

assay (See Materials and Methods and Discussion for elaboration on these possible mechanisms).

My finding that SynSerB3 rescues the  $\Delta serB$  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., who did 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<sup>81</sup>. In the case of the  $\Delta serB$  auxotroph, Patrick et al. found that overexpression of HisB rescued the deletion. I have repeated this experiment, and observed the same result. In related experiments, Blank et al. searched for chromosomal mutations that rescue deletions<sup>82</sup>. In the case of  $\Delta serB$ , 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<sup>83</sup>, who reported a  $k_{cat}/K_M$  value of  $7.6 \text{ M}^{-1}\text{sec}^{-1}$ . 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 fact that HisB must be overexpressed in order for it to rescue  $\Delta serB$ . Indeed, the auxotrophy of the  $\Delta serB$  mutant shows explicitly that endogenous chromosomal expression of HisB is not sufficient to enable the growth of  $\Delta serB$  cells on minimal medium.

The experiments described above demonstrate that the *de novo* protein SynSerB3 increases expression of the *E. coli* gene, HisB. Since HisB encodes a phosphatase, these results suggested that rescue by SynSerB3 of the phosphoserine phosphatase deletion in  $\Delta serB$  is

Seeded Content – Center for Disease Control and Prevention *E.coli* (*Escherichia coli*)  
<https://www.cdc.gov/ecoli/general/index.html>

*Escherichia coli* (*E. coli*) bacteria normally live in the intestines of people and animals. Most *E. coli* are harmless and actually are an important part of a healthy human intestinal tract. However, some *E. coli* are pathogenic, meaning they can cause illness, either diarrhea or illness outside of the intestinal tract. The types of *E. coli* that can cause diarrhea can be transmitted through contaminated water or food, or through contact with animals or persons.

*E. coli* consists of a diverse group of bacteria. Pathogenic *E. coli* strains are categorized into pathotypes. Six pathotypes are associated with diarrhea and collectively are referred to as diarrheagenic *E. coli*.

- Shiga toxin-producing *E. coli* (STEC)—STEC may also be referred to as Verocytotoxin-producing *E. coli* (VTEC) or enterohemorrhagic *E. coli* (EHEC). This pathotype is the one most commonly heard about in the news in association with foodborne outbreaks.
- [Enterotoxigenic \*E. coli\* \(ETEC\)](#)
- Enteropathogenic *E. coli* (EPEC)
- Enteroaggregative *E. coli* (EAEC)
- Enteroinvasive *E. coli* (EIEC)
- Diffusely adherent *E. coli* (DAEC)

*Escherichia coli* (abbreviated as *E. coli*) are a large and diverse group of bacteria. Although most strains of *E. coli* are harmless, others can make you sick. Some kinds of *E. coli* can cause diarrhea, while others cause urinary tract infections, respiratory illness and pneumonia, and other illnesses. Still other kinds of *E. coli* are used as markers for water contamination—so you might hear about *E. coli* being found in drinking water, which are not themselves harmful, but indicate the water is contaminated. It does get a bit confusing—even to microbiologists.

Some kinds of *E. coli* cause disease by making a toxin called Shiga toxin. The bacteria that make these toxins are called “Shiga toxin-producing” *E. coli*, or STEC for short. You might hear these

bacteria called verocytotoxic *E. coli* (VTEC) or enterohemorrhagic *E. coli* (EHEC); these all refer generally to the same group of bacteria. The strain of Shiga toxin-producing *E. coli* O104:H4 that caused a large outbreak in Europe in 2011 was frequently referred to as EHEC. The most commonly identified STEC in North America is *E. coli* O157:H7 (often shortened to *E. coli* O157 or even just “O157”). When you hear news reports about outbreaks of “*E. coli*” infections, they are usually talking about *E. coli* O157.

In addition to *E. coli* O157, many other kinds (called serogroups) of STEC cause disease. Other *E. coli* serogroups in the STEC group, including *E. coli* O145, are sometimes called "non-O157 STECs." Currently, there are limited public health surveillance data on the occurrence of non-O157 STECs, including STEC O145; many STEC O145 infections may go undiagnosed or unreported.

Compared with STEC O157 infections, identification of non-O157 STEC infections is more complex. First, clinical laboratories must test stool samples for the presence of Shiga toxins. Then, the positive samples must be sent to public health laboratories to look for non-O157 STEC. Clinical laboratories typically cannot identify non-O157 STEC. Other non-O157 STEC serogroups that often cause illness in people in the United States include O26, O111, and O103. Some types of STEC frequently cause severe disease, including bloody diarrhea and hemolytic uremic syndrome (HUS), which is a type of kidney failure.

