nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-019-0894-z

In the format provided by the authors and unedited.

An *N*-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin

Tai L. Ng[1], Roman Rohac[2], Andrew J. Mitchell[3], Amie K. Boal[2,3]* & Emily P. Balskus[1]*

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [2]Department of Chemistry, The Pennsylvania State University, University Park, PA, USA. [3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA. *e-mail: akb20@psu.edu; balskus@chemistry.harvard.edu

Supplementary Information

An *N*-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin

Tai L. Ng[1], Roman Rohac[2], Andrew J. Mitchell[3], Amie K. Boal[2,3*], and Emily P. Balskus[1*]

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA
[2]Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania USA
[3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania USA

* To whom correspondence should be addressed:

Amie K. Boal, 326 Chemistry Building, University Park, Pennsylvania 16802, United States; akb20@psu.edu; phone: (814) 867-2863

Emily P. Balskus, 12 Oxford Street, Cambridge, Massachusetts 02138, United States; balskus@chemistry.harvard.edu; phone: (617) 496-9921; fax: (617-496-5866).

1. Supplementary Methods

SznF purification for crystallographic characterization

His_6-SznF purification. A pET29a plasmid encoding *Sa* SznF with an *N*-terminal hexahistidine affinity tag was incorporated into *E. coli* BL21(DE3) cells via heat-shock transformation. Successful transformants were selected overnight on kanamycin-supplemented (50 μg/mL) LB agar plates at 37 ºC and cultured in 1 L volumes of either rich LB medium (20 g/L yeast extract, 35 g/L tryptone, 5 g/L sodium chloride, pH 7.4) or M9 minimal medium[45]. The M9 cultures were used to generate selenomethionine (SeMet) SznF for initial phase determination. In these growths, a solution of amino acids [L-phenylalanine, L-lysine, L-threonine (100 mg/L) and L-valine, L-isoleucine, L-leucine, L–SeMet (50 mg/L)] was added to the cultures at the time of induction to ensure feedback inhibition of native methionine biosynthesis[46]. After growth at 37 ºC to an OD_{600} of 0.6 – 0.8, the cultures were cooled on ice for 15 min prior to induction of protein overexpression by addition of IPTG (1 mM final concentration) and continued growth at 18 ºC with shaking at 180 rpm overnight (16 – 20 h). Cells were harvested by centrifugation at 8000*g* and stored at – 80 ˚C. The average yield was 10 g of wet cell paste per 1 L of culture.

All protein purification procedures were performed at 4 ºC. Cell pellets were solubilized (3 mL/g) in buffer A (100 mM HEPES, 20 mM imidazole, 5 mM DTT, pH 7.5) supplemented with 1 mg/mL lysozyme and 1 mM PMSF. The suspended cells were disrupted via sonication (QSonica 750 W, 20 kHz, 2 s pulse, 6 s off, 50% amplitude, 6 min total sonication time) on ice, and the resulting cell debris was pelleted by centrifugation at 22,000*g*. The supernatant was applied to Ni-NTA resin and the bound proteins were eluted with buffer B (100 mM HEPES,

200 mM imidazole, 5 mM DTT, pH 7.5). Fractions containing SznF were pooled, concentrated, and dialyzed overnight into 100 mM HEPES, 50 mM NaCl, 5% (*v*/*v*) glycerol, 5 mM DTT, 5 mM EDTA, pH 7.5. The protein was further purified via gel filtration chromatography using an ÅKTA Pure FPLC outfitted with a Superdex 200pg 16/200 column (GE Healthcare Life Sciences) in a buffer containing 20 mM HEPES pH 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol. SznF elutes as a dimer and fractions containing the protein were pooled and concentrated using a 30K MWCO filtration device (PALL Corporation) prior to flash-freezing in liquid nitrogen and storage at –80 ºC. Protein concentrations were determined by UV-visible absorbance using a calculated extinction coefficient of 70,500 M^{-1} cm^{-1} and a predicted molecular weight of 55.6 kDa[47].

Tagless SznF purification. The coding sequence of *Sa* SznF lacking an affinity tag was amplified from the pET-29a/SznF plasmid template using the primers pET29b-sznF-F and pET29b-sznF-R (sequences in **Supplementary Table 4**) and Q5 DNA polymerase (NEB). The resulting amplicon was treated with *NdeI* and *EcoRI* restriction enzymes (NEB). The digested PCR product was purified by gel electrophoresis with a commercial extraction kit (Omega BioTek) and ligated into a linearized pET21b vector. The ligation products were used to transform DH5α *E. coli* competent cells (Invitrogen) and successful transformants were selected on LB agar plates supplemented with ampicillin (100 μg/mL). The DNA sequence of tagless SznF/pET21b was verified by Sanger sequencing (Penn State Nucleic Acid Facility).

