**Supporting Information**

# *In vitro* Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free Protein Synthesis

Anthony W. Goering[▽†‡], Jian Li[§‡], Ryan A. McClure[†▽], Regan J. Thomson[▽], Michael C. Jewett[§*] and Neil L. Kelleher[▽†*]

[†]Department of Molecular Biosciences, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States.
[▽]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States.
[§]Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States.

**Author Information**
**Corresponding Authors**
*N.L.K. Email: n-kelleher@northwestern.edu
*M.C.J. Email: m-jewett@northwestern.edu

[‡]A.W.Goering and J. Li are co-first authors.

## Supplemental Methods

### Quantitation of synthesized GrsA and GrsB1

The yields of synthesized GrsA and GrsB1 were quantified by the incorporation of $^{14}$C-leucine into trichloroacetic acid (TCA)-measuring precipitated radioactivity using a liquid scintillation counter (MicroBeta2, PerkinElmer, Waltham, MA). After CFPS, 100 µL of 0.1 M NaOH was added to each 15 µL reaction mixture. The NaOH treated samples were incubated at 37 °C for 20 min. Then, 50 µL of the treated sample was loaded onto each of two separate pieces of Whatman 3MM chromatography paper and dried under a heat lamp for 1 h. One piece of filter paper in each pair was placed into a beaker on ice and covered with 5% (v/v) TCA at 4 °C to precipitate the proteins onto the filter paper. After 15 min, the solution was exchanged with fresh TCA. This incubation was performed a total of three times. Following the third precipitation, the filter papers were covered once with 100% ethanol for 10 min at room temperature, and then they were removed from the beaker and allowed to dry under a heat lamp for another 1 h. Radioactivity of both the TCA-precipitated (washed) and non-TCA-precipitated (unwashed) samples was measured using a liquid scintillation counter. The fraction of incorporated $^{14}$C-leucine (washed/unwashed) was used to determine the amount of protein synthesized.

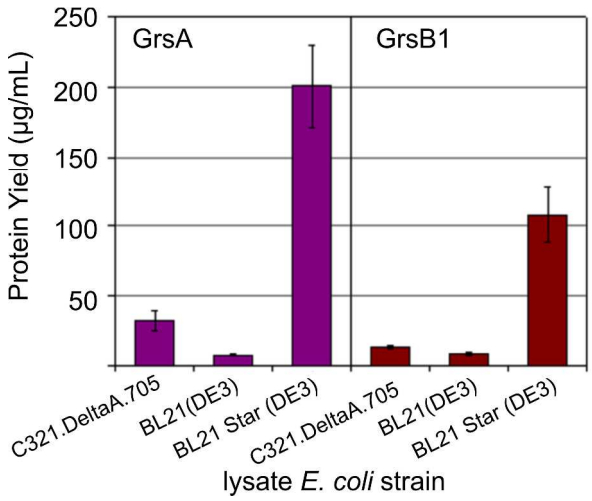



**Figure S1.** Production of NRPS proteins GrsA and GrsB1 with lysates prepared from different strains of *E. coli*. Y-axis shows the protein yields determined by incorporation of $^{14}$C-leucine and scintillation counting.

### Expression and purification of recombinant Sfp from *E. coli*

pET28-Sfp[1] was expressed and purified from BL21(DE3). Sfp was purified from the lysate of 250 mL culture over cobalt resin (Thermo Scientific), and washed in steps of 10, 25, 50 mM imidazole. Pure Sfp protein was eluted with 150 mM imidazole. Following desalting and buffer exchange using Zeba desalting spin columns (Thermo Scientific), the final concentration of Sfp was determined to be 2 mg/mL by the bicinchoninic acid assay. Sfp was stored in 10% glycerol at -80 °C prior to use.

