

Improving Designer Glycan Production in *Escherichia coli* through Model-Guided Metabolic Engineering

Joseph A. Wayman^a, Cameron Glasscock, Thomas J. Mansell, Matthew P. DeLisa and Jeffrey D. Varner

Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853

^a*School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853*

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.

Keywords: Glycosylation, Metabolic modeling, Microbial glycobiology,

N-linked glycan, Strain engineering

1 **1. Introduction**

2 Protein glycosylation is the attachment of glycans (mono-, oligo-, or
3 polysaccharide) to specific amino acid residues in proteins, most commonly
4 asparagine (*N*-linked) or serine and threonine (*O*-linked) residues. Roughly
5 three-quarters of eukaryotic proteins and more than half of prokaryotic pro-
6 teins are glycosylated [1]. Glycosylation is also vitally important to the
7 development of many protein biologics, and has been harnessed for enhanc-
8 ing therapeutic properties such as half-life extension [2, 3, 4, 5], antibody-
9 mediated cytotoxicity [6, 7], and immunogenicity [8, 9, 10].

10 Though once thought to occur only in eukaryotes, protein glycosylation
11 has now been discovered in all three domains of life, including bacteria [11].
12 The best characterized bacterial *N*-glycosylation system is that of the human
13 pathogen *Campylobacter jejuni* [12]. The *C. jejuni* glycan has the form of a
14 branched heptasaccharide Glc GalNAc₅ Bac, where Glc is glucose, GalNAc is
15 *N*-acetylgalactosamine, and Bac is bacillosamine. This glycan is assembled
16 on the lipid carrier undecaprenyl pyrophosphate (Und-PP) on the cytoplas-
17 mic face of the inner membrane by an enzymatic pathway encoded by the
18 *pgl* (protein glycosylation) locus (Fig. 1). The fully assembled glycan is
19 flipped across the membrane and transferred to asparagine residues in accep-
20 tor proteins by the oligosaccharyltransferase (OST) PglB. PglB attaches the
21 heptasaccharide to periplasmically-localized proteins containing the consen-
22 sus sequence D/E-X-N-Z-S/T, where X and Z are any residue except proline
23 [13, 14].

24 The functional transfer of this system into *E. coli* [15] has spurred in-
25 terest in recombinant production of glycans and ultimately therapeutic gly-
26 coproteins in this genetically tractable bacterial host [16, 17]. Along these
27 lines, glycosylation-competent *E. coli* cells have been used to produce a va-
28 riety of periplasmic and extracellular glycoproteins including antibodies [13]
29 and conjugate vaccine candidates [18]. The promiscuity of the PglB en-
30 zyme towards structurally diverse lipid-linked glycan substrates has been
31 exploited to further expand the *E. coli* platform, enabling the creation of
32 glycoproteins bearing different bacterial O-polysaccharide antigens [18, 19]
33 and even the eukaryotic trimannosyl core *N*-glycan produced by a synthetic
34 pathway comprised of four yeast glycosyltransferases [20]. However, while

PglB can efficiently glycosylate native *C. jejuni* acceptor proteins with cognate Glc GalNAc₅ Bac glycan in engineered *E. coli*, glycosylation of non-Campylobacter target proteins is often much less efficient [21], especially in combination with heterologous glycan structures [20].

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 [16, 22]. 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 [23]. Further, genome-wide screening of gene overexpression identified targets that increased glycoprotein production as well as glycosylation efficiency [24]; genes in pathways associated with glycan precursor synthesis (UDP-GlcNAc) as well as lipid carrier production (isoprenoid synthesis) were identified as bottlenecks. Improved glycosylation efficiency has also been achieved by supplementing growth media with GlcNAc [25] or increasing the expression of PglB via codon optimization [26]. 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* [20]. Consistent with simulations, the best model-predicted changes increased glycan production by nearly 3-fold compared with the glycan pro-

73 duction level in wild-type (wt) *E. coli* cells. Taken together, our results
74 reveal the significant impact that metabolic modeling can have on designing
75 chassis strains with enhanced *N*-linked protein glycosylation capabilities.

76 1.1. Results

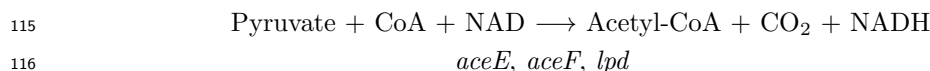
77 *Construction of a constraint-based model of N-linked glycosylation in E. coli*

78 A constraint-based model of N-glycosylation in *E. coli* was used to iden-
 79 tify genetic knockouts that coupled glycan biosynthesis with optimal growth.
 80 We augmented the existing genome-scale *E. coli* model iAF1260 from Palsson
 81 and coworkers [27] to include the reactions of the *C. jejuni* glycosylation path-
 82 way (Table 1). The adapted network consisted of 2395 reactions, 1271 open
 83 reading frames, and 1986 metabolites segregated into cytoplasmic, periplas-
 84 mic, and extracellular compartments. Added reactions included the biochem-
 85 ical transformations catalyzed by the glycosyltransferases (e.g., PglA, PglC)
 86 associated with glycan biosynthesis, PglK flippase-mediated translocation of
 87 the glycan into the periplasm, and PglB-mediated glycan conjugation to an
 88 acceptor protein (Fig. 1). In addition, we incorporated the transcriptional
 89 regulatory network of Covert *et al.*, consisting of 101 transcription factors,
 90 regulating the state of the metabolic genes [28]. This regulatory network
 91 imparts Boolean constraints on metabolic fluxes based upon the nutrient en-
 92 vironment. The model code is available for download under an MIT software
 93 license from the Varnerlab website [29].

94 *Identification of growth-coupled gene knockout strains*

95 To identify genetic knockouts that coupled optimal growth to glycan
 96 biosynthesis, we used heuristic optimization and the constraint-based model
 97 (see Materials and Methods). Coupling growth to glycan synthesis was de-
 98 sirable for several reasons. Foremost amongst these, growth-coupled strains
 99 create stoichiometric imbalances that reroute metabolic flux toward the de-
 100 sired product as a consequence of growth [30, 31]. Therefore, faster growth
 101 requires increased glycan formation. Thus, optimizing glycan production
 102 through adaptive evolution is made trivial by selecting for growth through
 103 serial passage [31, 32]. Several methods have been developed to estimate
 104 genetic knockouts using constraint-based models. In this study, we used
 105 simulated annealing to search over the states of metabolic enzyme and tran-
 106 scription factor (TF) genes to identify the desired phenotype (Fig. 2). The
 107 state of each gene was represented as a binary array, where a one indicated
 108 normal activity, while a zero indicated a genetic knockout or regulatory re-
 109 pression. Boolean rules informed by nutrient conditions controlled the TF
 110 genes, which in turn controlled the state of the metabolic genes. Once de-
 111 fined, 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:



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 [27]. The simulated annealing algorithm performed a random search of genetic knockouts, iteratively 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:

$$u_{\text{glycan}} = \frac{\Delta v_{\text{growth}}}{\Delta v_{\text{glycan}}} \quad (1)$$

where Δv_{growth} denotes the change in growth rate for a forced change in glycan flux Δv_{glycan} , and v_{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 $u_{\text{glycan}} > 0$, 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 $\Delta sdh \Delta gnd \Delta pta \Delta eutD$), the strain with the highest simulated glycan yield, the optimal growth rate occurs at a non-zero glycan flux (Fig.

