Multiple Optimal Phenotypes Overcome Redox and Glycolytic Intermediate Metabolite Imbalances in Escherichia coli pgi Knockout Evolutions

McCloskey, Douglas; Xu, Sibei; Sandberg, Troy E.; Brunk, Elizabeth; Hefner, Ying; Szubin, Richard; Feist, Adam M.; Palsson, Bernhard O.

Published in:
Applied and Environmental Microbiology

Link to article, DOI:
10.1128/AEM.00823-18

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Multiple Optimal Phenotypes Overcome Redox and Glycolytic Intermediate Metabolite Imbalances in *Escherichia coli pgi* Knockout Evolutions

Douglas McCloskey,a,b Sibei Xu,a Troy E. Sandberg,a Elizabeth Brunk,a Ying Hefner,a Richard Szubin,a Adam M. Feist,a,b Bernhard O. Palsson,a,b

aDepartment of Bioengineering, University of California, San Diego, La Jolla, California, USA
bNovo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

**ABSTRACT** A mechanistic understanding of how new phenotypes develop to overcome the loss of a gene product provides valuable insight on both the metabolic and regulatory functions of the lost gene. The *pgi* gene, whose product catalyzes the second step in glycolysis, was deleted in a growth-optimized *Escherichia coli* K-12 MG1655 strain. The initial knockout (KO) strain exhibited an 80% drop in growth rate that was largely recovered in eight replicate, but phenotypically distinct, cultures after undergoing adaptive laboratory evolution (ALE). Multi-omic data sets showed that the loss of *pgi* substantially shifted pathway usage, leading to a redox and sugar phosphate stress response. These stress responses were overcome by unique combinations of innovative mutations selected for by ALE. Thus, the coordinated mechanisms from genome to metabolome that lead to multiple optimal phenotypes after the loss of a major gene product were revealed.

**IMPORTANCE** A mechanistic understanding of how microbes are able to overcome the loss of a gene through regulatory and metabolic changes is not well understood. Eight independent adaptive laboratory evolution (ALE) experiments with *pgi* knockout strains resulted in eight phenotypically distinct endpoints that were able to overcome the gene loss. Utilizing multi-omics analysis, the coordinated mechanisms from genome to metabolome that lead to multiple optimal phenotypes after the loss of a major gene product were revealed.

**KEYWORDS** *Escherichia coli*, adaptive laboratory evolution, multi-omics analysis, mutation analysis, *pgi* gene knockout, systems biology

The flux split between upper glycolysis and the oxidative pentose phosphate pathway (oxPPP) at the glucose 6-phosphate (G6P) node is a major determinant of the flux state of a cell’s core metabolic function. This is particularly true when a cell is exposed to glucose, which is commonly used in many laboratory medium formulations and is found under many environmental conditions in nature. The loss of phosphoglucone isomerase (PGI), encoded by *pgi*, induces detrimental physiological consequences (1–5). The removal of *pgi* generates an imbalance in glycolytic intermediates from the loss of upper glycolytic flux that leads to a loss of fitness and induces a sugar phosphate stress response. The sugar phosphate stress response involves the actions of both small RNAs (sRNAs) and transcription factors (TFs) that induce transcription-level changes aimed at alleviating the imbalance (6–8). The removal of *pgi* also generates an overabundance of NADPH and redox imbalance by redirecting glycolytic flux into the oxPPP. NADPH provides reducing equivalents for biosynthesis. In addition, NADPH plays an important role in reactive oxygen species (ROS) detoxification by regenerating reduced glutathione (gthrd) (9). Increased availability of NADPH in *pgi* knockout

Received 15 April 2018  Accepted 19 July 2018  Accepted manuscript posted online 27 July 2018


Editor Maia Kivisaar, University of Tartu

Copyright © 2018 American Society for Microbiology. All Rights Reserved.

Address correspondence to Bernhard O. Palsson, palsson@ucsd.edu.
backgrounds has proven useful in various biotechnology applications in order to increase cofactor and heterologous pathway production (1, 2, 10).

Adaptive laboratory evolution (ALE) of pgi mutants has been carried out to better understand the physiological changes required to overcome genetic perturbation (3, 4). ALE is an experimental method that introduces selection pressure (e.g., growth rate selection) in a controlled environmental setting (11–13). Using ALE, organisms can be perturbed from their evolutionary optimized homeostatic states, and their readjustments can be studied during the course of adaptation to reveal novel and nonintuitive component functions and interactions (14). Previous ALEs of pgi mutants have demonstrated a rewiring of central metabolic fluxes (4) and diversity in endpoint physiological phenotypes (3). The diversity in endpoint physiological phenotypes is directly attributed to the existence of alternate optimal metabolic and regulatory network states that can achieve the same physiological function (3). However, the mechanisms and coordination of the regulatory and metabolic network required to produce physiologically distinct, yet fit, phenotypes are not well understood. In addition, these studies were conducted with a starting strain that was not previously optimized to the growth conditions of the experiment. This confounds the interpretation of the experimental results, because adaptations to the growth conditions and loss of the gene occur simultaneously.

The consequence of the loss of a major metabolic gene can be studied at the systems level through the integration of multi-omics data sets (i.e., metabolomics, fluxomics, proteomics, and transcriptomics) to gain deeper insight into the function of the gene in the context of the biological system as a whole. Previous work has found that the metabolic network is robust to perturbations through adjustments made at the regulatory level that coordinate rerouting of flux with enzyme level (5, 15). While these studies reveal insights into the immediate response of gene loss, the adaptive changes required to overcome the loss were not explored. In addition, improvements in omics data acquisition and analysis methods could improve and reveal new relationships between changes in omics data at one layer of the system to another.

