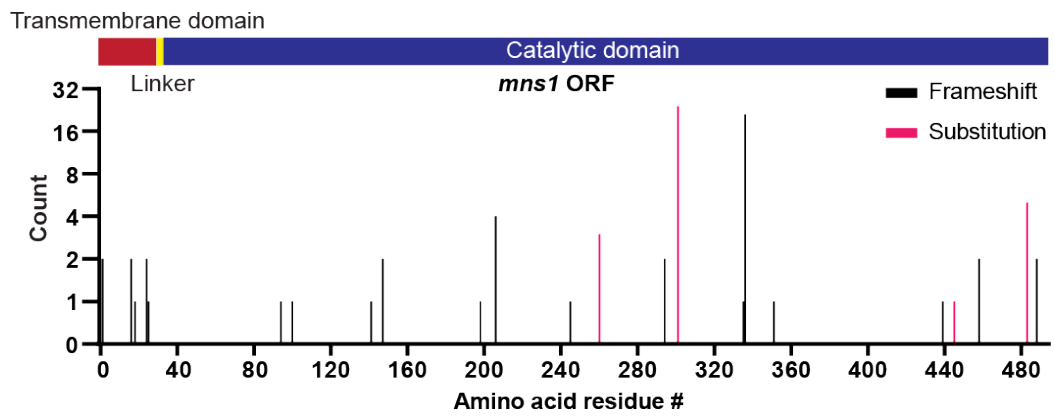
Figure A4.2. Spontaneous mutations in *MNS1* open reading frame in GlcNAcMan5-glycoengineered strains.

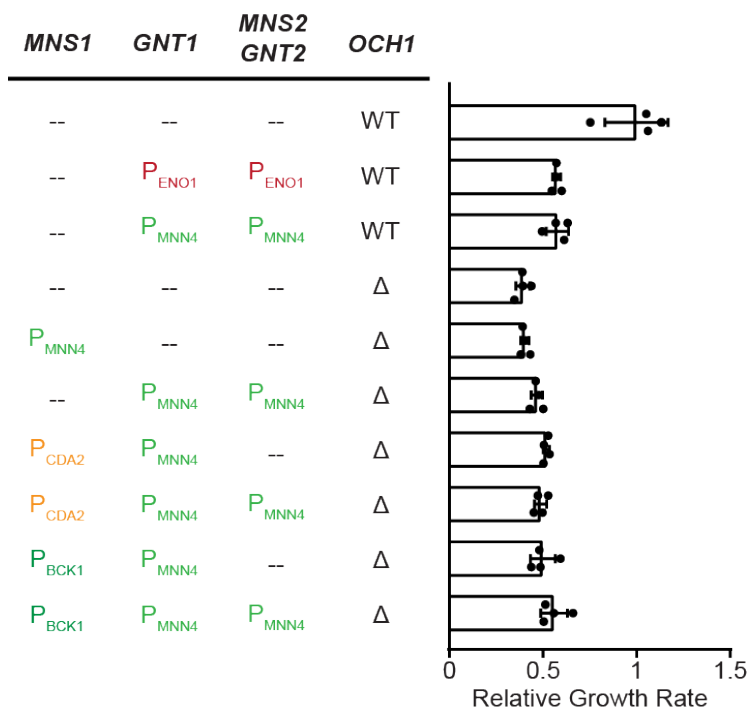| MNS1	GNT1	MNS2 GNT2	OCH1
--	--	--	WT
--	P_{ENO1}	P_{ENO1}	WT
--	P_{MNN4}	P_{MNN4}	WT
--	--	--	Δ
P_{MNN4}	--	--	Δ
--	P_{MNN4}	P_{MNN4}	Δ
P_{CDA2}	P_{MNN4}	--	Δ
P_{CDA2}	P_{MNN4}	P_{MNN4}	Δ
P_{BCK1}	P_{MNN4}	--	Δ
P_{BCK1}	P_{MNN4}	P_{MNN4}	Δ

Relative Growth Rate

Figure A4.3. Growth rate comparison of glycoengineered strains.
Note the significant decrease in OCH1-intact strains after the integration of heterologous glycosylation pathway genes. Additional engineering (integration of *MNS2* and *GNT2*) does not appear to significantly affect growth rate in Δ*och1* strains.
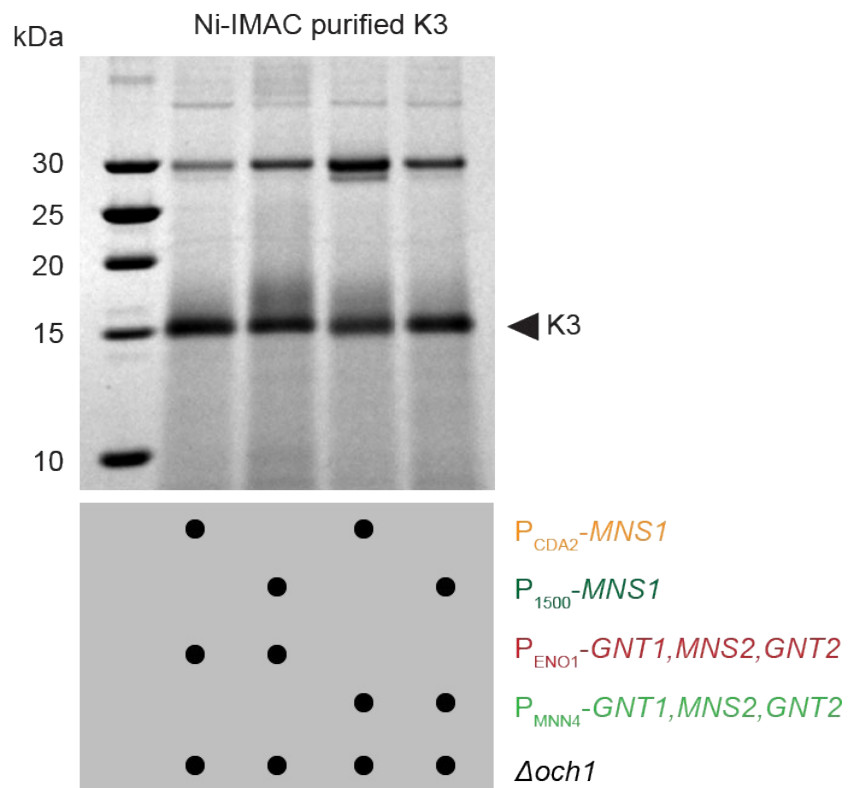
162