146 S1B). All growth-coupled strains contained a knockout of succinate dehy-
147 drogenase (*sdh*) and truncated pentose phosphate pathway (PPP) flux at ei-
148 ther glucose 6-phosphate-1-dehydrogenase (*zwf*), 6-phosphogluconolactonase
149 (*pgl*), or 6-phosphogluconate dehydrogenase (*gnd*).

150 *Flux analysis of N-glycan production in growth-coupled strains*

151 Growth-coupled glycan producing strains had increased glycolytic flux,
152 and decreased amino acid biosynthesis compared to glycan production in the
153 wt strain background (Fig. 3). We compared the normalized flux values for
154 EcGM2 with the wt strain. Normalizing all fluxes to glucose uptake rate,
155 EcGM2 displayed greater flux through glycolysis by cutting off the PPP via
156 knockout of NADPH-producing *gnd* (Fig. 3A). EcGM2 also had decreased
157 synthesis of every amino acid except for glutamine, indicating a source of
158 stoichiometric imbalance that may be relieved by synthesis of the glycan
159 precursor UDP-GlcNAc. Further, the PEP-pyruvate node acted as a switch
160 point in central carbon metabolism (Fig. 3B). Here, PEP and pyruvate, the
161 products of glycolysis, enter the TCA cycle through decarboxylation of pyru-
162 vate to acetyl-CoA (ACCoA) and carboxylation of PEP to form oxaloacetate
163 (OAA) [33]. The latter replenishes TCA cycle intermediates that exited TCA
164 for anabolic processes. EcGM2, with a diminished anabolic capacity for cell
165 growth, displayed lower flux through PEP carboxylase (*ppc*). However, as
166 the result of high glycolytic flux, EcGM2 had increased flux through pyru-
167 vate dehydrogenase (*aceEF*), sending carbon into the oxidative branch of
168 the TCA cycle. It is known that high glucose uptake rates result in excess
169 acetyl-CoA, surpassing the capacity of the TCA cycle. Because of this excess
170 flux, wt *E. coli* grown on glucose commonly displays acetate fermentation,
171 even under aerobic conditions [34]. We observed increased acetate secretion
172 in EcGM1 simulations, but through a route differing from wild-type cells.
173 The knockouts Δpta and $\Delta eutD$ prevented ATP-generating acetate secre-
174 tion. Flux was instead routed through the redox-neutral reactions initiated
175 by acetaldehyde dehydrogenase (*mhpF*). Excess acetyl-CoA was also utilized
176 in the pathway generating UDP-GlcNAc. Lastly, EcGM2 displayed a shift
177 in cofactor production (Fig. 3C). Higher flux through glycolysis naturally
178 led to NADH overproduction. On the other hand, the primary source of
179 NADPH shifted from PPP genes *zwf* and *gnd* to the membrane transhy-
180 drogenase *pnt*, capable of direct transfer of electrons from NADH to NADP.
181 Sauer *et al.* identified *pnt* as a major source of NADPH in *E. coli* (35-45%
182 of total) [35]. Thus, *pnt* is capable of carrying significant flux *in vivo*. Taken

183 together, these results suggested the model identified strains that promoted
184 glycan precursor synthesis, primarily UDP-GlcNAc, by creating a combina-
185 tion of metabolite and redox imbalance.

186 *Experimental validation of N-glycan-producing knockout strains*

187 Glycan production was measured in the mutant strains to validate the
188 model predictions (Fig. 4). Gene knockout strains were constructed using
189 the Keio collection of single gene knockouts *E. coli* BW25113 [36] as donor
190 strains for P1*vir* phage transduction. Mutants were constructed containing
191 single, double, and triple knockouts that appeared in growth-coupled strains
192 identified by the constraint-based model. We also performed simulations of
193 each single gene knockout to determine genes that prevented glycan synthe-
194 sis; *galU*, a key enzyme in the synthesis of glycan precursor UDP-glucose,
195 was the only non-lethal knockout that prevented glycan synthesis. Knock-
196 out strains were transformed with a plasmid constitutively expressing the
197 *C. jejuni* *pgl* locus. To quantify glycan production, we took advantage of
198 crosstalk between the glycosylation pathway and native lipopolysaccharide
199 (LPS) synthesis in *E. coli* [37]. Specifically, after the glycan is flipped into
200 the periplasm, it can be transferred to lipid A-core by the WaaL O-antigen
201 ligase and shuttled to the outer membrane by LPS pathway enzymes, where
202 it is displayed on the cell surface [16]. Labeling of these surface-displayed
203 N-glycans with fluorescently-tagged lectins can then be used to quantify the
204 amount of glycan displayed on the cell surface as a measure of glycan produc-
205 tion [20]. Here, we labeled *C. jejuni* glycans for detection by flow cytometry
206 with fluorophore-conjugated soybean agglutinin (SBA), a lectin specific to
207 terminal galactose and GalNAc residues. Prior to labeling, knockout strains
208 were grown in glucose minimal media and harvested during the exponential
209 growth phase, to most closely satisfy the pseudo-steady-state assumption of
210 model predictions.

211 A common feature of the predicted mutant strains was the deletion of
212 pentose phosphate pathway genes *zwf/pgl/gnd* in combination with Δ *sdh*.
213 Analysis of the metabolic flux distribution in these mutants suggested the
214 reducing state of the cell as well as the carbon flux was reprogrammed to
215 support enhanced glycan biosynthesis. While hypothetical knockouts such
216 as Δ *sdh* Δ (*zwf/pgl/gnd*) Δ *pta* Δ *eutD* were predicted to have higher glycan
217 yield, in this study we experimentally evaluated only the simplest growth-
218 coupled double knockout family, namely EcGM1. The EcGM1 family had
219 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 Δzwf , Δpgl , and Δgnd displayed greater fluorescence than wt cells, with Δgnd being the most significant. However, when these deletions were combined with Δsdh only the $\Delta sdhC \Delta gnd$ combination led to increased glycan biosynthesis compared to wt cells. The single Δgnd mutant increased glycan production by nearly 3-fold compared to the wt strain background, while the $\Delta sdhC \Delta gnd$ 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 $\Delta galU$ 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, $\Delta sdh \Delta (zwf/pgl/gnd)$. Of the model predicted changes, only Δgnd alone and $\Delta sdhC \Delta gnd$ 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 $\Delta sdhC \Delta gnd$ mutants gave a positive response.

