Biochemistry
Article
Cite This: *Biochemistry* XXXX, XXX, XXX−XXX
pubs.acs.org/biochemistry

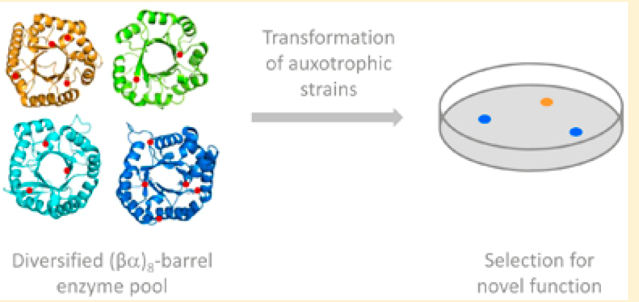
Library Selection with a Randomized Repertoire of $(\beta\alpha)_8$-Barrel Enzymes Results in Unexpected Induction of Gene Expression

Bettina Rohweder,[†] Gerhard Lehmann,[‡] Norbert Eichner,[‡] Tino Polen,[§] Chitra Rajendran,[†] Fabian Ruperti,[†] Mona Linde,[†] Thomas Treiber,[‡] Oona Jung,[∥] Katja Dettmer,[∥] Gunter Meister,[‡] Michael Bott,[§] Wolfram Gronwald,[∥] and Reinhard Sterner*[,†]

[†]Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany
[‡]Institute of Biochemistry, Genetics and Microbiology, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany
[§]IBG-1: Biotechnology, Institute of Bio- and Geosciences, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
[∥]Institute of Functional Genomics, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

Supporting Information

ABSTRACT: The potential of the frequently encountered $(\beta\alpha)_8$-barrel fold to acquire new functions was tested by an approach combining random mutagenesis and selection *in vivo*. For this purpose, the genes encoding 52 different phosphate-binding $(\beta\alpha)_8$-barrel proteins were subjected to error-prone PCR and cloned into an expression plasmid. The resulting mixed repertoire was used to transform different auxotrophic *Escherichia coli* strains, each lacking an enzyme with a phosphate-containing substrate. After plating of the different transformants on minimal medium, growth was observed only for two strains, lacking either the gene for the serine phosphatase SerB or the phosphoserine aminotransferase SerC. The same mutants of the *E. coli* genes *nan*E (encoding a putative *N*-acetylmannosamine-6-phosphate 2-epimerase) and *pdx*J (encoding the pyridoxine 5′-phosphate synthase) were responsible for rescuing both Δ*ser*B and Δ*ser*C. Unexpectedly, the complementing NanE and PdxJ variants did not catalyze the SerB or SerC reactions *in vitro*. Instead, RT-qPCR, RNAseq, and transcriptome analysis showed that they rescue the deletions by enlisting the help of endogenous *E. coli* enzymes HisB and HisC through exclusive up-regulation of histidine operon transcription. While the promiscuous SerB activity of HisB is well-established, our data indicate that HisC is promiscuous for the SerC reaction, as well. The successful rescue of Δ*ser*B and Δ*ser*C through point mutations and recruitment of additional amino acids in NanE and PdxJ provides another example for the adaptability of the $(\beta\alpha)_8$-barrel fold.



The mechanisms underlying the acquirement of new enzymatic functions during evolution are still debated. One established scenario is the duplication of existing genes. This duplication is followed by mutational diversification in one of the daughter genes that leads to the enhancement of an initially low promiscuous activity.[1] Hence, the enzymatic repertoire of a cell such as *Escherichia coli* could contain many promiscuous members. This hypothesis was tested by a comprehensive experiment in which researchers tried to complement the growth deficiency of 104 *E. coli* deletions strains through systematic overexpression of the whole *E. coli* proteome ("multi-copy suppression experiment").[2] For this purpose, a complete set of 4123 *E. coli* K-12 genes (the ASKA library),[3] which were cloned into an expression vector, was used. For as many as 21 deletion strains, growth could be restored not only by expressing the respective wild-type gene but also due to the expression of other *E. coli* genes encoded on an ASKA plasmid. A method for predicting promiscuous functions, which is based on an unsupervised BLAST-search, successfully recapitulated known multicopy suppression events and predicted additional ones, several of which were validated *in vitro*.[4]

We wanted to pursue a similar approach, as described.[2] However, instead of using wild-type proteins to restore growth of deletion strains, we aimed to increase the tested sequence space through random mutation further. In order to keep library size manageable, we focused on enzymes belonging to the $(\beta\alpha)_8$-barrel fold (Figure 1A). The $(\beta\alpha)_8$-barrel or TIM barrel fold is the most common enzyme fold and is found in five of the six enzyme classes defined by the Enzyme Commission (oxidoreductases, transferases, lyases, hydrolases, and isomerases), illustrating its high adaptability and versatility.[5] It is therefore interesting to check the evolutionary potential of this fold by laboratory experiments. Due to their

Received: July 7, 2019
Revised: September 25, 2019
Published: September 26, 2019

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

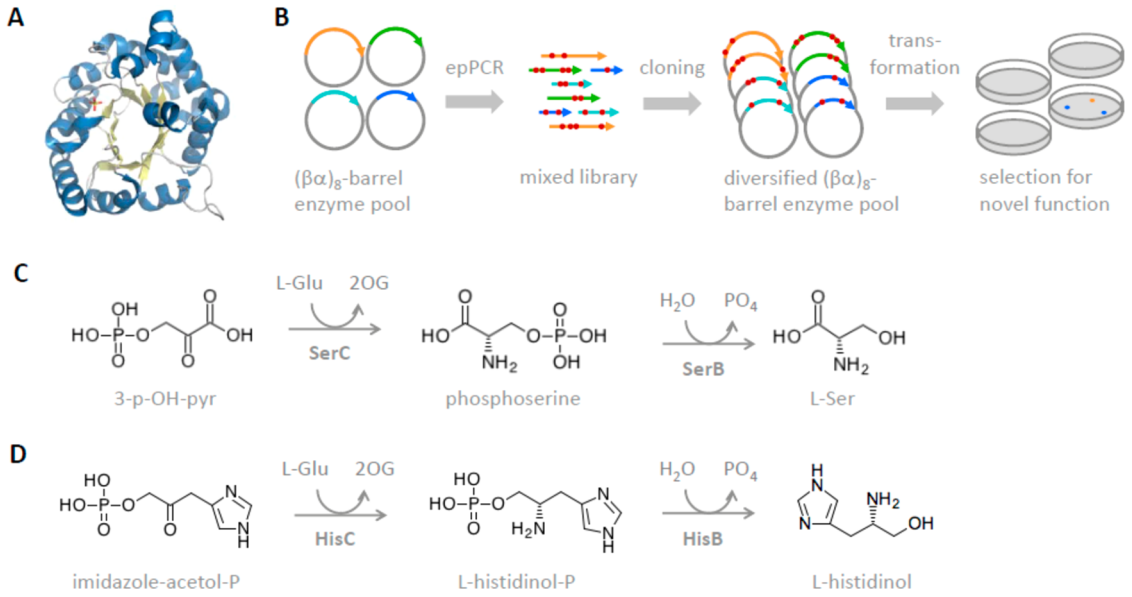


Figure 1. (A) Cartoon representation of the name-giving member of the $(\beta\alpha)_8$-barrel enzyme superfamily, triose phosphate isomerase TIM (PDB code 4IOT). The eight central β-sheets forming the barrel structure are colored yellow, while the eight surrounding helices are colored blue. A sulfate ion that occupies the phosphate-binding pocket formed by main chain NH-groups of active-site loops is shown as a stick model. (B) Overview of library generation and selection for novel function by complementation of auxotrophic *E. coli* strains. (C) Reactions catalyzed by phosphoserine aminotransferase (SerC) and phosphoserine phosphatase (SerB). Abbreviations: 3-*p*-OH-pyr: 3-phosphohydroxypyruvate; Glu: glutamate; 2OG: 2-oxoglutarate. (D) The reactions catalyzed by histidinol-phosphate aminotransferase (HisC) and L-histidinol-phosphate phosphohydrolase (HisB) are chemically similar to the reactions catalyzed by SerC and SerB.

intrinsic stability to mutations, $(\beta\alpha)_8$-barrels have been proven ideal targets for directed evolution or protein design approaches.[6,7] We further limited our initial gene pool to $(\beta\alpha)_8$-barrel enzymes containing a phosphate-binding site in their active center. Phosphate is the most common functional group in the *E. coli* metabolome, with 35−40% of all metabolites containing a phosphate group.[8] We reasoned that this existing binding site could function as an anchor that helps to bind novel substrates and thus facilitates the enhancement of promiscuous activity levels beyond the detection limit of the growth assays.