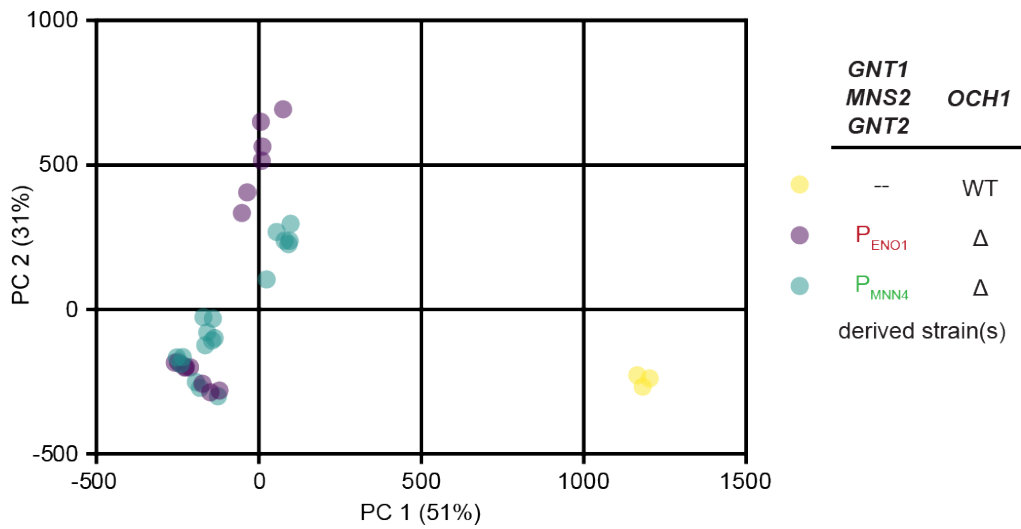**Figure A4.4. Secretion of K3 peptide in G0 glycoengineered strains with WT *MNS1*.**

