Comparative membrane proteomics reveals a nonannotated *E. coli* heat shock protein

Peijia Yuan†‡, Nadia G. D'Lima†‡, Sarah A. Slavoff†‡⊥*

† Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

‡ Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States

⊥Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06529, United States

Supporting Information

Experimental Procedures

Strains and Constructs
E. coli K12 substr. MG1655 and pKD46 plasmids were obtained as a gift from Jason Crawford (Yale University). The method of Uzzau et al. was used to introduce an SPA epitope tag and kanamycin selection marker to the C-terminal end of the nonannotated protein[1]. Colonies were screened for recombinants on LB/kanamycin plates, and the presence of the SPA epitope tag was checked by PCR and verified by Sanger sequencing. Primers to introduce the genomic SPA tag and for integration check PCR are listed in Table S1.

For recombinant expression, the *gnd* gene both with and without the *gnd* start codon up to the C-terminus of the non-annotated peptide was amplified from an *E. coli* K12 substr. MG1655 colony, and then cloned into pET21a using restriction sites BamHI and HindIII (New England Biolabs). The C-terminal GFP tag in-frame with the nonannotated small open reading frame was introduced using restriction sites HindIII and NotI. A control construct encoding GFP alone was generated in parallel using restriction sites HindIII and NotI of pET21a.

Stress conditions for mass spectrometry
Stress conditions were adapted from Hemm *et al*[2]*.* An overnight culture of *E. coli* K12 substr. MG1655 was prepared. A flask containing 1 L of LB was inoculated with a 1:100 dilution of overnight culture and grown at 30 ^{o}C with shaking until an OD_{600} between 0.4 and 0.5 was reached. At this point, 500 mL of the culture were transferred to a flask containing a stir bar in a water bath set at 45 ^{o}C for 20 minutes (starting from the time at which the culture reached 45 ^{o}C), while the remaining 500 mL of culture continued to be grown at 30 ^{o}C. Afterward, both fractions were pelleted at 4,000*g* for 10 minutes at room temperature. The fractions were washed with 20 mL PBS each and spun again at 4,000*g* at room temperature for 10 minutes. The supernatant was then removed. Pellets were stored at -80 ^{o}C after being flash frozen.

Cell lysis and membrane extraction
Frozen pellets were resuspended in 20 mL PBS. The cell suspensions were sonicated at 45% amplitude for ten 15-second bursts with 20-second rest intervals on a Fisher Scientific Model 120 Sonic Dismembrator. The lysates were centrifuged at 20,000*g* for 15 minutes at 4 ^{o}C. The pellet was discarded, and the supernatant was centrifuged at 100,000*g* for 90 minutes at 4 ^{o}C. Following the second centrifugation step, the supernatant was discarded and the pellets were each resuspended in 200 µL of 0.1% SDS and incubated overnight at room temperature on a rotator.

Protein size selection
The entire volumes of samples from the membrane extraction were separated on a 15% SDS gel. The gel was stained with Coomassie, and the bands corresponding to sizes of 15 kDa and 10 kDa were cut from the gel for digestion. The cut gel pieces were destained, then washed 3x with 200 µL of a 50 mM 1:1 v/v solution of ammonium bicarbonate/acetonitrile on a rotator. The first wash was performed for 2 minutes and the second and third for 30 minutes each. Then, 500 µL acetonitrile was added, and the samples were vortexed for 10 seconds. The supernatant was then discarded, and the gel pieces were dried by Speedvac.

Digestion of samples for mass spectrometry
The gel pieces were covered with 600 µL of a solution consisting of 16 µL of 0.5 mg/mL trypsin (Promega) and 584 µL of 9:1 50 mM ammonium bicarbonate:acetonitrile. The digest solution was incubated for 1 hour on ice, then for 16 hours at 37 ^{o}C. The following day, the supernatant was transferred to an Eppendorf tube, and to the gel pieces was added 200 µL of 1:2 v/v 5%

formic acid/acetonitrile. The gel pieces were incubated in solution for 15 minutes on a rotator, then the supernatant was combined with the supernatant from the previous step. The solution was dried on Speedvac. Then 5.45 µL of 70% formic acid was added, followed by 14.55 µL of 0.1% trifluoroacetic acid, with vortexing between each step. The supernatant was removed and applied to a custom-made stage tip (2 × 1.06 mm punches of Empore C18 extraction disks [3M] in a 200 µl pipette tip) activated with methanol and 0.1% TFA. The tips were washed twice with 50 µL 0.1% TFA, and then eluted with 30 µL of 80% acetonitrile/0.1% TFA. Eluents were dried on Speedvac, then resuspended in 85% acetonitrile/0.1% formic acid and fractionated.

