ACS Synthetic Biology

Article

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

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

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.6b00336 • Publication Date (Web): 05 Jan 2017

ACS Publications

ACS Synthetic Biology is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

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

Katie Digianantonio[1], Maria Korolev[2], Michael H. Hecht*

Department of Chemistry

Princeton University

Princeton, NJ 08540, USA

*Corresponding Author: Michael H. Hecht, hecht@princeton.edu , (609) 258-2901

[1]Currently at the Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT

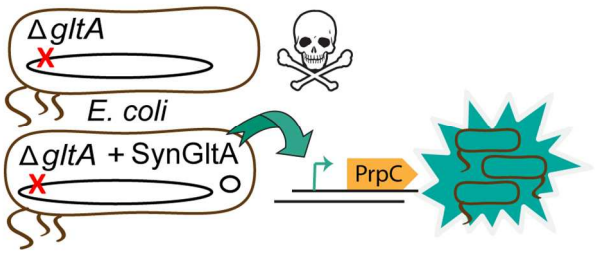
[2]Currently at the Dept. of Chemistry, University of Florida, Gainesville, FL

SUMMARY/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 Δ*gltA*. 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 acts by an unknown mechanism to enhance 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. These results 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

For Table of Contents Use only:



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

Katie Digianantonio, Maria Korolev, Michael H. Hecht

INTRODUCTION

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* [6]. 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 non-polar, 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 (Fig. 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 for generating energy, this deletion is not lethal to cells growing in rich medium. In minimal medium, however, deletion of *gltA* is lethal. This is because intermediates

in the TCA cycle serve as precursors for the biosynthesis of essential metabolites (Fig. 1A). In particular, the TCA intermediate α-ketoglutarate is a precursor for the synthesis of glutamate, and cells without a functional TCA cycle cannot grow in the absence of glutamate. Indeed the gene encoding citrate synthase is named *gltA* because mutations in this gene are glutamate auxotrophs [19].

Because of this glutamate auxotrophy, one can search for novel proteins that compensate for the deletion of citrate synthase -- the entry point to the TCA cycle -- by screening libraries of sequences for those that enable the growth of Δ*gltA* cells on minimal medium.

Previously, we reported a series of life-or-death selections that searched a library of 1.5 x 10^6 binary patterned sequences for proteins that rescued the growth of various auxotrophs on minimal medium. While most strains were not rescued, we found four auxotrophs that were reproducibly rescued by *de novo* proteins from our library [6]. Several binary patterned proteins rescued Δ*fes,* Δ*ilvA,* and Δ*serB*. These genes encode proteins involved in the assimilation of iron, and the biosynthesis of isoleucine and serine, respectively. However despite extensive searches, we found only one sequence that rescued the deletion of citrate synthase in Δ*gltA* strains. We named this *de novo* sequence SynGltA. Prior to the current study, the mechanism by which SynGltA rescued Δ*gltA* was not known.

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 Δ*gltA* cells as depicted in Fig. 1B. As a positive control, Δ*gltA* cells were transformed with a plasmid encoding the native *E. coli* GltA (WT-GltA). These cells formed visible colonies on minimal medium in two days. The negative control, Δ*gltA* 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, Δ*gltA* cells transformed with a plasmid encoding SynGltA grow in four days (Fig. 1B, Table I row 1).

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 Δ*gltA* 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 Δ*gltA* cells.

In principle, the SynGltA protein could rescue the glutamate auxotrophy either by enabling the synthesis of citrate (the reaction deleted in Δ*gltA*), 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 Δ*icd*, which encodes isocitrate dehydrogenase [6] (Fig. 1A). Thus, SynGltA does not bypass the natural glutamate biosynthesis pathway. Instead, it rescues Δ*gltA* by enabling the synthesis of citrate, thereby overcoming a defect in central carbon metabolism.

Purified SynGltA protein has no detectable citrate synthase activity.

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 Δ*gltA* cells. The chromatographically purified SynGltA was incubated with ^{13}C-acetylCoA and

oxaloacetate in a variety of buffers, and formation of citrate was monitored by ^{13}C-NMR. No product was detected, indicating that SynGltA does not function by direct catalysis.

SynGltA alters gene expression in *E. coli*.

The inability of purified SynGltA to function as an enzyme suggests that the *de novo* protein functions by an indirect mechanism, presumably involving endogenous *E. coli* genes and proteins. To determine if SynGltA alters the expression of endogenous genes, the entire transcriptome was assayed using quantitative RNA sequencing (RNAseq). To ensure these assays were relevant to the mechanism of auxotroph rescue, RNAseq was performed on Δ*gltA* cells grown in minimal medium. This experiment allowed us to compare cells that relied on SynGltA for growth to the same cells relying on exogenous glutamate added to the minimal medium. (Assays were also performed in the pseudo-while type strain BW25113, and qualitatively similar results were observed. *See Dataset*.)

Analysis of the RNAseq data revealed many genes with altered expression: 343 genes were overexpressed, and 240 genes were underexpressed in Δ*gltA* cells expressing SynGltA relative to the same cells not expressing SynGltA. Figure 2A shows that the most significantly overexpressed genes encode proteins involved in purine biosynthesis and propionate metabolism. This latter class of genes is encoded by the *prp* operon. These are highlighted red in Fig. 2A, and will be discussed below.

In contrast, genes that were underexpressed in SynGltA cells relative to cells supplemented with added glutamate include those involved in glutamate metabolism and chemotaxis. Indeed, it is not surprising that cells grown in media containing added glutamate would express these genes at higher levels.

The RNAseq data demonstrate that expression of SynGltA has a dramatic effect on the expression of endogenous *E. coli* genes. While RNAseq probes the entire transcriptome in one experiment, it does not indicate which changes are responsible for the observed phenotype (growth on minimal). However, the observation that the *prp* operon was turned up more that 40-fold led us to hypothesize that overexpression of *prpC*, which encodes methylcitrate synthase (Fig. 2B), might rescue the deletion of citrate synthase. To test this hypothesis, we performed several orthogonal experiments, as described below.