Hence, the ASKA library plasmids corresponding to 52 phosphate-binding $(\beta\alpha)_8$-barrel proteins from *E. coli* were pooled, diversified by error-prone PCR, and used for the selection of novel functions through complementation of auxotrophic *E. coli* deletion strains. Cell growth was observed for two strains, lacking either the gene for the serine phosphatase SerB or the phosphoserine aminotransferase SerC. It turned out that both, the mutated gene *pdx*J (encoding the pyridoxine 5′-phosphate synthase) and the mutated gene *nan*E (encoding a putative *N*-acetylmannosamine-6-phosphate 2-epimerase) were responsible for rescuing both Δ*ser*B and Δ*ser*C. However, instead of catalyzing the SerB and SerC reactions, the selected PdxJ and NanE variants unexpectedly induced the expression of the histidine operon, thus enlisting the help of the endogenous *E. coli* enzymes histidinol phosphatase HisB and histidinol phosphate aminotransferase HisC. While it is established that HisB is promiscuous for the SerB reaction[2,9] our data shows as well that HisC is promiscuous for the SerC reaction.

MATERIALS AND METHODS

Oligonucleotides, Growth Media, and Reagents. Oligonucleotides were obtained from biomers.net. The ASKA library-gfp (a complete set of *E. coli* K-12 ORF Archive)[3] was obtained from the National Institute of Genetics (1111, Yata, Mishima, Shizuoka 411-8540 Japan). Strains were grown in LB medium or M9-glucose minimal medium (1× M9-salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% Glucose). Antibiotics and additional substrates were added to the following concentrations: ampicillin (150 μg/mL), kanamycin (50 μg/mL), chloramphenicol (30 μg/mL), IPTG (0.5 mM), pyridoxine (0.1 mg/L), L-histidine, L-histidinol, L-alanine, L-threonine, L-lysine, sodium acetate, propionic acid, maltotriose, and 2,6-diaminopimelic acid (12 μg/mL). For Δ*ser*C deletions, pyridoxine was added to the plates (Table S1). SerC has two enzymatic functions: Apart from serine biosynthesis, it also catalyzes a step in the biosynthesis of pyridoxal-5-phosphate. For Δ*his*B and Δ*his*C deletions, L-histidinol was added (Table S1), to allow selection for complementation in the absence of HisB or HisC, respectively. We used L-histidinol (the product of the HisB reaction) instead of L-histidine, as L-histidine functions as a native regulator that downregulates the expression of the *his* operon. (With L-histidine added to the growth plate, wild-type HisB fails to complement Δ*ser*BΔ*his*B.) Plates were kept at 37 °C up to 4 weeks and checked regularly for the appearance of colonies. For growth on solid media, 15 mg/mL agar was added to the growth medium. ^{14}C-phosphoserine was obtained from Hartmann Analytics. All other reagents were purchased from Sigma.

Strains. The parent strain *E. coli* BW25113 and the deletion strains Δ*ser*B, Δ*ser*C, Δ*aro*A, Δ*aro*C, Δ*pdx*A, Δ*pur*E, Δ*pyr*B, and Δ*ser*A in the BW25113 background (all *gene*::kan), were obtained from the National BioResource Project (NIG, Japan): *E. coli* (1111, Yata, Mishima, Shizuoka 411−8540 Japan).[10] All other deletions strains (Δ*ser*BΔ*his*B [Δ*ser*B::kan Δ*his*B::cat], Δ*ser*BΔ*gph* [Δ*ser*B::kan Δ*gph*::*cat*], Δ*ser*BΔ*ytj*C [Δ*ser*B::kan Δ*ytj*C], Δ*ser*BΔ*ser*C [Δ*ser*B::kan Δ*ser*C], and Δ*ser*CΔ*his*C [Δ*ser*C Δ*his*C::kana]) were constructed from the parent strain BW25113 by the P1 transduction method.[11]

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

Primers used for strain construction and PCR confirmation are listed in Table S2.

Library Construction and *in Vivo* Complementation. ASKA library clones (vector pCA24N-gfp) were grown overnight in LB medium with added chloramphenicol. All 52 plasmids (complete list in Table S3) harboring a gene encoding for a phosphate-binding $(\beta\alpha)_8$-barrel were prepared individually (GeneJET Plasmid Miniprep Kit, Thermo Fisher) and subsequently mixed in equimolar concentrations. This pool (150 ng/μL) was used as a template for the randomization by the error-prone PCR reaction. For the PCR reaction, 50 ng DNA template, 15 U GoTaq, 1× GoTaq buffer, 0.8 mM $MnCl_2$, 1 mM $MgCl_2$, 0.35 mM dATP, 0.40 mM dCTP, 0.20 mM dGTP, 1.35 mM dTTP, and 0.2 μM of each primer were added to a total volume of 50 μL. Thirty PCR cycles were performed (denaturation 95 °C for 45 s, annealing 56 °C for 45 s, elongation 72 °C for 5 min). The resulting mixed repertoire was cloned into the expression plasmid pTNA_*Bsa*I[12] allowing for low, constitutive expression in *E. coli*. The plasmids were used to transform *E. coli* BW25113 cells, which were then plated on medium containing ampicillin and incubated at 37 °C. Grown colonies were scratched off the plates and suspended, followed by the isolation of the $(\beta\alpha)_8$-barrel enzyme library. As estimated from the number of grown colonies and the ligation efficiency tested by colony PCR, the library contained about 1.4×10^8 independent mutants. To test for *in vivo* complementation, electrocompetent cells from nine different auxotrophic strains were transformed with the diversified $(\beta\alpha)_8$-barrel enzyme library, according to standard protocols. After recovery (1 h shaking in SOC), cells were washed three times with 1% NaCl and plated on solid M9-glucose minimal medium supplemented with additives if necessary and incubated at 37 °C.

Site-Directed Mutagenesis. Single point mutations and frameshift mutations were introduced by the QuickChange Site-Directed Mutagenesis System (QCM) protocol developed by Stratagene (La Jolla, CA) with the primers listed in Table S2. For the purification of FLAG-tagged proteins PdxJ(X3) and wild-type PdxJ(+Ext), the corresponding genes were cloned from pTNA_BsaI into vector pUR21 (Table S4) and an N-terminal Flag-tag was introduced with the forward primer.

Protein Expression, Purification, and Characterization. Selected mutant genes were recloned for expression into vector pET21a and used to transform *E. coli* BL21 cells. The LB medium (1 L) was inoculated with an overnight culture and induced with IPTG at an optical density at 600 nm of about 0.5. Protein expression occurred overnight at 20 °C. Cells were harvested and disrupted by sonification in 50 mM Tris-HCl (pH 7.5), 300 mM KCl, and 20 mM imidazole. Proteins were purified from the soluble fraction of the cell extracts by Ni^{2+}-affinity (all variants retain the N-terminal hexahistidine tag of the ASKA library) in 50 mM Tris-HCl (pH 7.5), 300 mM KCl, using a linear gradient from 20 mM to 500 mM imidazole for protein elution followed by preparative size exclusion chromatography (25 mM Tris-HCl (pH 7.5), 300 mM KCl). Fractions containing at least 95% pure protein as judged by SDS-PAGE were pooled, dialyzed against 25 mM Tris-HCl (pH 7.5), 100 mM KCl, and stored at −80 °C after flash-freezing in liquid nitrogen. Protein concentrations were determined by absorption spectroscopy using a molar extinction coefficient at 280 nm that was calculated from the amino acid sequence.[13]

Enzyme Activity Assays. SerB activity was assayed by monitoring released phosphate with a malachite green assay, performed with purified protein, as described previously.[14] Additionally, we measured phosphate release from ^{14}C-phosphoserine in a radiometric assay, and 1−4 μM purified protein was incubated with 4.5 μM ^{14}C-phosphoserine in 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/mL BSA (pH 7.9) for up to 15 h at 37 °C. The products were analyzed and visualized, as described previously.[15]

Protein Crystallization. PdxJ(11) was crystallized using standard hanging drop vapor diffusion methods. The protein (9.6 mg/mL in 10 mM Tris-HCl (pH 7.5), 30 mM NaCl) was supplemented with 5 mM phosphoserine, incubated at 4 °C for 15 min before adding the crystallization buffer (7.5% PEG 6000 and 1.7 M NaCl), and incubation was continued at 18 °C. Crystals of PdxJ(11) grew within 1 week. After transferring the crystals to 30% ethylene glycol as cryoprotectant, they were flash-frozen in liquid nitrogen.

