A protein constructed de novo enables cell growth by altering gene regulation

Katherine M. Digianantonio[a] and Michael H. Hecht[a,1]

[a]Department of Chemistry, Princeton University, Princeton, NJ 08540

Edited by Richard E. Lenski, Michigan State University, East Lansing, MI, and approved January 29, 2016 (received for review January 14, 2016)

Recent advances in protein design rely on rational and computational approaches to create novel sequences that fold and function. In contrast, natural systems selected functional proteins without any design a priori. In an attempt to mimic nature, we used large libraries of novel sequences and selected for functional proteins that rescue *Escherichia coli* cells in which a conditionally essential gene has been deleted. In this way, the de novo protein SynSerB3 was selected as a rescuer of cells in which *serB*, which encodes phosphoserine phosphatase, an enzyme essential for serine biosynthesis, was deleted. However, SynSerB3 does not rescue the deleted activity by catalyzing hydrolysis of phosphoserine. Instead, SynSerB3 upregulates *hisB*, a gene encoding histidinol phosphate phosphatase. This endogenous *E. coli* phosphatase has promiscuous activity that, when overexpressed, compensates for the deletion of phosphoserine phosphatase. Thus, the de novo protein SynSerB3 rescues the deletion of *serB* by altering the natural regulation of the *His* operon.

de novo protein design | serB | hisB | auxotroph | synthetic biology

One of the key goals of synthetic biology is to design novel proteins that fold and function in vivo. A particularly challenging objective would be to produce nonnatural proteins that do not merely generate interesting phenotypes but actually provide essential functions necessary for the growth of living cells. The successful design of such life-sustaining proteins would represent an initial step toward constructing artificial "proteomes" of nonnatural sequences.

We initiated work toward artificial proteomes by constructing large combinatorial libraries of novel sequences designed to fold into stable four-helix structures (1, 2). Our libraries were based on a strategy for protein design, which assumes that the overall fold of a simple structure can be specified by the pattern of polar and nonpolar residues in the linear sequence. Because only the type of residue—polar versus nonpolar—is specified, this strategy has been called a "binary code" for protein design (1–3). At the same time, because the exact identities of the side chains at each polar and nonpolar position are not specified explicitly, this strategy is well suited for constructing large combinatorial libraries of novel sequences (3). To express these libraries of binary-patterned sequences in vivo, we construct collections of synthetic genes using degenerate DNA codons. For example, the degenerate codon NTN (N = A,T,C,G) is used to encode five nonpolar residues, and the degenerate codon VAN (V = A,C,G) is used to encode six polar residues.

We have shown previously that several proteins from these binary patterned libraries fold into stable four-helix bundles, and both crystallographic and NMR structures have been determined (4–6). Moreover, in initial steps probing the potential for functional activity, proteins from these libraries were shown to bind small molecules, including cofactors, and to catalyze rudimentary reactions (7, 8). Although those experiments screened a subset of library proteins for activity in vitro, they did not probe for function in vivo.

To mount an unbiased search for proteins that provide life-sustaining functions in vivo, we used life-or-death selections. Specifically, we selected for proteins that rescue the deletion of conditionally essential genes in *E. coli*. We transformed a library of 1.5×10^6 binary patterned de novo sequences into strains of *E. coli* that contain deletions of a gene encoding a protein required for survival on minimal medium (9). Such strains, called auxotrophs, grow on rich medium but are unable to grow on minimal medium because a protein involved in the biosynthesis or absorption of an essential metabolite has been deleted. Our initial studies tested numerous auxotrophic strains from the Keio collection, which contains all viable single-deletion strains of *E. coli* (10). Although most of the auxotrophs were not rescued by proteins from our library, we found four auxotrophic strains that were reproducibly rescued by novel sequences from our binary-patterned library (9).

This report describes the mechanism by which one of our novel proteins supports cell growth. We show that the de novo protein SynSerB3 rescues the deletion of the native *E. coli* SerB (which encodes phosphoserine phosphatase), not by catalyzing the phosphoserine phosphatase reaction but rather by performing a regulatory function. [This sequence, first isolated by Fisher et al. (9), was named "SynSerB3" because it was the third sequence isolated in a selection for *syn*thetic sequences that rescue the Δ*serB* auxotroph.] SynSerB3 causes elevated expression of HisB, which encodes histidinol phosphate phosphatase. This phosphatase has promiscuous activity and, when expressed at high levels, is capable of hydrolyzing enough phosphoserine to enable the growth on minimal medium of otherwise moribund Δ*serB* cells. These results demonstrate that a nonnatural de novo protein enables cell growth by altering gene regulation.

Results

The de Novo Protein SynSerB3 Enables Growth of Δ*serB* Cells on Minimal Medium. The *serB* gene in *E. coli* encodes phosphoserine phosphatase, which catalyzes the final step in serine biosynthesis (Fig. 1*A*). Because serine is an essential metabolite, Δ*serB* cells are auxotrophs and cannot grow on minimal medium. We reported previously that several de novo proteins from our binary patterned

Significance

We describe a novel regulatory protein that was discovered in a library of de novo-designed proteins never sampled by nature. This de novo protein, SynSerB3, rescues a conditionally lethal mutation in *Escherichia coli* cells that prevents growth in the absence of serine. We found that SynSerB3 does not catalyze the biosynthesis of serine but rather upregulates an endogenous protein, histidinol phosphate phosphatase, which can synthesize serine via a promiscuous catalytic activity. This regulatory function of SynSerB3 sustains life in *E. coli* cells under conditions that are otherwise lethal and is a synthetic addition to natural gene regulation.