First, to validate that expression of SynGltA indeed upregulates *prpC*, we performed RT-qPCR on *ΔgltA* cells expressing SynGltA and compared these to the same cells expressing wildtype GltA. As shown in Fig 2C, RT-qPCR showed that SynGltA increases expression of *prpC* by approximately 1000-fold.

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 Δ*gltA* 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 (Fig. 2), but the product of its catalytic activity, 2-methylcitrate, is also synthesized abundantly (Fig. 3).

Methylcitrate synthase encoded by PrpC has promiscuous catalytic activity.

Our finding that SynGltA rescues Δ*gltA* 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 (Figs. 1 & 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, Δ*gltA* cells are auxotrophs.

To test if overexpression of PrpC – or any other endogenous *E. coli* protein – can produce enough citrate to rescue Δ*gltA* , we transformed the ASKA library, which contains overexpression plasmids encoding all *E.coli* open reading frames[24], into Δ*gltA* cells and plated on minimal medium. Figure 4 shows the three ORFs that were found to rescue Δ*gltA:* These are GltA, PrpC, and the zinc-associated repressor ZraR. As expected, WT- GltA rescues Δ*gltA* cells by directly providing citrate synthase activity. We found that Δ*gltA* 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 Δ*gltA* was rescued by mutations that enhance expression of PrpC.

E. coli PrpC is required for the rescue of Δ*gltA* cells by SynGltA

The experiments described above demonstrate that (i) overexpression of PrpC, an *E. coli* methylcitrate synthase with promiscuous activity, can rescue the Δ*gltA* auxotroph; and (ii) the *de novo* protein SynGltA increases expression of PrpC. Together, these results suggest that SynGltA rescues Δ*gltA* 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 Δ*gltA* auxotroph by SynGltA cannot occur in the absence of PrpC. Therefore, we constructed the double-deletion strain Δ*gltA*Δ*prpC*, and tested whether the *de novo* protein could still rescue the glutamate auxotroph in this genetic background. (Δ*prpC* itself is not auxotrophic, so the double knockout experiment only tests for the ability of SynGltA to rescue Δ*gltA*.) As shown in Table I row 2, Δ*gltA*Δ*prpC* 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 Δ*gltA* by enhancing expression of the promiscuous methylcitrate synthase, PrpC.

Rescue of Δ*gltA* by SynGltA requires native regulation of PrpC.

The results presented above indicate that SynGltA rescues Δ*gltA* 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 GltA [20-21, 28]], which binds the transcription factor PrpR. This binding activates PrpR to stimulate transcription of the *prp* operon (Fig. 5). To test whether upregulation of *prpC* by SynGltA depends on this transcriptional regulation by PrpR, we constructed the Δ*gltA*Δ*prpR* double knockout, and tested whether SynGltA could rescue this double knockout. As shown in Table I row 3, SynGltA *fails* to rescue Δ*gltA*Δ*prpR*, thereby demonstrating that rescue of Δ*gltA* by SynGltA involves the endogenous transcription factor, PrpR.

E. coli ZraR is not required for rescue of Δ*gltA* cells by SynGltA

Our screen of overexpressed ORFs also revealed that a second E. coli protein, ZraR, was able to rescue Δ*gltA* cells. ZraR is a zinc-associated transcriptional regulator, so it was surprising that its overexpression rescued Δ*gltA*. To assay the relevance of ZraR to the rescue of Δ*gltA* by SynGltA, we constructed the Δ*gltA*Δ*zraR* double knockout. As shown in Table I row 4,

SynGltA rescues this double knockout, demonstrating that ZraR is *not* required for the rescue of Δ*gltA* 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 Δ*gltA* by acting in place of PrpR to upregulate expression of the *prp* operon, including PrpC, which would rescue Δ*gltA* as described above. To explicitly test whether ZraR requires PrpC to rescue Δ*gltA*, we transformed Δ*gltA*Δ*prpC* 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 Δ*gltA*. This is consistent with the hypothesis that ZraR upregulates the *prp* operon to rescue Δ*gltA*. Since ZraR and the natural regulator of the *prp* operon, PrpR, are both σ^{54} associated regulators, we asked whether ZraR can substitute for PrpR to stimulate transcription of the *prp* operon. To test this possibility, the Δ*gltA*Δ*prpR* double knockout strain was transformed with a plasmid encoding ZraR, and plated on minimal medium. ZraR rescued Δ*gltA*Δ*prpR* 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 x10^6 *de novo* designed sequences, we found one protein, SynGltA, that rescued Δ*gltA* (Fig. 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 Δ*gltA*. 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 Δ*gltA* 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 primary activities, GltA combines oxaloacetate with acetyl-CoA to make citrate, while PrpC combines oxaloacetate with propionyl-CoA to make 2-methylcitrate. Both enzymes have weak promiscuous activity for the alternative substrate[20-21, 28]. The catalytic promiscuity demonstrated by both enzymes hints that they may have evolved from an ancestral enzyme with dual specificity, which ultimately diverged into two distinct enzymes with specificities favoring one or the other metabolic pathway.

The regulation of the two enzymes has also diverged. Consistent with its role in central metabolism, GltA is constitutively expressed in growing cells. However, PrpC is expressed only in the presence of propionate. The availability of propionate is sensed by the regulator protein, PrpR, which binds 2-methylcitrate, a downstream product of propionate metabolism. It is not clear why PrpR senses a product of propionate metabolism, rather than propionate itself, especially considering that the enzymes that transform propionate into 2-methylcitrate, PrpE and PrpC, are not basally expressed. However, as mentioned above, the weak promiscuous activity of GltA can produce 2-methylcitrate. Therefore, in wild-type strains, the presence of propionate leads to low levels of 2-methylcitrate, which activates PrpR thereby upregulating the *prp* operon, which encodes a set of enzymes dedicated to propionate catabolism. It has been proposed that 2-methylcitrate is toxic to the cell, and sensing toxic concentrations of this metabolite rather than propionate itself may give the cell more sensitive control over regulation [29-30].