Data Collection, Structure Solution, and Refinement. Data collection was done at Swiss Light Source (SLS), Switzerland at beamline PXIII and PXI at cryogenic temperature. The data processing was done using XDS,[16] and the data quality assessment was done using phenix.xtriage. Molecular replacement was performed with Phaser[17] within the CCP4i[18] suite using 1HO1 and 1M5W as search models. Initial refinement was performed using REFMAC.[19] The model was further improved in several refinement rounds using automated restrained refinement with the program PHENIX[20] and interactive modeling with Coot.[21] The refinement statistics are listed in Table S5. The final model was analyzed using the program MolProbity[22] and deposited in the Protein Data Bank (6RG0).

RT-qPCR. Cells were grown in LB medium to mid log phase. After harvesting, cells were washed three times with 1% NaCl and used to inoculate minimal M9-glucose medium and shaken at 30 °C. After the culture reached an optical density at 600 nm of 0.5 cells were mixed with two volumes of RNAprotect Bacteria Reagent (Qiagen), harvested and stored at −80 °C. Cells were lysed in RNase-free TE buffer (30 mM Tris-HCl, 1 mM EDTA, pH 8.0) containing 1 mg/mL lysozyme for 30 min at room temperature. Total RNA was prepared using an RNeasy Mini Kit (Qiagen). cDNA was prepared from 1 μg RNA with the QuantiTect Reverse Transcription Kit (Qiagen). Genomic DNA was eliminated prior to the reverse transcriptase reaction according to the Kit's protocol. RT-qPCR reactions were carried out in a final volume of 20 μL containing 1 μL of 1:5 diluted cDNA, 1 μL of 10 μM stock of each primer (primer pairs listed in Table S2), and 10 μL of SsoFast EvaGreen Supermix (BIO-RAD). The reactions were measured in a Bio-Rad Cfx96 qPCR system using 40 cycles of a two-step amplification protocol, with two technical replicates per sample. To determine relative expression levels, transcript levels were normalized to the 23S RNA. Samples without reverse transcriptase enzyme were used to control for specificity.

RNAseq. Cells were grown in LB medium to mid log phase. After harvesting, cells were washed three times with 1% NaCl and used to inoculate minimal M9-glucose medium and shaken at 30 °C. After the culture reached an optical density of 0.5 cells were mixed with two volumes of RNAprotect Bacteria Reagent (Qiagen), harvested and stored at −80 °C. Cells were lysed in RNase-free TE buffer (30 mM Tris-HCl, 1 mM

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

EDTA, pH 8.0) containing 1 mg/mL lysozyme for 30 min at room temperature. Total RNA was prepared using an RNeasy Mini Kit (Qiagen). Genomic DNA was removed by RNase-free DNase (Qiagen). The RNA quality was assayed using a NanoDrop and a 4200 TapeStation system (Agilent Technologies Inc.). Ribosomal RNA was depleted using a Ribo-Zero rRNA Removal Kit Bacteria (Illumina) and successful rRNA depletion verified by TapeStation. The library was prepared using a NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) with NEBNext Multiplex Oligos for Illumina (Index Primer Set 1) (NEB). After quality and quantity assessment on the TapeStation, the libraries were pooled equimolar. The resulting library pool was quantified with KAPA library quantification kit (Roche) and finally sequenced on a MiSeq sequencer (Illumina) performing a 2 × 80 cycles paired-end run. Of the sequences obtained, the adapters were trimmed with SeqPrep v1.3.2−1 (github.com/jstjohn/SeqPrep). The trimmed reads were mapped using the Bowtie 2 aligner[23] with default settings to the *E. coli* genome (www.ecogene.org). Samtools v1.3.1[24] was used to remove optical duplicates. For counting HTseq-count v0.9.1[25] was used against an annotation database (www.ecogene.org_Ecoli_annotation_110217-065029). Differential analysis was carried out with R (www.R-project.org) and the Bioconductor[26] package DESeq2.[27]

DNA Microarray Analysis. *E. coli* BW25113 cells were grown in minimal M9-glucose medium (with no additional supplements added) as described above. When the culture reached an optical density of 0.5, cells were harvested by centrifugation (3 min, 5000*g*) with crushed ice to ensure rapid cooling. The pelleted cells were shock-frozen in liquid nitrogen and stored at −80 °C. Total RNA was isolated using the RNeasy system (Qiagen) and quality-checked as described previously.[28] The synthesis of Cy3- or Cy5-labeled cDNA and two-color hybridization to DNA microarrays was carried out, as described.[28] Custom-made 4 × 44K DNA microarrays were obtained from Agilent Technologies and were designed using Agilent's eArray platform (https://earray.chem.agilent.com/earray). The array design comprises oligonucleotides for the annotated genes of *E. coli* MG1655 and other bacterial genomes, as well as Agilent's control spots. After hybridization, the arrays were washed using Agilent's wash buffer kit, and the fluorescence was determined at 532 nm (Cy3-dUTP) and 635 nm (Cy5-dUTP) at 5 μm resolution (GenePix 4000B laser scanner, GenePix Pro 6.0 software, Molecular Devices). After quantitative image analysis using the corresponding Agilent's gene array list, results containing the non-normalized ratio of median values were saved as GPR file (GenePix Pro 6.0). Subsequently, BioConductor R-packages limma and marray (http://www.bioconductor.org) were used to achieve background correction, loess-normalization of ratios, and diagnostic plot generation for array quality control. To check for differentially expressed genes according to the normalized ratios the quality criteria (i) Flags ≥ 0 and (ii) signal/noise of Cy5 (F635Median/B635Median) or Cy3 (F532Median/B532Median) ≥ 3 were applied. For each comparison *pdxJ*(11)/*pdxJ*(11-Ext) and *pdxJ*(X3)/wild-type *pdxJ*(+Ext) two independent biological samples obtained from separate cultures were used.

Protein Complex Immunoprecipitation (Co-IP). The FLAG-tagged proteins PdxJ(X3) and wt PdxJ(+Ext) were overexpressed for 9h in *E. coli* BW25113 cells in minimal M9-glucose medium (with no additional supplements added). For Co-IP, anti-FLAG M2 agarose (Sigma-Aldrich, Deisenhofen) was used. Prior to use, the matrix was washed twice with cold PBS and once with the lysis buffer (150 mM KCl, 25 mM Tris pH 7.5, 2 mM EDTA, 1 mM NaF, 0.5% NP-40, 1 mM DTT, 1 mM AEBSF). The *E. coli* cultures were washed twice in ice-cold PBS and lysed by resuspending in lysis buffer. Input samples of the lysates were taken, mixed with SDS-PAGE sample buffer, and stored at −20 °C. For immunoprecipitation, 1 mL of the lysate was incubated with 50 μL of the antibody-coupled beads for 2 h while rotating at 4 °C. After incubation, the affinity matrix was sedimented by centrifugation for 1 min at 1000*g*, and the supernatant was removed. The beads were incubated four times with the wash buffer (300 mM NaCl, 50 mM Tris (pH 7.5), 0.5% NP-40, 1 mM $MgCl_2$). After the third washing step, the samples were transferred to new reaction tubes to minimize contamination by unspecific protein binding to the tube material. After a final washing step with PBS, the supernatant was removed quantitatively, and the beads were eluted by addition of 5 μL of 3× Flag peptide (Sigma-Aldrich, Deisenhofen) and 35 μL of PBS and rotating for 2 h at 4 °C. After elution, the affinity matrix was sedimented by centrifugation for 1 min at 1000*g*, and the supernatant was transferred into a new cup. A second elution step with 20 μL of 0.1 M glycine (pH 2.6) for 5 min at room temperature (stopped by the addition of 1 μL of 1 M NaOH), was applied. Again, the affinity matrix was sedimented by centrifugation for 1 min at 1000*g* and the supernatant was transferred into a new cup. The input (about 15 μg) and the supernatants of the elution steps were analyzed by SDS-PAGE. Of the supernatant of the first elution step (approximately 50 μL), 25 μL were loaded into a gel pocket, and voltage was applied. After near-complete migration of the applied volume into the stacking gel, the remaining 25 μL of the first elution step were loaded into the same gel pocket. Detected protein bands were cut out, digested with trypsin and analyzed via mass spectrometry.

NMR Metabolomics Analysis. *E. coli* BW25113 cells were grown in minimal M9-glucose medium (with no additional supplements added) as described above. Pelleted cells were washed three times with 1% NaCl. Cell pellet and liquid culture supernatant were stored at −20 °C. For metabolite extraction, cell pellets were incubated overnight in 600 μL of 80% aqueous methanol. To each sample, 10 μL of 80 mM nicotinate solution were added as extraction standard followed by centrifugation at 9000*g* for 5 min. Supernatants were collected, and remaining pellets were resuspended twice in 200 μL of 80% aqueous methanol followed by centrifugation and collection of supernatants as described above. For each sample, supernatants were combined and dried under a gentle stream of nitrogen to complete dryness. Resulting pellets were dissolved in 400 μL of deionized water and mixed with 200 μL of 0.1 mol/L phosphate buffer (pH 7.4) and 50 μL of 37.04 mmol/L trimethylsilyl-2,2,3,3-tetradeuteropropionate (TSP) dissolved in deuterium oxide, the latter serving as an internal standard for referencing and quantification. For the analysis of cell culture supernatants, 400 μL of supernatant was mixed with buffer and deuterium oxide as described above. Here, as an internal standard for referencing and quantification, 10 μL of 81.97 mmol/L formic acid was additionally added.

