Cellular function of the GndA small open reading frame-encoded polypeptide during heat shock

Jessica J. Mohsen[1,2], Michael G. Mohsen[3,4], Kevin Jiang[1,2], Ane Landajuela[5,6], Laura Quinto[3,7], Farren J. Isaacs[3,7], Erdem Karatekin[5,6,8,9,10], Sarah A. Slavoff[1,2,10,11]

1 Department of Chemistry, Yale University, New Haven, CT 06511
2 Institute for Biomolecular Design and Discovery, Yale University, West Haven, CT 06516
3 Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511
4 Howard Hughes Medical Institute, Yale University, New Haven, CT 06511
5 Department of Cellular and Molecular Physiology, Yale School of Medicine, New Haven, CT 06510
6 Nanobiology Institute, Yale University, West Haven, CT 06516
7 Systems Biology Institute, Yale University, West Haven, CT 06516
8 Wu Tsai Institute, Yale University, New Haven, CT 06511
9 Université de Paris, Saints-Pères Paris Institute for the Neurosciences (SPPIN), Centre National de la Recherche Scientifique (CNRS), 75006 Paris, France
10 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511
11 To whom correspondence should be addressed:

Sarah A. Slavoff
Yale University West Campus
Chemical Biology Institute
Molecular Innovations Center 232
600 West Campus Drive
PO Box 27392
West Haven, CT 06516
Phone: 203-737-8670
sarah.slavoff@yale.edu

Classification: Biological Sciences; Biochemistry
Keywords: Small open reading frame (sORF); SEP; aerobic respiration; respiratory complex I; nuo; 6PGD; GndA; heat shock

Abstract

Over the past 15 years, hundreds of previously undiscovered bacterial small open reading frame (sORF)-encoded polypeptides (SEPs) of fewer than fifty amino acids have been identified, and biological functions have been ascribed to an increasing number of SEPs from intergenic regions and small RNAs. However, despite numbering in the dozens in *Escherichia coli*, and hundreds to thousands in humans, same-strand nested sORFs that overlap protein coding genes in alternative reading frames remain understudied. In order to provide insight into this enigmatic class of unannotated genes, we characterized GndA, a 36-amino acid, heat shock-regulated SEP encoded within the +2 reading frame of the *gnd* gene in *E. coli* K-12 MG1655. We show that GndA pulls down components of respiratory complex I (RCI) and is required for proper localization of a RCI subunit during heat shock. At high temperature GndA deletion (ΔGndA) cells exhibit perturbations in cell growth, $NADH^+$/NAD ratio, and expression of a number of genes including several associated with oxidative stress. These findings suggest that GndA may function in maintenance of homeostasis during heat shock. Characterization of GndA therefore supports the nascent but growing consensus that functional, overlapping genes occur in genomes from viruses to humans.

Significance Statement

Same-strand overlapping, or nested, protein coding sequences optimize the information content of size-constrained viral genomes, but were previously omitted from prokaryotic and eukaryotic genome annotations. It was therefore surprising when dozens of nested sORFs were recently discovered in bacteria. Our case study of *E. coli* GndA supports the hypothesis that overlapping genes may exist because they encode proteins with related functions. More broadly, characterization of nested sORFs may revise our understanding of the architecture of bacterial and eukaryotic genes.

Main Text

Introduction

Small open reading frames (sORFs) are generally defined as short coding sequences (CDSs) that encode polypeptides (sORF-encoded polypeptides, or SEPs) with functions across biology (1–6). Due to an arbitrary 50-amino acid (aa) cutoff for bacterial genome annotation, bacterial sORFs have been overlooked until recently (3). This cutoff was applied because short protein sequences are refractory to strategies for prediction and experimental detection of canonical (larger) proteins, which include bioinformatic tools that assess sequence homology (7), structural homology (8), codon usage (9), and phylogenetic conservation (10). SEPs are also difficult to detect using standard proteomic workflows (11). SEPs, also referred to as microproteins (5), miniproteins (12), micropeptides (13), or small proteins (2, 3), are now known to be abundant, to have low molecular weights, and are often membrane-associated (14). Some SEPs also exhibit limited conservation even amongst related species (15) which is surprising given their biological functionality (16, 17). As a result of these properties, SEPs are challenging to detect, and their high-confidence annotation necessitates the development of new genomic (18, 19) and proteomic technologies (11). A combination of mass spectrometry (11, 20), ribosome profiling (18, 19), in vivo contextualization (21–23), and computation (24) should be leveraged to illuminate both the existence and function of SEPs within the cellular landscape.

The functions of SEPs in carbon metabolism and stress responses are beginning to be elucidated. A foundational study reported upregulation of multiple SEPs in *Escherichia coli*

during the heat shock response at 45 °C, suggesting that SEPs may be broadly involved in cellular stress responses (25). SEPs have also been implicated in cellular metabolism. In human cells, mitochondrial SEPs (mito-SEPs), including BRAWNIN (26), MOCCI (27), SMIM4 and UQCC1 (28), regulate assembly and/or activity of the electron transport chain. Additionally, mito-SEP SLC35A4-MP localizes to the inner mitochondrial membrane where it affects cellular respiration via an uncharacterized mechanism (29). *E. coli* SEPs AzuC (30), SgrT (31), and SpfP (32), which are dually encoded on functional bacterial small RNAs (sRNAs), regulate metabolism of alternative carbon sources. SEP upregulation can therefore potentiate cellular stress responses by regulating metabolic reprogramming to sustain homeostasis.

When *E. coli* cells are exposed to a rapid increase in environmental temperature to at least 42 °C, the cellular heat shock response directs resources to de-aggregate, fold, or proteolyze nascent and denatured proteins (33–35). The heat shock response is regulated by σ^{32}, encoded by *rpoH*, which binds to RNA polymerase to increase transcription of heat shock proteins (HSPs) aiding in protein quality control (36, 37). In addition to the canonical HSPs (35), expression of some bacterial SEPs are upregulated in response to heat shock (25).

GndA is a 36-aa SEP with a single predicted alpha-helix that is upregulated during heat shock in *E. coli* (38). The initial detection of GndA was achieved by comparative proteomics of SEPs from the membrane fraction of heat-shocked cells. GndA is encoded in a sORF nested within the CDS of the *gnd* gene in the +2 reading frame (Fig. 1*A*). The UUG start codon of the *gndA* sORF was subsequently identified with retapamulin-enhanced Ribo-seq (Ribo-RET) (19). In addition to GndA, the *gnd* locus canonically encodes 6-phosphogluconate dehydrogenase (6PGD), an enzyme of the pentose phosphate pathway. The 6PGD homodimer catalyzes decarboxylation of two equivalents of 6PG to produce ribulose 5-phosphate coupled to the release of CO_2 and reduction of two equivalents of $NADP^+$ (39). However, the function of the overlapping *gndA* sORF was unknown. In this work, we sought to establish the function of the GndA SEP in the context of the bacterial heat shock response, as well as the relevance of its nested gene architecture. Nested sORFs, such as MINAS-60 in mammals (23) and *comS* in *B. subtilis* (40–42), are associated with phenotypes related to the main ORF in which they are encoded, suggesting that GndA may have a role broadly related to that of the *gnd* gene it is nested within. Our data support the hypothesis that GndA associates, directly or indirectly, with RCI and promotes cell growth and homeostasis at elevated temperature, suggesting that the *gnd* locus encodes two proteins involved in metabolism.

Results

Given that GndA is nested within 6PGD and upregulated during heat shock, we hypothesized that this SEP might play a role in the heat shock response in *E. coli*, and further that its function may be related to metabolism during stress. To determine what its role may be, we first examined its cellular localization. Many functional SEPs exhibit single transmembrane helices (43–45), and AlphaFold 3 (46) predicts the GndA peptide to form an alpha helix extending from residues 7 to 35 (Fig. 1*B*), consistent with our prior bioinformatic prediction of an alpha-helix in the C-terminal region of GndA (38). Our earlier study also found that GndA co-sediments with the plasma membrane (38), consistent with the presence of a transmembrane helix. We therefore carried out subcellular fraction, revealing association of GndA with the inner membrane when overexpressed from a pET vector in *E. coli* BL21(DE3) cells grown at 37 °C, and with both the inner membrane and the pellet during heat shock (Fig. S1*A*). It is likely that the sedimentation with the pellet during heat shock is a consequence of overexpressed GndA oligomerization or aggregation (as also documented for the mito-SEP mitoregulin (47, 48)) and may not reflect its endogenous localization. Next, we wished to determine whether GndA and

the protein it overlaps, 6PGD, are co- or oppositely regulated. We previously reported that the GndA SEP is upregulated during heat shock (38). We therefore examined levels of the 6PGD protein. Endogenous 6PGD remained static at the protein level before and after heat shock (Fig. S1*B*), indicating decoupled expression of 6PGD and GndA during heat shock.

While evolutionary conservation is not required for SEP functionality (16), many functional SEPs are conserved (49–51), and conservation is a robust metric to identify SEPs with phenotypic effects (52). We thus hypothesized that if GndA homologs are present in multiple bacterial lineages, it is likely to play a role in cells. To examine *gndA* conservation, we first performed a translated nucleotide BLAST (tblastn) search using the *E. coli* GndA SEP sequence. We then retrieved the genomic DNA sequences of species that returned putative homologs in order to manually determine whether syntenic Shine Dalgarno sequence, start and stop codons were present. 23 GndA homologs with > 50% sequence positives were identified in Gram-negative bacteria (Fig. *1C-D*, Fig. S1*C*). The evolutionary conservation and transmembrane helix of GndA are consistent with functionality of this SEP in bacterial cells.

Many SEPs bind to and regulate the function of macromolecular complexes (53), leading us to examine the GndA interactome. We observed that GndA is highly insoluble, so attempts to immunoprecipitate endogenous or near-endogenously expressed, FLAG-tagged GndA were unsuccessful. We therefore co-immunoprecipitated overexpressed (pET28a) FLAG-GndA from *E. coli* BL21(DE3) cells followed by quantitative proteomics. This experiment revealed enrichment of multiple endogenous proteins, including several *nuo* operon proteins (Fig. 2*A* and Table S2). The *nuo* operon encodes the protein components of respiratory complex I (RCI), NADH:quinone oxidoreductase, or NDH-1, an inner membrane complex that catalyzes the first step in the electron transport chain (54–59). We also detected enrichment of NuoJ by GndA (Table S2), which we focused on for follow-up validation because it, like GndA, is a transmembrane protein. The enrichment of NuoJ-Myc by FLAG-GndA was validated by co-immunoprecipitation and western blotting when both proteins were overexpressed in *E. coli* (Fig. 2B). GndA therefore associates with RCI when it is overexpressed in *E. coli*, though further experimentation will be necessary to determine whether this association is direct or indirect, and whether it occurs at endogenous expression levels when GndA is expressed during heat shock.