Figure A4.5. Principal component analysis of transcriptomic dataset.
Note the separation by the first PC by wildtype vs. Δ*och1*/glycoengineered strains.
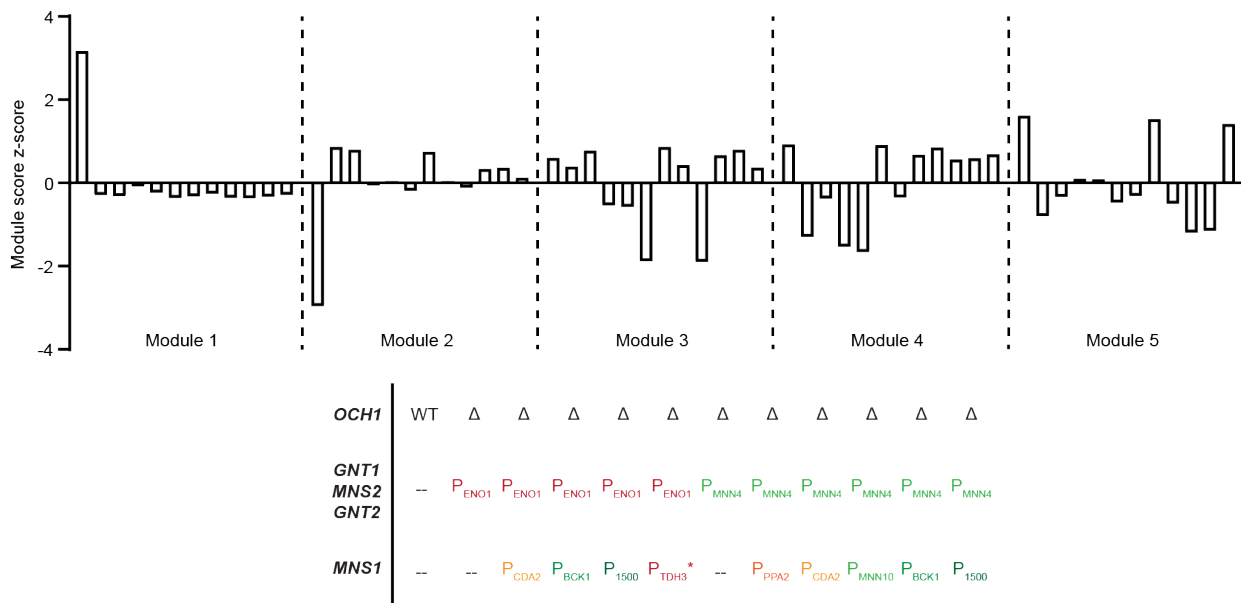
Figure A4.6. Scores of five different modules defined by sPCA.
Module scores are calculated by the matrix multiplication of module loading and gene expression.
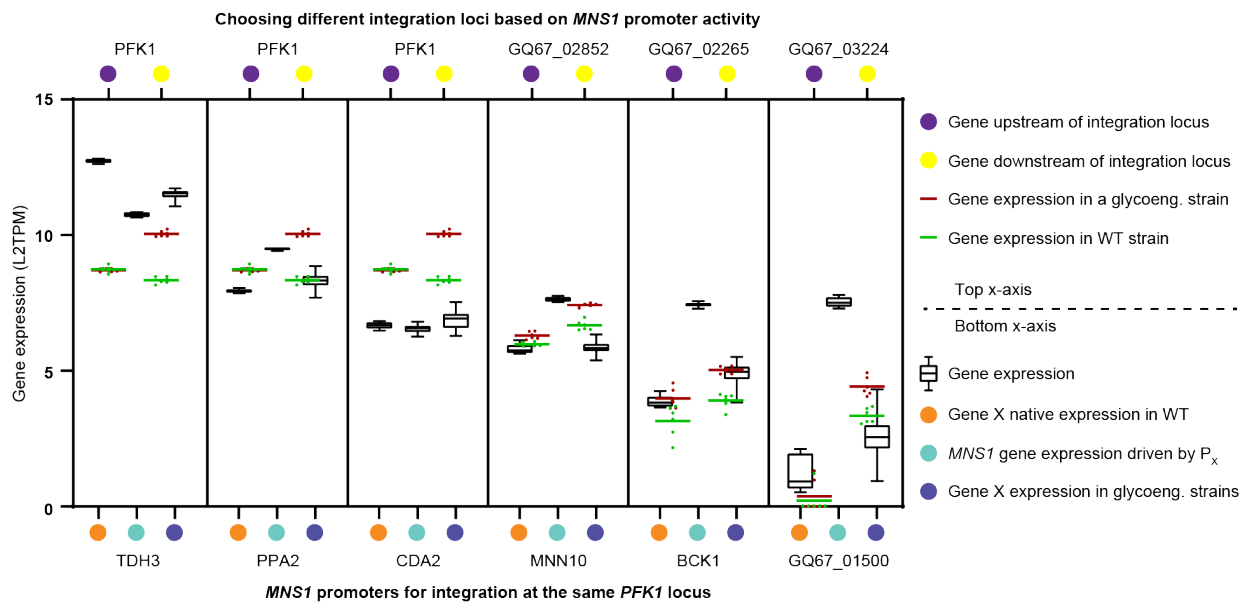
Figure A4.7. Selection of alternative integration loci for *MNS1*.
New loci (near genes *GQ67_02852*, *GQ67_02265*, and *GQ67_03224*) are chosen based on
their upstream and downstream gene expression levels, which ideally match up with *MNN10*,
BCK1, and *GQ67_01500* native gene expression, respectively.
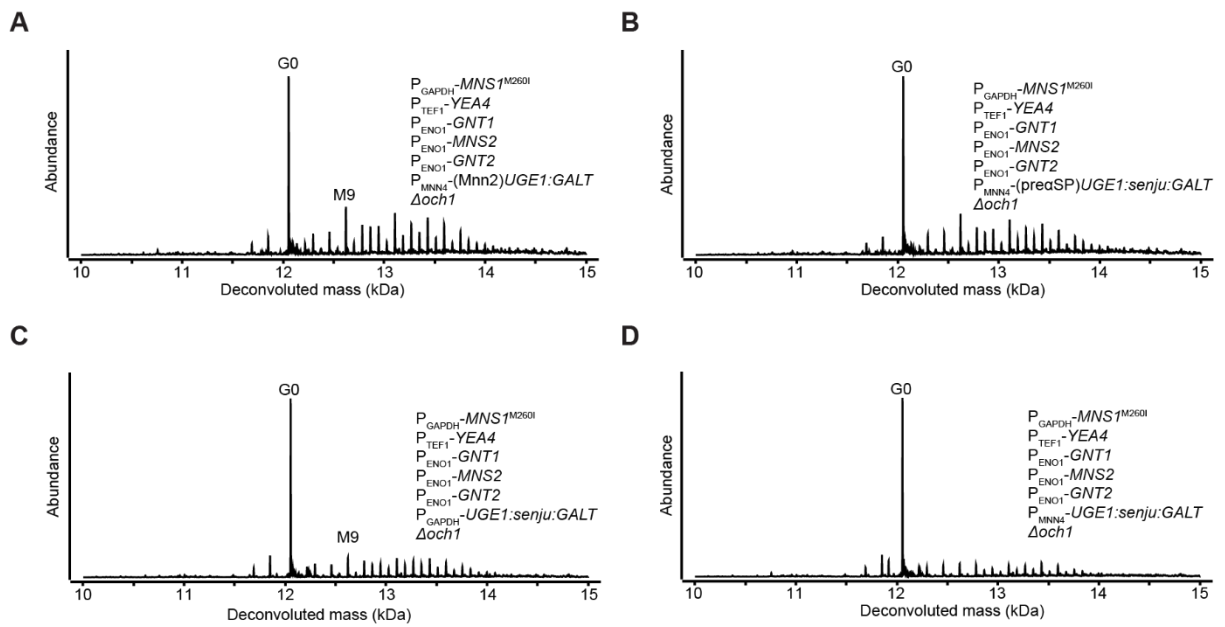
Figure A4.8. Mass spectra of additional galactosylation-engineered strains.
(A) P_{MNN4} was likely too weak to support galactosylation. (B), (C), and (D) tripartite protein fusion's lack of enzymatic activity was possibly due to protein misfolding or incorrect subcellular localization.
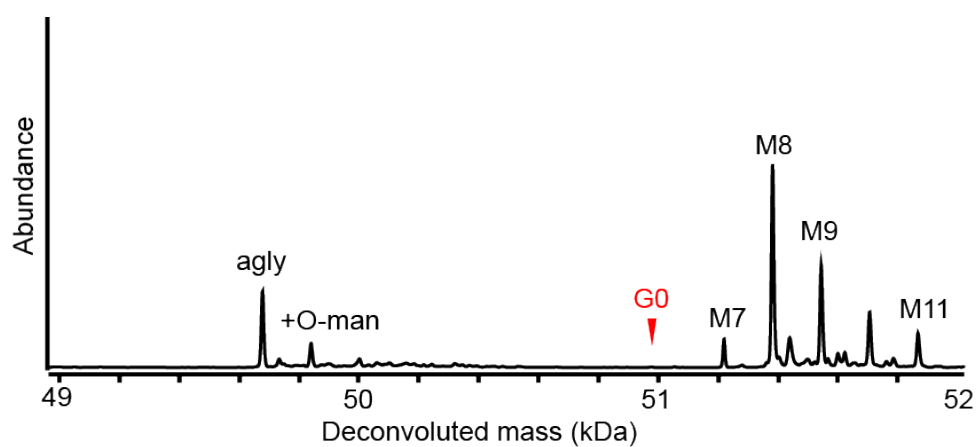
Figure A5.1. Mass spectrum of trastuzumab heavy chain.
Trastuzumab integration was carried out using the dual expression vector. EAEA residues were added to both LC and HC.
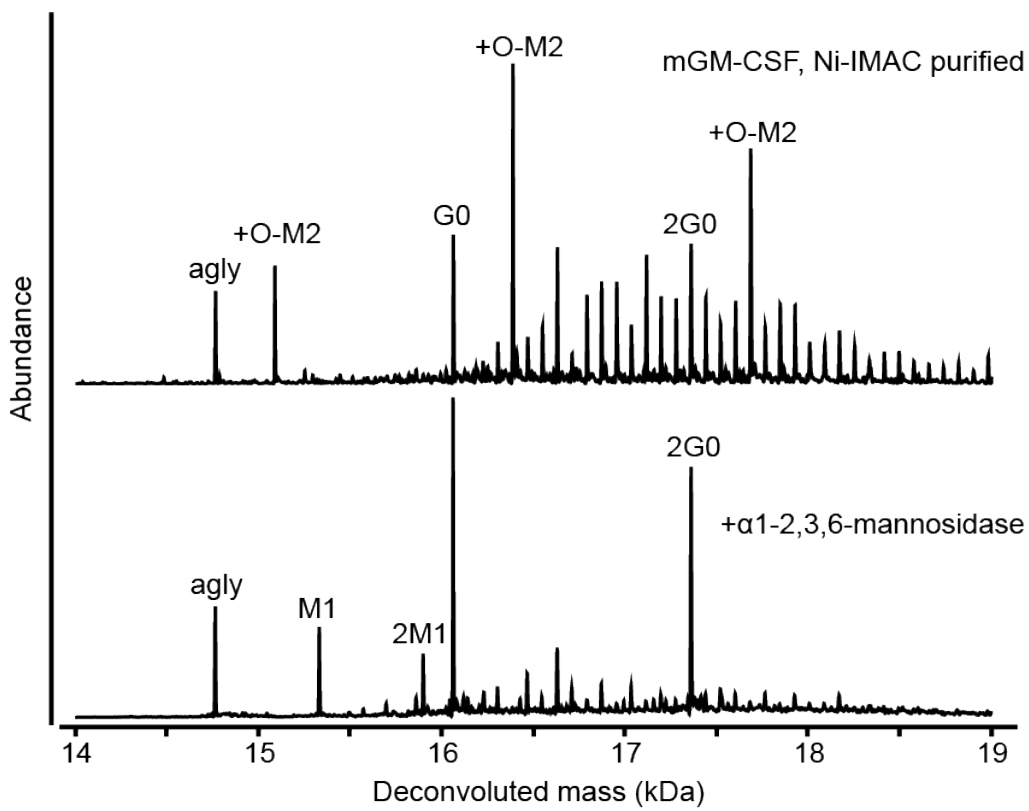
Figure A5.2. Mass spectra of mGM-CSF with mannosidase digestion.