Most of what we know about STEC comes from studies of *E. coli* O157 infection, which was first identified as a pathogen in 1982. Less is known about the non-O157 STEC, partly because older laboratory practices did not identify non-O157 infections. As a whole, the non-O157 serogroups are less likely to cause severe illness than *E. coli* O157, though sometimes they can. For example, *E. coli* O26 produces the same type of toxins that *E. coli* O157 produces, and causes a similar illness, though it is typically less likely to lead to kidney problems (called hemolytic uremic syndrome, or HUS).

People of any age can become infected. Very young children and the elderly are more likely to develop severe illness and hemolytic uremic syndrome (HUS) than others, but even healthy older children and young adults can become seriously ill.

The symptoms of STEC infections vary for each person but often include severe stomach cramps, diarrhea (often bloody), and vomiting. If there is fever, it usually is not very high (less than 101°F/less than 38.5°C). Most people get better within 5–7 days. Some infections are very mild, but others are severe or even life-threatening.

Around 5–10% of those who are diagnosed with STEC infection develop a potentially life-threatening complication known as hemolytic uremic syndrome (HUS). Clues that a person is developing HUS include decreased frequency of urination, feeling very tired, and losing pink color in cheeks and inside the lower eyelids. Persons with HUS should be hospitalized because their kidneys may stop working and they may develop other serious problems. Most persons with HUS recover within a few weeks, but some suffer permanent damage or die.

The time between ingesting the STEC bacteria and feeling sick is called the “incubation period.” The incubation period is usually 3–4 days after the exposure, but may be as short as 1 day or as long as 10 days. The symptoms often begin slowly with mild belly pain or non-bloody diarrhea that worsens over several days. HUS, if it occurs, develops an average 7 days after the first symptoms, when the diarrhea is improving.

STEC live in the guts of ruminant animals, including cattle, goats, sheep, deer, and elk. The major source for human illnesses is cattle. STEC that cause human illness generally do not make animals sick. Other kinds of animals, including pigs and birds, sometimes pick up STEC from the environment and may spread it.

Infections start when you swallow STEC—in other words, when you get tiny (usually invisible) amounts of human or animal feces in your mouth. Unfortunately, this happens more often than we would like to think about. Exposures that result in illness include consumption of contaminated food, consumption of unpasteurized (raw) milk, consumption of water that has not been disinfected, contact with cattle, or contact with the feces of infected people. Some foods are considered to carry such a high risk of infection with *E. coli* O157 or another germ that health officials recommend that people avoid them completely. These foods include unpasteurized (raw) milk, unpasteurized apple

cider, and soft cheeses made from raw milk. Sometimes the contact is pretty obvious (working with cows at a dairy or changing diapers, for example), but sometimes it is not (like eating an undercooked hamburger or a contaminated piece of lettuce). People have gotten infected by swallowing lake water while swimming, touching the environment in petting zoos and other animal exhibits, and by eating food prepared by people who did not wash their hands well after using the toilet. Almost everyone has some risk of infection.

Because there are so many possible sources, for most people we can only guess. If your infection happens to be part of the about 20% of cases that are part of a recognized outbreak, the health department might identify the source.

An estimated 265,000 STEC infections occur each year in the United States. STEC O157 causes about 36% of these infections, and non-O157 STEC cause the rest. Public health experts rely on estimates rather than actual numbers of infections because not all STEC infections are diagnosed, for several reasons. Many infected people do not seek medical care; many of those who do seek care do not provide a stool specimen for testing, and many labs do not test for non-O157 STEC. However, this situation is changing as more labs have begun using newer, simpler tests that can help detect non-O157 STEC.

STEC infections are usually diagnosed through laboratory testing of stool specimens (feces).

Identifying the specific strain of STEC is essential for public health purposes, such as finding outbreaks. Many labs can determine if STEC are present, and most can identify *E. coli* O157. Labs that test for the presence of Shiga toxins in stool can detect non-O157 STEC infections. However, for the O group (serogroup) and other characteristics of non-O157 STEC to be identified, Shiga toxin-positive specimens must be sent to a state public health laboratory.