In this study, a combination of experimental design (i.e., starting with a strain that was pre-evolved on glucose M9 minimal medium) and systems analysis from multi-omics data was used to mechanistically investigate how multiple phenotypes can overcome the loss of pgi. First, the reduction in fitness after the pgi KO was found to be attributed to malfunctions in the regulatory and metabolic network that were incapable of handling the redox and glycolytic intermediate metabolite imbalance induced from major shifts in central metabolic flux. Second, all evolved pgi KO lineages regained a substantial portion of fitness but were found to display unique genotypes and displayed unique physiologies. Third, the recovery in fitness was enabled by mutations that were selected for by ALE that altered the transcription regulatory network (TRN) and metabolic fluxes to alleviate the redox and sugar phosphate imbalance. These regulatory and metabolic alterations were unique across all endpoints, which led to the emergence of multiple optimal phenotypes.

RESULTS

Diversity in ALE endpoint phenotypes points to multiple optimal selection outcomes. To eliminate the confounding variable of adaptation to the growth conditions of the experiment, a wild-type E. coli K-12 MG1655 strain previously evolved under glucose minimal medium at 37°C (16) (denoted “Ref”) was selected as the starting strain (Fig. 1A). This selection was made to separate changes caused by adaptation to the loss of a gene product from those caused by adaptation to the growth conditions of the experiment.

PGI (encoded by pgi, phosphoglucone isomerase) was removed from the Ref strain to generate strain uPgi (for “unevolved pgi knockout strain”) (Fig. 1B). The loss of pgi resulted in an 81% loss in growth rate (Fig. 1C and D). Eight uPgi independently inoculated starting cultures were simultaneously evolved on glucose minimal medium at 37°C in an automated ALE platform (16, 17) denoted “evolved pgi knockout strains”
A statistically significant increase in final growth rate (Student’s t test, \( P < 0.05 \)) was found in all ALE endpoints of the ePgi lineages (average standard deviation [SD], 284%, 20% increase in growth rate) compared to the uPgi strain (Fig. 1D). Metabolomics, fluxomics, transcriptomics, genomics, and phenomics data were collected from exponentially growing cultures inoculated in triplicate from the Ref strain, the uPgi strain, and each of the 8 independently evolved endpoint lineage populations (ePgi01 to ePgi08). It is important to note that the data presented below were derived from endpoint populations as opposed to isolated clones.

Statistically significant variabilities in growth rate, acetate secretion, and glucose uptake rate were found in the ePgi samples (Fig. 1D; see also Table S1 in the supplemental material). Specifically, replicates 3, 4, 7, and 8 excreted acetate. Replicate 4, in particular, had acetate secretion levels similar to those of the uPgi strain and the highest growth and glucose consumption rates of all endpoints. The overall mutation load also differed across lineages. The overall numbers of mutations (mutation frequency, \( < 0.2 \)) for each of the endpoint replicates were 15, 12, 9, 13, 7, 14, 9, and 7, respectively (Table S7).

The different phenotypes and genotypes displayed by the endpoints raised two defining questions: what metabolic and regulatory changes occurred to allow for a large improvement in fitness without the use of upper glycolysis? And how were a diversity of endpoint physiologies capable of overcoming the loss of PGI? To answer these questions, intracellular metabolite levels, gene expression levels, and flux levels were measured for the Ref strain, uPgi strain, and ePgi endpoint populations (Tables S1 to S7).

**PGI KO shifted metabolic flux.** Genome-scale metabolic flux analysis (MFA) (18) found significant shifts in flux splits throughout central metabolism in response to the loss of PGI (Fig. 2 and Table S5). Note that all fluxes discussed in the text passed observability criteria described previously (18). Flux splits included the distribution of flux through the oxidative pentose phosphate pathway (oxPPP) (phosphogluconate dehydrogenase [EDD], 6-phosphogluconate dehydrogenase [GND], and 6-phosphogluconolactonase [PGL]), flux through the nonoxidative branch (nonOxPPP), flux around the anaerobic reactions

![FIG 1](http://aem.asm.org/) Evolution of knockout (KO) strains from a preevolved (i.e., optimized) wild-type strain. (A) Wild-type (wt) *E. coli* (MG1655 K-12) was previously evolved on glucose minimal medium at 37°C (16). An isolate from the endpoint of the evolutionary experiment was selected as the starting strain for subsequent KO of *pgi* and adaptive laboratory evolution (ALE). (B) Adaptive laboratory evolution trajectories of the evolved knockout lineages. Omics data were collected from the fresh KO, and endpoint lineages included metabolomics, fluxomics, physiology, DNA resequencing, and transcriptomics. (C) Phosphoglucone isomerase (PGI) was disabled by the gene KO. PGI is the first step in glycolysis and converts glucose 6-phosphate (G6P) to fructose 6-phosphate (F6P). (D) Growth rate and glucose (glc-D) uptake and acetate (ac) excretion rates for unevolved KO (uPgi) and evolved KOs (ePgi). Error bars denote 95% confidence intervals from biological triplicates. succ, succinate.
(phosphoenolpyruvate carboxylase [PPC], phosphoenolpyruvate carboxylase kinase [PPCK], malate dehydrogenase [MALS], NADP-dependent malic enzyme 1 [ME1], and ME2), flux into the tricarboxylic acid (TCA) cycle or toward acetate secretion (citrate synthase [CS], acetate secretion [ACt2rpp], and pyruvate dehydrogenase [PDH]), and flux through the lower glyoxylate shunt or through the lower TCA cycle (aconitase [ACONTb], isocitrate dehydrogenase [ICDHyr], and isocitrate lyase [ICL]). A massive increase of over 10,000-fold in flux per mole glucose through the Entner-Doudoroff (ED) pathway (EDD) was found (note that the Ref strain had only minimal flux through the EDD), while a minor 15.6% drop in flux per mole glucose through GND was found in the uPgi strain compared to that with the Ref strain. Redistribution of flux through the nonOxPPP was found. A 79% increase in flux per mole glucose into the TCA through CS was offset by an 87% drop in flux per mole glucose into the TCA through PPC. A minor 6% drop in acetate secretion per mole glucose was found. Further, a significant 380.8% increase in flux per mole glucose through the glyoxylate shunt was found. These changes in major metabolic pathways and flux splits in the cell quantify the magnitude of the perturbation and the initial response of the cell to losing the \textit{pgi} gene function.