Author contributions: K.M.D. and M.H.H. designed research; K.M.D. performed research; K.M.D. and M.H.H. analyzed data; and K.M.D. and M.H.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: hecht@princeton.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600566113/-/DCSupplemental.

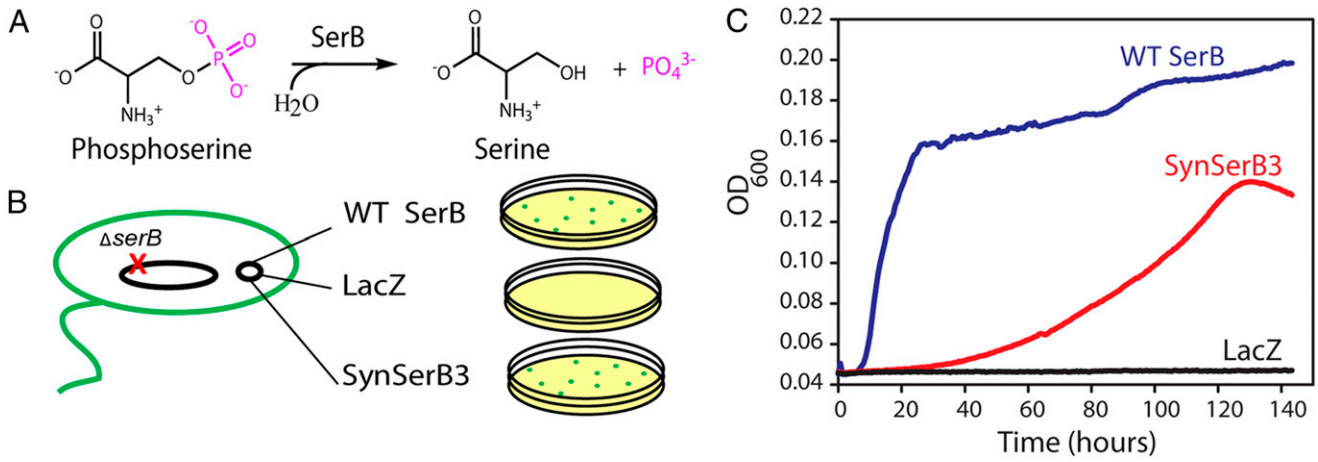

Fig. 1. SynSerB3 enables the growth of Δ*serB* cells in minimal medium. (*A*) The phosphoserine phosphatase reaction catalyzed by the enzyme encoded by *serB*. (*B*) An overview of the auxotroph screen: A strain of *E. coli* in which *serB* is deleted cannot grow on minimal medium. A plasmid encoding the negative control (LacZ) fails to support growth, whereas both the native *E. coli* SerB (positive control), and the de novo protein SynSerB3 support growth on minimal medium. (*C*) The de novo protein SynSerB3 allows the auxotrophic strain Δ*serB* to grow in liquid minimal medium. Strains expressing the natural *E. coli* enzyme encoded by *serB* grow more rapidly, whereas cells expressing the control protein LacZ failed to grow over the course of 6 d.

libraries rescue the *serB* deletion on minimal medium (9). We chose one of these proteins, SynSerB3, for detailed characterization because it rescues Δ*serB* in fewer days than the other SynSerB proteins.

Before setting out to determine the mechanism of rescue, we sought to confirm that SynSerB3 is indeed responsible for the rescue of the Δ*serB* auxotroph. As depicted in Fig. 1*B*, Δ*serB* cells were transformed with plasmids expressing SynSerB3, LacZ (negative control), or WT *E. coli* SerB (positive control). Transformants were plated on M9/glucose minimal plates containing isopropyl-β-D-1thiogalactopyranoside (IPTG) to induce expression. As expected, Δ*serB* cells expressing LacZ failed to grow, even after 14 d, whereas those expressing the native *E. coli* SerB formed colonies in 2 d. The same cells expressing SynSerB3 formed colonies in 4 d, confirming that SynSerB3 rescues the Δ*serB* deletion. Not surprisingly, growth sustained by the de novo protein is slower than that of the natural protein, which had been selected by billions of years of evolution.

To confirm the results observed on plates and to quantify the growth rates of the rescued cells more accurately, we assayed growth in liquid cultures. As shown in Fig. 1*C*, the controls produce the expected growth curves in minimal medium: Δ*serB* cells expressing the natural SerB gene showed exponential growth with a relatively short lag time, and cells expressing LacZ did not grow at all. As expected from our results on solid medium, cells expressing the de novo protein SynSerB3 grew in minimal medium but did not grow as rapidly as cells expressing the native *E. coli* protein.