All subsequent NMR analyses were performed on a 600 MHz Avance III spectrometer (Bruker BioSpin, Rheinstetten, Germany) employing a triple resonance (^1H, ^{13}C ^{15}N, ^2H lock) cryogenic probe equipped with *z*-gradients and an automatic cooled sample changer. Following established protocols, 1D

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

^{1}H NMR spectra were acquired employing either a 1D CPMG pulse sequence[29] for the analysis of cell culture supernatants or a 1D NOESY pulse sequence[30] for samples of cell pellet extracts. Additionally, for unambiguous metabolite identification, a 2D ^{1}H−^{13}C HSQC spectrum was acquired for each sample. Metabolites were identified by comparison with spectra of pure reference compounds employing CHENOMX 8.1 (Chenomx Inc., Edmonton, Canada), AMIX 3.9.13 (BrukerBioSpin, Rheinstetten, Germany) and Metabominer.[31] Metabolites were quantified from 1D ^{1}H spectra using AMIX 3.9.13 and Chenomx 8.1. To allow for the accurate determination of peak integrals in the case of partially overlapping signals, spectral deconvolution employing CHENOMX 8.1 was used. Metabolites of cell culture supernatants were quantified relative to the formic acid reference signal, while for cell pellet extracts the TSP reference signal was employed. Data from cell pellet extracts were normalized relative to the extraction standard. For the identification of signals corresponding to an unknown compound not available in the used spectral libraries, additional 2D ^{1}H−^{13}C HSQC, 2D ^{1}H−^{1}H TOCSY, and 2D ^{1}H−^{1}H NOESY spectra were acquired. Compound identification was further aided by hyphenated mass spectrometry employing a Bruker maXis-Impact quadrupole time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) coupled to a Dionex Ultimate 3000 high-performance liquid chromatography system (Thermo Fisher Scientific, Idstein, Germany). To this end, samples were diluted with deionized water (1:4). Based on the exact mass and fragment patterns together with the corresponding NMR data the compound was unambiguously identified as *N*-succinyl-L,L-2,6-diaminopimelate. Discriminating metabolites were identified by a two-sided heteroscedastic Student's *t* test while controlling the false discovery rate at the 5% level according to the method of Benjamini and Hochberg.[32]

Accession Numbers. The microarray data are accessible in NCBI's Gene Expression Omnibus[33] through accession no. GSE126228. The RNAseq data are accessible in NCBI's Gene Expression Omnibus through accession no. GSE133289. The structure of PdxJ(11) was deposited in the Protein Data Bank under accession code 6RG0.

RESULTS

***In Vivo* Complementation of Auxotrophic Deletions Strains by a Mixed Library of $(\beta\alpha)_8$-Barrel Enzymes.** To construct a gene library of diversified $(\beta\alpha)_8$-barrel enzymes we selected 52 pCA24N vectors from the ASKA library,[3] each harboring a gene for an *E. coli* phosphate-binding $(\beta\alpha)_8$-barrel enzyme (Table S3). This mixed pool was amplified and randomized by error-prone PCR, and the resulting mutants were cloned into the expression vector pTNA_*Bsa*I (Table S4). This vector contained two distinct *Bsa*I restriction sites and was designed for the easy generation of large gene libraries.[12] The resulting diversified $(\beta\alpha)_8$-barrel enzyme repertoire contained 1.4×10^8 different gene variants. Sequencing of 20 randomly picked clones revealed that library members contained 5.5 mutations per 1000 base pairs on average, corresponding to 4.1 amino acid exchanges. The 20 randomly picked clones belonged to 20 different genes, indicating that all genes from the initial pool are still present and equally distributed in the library. Along the same lines, more than 200 colony PCRs of random clones revealed a broad range of insert sizes.

To select for novel functions, different auxotrophic *E. coli* strains were obtained either from the Keio collection[10] or constructed by using standard P1 transduction methods.[11] We chose auxotrophic strains deleted for enzymes that fulfill a role in primary metabolism and bind substrates with a phosphate moiety. The deleted enzymes had to adopt a different fold than the $(\beta\alpha)_8$-barrel to exclude wild-type proteins from the library. Last, in order to decrease the complexity of the selected reactions, we tried to choose enzymes with as few cofactors as possible. All used strains (Δ*aro*A, Δ*aro*C, Δ*pdx*A, Δ*pur*E, Δ*pyr*B, Δ*ser*A, Δ*ser*B, Δ*ser*C, Δ*deo*C) were transformed with 2 μg (20 × 100 ng) of the diversified $(\beta\alpha)_8$-barrel enzyme library, followed by the selection for growth on minimal medium (Figure 1B). The medium was supplemented with glucose and, in some cases, additional nutrients according to the growth requirements of the individual strains (Table S1). By plating an aliquot of the reaction mixture on rich medium containing ampicillin, the transformation efficiency was determined to be higher than 10^9 cells per μg of plasmid DNA, ensuring that the entire library was successfully transformed into each auxotroph.

Colony growth on the minimal medium was observed only for the strains Δ*ser*B and Δ*ser*C, lacking the serine phosphatase SerB or the serine phosphate aminotransferase SerC, respectively (Table S1). SerC and SerB catalyze the final two, consecutive steps in the biosynthesis of serine: the conversion of 3-phosphohydroxypyruvate to phosphoserine and the dephosphorylation of phosphoserine (Figure 1C). Several hundred colonies appeared within 4−8 days for the Δ*ser*B complementation and several dozen colonies in 14−20 days for the Δ*ser*C complementation. Prior to plasmid isolation, about 50 arbitrarily selected colonies of Δ*ser*B and about 20 arbitrarily selected colonies of Δ*ser*C were restreaked on fresh minimal medium. Plasmids isolated from these fresh colonies were used for the retransformation of Δ*ser*B and Δ*ser*C. These retransformations confirmed the rescue of nine Δ*ser*B and the rescue of two Δ*ser*C colonies; colonies consistently grew within two to 5 days. The inserts of the plasmids from these colonies were sequenced (Table S6).

In order to ensure that only the $(\beta\alpha)_8$-barrel enzyme coding sequence was responsible for the rescue of Δ*ser*B and Δ*ser*C, the genes encoded by the selected plasmids were amplified by PCR and recloned into the vector pTNA_NN (Table S4). The genes were again sequenced and once more tested for complementation of Δ*ser*B and Δ*ser*C. Colonies again consistently grew within 2−5 days.

Variants of *nan*E and *pdx*J Rescue Δ*ser*B and Δ*ser*C Strains on Minimal Media. The sequencing results showed that variants of two different genes, *nan*E (encoding a putative *N*-acetylmannosamine-6-phosphate 2-epimerase) and *pdx*J (encoding the pyridoxine 5′-phosphate synthase), were responsible for the observed *in vivo* complementation (Table S6). The two *nan*E variants (*nan*E(7) and *nan*E(25)), which were identified through their ability to rescue Δ*ser*B, contained 3 and 6 amino acid exchanges. Variants of *pdx*J were identified through their rescue of Δ*ser*B (7 variants) and or through their rescue of Δ*ser*C (2 variants). Remarkably, apart from one to four different amino acid exchanges, all *pdx*J variants contained a ten amino acid extension at the C-terminus (Table S6). This extension was recruited from the cloning vector pCA24N. Two deletions in the sequence of the synthesized primers resulted in a frameshift, leading to the loss of the wild-type stop codon and the recruitment of an alternative stop codon from the