**Perturbed glycolytic intermediates generated a sugar phosphate stress response.** Shifts in central metabolic fluxes imbalanced central metabolic intermediate metabolite levels, leading to a sugar phosphate stress response. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to quantify the absolute metabolite concentrations of glycolytic intermediates, pentose phosphate pathway (PPP), and TCA cycle intermediates (Table S2), and transcriptomics was used to quantify the relative shifts in genes targeted by transcription factors (Tables S2 to S4). All measured glycolytic and PPP intermediates changed significantly in the uPgi strain compared to the Ref strain (Fig. 3 and Table S2). In particular, an approximate 5-fold increase in glucose 6-phosphate (G6P) was found in the uPgi strain.
compared to the Ref strain (Fig. 4). G6P is the sole substrate of the 
pgi gene. Abnormal elevations in G6P and an imbalance of the glycolytic intermediates in the uPgi strain were found to induce the sugar phosphate toxicity response transcription factor (TF) SgrR (6–8), as measured through the transcriptomic measurements. SgrR is thought to bind hexose phosphates and induce the expression of the small RNA sgrS (6–8) (Fig. 4D and E). ePgi strains dampened sgrS expression, which was highly overexpressed in the uPgi strain. Further, sgrS transcriptionally regulates a number of genes, including the pur regulon, ptsG, and genes involved in biofilm formation and curli formation (7, 8, 19–21).

Imbalances in central carbon intermediates induced TRN responses that were consistent with the literature. Many E. coli TFs are activated by metabolites (22–26). Thus, changes in metabolite levels were investigated to reveal potential TRN responses as measured by changes in expression profiles associated with well-known TF regulons. Many of the measured expression changes appeared to conflict with optimal fitness. Specifically, the glp regulon required for glycerol import and catabolism was upregulated by C-reactive protein (CRP)-cAMP (27). cAMP was significantly elevated in the uPGi strain (supplemental material). This hard-wired regulation led to massive upregulation of the glp regulon in the uPGi strain, which could potentially have led to counterproductive allocation of the proteome to glycerol metabolism.
Interestingly, the hexose phosphate importer gene, *uhpT*, was overexpressed in the *uPgi* strain compared to the Ref strain. High periplasmic G6P plausibly activated the *uhpAB* two-component system, which in turn upregulates the expression of the hexose phosphate importer gene *uhpT* (28–30). This result suggests that the concentration build-up in the *uPgi* strain was so great that G6P spilled over into the periplasmic space (31, 32)(Fig. 3, top). Increased expression of *uhpT* could have generated a loop whereby excessive G6P that spilled into the periplasmic space would be reimported into the cytosol. The transcriptional attenuation of *ptsG* by *sgrS* may act to compensate for this futile cycle.

It should be noted that in the present context, the two examples given above (i.e., the *glp* regulon and G6P cycling) could be considered counterproductive and not aligning with optimal growth. However, in a different environment, these hard-wired regulatory circuits could play a critical role in maintaining optimal physiology. This provides evidence that the combination of tightly controlled laboratory setting, gene knockout, and multi-omics data collection and analysis described here provided a useful means to reveal these hard-wired responses.

**Imbalances in central carbon intermediates were mirrored in amino acid pools.**

In addition to regulatory shifts, biomass components directly reflected the levels of their biosynthetic precursors (Fig. 3B and C). The aromatic amino acids L-tyrosine (*tyr-L*), L-phenylalanine (*phe-L*), and L-tryptophan (*trp-L*) are derived from phosphoenolpyruvate (PEP). A decrease in PEP levels in the *uPgi* strain and a rise in PEP levels in the *ePgi* strains were mirrored by all three of the amino acids (Fig. 3C). An increase in ribose 5-phosphate (*r5p*) levels in the *uPgi* and *ePgi* strains were mirrored by the downstream amino acid L-histidine (*his-L*) and nucleotide UMP (Fig. 3B).

Similar trends were found for amino acid and precursor pairs L-serine (*ser-L*) and 2-phosphogluconate (2pg), L-aspartate (asp-L) and oxaloacetate (oaa), L-alanine (ala-L) and pyruvate (pyr), and L-glutamate (glu-L), L-glutamine (glu-L), and alpha-ketoglutarate (akg), respectively (Table S2). Perturbations in the distribution and abundance of proteogenic amino acids have been shown to alter protein synthesis rates, leading to...
a drop in the growth rate (33–37). The drop in the growth rate of the uPgi strain and regain of fitness in the ePgi strains provide evidence that an imbalance in glycolytic intermediates directly alters growth rate via manipulating proteogenic amino acid levels.