We created a series of strains to examine the function of GndA in heat shock, as well as to separate the phenotypes associated with co-encoded GndA and 6PGD (Table S3, Fig. S3*A-I*). *E. coli* K-12 MG1655 cells (WT) were engineered to lack the entire *gnd* gene and thus both CDSs (gKO), to delete 6PGD alone (6KO-GndA$_f$), or to specifically abrogate the GndA reading frame (aKO-6PGD$_f$, which also bears a S116L mutation in the 6PGD reading frame). Additionally, to assign phenotypes to the *gndA* sORF, rescue cells in which GndA was expressed from an arabinose-inducible plasmid were created from aKO-6PGD$_f$ (aKO-Rescue, Fig. S2). We first tested the hypothesis that GndA may have a role in cell growth at elevated temperature. aKO-6PGD$_f$ cells (ΔGndA) exhibit a reduced growth phenotype of 37.8 ± 6.2% during stationary phase (20 h) at 45 °C relative to WT cells (Fig. 3*A*) that is rescued in aKO-Rescue cells wherein GndA is reintroduced on an arabinose-inducible plasmid (Fig. 3*B*). A similar growth defect of 40.3 ± 7.8% during stationary phase was observed when only 6PGD was deleted (6KO-GndA$_f$, Fig. S3*A*), suggesting that both ORFs independently support growth at high temperature. Cells sustaining a deletion of both ORFs (gKO) exhibit a lesser growth defect of 11.0 ± 2.7% during stationary phase relative to WT cells (Fig. 3*C*), possibly indicating a partial rescue of the *gndA* deletion phenotype by loss of 6PGD. In accordance with a specific role for GndA during growth at elevated temperature, no strains or controls exhibit changes in growth at 30 °C (Fig. S4*B*). The status of *tolC*, which was leveraged to edit the *E. coli* genome (see Materials and Methods), had a small deleterious effect on growth of 6.2 ± 1.8% at

stationary phase at 45 °C relative to WT cells (Fig. S4*C*). These results indicate that GndA contributes to fitness at elevated temperature.

We sought to further investigate the subcellular localization and association of GndA and NuoJ using fluorescence microscopy. Biarsenical dyes have previously been applied to visualization of endogenous *E. coli* and *S. flexneri* proteins (60, 61). We created *E. coli* strains bearing fusion of the tetracysteine binding peptide to the N-terminus of the genomic copy of *gndA* ($GndA_t$), as well as to the C-terminus of the endogenous copy of *nuoJ* ($NuoJ_t$) (Table S3, Fig. S3*G-I*). At 45 °C, GndA fluorescence was observed in puncta at the cell poles (Fig. 4*A*, S5), as was endogenous NuoJ (Fig. 4*B*, S6). Deletion of *gnd* led to a 2.2-fold increase in NuoJ puncta at high temperature, which mislocalize throughout the cell (Fig. 4*C-D*, S7*A-C*). Tagged GndA and NuoJ therefore both localize to the cell pole at 45 °C. Furthermore, loss of *gnd* alters NuoJ localization at high temperature, consistent with the hypothesis that GndA directly or indirectly affects RCI function during heat shock.

Given that GndA co-immunoprecipitation enriches RCI, we hypothesized that GndA might affect cellular respiration and ATP generation at high temperature. RCI catalyzes the oxidation of NADH to NAD^+, concomitant with reduction of ubiquinone to ubiquinol through the coordinated action of approximately eight iron-sulfur clusters, resulting in the pumping of four protons into the periplasm, which generates the electrochemical gradient required for ATP synthesis (62). We therefore used luminescence-based assays to assess relative ATP concentrations and NAD^+/NADH ratios in cells expressing or lacking GndA. We observed a statistically significant 10.9% reduction in ATP at 45 °C in aKO-$6PGD_f$ cells (ΔGndA), and a return to wild-type ATP levels at 45 °C when GndA is reintroduced on an arabinose-inducible plasmid in these cells (Fig. 5*A-C*). No statistically significant change in ATP at 45 °C was observed when the entire *gnd* gene, or 6PGD alone, is deleted (Fig. 5*D-E*). This suggests that GndA has a small but supportive role in aerobic ATP generation at high temperature. Another indicator of aerobic respiration is a high cellular ratio of NAD^+/NADH, which decreases during anaerobic growth (63). In ΔGndA cells, we observed 44% and 27% decreases in NAD^+/NADH ratios relative to WT cells at 30 °C and 45 °C, respectively (Fig. 6*A-B*), suggesting impaired aerobic respiration. Decreases in NAD^+/NADH ratios of 87.7% and 87.2% at 30 °C and 45 °C, respectively, were observed in gKO cells, consistent with a transition to anaerobic respiration in the absence of both GndA and 6PGD (Fig. 6*C*). In addition, a significant difference in NAD^+/NADH ratios between 30 °C and 45 °C was observed in aKO-$6PGD_f$ cells (ΔGndA). These changes in cellular ATP and NAD^+/NADH ratios support a role for GndA in aerobic ATP generation, and a shift to anaerobic respiration if both *gnd* gene products are absent.

In order to maintain homeostasis, gene regulatory networks are reshaped in response to stress and genetic perturbations. We hypothesized that the transcriptional state of cells would provide additional insight into the role of GndA in the heat shock response. We performed RNA sequencing (RNA-seq) in WT, aKO-$6PGD_f$, aKO-Rescue, and gKO cells immediately after heat shock (Fig. 7). Principal component analysis reveals clustering of individual RNA-seq replicates, as expected (Fig. S8). aKO-$6PGD_f$ cells (ΔGndA) exhibit upregulation (> 3.5-fold change and p-value < 0.05) of 116 genes upon heat shock relative to WT cells (Fig. 7*A*). 53 of these genes return to WT expression profiles (< 2-fold change) in aKO-Rescue at 45 °C, demonstrating that their upregulation is a direct consequence of the lack of GndA (Fig. 7*B*). Interestingly, upregulation of these genes is partially rescued in gKO cells lacking both GndA and 6PGD (Fig. 7*C*). Gene ontology (GO) analysis of genes upregulated in ΔGndA cells during heat shock reveals involvement in multiple pathways including cellular response to oxidative stress (Fig. 7*D*). Expression values and significance for specific oxidative stress genes are shown in Fig. 7*E*. Ablation of GndA leads to upregulation of oxidative stress genes *acnA* (aconitate

hydratase), *sodA* (superoxide dismutase), *ydeI* (stress protein), *katE* (catalase), and *sufABCD* (iron-sulfur cluster assembly complex) during heat shock (Fig. 7*A*-*C*,*E*). These genes are involved in general stress responses, as well as detoxification of reactive oxygen species (ROS), which are predominantly generated by RCI (64). Notably, *nuoJ* was also upregulated in aKO-6PGD$_f$ cells, and *nuoJ* expression level returned to WT values in aKO-Rescue cells (Fig. 7*D*), perhaps suggesting that *nuoJ* expression levels increase as a compensatory response to impaired RCI function at high temperature in the absence of GndA (see Discussion). It is also possible that GndA plays a more general role in cellular stress responses, and that other stress response pathways are upregulated to compensate for its absence. We conclude that the *E. coli* transcriptome is reshaped at high temperature in the absence of GndA to restore homeostasis, possibly as a result of RCI malfunction, or a more general defect in stress response pathways, under these conditions.

Discussion

In this work, we provide evidence for the cellular and molecular functions of the 36-aa GndA SEP. GndA was previously discovered as a membrane-associated SEP whose coding sequence is nested in an alternative reading frame within the 6PGD CDS (38). Internal, overlapping ORFs, once thought to primarily occur in viral genomes for efficient packing within viral capsids (65, 66), have now been identified in bacterial and eukaryotic genomes spanning evolutionary space (67). While only two bacterial nested out-of-frame ORFs were known prior to 2019 (*comS* in *B. subtilis* (40–42) and *rpmH* in *T. thermophilus* (68)), a recent study utilizing an inhibitor of initiating ribosomes identified > 70 nested, out-of-frame ORFs in *E. coli*, and mapped the start codon of *gndA* (19). The presence of overlapping frameshifted ORFs in cellular life is surprising, because these organisms are not under pressure to maintain compact genomes (69), and mutations creating start codons in alternative reading frames could be deleterious in the annotated reading frame. Why, then, might overlapping ORFs exist in bacteria and eukaryotes? Several recent studies suggest that SEPs and the canonical proteins that they overlap affect the same phenotype, either synergistically or antagonistically, despite their distinct amino acid sequences (23, 49, 70, 71). This suggests that overlapping gene arrangements may promote interconnected expression to tune phenotypic outputs, providing molecular logic as to how overlaps can benefit fitness. The two proteins encoded in the *gnd* locus, GndA and 6PGD, fit this model, as both may be involved in metabolism. Importantly, GndA is upregulated, while 6PGD does not change, during heat shock. We propose that the co-encoding of GndA within 6PGD enables decoupled expression of these two proteins, with each protein supporting metabolic homeostasis under the conditions in which it is expressed. It will be of future interest to delineate how the expression of 6PGD and GndA are independently controlled.

The question of how overlapping ORFs arise in bacteria remains to be addressed. Notably, while we demonstrate that GndA is present in multiple Gram-negative species, it is less widely distributed than 6PGD, which is conserved from bacteria to humans. We propose that the *gndA* sORF may have arisen *de novo* within the coenzyme domain of 6PGD (15, 72). The *gnd* locus has previously been found to be polymorphic within *E. coli*, possibly due to its proximity to the O-antigen-encoding *rfb* region (73–78). We therefore speculate that mutations in a tolerant region of the 6PGD coenzyme domain resulted in emergence of *gndA* in the +2 reading frame. It will be of interest to test this hypothesis in the context of GndA as well as other nested out-of-frame sORFs in *E. coli* in the future.