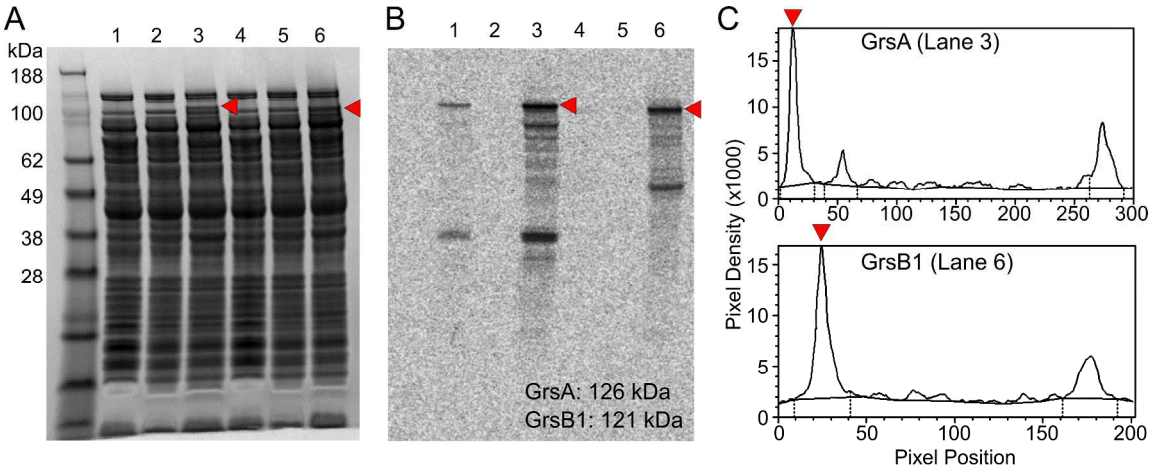



**Figure S2.** Autoradiography image showing the production of full-length GrsA and GrsB1 after incorporation of $^{14}$C-leucine. Panel A shows the coomassie stained gel with the expected GrsA and GrsB1 bands marked with a red triangle. Panel B shows the phosphor image of the same gel with GrsA and GrsB1 once again marked with the red triangle. The numbers indicate the following protein/*E. coli* lysate pairs: 1) GrsA, C321.DeltaA.705; 2) GrsA, BL21(DE3); 3) GrsA, BL21(DE3)star; 4) GrsB1, C321.DeltaA.705; 5) GrsB1, BL21(DE3); 6) GrsB1, BL21(DE3) star. Panel C charts the pixel density from which the relative percentage of full-length, soluble protein was calculated. Panel C was derived from the raw image data of Panel B using ImageQuantTL software (GE Healthcare Life Sciences). The relative contribution of the full-length band was determined by comparing the integration of the peak with the total area under the curve. An adjusted baseline, and dotted lines marking the integration area of each peak appear just above the x-axis. From the ratio of these integrations, and the total area under the curve, we estimated full-length GrsA and GrsB1 represent 53% and 71% of the radioisotope decay signal from the soluble protein fraction, respectively.

**Autoradiography analysis**

After the cell-free reaction, 3 µL of each sample was loaded on a 4-12% NuPAGE SDS-PAGE gel (Invitrogen). After running at 120 V for 1.5 h, the gel was stained using SimplyBlue$^{TM}$ SafeStain solution (Invitrogen) and destained in water. Then, the gel was fixed with cellophane films (Bio-Rad), dried overnight in a GelAir Dryer (Bio-Rad) without heating, and exposed for 48 h on a Storage Phosphor Screen (GE Healthcare Biosciences, Pittsburgh, PA). The autoradiogram was scanned using a Storm Imager (GE Healthcare Biosciences, Pittsburgh, PA) and analyzed using Quantity One software (Bio-Rad, Hercules, CA).