Figure A5.3. Alignment of an uncharacterized *K. phaffii* protein to *S. cerevisiae* Yos9. BLAST identified homology only within the glucosidase domain of ScYos9 and not in the OS9 domain. *K. phaffii* ERAD mechanism remains to be studied.
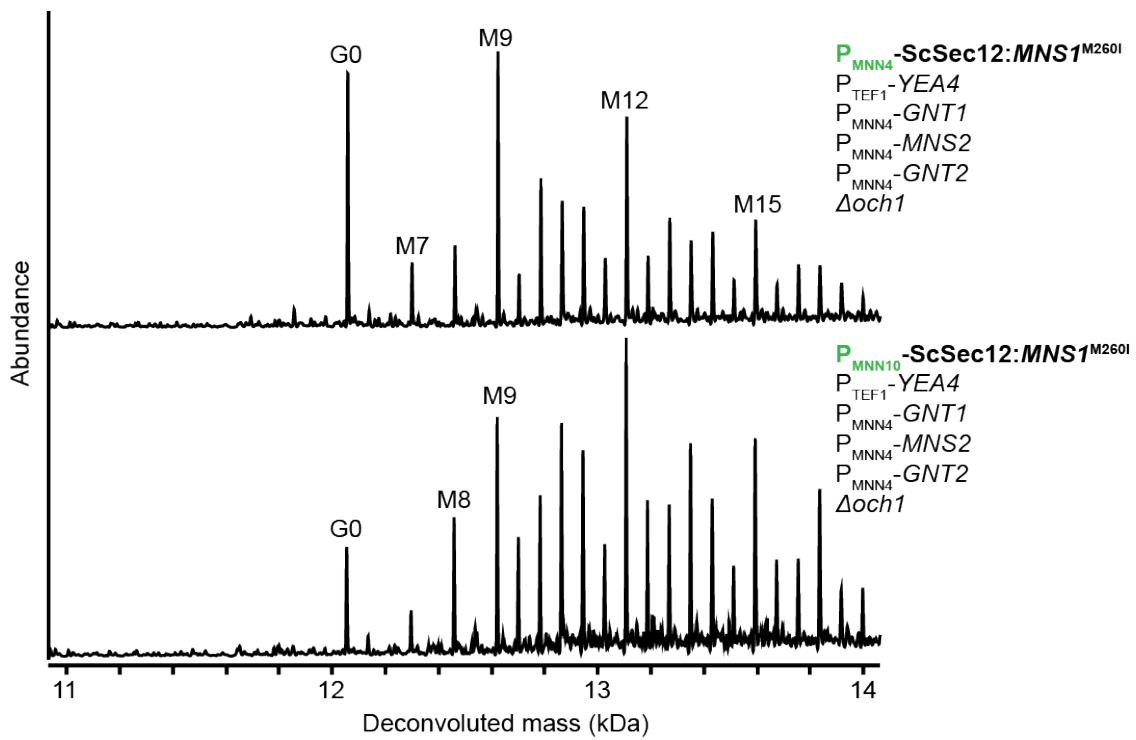
Figure A5.4. Mass spectra of K3 peptide in G0 glycoengineered strains with differently targeted mutant *MNS1*.
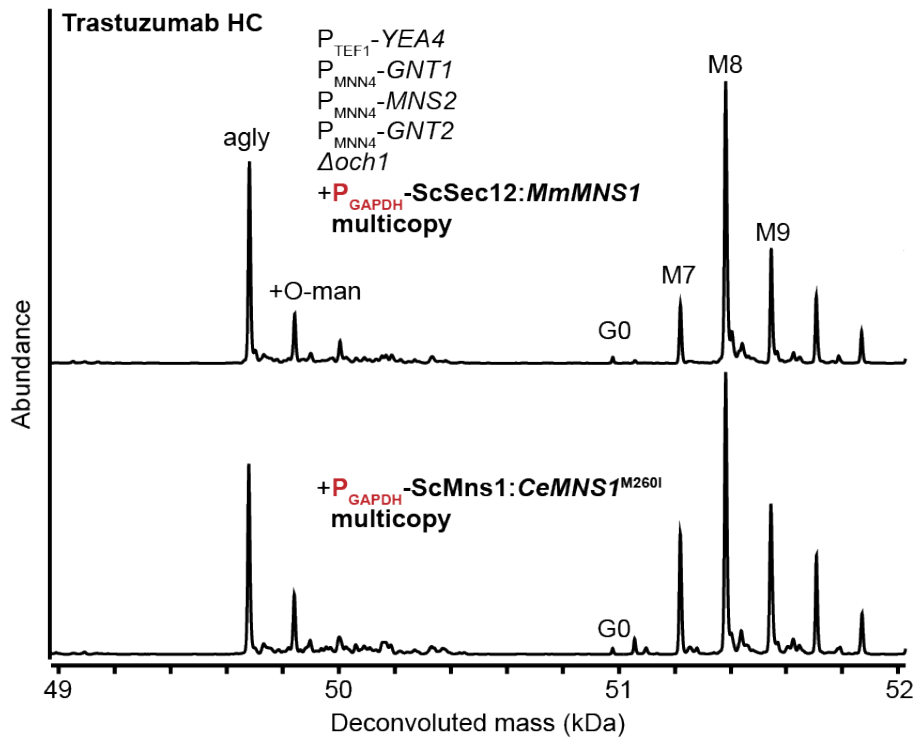
Figure A5.5. Mass spectra of trastuzumab HC in G0 glycoengineered strains with multi-copy integration of *MNS1*.
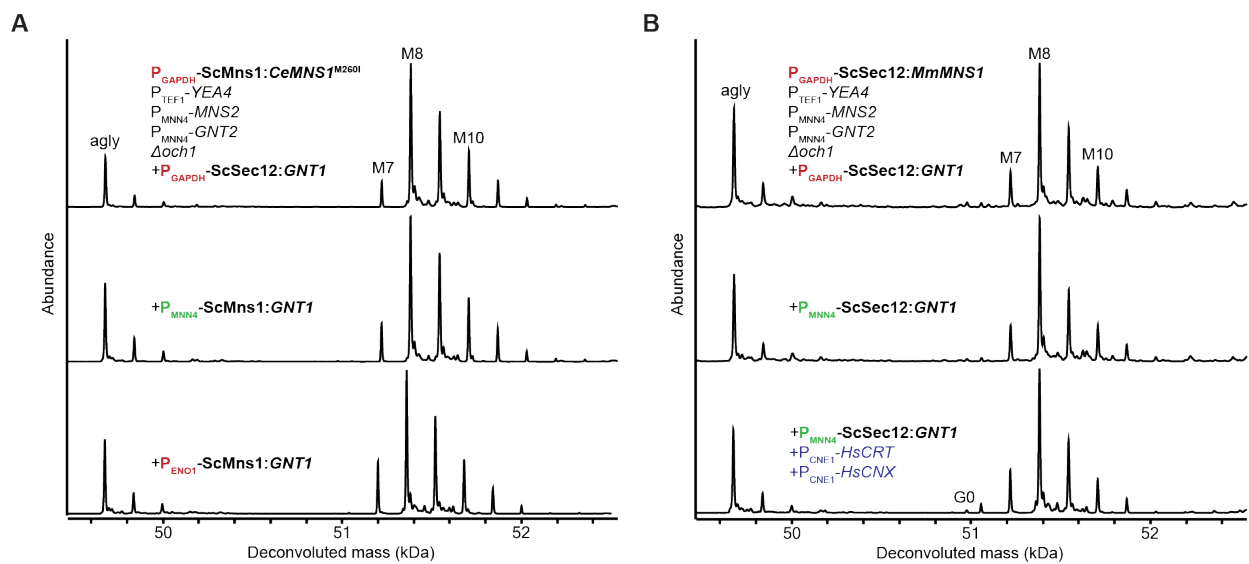
Figure A5.6. Mass spectra of trastuzumab HC in G0 glycoengineered strains with differently-targeted *GNT1*.
(A) in strains with ER-targeted mutant *MNS1*. (B) in strains with late ER/early Golgi-targeted *MmMNS1*.
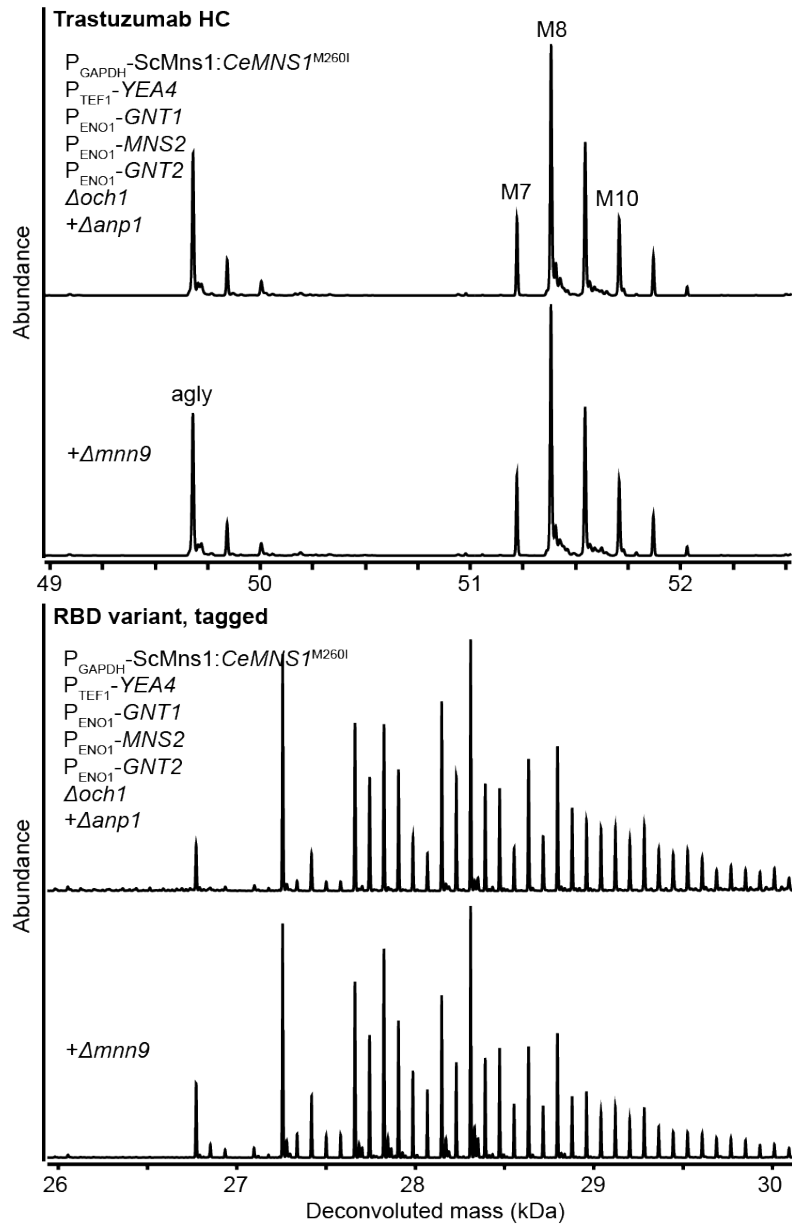
Figure A5.7. Mass spectra of trastuzumab HC and RBD in glycoengineered strains with additional mannosyltransferase knockout.
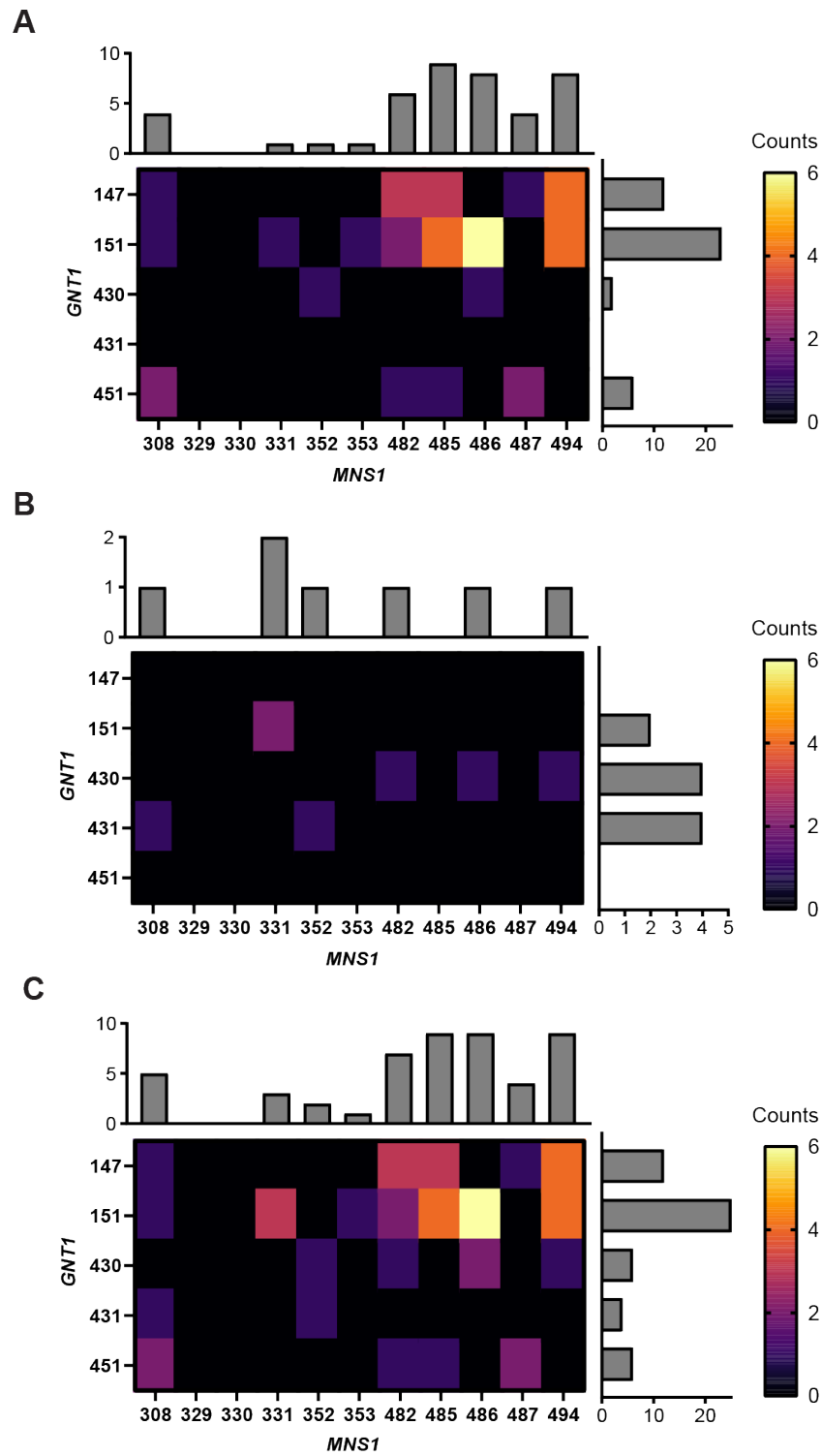
Figure A5.8. Library diversity assessment via _E. coli_ colony PCR sequencing.
(A) Round 1, (B) Round 2, and (C) combined diversity assessment through confirming the identity of individual integration constructs in distinct _E. coli_ colonies.