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

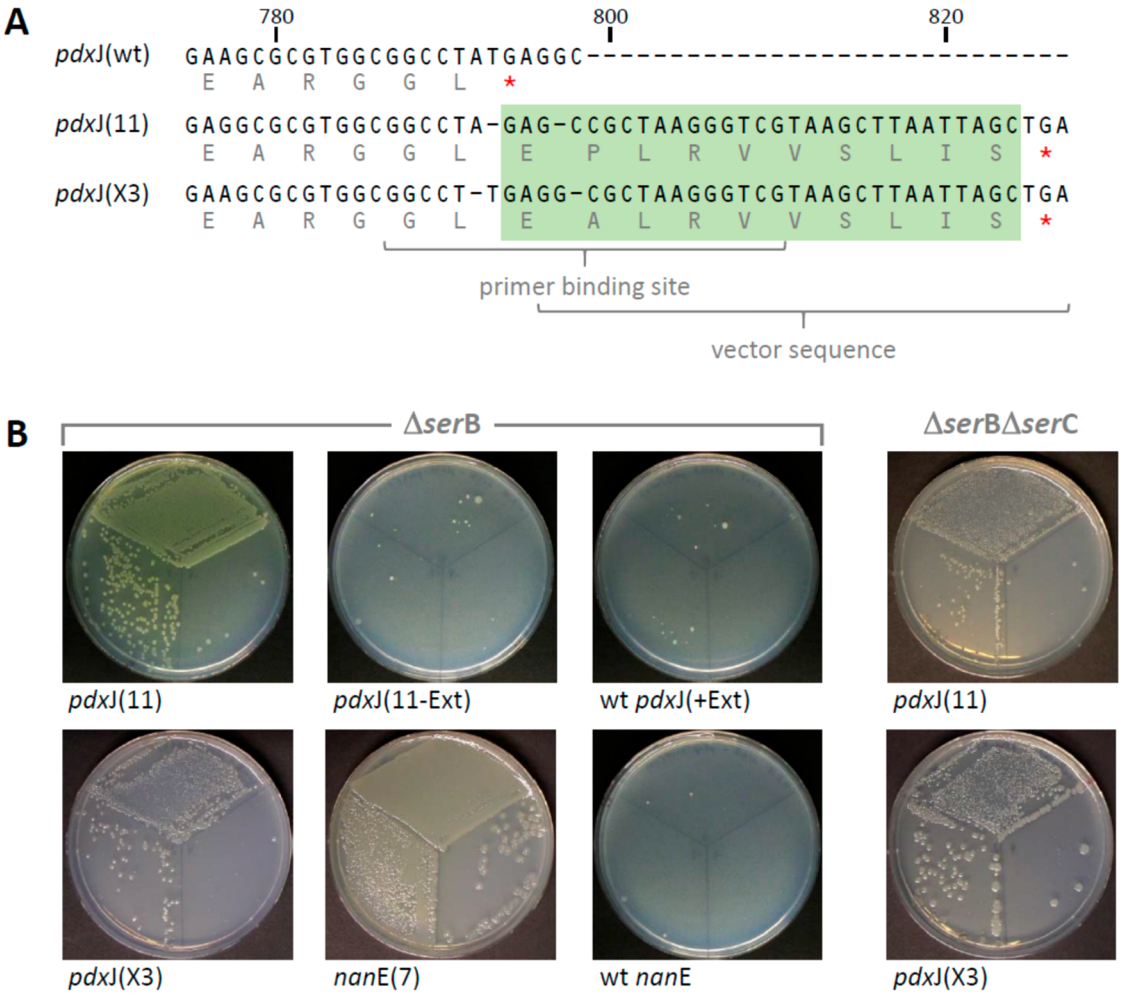


Figure 2. (A) Sequence of the ten amino acid extension of selected *pdx*J variants. The extension is shaded in green. The first 16 bases of the extension are recruited from the primer sequence; all point mutations or single nucleotide deletions that differentiate selected variants occurred in the primer sequence. The last 14 bases of the extension along with the stop codon are recruited from the sequence of vector pCA24N. Sequences of the C-terminal extension of additional *pdx*J variants are documented in Figure S1. (B) Growth of deletion strains on M9 minimal medium transformed with different constructs. The variants *pdx*J(11) and *pdx*J(X3) rescue the Δ*ser*B strain as well as the double deletion strain Δ*ser*BΔ*ser*C. *Nan*E(7) rescues Δ*ser*B. Constructs lacking the C-terminal extension (*pdx*J(11-Ext)), or wild-type *pdx*J(+Ext), and wild-type *nan*E do not rescue Δ*ser*B. Three different cell concentrations were plated: undiluted (top third of the plates), 10^{-2} dilution (lower left third of the plates), and 10^{-4} dilution (lower right third of the plates). The time until colonies appeared is given in Table 1. Additional complementation experiments are documented in Figure S2.

Table 1. Overview of *in Vivo* Complementation of Various Δ*ser*B and Δ*ser*C Single and Double Deletion Strains by Different Genes[a]

	gene										
deletion strain	*pdx*J(11)	*pdx*J(X3)	wt *pdx*J	wt *pdx*J(+Ext)	*nan*E(7)	wt *nan*E	wt *ser*B	wt *ser*C	wt *his*B	wt *his*C	wt *his*B*his*C
Δ*ser*B	2–3	2–3	–	–	3–5	–	1	–	1	–	1
Δ*ser*C	4–6	4–6	–	–	20	–	–	2	–	3	3
Δ*ser*BΔ*ser*C	4–6	4–6	–	–	–	–	–	–	–	–	3
Δ*ser*BΔ*gph*	2–3	2–3	NT	NT	3–5	NT	1	NT	1	NT	NT
Δ*ser*BΔ*ytj*C	2–3	2–3	NT	NT	3–5	NT	1	NT	1	NT	NT
Δ*ser*BΔ*his*B	–	–	NT	NT	–	NT	–	NT	1	–	2
Δ*ser*CΔ*his*C	–	–	NT	NT	–	NT	–	–	–	23	23

[a]The shown deletion strains were transformed with the different variants cloned in vector pTNA_*Bsa*I, and growth times on minimal M9 medium supplemented with glucose and supplements according to the requirements of the different strains (see Table S1) were recorded. The numbers correspond to the days until colony growth was observed. "–": no growth was observed after 30 days. "NT": not tested.

vector sequence (Figure 2A, Figure S1). During construction of the library and the selection experiments, we sequenced more than 70 genes that had been cloned with the same batch of primers. Yet, this ten amino acid extension was exclusively observed in the *pdx*J variants that were selected through complementation of Δ*ser*B and Δ*ser*C.

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX–XXX

Table 2. Variants of *pdx*J Most Frequently Referred to in This Study

gene	amino acid exchanges	C-terminal extension	initially selected by complementation of
*pdx*J(11)	E146 K, H198R, M224L, D229V	yes	Δ*ser*B
*pdx*J(X3)	M224L	yes	Δ*ser*C
wt *pdx*J	–	no	not applicable
wt *pdx*J(+Ext)	–	yes	not applicable
*nan*E(7)	S2L, Q6L, Q44R, I72T, L140P, H199R	no	Δ*ser*B
wt *nan*E	–	no	not applicable

Subsequent complementation experiments showed that the *nan*E and *pdx*J variants that were selected via the rescue of Δ*ser*B could also rescue Δ*ser*C, and vice versa. This was true for all variants and is paradigmatically shown for *pdx*J(11), *pdx*J(X3), and *nan*E(7) in Table 1.

Growth of the Δ*ser*B strain complemented by *pdx*J and nanE variants was faster (2−3 days and 3−5 days, respectively) than the growth of the Δ*ser*C strain (4−6 days and 20 days, respectively). The *pdx*J variants most frequently referred to in this study are listed in Table 2, and their *in vivo* complementation of Δ*ser*B is paradigmatically shown in Figure 2B.

Both, the mutations identified in the selected variants of *pdx*J and *nan*E and the C-terminal extension in the *pdx*J variants are required for the rescue of Δ*ser*B and Δ*ser*C: The wild-type enzymes *nan*E and *pdx*J, cloned into the same expression vector as the gene library, do not confer growth. Furthermore, wild-type *pdx*J with the ten amino acid C-terminal extension, but no additional mutations throughout the gene (wt *pdx*J(+Ext)), does not rescue the *ser*B or *ser*C deletion either (Table 1). Vice versa, the ten amino acid C-terminal extension is essential for the ability to complement Δ*ser*B and Δ*ser*C. Complete removal of the extension totally abolishes growth for all *pdx*J variants tested (Table S6).

We proceeded with mutating the coding sequence of the ten amino acid C-terminal extension of one fast-growing *pdx*J variant, *pdx*J(11). If the extension is shortened by two or more amino acids, the ability to rescue Δ*ser*B and Δ*ser*C is completely lost (Table S7). Additionally, we found that although the extension can be lengthened by one amino acid without losing the ability to rescue Δ*ser*B and Δ*ser*C, introducing the bulky amino acid tryptophan at the last position of the extension results in total loss of the ability to rescue Δ*ser*B and Δ*ser*C. Next, we constructed a gene library of the variant *pdx*J(11) where we randomized the amino acid at position 10 of the extension and selected for complementation of Δ*ser*B. While most amino acids at this position led to rescue of the deletion strain, a few, mostly bulky, amino acids (phenylalanine, glutamate, glutamine, lysine, valine, and cysteine), were never found at this position.