To further confirm that the sequence of SynSerB3 is responsible for the rescue, we performed following additional experiments:

i) The DNA fragment encoding SynSerB3 was recloned into the expression vector and retransformed in Δ*serB* cells to confirm that the SynSerB3 sequence, and no other sequences on the plasmid or in the host strain, is responsible for rescue.

ii) Many other de novo sequences from our binary patterned libraries were shown to be unable to rescue Δ*serB*, thereby demonstrating that expression of an arbitrary binary patterned protein does not induce a generic response responsible for rescue.

iii) One particular binary patterned protein, SynGltA, which rescues the deletion of citrate synthase (essential for glutamate biosynthesis) was used as a control for many experiments in this study and was shown to be unable to rescue Δ*serB*.

iv) Single base changes causing either stop codons or frameshifts were introduced into SynSerB3 and were shown to knock out activity, thereby demonstrating that expression of the SynSerB3 protein, not merely its mRNA, is required for rescue.

v) A codon-optimized gene encoding the SynSerB3 amino acid sequence was synthesized and shown to rescue Δ*serB* cells with the same growth rate as the original SynSerB3 gene, further confirming that the SynSerB3 protein, not its mRNA, mediates rescue.

vi) Single amino acid changes introduced into SynSerB3 were shown to prevent rescue, thereby demonstrating that the phenotype conferred by SynSerB3 depends upon its exact amino acid sequence.

The SynSerB3 Protein Has No Detectable Phosphoserine Phosphatase Activity. In principle SynSerB3 could rescue ΔserB cells either by enabling the conversion of phosphoserine to serine (the reaction deleted by Δ*serB*) or by facilitating a novel pathway that bypasses this step. If SynSerB3 facilitated a bypass pathway, it would be expected to rescue the deletion of other enzymes in the serine biosynthesis pathway. However, as reported by Fisher et al. (9), SynSerB3 does not rescue Δ*serC*, which encodes the enzyme that catalyzes the step before SerB. Thus, SynSerB3 does not bypass the natural serine biosynthesis pathway; it rescues Δ*serB* cells by enabling the same step catalyzed by the deleted enzyme, phosphoserine phosphatase.

Because expression of SynSerB3 enables the conversion of phosphoserine to serine, we began our studies by testing whether the de novo protein accomplished this conversion by direct action, i.e., by catalyzing this reaction. We purified the SynSerB3 protein using affinity chromatography followed by size-exclusion chromatography (*SI Materials and Methods*). To ensure there was no possibility of contamination by the natural *E. coli* phosphoserine phosphatase enzyme, all purifications were done from Δ*serB* cells. The purified SynSerB3 protein was incubated with phosphoserine in a variety of buffers, and liberation of phosphate was assayed using a standard malachite green assay (11). The positive control, SerB from *E. coli*, showed high levels of activity. However, the de novo SynSerB3 protein displayed no detectable activity under any of the conditions tested.

Because the purified SynSerB3 protein was not enzymatically active, we considered the possibility that SynSerB3 might require an endogenous molecular partner—either another protein or a cofactor —to catalyze the reaction. To test this possibility, we assayed for phosphoserine phosphatase activity in cell lysates. Because the malachite green assay is unreliable in lysates, we used ^{13}C NMR to assay the reaction. As expected, lysates containing the positive control protein, native *E. coli* SerB, showed rapid conversion of phosphoserine to serine. Unfortunately, the negative control containing LacZ also showed low levels of activity. We attribute this activity to nonspecific *E. coli* phosphatases in the lysate. For example, alkaline phosphatase can catalyze the phosphoserine phosphatase reaction but does not rescue Δ*serB* in vivo because it resides in the periplasm. Lysates from cells expressing SynSerB3 showed activity similar to those from cells expressing LacZ, and we presume this activity is also caused by endogenous nonspecific phosphatases in the lysate. Not surprisingly, all lysates also showed the eventual disappearance of serine, which can be attributed to downstream metabolic processes that use serine to synthesize other metabolites (Fig. S1) (12).

SynSerB3 Alters Gene-Expression Profiles in *E. coli*. Our finding that purified SynSerB3 is not active as a phosphoserine phosphatase suggests that this de novo protein rescues Δ*serB* cells by an indirect mechanism involving regulation and/or activation of endogenous *E. coli* genes or proteins. To probe for altered gene regulation in a model-independent and unbiased way, we performed quantitative RNA sequencing (RNAseq). RNAseq profiles of cells expressing SynSerB3 were compared with those of controls expressing WT *E. coli* SerB from the same vector. These experiments were performed in two cellular backgrounds, the Δ*serB* auxotroph and the pseudo-WT Keio parent strain, BW25113. Cells were grown in minimal medium, and samples were collected during mid-logarithmic growth. See Fig. S2 for all-sample quality control.

In the parental (nondeletion) strain BW25113 we observed dramatic differences between cells expressing SynSerB3 and control cells expressing *E. coli* SerB. Because the chromosome of this pseudo-WT strain encodes all the genes required for growth in minimal medium, the observed differences must be caused by the expression of different proteins on the plasmid. Analysis of the RNAseq data revealed that BW25113 cells expressing the de novo protein SynSerB3 contain 10-fold higher levels of mRNAs from the histidine biosynthetic operon (Fig. 2*A*). Expression of SynSerB3 also led to the down-regulation of aromatic amino acid biosynthetic genes (Dataset S1, table 2). [When comparing cells expressing SynSerB3 with those expressing *E. coli* SerB, it can be difficult to determine whether one protein causes down-regulation or the other causes up-regulation. In the current example, we state that SynSerB3 alters regulation because separate control experiments showed that overexpressing *E. coli* SerB in BW25113 cells has little effect on the expression profile (Dataset S1, table 1).]

