A Non-natural Protein Rescues Cells Deleted for a Key Enzyme in Central Metabolism

Katherine M. Digianantonio, Maria Korolev, and Michael H. Hecht

Department of Chemistry Princeton University, Princeton, New Jersey 08540, United States

Supporting Information

ABSTRACT: An important goal of synthetic biology is to create novel proteins that provide life-sustaining functions in living organisms. Recent attempts to produce novel proteins have focused largely on rational design involving significant computational efforts. In contrast, nature does not design sequences a priori. Instead, nature relies on Darwinian evolution to select biologically functional sequences from non-designed sequence space. To mimic natural selection in the laboratory, we combed through libraries of novel sequences and selected proteins that rescue E. coli cells deleted for conditionally essential genes. One such gene, gltA, encodes citrate synthase, the enzyme responsible for metabolic entry into the citric acid cycle. The de novo protein SynGltA was isolated as a rescuer of ΔgltA. However, SynGltA is not an enzyme. Instead, SynGltA allows cells to recover from a defect in central carbon and energy metabolism by altering the regulation of an alternative metabolic pathway. Specifically, SynGltA dramatically enhances the expression of prpC, a gene encoding methylcitrate synthase in the propionate degradation pathway. This endogenous protein has promiscuous catalytic activity, which when overexpressed, compensates for the deletion of citrate synthase. While the molecular details responsible for this overexpression have not been elucidated, the results clearly demonstrate that non-natural proteins—unrelated to sequences in nature—can provide life-sustaining functions by altering gene regulation in natural organisms.

KEYWORDS: de novo proteins, synthetic biology, binary code, auxotrophic E. coli, gltA, prpC

Synthetic biology aims to impart novel functions into living organisms. In most cases, these functions are assembled from sequences that already exist in nature. Recently, however, advances in protein design have made it possible to construct entirely novel sequences that fold and function.1–4 Moreover, in some cases, novel sequences have been shown to provide essential functions to living cells.5–8 These advances open the possibility of creating "artificial proteomes" comprising sequences that did not arise in nature, but which nonetheless sustain the growth of living organisms.

In an initial step toward constructing artificial proteomes, we reported that several proteins designed de novo conferred viable metabolic outcomes on E. coli.9 These novel proteins were isolated from a library of sequences designed to fold into stable 4-helix bundles. The library was designed using a strategy called binary patterning, which explicitly specifies each position in a sequence as either polar or nonpolar, but allows the identity of the amino acid at each position to vary combinatorially.9–12 To confirm that binary patterned proteins behaved as designed, several were characterized biophysically and shown to fold into soluble α-helical structures.13–16

While the binary code aims to specify a particular fold, it does not design for function. Nonetheless, proteins from our libraries bind small molecules and catalyze reactions in vitro.17,18 Moreover, several binary patterned proteins function in vivo. In one study, we showed that a de novo protein rescued E. coli from toxic levels of copper,7 and in other studies we demonstrated that de novo proteins can rescue conditionally lethal gene deletions in E. coli.5,6,8 A gene deletion is conditionally lethal if it allows a strain to live on rich medium, but not minimal medium; such strains are called auxotrophs.

In many auxotrophs, the deleted gene encodes a protein responsible for the synthesis or uptake of an essential nutrient, such as an amino acid. Novel proteins encoding many such functions will be required for the eventual construction of artificial proteomes. However, progress toward artificial proteomes will also require novel proteins involved in central carbon and energy metabolism.

Energy production in virtually all aerobic organisms relies on the citric acid cycle (also called the TCA cycle). The centrality of this pathway to metabolism and its presence across a broad range of living systems suggest it arose early in the evolution of natural biology, and make it an attractive target for forays into synthetic biology.

