Improving designer glycan production in Escherichia coli through model-guided metabolic engineering

Joseph A. Waymana, Cameron Glasscockb, Thomas J. Mansella, Matthew P. DeLisab, Jeffrey D. Varnerb,*

a School of Applied and Engineering Physics, Cornell University, Ithaca, NY, 14853, USA
b Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA

ARTICLE INFO

Keywords:
Glycosylation
Metabolic modeling
Microbial glycobiology
N-linked glycan
Strain engineering

ABSTRACT

Asparagine-linked (N-linked) glycosylation is the most common protein modification in eukaryotes, affecting over two-thirds of the proteome. Glycosylation is also critical to the pharmacokinetic activity and immunogenicity of many therapeutic proteins currently produced in complex eukaryotic hosts. The discovery of a protein glycosylation pathway in the pathogen Campylobacter jejuni and its subsequent transfer into laboratory strains of Escherichia coli has spurred great interest in glycoprotein production in prokaryotes. However, prokaryotic glycoprotein production has several drawbacks, including insufficient availability of non-native glycan precursors. To address this limitation, we used a constraint-based model of E. coli metabolism in combination with heuristic optimization to design gene knockout strains that overproduced glycan precursors. First, we incorporated reactions associated with C. jejuni glycan assembly into a genome-scale model of E. coli metabolism. We then identified gene knockout strains that coupled optimal growth to glycan synthesis. Simulations suggested that these growth-coupled glycan overproducing strains had metabolic imbalances that rerouted flux toward glycan precursor synthesis. We then validated the model-identified knockout strains experimentally by measuring glycan expression using a flow cytometric-based assay involving fluorescent labeling of cell surface-displayed glycans. Overall, this study demonstrates the promising role that metabolic modeling can play in optimizing the performance of a next-generation microbial glycosylation platform.

1. Introduction

Protein glycosylation is the attachment of glycans (mono-, oligo-, or polysaccharide) to specific amino acid residues in proteins, most commonly asparagine (N-linked) or serine and threonine (O-linked) residues. Roughly three-quarters of eukaryotic proteins and more than half of prokaryotic proteins are glycosylated (Dell et al., 2010). Glycosylation is also vitally important to the development of many protein biologics, and has been harnessed for enhancing therapeutic properties such as half-life extension (Elliott et al., 2003; Flintegaard et al., 2010; Ilyushin et al., 2013; Lindhout et al., 2011), antibody-mediated cytotoxicity (Li et al., 2017; Lin et al., 2015), and immunogenicity (Lipinski et al., 2013; Sadoulet et al., 2007; Wacker et al., 2014).

Though once thought to occur only in eukaryotes, protein glycosylation has now been discovered in all three domains of life, including bacteria (Nothaft and Szymanski, 2010). The best characterized bacterial N-glycosylation system is that of the human pathogen Campylobacter jejuni (Szymanski et al., 1999). The C. jejuni glycan has the form of a branched heptasaccharide Glc GalNAc 2 Bac, where Glc is glucose, GalNAc is N-acetylgalactosamine, and Bac is bacillosamine. This glycan is assembled on the lipid carrier undecaprenyl pyrophosphate (Und-PP) on the cytoplasmic face of the inner membrane by an enzymatic pathway encoded by the pgl (protein glycosylation) locus (Fig. 1). The fully assembled glycan is flipped across the membrane and transferred to asparagine residues in acceptor proteins by the oligosaccharyltransferase (OST) PglB. PglB attaches the heptasaccharide to periplasmically-localized proteins containing the consensus sequence D/E-X-N-Z-S/T, where X and Z are any residue except proline (Fisher et al., 2011; Kowarik et al., 2006).

The functional transfer of this system into E. coli (Wacker et al., 2002) has spurred interest in recombinant production of glycans and ultimately therapeutic glycoproteins in this genetically tractable bacterial host (Merritt et al., 2013; Baker et al., 2013). Along these lines, glycosylation-competent E. coli cells have been used to produce a variety

* Corresponding author.