Next we examined expression profiles in the deletion strain, Δ*serB*, and compared cells expressing SynSerB3 with the same cells expressing native *E. coli* SerB. In this strain, many more genes were differentially expressed, presumably because of the added complexities associated with the chromosomal deletion of the *serB* gene. Analysis of the RNAseq data revealed that expression of 633 genes was altered fourfold or more in Δ*serB* cells expressing SynSerB3 versus the control. Some genes, such as those involved in fermentation and other carboxylic acid biosynthetic processes, were down-regulated; others, including those involved in the SOS response, were up-regulated. These results are discussed further below and in *SI Discussion*.

An advantage of the RNAseq method is its ability to probe the entire transcriptome in an experiment that is not biased by models or expectations. On the other hand, a disadvantage of RNAseq is that it can be difficult to determine which of the observed changes are responsible for the biological phenotype. In the current situation, several orthogonal experiments described in the next sections suggest that up-regulation of the *his* operon is responsible for the rescue of Δ*serB* cells by SynSerB3. In particular, we hypothesized that overexpression of HisB, which encodes histidinol phosphate phosphatase, might rescue the deletion of phosphoserine phosphatase. Therefore it was gratifying to see that expression of the *his* operon was enhanced 10-fold by expression of SynSerB3 in BW25113 cells (Fig. 2*A*). However, in the Δ*serB* strain, enhanced expression of the His operon was not observed by RNAseq. Therefore, we used a second method, quantitative PCR (qPCR), to measure changes in the expression of HisB in both strains. As shown in Fig. 2*B*, qPCR showed that *hisB* transcripts were increased in response to SynSerB3 in both cellular backgrounds: fivefold in Δ*serB* cells and 25-fold in BW25113 cells (Fig. 2*B*).

In summary, both model-free assays (RNAseq) and focused experiments (qPCR) demonstrated that the de novo protein SynSerB3 leads to an increase in the transcription of HisB, which encodes histidinol phosphate phosphatase.

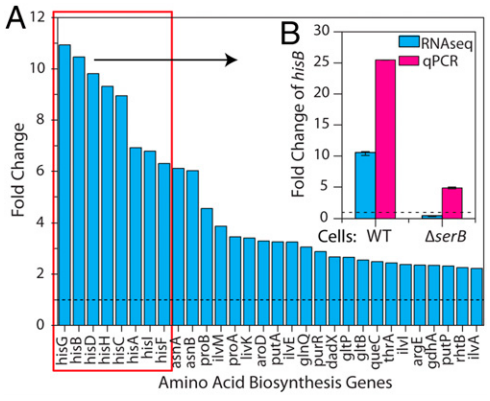

Fig. 2. The de novo protein SynSerB3 increases expression of *hisB*, which encodes histidinol phosphate phosphatase. (*A*) Bars show the top 30 up-regulated transcripts for amino acid biosynthesis genes in pseudo-WT strain BW25113 expressing SynSerB3, relative to the same cells expressing native *E. coli* SerB (using RNAseq). (*B*) The abundance of *hisB* transcripts measured using both RNAseq and qPCR in both the pseudo-WT strain BW25113 and in the Δ*serB* auxotroph. The ratio of abundance is shown for cells expressing SynSerB3 relative to the same cells expressing native *E. coli* SerB. The dotted line represents a fold change of 1, indicating no change. Error bars represent the SE.

SynSerB3 Activates Transcription of the *His* Operon. The RNAseq and qPCR results described in the preceding section demonstrate that SynSerB3 increases the abundance of mRNA transcripts from the *his* operon. In principle, abundance could be increased either by stimulating the transcription of the mRNA or by slowing its degradation. To distinguish between these possibilities, we replaced the sequence of the long polycistronic *his* mRNA with GFP. As shown Fig. 3*A*, this construct contains all the *his* operon regulatory signals for transcription and translation but does not contain the *his* structural genes. Because these latter sequences presumably dictate the rate of *his* operon mRNA degradation, the induction of increased GFP fluorescence by SynSerB3 would indicate an increase in the rate of transcription initiated at the *his* regulatory region. As shown in the center pair of bars in Fig. 3*B*, SynSerB3 increased GFP fluorescence fivefold above controls, demonstrating that the de novo protein increases transcription of the *his* operon.

Transcription of the *his* operon in *E. coli* is natively regulated in two ways. First, transcription is stimulated in the stringent response (induced by amino acid starvation) by allosteric binding of ppGpp-DksA to RNA polymerase. Second, transcription is attenuated by high intracellular concentrations of histidine. Both mechanisms are