Offline fractionation of peptides
Prior to LC-MS/MS, samples were fractionated via electrostatic repulsion-hydrophilic interaction chromatography (ERLIC)[3]. Samples were loaded in 85% acetonitrile/0.1% formic acid onto a polyWAX LP column (150 x 1.0-mm; 5um 300 Å; PolyLC) attached to an Agilent 1100 HPLC at a flow rate of 0.05 mL/min. The samples were run with a 90-minute method as follows (Solvent A: 80% acetonitrile, 0.1% formic acid; Solvent B: 30% acetonitrile; 0.1% formic acid) with a flow rate of 0.3 mL/min: Isocratic flow was maintained at 100% A for 5 minutes, followed by a 17-minute linear gradient to 8% B, then a 25-minute linear gradient to 45% B. A 10-minute gradient to 100% B was then applied, followed by a 5-minute hold at 100% B. Finally, a 10-minute linear gradient back to 100% A was applied, and then held at 100% A until the end of the run. Fractions for the heat shock experiments were collected in two minute intervals for the first 10 minutes, then in 3 minute intervals for 21 minutes. Fractions were collected every 9 minutes for the next 27 minutes. This was followed by an 11-minute fraction and a 20-minute fraction. The fractions were dried on Speedvac and resuspended in 7 µL 3:8 70% formic acid:0.1% TFA.

LC-MS/MS analysis
LC-MS/MS methods were performed as in D'Lima et al., *J. Proteome Res.* **2017**, and were based on a previous report[4].

Data analysis
The programs ProteoWizard MS Convert[5] and Mascot Version 2.5.1 (Matrix Science, Inc., London, UK)[6] were used for peak picking and file analysis, respectively. Modifications set were Carbamidomethyl (C) (fixed) and carbamyl (K and N-term), oxidation (M), and phospho (STY) (variable). The peptide mass error tolerance used was 20 ppm, the maximum number of missed cleavages was 3, and the parameters were set to a semi-tryptic digest. A six-frame translation database of the *E. coli* K12 MG1655 genome (accession number NC_000913.3 in NCBI) as well as the common contaminant database were used. Annotated peptide sequences were excluded with a string-matching algorithm[7] using the E. coli K12 MG1655 proteome as a reference. Only unique peptides of greater than 8 amino acids, with a MASCOT score of at least 50, at least five consecutive cleavages, and both y- and b-ion fragmentation were considered. Identified peptides were checked for annotation in the RefSeq database for MG1655. Putative non-annotated hits were searched in NCBI Blast, and those that only contained one amino acid mismatch relative to annotated proteins were removed from consideration.

Protein expression
To test non-annotated protein upregulation during heat shock, 5 mL of LB were inoculated with 100 µL of a genomically SPA-tagged culture grown overnight to saturation at 37 ^{o}C. Wild-type *E. coli* K12 MG1655 was used as a control. The cultures were grown at 30 ^{o}C with shaking until OD_{600} 0.4-0.5 was reached. A culture of each strain was transferred to a water bath at 30 ^{o}C (control) or 45 ^{o}C (heat shock) for 20 minutes from the time the heat shock culture reached 45 ^{o}C. In order to compare relative protein expression, an aliquot from each tube corresponding to 0.2 OD_{600} units was taken, trichloroacetic acid (TCA) was added immediately to a final concentration

of 8 % and centrifuged at 21,000 x g for 15 minutes at 4 ºC. Pellets were washed with acetone, dried on Speedvac, and resuspended in 30 µL SDS gel loading buffer on a rotator overnight. 10 µL of each sample was loaded on a 16% tricine gel.