To confirm that the variant protein PdxJ, and not the mRNA, is responsible for the observed phenotype, we mutated the first bases of *pdx*J(11) to introduce either an early stop codon or an early frame shift. In both cases, the sequence of the translated protein is severely affected, whereas the mRNA remains practically identical. Growth assays showed that both mutants no longer rescue Δ*ser*B or Δ*ser*C.

Furthermore, we performed complementation experiments with a newly constructed double deletion strain Δ*ser*BΔ*ser*C. As expected, the *pdx*J(11) and *pdx*J(X3) variants could rescue Δ*ser*BΔ*ser*C, illustrating that these variants are simultaneously capable of complementing the lack of both serine biosynthesis enzymes (Figure 2B, Table 1). The *nan*E variants, which were selected for their ability to complement Δ*ser*B, can only complement Δ*ser*C with slower growth times compared to the *pdx*J variants and were not able to complement the double deletion strain Δ*ser*BΔ*ser*C, for unknown reasons.

Characterization of Purified NanE and PdxJ Variants. Both isolated *nan*E variants and three isolated *pdx*J variants (with or without the C-terminal extension), along with the wild-type genes *ser*B, *ser*C, *nan*E, and *pdx*J (with and without the extension), were cloned into the expression vector pET21a. The corresponding recombinant proteins containing an N-terminal hexahistidine tag were expressed in *E. coli* and could readily be purified to homogeneity from the soluble cell extract using metal chelate affinity chromatography. Circular dichroism spectroscopy and analytical gel filtration chromatography showed that all purified proteins were well folded and confirmed that the amino acid exchanges and the C-terminal extension did not change the oligomerization state compared to the wild-type proteins. Along the same lines, the structure of the variant PdxJ(11) as determined by X-ray crystallography (PDB code 6RG0) turned out to be similar to the structure of wild-type PdxJ (PDB code 1M5W), with an overall rmsd of 2.0 Å for 239 Cα-atoms (Figure S3). Unfortunately, the 10 amino C-terminal extension is not visible in the structure, indicating high structural flexibility.

The results of the *in vivo* complementation described above suggested that the selected NanE and PdxJ variants possess SerB activity. In order to test this, we performed two different assays with the purified proteins. However, neither in end-point assays using radio-labeled phosphoserine as the substrate, nor by detecting free phosphate cleaved from phosphoserine with malachite green,[14] could we detect any enzymatic activity. Addition of various divalent metal ions or of crude cell extract did not result in activity either. Wild-type SerB from *E. coli* was used as positive control and showed activity levels comparable to previous studies.[34,35] We concluded that the selected *nan*E and *pdx*J variants do not rescue Δ*ser*B by providing the missing enzymatic function.

The NanE and PdxJ Variants Rescue Δ*ser*B and Δ*ser*C by Recruiting HisB and HisC. Previous studies have shown that novel sequences can rescue deletion strains not by providing the missing enzymatic function, but by recruiting an endogenous, promiscuous *E. coli* enzyme.[36,37] In *E. coli*, three phosphatases, HisB, Gph, and YtjC, can rescue Δ*ser*B when overexpressed.[2] To test whether any of these phosphatases play a role in the rescue of Δ*ser*B by the *nan*E and *pdx*J variants, we constructed the double deletion strains Δ*ser*BΔ*his*B, Δ*ser*BΔ*gph*, and Δ*ser*BΔ*ytj*C. Whereas, the additional deletion of the phosphoglycolate phosphatase Gph or the putative phosphoglycerate mutase YtjC had no effect on the rescue of Δ*ser*B, the double deletion strain Δ*ser*BΔ*his*B could no longer be complemented by our selected *nan*E or *pdx*J variants (Table 1). This finding suggests that the *nan*E and *pdx*J variants recruit the promiscuous histidinol phosphatase HisB, which

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

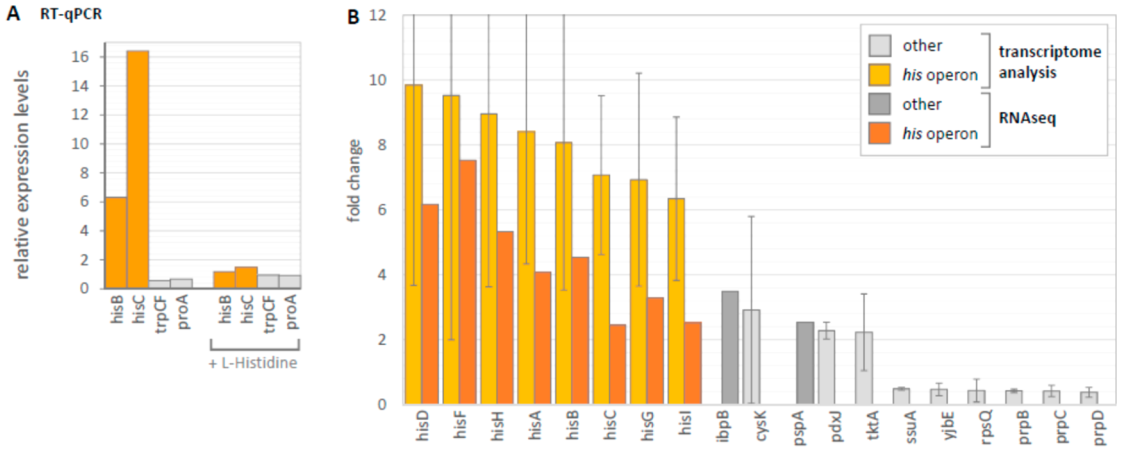


Figure 3. Expression of *pdx*J(11) increases histidine operon mRNA levels up to 7−16-fold. Bars show the abundance of transcripts in cells expressing *pdx*J(11) relative to the same cell line expressing *pdx*J(11-Ext) as a negative control. Genes of the histidine operon are colored orange. (A) qRT-PCR identifies changes in relative mRNA levels of the four amino acid biosynthesis genes *his*B, *his*C, *trp*CF, and *pro*A upon expression of *pdx*J(11). Addition of L-histidine (12 μg/mL) to the growth medium prevents induction of *his* operon expression through overexpressed *pdx*J(11). Expression levels were normalized with the housekeeping 23S rRNA gene. (B) DNA microarray analysis and RNAseq identifies fold changes in mRNA levels. Shown are differentially expressed genes (more than 2-fold change in DNA microarray analysis and *p*-adjust value < 0.01 and more than a 2-fold change in RNAseq) in cells expressing *pdx*J(11) compared to the negative controls lacking the extension (*pdx*J(11-Ext)). DNA microarray: Values given correspond to average fold changes and the standard deviation calculated from two arrays of two independent biological replicates. RNAseq: Values are based on two to three independent experiments, each with biologically independent cultures. All DNA microarray data are shown in Table S8, and all RNAseq data are shown in Table S9.

dephosphorylates histidinol-phosphate in the biosynthesis of histidine (Figure 1D) and is known to catalyze the SerB reaction.[9,36]

However, this did not explain the rescue of Δ*ser*C through the expression of the *nan*E and *pdx*J variants. Still, the similarity of the enzymatic reactions of SerC and the histidinol phosphate aminotransferase HisC (Figure 1D) let us construct the double deletion strain Δ*ser*CΔ*his*C. And, as was the case with Δ*ser*BΔ*his*B, the selected *nan*E and *pdx*J variants could not confer growth to this strain (Table 1). Furthermore, we could confirm that overexpression of *his*C can indeed rescue the Δ*ser*C strain and that the combined overexpression of both *his*B and *his*C rescues the double deletion strain Δ*ser*BΔ*ser*C (Table 1). These results strongly indicate that HisC is promiscuous for the SerC reaction, just as HisB has promiscuous SerB activity.[2,34] This promiscuity of HisC was probably missed in the previous large-scale complementation analysis[2] due to a deletion in the coding sequence of *his*C in the ASKA vector pCA24N-*his*C, which we detected by sequencing.

The NanE and PdxJ Variants Induce the Expression of the Histidine Operon. We were interested in uncovering the mechanism by which the selected variants induce the overproduction of the endogenous *E. coli* HisB and HisC proteins. In a first step to solve this enigma, we performed RT-qPCR to quantify possible expression changes of the *his*B and *his*C genes in the presence of PdxJ variants. Due to the very slow growth of the deletion strains in liquid minimal medium, total RNA was isolated from the pseudo-wild-type strain BW25113 (the parent strain of the deletion strains used in this study), expressing either *pdx*J(11) or *pdx*J(X3). Expression of either *pdx*J variant led to 7 to 16-fold increased *his*B and *his*C mRNA levels, whereas other amino acid biosynthetic genes were not affected (Figure 3A).