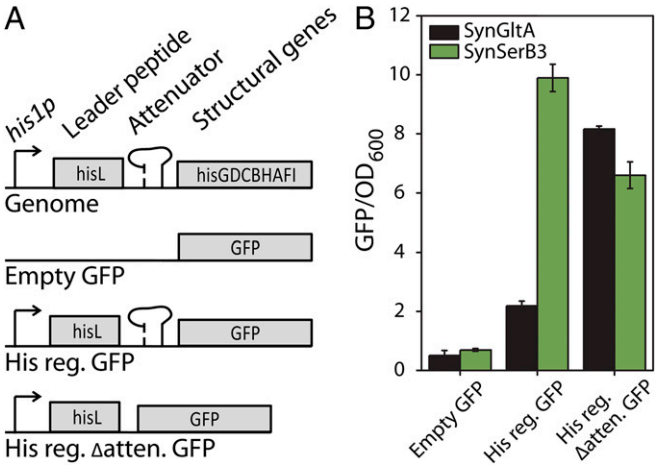

Fig. 3. SynSerB3 increases transcription of the *his* operon. (*A*) Fusion of GFP to the regulatory region of the histidine biosynthesis operon. In His reg. GFP, the entire regulatory region is retained; in His reg.ΔattenGFP, the attenuator sequence is deleted. In both constructs, the structural genes of the *his* operon are replaced by GFP. (*B*) GFP fluorescence normalized for cell density (OD_{600}) from overnight cultures. The pair of bars on the left show results for a control plasmid containing none of the *his* operon regulatory region. For this vector, minimal GFP is expressed in the presence of SynSerB3 and in the control protein SynGltA. The pair of bars in the middle correspond to the His reg.GFP construct shown in *A* containing all the *his* operon regulatory region. In this case, the presence of SynSerB3 increases GFP levels dramatically above those in the control. The pair of bars on the right correspond to the His reg.ΔattenGFP construct, which lacks the attenuator sequence. In this case, the presence of SynSerB3 does not increase GFP levels above those in the control. Error bars represent SE.

regulated by sequences upstream of the *his* structural genes in a regulatory region that includes the promoter, *his* leader peptide (*hisL*), and attenuator sequence (13).

We performed several experiments to determine whether SynSerB3 affects regulation via the stringent response or the attenuation mechanism. First, to assess whether SynSerB3 activates the stringent response, we compared the known gene-expression signatures of this response, which includes increased expression of amino acid biosynthetic genes and decreased expression of stable RNAs (rRNA and tRNA) (14–16), with the RNAseq data from the current study (Fig. S3) Although there are some correlations between the stringent response and the RNAseq results for Δ*serB* cells expressing SynSerB3, these correlations are not observed when SynSerB3 is expressed in pseudo-WT BW25113 cells. Thus, the genetic background of the cells, Δ*serB*, which leads to a shortage of serine, rather than expression of the de novo protein per se, is responsible for the observed correlation. (See further discussion in *SI Discussion*.)

For further confirmation that the stringent response is not regulated by SynSerB3, we constructed several different fusions of the *his* regulatory region to GFP. As shown in the center pair of bars in Fig. 3*B*, SynSerB3 has a substantial impact on expression of GFP from the full-length *his* regulatory sequence. However, deletion of the attenuator sequence equalizes GFP fluorescence between cells expressing SynSerB3 and those expressing the control protein (Fig. 3*B*, right pair of bars). These results indicate that SynSerB3 alters regulation via the attenuation mechanism and not by the ppGpp-mediated stringent response. (In constructs where the attenuation sequences are deleted, GFP fluorescence is high in cells with or without SynSerB3 because the *his* promoter is a strong promoter, and there is no stem-loop to attenuate transcription.) From these results we conclude that SynSerB3 activates transcription of the *his* operon by deattenuation.

In principle, there are several ways that SynSerB3 could cause deattenuation. For example, SynSerB3 could reduce the abundance of histidine, which is required for the attenuation mechanism. To address this possibility, we measured the concentration of histidine in cells expressing SynSerB3, relative to controls expressing WT SerB. As shown in Fig. 4, LC/MS showed that cells expressing SynSerB3 contain higher amounts of histidine, presumably because of the overexpression of the histidine biosynthesis genes. Therefore we can rule out a mechanism by which SynSerB3 reduces the abundance of histidine.

Another mechanism by which SynSerB3 could cause deattenuation would be by decreasing the amount of the single ^{his}tRNA, encoded by *hisR*. To address this possibility, we used RT-qPCR to measure the amount of ^{his}tRNA in both Δ*serB* cells and pseudo-WT BW25113 cells. We found no significant difference in abundance between cells expressing SynSerB3 and controls expressing native SerB, thereby demonstrating that the de novo protein does not function by altering ^{his}tRNA levels.

Although SynSerB3 does not alter the intracellular abundance of histidine or ^{his}tRNA, in principle it could alter the functional pool of ^{his}tRNA by inhibiting successful charging of the ^{his}tRNA. Alternatively, SynSerB3 could disrupt attenuation by binding the attenuator RNA.

Histidinol Phosphate Phosphatase Encoded by HisB Is Promiscuous and Catalyzes Hydrolysis of Serine Phosphate. Our finding that SynSerB3 rescues the Δ*serB* auxotroph by causing overexpression of HisB suggests that histidinol phosphate phosphatase might have a promiscuous activity capable of catalyzing the hydrolysis of serine phosphate. This suggestion can be tested both genetically and biochemically. The genetic studies were first reported by Patrick et al. (17), who conducted a large-scale study to test whether chromosomal deletions of single genes in *E. coli* could be rescued by other *E. coli* genes overexpressed from a plasmid. In the case of the Δ*serB* auxotroph, Patrick et al. (17) found that overexpression of HisB rescued the deletion. We repeated this experiment and observed the same result. In related experiments, Blank et al. (18) searched for chromosomal mutations that rescue deletions. In the case of Δ*serB,* they also found rescuing mutations that either enhanced expression of HisB or relaxed its specificity.