The main entry point into the citric acid cycle is the synthesis of citrate from oxaloacetate and acetyl CoA (Figure 1A). The enzyme responsible for this reaction, citrate synthase, is encoded by the gltA gene in E. coli. Deletion of gltA disables the TCA cycle. However, because E. coli has alternate pathways
In the current study, we show that deletion of citrate synthase—the conditionally essential enzyme that catalyzes entry into the TCA cycle—can be rescued by a novel protein that alters regulation of an alternative metabolic pathway. Specifically, SynGltA upregulates the propionate degradation pathway, encoded by the prp operon. One of the enzymes in this pathway, methylcitrate synthase (encoded by prpC) has weak promiscuous activity capable of synthesizing citrate. Thus, the binary patterned de novo protein, SynGltA, compensates for a defect in central metabolism by upregulating an alternative metabolic pathway, which includes a promiscuous enzyme that produces enough citrate to sustain cell growth.

**RESULTS**

The *de novo* Protein SynGltA Rescues an Auxotroph Deleted for Citrate Synthase. First, we wanted to confirm that SynGltA rescues ΔgltA cells as depicted in Figure 1B. As a positive control, ΔgltA cells were transformed with a plasmid encoding the native *E. coli* GltA (WT-GltA). These cells formed visible colonies on minimal medium in 2 days. The negative control, ΔgltA cells transformed with a plasmid encoding LacZ, produced no colonies in 14 days. Another negative control, cells expressing SynSerB3, a binary patterned protein from the same library, also failed to grow. In contrast, ΔgltA cells transformed with a plasmid encoding SynGltA grow in 4 days (Figure 1B, Table 1 row 1).

**Table 1. Growth of Auxotrophic *E. coli* on Minimal Media**

<table>
<thead>
<tr>
<th>plasmid expressing</th>
<th>strain</th>
<th>LacZ</th>
<th>WT GltA</th>
<th>WT PrpC</th>
<th>WT ZraR</th>
<th>SynGltA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔgltA</td>
<td>X</td>
<td>2 days</td>
<td>3 days</td>
<td>3 days</td>
<td>4 days</td>
<td></td>
</tr>
<tr>
<td>ΔgltAΔprpC</td>
<td>X</td>
<td>2 days</td>
<td>2 days</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ΔgltAΔprpR</td>
<td>X</td>
<td>2 days</td>
<td>2 days</td>
<td>2 days</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ΔgltAΔzraR</td>
<td>X</td>
<td>2 days</td>
<td>2 days</td>
<td>2 days</td>
<td>3 days</td>
<td></td>
</tr>
</tbody>
</table>

*X = no growth. The *de novo* protein SynGltA rescues the ΔgltA auxotroph only in the presence of chromosomal copies of the endogenous *E. coli* prpC and prpR genes. Plates were monitored for colonies with diameters >1 mm for 14 days.*

To confirm that the protein sequence of SynGltA—rather than the mRNA—is responsible for the observed rescue, we made two different mutants that change the mRNA sequence by only one nucleotide, but completely prevent translation of the protein. Tyrosine at the second residue was replaced by the translational stop codon, TAA; and in a separate experiment, a frameshift was introduced at this codon by inserting an extra base (TAT → TATG). Both mutants failed to rescue ΔgltA cells, indicating that translation of the SynGltA protein is required for rescue. Next, to demonstrate that the exact amino acid sequence of SynGltA is required for rescue, we made single amino acid changes and showed that they prevented rescue by SynGltA. From these results, we conclude that the amino acid sequence of the SynGltA protein provides a biological function that rescues ΔgltA cells.

In principle, the SynGltA protein could rescue the glutamate auxotrophy either by enabling the synthesis of citrate (the reaction deleted in ΔgltA), or by facilitating a novel pathway for the biosynthesis of glutamate. If SynGltA facilitated a novel pathway that bypassed the TCA cycle, then this artificial protein would also be expected to rescue the deletion of other enzymes upstream of α-ketoglutarate in the TCA cycle. However, SynGltA does not rescue the glutamate auxotrophy
of ΔcitD, which encodes isocitrate dehydrogenase (Figure 1A). Thus, SynGltA does not bypass the natural glutamate biosynthesis pathway. Instead, it rescues ΔgltA by enabling the synthesis of citrate, thereby overcoming a defect in central carbon metabolism.