To analyze alterations of the transcriptome in the presence of PdxJ and NanE variants in an unbiased manner, we performed both DNA microarray and RNAseq analysis. Analogous to the RT-qPCR measurements described above, total RNA was isolated from the pseudo wild-type strain BW25113 expressing either *pdx*J(11), *pdx*J(X3), *nan*E(7), or the negative controls *pdx*J lacking the C-terminal extension (*pdx*J(11-Ext)), wild-type *pdx*J with the C-terminal extension (wt *pdx*J(+Ext)), and wild-type *ser*B. Although the transcription of a large number of genes was found to be altered through expression of *pdx*J(11), *pdx*J(X3), and *nan*E(7), only eight genes belonging to the histidine operon (*his* operon) were found to be significantly and exclusively overexpressed under all conditions tested (Table S8, Table S9, Figure 3B). Thus, expression of *pdx*J and *nan*E variants leads to selective induction of *his* operon expression and, through recruitment of promiscuous HisB and HisC, to the rescue of the Δ*ser*B and Δ*ser*C strains.

Attempts To Clarify How the NanE and PdJ Variants Induce the Expression of the Histidine Operon. In *E. coli*, transcription of the *his* operon is stimulated by guanosine 5′-diphosphate 3′-diphosphate (ppGpp), the effector of the stringent response,[38] under conditions of amino acid starvation and in cells grown in minimal medium. Other amino acids biosynthetic operons are simultaneously up-regulated through ppGpp, while the translation apparatus is subject to a large-scale downregulation.[39] In cells expressing *pdx*J(11), *pdx*J(X3), and *nan*E(7), we did not observe the typical gene expression pattern of the stringent response (Table S8, Table S9, Figure 3B). Thus, we conclude that ppGpp is not involved in the rescue mechanism.

Furthermore, transcription of the *his* operon is regulated by attenuation.[40] Attenuation is based on the finding that an excess of the amino acid L-histidine leads to the production of the leader peptide HisI, which results in an mRNA secondary structure that functions as a terminator signal for the RNA polymerase. If L-histidine is limited, the ribosome stalls at the histidine codons of HisI, leading to the formation of an alternative secondary structure (the "antiterminator") and subsequent transcription of the whole operon. We tested

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

whether the presence of an excess concentration of L-histidine affects the transcription of the *his* operon when *pdx*J and *nan*E variants are overexpressed. We found that *pdx*J(11), *pdx*J(X3), and *nan*E(7) could no longer rescue the Δ*ser*B and Δ*ser*C strains when L-histidine is added to the growth medium. RT-qPCR verified these results: In the presence of L-histidine, the genes *his*B and *his*C of the *his* operon are down-regulated even under overexpression of *pdx*J(11) (Figure 3A) or *pdx*J(X3) (Figure S4). Thus, the endogenous attenuation mechanism is probably unaffected by the *pdx*J and *nan*E variants. However, we cannot rule out that *pdx*J(11) and *pdx*J(X3) relieve background attenuation with low efficiency, and that such a weak effect can be overridden by the attenuation that occurs in the presence of L-histidine.

In order to identify potential interaction partners of PdxJ(X3) that might mediate up-regulation of the *his* operon, protein complex immunoprecipitation (Co-IP) was performed; wild-type *pdx*J(+Ext), which lacks the M224L exchange, was used as a negative control. An N-terminal Flag-tag was attached to both proteins, which were then expressed in BW25113 cells. Following cell lysis, Co-IP was performed with the help of anti-Flag-tag agarose beads. Whole protein extracts (input), as well as the supernatants of two subsequent elution steps, were analyzed by SDS-PAGE. Identified bands were cut out, digested with trypsin and analyzed via mass spectrometry. The input and both elution steps contained main bands corresponding to PdxJ(X3) and wild-type *pdx*J(+Ext), which shows that both proteins were overexpressed and successfully bound to and eluted from the anti-Flag-tag agarose beads. Two additional minor bands corresponded to the heavy and light chains of the antibody coupled to the beads. However, no interaction partner of PdxJ(X3) could be detected.

We hypothesized that small molecules from the *E. coli* metabolome could be recruited by the selected variants and up-regulate the transcription of the *his* operon. In order to identify such small-molecule regulators, we compared the total metabolome of BW25113 cells expressing either *pdx*J(11) or the negative control *pdx*J(11-Ext), lacking the C-terminal extension, by a global NMR based metabolic screen. We analyzed both cellular metabolites and the culture supernatant of cells grown in minimal medium in the exponential growth phase. In each case, three biological replicates were investigated per group. Following metabolite identification and quantification, discriminating metabolites were identified by a two-sided heteroscedastic Student's *t* test. While we did not identify metabolites that are more abundant in cells expressing *pdx*J(11), we found 5 small molecules (alanine, threonine, lysine, maltotriose, and the intermediate of lysine biosynthesis *N*-succinyl-L,L-2,6-diaminopimelate (SDAP)) that are significantly over-represented in whole cells expressing the control *pdx*J(11-Ext), and two metabolites (propionate and acetate) that are significantly over-represented in the culture supernatant of the same cells (Table S10). We speculated that one (or more) of the compounds that accumulated in cells expressing *pdx*J(11-Ext) might function as negative regulators for *his* operon transcription. Expression of *pdx*J(11) would lead to a drop in their concentration and hence indirectly up-regulate the *his* operon. To test this hypothesis, we added the identified compounds to the growth medium of Δ*ser*B cells expressing *pdx*J(11) (to a concentration of 12 μg/mL or 120 μg/mL). In the case of SDAP, we had to substitute the originally identified substance with the structurally related and commercially available desuccinylated 2,6-diaminopimelic acid (DAP). We expected that, if one of the identified metabolites had a regulatory function, the rescue of Δ*ser*B would be impaired by its addition (analogous to the addition of L-histidine to the growth medium). Unfortunately, none of the compounds had such an effect. Δ*ser*B cells expressing *pdx*J(11) grow normally on minimal medium even when all seven compounds are combined in high concentrations (12 μg/mL or 120 μg/mL).

DISCUSSION

By generating a large library of diversified $(\beta\alpha)_8$-barrel enzymes, we sought to select for novel enzymatic functions through auxotrophy complementation. We were able to identify variants in our library that allowed the growth of the deletion strains Δ*ser*B and Δ*ser*C. However, in contrast to our initial expectations, these variants do not substitute for the missing SerB or SerC by catalyzing the same enzymatic reaction. Obviously, although the $(\beta\alpha)_8$-barrel fold is thought to be very versatile, a pool of only 52 enzymes containing a few mutations each is still a very limited set and possibly too small for the evolution of a novel catalytic activity or substrate specificity. Another reason for our lack of success might be a too-small number of introduced mutations. On average, each member of our mixed library contained about four amino acid exchanges. Assuming that the most promising substitutions are located in active site loops, which comprise no more than about 25% of all residues of the $(\beta\alpha)_8$-barrel scaffold, a single exchange would have had to suffice for establishing a novel catalytic activity. Instead of identifying such a new enzymatic activity, we found that the selected PdxJ and NanE variants lead to the recruitment of HisB and HisC.

It has been shown previously that overexpression of *his*B is sufficient to rescue the *ser*B deletion[2] and that HisB has promiscuous SerB activity.[9] SerB and HisB are both members of the HAD-like hydrolase superfamily, SCOP classification c.67.1,[41] that share the same protein fold and catalyze similar reactions (EC 3.1.3), the dephosphorylation of phosphoserine (Figure 1C) and the dephosphorylation of histidinol phosphate (Figure 1D), respectively. However, in this work, we could show that the relationship between the serine and histidine biosynthetic pathways extends to SerC and HisC, as well. Analogous to SerB and HisB, SerC and HisC are structurally and functionally related: Both enzymes belong to the same SCOP[41] superfamily c.67.1 and both function as PLP-dependent aminotransferases (EC 2.6.1) that catalyze the transfer of an amino group from donor L-glutamate to their respective substrates (Figures 1C,D). Overexpression of *his*C is sufficient to rescue the *ser*C deletion, illustrating that HisC retains a promiscuous activity for the SerC reaction. Our results confirm the common evolutionary origin of histidine and serine biosynthetic genes as postulated early on.[42]

*Pdx*J and *nan*E variants described in this study rescue the deletion of SerB or SerC through exclusive up-regulation of histidine operon expression. However, the mechanism behind this up-regulation remains unclear. While we have ruled out any involvement of the stringent response effector ppGpp, it is principally possible that PdxJ and NanE variants bind directly to the regulatory region of the *his* operon DNA or mRNA. However, this seems unlikely as the acquired mutations (both in the gene sequence and the C-terminal extension) do not add positive charges that might be expected for a novel nucleic acid binding site. Furthermore, our variants are not the only examples of novel proteins that can rescue Δ*ser*B through up-

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

regulation of *his* operon mRNA: The 4-helix bundle SynSerB3, that shares no sequence or structural identity with our variants nor with the C-terminal extension, and adopts a completely different protein fold, analogously stimulates the expression of the *his* operon.[36] Therefore, it seems that an unknown and specific activator or repressor of the *his* operon exists in the *E. coli* proteome and that this potential activator or repressor might be the target of both PdxJ and NanE variants and SynSerB3. However, our Co-IP experiments performed with PdxJ(X3) failed to detect such a hypothetical interaction partner.