E-mail address: jdv27@cornell.edu (J.D. Varner).

https://doi.org/10.1016/j.mec.2019.e00088

Received 7 July 2017; Received in revised form 13 March 2019; Accepted 14 March 2019

© 2019 The Authors. Published by Elsevier B.V. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-ND license (http://creativecommons.org/licenses/by-nd/4.0/).
of periplasmic and extracellular glycoproteins including antibodies (Fisher et al., 2011) and conjugate vaccine candidates (Feldman et al., 2005). The promiscuity of the PglB enzyme towards structurally diverse lipid-linked glycan substrates has been exploited to further expand the *E. coli* platform, enabling the creation of glycoproteins bearing different bacterial O-polysaccharide antigens (Feldman et al., 2005; Ihssen et al., 2010) and even the eukaryotic trimannosyl core N-glycan produced by a synthetic pathway comprised of four yeast glycosyltransferases (Valderrama-Rincon et al., 2012). However, while PglB can efficiently glycosylate native *C. jejuni* acceptor proteins with cognate GlcGalNAc5 Bac glycan in engineered *E. coli*, glycosylation of non-Campylobacter target proteins is often much less efficient (Schwarz et al., 2010), especially in combination with heterologous glycan structures (Valderrama-Rincon et al., 2012).

In engineered *E. coli*, protein glycosylation is affected by the availability of lipid carriers, and the availability of nucleotide-activated sugar substrates serving as glycan precursors (Merritt et al., 2013; Jaffé et al., 2014). Hence, a plausible strategy for increasing glycosylation efficiency is to optimize the levels of these key reaction intermediates and their related biosynthetic pathways. Along these lines, Wright and coworkers applied genome-scale metabolic engineering techniques to improve glycosylation efficiency in *E. coli*. Using a high-throughput proteomic screening and probabilistic metabolic network analysis, they showed that upregulation of the glyoxylate cycle by overexpression of isocitrate lyase (aceA/icl) increased glycosylation efficiency of a prototypic protein by three-fold (Pandhal et al., 2011). Further, genome-wide screening of gene overexpression identified targets that increased glycoprotein production as well as glycosylation efficiency (Pandhal et al., 2013); genes in pathways associated with glycan precursor synthesis (UDP-GlcNAc) as well as lipid carrier production (isopenoid synthesis) were identified as bottlenecks. Improved glycosylation efficiency has also been achieved by supplementing growth media with GlcNAc (Kämpf et al., 2015) or increasing the expression of PglB via codon optimization (Pandhal et al., 2012). These studies and others have demonstrated the complex interplay between recombinant protein production, glycan synthesis and assembly, and glycosylation efficiency.

In this study, we addressed one of the challenges facing high-level glycoprotein production in engineered *E. coli*, namely the availability of glycan precursors, using constraint-based modeling. In particular, we used a constraint-based model of *E. coli* metabolism, in combination with heuristic optimization, to design gene knockout strains that overproduced glycan precursors. First, we incorporated reactions associated with *C. jejuni* glycan assembly into a genome-scale model of *E. coli* metabolism. We then used a combination of constraint-based modeling and simulated annealing to identify gene knockout strains that coupled optimal growth to glycan synthesis. Simulations suggested that these growth-coupled glycan overproducing strains had metabolic imbalances that rerouted flux toward glycan precursor synthesis. We then experimentally validated the model-identified metabolic designs using a flow cytometric-based assay for quantifying cellular N-glycans in *E. coli* (Valderrama-Rincon et al., 2012). Consistent with simulations, the best model-predicted changes increased glycan production by nearly 3-fold compared with the glycan production level in wild-type (wt) *E. coli* cells. Taken together, our results reveal the significant impact that metabolic modeling can have on designing chassis strains with enhanced N-linked protein glycosylation capabilities.