**SynGltA Alters Gene Expression in E. coli.** To test the possibility that SynGltA replaces WT-GltA by performing the same enzymatic activity, we assayed the ability of purified SynGltA to catalyze the synthesis of citrate from oxaloacetate and acetyl CoA. To ensure the SynGltA protein was not contaminated by the natural citrate synthase enzyme, we purified the de novo protein following expression in ΔgltA cells. The chromatographically purified SynGltA was incubated with 13C-acetylCoA and oxaloacetate in a variety of buffers, and formation of citrate was monitored by 13C NMR. No product was detected, indicating that SynGltA does not function by direct catalysis.

**Figure 2.** The de novo protein SynGltA increases expression of prpC, which encodes methyl citrate synthase. (A) Bars show the top 15 upregulated transcripts in ΔgltA cells expressing SynGltA relative to ΔgltA cells with added glutamate. Transcripts were quantified by RNaseq. (B) The condensation of propionyl-CoA and oxaloacetate catalyzed by the enzyme methyl citrate synthase (PrpC). (C) The abundance of prpC transcripts in ΔgltA cells measured using both RNaseq and RT-qPCR. The ratio of abundance is shown for cells expressing SynGltA relative to the same cells expressing native E. coli GltA. Error bars represent SD.

In summary, both RNaseq and RT-qPCR demonstrate that the de novo protein SynGltA leads to a substantial increase in the transcription of prpC, which encodes methylcitrate synthase.

**SynGltA Alters the Metabolome of E. coli.** To probe the metabolic consequences of rescue by SynGltA, LC/MS was used to compare the metabolome of ΔgltA cells expressing SynGltA to the same cells expressing native E. coli GltA. As shown in Figure 3, expression of SynGltA caused a 1000-fold increase in the abundance of 2-methylcitrate. This result shows that not only is the PrpC enzyme overexpressed in response to SynGltA (Figure 2), but the product of its catalytic activity, 2-methylcitrate, is also synthesized abundantly (Figure 3).

**Figure 3.** Expression of SynGltA increases abundance of 2-methylcitrate. Counts from LC/MS probing ΔgltA cells expressing either native E. coli GltA (WT-GltA) or SynGltA. There is far more 2-methylcitrate in ΔgltA cells expressing SynGltA (blue bar) than in cells expressing WT GltA. Error bars represent SD.

**Methylcitrate Synthase Encoded by PrpC Has Promiscuous Catalytic Activity.** Our finding that SynGltA rescues ΔgltA by causing overexpression of PrpC suggests that the methylcitrate synthase encoded by PrpC may have promiscuous activity capable of synthesizing citrate by combining oxaloacetate with acetyl-CoA, rather than with its usual substrate, propionyl-CoA (Figures 1 and 2). This suggestion can be tested both in vitro and in vivo: The ability of the PrpC enzyme to synthesize citrate was confirmed in vitro...
by several groups who reported that methylcitrate synthase indeed catalyzes formation of citrate, albeit at a slower rate than it forms methylcitrate. 20,21

The promiscuous citrate synthase activity of PrpC is ∼100-fold slower than the dedicated citrate synthase activity of GltA. 22,23 Therefore, the low level of citrate synthase activity provided by the chromosomally expressed prpC gene is not sufficient to support growth on minimal medium. Hence, ΔgltA cells are auxotrophs.

To test if overexpression of PrpC—or any other endogenous E. coli protein—can produce enough citrate to rescue ΔgltA, we transformed the ASKA library, which contains overexpression plasmids encoding all E. coli open reading frames, 24 into ΔgltA cells and plated on minimal medium. Figure 4 shows the three ORFs that were found to rescue ΔgltA: These are GltA, PrpC, and the zinc-associated repressor ZraR. As expected, WT-GltA rescues ΔgltA cells by directly providing citrate synthase activity. We found that ΔgltA cells expressing PrpC or ZraR also grew on minimal medium. This was surprising, since Patrick et al. had performed a similar screen, and found no rescuers other than WT-GltA.25 However, in related experiments, Blank et al.26 and Guzman et al.27 searched for chromosomal mutations (rather than ORFs overexpressed from plasmids), and found that ΔgltA was rescued by mutations that enhance expression of PrpC.