The ability of histidinol phosphate phosphatase to hydrolyze serine phosphate was confirmed biochemically by Yip and Matsumura (19), who reported a $k\text{cat}/K_m$ value of 7.6 M/s. This promiscuous off-target activity of the HisB-encoded enzyme is 10,000-fold lower than the native activity of the *E. coli* SerB enzyme for the same substrate. This dramatic difference in catalytic activity is consistent with the requirement that HisB must be overexpressed for it to rescue Δ*serB*. Indeed, the auxotrophy of the Δ*serB* mutant shows explicitly that endogenous chromosomal expression of HisB is not sufficient to enable the growth of Δ*serB* cells on minimal medium.

The *E. coli* HisB Gene Is Required for Rescue of Δ*serB* Cells by SynSerB3. The experiments described above demonstrate that the de novo protein SynSerB3 increases expression of the *E. coli* gene HisB. Because HisB encodes a phosphatase, these results

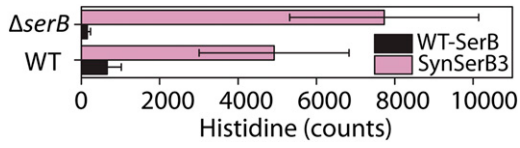

Fig. 4. Histidine counts from LC/MS in pseudo-WT and Δ*serB* cells expressing either native *E. coli* SerB (WT SerB) or SynSerB3. There is more histidine in cells expressing SynSerB3 (purple bar) than in cells expressing WT SerB. Error bars represent the SE.

suggested that rescue by SynSerB3 of the phosphoserine phosphatase deletion in Δ*serB* is mediated by the overexpression of an alternative *E. coli* phosphatase with promiscuous activity.

However, other transcripts also were increased or decreased in response to SynSerB3, and the observation of enhanced transcription of HisB does not explicitly prove that it is responsible for the rescue phenotype. To confirm that increased expression of HisB is essential for the rescue, it is crucial to show that rescue of the Δ*serB* auxotroph by SynSerB3 cannot occur in the absence of HisB. Therefore, we constructed the double-deletion strain, Δ*serB*Δ*hisB*, and tested whether the de novo protein still could rescue the serine auxotroph in this genetic background. Because the deletion of *hisB* causes a requirement for histidine, which is tangential to the current study, we added histidine to the medium. The plasmid encoding SynSerB3 was transformed into the double-knockout strain, and no growth was observed after 14 d. This finding and appropriate controls are summarized in Table 1. Thus, in agreement with the hypothesis that SynSerB3 rescues Δ*serB* by enhancing expression of histidinol phosphate phosphatase, the presence of the HisB gene is indeed required for auxotroph rescue by SynSerB3.

The SOS Response Is Induced by Expressing SynSerB3 in Δ*serB* Cells Grown in Minimal Medium. Analysis of RNAseq data for Δ*serB* cells expressing SynSerB3 showed enhanced expression of genes associated with the SOS response (Fig. 5*A*) (20). We further assayed this response using qPCR to measure the induction of two proteins associated with SOS: RecA, a central regulator of SOS, and SulA, a known inhibitor of cell division. We found that *recA* and *sulA* are induced when Δ*serB* cells expressing SynSerB3 are grown in minimal medium. Likewise a filamentous cell shape, also associated with SOS (21), was observed for Δ*serB* cells expressing SynSerB3 in minimal medium (Fig. S4). Both the filamentous morphology and the induction of *recA* and *sulA* were observed only in Δ*serB* cells expressing SynSerB3 in minimal medium. They were not observed when stress was relieved by using the pseudo-WT strain (in minimal medium) or by expressing SynSerB3 (in either strain) in rich medium (Fig. 5*B* and Fig. S4). Thus, it appears that overall stress, rather than the expression of SynSerB3, induces SOS.

Nonetheless, we considered the possibility that the SOS response might be responsible for rescuing the Δ*serB* auxotroph. To test this possibility, we plated Δ*serB* cells on minimal medium spiked with sublethal concentrations (2–10 μg/mL) of nalidixic acid, a known inducer of SOS (22, 23). These cells did not grow on minimal medium, thereby confirming that induction of the SOS response is not responsible for rescue of the Δ*serB* auxotroph.

Further evidence that induction of the *his* operon by SynSerB3—and not the SOS response—is responsible for the rescue of Δ*serB* cells comes from several of the results described above. (*i*) Both RNAseq and qPCR (Fig. 2) show that SynSerB3 induces

Table 1. Growth of auxotrophic strains of *E. coli* in minimal medium

Strain	Plasmid expressing	No amino acid	+His
Δ*serB*	LacZ	X	X
	SynSerB3	**G**	**G**
	HisB	G	–
Δ*hisB*Δ*serB*	LacZ	X	X
	SynSerB3	**X**	**X**
	HisB	G	–
	SerB	X	G

As highlighted in bold and underlined, the de novo protein SynSerB3 rescues the Δ*serB* auxotroph only in the presence of a chromosomal copy of the endogenous *E. coli HisB* gene. Test tubes were monitored for 10 d, and plates were monitored for colonies with diameters >1 mm for 14 d. In the experiments described, growth occurred within 4 d. G, growth; X, no growth; –, experiment not performed.