### 2. Results

#### 2.1. Construction of a constraint-based model of N-linked glycosylation in *E. coli*

A constraint-based model of N-glycosylation in *E. coli* was used to identify genetic knockouts that coupled glycan biosynthesis with optimal growth. We augmented the existing genome-scale *E. coli* model iAF1260 from Palsson and coworkers (Peist et al., 2007) to include the reactions of the *C. jejuni* glycosylation pathway (Table 1). The adapted network consisted of 2395 reactions, 1271 open reading frames, and 1986 metabolites segregated into cytoplasmic, periplasmic, and extracellular compartments. Added reactions included the biochemical transformations catalyzed by the glycosyltransferases (e.g., PglA, PglC) associated with glycan biosynthesis, PglK flipase-mediated translocation of the glycan into the periplasm, and PglB-mediated glycan conjugation to an acceptor protein (Fig. 1). In addition, we incorporated the transcriptional regulatory network of Covert et al., consisting of 101 transcription factors, regulating the state of the metabolic genes (Covert et al., 2004). This regulatory network imparts Boolean constraints on metabolic fluxes based upon the nutrient environment. The model code is available for download under an MIT software license from the Varnerlab website (http://www.varnerlab.org/).

#### 2.2. Identification of growth-coupled gene knockout strains

To identify genetic knockouts that coupled optimal growth to glycan
biosynthesis, we used heuristic optimization and the constraint-based model (see Materials and Methods). Coupling growth to glycan synthesis was desirable for several reasons. Foremost amongst these, growth-coupled strains create stoichiometric imbalances that reroute metabolic flux toward the desired product as a consequence of growth (Burgard et al., 2003; Feist et al., 2010). Therefore, faster growth requires increased glycan formation. Thus, optimizing glycan production through adaptive evolution is made trivial by selecting for growth through serial passage (Feist et al., 2007). Therefore, faster growth requires increased glycan formation. Thus, optimizing glycan production through adaptive evolution is made trivial by selecting for growth through serial passage (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]

Thus, if any of the genes aceE, aceF, or lpd was knocked out or transcriptionally repressed, the flux through this reaction was bound to zero. Gene-protein-reaction (GPR) associations from the iAF1260 network were used in this study (Feist et al., 2007). The simulated annealing algorithm performed a random search of genetic knockouts, iteratively indicated a genetic knockout or regulatory repression. Boolean rules informed by nutrient conditions controlled the TF genes, which in turn controlled the state of the metabolic genes. Once defined, the genetic state of the model modified the flux constraints placed on each reaction. For example, the reaction governed by pyruvate dehydrogenase, a multi-component enzyme, relied on the assembly of three enzymes: AceE, AceF, and Lpd. This reaction was encoded as:

\[
\text{Pyruvate} + \text{CoA} + \text{NAD} \rightarrow \text{Acetyl-CoA} + \text{Acetate} + \text{NADH}
\]
applying flux constraints based on the genetic state, then performing a flux balance analysis simulation. To identify growth-coupled glycan producing strains, we optimized the shadow price given by:

$$u_{glycan} = \frac{\Delta r_{growth}}{\Delta r_{glycan}}$$

(1)

where $\Delta r_{growth}$ denotes the change in growth rate for a forced change in glycan flux $\Delta r_{glycan}$, and $u_{glycan}$ denotes the flux representing the fully assembled C. jejuni glycan flipped into the periplasm. The shadow price $u_{glycan}$ was calculated for a particular knockout strain by first calculating the optimal growth with the glycan flux constrained to zero. A second simulation was then performed with a forced incremental change in the glycan flux in order to obtain the difference in growth rate. The search algorithm continued until $u_{glycan} > 0$, indicating a growth-coupled phenotype (Supplementary Fig. S1).

We identified growth-coupled knockout strains with four or fewer knockouts for growth on glucose as the sole source of carbon and energy (Table 2). We performed optimization simulations using boundary conditions representing minimal medium with a single 6-, 5-, and 3-carbon substrate. A well-defined minimal media allowed for precise control over nutrient conditions experimentally, and was accurately simulated, particularly for the transcriptional regulatory network. For each substrate, we performed ten independent optimization simulations to identify growth-coupled strains. We considered growth-coupled strains with four or fewer knockouts (those most likely to be experimentally viable) by restricting the formation of extracellular byproducts to acetate. For example, for E. coli glycosylation mutant 2 (EcGM2; E. coli iAF1260 Δsdh Δgnd Δpta ΔeutD), the strain with the highest simulated glycan yield, the optimal growth rate occurs at a non-zero glycan flux (Supplementary Fig. S1B). All growth-coupled strains contained a knockout of succinate dehydrogenase ($sdh$) and truncated pentose phosphate pathway (PPP) flux at either glucose 6-phosphate-1-dehydrogenase ($swf$), 6-phosphogluconolactonase ($pgl$), or 6-phosphogluconate dehydrogenase ($gnd$).