GndA enriches several subunits of RCI from cell extracts, and colocalizes in puncta with RCI subunit NuoJ in *E. coli* cells at high temperature. In addition, GndA and RCI (57, 62, 79) both localize to the inner membrane. We therefore conclude that GndA is associated with RCI,

though it is not yet clear whether this association is direct or indirect. Future studies will be required to determine whether GndA binds to a specific component of RCI, or to another RCI-associated protein. Nonetheless, our cellular assays provide additional genetic and molecular links between GndA expression and RCI expression and function at high temperature. Remarkably, multiple SEPs in mammalian mitochondria have been reported to associate with respiratory complexes. In addition, the bacterial SEPs CydX and CydH promote the activity of cytochrome *bd* oxidase, a terminal electron acceptor of the respiratory chain (80, 81). While bacterial GndA, CydX and CydH, and mito-SEPs MOCCI and BRAWNIN are distinct in their sequences, binding partners, and mechanisms, it is possible that a process of convergent evolution has repeatedly recruited SEPs to support the assembly and activity of respiratory chain complexes in multiple organisms.

The transcriptional changes observed in response to elevated temperature in ΔGndA cells provide a snapshot of how *E. coli* may attempt to maintain homeostasis in the absence of GndA. We observe increased expression of genes involved in cellular stress responses (Fig. 7*A*), and GO analysis of transcripts upregulated in heat shocked aKO-6PGD$_f$ cells reveal enrichment for oxidative stress response proteins (Fig. 7*D*, red). ROS have previously been detected in *E. coli* after exposure to elevated temperatures (82). Additionally, heat injured *E. coli* struggle to grow in aerobic conditions unless they are first allowed to recover anaerobically (83). Considering the links between heat shock, oxidative stress, and RCI, and integrating our biochemical and transcriptomic data, we propose that GndA promotes proper RCI function during heat shock, and that in its absence, aberrant RCI function at high temperature leads to ROS production and decreased ATP generation. *E. coli* cells may compensate for GndA loss by upregulating RCI expression and switching to an anaerobic metabolic program to rescue cellular ATP levels—also explaining the small effect on overall cellular ATP levels observed in cells lacking GndA expression. However, whether and how GndA directly affects RCI is not yet clear, and we cannot fully exclude the alternative hypothesis that GndA plays a more general role in the heat shock response, which indirectly alters RCI expression and cellular metabolism. Regardless, this study demonstrates a role for GndA in the response to heat shock that future investigation should continue to elucidate.

In conclusion, GndA functions in cellular stress response and this work provides further evidence that overlapping, frameshifted ORFs may broadly encode proteins with important cellular functions.

Materials and Methods

Bioinformatics

Next-generation sequencing (NGS): For each sample type consisting of 3 replicates each, 2 µg of extracted RNA per replicate was submitted to the Yale Center for Genome Analysis, where ribosomal RNA depletion and further sample processing was performed. Sequencing was performed with the Illumina NovaSeq system at a depth of approximately 20 million reads per sample. Paired-end reads were sequenced with a read length of 150 base pairs.

Computational analysis of sequencing results: Sequencing data was obtained in gzipped fastq files (.fastq.gz). Read files were trimmed to deplete adapter sequences using bbduk.sh, part of the BBMap array of tools (84). The resulting trimmed read files were then assessed for quality control with fastQC. Subsequently, the reads were mapped to the *E. coli* str. K-12 substr. MG1665 genome using STAR (85). The resulting alignment files (.BAM) were analyzed using featureCounts (86).

The featureCounts program outputs a tab-separated column (.tsv) file. Microsoft Excel was used for downstream analysis. To normalize for gene length, reads per kilobase (RPK) values were calculated for each replicate by dividing the reads for each gene by the length of the gene in kilobases. To normalize for read depth, first the scaling factor for each replicate was calculated by summing up all RPK values and dividing by 10^6. Then, transcripts per million (TPM) values were calculated by dividing the RPK value for each gene by the scaling factor for the replicate. TPM values were used in all subsequent analyses (*i.e.*, volcano plots).

GndA homology studies: tblastn searches were run against the *E. coli* K-12 GndA amino acid sequence (87). Nucleic acid sequences of top hits were aligned with the Clustal Omega multiple sequence alignment program (88).

Protein secondary structure prediction: The AlphaFold server was used to generate a predicted structure for GndA based on the *E. coli* K-12 GndA amino acid sequence (46).

Protein extraction and co-immunoprecipitation (co-IP)

Protein extraction: Equal numbers of cells normalized by OD_{600} were taken from control and heat shocked cultures and pelleted by centrifugation at 21 °C at 4,000 rpm for 5 min. Media was decanted and pellets were resuspended in 1 mL of BPER lysis buffer (Thermo Scientific catalog# 90078) containing DNase I, Lysozyme, and Roche complete protease inhibitor (Sigma catalog #11836170001). Cells were lysed by sonication with a QSonica Misonix Microson Ultrasonic Cell Disruptor XL-2000 for 5 rounds, 5 s on 25 s off, on ice at medium (5) setting. Lysate was clarified by centrifugation at 4 °C at 21,000 rcf for 5 min. Clarified lysate containing solubilized protein was removed from the tube with pelleted material and transferred to a fresh tube.

Co-immunoprecipitation (co-IP): For each clarified lysate, 25 µL of anti-FLAG M2 affinity agarose gel (Sigma catalog #A2220) was washed in 1 mL ice-cold BPER and collected at 1,000 rcf for 1.5 min at 4 °C. Wash supernatant was removed without disrupting agarose gel, and the protein containing clarified lysate was added. The protein lysate was incubated with agarose gel at 4 °C with rotation for 1 h. Samples were spun down at 1,000 rcf for 1.5 min at 4 °C to collect beads bound to FLAG fusion protein and binding partners. To remove nonspecific binding, supernatant was removed and beads were washed twice with TBS containing 1% Triton X-100 (Sigma catalog #T8787) and 500 mM NaCl at 4 °C with rotation for 3 min. A final wash with 10 inversions by hand was done in TBS containing 1% Triton X-100 and 150 mM NaCl to return proteins to physiological salt concentration. Bound protein was eluted off agarose gel with 30-50 µL 1X 3XFLAG peptide (Sigma catalog #F4799) at 4 °C with rotation for 1 h. Agarose beads were collected at 4 °C 1,000 rcf for 2 min, and supernatant containing immunoprecipitated protein was removed and stored at -80 °C for downstream western blotting and LC-MS/MS.

pETDuet-1 Co-IP: pETDuet-1 expression vectors encoding 1) FLAG-GndA alone, 2) NuoJ-Myc alone, or 3) both constructs, were transformed into BL21(DE3) cells. Single colonies were inoculated in Luria-Bertani (LB) broth with 100 µg/mL carbenicillin (Sigma catalog #C1389-5G) overnight. After a 1:100 dilution and growth to exponential phase (OD = 0.3-0.5), T7 protein expression was induced at 37 °C with 0.5 mM IPTG. Total protein was extracted by sonication and lysates were clarified twice by centrifugation at 20,000 rcf for 3 min at 4 °C. FLAG-GndA was immunoprecipitated with anti-FLAG M2 affinity agarose gel and samples were eluted with 3XFLAG peptide. Samples were run on both 15% Glycine-SDS-PAGE to resolve NuoJ, and 16% Tricine-SDS-PAGE to resolve GndA. DYKDDDDK tag rabbit monoclonal antibody (Cell Signaling catalog #14793S) and Myc tag rabbit polyclonal antibody (Rockland catalog #600-401-381) were used for Western blotting primary immunostaining. Secondary staining was

performed with horseradish peroxidase conjugated rabbit IgG (H&L) (Rockland catalog #611-1302) and visualized with clarity western ECL substrate (BioRad catalog #1705061).

Dual selection for bacterial genome editing

Scarless mutants at the *gnd* and *nuo* loci were generated with an SDS/colicin E1 dual-selectable system (89, 90). WT *E. coli* K-12 MG1655 cells were transformed with a ColE1 kanamycin-resistant plasmid (91) carrying anhydrotetracycline-inducible λ-Red recombineering genes (Exo, Beta, and Gam). To prepare cell cultures for gene editing, a single colony streaked out on LB agar with kanamycin was inoculated in LB broth with 50 µg/mL kanamycin at 37 °C and shaken at 200 rpm overnight. The culture was diluted 1:100 in LB supplemented with 50 µg/mL kanamycin, and 100 ng/mL of anhydrotetracycline (Sigma catalog #37919) was added after 45 min of incubation to induce the λ-Red recombineering genes. Cultures were grown to OD_{600} = 0.5 and chilled on ice for 20 min. All tips, tubes, cuvettes, and autoclaved milli-Q water were chilled on ice. 1 mL of chilled culture was spun down at 3,000 rpm for 5 min at 4 °C and washed 3 times with chilled milli-Q water. To generate a Δ*tolC* strain from WT MG1655, an oligo mixture was prepared with 1 µL 100 µM oligonucleotide designed to disrupt the *tolC* locus added to 40 µL chilled milli-Q water. Washed cells were gently resuspended in oligo mixture and incubated on ice for 1 min. 50 µL of the cell-oligonucleotide mixture was transferred to a pre-chilled 1 mm electroporation cuvette (BioRad catalog #1652083) and transformed with 1.8 kV, 200 Ω, and 25 µF in a MicroPulser Electroporation Apparatus. Immediately after electroporation, cells were resuspended in 1 mL room temperature SOC medium (Sigma catalog #S1797-10X5ML). 2 mL LB Miller media was additionally added for a total recovery culture volume of 3 mL. Cells were recovered for a minimum of 2 h at 37 °C with 200 rpm shaking to allow turnover of transmembrane TolC before selection. Recovery cultures were diluted 1:100 in LB supplemented with approximately 10 µg/mL colicin E1 (prepared as described previously (92) in strain JC411) and 64 µg/mL vancomycin to select for *tolC* marker displacement, and incubated for 6-12 h at 37 °C with shaking at 200 rpm. Δ*tolC*-selected cultures were plated on antibiotic-free LB agar plates and individual colonies were screened for *tolC* knockout by both negative selection in LB supplemented with 0.005% SDS and colony PCR showing a lack of amplicon from the *tolC* locus.

All mutant strains, including Δ*tolC*, gKO, nKO, 6KO-$GndA_f$, aKO-$6PGD_f$, $GndA_t$, $NuoJ_t$, and gKO-$NuoJ_t$, were generated using sequential *tolC* selection/counterselection. At each target locus (*gnd* and *nuoJ*), the target gene was first knocked out in the MG1655 Δ*tolC* strain by knocking in the *tolC* cassette. This then served as a displaceable marker to select for knock in of tagged and/or mutant genes. First, gKO and nKO strains were constructed via double-stranded DNA (dsDNA) recombineering, replacing the native locus with *tolC*. Recombineering cassettes were generated by amplifying the *tolC* cassette from MG1655 cells, with primers encoding 50-bp overhangs surrounding the gene targeted for deletion (Table S1). PCR products were purified by PCR purification (Qiagen catalog #28104) and full-length cassettes were subsequently isolated by agarose gel purification (Qiagen catalog #28506) and verified with Sanger sequencing. The Δ*tolC* strain bearing the recombineering plasmid was cultured and induced as described above, then electroporated with 1 µL of dsDNA substrate (100 – 400 ng of dsDNA). For positive *tolC* selection, cells were recovered for 2 h in SOC with LB, then plated on LB agar plates supplemented with 0.005% SDS. Individual knockout strains were confirmed for *tolC* integration by colony PCR and Sanger sequencing.