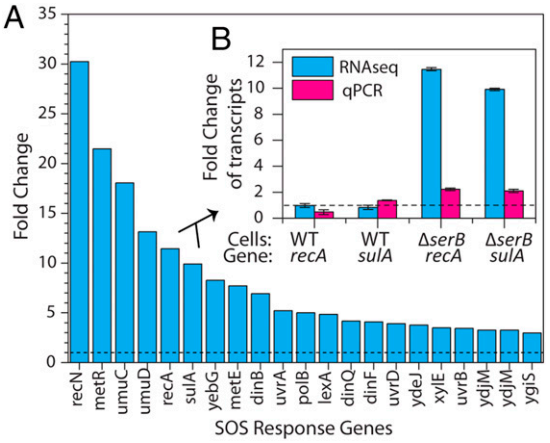

Fig. 5. The SOS response is turned on in Δ*serB* cells but not in the pseudo-WT strain. (*A*) Many genes in the SOS regulon are up-regulated (above the dotted line) in Δ*serB* cells expressing SynSerB3 in minimal medium. (*B*) Transcript levels of *sulA* and *recA* as measured by RNAseq and qPCR in both the pseudo-WT strain BW25113 and in Δ*serB* cells. In both *A* and *B*, bars show the abundance of transcript in cells expressing SynSerB3 relative to the same cells expressing native *E. coli* SerB. Error bars represent the SE.

expression of the *his* operon in the BW25113 strain; however SynSerB3 does not induce filamentation, *recA*, or *sulA* in this strain (Fig. 5). (*ii*) GFP fusions to the regulatory region of the *his* operon show that SynSerB3 alters the regulation of this operon (Fig. 3) under conditions in which SOS and filamentation are not induced in BW25113 the strain. Thus, we conclude that the de novo protein SynSerB3 rescues Δ*serB* cells by inducing the promiscuous phosphatase encoded by HisB and that the SOS response results from the overall stress these cells experience while growing in minimal medium using a rescue mechanism that is not robust, but is just sufficient to sustain life.

Discussion

The number of sequences that can be encoded by an alphabet of 20 amino acids far exceeds the number of atoms in the universe. Even for relatively short sequences of 102 amino acids, there are many more possibilities (5×10^{132}) than could have been sampled by evolution. From this large universe of possibilities, nature selected relatively small collections of sequences (proteomes) to sustain the growth of living organisms. Thus, the *E. coli* genome encodes only 4,300 proteins (24), and even the human genome contains only ~20,000 sequences (25, 26).

These considerations led us to question whether novel protein sequences, never sampled by nature, might be able to provide essential functions necessary to sustain life. To address this possibility, we imposed life-or-death selections to isolate functional proteins from a library of 1.5 million de novo sequences. Several novel sequences were discovered which rescued a range of different deletions (9). Among these, SynSerB3 was isolated as a rescuer of Δ*serB*, an auxotroph in which phosphoserine phosphatase, which catalyzes the last step in serine biosynthesis, was deleted.

At the outset, we anticipated that the novel protein would rescue Δ*serB* by functioning in the same way as the deleted enzyme, i.e., by catalyzing the hydrolysis of phosphoserine. However, extensive studies demonstrated that SynSerB3 is not active as a phosphoserine phosphatase and therefore must exert its life-sustaining phenotype by functioning as a regulator of endogenous genes and/or proteins.

Through a series of experiments including both unbiased searches (RNAseq) and targeted analyses (qPCR and GFP fusions), we showed that SynSerB3 enhances expression of HisB, which encodes histidinol phosphate phosphatase, by deattenuating the *his* operon.

Further experiments showed that (*i*) histidinol phosphate phosphatase has promiscuous activity that can hydrolyze phosphoserine in vitro, and (*ii*) this promiscuous activity, when expressed at high levels, is sufficient in vivo to sustain the growth of Δ*serB* cells on minimal medium. After showing that HisB was sufficient for rescue, we confirmed that it was necessary by demonstrating that the ability of SynSerB3 to rescue Δ*serB* cells absolutely requires the presence of the HisB gene. Taken together, these findings demonstrate that a novel protein, unrelated to sequences in nature, can enable cell growth by altering the natural regulatory landscape.

Evolution is an opportunistic process. When presented with environmental challenges, organisms adapt by using a range of strategies, including new enzymatic functions and novel regulatory processes. In natural systems, there are numerous precedents showing that selecting for growth under specified conditions can yield mutations that rewire gene regulation (27–31). For example, Taylor et al. (27) showed that immotile mutants of *Pseudomonas fluorescens* subjected to selections for motility evolve by repurposing a protein that normally functions in nitrogen uptake toward a new function involving flagellar regulation.

Our studies demonstrate that nonnatural, synthetic biological systems can also surmount environmental challenges. Rather than mutating a naturally occurring gene, as in the above examples, we demonstrate that a de novo-designed protein can drive adaptive changes in gene expression. To the best of our knowledge, this is the first example of a de novo protein that provides a life-sustaining regulatory function.

Materials and Methods

For additional information on cell growth and manipulation, protein purification, biochemical assays, differential gene expression analysis, and metabolomics profiling, please see *SI Materials and Methods*. Datasets associated with gene expression are given in Dataset S1. Primers are listed in Table S1.

ACKNOWLEDGMENTS. We thank Dr. 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; Lance Parsons for help in analyzing the RNAseq data; and Xin Teng and Prof. Josh Rabinowitz for performing metabolite LC/MS. This research was funded by National Science Foundation (NSF) Grant MCB-1050510. K.M.D. was supported by an NSF Graduate Research Fellowship.

1. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262(5140):1680–1685.
2. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. *Protein Sci* 13(7):1711–1723.
3. Bradley LH, Kleiner RE, Wang AF, Hecht MH, Wood DW (2005) An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences. *Protein Eng Des Sel* 18(4):201–207.
4. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a de novo protein from a designed combinatorial library. *Proc Natl Acad Sci USA* 100(23):13270–13273.
5. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. *Protein Sci* 17(5):821–832.
6. Arai R, et al. (2012) Domain-swapped dimeric structure of a stable and functional de novo four-helix bundle protein, WA20. *J Phys Chem B* 116(23):6789–6797.
7. Patel SC, Bradley LH, Jinadasa SP, Hecht MH (2009) Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. *Protein Sci* 18(7):1388–1400.
8. Cherny I, Korolev M, Koehler AN, Hecht MH (2012) Proteins from an unevolved library of de novo designed sequences bind a range of small molecules. *ACS Synth Biol* 1(4): 130–138.
9. Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH (2011) De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. *PLoS One* 6(1):e15364.
10. Baba T, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* 2:0008.
11. Geladopoulos TP, Sotiroudis TG, Evangelopoulos AE (1991) A malachite green colorimetric assay for protein phosphatase activity. *Anal Biochem* 192(1):112–116.
12. Nelson DL, Cox MM (2008) *Principles of Biochemistry* (W.H. Freeman and Company, New York), 5th Ed.
13. Alifano P, et al. (1996) Histidine biosynthetic pathway and genes: Structure, regulation, and evolution. *Microbiol Rev* 60(1):44–69.
14. Keseler IM, et al. (2011) EcoCyc: A comprehensive database of Escherichia coli biology. *Nucleic Acids Res* 39(Database issue):D583–D590.
15. Durfee T, Hansen AM, Zhi H, Blattner FR, Jin DJ (2008) Transcription profiling of the stringent response in Escherichia coli. *J Bacteriol* 190(3):1084–1096.
16. Traxler MF, et al. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. *Mol Microbiol* 68(5):1128–1148.
17. Patrick WM, Quandt EM, Swartzlander DB, Matsumura I (2007) Multicopy suppression underpins metabolic evolvability. *Mol Biol Evol* 24(12):2716–2722.
18. Blank D, Wolf L, Ackermann M, Silander OK (2014) The predictability of molecular evolution during functional innovation. *Proc Natl Acad Sci USA* 111(8):3044–3049.
19. Yip SH-C, Matsumura I (2013) Substrate ambiguous enzymes within the Escherichia coli proteome offer different evolutionary solutions to the same problem. *Mol Biol Evol* 30(9):2001–2012.
20. Fernández De Henestrosa AR, et al. (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol* 35(6):1560–1572.
21. Little JW, Mount DW (1982) The SOS regulatory system of Escherichia coli. *Cell* 29(1): 11–22.
22. O'Byrne CP, Ní Bhriain N, Dorman CJ (1992) The DNA supercoiling-sensitive expression of the Salmonella typhimurium his operon requires the his attenuator and is modulated by anaerobiosis and by osmolarity. *Mol Microbiol* 6(17):2467–2476.
23. Schoemaker JM, Gayda RC, Markovitz A (1984) Regulation of cell division in Escherichia coli: SOS induction and cellular location of the sulA protein, a key to lon-associated filamentation and death. *J Bacteriol* 158(2):551–561.
24. Blattner FR, et al. (1997) The complete genome sequence of Escherichia coli K-12. *Science* 277(5331):1453–1462.
25. Lander ES, et al.; International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.
26. Venter JC, et al. (2001) The sequence of the human genome. *Science* 291(5507): 1304–1351.
27. Taylor TB, et al. (2015) Evolution. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science* 347(6225):1014–1017.
28. Li D, et al. (2010) A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. *Cell Res* 20(4): 408–420.
29. Li D, Yan Z, Lu L, Jiang H, Wang W (2014) Pleiotropy of the de novo-originated gene MDF1. *Sci Rep* 4:7280.
30. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* 489(7417):513–518.
31. Rodríguez-Verdugo A, Tenaillon O, Gaut BS (2016) First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes. *Mol Biol Evol* 33(1):25–39.
32. Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
33. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* 97(12):6640–6645.
34. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
35. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
36. Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
37. Blankenberg D, et al. (2010) Galaxy: A web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, eds Ausubel FM, et al. Chapter 19: Unit 19.10.11–21.
38. Giardine B, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455.
39. Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
40. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
41. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402–408.
42. Zaslaver A, et al. (2006) A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods* 3(8):623–628.
43. Lu W, et al. (2010) Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal Chem* 82(8): 3212–3221.
44. Melamud E, Vastag L, Rabinowitz JD (2010) Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* 82(23):9818–9826.
45. Stomel JM, Wilson JW, León MA, Stafford P, Chaput JC (2009) A man-made ATP-binding protein evolved independent of nature causes abnormal growth in bacterial cells. *PLoS One* 4(10):e7385.
46. Winkler ME, Ramos-Montañez S (2009) Biosynthesis of histidine. *Ecosal Plus* 3(2):1–34.
47. Frandsen N, D'Ari R (1993) Excess histidine enzymes cause AICAR-independent filamentation in Escherichia coli. *Mol Gen Genet* 240(3):348–354.