Contact your healthcare provider if you have diarrhea that lasts for more than 3 days, or is accompanied by high fever, blood in the stool, or so much vomiting that you cannot keep liquids down and you pass very little urine.

allowed the extremely abundant Omp proteins to stick to SynSerB3 non-specifically in the pulldown reaction.

## DISCUSSION

The aim of this project was to determine if *de novo* designed proteins that have not endured natural selection in a cellular environment could function productively in the cell. At the outset, we anticipated that the novel protein SynSerB3 would rescue  $\Delta serB$  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. In the case presented here, SynSerB3 is regulating the endogenous histidine biosynthetic operon. This finding is entirely consistent with the fact that we selected for growth, not phosphoserine phosphatase activity.

In addition, I found many side effects of SynSerB3 expression in cells, namely, induction of the SOS response. This is a firm reminder of the importance of negative selection in evolution. Naturally-occurring proteins have not only evolved to perform a specialist function very well, but they also have minimized off-target and wasteful functions, and other burdensome effects that reduce the cells overall fitness. In the case of SynSerB3, induction of the SOS response leads to impaired growth and division, and upregulation of mutagenic DNA polymerases (UmuCD, DinB).

The aberrant cellular morphology caused by SynSerB3 has been observed in other studies with *de novo* proteins as well. Stomel *et al.* reported that expression in *E. coli* of an artificial ATP binding protein DX also caused extensive filamentation<sup>13,37</sup>. ATP sequestration by an

artificial protein or by excessive histidine biosynthesis (this study) (which requires 41 ATP molecules per molecule of histidine<sup>75</sup>) could cause filamentation in addition to the SOS response. Since SynSerB3 and DX were selected for functions regardless of cellular morphology, the filamentation phenotype is probably not an intrinsic phenotype of *de novo* protein expression.

There is also evidence that overexpression of the histidine biosynthetic enzymes HisH and HisF can cause filamentation in *E. coli*<sup>87</sup>. However, the data presented here shows that pseudo-WT cells expressing SynSerB3 do overexpress His biosynthetic enzymes but do not cause filamentation. These results indicate that the aberrant morphology apparently results from the overall stress associated with growing the deletion strain in minimal medium, while relying on a relatively inefficient rescue mechanism.

The data show that SynSerB3 interferes with the natural regulation of the *his* operon by deattenuating, or derepressing, the *his* operon. Many studies have explored derepression of the *his* operon in *E. coli* and *S. typhimurium*. In 1987, five loci unlinked to the *his* operon were listed that derepress the *his* operon, named hisR, S, T, U, and W<sup>88</sup>. HisR is the sole <sup>his</sup>tRNA in the cell, HisS is the histidinyl tRNA synthetase. HisT is tRNA pseudouridine synthase and has been renamed TruA. HisU and HisW are GyrB and GyrA respectively, which encode the two subunits of DNA gyrase. These genes derepress the *his* operon by decreasing the functional pool of histidinyl-<sup>his</sup>tRNA. The functional pool of histidinyl-<sup>his</sup>tRNA is decreased by inhibiting its synthesis/transcription (*hisR*, *gyrA*, and *gyrB*), inhibiting its proper charging (*hisS*), or by inhibiting its proper modification (*truA*). The decreased pool of charged histidinyl-<sup>his</sup>tRNA stalls the ribosome on the leader peptide mRNA, blocking attenuator formation, and results in derepression of the *his* operon. Since SynSerB3 derepresses the *his* operon, it is reasonable to assume that it functions like one of the aforementioned mutations to decrease the functional pool

of histidyl-<sup>his</sup>tRNA. This could occur by blocking (directly or allosterically) catalytic function of one of the <sup>his</sup>tRNA-modifying enzymes, which has not been tested directly. Alternatively, SynSerB3 could function by altering the expression of one of the aforementioned genes; upon examination of the RNAseq data, the expression of all of these genes is unaltered.

To indirectly test whether SynSerB3 is interfering with a <sup>his</sup>tRNA-modifying enzyme, we used a coupled transcription-translation assay that measures where ribosomes are stalled on a mRNA transcript<sup>89</sup>. In particular, we assayed where ribosomes stall on the *hisL* leader mRNA in the presence of SynSerB3. Assuming SynSerB3 was artificially lowering the amount of charged-<sup>his</sup>tRNA, it could be expected that the ribosome would be stalled prior to the stretch of histidines encoded in the *hisL* leader peptide. The results obtained thus far show that SynSerB3 has no effect on the efficiency of *hisL* translation. From these results I conclude that SynSerB3 is not altering the pool of charged-<sup>his</sup>tRNA. This experiment was done in collaboration with Dr. Nora Vazquez-Laslop (See Materials and Methods).