244 Discussion

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

261 We used simulated annealing and flux balance analysis to search for
 262 metabolic and regulatory gene knockouts that produced a growth coupled
 263 phenotype. Several constraint-based methods have been developed previ-
 264 ously to identify gene knockouts that coupled production to growth e.g.,
 265 [30, 38, 39]. Most of these methods rely on an OptKnock-like approach,
 266 whereby a bi-level mixed integer optimization problem is solved to identify
 267 the optimal set of gene knockouts. This class of method guarantees identifica-
 268 tion of the global optimum; however, it suffers from a few limitations. First,
 269 search time for OptKnock-like algorithms scales exponentially with system
 270 size and number of gene knockouts, making them unable to handle very large
 271 metabolic networks. Second, only linear engineering objectives (e.g., target
 272 production flux) can be searched over. In contrast, heuristic optimization is
 273 an effective approach for searching large networks while simultaneously con-
 274 sidering non-linear objective functions. Though identification of the global
 275 optimum is not guaranteed with these methods, desirable sub-optimal solu-
 276 tions can be found quickly [38, 40]. Also, heuristic optimization can search
 277 efficiently for gene knockouts rather than reaction knockouts. This is an
 278 important distinction because the mapping of genes to reactions is not nec-
 279 essarily one to one. Thus, experimentally, many reactions may be difficult
 280 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.

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 [23, 26, 24]. 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 [41]. 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 [42]. *E. coli* robustly controls metabolic flux using allosteric, transcriptional and post-transcriptional regulatory, and post-translational modification systems [43, 44]. 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 description of gene expression to a stoichiometric model may be an effective

319 strategy for optimizing these opposing processes. Other strategies that may
320 be helpful for optimization of this system include the enhancement of glycan
321 precursor pathways, such as hexosamine synthesis, as well as the removal of
322 competing pathways.

323 Materials and Methods

324 *Flux balance analysis and heuristic optimization.*

Reactions encoding *C. jejuni* glycan formation (Table 1) were added to the genome-scale metabolic model of *E. coli* iAF1260 [27]. The combined model was then used to determine growth coupled gene knockouts that improved glycan production flux. Metabolic fluxes were estimated using flux balance analysis. Flux balance analysis requires two assumptions. First, the cell was assumed to operate at a pseudo-steady-state, where the rate of production of every intracellular metabolite was equal to its consumption. Second, the cell has evolved to operate optimally to achieve a cellular objective. Though many objectives have been proposed, we use the most common, namely growth rate (i.e., biomass formation) maximization [45]. The determination of a flux distribution satisfying these assumptions was formulated as a linear optimization problem:

$$\begin{aligned} \max_{\mathbf{v}} (v_{growth} = \mathbf{c}^T \mathbf{v}) \\ \text{Subject to : } \mathbf{S} \mathbf{v} = 0 \\ \alpha_i \leq v_i \leq \beta_i \end{aligned}$$

325 where $\mathbf{v}$ is the steady-state flux vector and α_i and β_i are the lower and up-
326 per limits for the individual flux values, respectively. The quantity, v_{growth} ,
327 denotes the specific growth rate where $\mathbf{c}$ is a vector containing the stoichio-
328 metric contribution of each metabolic species to biomass. The stoichiometric
329 matrix $\mathbf{S}$ encodes all biochemical reaction connectivity considered in the
330 model. Each row of $\mathbf{S}$ describes a metabolite, while each column describes
331 a particular reaction. The (i, j) element of $\mathbf{S}$, denoted by σ_{ij} , describes how
332 species i participates in reaction j . If $\sigma_{ij} > 0$, species i is produced by re-
333 action j . Conversely, if $\sigma_{ij} < 0$, then species i is consumed by reaction j .
334 Lastly, if $\sigma_{ij} = 0$, then species i is not involved in reaction j . The maximum
335 substrate and oxygen uptake rates were set at 10 mmol/gDW/hr. Boundary
336 conditions were set to allow for the unrestricted formation of acetate. All
337 genes found to be essential for growth on Luria-Bertani (LB) medium were
338 excluded from the search [36].

339 We used the FastPros algorithm developed by Ohno *et al.*, in combination
340 with a shadow price objective, to estimate genetic knockouts [46]. Simulated
341 annealing identified growth-coupled genetic knockouts with improved glycan
342 production [47]. 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. New knockout arrays were generated through crossover and mutation operators that randomly introduced new knockouts [38]. 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:

$$\mathcal{P}(\text{accept}) = e^{-\Delta u_{glycan}/T} \quad (2)$$

where Δu_{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 $T_{k+1} = \alpha T_k$, where k denotes the iteration index and α denotes the cooling rate defined as [40]:

$$\alpha = \exp\left(\frac{\log T_f - \log T_o}{N_{max}/N_\alpha}\right) \quad (3)$$

The term N_{max} denotes the maximum allowable number of objective function evaluations ($N_{max} = 10,000$), and N_α denotes the number of objective function evaluations performed at each distinct temperature value ($N_\alpha = 1$). The initial temperature T_o was defined as $T_o = -\frac{\Delta u_{glycan,o}}{\log 0.5}$, while the final temperature T_f was given by $T_f = -\frac{\Delta u_{glycan,f}}{\log 0.5}$. Lastly, $\Delta u_{glycan,o}$ denotes the difference in shadow price corresponding to an acceptance probability of worse solutions of 50% at the beginning of the search, and $\Delta u_{glycan,f}$ 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: $\Delta u_{glycan,o} = 0.005$, $\Delta u_{glycan,f} = 0.0005$.

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 growth-coupled knockout strains by iteratively knocking in each knockout gene to estimate knockouts that did not affect the phenotype. In this way we identified the smallest number of gene knockouts that produced enhanced glycan flux at optimal growth. Each optimization run required approximately six hours on a single CPU Apple workstation (Apple, Cupertino, CA, USA; OS X v10.10). All model and optimization code is available in the MATLAB (The Mathworks, Natick MA) programming language, and free to download under an MIT software license from Varnerlab.org [29].

Bacterial strains and media

For surface-labeled glycan fluorescence measurements, we used the *E. coli* strain BW25113 as our wild-type case [36]. BW25113 was used as the parent strain to construct all gene knockout strains. Plasmid pCP20 was used to excise KmR cassette [48]. 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 µg/mL ampicillin (Amp), 25 µg/mL chloramphenicol, and 50 µg/mL kanamycin (Kan). Growth was monitored by measuring optical density at 600 nm (OD₆₀₀).

Flow cytometry

BW25113-based knockout strains were transformed with plasmid pACYCp_{gl}, constitutively expressed by the *C. jejuni* *p_{gl}* locus. Cultures were inoculated from frozen stock in LB and grew for 3-6 hours. Cells were subcultured 1:100 in minimal media overnight and then transferred to fresh minimal media to an OD₆₀₀ of 0.1. 300 µL cells were harvested during exponential growth phase (OD₆₀₀ ≈ 0.6). Cells were washed with PBS then incubated in the dark for 15 minutes at 37°C. Cells were resuspended in 5 µg/mL SBA-Alexa Fluor 488 (Invitrogen) and 500 µL PBS and analyzed using a FACSCalibur (Becton Dickinson). Geometric mean fluorescence was determined from 100,000 events.