This divergent regulation between GltA and PrpC is key in allowing the *de novo* protein SynGltA to enable cells to rebound from a knockout of the TCA cycle: SynGltA rescues Δ*gltA* by increasing transcription of the *prp* operon *only* when the PrpR regulatory protein is present; but fails to rescue the double deletion, Δ*gltA*Δ*prpR*. These results demonstrate that the PrpR activator protein is involved in the SynGltA mediated rescue. At this point, however, the molecular details of this involvement are unclear. One possibility would be that SynGltA activates PrpR in a manner analogous to the activation by 2-methylcitrate. Alternatively, it is also possible that SynGltA increases the pool of cellular propionate or 2-methylcitrate thereby increasing the transcriptional activity of PrpR. However, propionate and 2-methylcitrate are products of lipid 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 Δ*serB* cells by increasing expression of HisB, which encodes a promiscuous phosphatase in the histidine biosynthetic pathway [5]. Together, these two *de novo* proteins represent the first members of a toolbox of non-natural proteins that can be used 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 x M9 salts, 0.4% glucose, 2 mM $MgSO_4$, 100 μM $CaCl_2$). Selective agents and inducers were used at the concentrations listed: kanamycin (kan, 30 μg/mL), chloramphenicol (30 μg/mL), IPTG (50μM). Acetyl-CoA, (1,2)-^{13}C-acetyl-CoA, and oxaloacetate were purchased from Sigma.

Strains. Keio parent cells are strain BW25113 and Δ*gltA* cells are [ΔgltA770::kan] in the BW25113 background; both strains were obtained from the *E. coli* Genetic Stock Center (http://cgsc.biology.yale.edu/). Δ*gltA*Δ*prpC*, Δ*gltA*Δ*prpR*, and Δ*gltA*Δ*zraR* cells are [ΔprpC774, ΔgltA770::kan], [ΔprpR772, ΔgltA770::kan], and [ΔzraR775, ΔgltA770::kan] in the BW25113 background respectively; these strains were made using standard P1 transduction methods [31]. Briefly, using the Δ*gltA*Δ*prpC* strain as an example, P1 virus lysate was made using the donor strain Δ*gltA::kan.* The kanamycin cassette in Δ*prpC::kan* cells was excised using plasmid pCP20 [32]. Δ*prpC* cells were then transduced with Δ*gltA::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 [33]. After transformation, cells recovered in SOC for one hour while shaking. For rescue experiments, cells were then washed twice with 1xM9 before plating. LacZ, WT-GltA, SynGltA and its mutants were expressed from vector p3Glar, 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. Δ*gltA* cells containing plasmid-born SynGltA or control plasmids were used to inoculate a 10 mL starter culture overnight. The starter culture was used to inoculated 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 butter 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 additional 500 mM imidazole. Finally, eluates were loaded onto a HiLoad 16/600 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 50 mM Tris + 300 mM NaCl pH 7.5 buffer, 40 µL D_2O, 40 µL purified protein sample, 14 µL 30 mM (1,2)-^{13}C-Acetyl-CoA, and 70 µL oxaloacetate. ^{13}C NMR was performed on a 500 MHz (125 MHz ^{13}C) Bruker Avance III instrument with a cryoprobe with a 3 second relaxation delay for qualitative experiments, and a 14.3 second relaxation time for quantitative experiments.

RNAseq and Differential Expression Analysis. Cells were grown in an overnight culture and washed twice with 1xM9 salts. They were then used to inoculate 10 mL minimal M9-glucose medium, which grew until cells reached mid-log. Grown cells were incubated with 2 volumes of RNAlater reagent (Thermo Fisher Scientific) and stored at -20° until all samples could be extracted for RNA at the same time. Total RNA was prepared using an RNeasy Mini Kit (Qiagen). RNA quality was assayed using a NanoDrop and Agilent 2100 Bioanalyzer. The Ribo-Zero rRNA removal kit for bacteria (Illumina) was used to extract ribosomal RNA. The RNAseq library was prepared using the Illumina TruSeq protocol, and an Illumnia HiSeq instrument was used for sequencing. The *E.coli* genome (NCBI reference number NC_000913.2) was used to map reads using TopHat2 [34] and read counts for each gene were obtained using htseq-count[35] on a Galaxy server [36-38]. DEseq2 [39] was used to perform the differential expression analysis. The design formula for pairwise differential expression analysis was ~batch + treatment. Using p-value cutoff $p < 0.05$ and a magnitude cutoff for the $\log_2$(Fold Change) $|\log_2 \text{FoldChange}| > 2$ we analyzed genes using DAVID (http://david.abcc.ncifcrf.gov/).

RT-qPCR. RNA was isolated as described above, and the TURBO DNA-free kit (Ambion) was used to remove DNA. cDNA was prepared from 2 µg RNA using SuperScript First-Strand Synthesis System (Life Technologies). Then, RT-qPCR reactions were made containing 10 µL Power SYBR Green PCR master mix (Applied Biosystems), 2.4 µL of 10 µM stocks of each amplicon primer, and 4 µL cDNA in a final volume of 20 µL. Primers used are in Supporting Table S1. An Applied Biosystems ABI 7900 instrument was used for qPCR with conditions as follows: 50°C 2 min, 95°C 10 min, [95°C 15 sec, 60°C 1 min] x40, 95°C 15 sec, 60°C 15 sec, 95°C 15 sec. Each sample had three technical replicates, and one plate contained three samples grown from three separate cultures. Relative expression levels using the $2^{-\Delta\Delta Cq}$ method were determined after normalizing to the *rrs* mRNA [40].

LC/MS Metabolomic Analysis. Cells were grown in minimal medium exactly as for RNA transcript profiling. To extract cellular metabolites, 3.0 mL of cells at OD_{600} = 0.3-0.5 were quickly vacuum-filtered onto nylon membranes (0.45-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 minutes. 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 (500µL) 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 [41]. Metabolite peaks were quantified using MAVEN [42], and normalized to cell density (OD_{600}). Three biological replicates for each sample were analyzed.

Supporting Information. Primers used for strain checking and RT-qPCR, RNAseq dataset

Author Information. KD is currently at the Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT. MK is currently at the Dept. of Chemistry, University of Florida, Gainesville, FL.

Author Contribution. KD and MK performed research, KD and MH wrote manuscript, all authors contributed ideas, discussed results, and commented on manuscript.