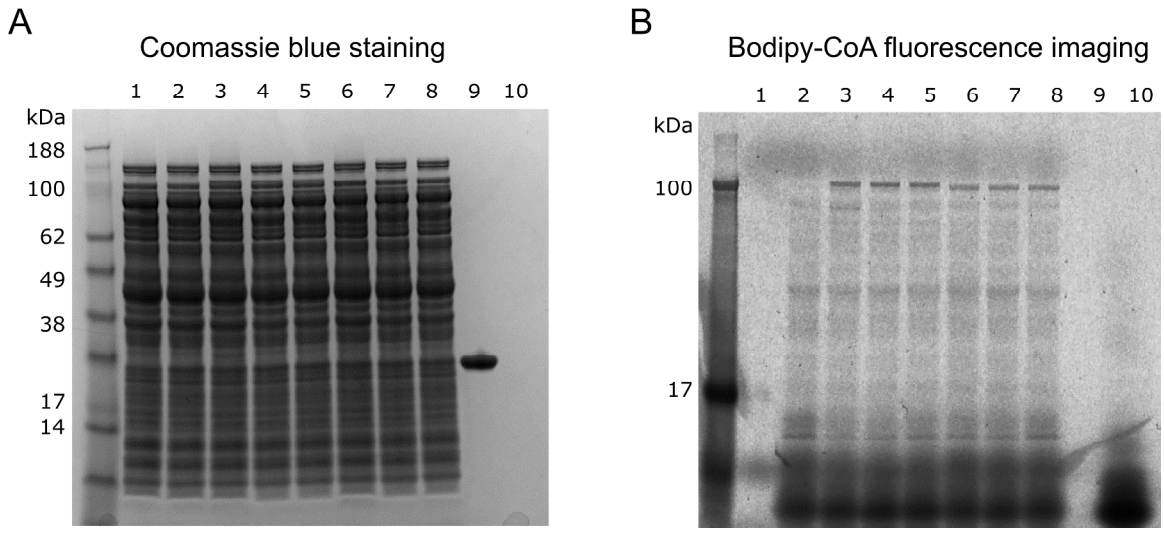
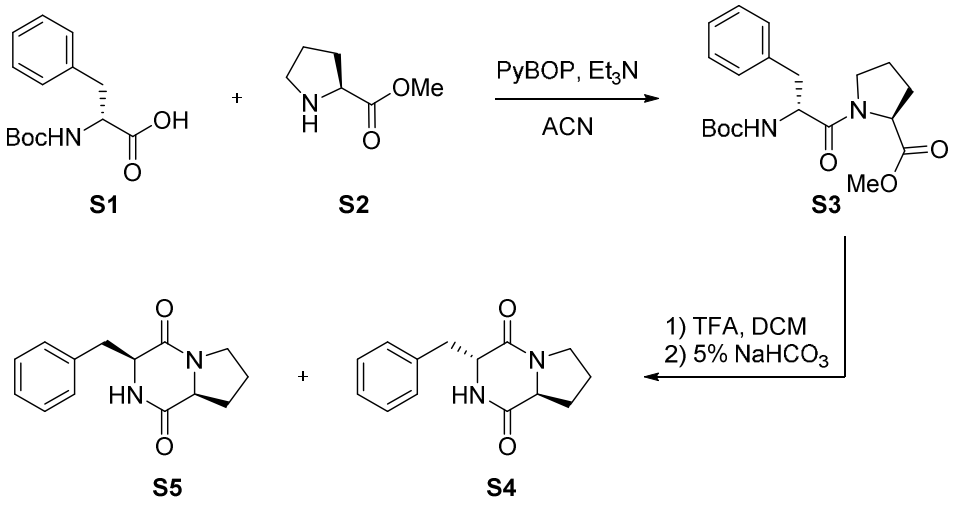



**Figure S3.** Labeling of NRPS proteins GrsA and GrsB1 with conjugated Bodipy-CoA. Panel A shows the coommassie stained gel, and panel B shows the fluorescence image acquired at 520 nm. The numbers above the lanes on both images indicate the following conditions: **1)** no plasmid, no Bodipy-CoA, no Sfp; **2)** no plasmid, 1 µL Bodipy-CoA, 1 µL Sfp; **3)** GrsA, after 17 h added 1 µL Bodipy-CoA , 1 µL Sfp, at 30 °C for another 3 h; **4)** GrsA, after 17 h, centrifuge at 12000 g, 10 min, 4 °C , take out 13 µL supernatant, then add 1 µL Bodipy-CoA, 1 µL Sfp, at 30 °C for another 3 h; **5)** GrsA, after 17 h, centrifuge at 12000 g, 10 min, 4 °C , remove 13 µL supernatant, then add 1 µL Bodipy-CoA, 1 µL Sfp, at 37 °C for another 3 h; **6-8)**. GrsB1, with the same respective conditions as **3-5**); **9)** Sfp; **10)** Bodipy-CoA

**Synthetic preparation of DKP**




**S1** **S2** **S3** **S5** **S4**

**Scheme S1.** Preparation of synthetic DKP

Boc-D-Phe-OH (0.1 mmol) and L-proline methyl ester (0.1 mmol) were dissolved in 5 mL of acetonitrile. PyBOP (0.12 mmol) was added to the reaction mixture and allowed to stir at room temperature. Triethylamine (0.2 mmol) was added to the flask. After 12 h, saturated NaCl solution (10 mL) was added to the mixture and extracted with ethyl acetate (3 x 10 mL). The organic layer was washed with 1 M HCl, 5% $NaHCO_3$, and $H_2O$ sequentially. The organic layer was dried over sodium sulfate and solvent removed under reduced pressure to yield 75 mg of crude material, which was used without further purification. The Boc-protected dipeptide (**S3**) was dissolved in 5 mL of dichloromethane and 5 mL of trifluoroacetic acid and then stirred at room temperature for 1 hour. The solvent was removed under reduced pressure to yield a colorless oil, which was treated with 10 mL of 5% $NaHCO_3$ solution. After stirring for 12 hours, the mixture was extracted with dichloromethane (3 x 10 mL). The solvent was removed under reduced pressure—resulting in 48 mg of crude product. The product was determined to be a mixture of the diastereomers, **S4** and **S5**, and provided a match to previously reported spectra[2]. The diastereomers were separated by HPLC on a 250 mm x 10 mm reversed-phase (C18) column (Phenomonex). 0.13 mg (measured on an ATI Cahn CA-126 microbalance) of D-Phe-L-Pro DKP was purified.