2.3. Flux analysis of N-glycan production in growth-coupled strains

Growth-coupled glycan producing strains had increased glycolytic flux, and decreased amino acid biosynthesis compared to glycan production in the wt strain background (Fig. 3). We compared the normalized flux values for EcGM2 with the wt strain. Normalizing all fluxes to glucose uptake rate, EcGM2 displayed greater flux through glycolysis by cutting off the PPP via knockout of NADPH-producing $gnd$ (Fig. 3A). EcGM2 also had decreased synthesis of every amino acid except for glutamine, indicating a source of stoichiometric imbalance that may be relieved by synthesis of the glycan precursor UDP-glucose. Further, the PEP-pyruvate node acted as a switch point in central carbon metabolism (Fig. 3B). Here, PEP and pyruvate, the products of glycolysis, enter the TCA cycle through decarboxylation of pyruvate to acetyl-CoA (ACCOA) and carboxylation of PEP to form oxaloacetate (OAA) (Sauer and Eikmanns, 2005). The latter replenishes TCA cycle intermediates that exited TCA for anabolic processes. EcGM2, with a diminished anabolic capacity for cell growth, displayed lower flux through PEP carboxylase ($pcp$). However, as the result of high glycolytic flux, EcGM2 had increased flux through pyruvate dehydrogenase (aceEF), sending carbon into the oxidative branch of the TCA cycle. It is known that high glucose uptake rates result in excess acetyl-CoA, surpassing the capacity of the TCA cycle. Because of this excess flux, wt E. coli grown on glucose commonly displays acetate fermentation, even under aerobic conditions (Gosset, 2005). We observed increased acetate secretion in EcGM1 simulations, but through a route differing from wild-type cells. The knockouts $\Delta$pta and $\Delta$eutD prevented ATP-generating acetate secretion. Flux was instead routed through the redox-neutral reactions initiated by acetaldelyde dehydrogenase ($mhpF$). Excess acetyl-CoA was also utilized in the pathway generating UDP-GlcNAc. Lastly, EcGM2 displayed a shift in cofactor production (Fig. 3C). Higher flux through glycolysis naturally led to NADH overproduction. On the other hand, the primary source of NADPH shifted from PPP genes $swf$ and $gnd$ to the membrane transhydrogenase $pnt$, capable of direct transfer of electrons from NADH to NADP. Sauer et al. identified $pnt$ as a major source of NADPH in E. coli (35–45% of total) (Sauer et al., 2004). Thus, $pnt$ is capable of carrying significant flux in vivo. Taken together, these results suggested the model iHug and strains that promoted glycan precursor synthesis, primarily UDP-GlcNAc, by creating a combination of metabolite and redox imbalance.

2.4. Experimental validation of N-glycan-producing knockout strains

Glycan production was measured in the mutant strains to validate the model predictions (Fig. 4). Gene knockout strains were constructed using the Keio collection of single gene knockouts E. coli BW25113 (Baba et al., 2006) as donor strains for P1vir phage transduction. Mutants were constructed containing single, double, and triple knockouts that appeared in growth-coupled strains identified by the constraint-based model. We also performed simulations of each single gene knockout to determine genes that prevented glycan synthesis; galU, a key enzyme in the synthesis of glycan precursor UDP-glucose, was the only non-lethal knockout that prevented glycan synthesis. Knockout strains were transformed with a plasmid constitutively expressing the C. jejuni $pgl$ locus. To quantify glycan production, we took advantage of crossstalk between the glycosylation pathway and native lipopolysaccharide (LPS) synthesis in E. coli (Hug and Feldman, 2011). Specifically, after the glycan is flipped into the periplasm, it can be transferred to lipid A-core by the Waal O-antigen ligase and shuttled to the outer membrane by LPS pathway enzymes, where it is displayed on the cell surface (Merritt et al., 2013). Labeling of these surface-displayed N-glycans with fluorescently-tagged lectins can then be used to quantify the amount of glycan displayed on the cell surface as a measure of glycan production (Valderrama-Rincon et al., 2012). Here, we labeled C. jejuni glycans for detection by flow cytometry with fluorophore-conjugated soybean agglutinin (SBA), a lectin specific to terminal galactose and GalNAc residues. Prior to labeling, knockout strains were grown in glucose minimal media and harvested during the exponential growth phase, to most closely satisfy the pseudo-steady-state assumption of model predictions.