409 *Author contributions.*

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

415 *Funding.*

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

420 *Conflicts of interest.*

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

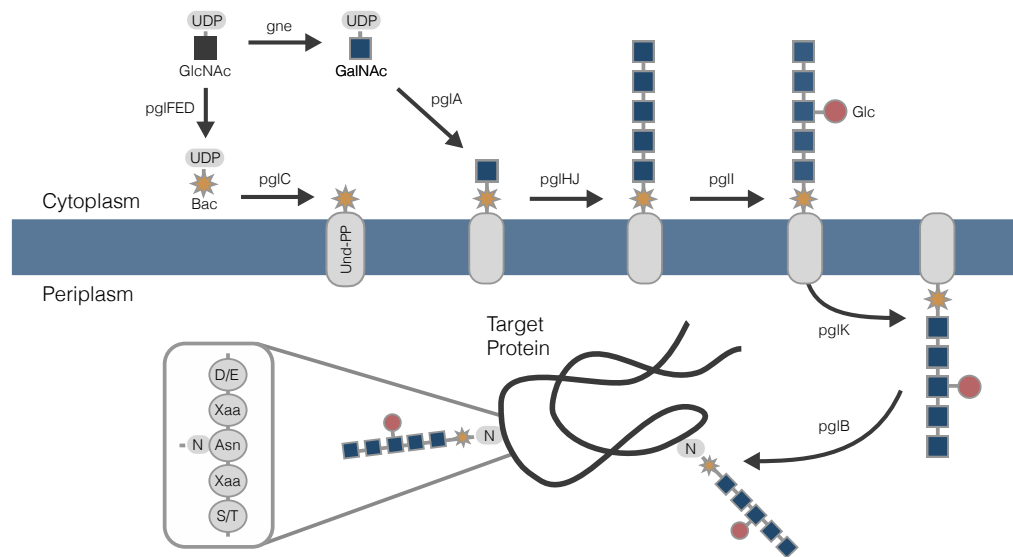


Figure 1: Glycosylation pathway in *C. jejuni* and *E. coli*. Glycan assembly, facilitated by *pgl* locus enzymes, takes place on a lipid carrier, undecaprenyl pyrophosphate (Und-PP), from cytoplasmic pools of nucleotide-activated sugars N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), and glucose (Glc). The glycan is then flipped onto the periplasmic side of the inner membrane, where it is transferred to an asparagine residue on a glycoprotein acceptor motif.

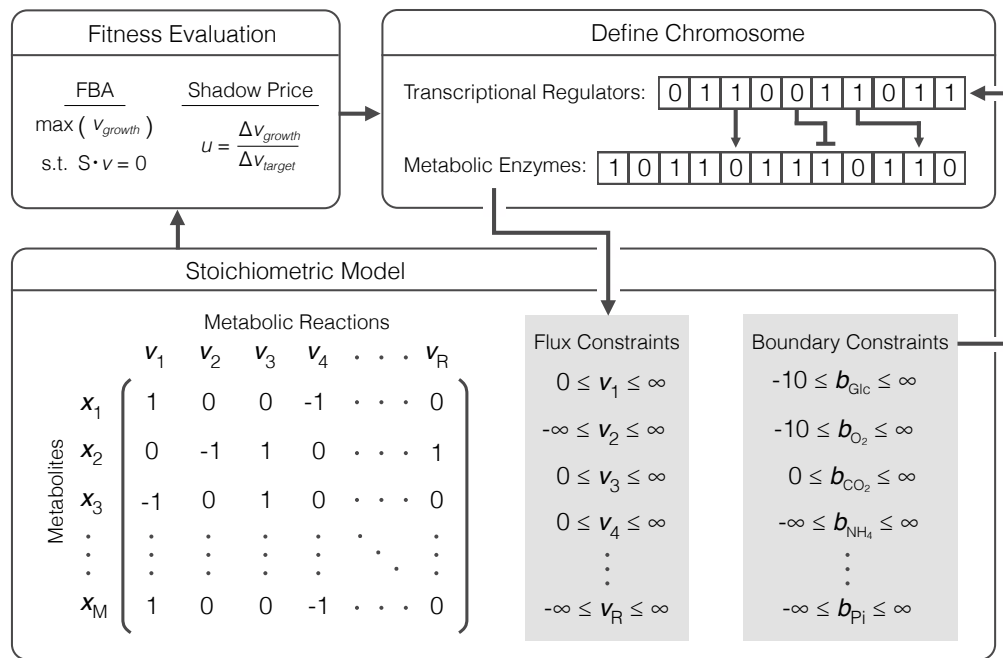


Figure 2: Heuristic optimization approach used to identify strains coupling growth to glycan production. The chromosome is defined as two separate binary arrays, one defining the state of metabolic enzyme expression and another defining the state of transcriptional regulator activation. Gene repression and knockouts are designated by zeros. Nutrient conditions define the boundary constraints within the stoichiometric model which in turn affect the state of the metabolic enzyme chromosome. Gene repression and knockouts determine the constraints placed on fluxes in the stoichiometric model. Nutrients are mapped to the state of transcriptional regulators and genes are mapped to the state of flux constraints using Boolean rules as defined in [27, 28]. Flux balance analysis is used to maximize growth rate under the constraints imposed by the mutant strain and transcriptional regulation and the fitness objective is calculated. Here, we use shadow price; the strain is accepted or rejected based on the change in fitness and a Boltzmann criterion. New mutant strains are randomly generated from accepted ones. The search continues until a positive shadow price is achieved.