Acknowledgements. Research was funded by NSF grant MCB-1050510, and KD 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.

REFERENCES

1. Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D., Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** *302* (5649), 1364-1368.

2. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-U4.

3. Joh, N. H.; Wang, T.; Bhate, M. P.; Acharya, R.; Wu, Y.; Grabe, M.; Hong, M.; Grigoryan, G.; DeGrado, W. F., De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science* **2014,** *346* (6216), 1520-1524.

4. Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A. D.; Hilvert, D., Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. *Protein Eng Des Sel* **2016,** *29* (9), 355-66.

5. Digianantonio, K. M.; Hecht, M. H., A protein constructed de novo enables cell growth by altering gene regulation. *Proc Natl Acad Sci U S A* **2016,** *113* (9), 2400-5.

6. Fisher, M. A.; McKinley, K. L.; Bradley, L. H.; Viola, S. R.; Hecht, M. H., De Novo Designed Proteins from a Library of Artificial Sequences Function in Escherichia Coli and Enable Cell Growth. *Plos One* **2011,** *6* (1).

7. Hoegler, K. J.; Hecht, M. H., A De Novo Protein Confers Copper Resistance in Escherichia Coli. *Protein Sci* **2016**.

8. Smith, B. A.; Mularz, A. E.; Hecht, M. H., Divergent evolution of a bifunctional de novo protein. *Protein Science* **2015,** *24* (2), 246-252.

9. Kamtekar, S.; Schiffer, J. M.; Xiong, H. Y.; Babik, J. M.; Hecht, M. H., Protein Design by Binary Patterning Of Polar and Nonpolar Amino-acids. *Science* **1993,** *262* (5140), 1680-1685.

10. Bradley, L. H.; Kleiner, R. E.; Wang, A. F.; Hecht, M. H.; Wood, D. W., An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences. *Protein Engineering Design & Selection* **2005,** *18* (4), 201-207.

11. Hecht, M. H.; Das, A.; Go, A.; Bradley, L. H.; Wei, Y. N., De novo proteins from designed combinatorial libraries. *Protein Science* **2004,** *13* (7), 1711-1723.

12. Wei, Y. N.; Liu, T.; Sazinsky, S. L.; Moffet, D. A.; Pelczer, I.; Hecht, M. H., Stably folded de novo proteins from a designed combinatorial library. *Protein Science* **2003,** *12* (1), 92-102.

13. Arai, R.; Kobayashi, N.; Kimura, A.; Sato, T.; Matsuo, K.; Wang, A. F.; Platt, J. M.; Bradley, L. H.; Hecht, M. H., Domain-Swapped Dimeric Structure of a Stable and Functional De

Novo Four-Helix Bundle Protein, WA20. *Journal of Physical Chemistry B* **2012,** *116* (23), 6789-6797.

14. Wei, Y. N.; Kim, S.; Fela, D.; Baum, J.; Hecht, M. H., Solution structure of a de novo protein from a designed combinatorial library. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (23), 13270-13273.

15. Murphy, G. S.; Greisman, J. B.; Hecht, M. H., De Novo Proteins with Life-Sustaining Functions Are Structurally Dynamic. *Journal of Molecular Biology* **2016,** *428* (2), 399-411.

16. Go, A.; Kim, S.; Baum, J.; Hecht, M. H., Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. *Protein Science* **2008,** *17* (5), 821-832.

17. Cherny, I.; Korolev, M.; Koehler, A. N.; Hecht, M. H., Proteins from an Unevolved Library of de novo Designed Sequences Bind a Range of Small Molecules. *Acs Synthetic Biology* **2012,** *1* (4), 130-138.

18. Patel, S. C.; Bradley, L. H.; Jinadasa, S. P.; Hecht, M. H., Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. *Protein Science* **2009,** *18* (7), 1388-1400.

19. Davis, B. D.; Gilvarg, C., The role of the tricarboxylic acid cycle in acetate oxidation in Escherichia coli. *J Biol Chem* **1956,** *222* (1), 307-19.

20. Gerike, U.; Hough, D. W.; Russell, N. J.; Dyall-Smith, M. L.; Danson, M. J., Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships. *Microbiology-Uk* **1998,** *144*, 929-935.

21. Textor, S.; Wendisch, V. F.; De Graaf, A. A.; Müller, U.; Linder, M. I.; Linder, D.; Buckel, W., Propionate oxidation in Escherichia coli: evidence for operation of a methylcitrate cycle in bacteria. *Arch Microbiol* **1997,** *168* (5), 428-36.

22. Anderson, D. H.; Duckworth, H. W., In vitro mutagenesis of Escherichia coli citrate synthase to clarify the locations of ligand binding sites. *J Biol Chem* **1988,** *263* (5), 2163-9.

23. Horswill, A. R.; Escalante-Semerena, J. C., Salmonella typhimurium LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J Bacteriol* **1999,** *181* (18), 5615-23.

24. Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H., Complete set of ORF clones of Escherichia coli ASKA library (A complete Set of E. coli K-12 ORF archive): Unique resources for biological research. *DNA Research* **2005,** *12* (5), 291-299.

25. Patrick, W. M.; Quandt, E. M.; Swartzlander, D. B.; Matsumura, I., Multicopy suppression underpins metabolic evolvability. *Molecular Biology and Evolution* **2007,** *24* (12), 2716-2722.

26. Blank, D.; Wolf, L.; Ackermann, M.; Silander, O. K., The predictability of molecular evolution during functional innovation. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (8), 3044-3049.

27. Guzman, G. I.; Utrilla, J.; Nurk, S.; Brunk, E.; Monk, J. M.; Ebrahim, A.; Palsson, B. O.; Feist, A. M., Model-driven discovery of underground metabolic functions in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **2015,** *112* (3), 929-934.

28. Man, W. J.; Li, Y.; O'Connor, C. D.; Wilton, D. C., The binding of propionyl-CoA and carboxymethyl-CoA to Escherichia coli citrate synthase. *Biochim Biophys Acta* **1995,** *1250* (1), 69-75.