To test for translation of GndA independent of translation from the first ATG start codon of *gnd*, pet21A expression constructs encoding GFP alone, GndA fused to GFP with the upstream *gnd* sequence intact, and GndA fused to GFP with the upstream *gnd* sequence present but lacking the first ATG codon were transformed into BL21 cells. Single colonies were inoculated into 10 mL of LB with ampicillin (100 µg/mL) and grown to log phase at 37 ^{o}C with shaking. Then, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to a final concentration of 1 mM, and the culture was allowed to continue growing at 37 ^{o}C for 3.5 hours. To 1 mL of each culture (not normalized) was added 10% TCA, followed by incubation on ice for 20 min, and centrifuged at 21,000 x g for 15 minutes at 4 ºC. Pellets were twice washed with acetone and again centrifuged at 21,000 x g for 15 minutes at 4 ºC, followed by air drying. Each sample was resuspended in 30 µL 1X SDS gel loading buffer and electrophoresed on a 4-20% Tris/glycine gel.

To test for enrichment of the non-annotated protein in the membrane, a single colony of BL21 cells with the non-annotated peptide C-terminally tagged with GFP in pET21a was inoculated into 15 mL of LB with ampicillin (100 µg/mL) and grown overnight at 37 ^{o}C. The next day, 500 mL of LB/ampicillin were inoculated with 10 mL of the overnight culture and then grown to log phase at 37 ^{o}C with shaking. Then, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to a final concentration of 1 mM, and the culture was allowed to continue growing at 37 ^{o}C for 3.5 hours. At this point, the culture was pelleted at 4,000*g* for 10 minutes at 4 ^{o}C. The fractions were washed with 20 mL PBS each and spun again at 4,000*g* at 4 ^{o}C for 10 minutes. The supernatant was then removed. Pellets were stored at -80 ^{o}C after being flash frozen. They were then subjected to cell lysis and membrane extraction as above. Aliquots of the cellular lysate and the supernatant from the ultracentrifugation step were flash frozen. The membrane pellet from the ultracentrifugation step was resuspended in 1 mL 0.1% SDS and rotated overnight at room temperature. The concentration of each fraction was determined by BCA assay and 20 ug of each were loaded onto a gel. All gels were run in duplicate, where one was stained with Coomassie and the other was subjected to Western blotting.

Western Blotting

SDS and tricine gels were transferred to BioTrace nitrocellulose membranes (VWR) at 100 V for 1 hour. Membranes were blocked in 3% BSA for 1 hour at room temperature with shaking. To probe for SPA-tagged proteins, membranes were incubated with mouse monoclonal anti-FLAG M2 (Sigma) primary antibody diluted 1:1000 for 1 hour at room temperature, then washed with Tris buffered saline containing 0.1 % Tween 20 (TBS-T). The membranes were then incubated with horseradish peroxidase (HRP) anti-mouse secondary antibody (Rockland) diluted 1:10,000 for 1 hour at room temperature, then washed with TBS-T. Clarity ECL Western Blotting Substrate (Bio-Rad) was used to develop the blots, and imaging was performed using a ChemiDoc imaging system (Bio-Rad) and Image Lab software (BioRad).