To generate tagged and mutant strains (6KO-$GndA_f$, aKO-$6PGD_f$, $GndA_t$, $NuoJ_t$, and gKO-$NuoJ_t$), the gKO and nKO strains were subjected to another round of dsDNA recombineering to displace the *tolC* marker with the desired mutant sequence. Tagged and mutant cassettes (Table S1) were generated by colony PCR and overlap extension PCR with Phusion

polymerase, or amplification from plasmids purchased from Genscript, and confirmed by Sanger sequencing. gKO and nKO strains were cultured and induced as described above. Purified cassettes were electroporated into the strains, and cells were subjected to colicin E1 and vancomycin counterselection as in the initial *tolC* knockout experiment. After counterselection, mutant colonies were plated on LB agar and tested for desired mutations by negative selection in LB supplemented with 0.005% SDS as well as colony PCR and Sanger sequencing.

Proteomics LC-MS/MS

Quantitative proteomics was performed as previously reported (93). Gel slices containing entire lanes were digested with trypsin at 37 °C for 14-16 h. The resulting peptide mixtures were extracted from the gel, dried, subjected to ethyl acetate extraction to remove residual detergent, de-salted with a peptide cleanup C18 spin column (Thermo Scientific catalog #89870), then resuspended in 35 µL 0.1% formic acid (FA), followed by centrifugation at 15,000 rcf at 4 °C for 30 min. A 5-µL aliquot of each sample was injected onto a pre-packed column attached to an Easy-nLC 1200 (Thermo Scientific) in-line with a Thermo Scientific Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. A 125-min gradient was used to further separate the peptide mixtures as follows (solvent A: 0.1% FA; solvent B: acetonitrile (ACN) with 0.1% FA): linear gradient from 5% B to 45% B over 90 min, 45% B to 85% B over 1 min, isocratic flow at 85% B was maintained for 10 min, followed by a gradient from 85% B to 5% B over 1 min, and isocratic flow at 5% B over 23 min. The full MS was collected over the mass range of 300-1,700 m/z with a resolution of 70,000 and the automatic gain control (AGC) target was set as 3×10^6. MS/MS data was collected using a top 20 high-collisional energy dissociation method in data-dependent mode with a normalized collision energy of 28.0 eV and a 1.6 m/z isolation window. MS/MS resolution was 17,500 and dynamic exclusion was 90 s. Files were analyzed using MaxQuant, and oxidation of methionine and N-terminal acetylation were set as variable modifications. *E. coli* K-12 Uniprot plus NuoJ-FLAG and GndA was used as the database for searching. Protein quantitation was accomplished with MaxQuant LFQ (version 2.0.2.0) or spectral counting [PMID: 24610814].

Bacterial heat shock

Single colonies were inoculated in LB broth and grown overnight with shaking at 200 rpm at 30 °C or 37 °C. Cultures were then diluted 1:100 in fresh LB broth and grown to exponential phase (OD_{600} = 0.3-0.5). Each culture was split into two. One half (control) was returned to the original incubator shaker (30-37 °C). The other half (heat shock culture) was warmed for 5 min in a 45 °C water bath with shaking by hand until an internal temperature of 45 °C was reached. Then, the heat shock culture was kept in a 45 °C water bath for 20 min with shaking by hand. Control and heat shock cultures were harvested simultaneously. Control cultures were grown at 30 °C to observe the heat shock effect over a larger temperature change.

RNA-seq

RNA extraction after heat shock was carried out using the RNeasy Protect Bacteria kit (Qiagen catalog #74524). 5×10^8 *E. coli* cells were harvested from each sample using an RNeasy mini spin column. Qiagen protocol 4 was followed for enzymatic lysis with lysozyme (DOT Scientific catalog #DSL38100) and 20 µL proteinase K (NEB catalog #P8107S). Qiagen protocol 7 was followed for RNA purification including DNase I digestion on-column (Qiagen catalog# 79254). Total RNA was eluted in 50 µL RNase free water.

ATP luminescence assays

BacTiter-Glo Microbial Cell Viability Assay (Promega catalog #G8230) was used to compare total cellular ATP in WT and GndA mutant lines between growth at 37 °C and heat shock conditions. Promega protocols 3.A and 3.B were followed. After inoculating a single colony per

replicate in LB and growth overnight at 37 °C at 200 rpm in an incubator shaker, cultures were diluted 1:100 and grown to exponential phase (OD_{600} = 0.3-0.5). Each culture was split into two and the control half was returned to the incubator shaker while the other half was subjected to heat shock (see above). Immediately after heat shock all cultures were normalized to OD_{600} = 0.4 with room temperature LB to a final volume of 2 mL. 100 µL per replicate was aliquoted into a 96-well white Lumitrac polystyrene microplate (Greiner catalog# 655075), and 100 µL of BacTiter-Glo reagent was added to each well. The plate was loaded into a BioTek Synergy Neo2 Multimode Reader set to 23 °C. Orbital shaking at 425 cpm was applied for 10 s. Luminescence was recorded once per min for 20 min with an integration time of 40 ms per well, read height 4.5 mm, and gain of 135. BacTiter-Glo reagent was added to 100 µL LB to measure background luminescence which was subtracted from experimental readings. The 5-min timepoint was used for analysis of four replicates.

NAD^+/NADH luminescence assays

NAD/NADH-Glo Assay (Promega catalog #G9071) was used to measure NAD^+/NADH ratios in WT and GndA mutant cell lines for growth at 30 °C and heat shock. Promega protocols 3.A, 3.B, and 5.A were followed with modification. A single colony per replicate was inoculated in LB and grown overnight at 30 °C at 200 rpm in an incubator shaker. Cultures diluted 1:100 were grown to exponential phase (OD_{600} = 0.3-0.5) and split into two. The control half was returned to the 30 °C incubator shaker while the other half was subjected to heat shock. Immediately after heat shock all cultures were normalized to OD_{600} = 0.4 with room temperature LB to a final volume of 2 mL. 25 µL of each culture was added to a 96-well white Lumitrac polystyrene microplate and 25 µL of bicarbonate base buffer with 1% dodecyltrimethylammonium bromide (DTAB) (Sigma catalog #D8638) was added for cell lysis. These 50 µL base treated samples were mixed with agitation by hand. 25 µL of the base treated samples were moved to new wells, and 12.5 µL of 0.4 M HCl was added to generate the acid treated samples. The 96-well plate containing both the acid and base treated samples was covered with a lid and incubated in a metal bead bath that maintained 60-85 °C for 15 min. The samples were brought to room temperature on the benchtop for 10 min. Neutralization of both reactions was achieved by adding 12.5 µL of 0.5 M tris base to acid treated samples (containing cellular NAD^+ only) and 25 µL of HCl/tris solution to base treated samples (containing cellular NADH only). A blank was generated with sterile LB media lacking *E. coli* and subjected to the same acid base treatment to measure background luminescence and subtracted from experimental readings. 50 µL NAD/NADH-Glo Detection Reagent was added to each well. The plate was loaded into a BioTek Synergy Neo2 Multimode Reader set to 23 °C. Orbital shaking was applied for 10 s, and luminescence was recorded every 15 min for 1.5 h. The luminescence reading at 1 h was used to calculate NAD+/NADH ratios for four replicates.

FlAsH imaging

The TC-FlAsH II In-Cell Tetracysteine Tag Detection Kit (Invitrogen catalog #T34561) was used to obtain FlAsH-EDT_2 and dimercaprol/British anti-Lewisite (BAL) reagents for in vivo FlAsH labeling (94, 95). *E. coli* cells were mutated to endogenously express TC-GndA or NuoJ-TC chromosomally, appended with a tetracysteine (TC) tag sequence (FLNCCPGCCMEP) optimized for in vivo FlAsH imaging at low levels of protein expression (60, 96). Cells were streaked out on antibiotic free LB agar plates and grown overnight at 30 °C. A single colony was inoculated in LB and grown overnight in an incubator shaker at 30 °C at 200 rpm. Overnight cultures were diluted 1:100 in LB and grown to an OD_{600} of 0.4-0.5 at 30 °C with shaking at 200 rpm. Cultures were then pelleted at 4,000 rpm for 5 min at 21 °C. LB media was decanted and cell pellets were gently resuspended in M9 minimal media (Sigma catalog #M6030) supplemented with MEM Vitamins Solution (100X) (Gibco catalog #11120052) and MEM Amino Acids Solution (50X) (Gibco catalog #11130051) to reduce background fluorescence while

imaging *E. coli* (60). M9 Cultures were heat shocked at 45 °C as stated. After heat shock, cultures were normalized to OD_{600} = 0.4 for FlAsH labeling. TCEP was added to a final concentration of 1 µM to protect the TC motif, and FlAsH was added to a final concentration of 2.5-20 µM (61) to 1 mL of each OD_{600} = 0.4 culture and immediately removed from light. *E. coli* cells were incubated in the dark with FlAsH labeling reagent for 45 min with shaking at 200 rpm at 45 °C in an incubator shaker. FlAsH labeled *E. coli* were spun for 5 min at 21 °C at 4,000 rpm, and the FlAsH labeling solution was decanted. Pelleted cells were washed twice in 2 mL M9 media with 250 µM BAL for 10 min at 45 °C with shaking. A final wash was performed in M9 media to remove residual BAL and maintain in vivo FlAsH TC interactions. Cells were spun down a final time and resuspended in approximately 20 µL M9 media from residual after decanting. M9 2% agarose pads were prepared within gene frames (Thermo Scientific catalog #AB0578) mounted on glass microscope slides. In vivo FlAsH-labeled bacteria were spread onto the agarose pad and cover glass was secured to the gene frame. Samples were imaged on a Leica DMi8 widefield fluorescence microscope at 100X magnification with oil immersion.