A common feature of the predicted mutant strains was the deletion of pentose phosphate pathway genes $swf/pgl/gnd$ in combination with $sdh$. Analysis of the metabolic flux distribution in these mutants suggested the reducing state of the cell as well as the carbon flux was reprogrammed to

| Table 2 |
| --- |
| Growth-coupled strains producing C. jejuni glycan identified by flux balance analysis and heuristic optimization using single carbon substrate. Knockouts listing multiple genes indicate that knockout of any one of those genes produces the same phenotype in the model. Abbreviations: D-Glucose, Glc; E. coli Wild type, EcWT; E. coli glycosylating mutant, EcGM. |
| **Strain** | **Substrate** | **Genotype** | **Growth rate (mmol/gDW/hr)** | **Glycan flux (mmol/gDW/hr)** | **Yield (mmol/gDW)** |
| --- | --- | --- | --- | --- | --- |
| EcWT | Glucose | Wild type | 0.78 | 0 | 0 |
| EcGM1 | Glucose | Δsdh Δzwf/pga/gnd | 0.65 | 0.012 | 0.018 |
| EcGM2 | Glucose | Δsdh Δzwf/pga/gnd Δpta ΔeutD | 0.53 | 0.098 | 0.185 |
| EcGM3 | Glucose | Δsdh Δzwf/pga/gnd ΔpykAF Δmdh | 0.64 | 0.016 | 0.025 |
support enhanced glycan biosynthesis. While hypothetical knockouts such as $\Delta$sdh $\Delta$zwf/pgl/gnd $\Delta$pta $\Delta$eutD were predicted to have higher glycan yield, in this study we experimentally evaluated only the simplest growth-coupled double knockout family, namely EcGM1. The EcGM1 family had the largest predicted growth rate, was more experimentally tractable than the triple and quad knockouts, and was an unambiguous test of the reducing power hypothesis without the complication of the additional deletions. Thus, while the EcGM2 and EcGM3 families could potentially give higher glycan flux, the EcGM1 family gave the clearest evaluation of the influence of the pentose phosphate pathway deletions.

As predicted, single pentose phosphate knockouts $\Delta$zwf, $\Delta$pgl, and $\Delta$gnd displayed greater fluorescence than wt cells, with $\Delta$gnd being the most significant. However, when these deletions were combined with $\Delta$sdh only the $\Delta$sdhC $\Delta$gnd combination led to increased glycan biosynthesis compared to wt cells. The single $\Delta$gnd mutant increased glycan production by nearly 3-fold compared to the wt strain background, while the