The regulatory response to elevated G6P levels and the relationship between biomass components and their precursors reflected the importance of balancing glycolytic, PPP, and TCA cycle intermediates to maintain balanced ratios of amino acid levels for protein biosynthesis and purine/pyrimidine mononucleotides for nucleotide biosynthesis.

Mutations that targeted alternative glucose import systems corrected TRN responses and helped rebalance glycolytic intermediate levels. Major alterations in expression profiles were found in the evolved KO strains that correlated well with mutations detected in TFs. These included mutations to galR (Fig. 5) and malT (Fig. 6) in the ePgi strains. A 22-nucleotide deletion in the small-molecule-binding domain of galR (DEL) 322Δ33nt negates galR repression in ePgi07. (A) Regulatory network specifically controlled by cAMP-CRP, galR, and galS (76–78). cAMP-CRP can both positively and negatively regulate the expression of galR, galS, galETKM, galP, and mgiBAC; GalR and GalS act as repressors; and GalR and GalS bound to galactose active primarily as activators. (B) Crystal structure of the galR transcription factor (70). The position of the deletion is highlighted in red, the small-molecule-binding domain is highlighted in cyan, and the H-T-H DNA-binding region is highlighted in magenta. Images created using the VMD software tool (v. 1.9.2; http://www.ks.uiuc.edu/Research/vmd/). (C) Mutation frequency for galR, metabolite concentration for cAMP, and expression profiles of galR-controlled operons. Note the increased expression of galP and galETKM in ePgi07. Metabolite concentrations are derived from averages taken from triplicate cultures that were analyzed in duplicate (n = 6). Gene expression values are derived from averages of biological duplicates.
galR in ePgi07 appears to negate the repression of the galR-controlled operons (Fig. 5). These include galETKM, galP, and mglBAC that encode enzymes for galactose catabolism, symport, and ABC transport, respectively (38). These operons are also regulated by CRP-cAMP and were not expressed in the Ref strain. The galactose importers have lesser affinity for the transport of glucose, which may give ePgi07 an additional route to import and catabolize glucose from the environment. In addition, the mutation may have aided in conserving PEP for aromatic amino acid production, which was limiting fitness in all of the pgi strains (as discussed previously).

In another example, a mobile element insertion (MOB) that truncated the MalT TF in ePgi06 was found that appeared to silence the expression of MalT-controlled operons. (Fig. 6). A schematic of the malT operon (39) and truncated MalT peptide. The mobile element insertion introduced a stop codon that reduced the MalT peptide from 901 amino acids to 29 amino acids. All binding domains and catalytic sites were cleaved (40, 41). Operons controlled by malT (39). All regulators except malT and CRP-cAMP have been omitted. MalT-controlled genes are involved in glycogen turnover and may give ePgi06 an advantage in controlling the levels of hexose phosphates that are converted to and broken down from glycogen. (C) Mutation frequency of malT, metabolite concentration for cAMP, and expression profiles for malT and malT-regulated genes. Note the significantly repressed gene expression levels of MalT-controlled genes in ePgi06. Metabolite concentrations are derived from averages taken from triplicate cultures that were analyzed in duplicate (n = 6). Gene expression values are derived from averages of biological duplicates.

An imbalance in redox carriers was compensated for by shifts in hydrogenase flux and buffered by glutathione. Genome-scale MFA (18) confirmed that removal of the pgi gene diverted all upper glycolytic flux into the oxidative pentose phosphate pathway (oxPPP) (Fig. 2). The loss of PGI resulted in a 556.7% increase in flux per mole glucose toward the oxPPP in the uPgi strain (14.4 and 94.7% flux per mole glucose in the Ref and uPgi strains, respectively). Note that as discussed previously, a large portion of the flux into the oxPPP was diverted down the ED pathway after the first NADPH-generating step to avoid generating additional NADPH via the second NADPH-generating step of the oxPPP (Fig. 2). Rearrangement of flux through hydrogenases to compensate for the increased flux toward NADPH generation were found (Fig. 2). Notable is the reversed utilization of the transhydrogenases from net NADPH to NADH.
generation in the uPgi and ePgi strains (approximately −8-fold change in the uPgi and ePgi strains in NADPH generation through pyridine nucleotide transhydrogenase [THD2pp], and −3-fold change in the uPgi strain and −1 to 4-fold change in ePgi strains in NADH generation through pyridine nucleotide transhydrogenase [NADTRHD]). Other hydrogenases significantly altered include serine dehydrogenase (LSERDHr, note the altered levels of ser-L mentioned previously), as well as isocitrate dehydrogenase (ICD) and glutamate synthase (GLUSy, note the altered levels of akg, gln-L, and glu-L mentioned previously).

The increased flux through the oxPPP would generate an increased abundance of NADPH and thus a redox imbalance. LC-MS/MS (42) was used to quantify the absolute metabolite concentrations of the redox carriers. While major shifts in the redox carriers were found (Fig. 3A), a statistically significant change in NADPH between the Ref and uPgi strains was not found. However, statistically significant changes in NADP and reduced and oxidized glutathione were found. This indicates a potential rapid buffering of NADPH by the glutathione via glutathione reductase (GTHOr) (Fig. 3A).

**High NADPH promoted activation of oxidative stress responders.** Mutations in soxR in ePgi02 and rseC in ePgi01 were found that altered the expression of oxidative stress genes (Fig. 7). The soxR mutation truncated the Fe-S cluster binding site of the SoxR peptide by introducing a premature stop codon (Fig. 7C). Cleavage of the Fe-S cluster does not affect DNA binding, but transcriptional activation of the target genes soxS and fumC and transcriptional deactivation of soxR are impaired (43–45). soxR was upregulated in ePgi02, which indicated that the mutation negated the self-regulation of soxR.