Growth curves

WT and GndA mutant *E. coli* cell lines were streaked out on LB agar plates and incubated at 30 °C or 37 °C overnight. aKO-Rescue cells containing a pBAD33 plasmid designed to rescue expression of GndA were grown on LB agar with 25 µg/mL chloramphenicol. Single colonies were inoculated in LB broth (with chloramphenicol for cells carrying pBAD33) and grown overnight at 30 °C or 37 °C with shaking at 200 rpm as appropriate for the experimental control conditions. Cultures were then diluted 1:100 and grown to exponential phase (OD_{600} = 0.3-0.6). For transient heat shock growth curves, exponential cultures were first heat shocked by hand as described, then all samples were normalized to OD_{600} = 0.01, and finally transferred to a sterile round bottom 96-well plate with lid (Corning catalog #3788). Four replicates per cell line were incubated at 30 °C with continuous orbital shaking at 205 cpm while growth was monitored in a BioTek Synergy Neo2 Multimode Reader. To monitor cell growth under prolonged heat shock for greater than 20 min, cultures were brought to 45 °C directly in the plate reader. Cultures were first grown to exponential phase (OD_{600} = 0.3-0.6) and then normalized to OD_{600} = 0.2 and transferred to a sterile round bottom 96-well plate with lid. 0.02% L-(+)-arabinose was added to aKO-Rescue cell cultures immediately before transferring to the 96-well plate. Four replicates per cell line were incubated at 45 °C with continuous orbital shaking while growth was monitored in a BioTek Synergy Neo2 Multimode Reader. OD_{600} readings were taken at 15 min intervals.

Western blotting

Standard western blotting procedures were followed with modification (97) using a Mini-PROTEAN Tetra Cell, Mini-Trans blot module, and power supply (Bio-Rad catalog# 1658033). Protein samples were resolved on 16% Tricine-SDS-PAGE (98) for proteins with molecular weights less than 20 kDa, and 15% Glycine-SDS-PAGE was used for proteins with molecular weights above 20 kDa. 4X SDS loading buffer was added to protein samples extracted from equivalent *E. coli* samples. PAGE was run at 50-80 V as proteins migrated through the stacking gel, and at 100-120 V through the resolving gel until the protein xtra prestained ladder (Bio-Rad catalog # 1610377) achieved good separation. Gels were equilibrated in pre-chilled transfer buffer with 15% methanol and then protein was transferred to a 0.45 µm nitrocellulose membrane (Sigma catalog# 10600002) via wet electroblotting on ice for 1 h at 300 mV. Membrane was blocked in 3% BSA in TBST for 2 h with rocking at room temperature. Primary antibody was added 1:3,000 in 3% BSA in TBST and incubated at 4 °C overnight with nutation. The membrane was washed three times in TBST for 5 min at room temperature with rocking. Secondary antibody was added 1:10,000 in 3% BSA in TBST and incubated 1 h at room temperature with rocking. Three washes in TBST were performed. Chemiluminescent imaging was done with horseradish peroxidase conjugated IgG and clarity western ECL substrate

(BioRad catalog# 1705061). Images were taken on BioRad ChemiDoc XRS+ Gel Imaging System.

Acknowledgments
We thank all current and former members of the Slavoff and Breaker labs for helpful discussions. This work was funded in part by a Mark Foundation for Cancer Research Emerging Leader Award, a Paul G. Allen Frontiers Group Distinguished Investigator Award, and a Sloan Research Fellowship (FG-2022-18417) (to S.A.S.). J.J.M. was supported in part by a Yale University fellowship associated with the NIGMS Chemistry-Biology Interface Training Program (5T32GM067543) and a Roberts Fellowship from the Yale University Department of Chemistry. M.G.M. is a Howard Hughes Medical Institute (HHMI) awardee of the Life Sciences Research Foundation, and received supplementary support from HHMI funds awarded to the Breaker lab.

Figures and Figure Legends

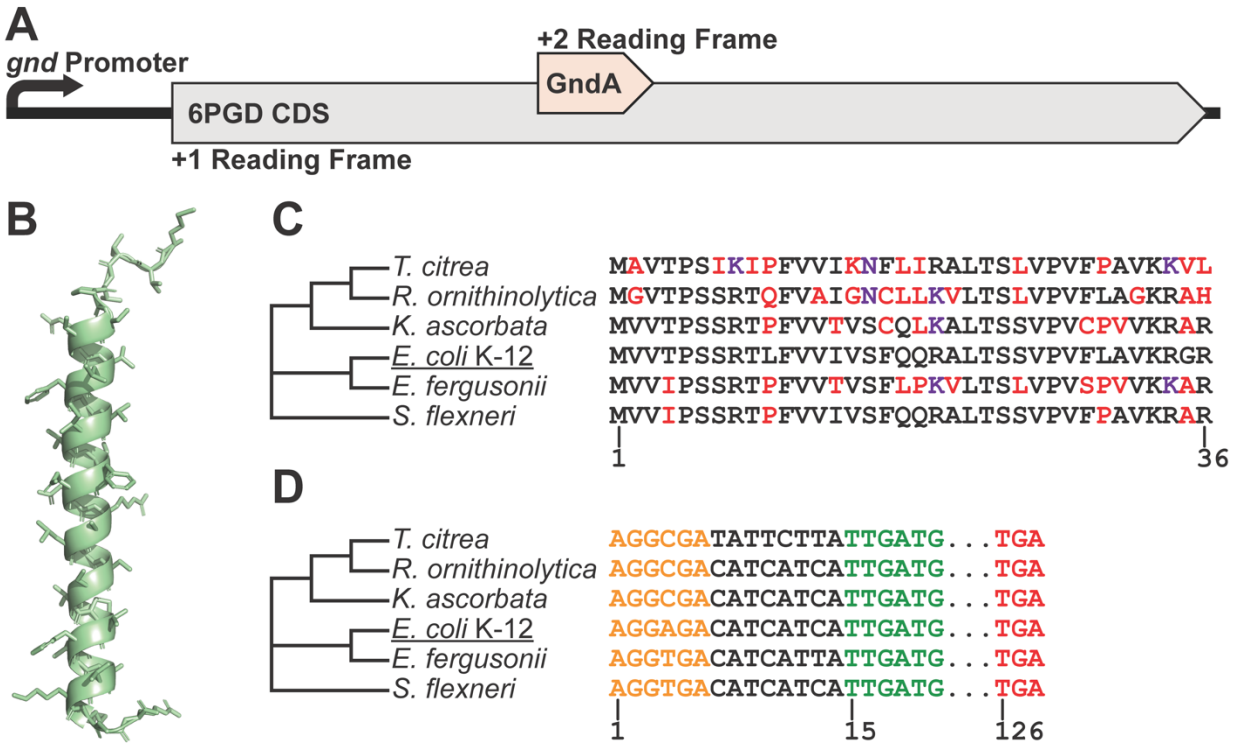

Fig. 1. The *gndA* small open reading frame (sORF) is nested within the 6-phosphogluconate dehydrogenase (6PGD) coding sequence. (*A*) The *E. coli* K-12 MG1655 *gnd* gene locus encodes two proteins with distinct amino acid sequences (38). Schematic of the *gnd* gene with the coding sequence (CDS) for 6PGD defined as the +1 reading frame. The GndA microprotein is encoded by a nested sORF in the +2 reading frame. (*B*) High confidence GndA secondary structure prediction generated with AlphaFold 3 (46). (*C*) GndA conservation in Gram-negative bacteria. Residues in black are conserved, in purple are positives (87), and in red are not conserved. (D) The coding and regulatory sequences for GndA homologs were aligned and Shine Dalgarno sequence (orange), cryptic TTG start codon (green), traditional ATG start codon (green), and stop codon (red) were manually identified.

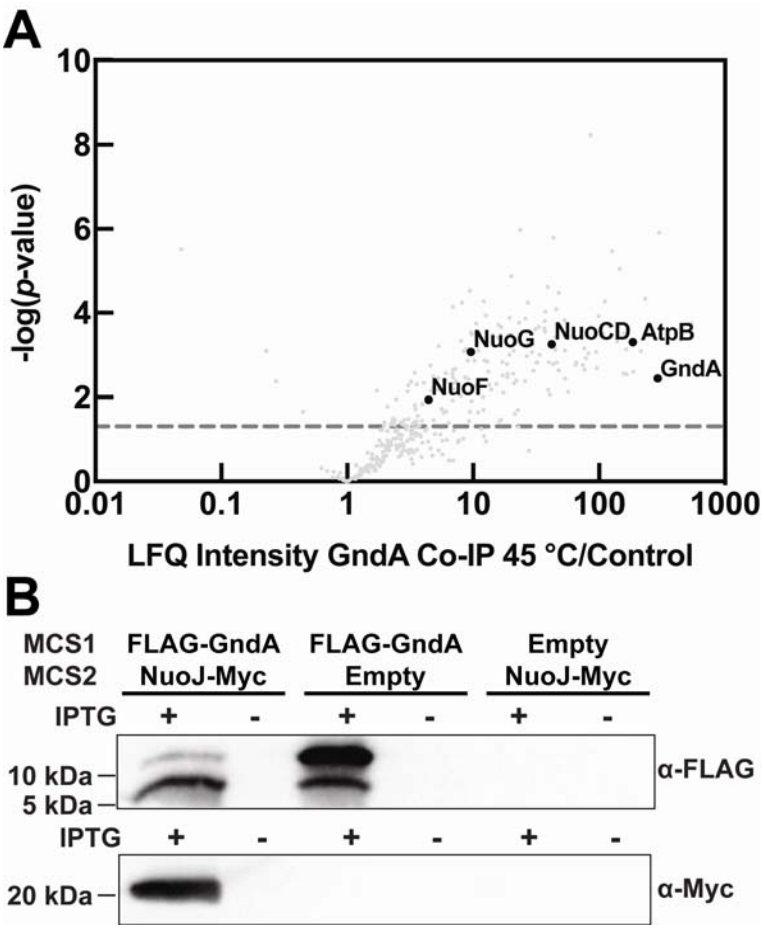

Fig. 2. GndA interaction with respiratory complex I. (*A*) Volcano plot of label free quantitative proteomic analysis of overexpressed FLAG-GndA co-immunoprecipitation with reference to no-induction control. *P*-values were calculated by performing a *t*-test assuming a two-tailed distribution and homoscedasticity (n = 3). (*B*) Western blot depicting NuoJ co-immunoprecipitation with GndA with anti-FLAG M2 affinity gel. FLAG-GndA and NuoJ-Myc were expressed on a pETDuet-1 vector in BL21(DE3) cells. Two control vectors were generated with empty multiple cloning sites to express either FLAG-GndA or NuoJ-Myc exclusively.