29. Horswill, A. R.; Dudding, A. R.; Escalante-Semerena, J. C., Studies of propionate toxicity in Salmonella enterica identify 2-methylcitrate as a potent inhibitor of cell growth. *J Biol Chem* **2001,** *276* (22), 19094-101.

30. Palacios, S.; Escalante-Semerena, J. C., 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of Salmonella enterica by the PrpR protein. *Microbiology* **2004,** *150* (Pt 11), 3877-87.

31. Miller, J. H., *Experiments in molecular genetics*. Cold Spring Harbor Laboratory: 1972.

32. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (12), 6640-6645.

33. Sambrook, J.; Russell, D. W., Molecular cloning: A laboratory manual. *Molecular cloning: A laboratory manual* **2001**.

34. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S. L., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **2013,** *14* (4).

35. Anders, S.; Pyl, P. T.; Huber, W., HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015,** *31* (2), 166-169.

36. Blankenberg, D.; Von Kuster, G.; Coraor, N.; Ananda, G.; Lazarus, R.; Mangan, M.; Nekrutenko, A.; Taylor, J., Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2010,** *Chapter 19*, Unit 19.10.1-21.

37. Giardine, B.; Riemer, C.; Hardison, R. C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; Miller, W.; Kent, W. J.; Nekrutenko, A., Galaxy: A platform for interactive large-scale genome analysis. *Genome Research* **2005,** *15* (10), 1451-1455.

38. Goecks, J.; Nekrutenko, A.; Taylor, J.; Galaxy, T., Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **2010,** *11* (8).

39. Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **2014,** *15* (12).

40. Livak, K. J.; Schmittgen, T. D., Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **2001,** *25* (4), 402-408.

41. Lu, W. Y.; Clasquin, M. F.; Melamud, E.; Amador-Noguez, D.; Caudy, A. A.; Rabinowitz, J. D., Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer. *Analytical Chemistry* **2010,** *82* (8), 3212-3221.

42. Melamud, E.; Vastag, L.; Rabinowitz, J. D., Metabolomic Analysis and Visualization Engine for LC-MS Data. *Analytical Chemistry* **2010,** *82* (23), 9818-9826.

FIGURES AND LEGENDS

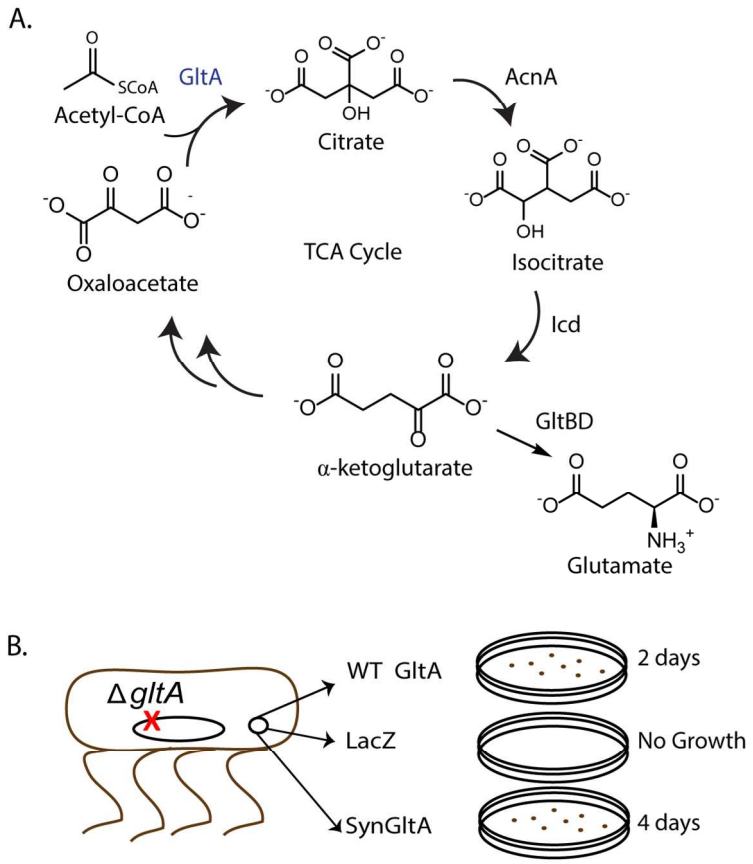


Figure 1: The *de novo* protein SynGltA enables the growth of Δ*gltA* cells on minimal medium. **(A)** The TCA cycle showing the condensation of acetyl-CoA and oxaloacetate performed by citrate synthase (GltA). Citrate is processed into α-ketoglutarate, which is a precursor of glutamate. **(B)** An overview of the auxotroph screen: A strain of *E. coli* in which *gltA* is deleted cannot grow on minimal medium. A plasmid encoding LacZ (negative control) fails to support growth, whereas both the native *E. coli* GltA (positive control) and the *de novo* protein SynGltA support growth of colonies on minimal medium

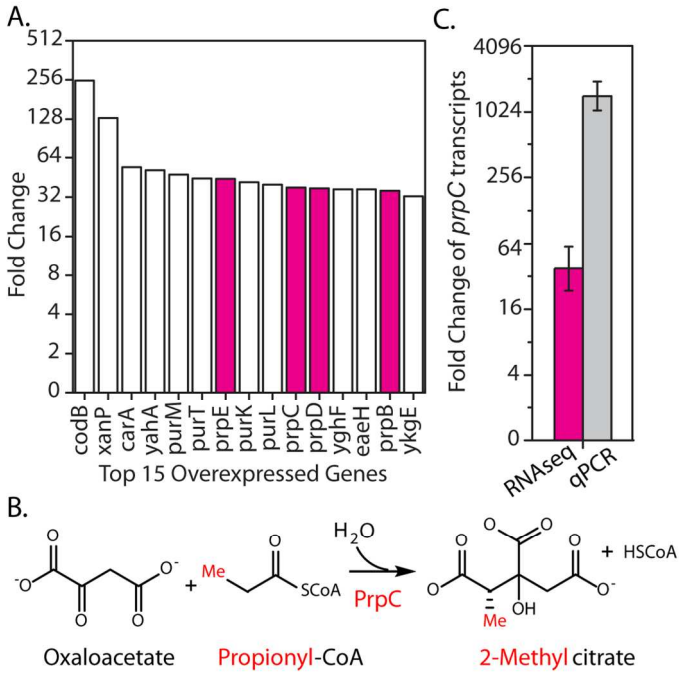


Figure 2: The *de novo* protein SynGltA increases expression of *prpC*, which encodes methyl citrate synthase. **(A)** Bars show the top 15 upregulated transcripts in Δ*gltA* cells expressing SynGltA relative to Δ*gltA* cells with added glutamate. Transcripts were quantified by RNAseq. **(B)** The condensation of propionyl-CoA and oxaloacetate catalyzed by the enzyme methyl citrate synthase (PrpC.) **(C)** The abundance of *prpC* transcripts in Δ*gltA* 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.