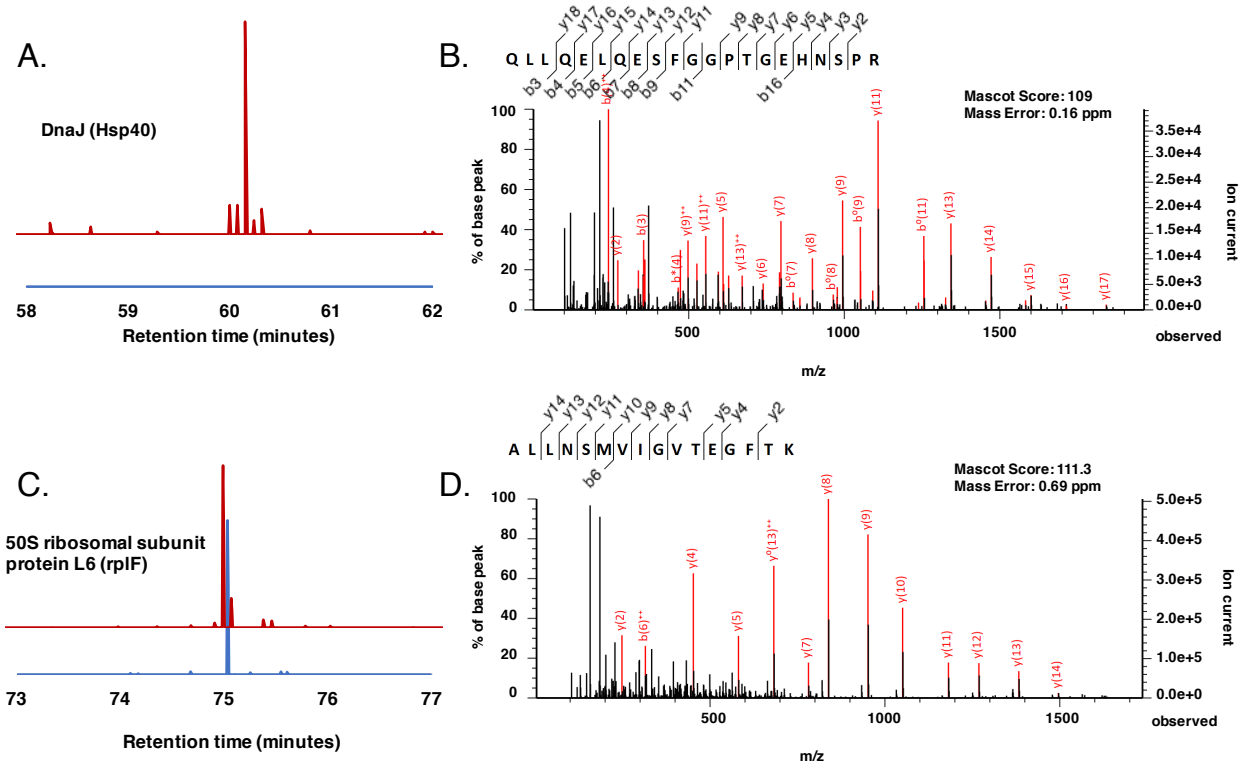

Figure S1. Quantitative detection of known heat shock and control proteins by LC-MS/MS. A. Extracted ion chromatograms (EICs) comparing peak areas corresponding to the DnaJ heat shock peptide ion detected in B. between heat shock (red trace) and control (blue trace) conditions at the same retention time. The same y-axis scale is used in both conditions. A viewing window of 1 Da around the parent ion mass is used. B. MS2 spectrum corresponding to the DnaJ peptide fragment is shown. Y-ions and b-ions are shown in red. C. EICs are shown for the corresponding peptide ion detected in D. belonging to the constitutively expressed ribosomal protein L6.