Lab-made, small proteins have been shown to provide a life-sustaining function by changing gene expression patterns in *E. coli*,[36,37] analogous to the results presented in this study. Whereas these designed proteins are 4-helix bundles with no sequence similarity to naturally occurring proteins, the variants presented in this study belong to the $(\beta\alpha)_8$-barrel fold. Natural $(\beta\alpha)_8$-barrels are enzymes belonging to five of the six enzyme classes, and the evolvability of this fold and its ability to acquire new enzymatic functions has been illustrated repeatedly.[43] To our knowledge, the variants described in this study are the first examples of $(\beta\alpha)_8$-barrel enzymes that function not as enzymes but have a regulatory function and lead to altered gene expression patterns in *E. coli*. As such, they provide a new example of the adaptability of the $(\beta\alpha)_8$-barrel fold.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.9b00579.

E. coli deletions strains and growth media used for selection; list of oligonucleotides used in this study; 52 phosphate-binding $(\beta\alpha)_8$-barrel genes pooled for the generation of the diversified library; list of expression vectors used in this study; crystal structure of PdxJ(11), data collection and refinement statistics; variants of *nan*E and *pdx*J selected by *in vivo* complementation; effect of the modification of the C-terminal extension of *pdx*J variants for *in vivo* complementation of Δ*ser*B; DNA microarray analysis to identify genes exhibiting more than a 2-fold change of mRNA levels; RNAseq to identify genes exhibiting more than a 2-fold change of mRNA levels; NMR analysis of metabolites in whole cells and culture supernatants; sequence of the ten amino acid extension of *pdx*J variants; growth of deletion strains on M9 minimal medium transformed with different constructs; overlay of variant PdxJ(11) with wild-type PdxJ; qRT-PCR identifies changes in relative mRNA levels of the four amino acid biosynthesis genes *his*B, *his*C, *trp*CF, and *pro*A upon expression of *pdx*J(X3) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: reinhard.sterner@ur.de. Phone: +49-941-943-3015.

ORCID

Michael Bott: 0000-0002-4701-8254

Reinhard Sterner: 0000-0001-8177-8460

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Joseph Sperl and Elisabeth Wörle for the *E. coli* deletion strains Δ*deo*C and Δ*ser*BΔ*gph* and Marc Creus for suggesting using DAP as a substitute for commercially unavailable SDAP. We are grateful to Ingrid Araya and Kevin Heizler for their advice on the Co-IP experiments.

■ REFERENCES

(1) Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem. 79*, 471−505.

(2) Patrick, W. M., Quandt, E. M., Swartzlander, D. B., and Matsumura, I. (2007) Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol. 24*, 2716−2722.

(3) Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2006) 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 Res. 12*, 291−299.

(4) Oberhardt, M. A., Zarecki, R., Reshef, L., Xia, F., Duran-Frigola, M., Schreiber, R., Henry, C. S., Ben-Tal, N., Dwyer, D. J., Gophna, U., and Ruppin, E. (2016) Systems-Wide Prediction of Enzyme Promiscuity Reveals a New Underground Alternative Route for Pyridoxal 5′-Phosphate Production in E. coli. *PLoS Comput. Biol. 12*, e1004705.

(5) Sterner, R., and Höcker, B. (2005) Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem. Rev. 105*, 4038−4055.

(6) List, F., Sterner, R., and Wilmanns, M. (2011) Related $(\beta\alpha)_8$-barrel proteins in histidine and tryptophan biosynthesis: a paradigm to study enzyme evolution. *ChemBioChem 12*, 1487−1494.

(7) Obexer, R., Godina, A., Garrabou, X., Mittl, P. R., Baker, D., Griffiths, A. D., and Hilvert, D. (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem. 9*, 50−56.

(8) Nobeli, I., Ponstingl, H., Krissinel, E. B., and Thornton, J. M. (2003) A structure-based anatomy of the E. coli metabolome. *J. Mol. Biol. 334*, 697−719.

(9) Yip, S. H., and Matsumura, I. (2013) Substrate ambiguous enzymes within the Escherichia coli proteome offer different evolutionary solutions to the same problem. *Mol. Biol. Evol. 30*, 2001−2012.

(10) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006.0008.

(11) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640−6645.

(12) Rohweder, B., Semmelmann, F., Endres, C., and Sterner, R. (2018) Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*. *BioTechniques 64*, 24−26.

(13) Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci. 4*, 2411−2423.

(14) Geladopoulos, T. P., Sotiroudis, T. G., and Evangelopoulos, A. E. (1991) A malachite green colorimetric assay for protein phosphatase activity. *Anal. Biochem. 192*, 112−116.

(15) Linde, M., Heyn, K., Merkl, R., Sterner, R., and Babinger, P. (2018) Hexamerization of geranylgeranylglyceryl phosphate synthase ensures structural integrity and catalytic activity at high temperatures. *Biochemistry 57*, 2335−2348.

(16) Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr. 26*, 795−800.

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX

(17) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.
(18) Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) The CCP4 molecular-graphics project. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1955−1957.
(19) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 240−255.
(20) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 1948−1954.
(21) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126−2132.
(22) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375−383.
(23) Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods 9*, 357−359.
(24) Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics 25*, 2078−2079.
(25) Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics 31*, 166−169.
(26) Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oles, A. K., Pages, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods 12*, 115−121.
(27) Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol. 15*, 550.
(28) Kranz, A., Steinmann, A., Degner, U., Mengus-Kaya, A., Matamouros, S., Bott, M., and Polen, T. (2018) Global mRNA decay and 23S rRNA fragmentation in Gluconobacter oxydans 621H. *BMC Genomics 19*, 753.
(29) Wallmeier, J., Samol, C., Ellmann, L., Zacharias, H. U., Vogl, F. C., Garcia, M., Dettmer, K., Oefner, P. J., and Gronwald, W. (2017) Quantification of Metabolites by NMR Spectroscopy in the Presence of Protein. *J. Proteome Res. 16*, 1784−1796.
(30) Gronwald, W., Klein, M. S., Kaspar, H., Fagerer, S. R., Nurnberger, N., Dettmer, K., Bertsch, T., and Oefner, P. J. (2008) Urinary metabolite quantification employing 2D NMR spectroscopy. *Anal. Chem. 80*, 9288−9297.
(31) Xia, J., Bjorndahl, T. C., Tang, P., and Wishart, D. S. (2008) MetaboMiner–semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinf. 9*, 507.
(32) Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate-a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B-Stat. Meth. 57*, 289−300.
(33) Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res. 30*, 207−210.
(34) Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V., Arrowsmith, C. H., Koonin, E. V., Edwards, A. M., and Yakunin, A. F. (2006) Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem. 281*, 36149−36161.
(35) Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E., and Kim, S. H. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. *J. Mol. Biol. 319*, 421−431.
(36) Digianantonio, K. M., and Hecht, M. H. (2016) A protein constructed de novo enables cell growth by altering gene regulation. *Proc. Natl. Acad. Sci. U. S. A. 113*, 2400−2405.
(37) Digianantonio, K. M., Korolev, M., and Hecht, M. H. (2017) A non-natural protein rescues cells deleted for a key enzyme in central metabolism. *ACS Synth. Biol. 6*, 694−700.
(38) Artz, S. W., and Holzschu, D. (1983) Histidine Biosynthesis and Its Regulation. In *Amino Acids: Biosynthesis and Genetic Regulation* (Somerville, K. M. H. a. R. L., Ed.), pp 379−404, Addison-Wesley Publishing Co, Reading, MA.
(39) Cashel, M., and Rudd, K. E. (1987) The Stringent Response. In *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E., Eds.), pp 1410−1438, American Society for Microbiology, Washington, D.C.
(40) Kolter, R., and Yanofsky, C. (1982) Attenuation in amino acid biosynthetic operons. *Annu. Rev. Genet. 16*, 113−134.
(41) Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res. 30*, 264−267.
(42) Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30*, 409−425.
(43) Gerlt, J. A., and Raushel, F. M. (2003) Evolution of function in $(\beta\alpha)_8$-barrel enzymes. *Curr. Opin. Chem. Biol. 7*, 252−264.

DOI: 10.1021/acs.biochem.9b00579
Biochemistry XXXX, XXX, XXX−XXX