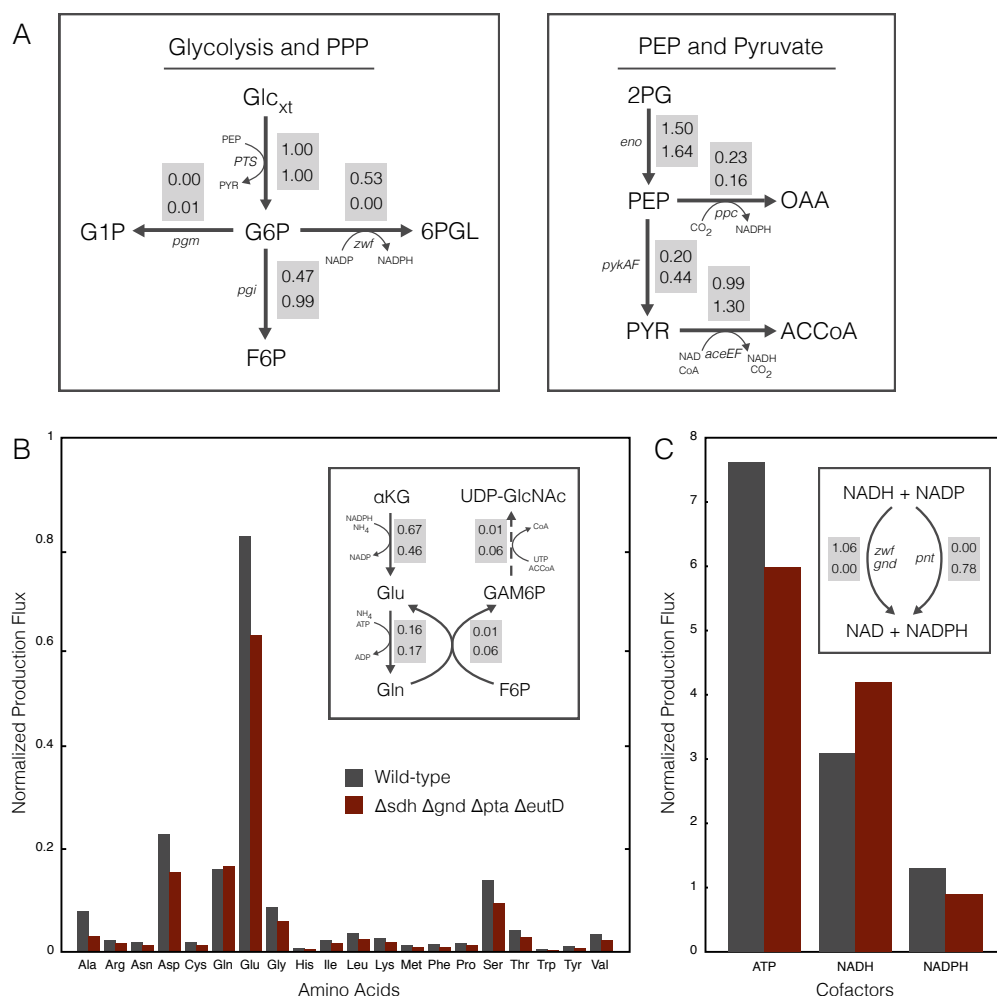


Figure 3: Comparison of fluxes between the wild-type case and glycan-producing strain of type EcGM3 as calculated by flux balance analysis. **(A)** Fluxes through key nodes of metabolism. Top fluxes correspond to the wild-type case, bottom fluxes are for strain EcGM3. Fluxes are normalized by the glucose uptake rate. **(B)** Total flux into each amino acid, normalized to glucose uptake rate. Inset shows fluxes associated with glutamate and glutamine synthesis along with the pathway to glycan precursor UDP-GlcNAc. The dotted arrow represents a lumped pathway of multiple enzymes leading to the glycan precursor. **(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, 2PG; Phosphoenolpyruvate, PEP; Pyruvate, PYR; Oxaloacetate, OAA; Acetyl-CoA, ACCoA; 2-Oxoglutarate, αKG ; Glucosamine 6-phosphate, GAM6P; UDP-N-acetyl-D-glucosamine, UDP-GlcNAc.

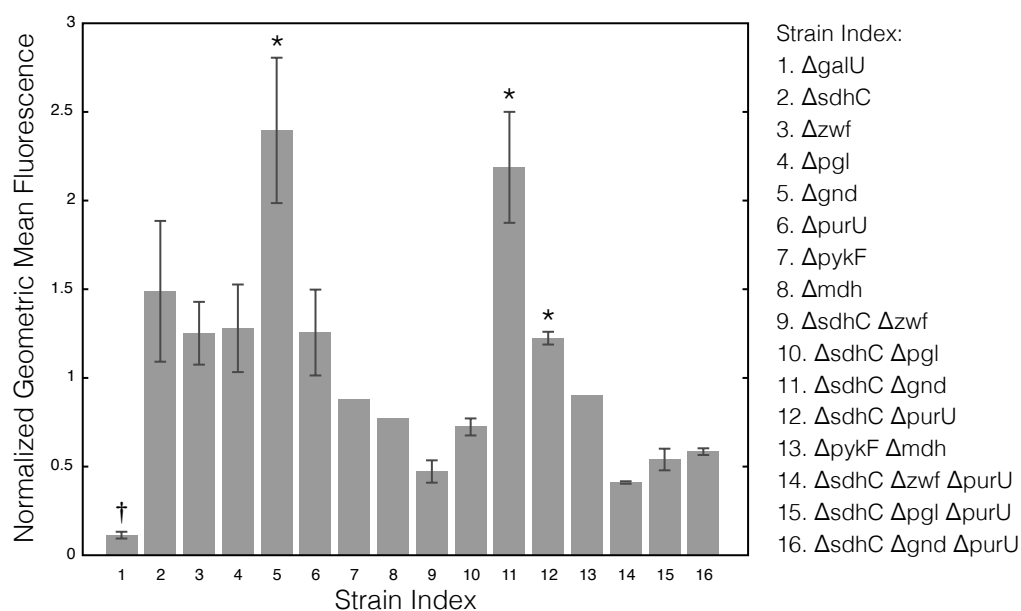


Figure 4: Geometric mean fluorescence, normalized to the wild-type value, from gene knockout strains appearing in growth-coupled strains identified by the constraint-based model. † indicates a strain predicted to eliminate glycan flux. Stars indicate statistically significant increases in fluorescences according to a t-test ($p < 0.05$). Error bars indicate the standard deviation of at least three replicates.

Table 1: Reactions added to the *E. coli* model iAF1260 [27] for biosynthesis of *C. jejuni* glycan. Species localized to the periplasm are denoted by (p), all others are cytoplasmic. Abbreviations: UDP-N-Acetyl-D-Glucosamine, UDP-GlcNAc; UDP-N-Acetyl-D-Galactosamine, UDP-GalNAc; UDP-2-acetamido-2,6-dideoxy- α -D-xylo-4-hexulose, KetoBac; L-Glutamate, Glu; UDP-N-Acetyl bacillosamine, AminoBac; α -ketoglutarate, α KG; Acetyl-CoA, ACCoA; UDP-N,N'-diacetyl bacillosamine, uBac; Coenzyme A, CoA; Undecaprenyl phosphate, Udcpp; *C. jejuni* glycan intermediates, UdcCjGlycan1, UdcCjGlycan6; Uridine monophosphate, UMP; Uridine diphosphate, UDP; UDP-Glucose, UDP-Glc; Lipid-linked *C. jejuni* glycan, UdcCjGlycan; Acceptor protein, AcceptorProt; GlycoProt, Glycoprotein; Undecaprenyl diphosphate, Udcdp.