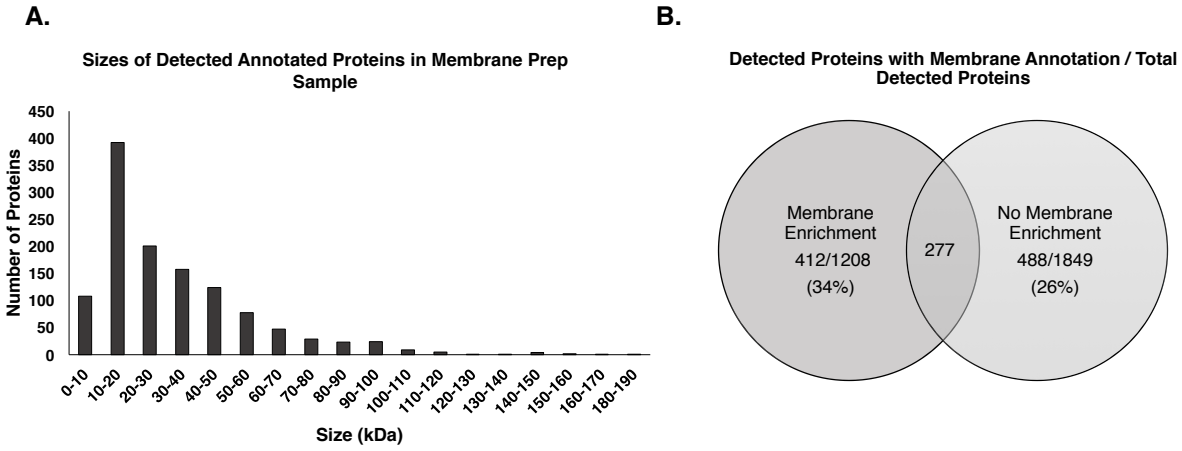

Figure S2. Membrane annotations and sizes of annotated proteins detected by LC-MS/MS. A. Bar chart showing the size distribution in kDa of detected annotated proteins in one replicate of proteogenomics with membrane enrichment step. B. Venn diagram showing the percentage of detected annotated proteins with a membrane annotation in EcoCyc for one replicate of control sample (growth at permissive temperature) proteogenomics with membrane enrichment step (412/1208; 34%) and without membrane enrichment step (488/1849; 26%). 277 proteins are overlapping between both sets.