The second alternative is that SynSerB3 blocks attenuator formation by binding the mRNA. To the best of our knowledge, no transcription factors bind to the attenuator sequence DNA or mRNA. On the other hand, there is speculation that the attenuator may be a genetic derivative of the <sup>his</sup>tRNA<sup>90</sup>, and therefore in theory, any proteins that bind the <sup>his</sup>tRNA could bind the attenuator region, such as the <sup>his</sup>tRNA modifying enzymes or HisG. None of this has been experimentally confirmed. Experiments that determine *in vivo* binding partners to the *his* operon attenuator could be interesting, with and without SynSerB3 present.

Do the other SynSerB proteins rescue in the same way as SynSerB3? We don't have conclusive evidence, but none of the SynSerB proteins can rescue the double knockout  $\Delta serB \Delta hisB$  (even when supplemented with histidine). Therefore, it is reasonable to assume that



SynSerB1, 2, and 4 rescue  $\Delta serB$  in the same way as SynSerB3. However, there may be subtle differences in how the SynSerB proteins effect cells overall. For example, SynSerB3 does not express well but rescues  $\Delta serB$  in minimal medium. But the other SynSerB proteins robustly grow on solid medium but not in liquid medium, and express well. Once a biochemical assay can quantify binding of SynSerB3 to something, then the other SynSerB proteins can be assayed for that activity as well.

Here I would like to raise an interesting question: does (i) SynSerB3 indirectly induce *his* operon expression or does it (ii) specifically induce *his* operon expression with confounding indirect effects? This is a difficult question to answer without more temporal proteomic/transcriptomic/metabolomics data. As stated previously, *his* operon expression is tightly regulated for a reason: it is energy-expensive to make histidine. For this reason, there are probably many side-effects to overexpressing this operon. In addition, out of 1.5 million *de novo* proteins that we screened, only four rescued  $\Delta serB$ . Presumably, any of these proteins could interfere with cellular metabolism so much as to render some indirect effect sufficient to rescue  $\Delta serB$ , yet only four did. Indeed, single-timepoint integrated gene and metabolite analysis do not show that histidine biosynthesis is the only thing altered by expression of SynSerB3 (Figure 2.17). A biochemical assay is critical to determine if SynSerB proteins function directly.

In conclusion, 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  $\Delta serB$  cells on minimal medium. After showing that HisB was sufficient for rescue, it was confirmed to be necessary by demonstrating that the ability

114. Ozenberger, B. A.; Nahlik, M. S.; McIntosh, M. A., Genetic organization of multiple *fep* genes encoding ferric enterobactin transport functions in *Escherichia coli*. *J Bacteriol* **1987**, *169* (8), 3638-46.
115. Neilands, J. B., Microbial envelope proteins related to iron. *Annu Rev Microbiol* **1982**, *36*, 285-309.
116. Slonczewski, J. L.; Rosen, B. P.; Alger, J. R.; Macnab, R. M., pH homeostasis in *Escherichia coli*: measurement by <sup>31</sup>P nuclear magnetic resonance of methylphosphonate and phosphate. *Proc Natl Acad Sci U S A* **1981**, *78* (10), 6271-5.
117. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T., Cell-free translation reconstituted with purified components. *Nat Biotechnol* **2001**, *19* (8), 751-5.
118. Patel, R. N., Chemo-enzymatic synthesis of pharmaceutical intermediates. *Expert Opin Drug Discov* **2008**, *3* (2), 187-245.
119. Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H., Directed evolution of the tryptophan synthase  $\beta$ -subunit for stand-alone function recapitulates allosteric activation. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112* (47), 14599-604.
120. Brock, M.; Maerker, C.; Schütz, A.; Völker, U.; Buckel, W., Oxidation of propionate to pyruvate in *Escherichia coli*. Involvement of methylcitrate dehydratase and aconitase. *Eur J Biochem* **2002**, *269* (24), 6184-94.
121. Sasson, S. J. A bacterial surface display library for the high-throughput screening and selection of De Novo designed four-helix bundle proteins. Thesis (Senior)-Princeton University, 2009, Princeton University, Dept. of Chemistry,, Princeton, NJ, 2009.