Gene	Enzyme	Reaction	Reference
gne	UDP-GlcNAc epimerase	UDP-GlcNAc $\rightarrow$ UDP-GalNAc	[49]
pglF	UDP-GlcNAc dehydratase	UDP-GlcNAc $\rightarrow$ KetoBac + H ₂ O	[50]
pglE	Aminotransferase	KetoBac + Glu $\leftrightarrow$ AminoBac + α KG	[50]
pglD	Acetyltransferase	AminoBac + ACCoA $\rightarrow$ uBac + CoA + H ⁺	[51]
pglC	Bacillosamine transferase	Udcpp + uBac $\rightarrow$ UdcCjGlycan1 + UMP	[52]
pglAHJ	GalNAc transferases	UdcCjGlycan1 + 5*UDP-GalNAc $\rightarrow$ UdcCjGlycan6 + 5*UDP + 5*H ⁺	[53]
pglI	Glucosyl transferase	UdcCjGlycan6 + UDP-Glc $\rightarrow$ UdcCjGlycan + UDP + H ⁺	[54]
pglK	ATP-driven flippase	UdcCjGlycan + ATP + H ₂ O $\rightarrow$ UdcCjGlycan(p) + ADP + H ⁺ + Pi	[54]
pglB	Oligosyltransferase	UdcCjGlycan(p) + AcceptorProt(p) $\rightarrow$ GlycoProt + Udcdp(p)	[55]

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* glycosylating mutant, EcGM

Strain	Substrate	Genotype	Growth rate (/hr)	Glycan flux (mmol/gDW/hr)	Yield (mmol/gDW)
EcGM1	Glucose	Δ sdh Δ (zwf/pgl/gnd)	0.65	0.012	0.018
EcGM2	Glucose	Δ sdh Δ (zwf/pgl/gnd) Δ pta Δ eutD	0.53	0.098	0.185
EcGM3	Glucose	Δ sdh Δ (zwf/pgl/gnd) Δ pykAF Δ mdh	0.64	0.016	0.025

- 423 [1] A. Dell, A. Galadari, F. Sastre, P. Hitchen, Similarities and differences
424 in the glycosylation mechanisms in prokaryotes and eukaryotes, *Int J*
425 *Microbiol* 2010 (2010) 148178.
- 426 [2] S. Elliott, T. Lorenzini, S. Asher, K. Aoki, D. Brankow, L. Buck,
427 L. Busse, D. Chang, J. Fuller, J. Grant, N. Hernday, M. Hokum, S. Hu,
428 A. Knudten, N. Levin, R. Komorowski, F. Martin, R. Navarro, T. Oss-
429 lund, G. Rogers, N. Rogers, G. Trail, J. Egrie, Enhancement of therapeu-
430 tic protein in vivo activities through glycoengineering, *Nat Biotechnol*
431 21 (2003) 414–21.
- 432 [3] T. V. Flintegaard, P. Thygesen, H. Rahbek-Nielsen, S. B. Levery,
433 C. Kristensen, H. Clausen, G. Bolt, N-glycosylation increases the cir-
434 culatory half-life of human growth hormone, *Endocrinology* 151 (2010)
435 5326–36.
- 436 [4] D. G. Ilyushin, I. V. Smirnov, A. A. Belogurov, Jr, I. A. Dyachenko,
437 T. I. Zharmukhamedova, T. I. Novozhilova, E. A. Bychikhin, M. V.
438 Serebryakova, O. N. Kharybin, A. N. Murashev, K. A. Anikienko, E. N.
439 Nikolaev, N. A. Ponomarenko, D. D. Genkin, G. M. Blackburn, P. Mas-
440 son, A. G. Gabibov, Chemical polysialylation of human recombinant bu-
441 tyrylcholinesterase delivers a long-acting bioscavenger for nerve agents
442 in vivo, *Proc Natl Acad Sci U S A* 110 (2013) 1243–8.
- 443 [5] T. Lindhout, U. Iqbal, L. M. Willis, A. N. Reid, J. Li, X. Liu, M. Moreno,
444 W. W. Wakarchuk, Site-specific enzymatic polysialylation of therapeutic
445 proteins using bacterial enzymes, *Proc Natl Acad Sci U S A* 108 (2011)
446 7397–402.
- 447 [6] T. Li, D. J. DiLillo, S. Bournazos, J. P. Giddens, J. V. Ravetch, L.-X.
448 Wang, Modulating igg effector function by fc glycan engineering, *Proc*
449 *Natl Acad Sci U S A* 114 (2017) 3485–3490.
- 450 [7] C.-W. Lin, M.-H. Tsai, S.-T. Li, T.-I. Tsai, K.-C. Chu, Y.-C. Liu, M.-
451 Y. Lai, C.-Y. Wu, Y.-C. Tseng, S. S. Shivatare, C.-H. Wang, P. Chao,
452 S.-Y. Wang, H.-W. Shih, Y.-F. Zeng, T.-H. You, J.-Y. Liao, Y.-C. Tu,
453 Y.-S. Lin, H.-Y. Chuang, C.-L. Chen, C.-S. Tsai, C.-C. Huang, N.-H.
454 Lin, C. Ma, C.-Y. Wu, C.-H. Wong, A common glycan structure on
455 immunoglobulin g for enhancement of effector functions, *Proc Natl Acad*
456 *Sci U S A* 112 (2015) 10611–6.

- 457 [8] T. Lipinski, A. Fitieh, J. St Pierre, H. L. Ostergaard, D. R. Bundle,
458 N. Touret, Enhanced immunogenicity of a tricomponent mannan tetanus
459 toxoid conjugate vaccine targeted to dendritic cells via dectin-1 by in-
460 corporating -glucan, *J Immunol* 190 (2013) 4116–28.
- 461 [9] M.-O. Sadoulet, C. Franceschi, M. Aubert, F. Silvy, J.-P. Bernard,
462 D. Lombardo, E. Mas, Glycoengineering of alphagal xenoantigen on re-
463 combinant peptide bearing the j28 pancreatic oncofetal glycotope, *Gly-*
464 *cobiology* 17 (2007) 620–30.
- 465 [10] M. Wacker, L. Wang, M. Kowarik, M. Dowd, G. Lipowsky, A. Farid-
466 moayer, K. Shields, S. Park, C. Alaimo, K. A. Kelley, M. Braun, J. Que-
467 batte, V. Gambillara, P. Carranza, M. Steffen, J. C. Lee, Prevention of
468 staphylococcus aureus infections by glycoprotein vaccines synthesized in
469 escherichia coli, *J Infect Dis* 209 (2014) 1551–61.
- 470 [11] H. Nothhaft, C. M. Szymanski, Protein glycosylation in bacteria: sweeter
471 than ever, *Nat Rev Microbiol* 8 (2010) 765–78.
- 472 [12] C. M. Szymanski, R. Yao, C. P. Ewing, T. J. Trust, P. Guerry, Evidence
473 for a system of general protein glycosylation in campylobacter jejuni,
474 *Mol Microbiol* 32 (1999) 1022–30.
- 475 [13] A. C. Fisher, C. H. Haitjema, C. Guarino, E. Çelik, C. E. Endicott, C. A.
476 Reading, J. H. Merritt, A. C. Ptak, S. Zhang, M. P. DeLisa, Production
477 of secretory and extracellular n-linked glycoproteins in escherichia coli,
478 *Appl Environ Microbiol* 77 (2011) 871–81.
- 479 [14] M. Kowarik, N. M. Young, S. Numao, B. L. Schulz, I. Hug, N. Calle-
480 waert, D. C. Mills, D. C. Watson, M. Hernandez, J. F. Kelly, M. Wacker,
481 M. Aebi, Definition of the bacterial n-glycosylation site consensus se-
482 quence, *EMBO J* 25 (2006) 1957–66.
- 483 [15] M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam,
484 S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, M. Aebi, N-
485 linked glycosylation in campylobacter jejuni and its functional transfer
486 into e. coli, *Science* 298 (2002) 1790–3.
- 487 [16] J. H. Merritt, A. A. Ollis, A. C. Fisher, M. P. DeLisa, Glycans-by-
488 design: engineering bacteria for the biosynthesis of complex glycans and
489 glycoconjugates, *Biotechnol Bioeng* 110 (2013) 1550–64.