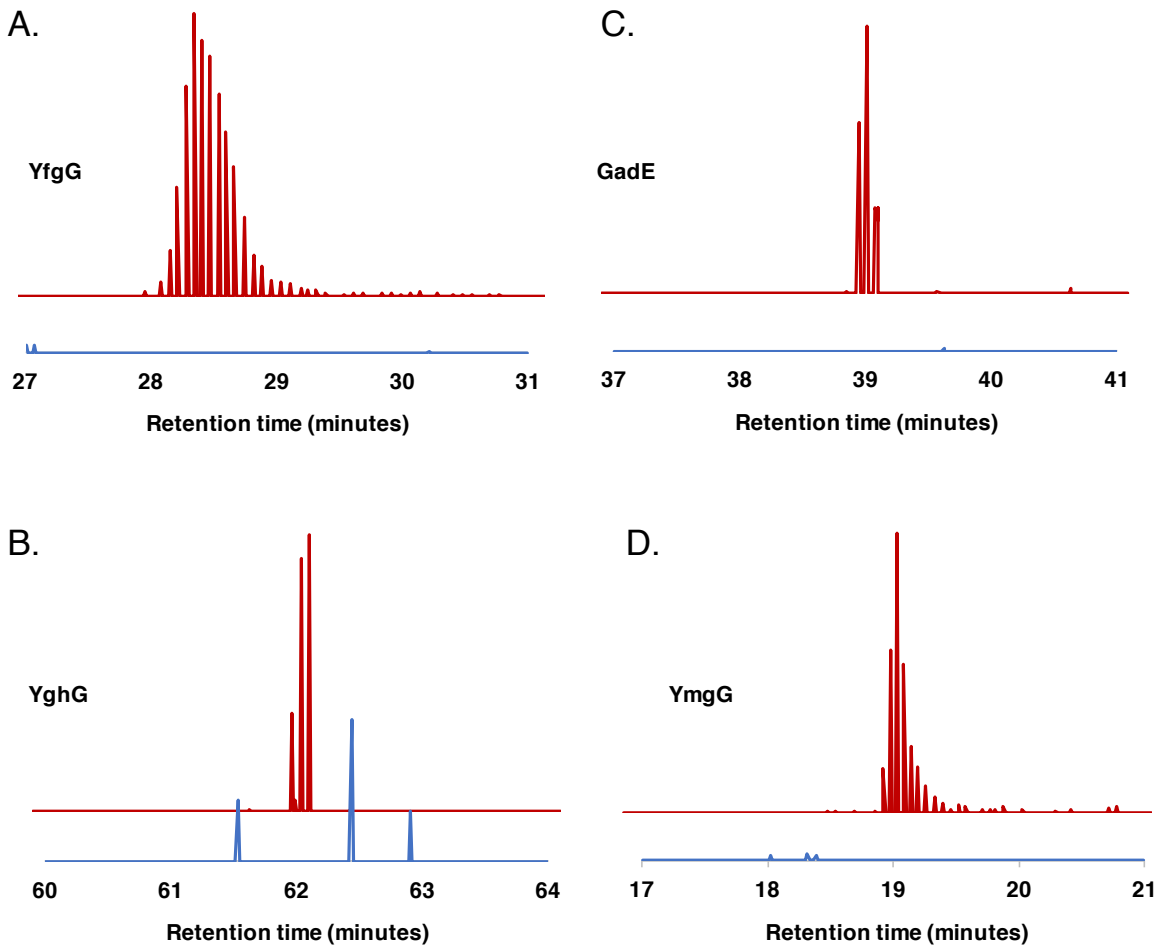

Figure S3. Quantitation of proteins without heat shock annotation by LC-MS/MS: extracted ion chromatograms. For four proteins without current heat shock annotations that were identified via proteomics search only in the heat shock sample and not the control, A. YfgG, B. YghG, C. GadE, and D. YmgG, extracted ion chromatograms (EICs) are presented comparing peak areas corresponding to a diagnostic tryptic peptide ion in heat shock (red trace) and control (blue trace) conditions at the same retention time. The same y-axis scale is used in both conditions. A viewing window of 1 Da around the parent ion mass is used.