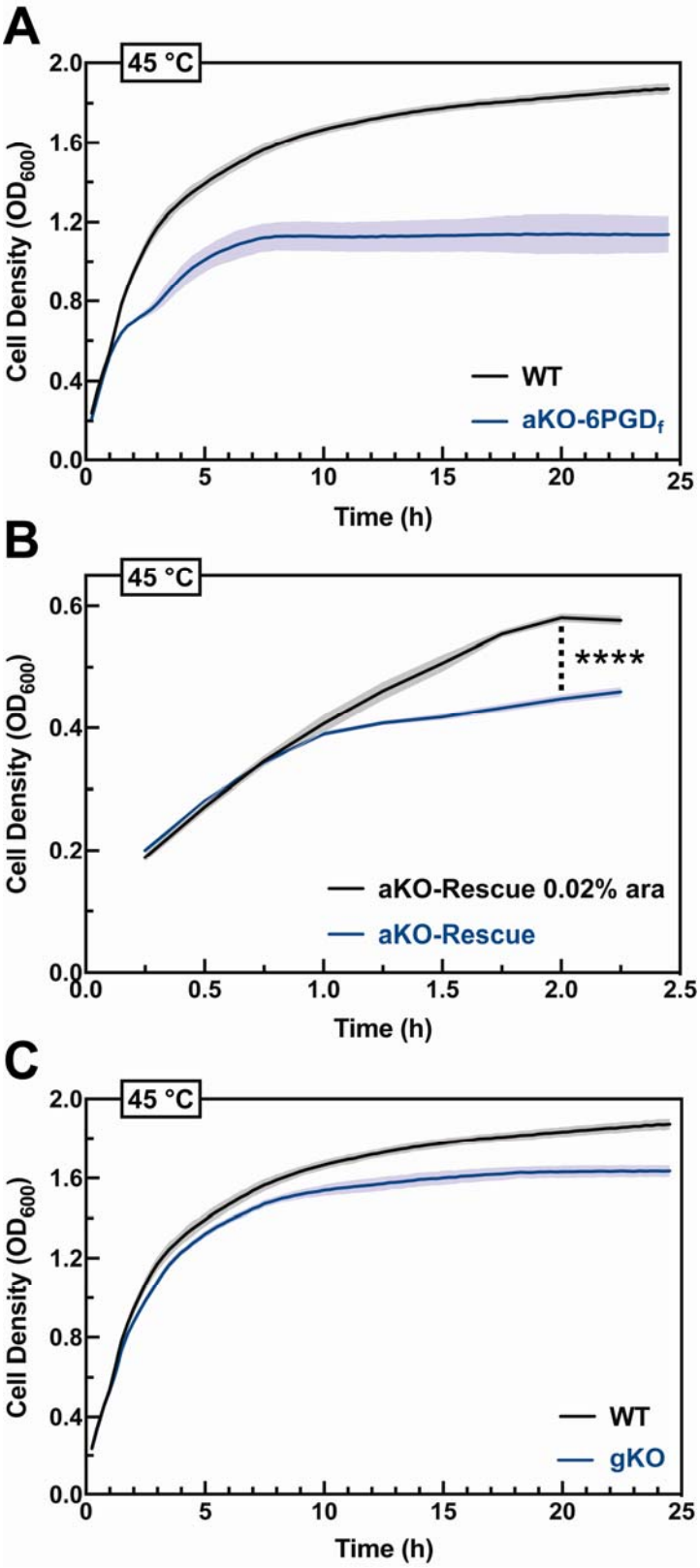

Fig. 3. GndA mutant *E. coli* exhibit a reduced growth phenotype at elevated temperature. (*A*) Growth curves (OD_{600}) for WT and aKO-6PGD$_f$ cells at 45 °C. (*B*) Growth curves comparing aKO-6PGD$_f$ cells transformed with a pBAD33 expression vector encoding GndA (aKO-Rescue)

rescued with 0.02% arabinose at 45 °C. (*C*) Growth curves comparing WT to gKO cells at 45 °C. Error (shading above and below curves) calculated from 4 replicates. **** *p*-value < 0.00005.

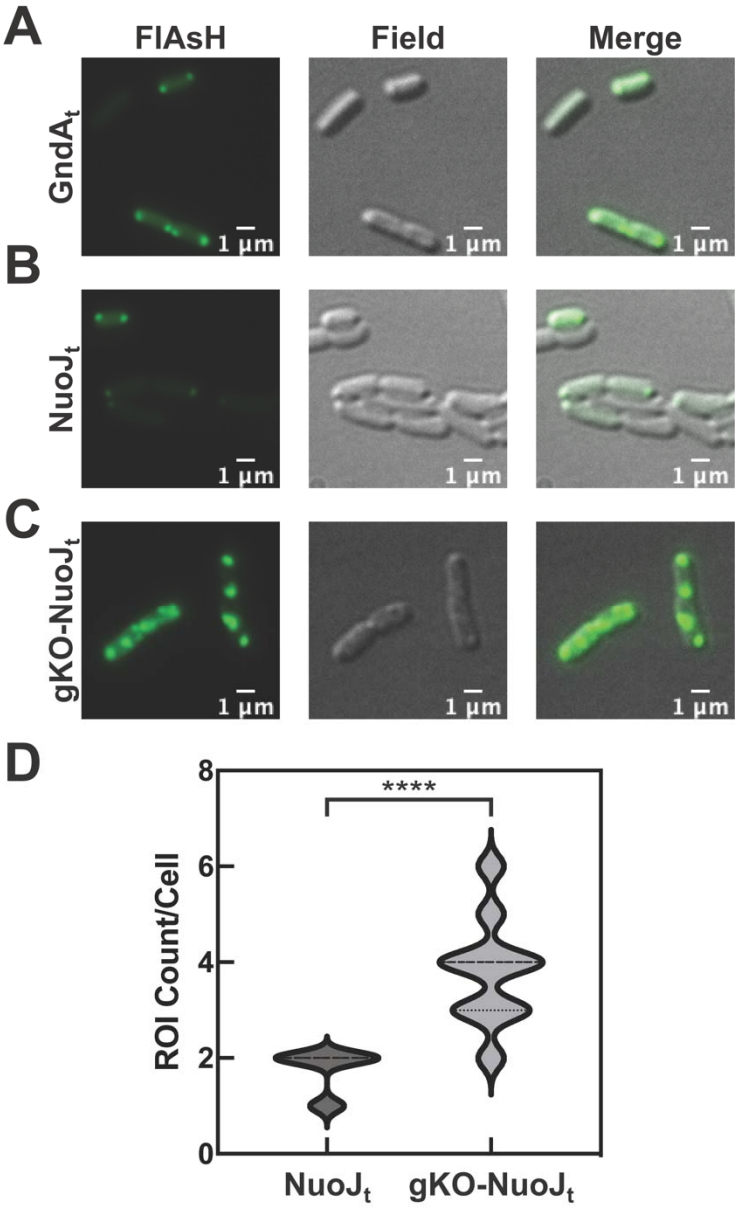

Fig. 4. Fluorescence imaging of chromosomally expressed tagged GndA and NuoJ with FlAsH at 45 °C. (*A*) Confocal microscopy of GndA$_t$ cells expressing tetracysteine-tagged GndA incubated at 45 °C with 20 µM FlAsH after heat shock from replicate two (n = 2). (*B*) Fluorescence imaging of NuoJ$_t$ cells expressing tetracysteine-tagged NuoJ incubated at 45 °C with 20 µM FlAsH after heat shock from replicate two (n = 2). (*C*) Fluorescence imaging of gKO-NuoJ$_t$ cells incubated at 45 °C with 20 µM FlAsH from replicate two (n = 2). (*D*) Quantification of regions of interest (ROI), *i.e.*, NuoJ-TC puncta, from gKO-NuoJ$_t$ replicate one and NuoJ$_t$ replicate two in *E. coli* expressing or lacking the *gnd* gene (n = 20). **** *p*-value < 0.00005.

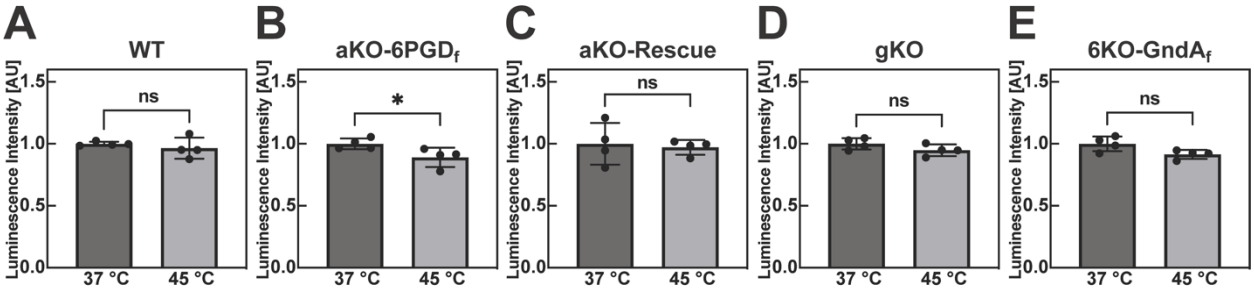

Fig. 5. ΔGndA cells exhibit decreased ATP levels under heat shock. (*A*) Cellular ATP was quantified with a luminescence assay at 37 °C or 45 °C in WT, (*B*) aKO-6PGD$_f$, (*C*) aKO-Rescue, (*D*) gKO, (*E*) and 6KO-GndA$_f$. Values were normalized to the mean value for 37 °C in each experiment. n = 4, * *p*-value < 0.05.

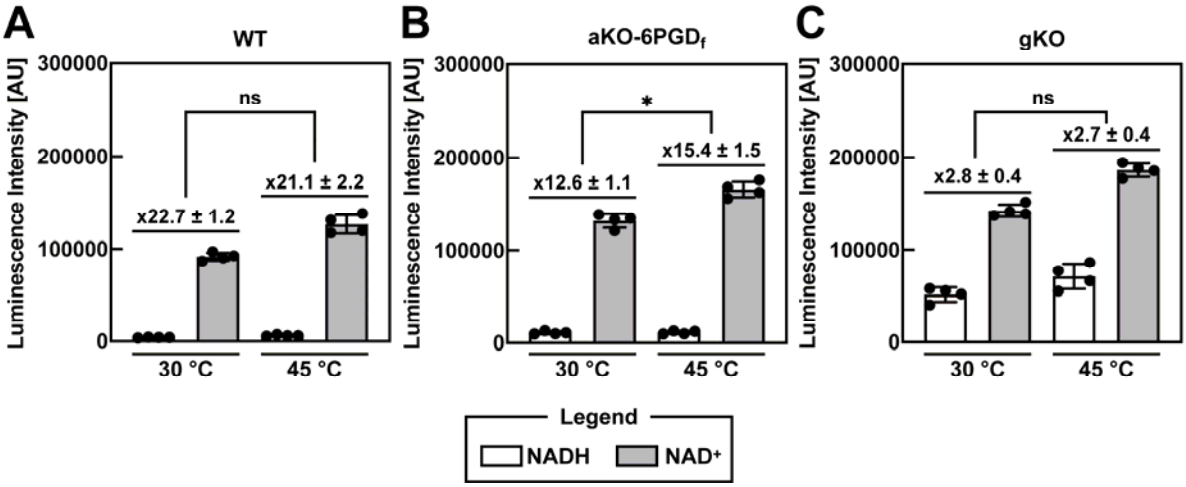

Fig. 6. NAD+/NADH ratio quantitation in GndA and 6PGD mutants. (A) Cellular NAD+ and NADH levels were measured by luminescence assay at 30 °C or 45 °C in WT, (B) aKO-6PGDf, (C) and gKO. NAD+/NADH ratios per experimental temperature were calculated based on 4 replicates and indicated above datapoints. * p-value < 0.05.