**Extraction and identification of DKP**

After DKP production, each 15 µL reaction mixture was immediately extracted with 3 volumes of *n*-butanol/chloroform (4:1, v/v). The organic layer was separated by centrifugation at 16,000 x g and 4 °C for 10 min and then transferred to a fresh microcentrifuge tube. After removal of the solvent under vacuum, samples were resuspended in 100 µL of 50% methanol for further analysis by LC-MS.

DKP samples were analyzed on a Q-exactive orbitrap instrument (Thermo) paired with an Agilent 1150 HPLC system. MS parameters were the following: 70 minute run-time, scan range 100-1000 *m/z*, resolution 35,000. After each full scan, the top five most intense ions selected for HCD fragmentation at 25% collisional energy. LC buffers were $H_2O$ + 0.1% formic acid and acetonitrile + 0.1% formic acid. For each sample 20 µL were injected into the LC system for a 70 minute gradient with a flow rate of 200 µL/min

across a 150 mm x 2.1 mm reversed-phase (C18) column (Phenomonex). SIC intensities were determined using Xcalibur software (Thermo Scientific) using the mass range of 245.127-245.129 *m/z*.

**Quantitation of DKP production**

A standard curve was prepared using synthetic DKP (preparation described above) by performing injections of 1 µg, 500 ng, 250 ng, 125 ng, 50 ng, and 25 ng on an Agilent 1150 HPLC stack coupled to a Q-Exactive mass spectrometer. Signal intensity was determined from the integrated abundance value of DKP over the course of a 70 minute experiment. A linear equation was fit to the standard curve and used to estimate the abundance of DKP in the sample prepared from CFPS by calculating the total amount of DKP present in a 45 µL butanol/chloroform extraction.

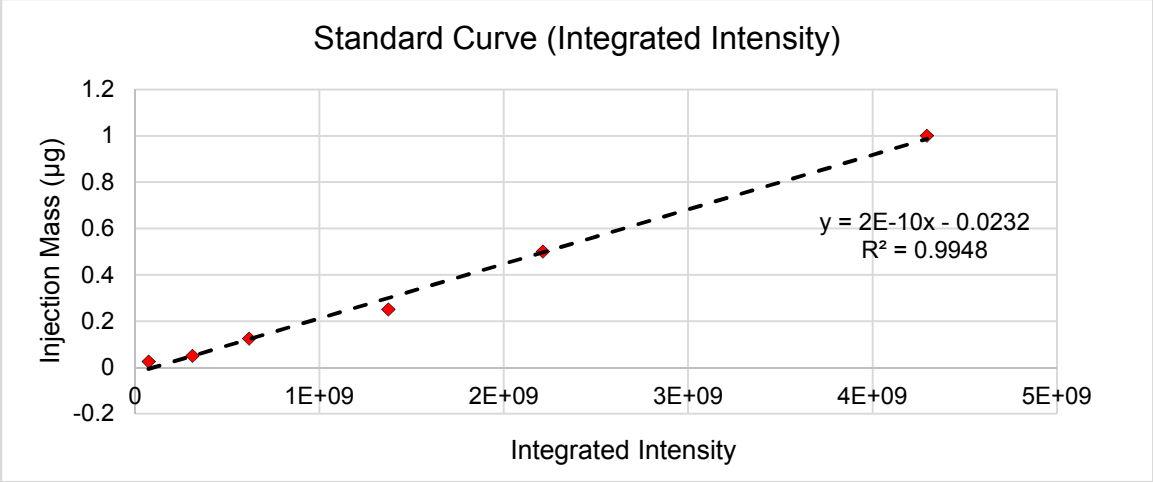



**Figure S4.** Standard curve for LC-MS-based quantitation of D-L-DKP.

## Bacterial strains

*Escherichia coli* stains C321.ΔA.705, BL21(DE3) and BL21 Star (DE3) were used for S30 cell extract preparation. C321.ΔA.705 was generated from a genetically encoded release factor 1 (RF1) deficient *E. coli* strain[3]. Both BL21(DE3) and BL21 Star (DE3) were purchased from Life Technologies (Grand Island, NY). *E. coli* DH5α was used for cloning and plasmid maintenance. BL21(DE3) was also used for Sfp expression and purification.