The rseC single-nucleotide deletion introduced a frameshift that truncated a large portion of the transmembrane helix region starting several amino acids from the initial deletion (Fig. 7D) that may affect Rsx-RseC complex formation or activity. rseC mutants were found previously to exhibit constitutive soxS expression by preventing the Rsx-RseC complex from inhibiting reduction and inactivating SoxR (46) (Fig. 7A). soxS, as well as many of its downstream activation targets, including acrA, acrB, fldA, fpr, inoA, and sodA, were upregulated in ePgi01, which indicated that the mutation promotes the expression of soxS. The Rsx-RseC complex prevents the reduction and inactivation of SoxR (46).

**Mutations in transhydrogenases helped alleviate redox imbalance.** Mutations selected during adaptive evolution also introduced innovations that targeted metabolic network elements involved in NADPH production. Mutated transhydrogenases included sthA, pntB, icd, and zwf. The soluble and membrane-bound transhydrogenases act to interconvert NADP(H) and NAD(H) (47, 48) (Fig. 8). Mutations in the soluble sthA (50) and membrane-bound pntB (51) transhydrogenase genes in ePgi07 and ePgi04, respectively, were found (Fig. 8). The sthA mutation appeared near the dimerization domain and may affect enzyme complex formation. The pntB mutation appeared in the transmembrane region and may affect catalytic activity or membrane association. It has been demonstrated that altered activities of sthA and pntAB confer a fitness advantage in pgi mutant strains by rebalancing the ratio of NADH to NADPH (47, 48). Interestingly, mutations in sthA and pntB were selected for in previous evolutions of a pgi strain (3). This observation provides further evidence that the sthA and pntB mutations provided a fitness advantage to ePgi07 and ePgi04 by rebalancing the ratios of NADH to NADPH via modulating the activities of the transhydrogenases. Note that ePgi07 and ePgi04 were also found to have the highest increases in flux through the soluble transhydrogenases.

Isocitrate dehydrogenase (ICD) catalyzes the conversion of isocitrate (icit) to 2-oxoglutarate (akg) while reducing NADP" to NADPH (Fig. 9). The activity of ICD also regulates the flux split between the full TCA cycle and the glyoxylate shunt (49–51). A point mutation at residue 395 that changed the amino acid from positively charged (l-arginine) to negatively charged (l-cysteine) in ICD was found in all ePgi replicates except replicate 7 (Fig. 9). The mutation occurs 4 Å from the phosphate moiety of
Residue 395 has been shown to be directly involved in NADP binding (52) and appears to allow the mutated enzyme to either utilize NAD as a cofactor or exclude NADPH from the active site by like-charge repulsion. The mutation was found to redirect flux through the glyoxylate shunt instead of the TCA cycle and may provide a fitness advantage to the ePgi strains by limiting the production of NADPH in the TCA cycle.

The metabolome, fluxome, transcriptome, and genome were unique in each ePgi strain. While the ePgi strains were not able to recover the fitness of the Ref strain, the ePgi strains exhibited upregulated expression of oxidative stress genes, which may indicate a compensatory mechanism to offset the loss of NADPH production.

**FIG 7** Mutations in *soxR* and *rseC* that altered the expression of oxidative stress genes. (A) Protein-protein interaction schema between SoxR and Rsx-RseC. In the reduced form, the iron sulfur clusters of the SoxR homodimers sense the presence of free radicals and ROS (79, 80). The oxidation of the iron sulfur clusters by free radicals and ROS induces a conformational change from the inactive form to the active form (81). While both reduced and inactive forms and oxidized and active forms of SoxR are capable of binding DNA, only the active form is capable of activating or inhibiting transcription (43, 82–85). The Rsx-RseC complex prevents reduction and inactivation of SoxR (46). (B) Regulatory schematic of a subset of SoxR and SoxS-controlled operons. (C) Crystal structure of SoxR (86). The *soxR* single-nucleotide polymorphism (SNP) eliminated the Fe-S cluster binding site of the SoxR peptide. The SoxR DNA-binding region in proximity to single-stranded DNA (sDNA) is shown below. (D) Crystal structure of *rseC*. The *rseC* mutation cleaved a large portion of the transmembrane helix region that may affect Rsx-RseC complex formation or activity. The mutated and/or cleaved residues are shown in red. AAs, amino acids. Images in panels C and D were created using the VMD software tool (v. 1.9.2; http://www.ks.uiuc.edu/Research/vmd/). (E) Mutation frequency and gene expression profiles. Gene expression values are derived from averages of biological duplicates.
they were able to recover the initial growth rate of wild-type MG1655. Many intermediate and cofactor levels, including G6P, remained perturbed in ePgi strains to various degrees. However, the majority of initially elevated transcription involving sugar phosphate stress, carbon catabolite repression, the uhpT transporter, and other general stress responders in the uPgi strain were dampened or completely shut down in ePgi strains. This indicated that the TRN had evolved to cope with the changed metabolome. Many of these changes to the TRN could be directly attributed to mutations.
The recovery to wild-type levels was in part made possible by a complete rewiring of central carbohydrate metabolic flux splits (Fig. 3). The rewiring differed substantially between ePgi strains. While an increase in flux through GND and a decrease in flux through the ED pathway occurred in all ePgi endpoints, the flux through each pathway differed substantially. GND flux increases varied from 389.9 to 604.6% per mole glucose compared to the Ref strain, and ED pathway flux increases varied from 150.9 to 4,463.8% per mole glucose compared to the Ref strain. NonOxPPP flux varied substantially among the strains, even altering between net forward and reverse utilization of transaldolase A (TALA). Of particular note, flux through PPC was regained in ePgi04 and ePgi07, and flux through the glyoxylate shunt was lost in ePgi07. The increased flux through the PPC and loss of flux through the glyoxylate shunt in particular correlate with mutations in the transhydrogenases and icd, respectively.