E. coli PrpC Is Required for the Rescue of ΔgltA Cells by SynGltA. The experiments described above demonstrate that (i) overexpression of PrpC, an E. coli methylcitrate synthase with promiscuous activity can rescue the ΔgltA auxotroph; and (ii) the de novo protein SynGltA increases expression of PrpC. Together, these results suggest that SynGltA rescues ΔgltA cells by causing overexpression of PrpC. However, other transcripts were also increased or decreased in response to SynGltA, and the observation of enhanced transcription of PrpC does not explicitly prove it is required for the rescue phenotype. To confirm that increased expression of PrpC is essential for rescue, it is crucial to show that rescue of the ΔgltA auxotroph by SynGltA cannot occur in the absence of PrpC. Therefore, we constructed the double-deletion strain ΔgltAΔprpC, and tested whether the de novo protein could still rescue the glutamate auxotroph in this genetic background. (ΔprpC itself is not auxotrophic, so the double knockout experiment only tests for the ability of SynGltA to rescue ΔgltA.) As shown in Table 1 row 2, ΔgltAΔprpC cells expressing SynGltA fail to grow. Thus, the presence of PrpC is indeed required for auxotroph rescue by SynGltA, thereby confirming the hypothesis that SynGltA rescues ΔgltA by enhancing expression of the promiscuous methylcitrate synthase, PrpC.

Rescue of ΔgltA by SynGltA Requires Native Regulation of PrpC. The results presented above indicate that SynGltA rescues ΔgltA cells by upregulating expression of the prp operon, including PrpC. In normal cells, regulation of the prp operon is controlled by the availability of propionate. When propionate is abundant, it is transformed into 2-methylcitrate [by the promiscuous activity of constitutively expressed GltA20,21,28], which binds the transcription factor PrpR. This binding activates PrpR to stimulate transcription of the prp operon (Figure 5). To test whether upregulation of prpC by SynGltA depends on this transcriptional regulation by PrpR, we constructed the ΔgltAΔprpR double knockout, and tested whether SynGltA could rescue this double knockout. As shown in Table 1 row 3, SynGltA fails to rescue ΔgltAΔprpR, thereby demonstrating that rescue of ΔgltA by SynGltA involves the endogenous transcription factor, PrpR.

E. coli ZraR Is Not Required for Rescue of ΔgltA Cells by SynGltA. Our screen of overexpressed ORFs also revealed that a second E. coli protein, ZraR, was able to rescue ΔgltA cells. ZraR is a zinc-associated transcriptional regulator, so it was surprising that its overexpression rescued ΔgltA. To assay the relevance of ZraR to the rescue of ΔgltA by SynGltA, we constructed the ΔgltAΔzraR double knockout. As shown in Table 1 row 4, SynGltA rescues this double knockout, demonstrating that ZraR is not required for the rescue of ΔgltA cells by the SynGltA.

Although ZraR is not required for the rescue of ΔgltA by SynGltA, we were curious to understand why ZraR came up in our screen for overexpressed ORFs that rescue ΔgltA. Because ZraR is a regulatory protein, we postulated that ZraR might rescue ΔgltA by acting in place of PrpR to upregulate expression of the prp operon, including PrpC, which would rescue ΔgltA as described above. To explicitly test whether ZraR requires PrpC to rescue ΔgltA, we transformed ΔgltAΔprpC cells with a plasmid encoding ZraR and plated them on minimal medium. These cells failed to grow, indicating that ZraR in itself does not have citrate synthase activity, and requires PrpC to rescue ΔgltA. This is consistent with the hypothesis that ZraR upregulates the prp operon to rescue ΔgltA. Since ZraR and the natural regulator of the prp operon, PrpR, are both σ44-associated regulators, we asked whether ZraR can substitute for PrpR to stimulate transcription of the prp operon. To test this possibility, the ΔgltAΔprpR double knockout strain was transformed with a plasmid encoding ZraR, and plated on minimal medium. ZraR rescued ΔgltAΔprpR cells on minimal medium, supporting the hypothesis that ZraR can replace PrpR as an activator of the prp operon.