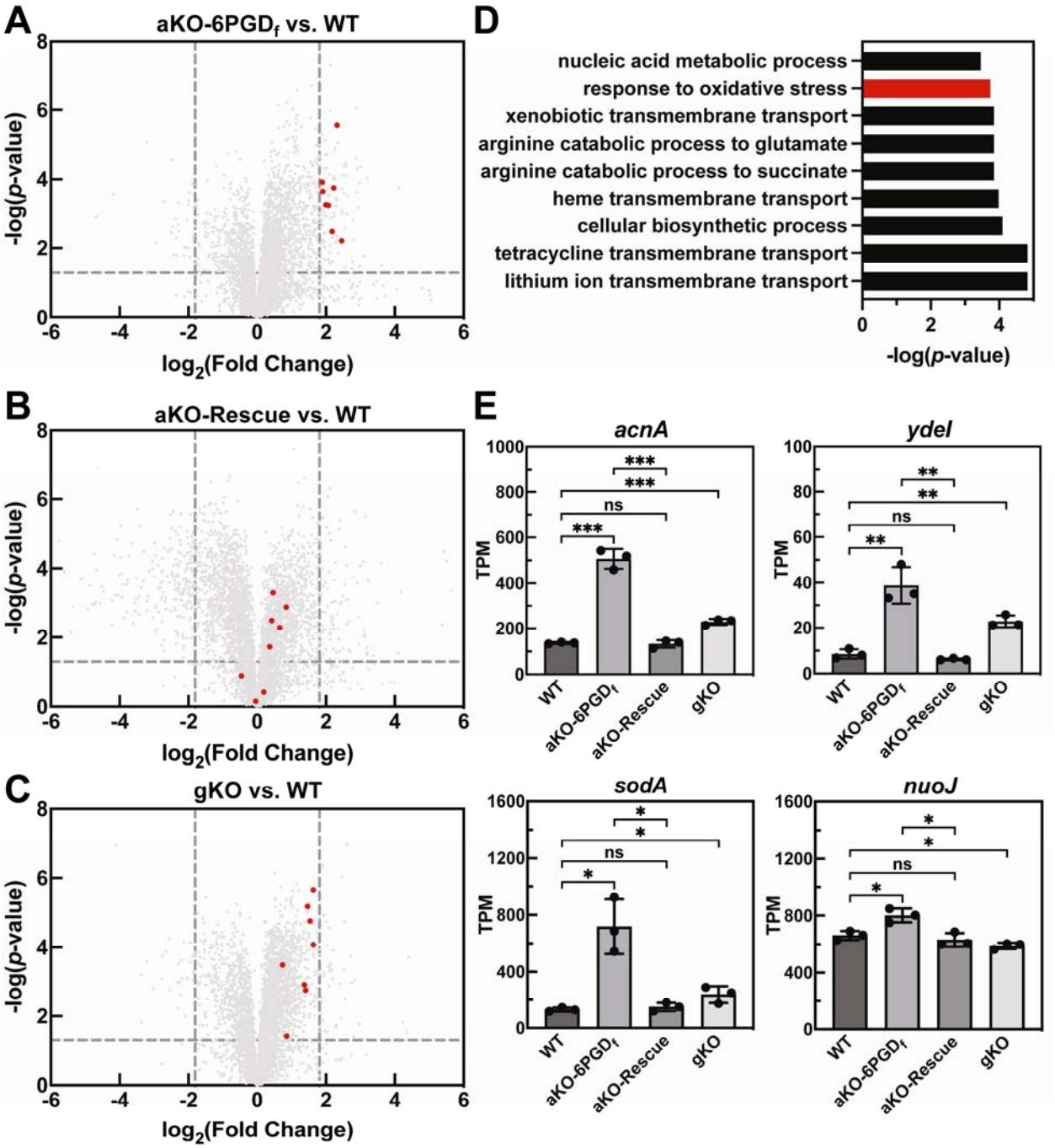

Fig. 7. GndA expression correlates with cellular stress. (*A*) RNA-seq analysis of gene expression changes in aKO-6PGD$_f$ vs. WT, (*B*) aKO-Rescue vs. WT, (*C*) and gKO vs. WT after heat shock. Oxidative stress genes of interest are in red. *P*-values were calculated by performing a *t*-test assuming a two-tailed distribution and homoscedasticity (n = 3). Horizontal dotted line signifies *p*-value of 0.05, and vertical dotted lines signifies a fold change of 3.5. One and twelve datapoints (including *ara* genes) are outside the axis limits of volcano plots *A* and *B* respectively. (*D*) Gene ontology (GO) analysis (99, 100) of TPM fold change greater than 3.5 for aKO-6PGD$_f$ vs. WT after heat shock. (*E*) Comparison of transcripts per million (TPM) levels between WT and GndA mutants for selected metabolic and stress associated genes.

References

1. G. Storz, Y. I. Wolf, K. S. Ramamurthi, Small proteins can no longer be ignored. *Annu. Rev. Biochem.* **83**, 753–777 (2014).
2. M. W. Orr, Y. Mao, G. Storz, S. B. Qian, Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* **48**, 1029–1042 (2020).
3. M. R. Hemm, J. Weaver, G. Storz, *Escherichia coli* Small Proteome . *EcoSal Plus* **9** (2020).
4. P. Garai, A. Blanc-Potard, Uncovering small membrane proteins in pathogenic bacteria: Regulatory functions and therapeutic potential. *Mol. Microbiol.* **114**, 710–720 (2020).
5. J. J. Mohsen, A. A. Martel, S. A. Slavoff, Microproteins—Discovery, structure, and function. *Proteomics* **23** (2023).
6. A. Khitun, T. J. Ness, S. A. Slavoff, Small open reading frames and cellular stress responses. *Mol. Omi.* **15**, 108–116 (2019).
7. W. R. Pearson, An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinforma.*, 1–8 (2013).
8. A. Waterhouse, *et al.*, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
9. P. Puigbò, I. G. Bravo, S. Garcia-Vallve, CAIcal: A combined set of tools to assess codon usage adaptation. *Biol. Direct* **3**, 1–8 (2008).
10. B. Alix, D. A. Boubacar, M. Vladimir, T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40**, 573–579 (2012).
11. S. A. Slavoff, *et al.*, Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **9**, 59–64 (2013).
12. F. Impens, *et al.*, N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in *Listeria monocytogenes*. *Nat. Microbiol.* **2**, 17005 (2017).
13. D. M. Anderson, *et al.*, A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
14. M. R. Hemm, B. J. Paul, T. D. Schneider, G. Storz, K. E. Rudd, Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.* **70**, 1487–1501 (2008).
15. A.-R. Carvunis, *et al.*, Proto-genes and de novo gene birth. *Nature* **487**, 370–374 (2012).
16. A. Wacholder, *et al.*, A vast evolutionarily transient translatome contributes to phenotype and fitness. *Cell Syst.* **14**, 363-381.e8 (2023).
17. S. B. Parikh, C. Houghton, S. B. Van Oss, A. Wacholder, A. Carvunis, Origins, evolution, and physiological implications of de novo genes in yeast. *Yeast* **39**, 471–481 (2022).
18. J. Weaver, F. Mohammad, A. R. Buskirk, G. Storz, Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes. *mBio* **10**, 1–21 (2019).
19. S. Meydan, *et al.*, Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. *Mol. Cell* **74**, 481-493.e6 (2019).
20. L. Cassidy, *et al.*, Bottom-up and top-down proteomic approaches for the identification, characterization, and quantification of the low molecular weight proteome with focus on short open reading frame-encoded peptides. *Proteomics* **21**, e2100008 (2021).
21. F. Fontaine, R. T. Fuchs, G. Storz, Membrane localization of small proteins in *Escherichia coli*. *J. Biol. Chem.* **286**, 32464–32474 (2011).
22. L. Lafranchi, D. Schlesinger, K. J. Kimler, S. J. Elsässer, Universal Single-Residue Terminal Labels for Fluorescent Live Cell Imaging of Microproteins. *J. Am. Chem. Soc.* **142**, 20080–20087 (2020).
23. X. Cao, *et al.*, Nascent alt-protein chemoproteomics reveals a pre-60S assembly checkpoint inhibitor. *Nat. Chem. Biol.* **18**, 643–651 (2022).