**Conclusion.** The loss of pgi induced massive perturbations to the metabolome, fluxome, and transcriptome in *E. coli* and led to a greatly retarded postknockout growth rate. In contrast to previous work (5), it was found that the loss of pgi induced major changes in all omics data measured not just local to the perturbation but also in distal network locations. Flux rerouting to compensate for the loss of pgi imbalanced the PPP and glycolytic intermediate levels, which led to a sugar phosphate stress response. The deleterious effects of this response were attributed to a misallocation of protein, a deleterious cycle of reimport of hexose phosphate, and alterations in the distributions and amount of proteogenic amino acids and nucleotides. Redistribution of glycolytic flux into the oxPPP generated an overabundance of NADPH, which led to a redox imbalance.

ALE selected for mutations that helped alleviate redox and glycolytic intermediate imbalances. Differences in the metabolome and mutation landscape led to a diversity of expression and fluxome profiles in ePgi strains. The multitude of hydrogenases and routes to generate glycolytic intermediates allowed for regulatory and metabolic

![Segment of TCA Cycle](http://aem.asm.org/)
flexibility in overcoming redox and glycolytic intermediate imbalance. Finally, the diversity in fixed mutations and the concomitant emergence of multiple optimal phenotypes was a manifestation of this metabolic flexibility. Additional studies utilizing ALE to uncover the response of a cell to a major network perturbation are likely to uncover additional mechanisms available to biological networks to overcome such major perturbations.

MATERIALS AND METHODS

Biological material, analytical reagents, and experimental conditions. (i) Biological material. A wild-type E. coli K-12 MG1655 strain previously evolved under glucose minimal medium at 37°C (ATCC 700926) (16, 17) served as the starting strain. Lambda Red-mediated DNA mutagenesis (53) was used to create the knockout strains (DNA mutagenesis and PCR confirmation primers are given in Table 1). Knockouts were confirmed by PCR and DNA resequencing. The pgil gene encoding triosephosphate isomerase (TPI) was removed. All cultures were grown in 25 ml of unlabeled or labeled glucose M9 minimal medium (54) with trace elements (55) and sampled from a heat block in 50-ml autoclaved tubes that were maintained at 37°C and aerated using magnets.

(ii) Materials and reagents. Uniformly labeled [13C6]glucose and 1-[13C6]glucose was purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA). Unlabeled glucose and other medium components were purchased from Sigma-Aldrich (St. Louis, MO). LC-MS reagents were purchased from Honeywell Burdick & Jackson (Muskegon, MI), Fisher Scientific (Pittsburgh, PA), and Sigma-Aldrich.

(iii) Reaction knockout selection. UO1366 (56) was used as the metabolic model for E. coli metabolism; GLPK (version 4.57) was used as the linear program solver. Markov chain Monte Carlo (MCMC) sampling (57) was used to predict the flux distribution of the optimized reference strain. Uptake, metabolism; GLPK (version 4.57) was used as the linear program solver. Markov chain Monte Carlo (MCMC) sampling (57) was used to predict the flux distribution of the optimized reference strain. The top 9 reaction deletions from the rank-ordered set of reactions that met the experimentally essential, or not actively expressed under the experimental growth conditions were not deletions were ranked by (i) averaged sampled flux, (ii) the number of immediate upstream and downstream metabolites that could be measured, and (iii) the number of genes required to produce a functional enzyme. The reactions involved in sampling loops that were computationally or experimentally essential, or not actively expressed under the conditions of growth, were not included in the analysis. Also, reactions that would require more than one genetic alteration to abolish activity were excluded. The top 9 reaction deletions from the rank-ordered set of reactions that met the above-mentioned criteria were chosen for implementation.

Adaptive laboratory evolution. Cultures were serially propagated (100 µl passage volume) in 15-ml (working volume) flasks of M9 minimal medium with 4 g/liter glucose, kept at 37°C, and well mixed for full aeration. An automated system passed the cultures to fresh flasks once they had reached an optical density at 600 nm (OD600) of 0.3 (equivalent to an OD600 of ~1 on a traditional spectrophotometer with a 1-cm path length; Tecan Sunrise plate reader), a point at which nutrients were still in excess and exponential growth had not started to taper off (confirmed with growth curves and high-performance liquid chromatography [HPLC] measurements). Four OD600 measurements were taken from each flask, and the slope of ln(OD600) versus time determined the culture growth rates. A cubic interpolating spline constrained to be monotonically increasing was fit to these growth rates to obtain the fitness trajectory curves.

Multi-omics data processing. (i) Phenomics. Physiological measurements for culture density were measured at 600 nm absorbance with a spectrophotometer and correlated to cell biomass. Samples to determine substrate uptake and secretion were filtered through a 0.22-µm filter (polyvinylidene difluoride [PVDF]; Millipore) and measured using refractive index (RI) detection by HPLC (Agilent 12600 Infinity) with a Bio-Rad Aminex HPX87-H ion exclusion column (injection volume, 10 µl) and 5 mM H2SO4 as the mobile phase (0.5 ml/min, 45°C). Growth, uptake, and secretion rates were calculated from a minimum of four steady-state time points.