DISCUSSION

The number of possible protein sequences is so large that a collection containing merely one molecule of every 102-residue protein (the length of SynGltA) would fill a volume far larger
than the known universe. From this vast number of possibilities, nature has explored only a miniscule fraction. Recently, however, thanks to substantial technological advances, it has become possible to construct large collections of novel proteins that were never sampled by nature. With these de novo proteins in hand, one can begin to explore the intersection between non-natural sequence space and natural cells that hitherto were sustained solely by sequences evolved in nature. In the current study, we probed how novel proteins might enable cells to rebound from the deletion of the gene encoding citrate synthase at the entry point of the TCA cycle.

By screening a library of 1.5 × 10⁶ de novo designed sequences, we found one protein, SynGltA, that rescued ΔgltA (Figure 1). Although we initially assumed SynGltA would substitute for the natural GltA enzyme by catalyzing the same reaction, we found that purified SynGltA was not enzymatically active. Therefore, we hypothesized that SynGltA might provide a novel regulatory function that rescues ΔgltA. Through a series of experiments utilizing both unbiased and targeted approaches, we determined that SynGltA enhances expression of methylcitrate synthase, encoded by prpC (Figure 2). This enzyme can promiscuously perform the citrate synthase reaction, and when overexpressed, can produce sufficient citrate to sustain the growth of ΔgltA cells on minimal medium. Thus, the non-natural protein, SynGltA, exerts a life-sustaining function by altering the regulation of an alternative metabolic pathway.

Citrate synthase, encoded by gltA, catalyzes the first step of the TCA cycle in central carbon metabolism, while methylcitrate synthase, encoded by prpC, catalyzes a reaction in propionate metabolism, and is not required for growth under most conditions. Yet these two enzymes likely share evolutionary history: At the amino acid sequence level, they are 30% identical. Moreover, they share catalytic activities: In their reaction, we found that purification of SynGltA did not enhance expression of propionate catabolism, and SynGltA did not enhance expression of lipid catabolic genes. Therefore, it seems unlikely that SynGltA functions by increasing the concentration of these metabolites.

Irrespective of the molecular details by which SynGltA stimulates the prp operon, the results described herein demonstrate that a lab-made protein, unrelated to naturally occurring sequences, can provide a life-sustaining function by altering the regulation of natural gene expression. As such, SynGltA joins SynSerB3, another binary patterned protein, which was shown previously to rescue the serine auxotrophy in ΔserB cells by increasing expression of HisB, which encodes a promiscuous phosphatase in the histidine biosynthetic pathway. Together, these two de novo proteins represent the first members of a toolbox of non-natural proteins that can be used to rewire gene regulation and alter metabolic outcomes in pathways ranging from amino acid auxotrophy to central carbon and energy metabolism.

### METHODS

**Reagents.** Oligonucleotide primers were purchased from Integrated DNA Technologies. Cultures were grown in LB medium or M9-glucose minimal medium (1 × M9 salts, 0.4% glucose, 2 mM MgSO₄, 100 μM CaCl₂). Selective agents and inducers were used at the concentrations listed: kanamycin (kan, 30 μg/mL), chloramphenicol (30 μg/mL), IPTG (50 μM), Acetyl-CoA, (1,2)-13C-acetyl-CoA, and oxaloacetate were purchased from Sigma.

**Strains.** Keio parent cells are strain BW25113 and ΔgltA cells are [ΔgltA770::kan] in the BW25113 background; both strains were obtained from the E. coli Genetic Stock Center (http://cgsc.biology.yale.edu/). ΔgltAΔprpC, ΔgltAΔprpR, and ΔgltAΔzraR cells are [ΔprpR772, ΔgltA770::kan], [ΔprpR772, ΔgltA770::kan], and [ΔzraR775, ΔgltA770::kan] in the BW25113 background, respectively; these strains were made using standard P1 transduction methods. Briefly, using the ΔgltAΔprpC strain as an example, P1 virus lysate was made using the donor strain ΔgltA::kan. The kanamycin cassette in ΔprpC::kan cells was excised using plasmid pCP20. P1ΔprpC cells were then transduced with ΔgltA::kan P1vir lysate, and plated on selective medium. Locus-specific primers were used to confirm the strain and are listed in Supporting Table S1.