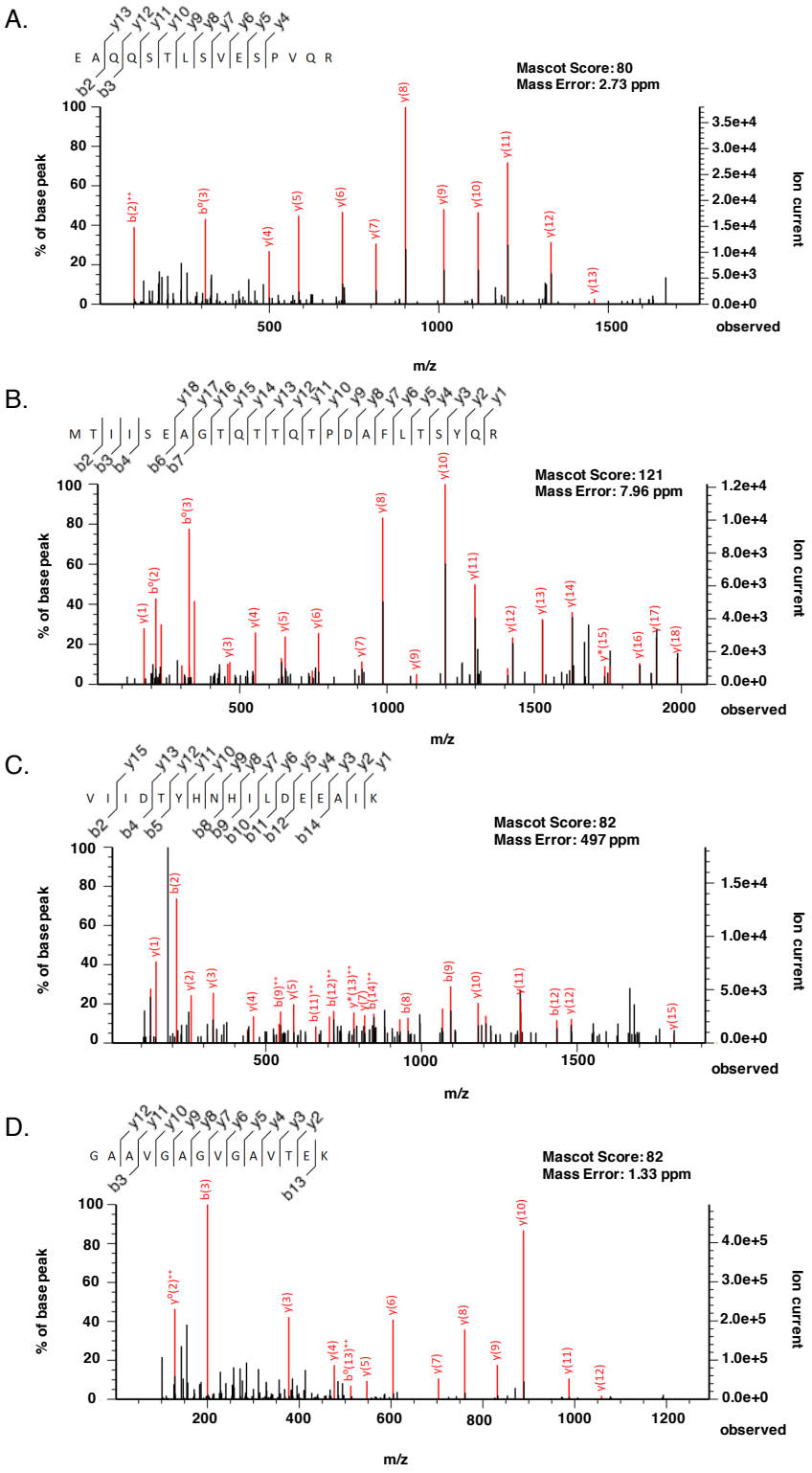

Figure S4. Quantitation of proteins without heat shock annotation by LC-MS/MS: MS2 spectra. The MS2 spectra used for EIC quantitation in Fig. S3 are presented, corresponding to A. YfgG, B. YghG, C. GadE, and D. YmgG. Y-ions and b-ions are shown in red and indicated on the identified peptide sequence.

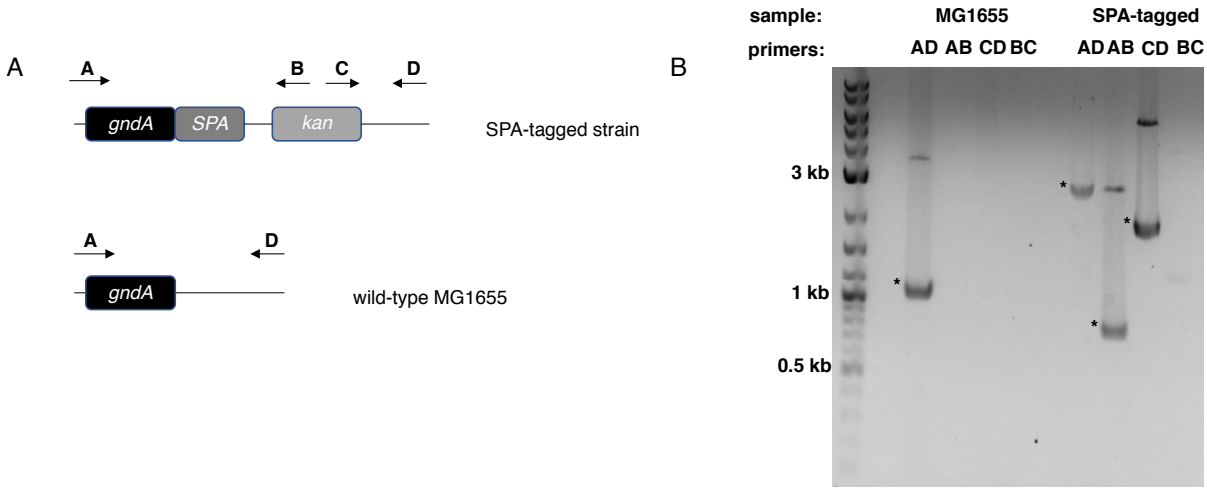

Figure S5. Validation of knock-in strain via integration check PCR (icPCR). (A) Gene locus diagrams and icPCR primer annealing sites. Primers A and D were designed to anneal to the 5' end of the coding sequence and the downstream genomic DNA sequence. Primers B and C specific to the knock-in construct (kanamycin selection marker) were also designed. (B) Agarose gel of icPCR results shows specific products at the expected sizes denoted by asterisks in the SPA-tagged strains. Correct integration was confirmed by Sanger sequencing of PCR products.