(ii) LC-MS/MS instrumentation and data processing. Metabolites were acquired and quantified on Qtrap 5500 mass spectrometer (AB Sciex, Framingham, MA) and processed using MultiQuant 3.0.1, as described previously (42). Mass isotopomer distributions (MIDs) were acquired on the same instrument and processed using MultiQuant 3.0.1 and PeakView 2.2, as described previously (58).

(iii) Metabolomics. Internal standards were generated as described previously (59). All samples and calibrators were spiked with the same amount of internal standard taken from the same batch of internal standards. Calibration curves were run before and after all biological and analytical replicates. The consistency of quantification between calibration curves was checked by running a quality control sample that was composed of all biological replicates twice a day. Solvent blanks were injected every

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>Confirmation Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_KO_primer</td>
<td>pgi_F_KO_primer</td>
<td>ACAATTCTCAAAAATCGAAGATTGTGTAAATGAAACATCAAT</td>
<td>F_conf_pr</td>
</tr>
<tr>
<td>R_KO_primer</td>
<td>pgi_R_KO_primer</td>
<td>CGCTTATACGCTCATTACATGAGCTGTTAAATGTTAGCTGGAGCGCTTTC</td>
<td>R_conf_pr</td>
</tr>
</tbody>
</table>

TABLE 1 Primers used to generate the pgil KO strains in this study.
ninth sample to check for carryover. System suitability tests were injected daily to check instrument performance.

Metabolomics samples were acquired from triplicate cultures (1 ml of cell broth at an OD600 of ~1.0) using a previously described method (60). A pooled sample of the filtered medium that was resampled using the fast Swinnex filtration (FSF) technique and processed in the same way as the biological triplicates was used as an analytical blank. Extracts obtained from triplicate cultures and refiltered medium were analyzed in duplicate. The intracellular values reported, unless otherwise noted, are derived from the average of the biological triplicates (n = 6). Metabolites in the pooled filtered medium with a concentration greater than 80% of that found in the triplicate samples were not analyzed. In addition, metabolites that were found to have a quantifiable variability (relative standard deviation [RSD], ≥50%) in the quality control samples or any individual components with an RSD of ≥80 were not used for analysis.

Missing values were imputed using a bootstrapping approach, as coded in the R package Amelia II (version 1.7.4, 1,000 imputations) (61). The remaining missing values were approximated as 1/2 the lower limit of quantification for the metabolite normalized to the biomass of the sample. Prior to statistical analyses, metabolite concentrations were log normalized to generate an approximately normal distribution using the R package LMGene (62) (version 3.3, “mult” = “TRUE,” “lowessnorm” = “FALSE”). A Bonferroni-adjusted P value cutoff of 0.01, as calculated from Student’s t test, was used to determine significance between metabolite concentration levels. The glog-normalized values or the median-normalized values to the reference strain (fold change [FC] of the median versus the Ref strain) were used for downstream statistical analyses.

(iv) Fluxomics. Fluxomics samples were acquired from triplicate cultures (10 ml of cell broth at an OD600 of ~1.0) using a modified version of the FSF technique, as described previously (58). MIDs were calculated from biological triplicates run in analytical duplicates (n = 6). MIDs with an RSD greater than 50% were excluded. In addition, MIDs with a mass that was found to have a signal greater than 80% in unlabeled or blank samples were excluded. A previously validated genome-scale MFA model of E. coli with minimal alterations was used for all MFA estimations using INCA (version 1.4) (63), as described previously (18). The model was constrained using MIDs, as well as measured growth, uptake, and secretion rates. Best flux values that were used to calculate the 95% confidence intervals were estimated from 500 restarts.

The 95% confidence intervals were used as lower and upper bound reaction constraints for further constraint-based analyses. MFA-derived constraints that violated optimality were discarded and resampled. The descriptive statistics (i.e., mean, median, interquartile ranges, minimum, maximum, etc.) for each reaction for each model were calculated from 5,000 points sampled from 5,000 steps using optisample (approximate) (64), which resulted in an approximately mixed fraction of 0.5 for all models. A permuted P value of <0.05 and geometric fold change of sampled flux values of >0.001 were used to determine differential flux levels, differential metabolite utilization levels, and differential subsystem utilization levels between models. Demand reactions and reactions corresponding to unassigned, transport; outer membrane porin, transport; inner membrane, inorganic ion transport and metabolism, transport; and outer membrane, nucleotide salvage pathway, oxidative phosphorylation were excluded from differential flux analysis. The geometric fold change of the mean between models and the reference model was used for hierarchical clustering; the median, interquartile ranges, minimum, and maximum values of each sampling distribution for each reaction and model were used as representative samples for downstream statistical analyses.

(v) Transcriptomics. Total RNA was sampled from triplicate cultures (3 ml of cell broth at an OD600 of ~1.0), immediately added to 2 volumes of Qiagen RNaseq protect bacteria reagent (6 ml), vortexed for 5 s, incubated at room temperature for 5 min, and immediately centrifuged for 10 min at 17,500 rpm. The supernatant was decanted, and the cell pellet was stored at −80°C. Cell pellets were thawed and incubated with Ready-Lyse lysosome, SUPERase In, and 20% SDS for 20 min at 37°C. Total RNA was isolated and purified using the Qiagen RNasy minikit columns and following vendor procedures. An on-column DNase treatment was performed for 30 min at room temperature. RNA was quantified using a NanoDrop spectrophotometer and assessed for quality by running an RNA Nano chip on a Bioanalyzer. The rRNA was removed using Epicentre’s Ribo-Zero rRNA removal kit for Gram-negative bacteria. A Stranded RNA sequencing (RNA-Seq) kit (catalog no. KK8401; Kapa Biosystems) was used, according to the manufacturer’s protocol, to create sequencing libraries with an average insert-length of around ~300 bp for two of the three biological replicates. Libraries were run on a MiSeq and/or HiSeq platform (Illumina).