Chapter 6 supplemental figure(s)
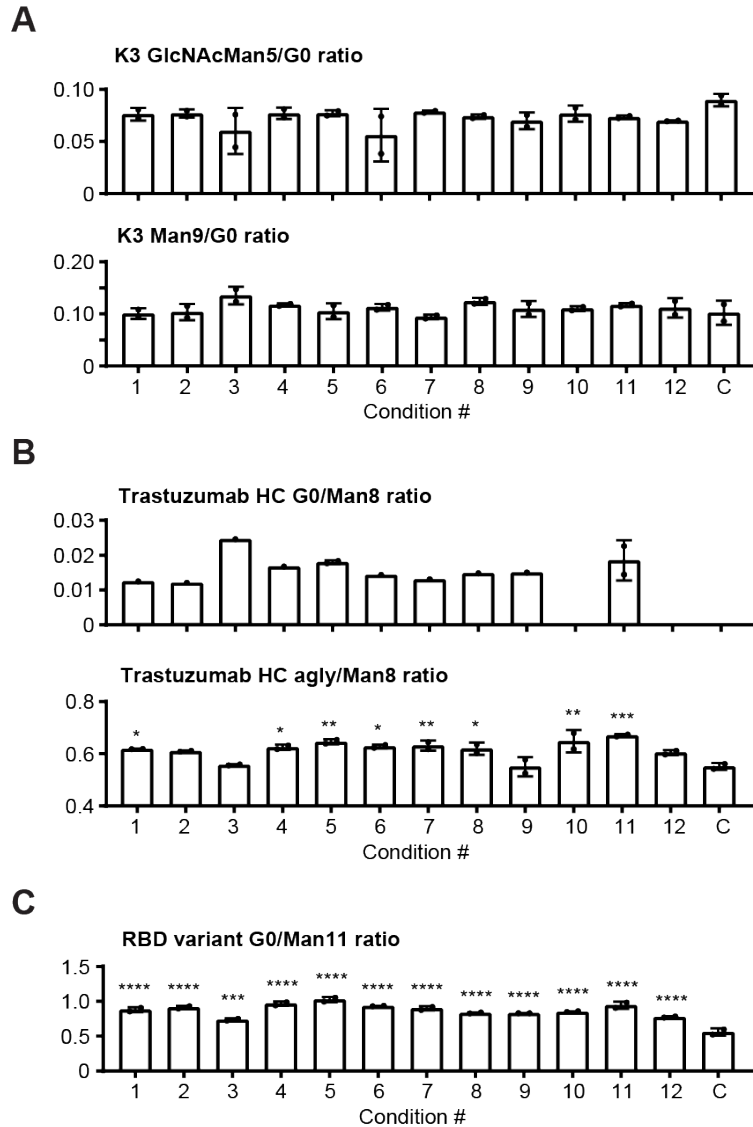
Figure A6.1. Observations of glycan profiles during initial exploration phase of round 1 optimization.
(A) K3 peptide glycan profile was robust against different media additives. (B) Improvements to trastuzumab G0 abundance were insignificant. In many conditions, media supplementation increased relative aglycosylated fraction. (C) RBD exhibited increased G0 glycosylation even during exploration phase. *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.0005$, ****: $p < 0.0001$, one-way ANOVA.