Fig. 3. Comparison of fluxes between the wild-type case and glycan-producing strain of type EcGM3 as calculated by flux balance analysis. (A) Fluxes through key nodes of metabolism. Top fluxes correspond to the wild-type case, bottom fluxes are for strain EcGM3. Fluxes are normalized by the glucose uptake rate. Inset shows fluxes associated with glutamate and glutamine synthesis along with the pathway to glycan precursor UDP-GlcNAc. (C) Total flux into selected cofactors, normalized to glucose uptake rate. Inset shows the primary modes of NADPH production in each strain. Abbreviations: Pentose phosphate pathway, PPP; Extracellular glucose, Glc$_{xt}$; Glucose-6-phosphate, G6P; Fructose 6-phosphate, F6P; 6-phospho D-glucono-1,5-lactone, 6PGL; Glucose 1-phosphate, G1P; Glycerate 2-phosphate, 2 PG; Phosphoenolpyruvate, PEP; Pyruvate, PYR; Oxaloacetate, OAA; Acetyl-CoA, ACCoA; 2-Oxoglutarate, $\alpha$KG; Glucosamine 6-phosphate, GAMP6P; UDP-N-acetyl-D-glucosamine, UDP-GlcNAc.
ΔsdhC Δgnd combination led to a nearly 2.5-fold increase over the wt strain. Lastly, we tested the non-lethal deletions that were predicted to remove glycan biosynthesis; the ΔgalU mutant showed no glycan production, thereby validating the model simulations. Taken together, constraint-based simulations predicted pentose phosphate pathway deletions in combination Δsdh (and potentially other genes) could improve glycan production by altering the redox state of the cell. We tested this hypothesis in the simplest possible experimental model, Δsdh Δzwf/Δpgl/Δgnd). Of the model predicted changes, only Δgnd alone and ΔsdhC Δgnd significantly increased glycan biosynthesis beyond the wt strain background. This suggested the model identified a potential axis for the improvement of glycan production, but results from the experimental system suggested this axis was likely more complicated as only the Δgnd and ΔsdhC Δgnd mutants gave a positive response.

3. Discussion

In this study we adapted a genome-scale model of E. coli metabolism for the simulation of heterologous synthesis of N-glycans. We applied heuristic optimization in combination with flux balance analysis to identify genetic knockouts that coupled C. jejuni glycan synthesis to growth. Simulations identified growth-coupled strains for minimal media growth on glucose as the sole source of carbon and energy. Flux analysis of these strains revealed two modes of flux redistribution that promoted glycan synthesis. For growth on glucose, simulations showed that maintaining high glycolytic flux and producing excess glutamine for the amination of glycan precursor sugars led to a growth-coupled phenotype. Simulations also identified the PPP as a primary target, suggesting the manipulation of the NADH/NADPH ratio influenced glycan synthesis. We validated model predictions by measuring cell surface-displayed N-glycans in E. coli mutants. In all growth conditions, the Δgnd mutant outperformed the wt strain in glycan synthesis. Overall, our model-guided strategy showed promise toward rationally designing a microbial glycosylation platform.

We used simulated annealing and flux balance analysis to search for metabolic and regulatory gene knockouts that produced a growth coupled phenotype. Several constraint-based methods have been developed previously to identify gene knockouts that coupled production to growth e.g., (Burgard et al., 2003; Patil et al., 2005; Nair et al., 2017). Most of these methods rely on an OptKnock-like approach, whereby a bi-level mixed integer optimization problem is solved to identify the optimal set of gene knockouts. This class of method guarantees identification of the global optimum; however, it suffers from a few limitations. First, search time for OptKnock-like algorithms scales exponentially with system size and number of gene knockouts, making them unable to handle very large metabolic networks. Second, only linear engineering objectives (e.g., target production flux) can be searched over. In contrast, heuristic optimization is an effective approach for searching large networks while simultaneously considering non-linear objective functions. Though identification of the global optimum is not guaranteed with these methods, desirable sub-optimal solutions can be found quickly (Patil et al., 2005; Rocha et al., 2008). Also, heuristic optimization can search efficiently for gene knockouts rather than reaction knockouts. This is an important distinction because the mapping of genes to reactions is not necessarily one to one. Thus, experimentally, many reactions may be difficult to knock out because they may be catalyzed by the products of many genes. Here, we used simulated annealing in combination with flux balance analysis to maximize the shadow price of growth with respect to glycan flux using a genome scale metabolic reconstruction. The approach identified PPP knockouts that altered the NADH/NADPH balance, and increased glycolytic flux leading to enhanced glycan production. Surprisingly, these knockouts were not in the same section of the metabolism compared with previous literature studies. However, this may be expected, as we searched for growth coupled solutions and did not simply increase glycan formation. These solutions, while more difficult to obtain, offer a significant future advantage; namely, optimization of glycan production could be improved by selecting for increased growth through serial passage.