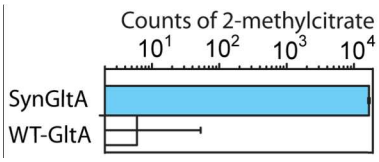


Figure 3: Expression of SynGltA increases abundance of 2-Methylcitrate. Counts from LC/MS probing Δ*gltA* cells expressing either native *E. coli* GltA (WT-GltA) or SynGltA. There is far more 2-methylcitrate in Δ*gltA* cells expressing SynGltA (blue bar) than in cells expressing WT GltA. Error bars represent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

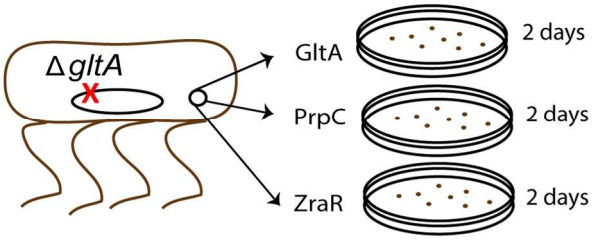


Figure 4: Three overexpressed *E coli* ORFS from the ASKA collection can rescue Δ*gltA* cells. GltA is citrate synthase, PrpC is methylcitrate synthase, and ZraR is a zinc-response associated protein.

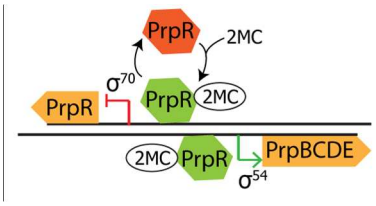


Figure 5: **Native regulation of the *prp* operon.** When PrpR is bound to 2-methylcitrate (2MC), it is activated to stimulate transcription of the *prp* operon including *prpC,* which encodes methyl citrate synthase. In addition, activated PrpR is an autorepressor.

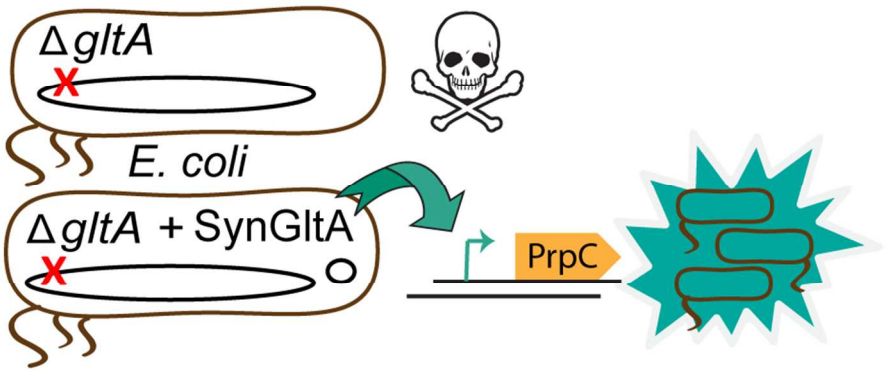
Table I: Growth of auxotrophic *E. coli* on minimal media					
Strain	Plasmid Expressing				
	LacZ	WT GltA	WT PrpC	WT ZraR	SynGltA
Δ*gltA*	X	2 days	3 days	3 days	4 days
Δ*gltA*Δ*prpC*	X	2 days	2 days	X	X
Δ*gltA*Δ*prpR*	X	2 days	2 days	2 days	X
Δ*gltA*Δ*zraR*	X	2 days	2 days	2 days	3 days

X = no growth

The *de novo* protein SynGltA rescues the Δ*gltA* 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.

Supporting Table S1: Primers used in this study	
Locus specific primers	
gltA for	GCGAGCCAAATAAAAAACG
gltA rev	GCTGATAATTTGAGCTGTTCTATTC
zraR for	CAACACGGTGGTACAATTCAG
zraR rev	ACGCGCCTATGCCTTAATG
prpR for	CAGTTAACGTTTCAGGCAATG
prpR rev	CTGAATCCTATGTAAACATCTCC
RT-qPCR primers	
rrs for	CGATGCAACGCGAAGAACCT
rrs rev	CCGGACCGCTGGCAACAAA
prpC for	ACCGAGATGTTCACACCACT
prpC rev	CATAATTGGCGGAAGGACGG
prpR for	TTTAAGTGTGGAACCGACGC
prpR rev	TGCCTGTTGTGGTGTCAGTA
zraR for	CCATTTTCTGCAGCGCTTTG
zraR rev	TATTTCCCGGCCAGTCGTAA

Primers used in this study for RT-qPCR and checking strains.



for TOC use only

88x34mm (300 x 300 DPI)