A

Trastuzumab HC agly/Man8 ratio

B

mGM-CSF occupancy score
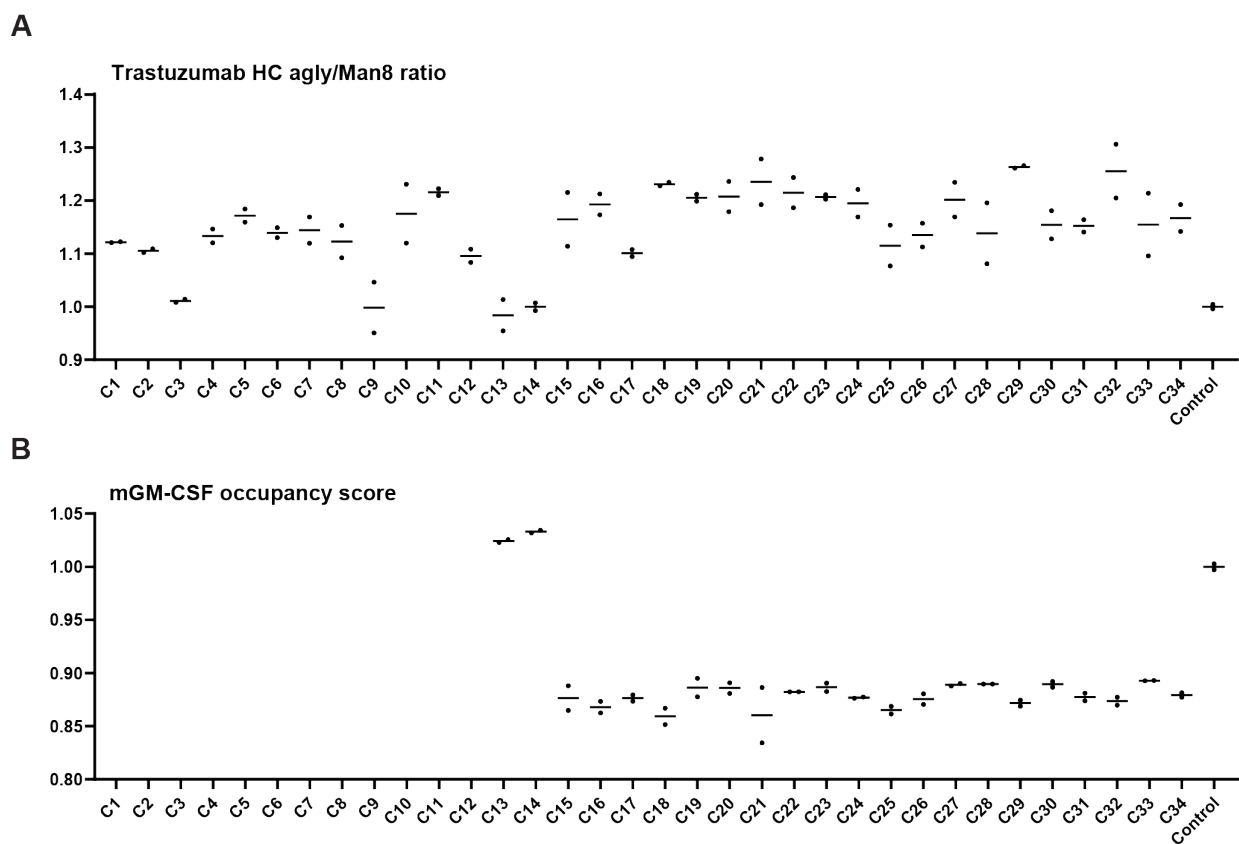
Figure A6.2. Effects of media supplementation on the glycan profile of non-target proteins.

mGM-CSF occupancy score is defined as:

$$\frac{2 * h_{2G0} + 2 * h_{2G0+OM2} + h_{G0} + h_{G0+OM2}}{2 * (h_{agly} + h_{agly+OM2} + h_{G0} + h_{G0+OM2} + h_{2G0} + h_{2G0+OM2})}$$

where h_X is the peak height of X, per LC-MS analysis.