- 490 [17] J. L. Baker, E. Çelik, M. P. DeLisa, Expanding the glycoengineering
491 toolbox: the rise of bacterial n-linked protein glycosylation, Trends
492 Biotechnol 31 (2013) 313–23.
- 493 [18] M. F. Feldman, M. Wacker, M. Hernandez, P. G. Hitchen, C. L. Marolda,
494 M. Kowarik, H. R. Morris, A. Dell, M. A. Valvano, M. Aebi, Engineering
495 n-linked protein glycosylation with diverse o antigen lipopolysaccharide
496 structures in escherichia coli, Proc Natl Acad Sci U S A 102 (2005)
497 3016–21.
- 498 [19] J. Ihssen, M. Kowarik, S. Dilettoso, C. Tanner, M. Wacker, L. Thöny-
499 Meyer, Production of glycoprotein vaccines in escherichia coli, Microb
500 Cell Fact 9 (2010) 61.
- 501 [20] J. D. Valderrama-Rincon, A. C. Fisher, J. H. Merritt, Y.-Y. Fan, C. A.
502 Reading, K. Chhiba, C. Heiss, P. Azadi, M. Aebi, M. P. DeLisa, An
503 engineered eukaryotic protein glycosylation pathway in escherichia coli,
504 Nat Chem Biol 8 (2012) 434–6.
- 505 [21] F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo,
506 S. Numao, D. Neri, M. Aebi, L.-X. Wang, A combined method for
507 producing homogeneous glycoproteins with eukaryotic n-glycosylation,
508 Nat Chem Biol 6 (2010) 264–6.
- 509 [22] S. R. Jaffé, B. Strutton, Z. Levarski, J. Pandhal, P. C. Wright, Es-
510 cherichia coli as a glycoprotein production host: recent developments
511 and challenges, Curr Opin Biotechnol 30C (2014) 205–210.
- 512 [23] J. Pandhal, S. Y. Ow, J. Noirel, P. C. Wright, Improving n-glycosylation
513 efficiency in escherichia coli using shotgun proteomics, metabolic net-
514 work analysis, and selective reaction monitoring, Biotechnol Bioeng 108
515 (2011) 902–12.
- 516 [24] J. Pandhal, L. B. A. Woodruff, S. Jaffe, P. Desai, S. Y. Ow, J. Noirel,
517 R. T. Gill, P. C. Wright, Inverse metabolic engineering to improve
518 escherichia coli as an n-glycosylation host, Biotechnol Bioeng 110 (2013)
519 2482–93.
- 520 [25] M. M. Kämpf, M. Braun, D. Sirena, J. Ihssen, L. Thöny-Meyer, Q. Ren,
521 In vivo production of a novel glycoconjugate vaccine against shigella

- 522 flexneri 2a in recombinant escherichia coli: identification of stimulating
523 factors for in vivo glycosylation, *Microb Cell Fact* 14 (2015) 12.
- 524 [26] J. Pandhal, P. Desai, C. Walpole, L. Doroudi, D. Malyshev, P. C.
525 Wright, Systematic metabolic engineering for improvement of glyco-
526 sylation efficiency in escherichia coli, *Biochem Biophys Res Commun*
527 419 (2012) 472–6.
- 528 [27] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce,
529 P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, B. Ø. Palsson, A
530 genome-scale metabolic reconstruction for escherichia coli k-12 mg1655
531 that accounts for 1260 orfs and thermodynamic information, *Mol Syst*
532 *Biol* 3 (2007) 121.
- 533 [28] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, B. O. Palsson,
534 Integrating high-throughput and computational data elucidates bacte-
535 rial networks, *Nature* 429 (2004) 92–6.
- 536 [29] <http://www.varnerlab.org/downloads/>, ????
- 537 [30] A. P. Burgard, P. Pharkya, C. D. Maranas, Optknock: a bilevel pro-
538 gramming framework for identifying gene knockout strategies for micro-
539 bial strain optimization, *Biotechnol Bioeng* 84 (2003) 647–57.
- 540 [31] A. M. Feist, D. C. Zielinski, J. D. Orth, J. Schellenberger, M. J. Her-
541 rgard, B. Ø. Palsson, Model-driven evaluation of the production po-
542 tential for growth-coupled products of escherichia coli, *Metab Eng* 12
543 (2010) 173–86.
- 544 [32] R. U. Ibarra, J. S. Edwards, B. O. Palsson, Escherichia coli k-12 under-
545 goes adaptive evolution to achieve in silico predicted optimal growth,
546 *Nature* 420 (2002) 186–9.
- 547 [33] U. Sauer, B. J. Eikmanns, The pep-pyruvate-oxaloacetate node as the
548 switch point for carbon flux distribution in bacteria, *FEMS Microbiol*
549 *Rev* 29 (2005) 765–94.
- 550 [34] G. Gosset, Improvement of escherichia coli production strains by mod-
551 ification of the phosphoenolpyruvate:sugar phosphotransferase system,
552 *Microb Cell Fact* 4 (2005) 14.