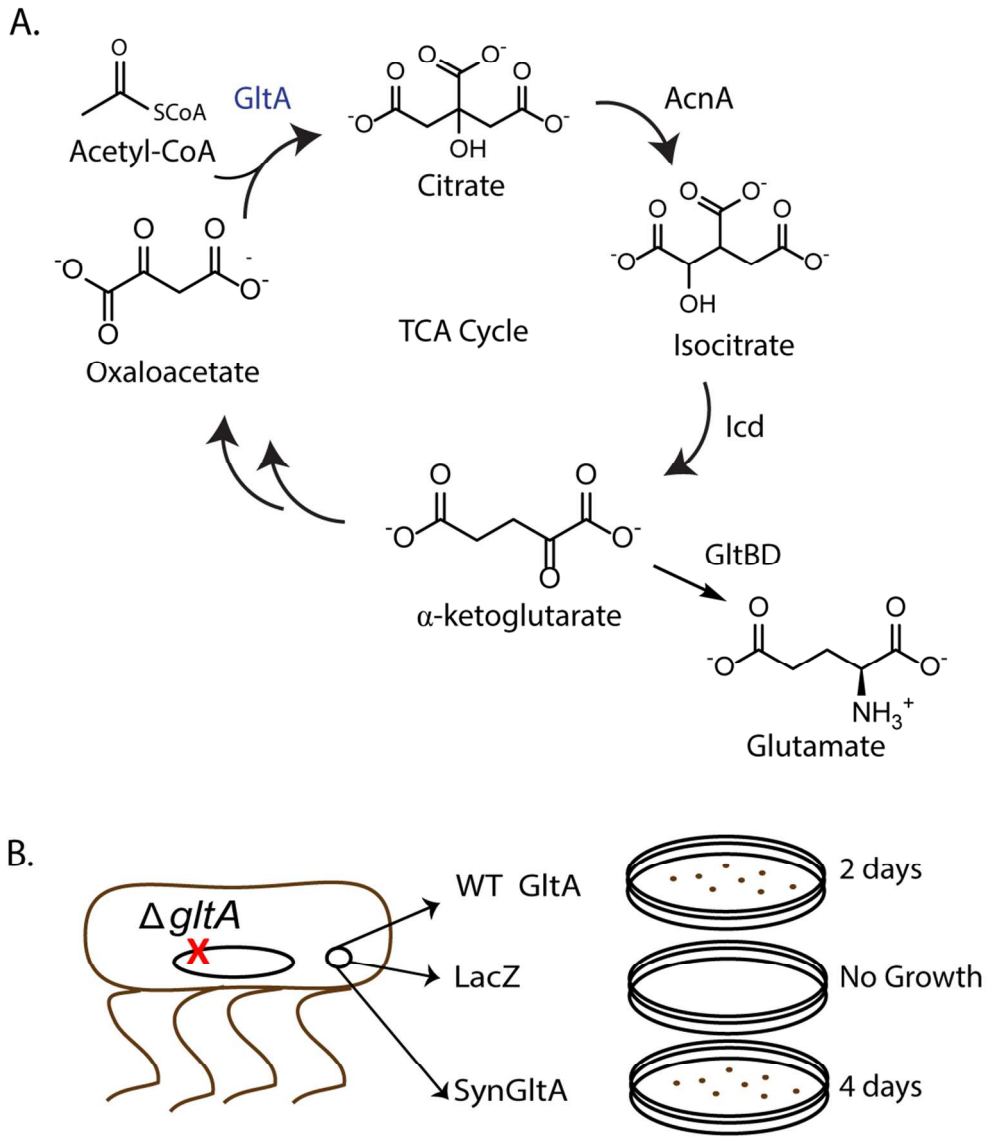


Figure 1: The de novo protein SynGltA enables the growth of ΔgltA cells on minimal medium. (A) The TCA cycle showing the condensation of acetyl-CoA and oxaloacetate performed by citrate synthase (GltA). Citrate is processed into α-ketoglutarate, which is a precursor of glutamate. (B) An overview of the auxotroph screen: A strain of E. coli in which gltA is deleted cannot grow on minimal medium. A plasmid encoding LacZ (negative control) fails to support growth, whereas both the native E. coli GltA (positive control) and the de novo protein SynGltA support growth of colonies on minimal medium

Figure 1
100x120mm (300 x 300 DPI)