## Plasmid construction

Plasmid construction was performed using a conventional restriction enzyme-ligation scheme, using the pET28a plasmid backbone (EMD Millipore). Primers were obtained from Integrated DNA technologies. GrsA and GrsB1 were first cloned from the genomic DNA of *Brevibacillus brevis*. PCR was performed using GoTaq Green Master Mix (Promega, 60μL reaction volume). PCR products and pET28 plasmid were digested with the enzymes NcoI and BamHI (New England Biolabs) and gel extracted using the QIAquick Gel Extraction kit (Qiagen). Digested PCR products and plasmid backbones were ligated to form whole constructs using T4 DNA ligase. The plasmids utilized the T7 promoter and terminator, and the ribosome binding site “aagagg”.

**Table S1.** Primers used for amplification of NRPS encoding genes *grsA* and *grsB1*. The CDS-binding portion of each primer is highlighted in red.

| Name | Sequence (5' -> 3') |
|---|---|
| NcoI-grsA-F | attattatccatgg**tgttaaacagttctaaaagtatattgattcatgc** |
| BamHI-grsA-R | aataatatggatcc**gttaatgaatcggccaacaaatc** |
| NcoI-grsB1-F | attattatccatgg**tgagtacatttaaaaaagaacatgttcagg** |
| BamHI-grsB1-R | aataatatggatcc**atataattagagatttcctgaatggttggtg** |

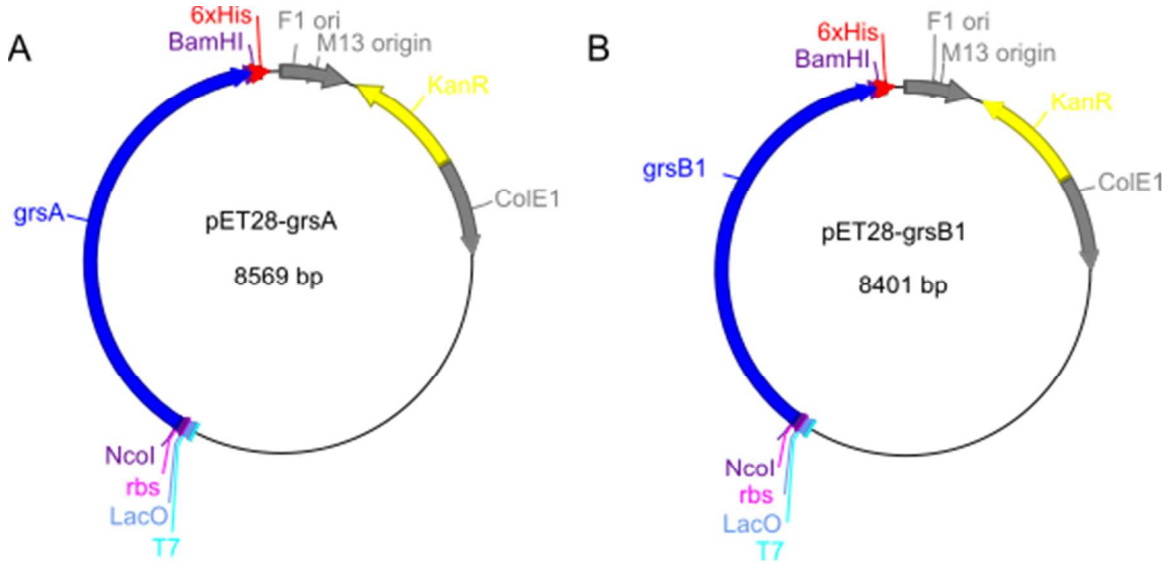



**Figure S5.** Plasmid maps of *grsA*(**A**) and *grsB1*(**B**) expression vectors

**Supplemental References**


1. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases, *Chem Biol 3*, 923-936.
2. Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., and Hilvert, D. (2014) Reprogramming nonribosomal peptide synthetases for "clickable" amino acids, *Angew Chem Int Ed Engl 53*, 10105-10108.
3. 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) Genomically recoded organisms expand biological functions, *Science 342*, 357-360.