Integration Check Primers	
A	ATGGTGGTAACACCTTCTTC
B	GCAATCCATCTTGTTCAAT
C	ATGACTGGGCACAACAGA
D	TTAATCCAGCCATTCGGTAT
SPA-tag Knock-In Primers	
Forward	GGGCTTTAACTTCATCGGTACCGGTGTTTCTGGCGGTGAAGAGGGGGCGCTCCATGGAAAAGAGAAGATGGAA
Reverse	TACCAATTCATAGGCTTCTTTCTGGCCACCAGGCATAATAGAAGGACCTTCATATGAATATCCTCCTTAGTTCCTATTCC

Table S1. Primers for genomic knock-in strain construction and icPCR validation.

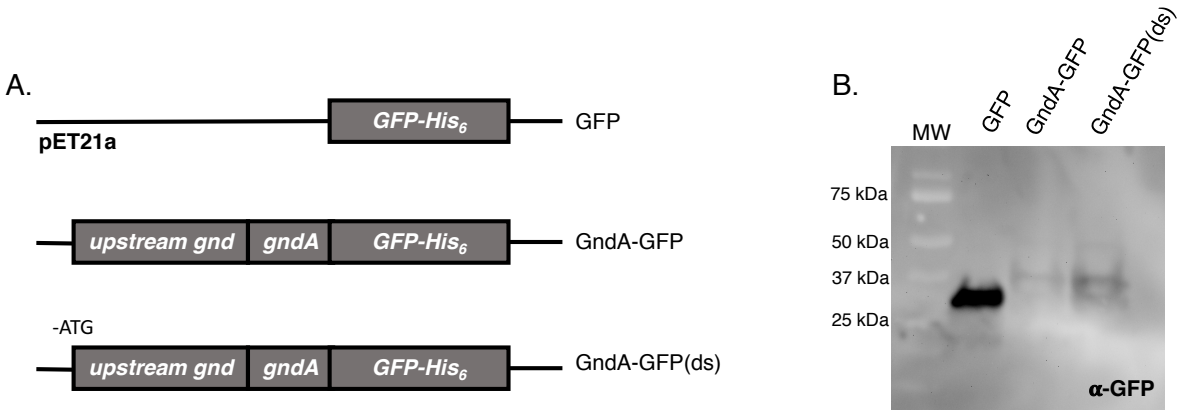

Figure S6. Expression of a GndA-GFP fusion protein. A. Constructs used to test expression of *gndA* within the upstream *gnd* sequence context as a GFP fusion protein. A pET21a expression vector was generated encoding hexahistidine-tagged (His_6) GFP alone as a control (GFP, top). The *gnd* coding sequence from the annotated start codon of *gnd* to the putative 3' end of *gndA* was cloned into pET21a to place *gndA* in frame with the GFP-His_6 reporter (GndA-GFP, middle). This permits expression of GndA-GFP without perturbing translation initiation, as the *gndA* start codon has not yet been conclusively established. The ATG start codon of *gnd* was subsequently deleted from the expression construct (GndA-GFP(ds), ds: "delete start", bottom). B. The three expression constructs shown in (A) were expressed in BL21 cells, and cell lysates were separated on a gradient SDS-PAGE gel and subjected to anti-GFP immunoblotting.

References:

[1] Uzzau, S., Figueroa-Bossi, N., Rubino, S., and Bossi, L. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 15264-15269.

[2] Hemm, M. R., Paul, B. J., Miranda-Rios, J., Zhang, A., Soltanzad, N., and Storz, G. (2010) *J. Bacteriol. 192*, 46-58.

[3] Hao, P., Ren, Y., Dutta, B., and Sze, S. K. (2013) *J. Proteomics 82*, 254-262.

[4] Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., and Isaacs, F. J. (2013) *Science 342*, 357-360.

[5] Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M. Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S. L., Nuwaysir, L. M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E. W., Moritz, R. L., Katz, J. E., Agus, D. B., MacCoss, M., Tabb, D. L., and Mallick, P. (2012) *Nat. Biotechnol. 30*, 918-920.

[6] Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) *Electrophoresis 20*, 3551-3567.

[7] Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., Karger, A. D., Budnik, B. A., Rinn, J. L., and Saghatelian, A. (2013) *Nat. Chem. Biol. 9*, 59-+.