Appendix A

Chapter 3 supplementary table(s)

Table B3.1. Carbon source and additives included in media screen for the discovery of native tunable promoters.

Chemical	as additive	as sole carbon source
Acetate		●
Arabinose	●	●
Aspartate		●
Caffeine	●	
Cellobiose		●
Dulcitol	●	●
Ethanol		●
Fructose		●
Galactose		●
Glucose		●
Glutamine		●
Inositol	●	●
Lactate		●
Lactose		●
Maltose		●
Mannitol	●	●
Mannose	●	
Mannose		●
Melibiose		●
Methanol	●	●
Pyruvate		●
Raffinose	●	●
Rhamnose		●
Ribose		●
Serine		●
Sorbitol	●	●
Succinate		●
Thiamine	●	
Trehalose		●
Xylitol		●
Xylose	●	●

Chapter 4 supplemental table(s)

Table B4.1. Transformation success of *MNS1* constructs with different promoters.
Parent strains are either P_{ENO1}-*GNT1/MNS2/GNT2* strain or P_{MNN4}-*GNT1/MNS2/GNT2* strain.

WT *MNS1* promoter	P_{ENO1}-G0	P_{MNN4}-G0
P_{ENO1}	✘	✘
P_{PPA2}	✘	✓
P_{CDA2}	✓	✓
P_{MNN4}	✘	✘
P_{MNN10}	✘	✓
P_{BCK1}	✓	✓
P_{1500}	✓	✓

Table B4.2. MAPK cascade targets for knockout.
* FUS3 knockout resulted in an amino acid substitution.

Gene name	Enrichment score	Non-essentiality score	Target for KO	KO success	Description
SLT2	29.97	0.871	✓		Serine/threonine MAP kinase
RTC1	29.25	0.785	✓	✓	Putative protein of unknown function
FUS3	28.69	0.998	✓	✓*	hypothetical protein PAS_chr2-1_0872
SST2	27.42	0.872	✓	✓	GTPase-activating protein for Gpa1p, regulates desensitization to alpha factor pheromone
STE3	24.11	0.819	✓	✓	Receptor for a factor receptor, transcribed in alpha cells and required for mating by alpha cells
MSB2	22.48	0.844	✓	✓	Mucin family member
GQ67_04899	21.62	0.583			Cytosolic aspartate aminotransferase, involved in nitrogen metabolism
GQ67_00187	20.47	0.769	✓	✓	Histidine kinase osmosensor that regulates a MAP kinase cascade
GQ67_03165	20.38	0.744			Glucose-repressible protein kinase involved in signal transduction during cell proliferation
NPR2	18.59	0.732			Protein with a possible role in regulating expression of nitrogen permeases
GQ67_02248	18.31	0.320			Hypothetical protein PAS_chr1-1_0399
HEM2	17.96	0.033			Delta-aminolevulinate dehydratase, a homo-octameric enzyme
GQ67_00249	17.02	0.073			Protein kinase involved in regulating diverse events including vesicular trafficking and DNA repair
GQ67_04558	14.11	0.935	✓	✓	Hypothetical protein PAS_chr4_0235
MRS6	13.77	0.442			Rab escort protein, forms a complex with the Ras-like small GTPase Ypt1p
STE50	13.69	0.742			Protein involved in mating response, invasive/filamentous growth, and osmotolerance
GQ67_00145	12.70	0.803			GTPase activating protein (GAP) for Rho1p, involved in signaling to the actin cytoskeleton
SLG1	12.33	0.704			Sensor-transducer of the stress-activated PKC1-MPK1 kinase pathway
GCN1	12.26	0.955			Positive regulator of the Gcn2p kinase activity, forms a complex with Gcn20p
GEA2	11.72	0.421			Guanine nucleotide exchange factor for ADP ribosylation factors (ARFs)

Table B4.3. Modules generated from sparse PCA analysis.

Module 1		Module 2		Module 3		Module 4		Module 5	
Gene	Loading	Gene	Loading	Gene	Loading	Gene	Loading	Gene	Loading
SSC1	-0.00279	KAR4	0.636132	MF(ALPHA)1	0.261436	FLO11-BSC1	0.660747	CTR1	0.595103
SSE1	-0.0235	FUS3	0.530194	ALD5	0.067553	USV1	0.3392	YFL040W	0.52904
ATP2	-0.03379	ASP3-4	0.238323	ADH2	0.037582	FLO9	0.228262	FRE1	0.41481
GUT1	-0.05968	PRM1	0.223132	PHO84	0.034055	MF(ALPHA)1	0.212193	MAC1	0.200819
PYC2	-0.06118	MF(ALPHA)1	0.213856	AQR1	0.030193	YFL040W	0.207765	SAH1	0.19466
FAS1	-0.06784	SST2	0.190155	MRP49	0.025317	BAG7	0.077969	ISR1	0.162165
KAR2	-0.07717	IME2	0.183556	ERG11	0.024884	SNN1	0.026187	GAP1	0.156609
ACO1	-0.10623	STE3	0.144467	YIM1	0.023672	ECI1	0.025926	ALD5	0.153168
SSA4	-0.18275	PRP18	0.051176	OCH1	0.021843	PDR15	0.002923	INO1	0.109415
HSP82	-0.20152	STE12	0.044055	CTR1	0.020519	ERG11	-0.0022	ROX1	0.096098
GUT2	-0.21007	RSN1	0.041337	DOG1	0.020178	OPI3	-0.00364	IME2	0.067513
EFT1	-0.26873	STE5	0.02611	GDT1	0.016155	CTR1	-0.00402	DIP5	0.065704
PMA1	-0.30405	DFG5	0.014971	YEL047C	0.014523	STE14	-0.00645	OPI3	0.058165
TEF1	-0.32136	SAH1	-0.01183	NMA1	0.014401	GDH3	-0.00856	GCW14	0.050018
PCK1	-0.52231	ADH2	-0.01578	FRE3	0.014109	YDL218W	-0.01615	MET6	0.048126
HSP104	-0.55701	CTR1	-0.02582	PDC1	0.010859	PRY2	-0.01979	MCH4	0.043698
		AQR1	-0.047	HEM13	0.009933	SPS4	-0.03223	YOL114C	0.027391
		ERG2	-0.05336	MET13	0.009689	RAM1	-0.03284	PDR12	0.019923
		FRE3	-0.08282	YIL166C	0.008835	ZRT1	-0.04471	RGS2	0.007575
		ALD5	-0.22593	UIP4	0.008187	STE6	-0.06016		
				AGC1	0.006784	CAR1	-0.0664		
				CTH1	0.006628	INO1	-0.09839		
				GCW14	0.00637	NIT1	-0.10228		
				YBR096W	0.005601	AXL1	-0.2988		
				ENO1	0.005005	STE2	-0.42219		
				MSF1	0.004109				
				ERG25	0.002325				
				YFH7	0.00146				
				YCR090C	0.001096				
				YPR098C	0.000707				
				CCP1	0.000474				
				YAH1	-0.00049				
				MSC1	-0.00202				
				SGA1	-0.00271				
				AGP2	-0.00316				
				SLT2	-0.00496				
				SRC1	-0.00622				
				YGR127W	-0.0069				
				CRH1	-0.00921				
				YOR1	-0.01325				
				YOR338W	-0.01367				
				ECI1	-0.0156				
				ISR1	-0.01976				
				FLO9-like2	-0.02054				
				RAM1	-0.02991				
				RAM2	-0.03132				
				STE14	-0.03423				
				PUN1	-0.0441				
				HSP12	-0.04836				
				ISU1	-0.05038				
				STE6	-0.05513				
				BAG7	-0.08652				
				FLO11-BSC1	-0.14344				
				AXL1	-0.17254				
				USV1	-0.17584				
				FLO9	-0.1946				
				STE2	-0.25708				
				FLO11-MUC1	-0.84361				