24. S. Miravet-Verde, *et al.*, Unraveling the hidden universe of small proteins in bacterial genomes. *Mol. Syst. Biol.* **15**, 1–17 (2019).
25. M. R. Hemm, *et al.*, Small stress response proteins in *Escherichia coli*: Proteins missed by classical proteomic studies. *J. Bacteriol.* **192**, 46–58 (2010).
26. S. Zhang, *et al.*, Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nat. Commun.* **11**, 1312 (2020).
27. C. Q. E. Lee, *et al.*, Coding and non-coding roles of MOCCI (C15ORF48) coordinate to regulate host inflammation and immunity. *Nat. Commun.* **12**, 1–22 (2021).
28. C. Liang, *et al.*, Mitochondrial microproteins link metabolic cues to respiratory chain biogenesis. *Cell Rep.* **40**, 111204 (2022).
29. A. L. Rocha, *et al.*, An inner mitochondrial membrane microprotein from the SLC35A4 upstream ORF regulates cellular metabolism. *J. Mol. Biol.*, 168559 (2024).
30. M. Raina, *et al.*, Dual-function AzuCR RNA modulates carbon metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2117930119 (2022).
31. C. R. Lloyd, S. Park, J. Fei, C. K. Vanderpool, The small protein SgrT controls transport activity of the glucose-specific phosphotransferase system. *J. Bacteriol.* **199** (2017).
32. J. J. Aoyama, M. Raina, A. Zhong, G. Storz, Dual-function Spot 42 RNA encodes a 15-amino acid protein that regulates the CRP transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–10 (2022).
33. S. Lindquist, The heat-shock response. *Annu. Rev. Biochem.* **55**, 1151–91 (1986).
34. T. Yura, H. Nagai, H. Mori, Regulation of the Heat-Shock Response In Bacteria. *Annu. Rev. Microbiol.* **47**, 321–50 (1993).
35. F. Arsène, T. Tomoyasu, B. Bukau, The heat shock response of *Escherichia coli*. *Int. J. Food Microbiol.* **55**, 3–9 (2000).
36. A. D. Grossman, D. B. Straus, W. A. Walter, C. A. Gross, Sigma 32 synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli*. *Genes Dev.* **1**, 179–184 (1987).
37. G. Storz, An RNA thermometer. *Genes Dev.* **13**, 633–636 (1999).
38. P. Yuan, N. G. D'Lima, S. A. Slavoff, Comparative Membrane Proteomics Reveals a Nonannotated *E. coli* Heat Shock Protein. *Biochemistry* **57**, 56–60 (2018).
39. Y. Y. Chen, *et al.*, Conformational changes associated with cofactor/substrate binding of 6-phosphogluconate dehydrogenase from *Escherichia coli* and *Klebsiella pneumoniae*: Implications for enzyme mechanism. *J. Struct. Biol.* **169**, 25–35 (2010).
40. C. D'Souza, M. M. Nakano, P. Zuber, Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9397–401 (1994).
41. L. W. Hamoen, H. Eshuis, J. Jongbloed, G. Venema, D. van Sinderen, A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol. Microbiol.* **15**, 55–63 (1995).
42. S. Meydan, N. Vázquez-Laslop, A. S. Mankin, Genes within Genes in Bacterial Genomes. *Microbiol. Spectr.* **6** (2018).
43. N. Vakirlis, *et al.*, De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nat. Commun.* **11**, 781 (2020).
44. E. C. Hobbs, X. Yin, B. J. Paul, J. L. Astarita, G. Storz, Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16696–16701 (2012).
45. D. Du, *et al.*, Structure of the AcrAB–TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).
46. J. Abramson, *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

47. A. Chugunova, *et al.*, LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. *Proc. Natl. Acad. Sci.* **116**, 4940–4945 (2019).
48. S. Zhang, *et al.*, LINC00116-encoded microprotein mitoregulin regulates fatty acid metabolism at the mitochondrial outer membrane. *iScience* **26**, 107558 (2023).
49. X. Cao, *et al.*, Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nat. Commun.* **12** (2021).
50. Z. Na, *et al.*, Mapping subcellular localizations of unannotated microproteins and alternative proteins with MicroID. *Mol. Cell* **82**, 2900-2911.e7 (2022).
51. Y. Chen, *et al.*, Unannotated microprotein EMBOW regulates the interactome and chromatin and mitotic functions of WDR5. *Cell Rep.* **42**, 113145 (2023).
52. T. F. Martinez, *et al.*, Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).
53. C.-L. Sandmann, *et al.*, Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Mol. Cell* **83**, 994-1011.e18 (2023).
54. U. Weidner, *et al.*, The Gene Locus of the Proton-translocating NADH□: Ubiquinone Oxidoreductase in *Escherichia coli*. *J. Mol. Biol.* **233**, 109–122 (1993).
55. T. Friedrich, The NADH:ubiquinone oxidoreductase (complex I) from *Escherichia coli*. *Biochim. Biophys. Acta - Bioenerg.* **1364**, 134–146 (1998).
56. J. M. Berrisford, C. J. Thompson, L. A. Sazanov, Chemical and NADH-induced, ROS-dependent, cross-linking between sublimits of complex I from *Escherichia coli* and *Thermus thermophilus*. *Biochemistry* **47**, 10262–10270 (2008).
57. R. G. Efremov, L. A. Sazanov, Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414–421 (2011).
58. H. Erhardt, *et al.*, Organization of the *Escherichia coli* aerobic enzyme complexes of oxidative phosphorylation in dynamic domains within the cytoplasmic membrane. *Microbiologyopen* **3**, 316–326 (2014).
59. T. Friedrich, D. K. Dekovic, S. Burschel, Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (respiratory complex I). *Biochim. Biophys. Acta - Bioenerg.* **1857**, 214–223 (2016).
60. Z. Zhao, *et al.*, Frequent pauses in *Escherichia coli* flagella elongation revealed by single cell real-Time fluorescence imaging. *Nat. Commun.* **9**, 1–10 (2018).
61. J. Enninga, J. Mounier, P. Sansonetti, G. T. Van Nhieu, Secretion of type III effectors into host cells in real time. *Nat. Methods* **2**, 959–965 (2005).
62. L. A. Sazanov, A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
63. M. R. Leonardo, Y. Dailly, D. P. Clark, Role of NAD in regulating the *adhE* gene of *Escherichia coli*. *J. Bacteriol.* **178**, 6013–6018 (1996).
64. D. Esterházy, M. S. King, G. Yakovlev, J. Hirst, Production of Reactive Oxygen Species by Complex I (NADH:Ubiquinone Oxidoreductase) from *Escherichia coli* and Comparison to the Enzyme from Mitochondria. *Biochemistry* **47**, 3964–3971 (2008).
65. T. E. Schlub, E. C. Holmes, Properties and abundance of overlapping genes in viruses. *Virus Evol.* **6**, 1–8 (2020).
66. C. W. Nelson, *et al.*, Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* **9**, 1–45 (2020).
67. B. W. Wright, M. P. Molloy, P. R. Jaschke, Overlapping genes in natural and engineered genomes. *Nat. Rev. Genet.* **23**, 154–168 (2022).
68. R. Feltens, M. Gossringer, D. K. Willkomm, H. Urlaub, R. K. Hartmann, An unusual mechanism of bacterial gene expression revealed for the RNase P protein of *Thermus* strains. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5724–9 (2003).
69. I. Sela, Y. I. Wolf, E. V. Koonin, Theory of prokaryotic genome evolution. *Proc. Natl.*

Acad. Sci. U. S. A. **113**, 11399–11407 (2016).
70. M. Gagnon, *et al.*, Potentiation of B2 receptor signaling by AltB2R, a newly identified alternative protein encoded in the human bradykinin B2 receptor gene. *J. Biol. Chem.* **296**, 100329 (2021).
71. M. A. Brunet, *et al.*, The FUS gene is dual-coding with both proteins contributing to FUS -mediated toxicity. *EMBO Rep.* **22** (2021).
72. A. Pavesi, Origin, Evolution and Stability of Overlapping Genes in Viruses: A Systematic Review. *Genes (Basel).* **12**, 809 (2021).
73. K. Nelson, R. K. Selander, Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10227–10231 (1994).
74. M. Biseric̀, J. Y. Feutrier, P. R. Reeves, Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic gnd locus. *J. Bacteriol.* **173**, 3894–3900 (1991).
75. D. E. Dykhuizen, L. Green, Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* **173**, 7257–7268 (1991).
76. P. I. Tarr, *et al.*, Acquisition of the *rfb-gnd* Cluster in Evolution of *Escherichia coli* O55 and O157. *J. Bacteriol.* **182**, 6183–6191 (2000).
77. J. R. Elder, M. Bugarel, H. C. den Bakker, G. H. Loneragan, K. K. Nightingale, Interrogation of single nucleotide polymorphisms in *gnd* provides a novel method for molecular serogrouping of clinically important Shiga toxin producing *Escherichia coli* (STEC) targeted by regulation in the United States, including the “big six” non-O157 . *J. Microbiol. Methods* **129**, 85–93 (2016).
78. A. L. Cookson, *et al.*, Culture independent analysis using *gnd* as a target gene to assess *Escherichia coli* diversity and community structure. *Sci. Rep.* **7**, 1–11 (2017).
79. P. Kolata, R. G. Efremov, Structure of *Escherichia coli* respiratory complex I reconstituted into lipid nanodiscs reveals an uncoupled conformation. *Elife* **10**, 1–32 (2021).
80. C. E. VanOrsdel, *et al.*, The *Escherichia coli* CydX Protein Is a Member of the CydAB Cytochrome *bd* Oxidase Complex and Is Required for Cytochrome *bd* Oxidase Activity. *J. Bacteriol.* **195**, 3640–3650 (2013).
81. S. Safarian, *et al.*, Active site rearrangement and structural divergence in prokaryotic respiratory oxidases. *Science* **366**, 100–104 (2019).
82. M. Marcén, V. Ruiz, M. J. Serrano, S. Condón, P. Mañas, Oxidative stress in *E. coli* cells upon exposure to heat treatments. *Int. J. Food Microbiol.* **241**, 198–205 (2017).
83. R. Bromberg, S. M. George, M. W. Peck, Oxygen sensitivity of heated cells of *Escherichia coli* O157:H7. *J. Appl. Microbiol.* **85**, 231–237 (1998).
84. B. Bushnell, J. Rood, E. Singer, BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
85. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
86. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
87. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
88. F. Madeira, *et al.*, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
89. J. A. Devito, Recombineering with *tolC* as a Selectable/Counter-selectable Marker: Remodeling the rRNA Operons of *Escherichia coli*. *Nucleic Acids Res.* **36**, 1–12 (2008).
90. C. J. Gregg, *et al.*, Rational optimization of *tolC* as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* **42**, 4779–4790 (2014).

91. R. Lutz, Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
92. S. A. Schwartz, D. R. Helinski, Purification and Characterization of Colicin E1. *J. Biol. Chem.* **246**, 6318–6327 (1971).
93. A. Khitun, S. A. Slavoff, Proteomic Detection and Validation of Translated Small Open Reading Frames. *Curr. Protoc. Chem. Biol.* **11**, 1–27 (2019).
94. S. R. Adams, *et al.*, New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *J. Am. Chem. Soc.* **124**, 6063–6076 (2002).
95. C. Hoffmann, *et al.*, Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat. Protoc.* **5**, 1666–1677 (2010).
96. B. R. Martin, B. N. G. Giepmans, S. R. Adams, R. Y. Tsien, Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat. Biotechnol.* **23**, 1308–1314 (2005).
97. P.-C. Yang, T. Mahmood, Western blot: Technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4**, 429 (2012).
98. H. Schägger, Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
99. M. Ashburner, *et al.*, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
100. S. A. Aleksander, *et al.*, The Gene Ontology knowledgebase in 2023. *Genetics* **224**, 1–14 (2023).