RNA-Seq reads were aligned using Bowtie (version 1.1.2, with default parameters) (65). Expression levels for individual samples were quantified using Cufflinks (version 2.2.1, library type fr-firststrand) (66). The quality of the reads was assessed by tracking the percentage of unmapped reads and expression level of genes that mapped to the ribosomal gene loci rRS to rMS and rNR to rRF. All samples had a percentage of unmapped reads of less than 7%. Differential expression levels for each condition (n = 2 per condition) compared to either the starting strain or initial knockout strain were calculated using Cuffdiff (version 2.2.1, library type fr-firststrand, library norm geometric) (66). Genes with a 0.05 false-discovery rate (FDR)-adjusted P value of <0.01 were considered differentially expressed. The expression levels for individual samples for all combinations of conditions tested in downstream statistical analyses were normalized using Cuffnorm (version 2.2.1, library type fr-firststrand, library norm geometric) (66). Genes with unmapped reads were imputed using a bootstrapping approach as coded in the R package Amelia II (version 1.7.4, 1,000 imputations). The remaining missing values were filled using the minimum expression level of the data set. Normalized fragments per kilobase per million (FPKM) values for gene expression were log-normalized to generate an approximately normal distribution prior to any statistical analysis. All replicates for a given condition were found to have a pairwise Pearson correlation coefficient of 0.95 or greater.
(vi) DNA resequencing. Total DNA was sample from an overnight culture (1 ml of cell broth at an OD_{600} of 2.0) and immediately centrifuged for 5 min at 8,000 rpm. The supernatant was decanted, and the cell pellet was frozen at −80°C. Genomic DNA was isolated using a NucleoSpin tissue kit (catalog no. 740952.50; Macherey-Nagel), according to the manufacturer’s protocol, including treatment with RNase A. Resequencing libraries were prepared using a Nextera XT kit (catalog no. FC-131-1024; Illumina), following the manufacturer’s protocol. Libraries were run on a MiSeq platform (Illumina). DNA resequencing reads were aligned to the E. coli reference genome (GenBank accession no. U00096.2) using bresseq (version 0.26.0) (67) as populations. Mutations with a frequency of less than 0.1, P value greater than 0.01, or quality score less than 6.0 were removed from the analysis. In addition, genes corresponding to crl, insertion elements (i.e., insH1, insB1, and insA), and the rhs and rrx gene loci were not considered for analysis due to repetitive regions that appear to cause frequent miscalls when using bresq. miRNA and peptide sequence changes were predicted using BioPython (https://github.com/biopython/biopython.github.io/). Large regions of DNA (minimum of 200 consecutive indices) corresponding to duplications were removed from the analysis. In addition, genes corresponding to crl, insertion elements (i.e., insH1, insB1, and insA), and the rhs and rrx gene loci were not considered for analysis due to repetitive regions that appear to cause frequent miscalls when using bresq. miRNA and peptide sequence changes were predicted using BioPython (https://github.com/biopython/biopython.github.io/). Large regions of DNA (minimum of 200 consecutive indices) where the coverage was two times greater than the average coverage of the sample were considered duplications.

Structural analysis. Corresponding Protein Data Bank (PDB) files for genes with a mutation of interest were downloaded from PDB (68, 69). Structural models for genes for which there were no corresponding PDB files were taken from I-TASSER-generated homology models (70) or generated using the I-TASSER protocol (71). The BioPython-predicted sequence changes and important protein features as listed in EcoCyc (72) were visualized and annotated using VMD (73).

SUPPLEMENTAL MATERIAL
Supplemental material for this article may be found at https://doi.org/10.1128/AEM.00823-18.

SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 3, CSV file, 0.4 MB.
SUPPLEMENTAL FILE 4, XLSX file, 6.7 MB.
SUPPLEMENTAL FILE 5, CSV file, 14.5 MB.
SUPPLEMENTAL FILE 6, CSV file, 0.9 MB.
SUPPLEMENTAL FILE 7, CSV file, 0.3 MB.
SUPPLEMENTAL FILE 8, CSV file, 0.1 MB.

ACKNOWLEDGMENTS
We thank José Utrilla for helpful discussion and guidance when implementing the knockouts in the preevolved strain. We thank Jamey Young for helpful discussions throughout the MFA. We thank Laurence Yang for helpful discussions regarding optimization and statistical analysis. This work was funded by the Novo Nordisk Foundation grant NNF10CC1016517. We declare no competing financial interests.

D.M. designed the experiments, generated the strains, conducted all aspects of the metabolomics, fluxomics, phenomics, transcriptomics, and genomics experiments, performed all multi-omics statistical, graph, and modeling analyses, and wrote the manuscript. T.E.S. ran the ALE experiments. T.E.S. assisted with structural analysis. R.S. processed the DNA and RNA samples. S.X. assisted with metabolomics and fluxomics data collection, sample processing, and peak integration. Y.H. assisted with fluxomics data collection and sample processing. A.M.F. designed and supervised the evolution experiments and contributed to the data analysis and the manuscript. B.O.P. conceived and outlined the study, supervised the data analysis, and cowrote the manuscript.

REFERENCES
5. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirawata T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi...