- 553 [35] U. Sauer, F. Canonaco, S. Heri, A. Perrenoud, E. Fischer, The soluble
554 and membrane-bound transhydrogenases *udha* and *pntab* have divergent
555 functions in nadph metabolism of *escherichia coli*, *J Biol Chem* 279
556 (2004) 6613–9.
- 557 [36] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A.
558 Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of *es-*
559 *cherichia coli* k-12 in-frame, single-gene knockout mutants: the keio
560 collection, *Mol Syst Biol* 2 (2006) 2006.0008.
- 561 [37] I. Hug, M. F. Feldman, Analogies and homologies in lipopolysaccharide
562 and glycoprotein biosynthesis in bacteria, *Glycobiology* 21 (2011) 138–
563 51.
- 564 [38] K. R. Patil, I. Rocha, J. Förster, J. Nielsen, Evolutionary programming
565 as a platform for in silico metabolic engineering, *BMC Bioinformatics* 6
566 (2005) 308.
- 567 [39] G. Nair, C. Jungreuthmayer, J. Zanghellini, Optimal knockout strategies
568 in genome-scale metabolic networks using particle swarm optimization,
569 *BMC Bioinformatics* 18 (2017) 78.
- 570 [40] M. Rocha, P. Maia, R. Mendes, J. P. Pinto, E. C. Ferreira, J. Nielsen,
571 K. R. Patil, I. Rocha, Natural computation meta-heuristics for the in
572 silico optimization of microbial strains, *BMC Bioinformatics* 9 (2008)
573 499.
- 574 [41] H. Link, D. Christodoulou, U. Sauer, Advancing metabolic models with
575 kinetic information, *Curr Opin Biotechnol* 29C (2014) 8–14.
- 576 [42] U. Sauer, D. R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyper-
577 ski, K. Wüthrich, J. E. Bailey, Metabolic flux ratio analysis of ge-
578 netic and environmental modulations of *escherichia coli* central carbon
579 metabolism, *J Bacteriol* 181 (1999) 6679–88.
- 580 [43] A. Kremling, K. Bettenbrock, E. D. Gilles, A feed-forward loop guar-
581 antees robust behavior in *escherichia coli* carbohydrate uptake, *Bioin-*
582 *formatics* 24 (2008) 704–10.

- 583 [44] H. Link, K. Kochanowski, U. Sauer, Systematic identification of al-
584 losteric protein-metabolite interactions that control enzyme activity in
585 vivo, *Nat Biotechnol* 31 (2013) 357–61.
- 586 [45] R. Schuetz, L. Kuepfer, U. Sauer, Systematic evaluation of objective
587 functions for predicting intracellular fluxes in escherichia coli, *Mol Syst*
588 *Biol* 3 (2007) 119.
- 589 [46] S. Ohno, H. Shimizu, C. Furusawa, Fastpros: screening of reaction
590 knockout strategies for metabolic engineering, *Bioinformatics* 30 (2014)
591 981–7.
- 592 [47] S. Kirkpatrick, C. D. Gelatt, Jr, M. P. Vecchi, Optimization by simu-
593 lated annealing, *Science* 220 (1983) 671–80.
- 594 [48] P. P. Cherepanov, W. Wackernagel, Gene disruption in escherichia coli:
595 Tcr and kmr cassettes with the option of flp-catalyzed excision of the
596 antibiotic-resistance determinant, *Gene* 158 (1995) 9–14.
- 597 [49] S. Bernatchez, C. M. Szymanski, N. Ishiyama, J. Li, H. C. Jarrell, P. C.
598 Lau, A. M. Berghuis, N. M. Young, W. W. Wakarchuk, A single bi-
599 functional udp-glcnac/glc 4-epimerase supports the synthesis of three
600 cell surface glycoconjugates in campylobacter jejuni, *J Biol Chem* 280
601 (2005) 4792–802.
- 602 [50] I. C. Schoenhofen, D. J. McNally, E. Vinogradov, D. Whitfield, N. M.
603 Young, S. Dick, W. W. Wakarchuk, J.-R. Brisson, S. M. Logan, Func-
604 tional characterization of dehydratase/aminotransferase pairs from he-
605 licobacter and campylobacter: enzymes distinguishing the pseudaminic
606 acid and bacillosamine biosynthetic pathways, *J Biol Chem* 281 (2006)
607 723–32.
- 608 [51] N. B. Olivier, M. M. Chen, J. R. Behr, B. Imperiali, In vitro biosyn-
609 thesis of udp-n,n'-diacetylbacillosamine by enzymes of the campylobac-
610 ter jejuni general protein glycosylation system, *Biochemistry* 45 (2006)
611 13659–69.
- 612 [52] K. J. Glover, E. Weerapana, M. M. Chen, B. Imperiali, Direct bio-
613 chemical evidence for the utilization of udp-bacillosamine by pglc, an
614 essential glycosyl-1-phosphate transferase in the campylobacter jejuni
615 n-linked glycosylation pathway, *Biochemistry* 45 (2006) 5343–50.

- 616 [53] K. J. Glover, E. Weerapana, B. Imperiali, In vitro assembly of the
617 undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic n-
618 linked glycosylation, *Proc Natl Acad Sci U S A* 102 (2005) 14255–9.
- 619 [54] J. Kelly, H. Jarrell, L. Millar, L. Tessier, L. M. Fiori, P. C. Lau, B. Allan,
620 C. M. Szymanski, Biosynthesis of the n-linked glycan in campylobacter
621 jejuni and addition onto protein through block transfer, *J Bacteriol* 188
622 (2006) 2427–34.
- 623 [55] D. Linton, N. Dorrell, P. G. Hitchen, S. Amber, A. V. Karlyshev, H. R.
624 Morris, A. Dell, M. A. Valvano, M. Aebi, B. W. Wren, Functional anal-
625 ysis of the campylobacter jejuni n-linked protein glycosylation pathway,
626 *Mol Microbiol* 55 (2005) 1695–703.

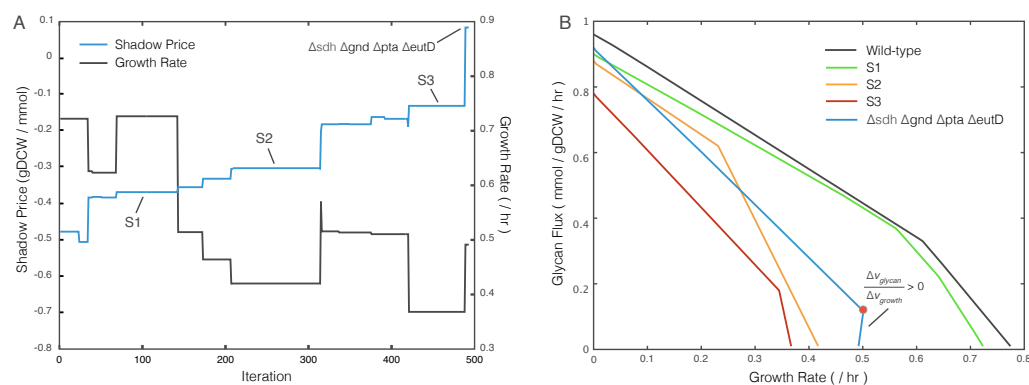


Figure S1: Representative simulated annealing simulation results identifying a growth-coupled strain of type EcGM2. **(A)** Shadow price and FBA-maximized growth rate versus iteration number for the identification of strain $\Delta sdh\Delta gnd\Delta pta\Delta eutD$. **(B)** Production envelopes of the strains (S1-S3) at the iteration points indicated in A. The simulation was terminated once a positive shadow price is found, visualized by the slope of the production envelope. The red dot indicates the optimal operating point of maximal growth rate for the growth-coupled strain.