Often times, due to the simplifying assumptions of the model, the predicted mutant strains end up being difficult or impossible to construct in the lab. Also, it is not known a priori which combinations of knockouts will result in unviable strains. This does not make higher order knockout solutions unhelpful in achieving a metabolic engineering objective. More detailed flux analysis, like that in Fig. 3, of experimentally intractable
strains, might guide exploration of alternative genetic perturbations, such as knock-down and overexpression, to achieve a similar flux profile and resulting phenotype. Our approach, which allows for an unbiased random search of many potential knockout combinations at once, is able to identify suboptimal and/or higher order knockout strains that may be of great interest to experimentalists. Incorporation of alternative genomics datasets, such as 13C-based fluxomics measurements, would highlight metabolic pathways in least agreement with model predictions, thereby identifying potential alternative target perturbations. The interpretation and application of metabolic flux models by leveraging other information sources to overcome model deficiencies continues to be an exciting area of research.

Many aspects of glycoprotein production in *E. coli* are amenable to investigation and engineering by metabolic modeling. This study focused on increasing the availability of glycan precursor metabolites through model-guided metabolic network manipulations. Other approaches in bacteria have focused on optimizing expression of glycosylation pathway enzymes and identification of metabolic reaction targets through proteomic and genome engineering (Pandhal et al., 2011, 2012, 2013). Despite these efforts, improving glycosylation efficiency in *E. coli* remains a significant challenge. To address this challenge, a more comprehensive mathematical description of the cell, one that couples metabolism with gene expression and metabolic demand, may be required to precisely model glycosylation in *E. coli*. Our approach does not explicitly consider the metabolic burden associated with heterologous expression of glycosylation pathway enzymes nor the expression of the acceptor glycoprotein. Also, flux balance analysis lacks a description of enzyme kinetics and metabolite concentrations. Predicting phenotypic changes to genetic perturbations is a primary challenge in model-guided metabolic engineering (Link et al., 2014). It has been shown that single knockouts in the central metabolism of *E. coli* do little to change the relative flux distribution in the organism (Sauer et al., 1999). *E. coli* robustly controls metabolic flux using allosteric, transcriptional and post-transcriptional regulatory, and post-translational modification systems (Kremling et al., 2008; Link et al., 2013). Thus, glycoprotein production in *E. coli* is a unique challenge in that it requires optimization of two opposing cellular processes. Recombinant protein production of a desired glycoprotein along with glycosylation pathway enzymes requires energy from catabolic processes. On the other hand, glycan precursor synthesis requires conservation of available sugars and anabolic processes. The addition of regulatory systems and an explicit formulation of these efforts, improving glycosylation efficiency, is able to identify suboptimal and/or higher order knockout strains that achieve a similar glycan production (Kirkpatrick et al., 1983). Prior to optimization, we removed all genes associated with dead end reactions, since knocking those out would have no effect on the network. Also, we removed duplicate genes, i.e., those that produced identical effects when knocked out. Finally, we removed genes whose knockout resulted in zero growth. We searched over both metabolic and regulatory genes; metabolic and transcriptional regulatory genes were represented by a binary array where 1 indicated the gene was expressed, and 0 zero indicated it was removed from the network (or transcriptionally repressed). A random initial gene knockout array was generated. We allowed for a maximum of 20 knockouts during the search. At each iteration, a new knockout array was generated through mutation operations that randomly introduced new knockouts and rearranged existing knockouts similar to (Patil et al., 2005). Briefly, new knockouts were introduced for each gene with probability \P(mutate) = 10^{-4}. If necessary, knockouts were randomly removed to limit their total number to 20. Then, a random number of knockouts were rearranged (i.e., removed from one gene and assigned to another). At each iteration, the fitness (shadow price) of an individual was computed using flux balance analysis. When an individual with a higher fitness was encountered (greater shadow price), that individual was accepted. However, when an individual with a lower fitness was encountered, we accepted this individual with a probability given by a Boltzmann factor:

\[
\P(accept) = e^{-\Delta \text{glycan} / T}
\]

where \(\Delta \text{glycan}\) denotes the change in shadow price between the current and previous solution, and the temperature \(T\) denotes the computational annealing temperature which decreased with the search iteration. The annealing temperature \(T\) decreased exponentially such that \(T_k = \alpha T_{k-1}\), where \(k\) denotes the iteration index and \(\alpha\) denotes the cooling rate defined as (Rocha et al., 2008):

\[
\alpha = \exp\left(\frac{\log T_f - \log T_0}{N_{max}/N_k}\right)
\]