**DNA Methods.** Transformations were done according to standard protocols. After transformation, cells recovered in SOC for 1 h while shaking. For rescue experiments, cells were then washed twice with 1 × M9 before plating. LacZ, WT-GltA, SynGltA, and its mutants were expressed from vector pGltA, a derivative of pCA24N. WT-PrpC and ZraR were expressed from vector pCA24N. Standard Quikchange PCR was used to make point mutations and the nucleotide insert mutants.

**Protein Expression and Purification.** ΔgltA cells containing plasmid-born SynGltA or control plasmids were used to inoculate a 10 mL starter culture overnight. The starter
culture was used to inoculate 1 L of LB medium, and the culture was induced with 50 μM IPTG when the OD_{600} ~ 0.5 after growing at 37 °C. The cultures were grown 8 additional hours at 18 °C to allow for protein expression. Cells were harvested, resuspended in 50 mM phosphate buffer pH 7.2 containing 50 mM NaCl, and lysed using an Emulsiflex (Avestin). Proteins were purified using a cation exchange column (GE); fractions were eluted with a gradient of 0–100% 50 mM phosphate buffer at pH 7.2 containing 1.5 M NaCl. Protein fractions containing SynGltA were further purified by a HisTrap FF crude column (GE) equilibrated with 200 mM phosphate buffer pH 7.2 containing 500 mM NaCl. Fractions were eluted in the same buffer containing an additional 500 mM imidazole. Finally, eluates were loaded onto a HiLoad 16/60 Superdex 75 pg column (GE) running in 100 mM phosphate buffer pH 7.2 containing 100 mM NaCl.

Enzymatic Assay. NMR samples contained 450 μL of 50 mM Tris +300 mM NaCl pH 7.5 buffer, 40 μL of D_2O, 40 μL of purified protein sample, 14 μL of 30 mM (1,2)-13C-Acetyl-CoA, and 70 μL of oxaloacetate. 13C NMR was performed on a 500 MHz (125 MHz 13C) Bruker Avance III instrument with a cryoprobe with a 3 s relaxation delay for qualitative experiments, and a 14.3 s relaxation time for quantitative experiments.

RNAseq and Differential Expression Analysis. Cells were grown in an overnight culture and washed twice with 1× M9 salts. They were then used to inoculate 10 mL minimal M9-glucose medium, which grew until cells reached mid log. Grown cells were quickly vacuum-porated through a 0.5 μm pore size, Millipore. Membranes were flipped into 60 mm Petri dishes containing 1.2 mL of cold (−20 °C) extraction solvent (40:40:20 methanol/acetonitrile/H_2O, HPLC grade) and metabolites were extracted at −20 °C for 15 min. Cold extraction solvent was used to wash the membranes in the dish. The extracts were collected in microcentrifuge tubes and centrifuged at 4 °C to remove cellular debris. A portion of the supernatant (300 μL) was transferred to a new tube and dried under N_2 gas. Metabolites were resuspended in HPLC-grade H_2O and analyzed by reversed-phase ion-pairing liquid chromatography coupled to a stand-alone Orbitrap mass spectrometer by negative-ion mode electrospray ionization. Metabolite peaks were quantified using MAVEN, and normalized to cell density (OD_{600}). Three biological replicates for each sample were analyzed.

**REFERENCES**


**ACKNOWLEDGMENTS**

Research was funded by NSF Grant MCB-1050510, and K.D. was supported by a NSF Graduate Research Fellowship. We thank Betsy Smith for plasmids encoding frameshift and stop codons; the microarray core facility at the Lewis-Sigler Institute for Integrative Genomics at Princeton for RNA sequencing; and Xin Teng and Prof. Josh Rabinowitz for performing metabolite LC/MS.
clones of Escherichia coli ASKA library (A complete Set of E. coli K-12 ORF archive): Unique resources for biological research. DNA Res. 12 (5), 291–299.