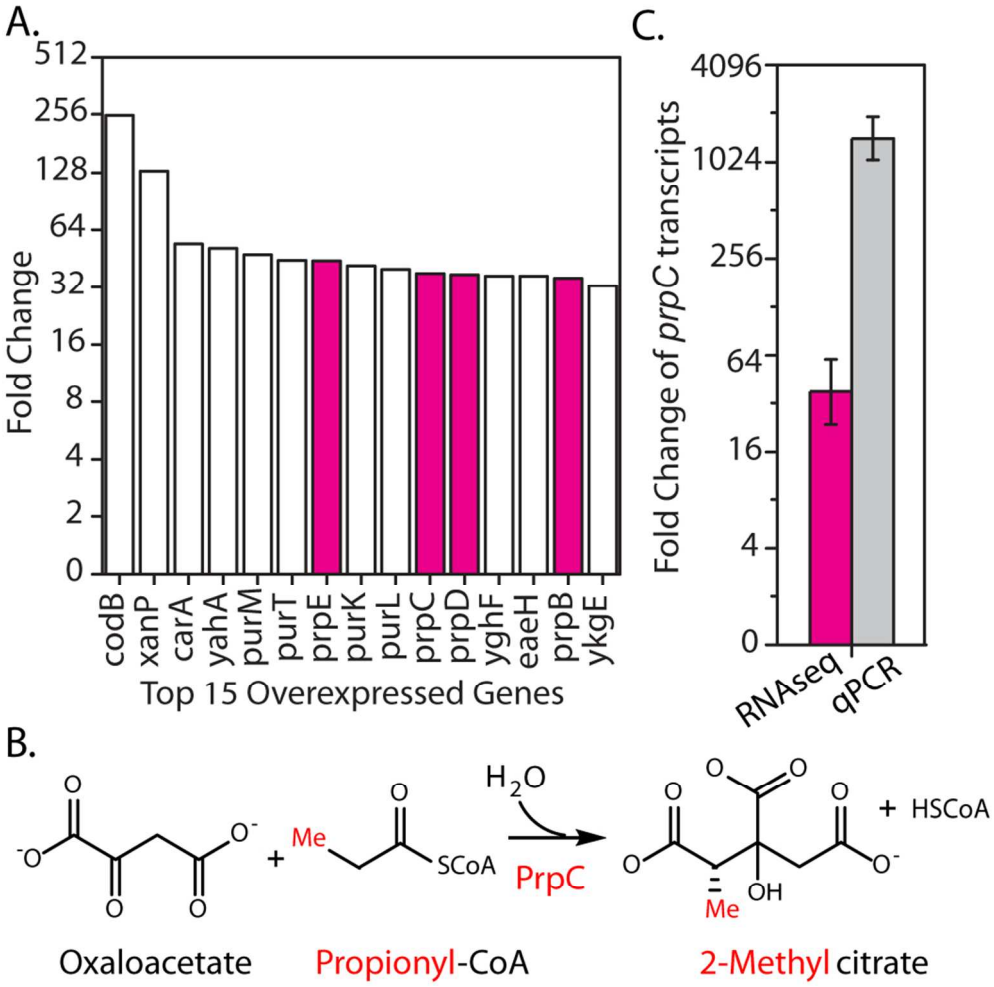


Figure 2: The de novo protein SynGltA increases expression of prpC, which encodes methyl citrate synthase. (A) Bars show the top 15 upregulated transcripts in ΔgltA cells expressing SynGltA relative to ΔgltA cells with added glutamate. Transcripts were quantified by RNAseq. (B) The condensation of propionyl-CoA 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.

Figure 2
90x90mm (300 x 300 DPI)

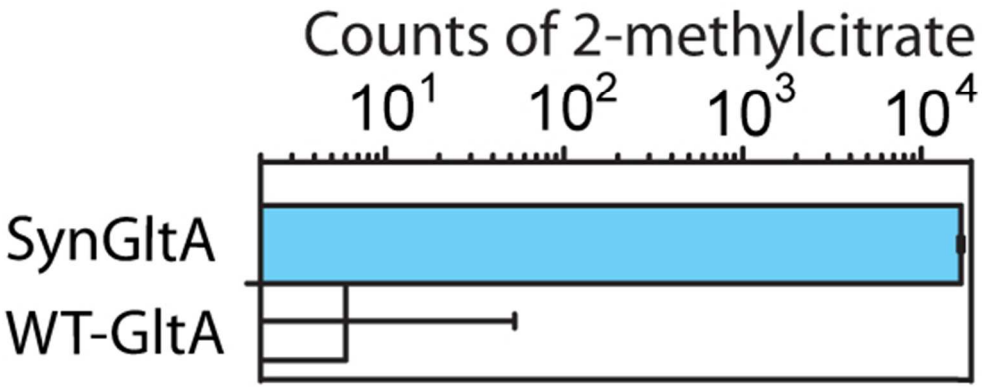


Figure 3: Expression of SynGltA increases abundance of 2-Methylcitrate. Counts from LC/MS probing ΔgltA cells expressing either native E. coli GltA (WT-GltA) or SynGltA. There is far more 2-methylcitrate in ΔgltA cells expressing SynGltA (blue bar) than in cells expressing WT GltA. Error bars represent SD.
Figure 3
50x20mm (300 x 300 DPI)

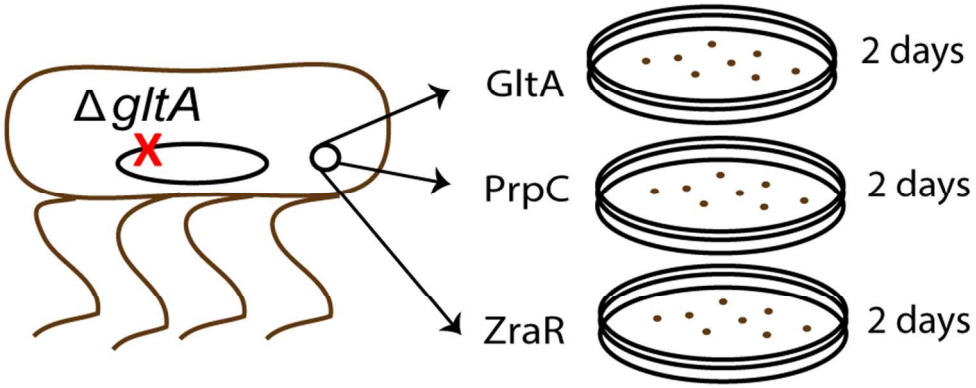


Figure 4: Three overexpressed E coli ORFS from the ASKA collection can rescue ΔgltA cells. GltA is citrate synthase, PrpC is methylcitrate synthase, and ZraR is a zinc-response associated protein.
Figure 4
80x33mm (300 x 300 DPI)

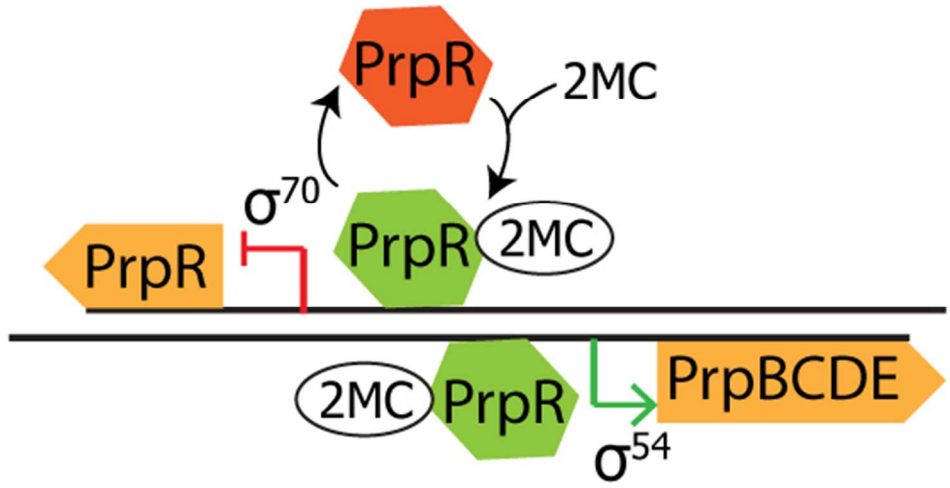


Figure 5: Native regulation of the prp operon. When PrpR is bound to 2-methylcitrate (2MC), it is activated to stimulate transcription of the prp operon including prpC, which encodes methyl citrate synthase. In addition, activated PrpR is an autorepressor.
Figure 5
50x26mm (300 x 300 DPI)