The term \(N_{max}\) denotes the maximum allowable number of objective function evaluations (\(N_{max} = 10,000\), and \(N_k\) denotes the number of objective function evaluations performed at each distinct temperature value (\(N = 1\)). The initial temperature \(T_0\) was defined as \(T_0 = - \frac{\Delta \text{glycan}}{\log 0.9}\) while the final temperature \(T_f\) was given by \(T_f = - \frac{\Delta \text{glycan}}{\log 0.5}\). Lastly, \(\Delta \text{glycan}_{\text{f}}\) denotes the difference in shadow price corresponding to an acceptance probability of worse solutions of 50% at the beginning of the search, and \(\Delta \text{glycan}_{\text{f}}\) is the shadow price difference giving a 50%
probability of accepting a worse solution by the end of the search. These values were approximated using the typical shadow price values of random knockout arrays: \( \Delta H_{\text{scan}} = 0.005 \), \( \Delta H_{\text{scan}} = 0.0005 \).

Though we sought to maximize glycan flux, we also wanted to identify experimentally viable strains. Thus, during an optimization search, we set a lower bound on the biomass reaction flux equal to 10% of the wild-type simulated growth rate. Strains that could not meet this constraint were ignored. The knockout search was terminated once a positive shadow price was found. After the optimization, we processed constraint were ignored. The knockout search was terminated once a

4.2. Bacterial strains and media

For surface-labeled glycan fluorescence measurements, we used the E. coli strain BW25113 as our wild-type case (Baba et al., 2006). BW25113 was used as the parent strain to construct all gene knockout strains. Plasmid pCP20 was used to excise KmR cassette (Cherepanov and Wackernagel, 1995). Minimal media consisted of 33.9 g/L Na2HPO4, 15.0 g/L KH2PO4, 5.0 g/L NH4Cl, and 2.5 g/L NaCl. Media was supplemented with 0.4% glucose. Growth medium was supplemented by appropriate antibiotic at: 100 \( \mu \)g/mL ampicillin (Amp), 25 \( \mu \)g/mL chloramphenicol, and 50 \( \mu \)g/mL kanamycin (Kan). Growth was monitored by measuring optical density at 600 nm (OD600).

4.3. Flow cytometry

BW25113-based knockout strains were transformed with plasmid pACycopgl, constitutively expressed by the C. jejuni pgl locus. Cultures were inoculated from frozen stock in LB and grew for 3–6 h. Cells were subcultured 1:100 in minimal media overnight and then transferred to fresh minimal media to an OD600 of 0.1. 300 \( \mu \)L cells were harvested during exponential growth phase (OD600 \( \approx \) 0.6). Cells were washed with PBS then incubated in the dark for 15 min at 37°C. Cells were resuspended in 5 \( \mu \)g/mL SBA-Alexa Fluor 488 (Invitrogen) and 500 \( \mu \)L PBS and analyzed using a FACS Calibur (Becton Dickinson). Geometric mean fluorescence was determined from 100,000 events.

Author contributions

J.V and M.P.D directed the study. J.W constructed the mathematical model and conducted the computational studies. T.M, C.G and J.W created the mutant strains and conducted the experimental studies. The manuscript was prepared and edited for publication by J.V, M.P.D, T.M and J.W. All authors reviewed the manuscript.

Conflicts of interest

M.P.D has a financial interest in Glycobia, Inc. J.V, T.M, C.G and J.W have no competing financial interests.

Funding

This study was supported by an award from the National Science Foundation (MCB-1411715) to M.D and J.V and by an award from the US Army (Systems Biology of Trauma Induced Coagulopathy W911NF-10-1-0376) to J.V. for the support of J.W.