Chapter 5 supplemental table(s)

Table B5.1. Constructs included in *MNS1-GNT1* combined library.

Gene	Construct #	Promoter	Localization		Catalytic	Note
MNS1	308	P_{GAPDH}	ScMns1	ER, tethered	*CeMNS1*[M260I]	Base construct for comparison
	329	P_{MNN4}	ScSec12	Late ER/Early Golgi	*CeMNS1*	Change *MNS1* localization from ER to late ER/early Golgi via different transmembrane domain
	330	P_{MNN10}	ScSec12	Late ER/Early Golgi	*CeMNS1*	
	331	P_{CDA2}	ScSec12	Late ER/Early Golgi	*CeMNS1*	
	352	P_{GAPDH}	ScSec12	Late ER/Early Golgi	*MmMNS1*	
	353	P_{MNN4}	ScSec12	Late ER/Early Golgi	*MmMNS1*	
	482	P_{BCK1}	HDEL	ER, lumenal	*CeMNS1*	Change *MNS1* from membrane-tethered to ER lumen-targeted
	485	P_{MNN4}	HDEL	ER, lumenal	*CeMNS1*	
	486	P_{ENO1}	HDEL	ER, lumenal	*CeMNS1*	
	487	P_{1500}	HDEL	ER, tethered	*CeMNS1*	Weaker promoter for WT *CeMNS1* expression
	494	P_{4776}	HDEL	ER, tethered	*CeMNS1*	
GNT1	147	P_{ENO1}	ScMnn2	Golgi	*HsGNT1*	Base constructs for comparison
	151	P_{MNN4}	ScMnn2	Golgi	*HsGNT1*	
	430	P_{GAPDH}	ScSec12	Late ER/Early Golgi	*HsGNT1*	Change *GNT1* localization from Golgi to late ER/early Golgi via different transmembrane domain
	431	P_{MNN4}	ScSec12	Late ER/Early Golgi	*HsGNT1*	
	451	P_{ENO1}	ScSec12	Late ER/Early Golgi	*HsGNT1*	

Chapter 6 supplemental table(s)

Table B6.1. Design space of additives in machine learning-guided media supplementation study.
*: adjusted to 1 mM after round 1. **: additive included in round 3.

Supplementation	Effect	Design space (mM)	
Mn^{2+}	Increased M5	0.001	0.04*
	Increased occupancy		
	Increased galactosylation		
Galactose	Increased galactosylation	0	100
NH_4^+	Decreased galactosylation	0	10
GlcNAc	Increased G0	0	20
ManNAc	Increased G0	0	20
Uridine	Potential synergy with Mn2+ and galactose	0	20
Ca^{2+}	*MNS1* cofactor	0	5**

Table B6.2. Media conditions tested based on the described Bayesian optimization algorithm.

	Condition #	Concentration (mM)						
		MnCl$_2$	Galactose	NH$_4$Cl	GlcNAc	ManNAc	Uridine	CaCl$_2$
Round 1	1	0.008	29.20	1.20	4.20	5.80	5.80	
	2	0.012	37.50	7.10	12.50	2.50	17.50	
	3	0.022	4.20	2.90	17.50	0.80	0.80	
	4	0.035	79.20	3.80	15.80	10.80	15.80	
	5	0.005	70.80	0.40	14.20	9.20	19.20	
	6	0.032	95.80	5.40	19.20	15.80	14.20	
	7	0.038	54.20	2.10	9.20	7.50	7.50	
	8	0.028	20.80	8.80	7.50	4.20	2.50	
	9	0.025	45.80	4.60	2.50	17.50	10.80	
	10	0.015	62.50	9.60	10.80	19.20	9.20	
	11	0.002	87.50	6.20	0.80	12.50	4.20	
	12	0.018	12.50	7.90	5.80	14.20	12.50	
Round 2	13	1.000	0.00	5.20	0.00	1.84	8.71	
	14	1.000	0.00	9.98	0.00	0.00	15.17	
	15	0.585	26.33	5.64	5.26	6.62	11.57	
	16	0.262	37.66	7.57	10.00	3.72	14.88	
	17	0.656	23.68	3.28	5.51	16.12	8.90	
	18	0.218	73.22	8.26	4.27	8.89	6.30	
	19	0.293	29.72	5.70	15.90	5.25	14.14	
	20	0.214	43.90	7.21	5.67	10.34	15.68	
	21	0.390	52.83	4.58	12.13	4.85	10.81	
	22	0.806	63.83	7.64	11.07	9.87	3.92	
	23	0.560	54.25	3.97	15.37	10.78	5.39	
Round 3	24	0.243	29.66	3.57	13.24	5.60	4.97	4.28
	25	0.003	6.21	3.96	15.26	15.96	1.46	4.77
	26	0.129	35.47	8.80	15.79	1.28	16.92	1.90
	27	0.905	12.04	3.75	8.33	1.60	2.36	4.11
	28	0.968	15.84	2.46	0.55	17.29	17.22	4.61
	29	0.238	88.35	9.60	4.60	8.99	15.68	4.19
	30	0.236	15.72	8.46	13.77	2.21	14.82	3.98
	31	0.913	73.27	0.91	3.11	10.99	1.55	1.37
	32	0.599	97.35	3.24	1.69	19.17	0.43	2.78
	33	0.840	19.34	6.98	0.16	6.19	16.67	4.70
	34	0.797	73.28	1.27	19.49	10.11	9.27	1.48
	Control	0	0	0	0	0	0	0