SznF/pET21b was used to transform *E. coli* BL21(DE3) cells via heat shock. Successful transformants were selected overnight on ampicillin supplemented (100 μg/mL) LB agar plates at 37 ºC. Protein expression was induced as described above for His-tagged SznF. Cell pellets (6

g wet weight per 1 L of culture) were solubilized (5 mL/g) in 100 mM HEPES pH 7.5 supplemented with 1 mg/mL lysozyme and 1 mM PMSF. The suspended cells were disrupted via sonication as described above and the resulting cell debris was pelleted by centrifugation at 50,000*g* at 4 ºC for 30 min. The lysate was subjected to two rounds of ammonium sulfate precipitation (39% and 55% saturation) followed by centrifugation at 50,000*g* at 4 ºC for 10 min. In the first step, SznF remains in the supernatant and the pellet is discarded. In the second step, SznF is found in the pellet and subsequently re-suspended in buffer C (50 mM Tris-HCl, 20 mM NaCl, pH 7.5). The protein solution was dialyzed (nitrocellulose dialysis tubing, MWCO 6-8 kDa) at 4 ºC against two changes of buffer C (sample-to-buffer ratio of 1:20). The dialyzed protein was applied to a HiPrep Q Sepharose Fast Flow 16/10 (GE Healthcare Life Sciences) anion exchange chromatography column (20 mL) equilibrated in buffer C. Bound proteins were eluted over 8 column volumes via a linear gradient of 0.02 0.5 M NaCl in buffer C. SznF eluted between 270 and 360 mM NaCl. The protein fractions were pooled and dialyzed overnight into 50 mM HEPES, 300 mM NaCl, 5% (*v*/*v*) glycerol, 5 mM EDTA, pH 8.0. The protein was additionally purified via gel filtration chromatography (Superdex 200pg 16/200) with isocratic elution in 50 mM HEPES, 300 mM NaCl, 5% (*v*/*v*) glycerol, pH 8. The tagless protein also migrates as a dimer. Fractions containing SznF were pooled and dialyzed into 50 mM HEPES, 300 mM NaCl, 5% (*v*/*v*) glycerol, pH 7.5 and concentrated using a 30K MWCO filtration device (PALL Corp) prior to flash-freezing in liquid nitrogen and storage at –80 ºC. Protein concentrations were determined by UV-visible absorbance using a calculated extinction coefficient[3] of 70,500 $M^{-1}cm^{-1}$.

An additional preparation of tagless SznF was prepared from a pET29a/SznFG co-expression construct in which the SznG component contains a C-terminal His_6-tag. The proteins were

purified as a complex by Ni-NTA affinity chromatography and subsequently separated by size exclusion chromatography using the procedures outlined above. All adventitiously-bound metal ions were stripped with 10 mM EDTA in storage buffer (20 mM HEPES, 50 mM NaCl, 10% glycerol, pH 8.0, and the active sites were reconstituted with 12 equivalents (equiv.) $Fe(NH_4)_2(SO_4)_2$ in 1 mM DTT. Crystallization and x-ray data collection of this sample yielded full reconstitution of the C-terminal cupin Fe-binding site and partial occupancy of the central 7-helix bundle domain, determined by inspection of Fe anomalous difference electron density maps (**Extended Data Fig. 7**). Importantly, a partially-occupied heavy atom can also be observed in the central domain in the original SeMet structure used for phasing. Both samples share in common inclusion of a thiol-based reductant, DTT, during iron reconstitution.

SznF crystallization and x-ray structure determination

All crystallization trials were carried out in an anoxic chamber (COY Laboratory Products). Aliquots of SznF were allowed to equilibrate in the chamber for ~30 minutes prior to setup. Protein samples were diluted to 5 – 15 mg/mL in 20 mM HEPES, pH 7.5, and pre-mixed with 2 molar equivalents of $Fe(NH_4)_2(SO_4)_2$ and 5 molar equivalents of NMA. The **1**-bound SznF structure was obtained by co-crystallizing tagless SznF with 2 molar equivalents of $Fe(NH_4)_2(SO_4)_2$ and 10 molar equivalents of **1** in place of L-NMA. The resulting mixture was combined in a 1:1 ratio with the well solution (500 μL) in hanging-drop vapor diffusion trays. SznF crystallized in two conditions: 1) 10 – 15 % (*v*/*v*) isopropanol, 0.1-0.15 M tri-sodium citrate pH 5.6, and 15 – 20 % (*w*/*v*) PEG 4000; 2) 0.1 M Tris pH 8.5, 0.1 M $MgCl_2$, and 20 – 30 % (*w*/*v*) PEG 4000. Crystals typically appeared within 4 days and continued to grow in size over 3 – 4 weeks. The initial SeMet-SznF crystals from condition 1 selected for x-ray diffraction analysis

were soaked in a cryoprotectant solution consisting of well solution supplemented with 20 – 25 % (*v*/*v*) glycerol for 1 – 2 minutes prior to harvest and cryocooling by direct plunge into liquid nitrogen. The **1**-bound protein crystals from condition 2 were looped and flash-frozen without addition of cryoprotectant.

X-ray diffraction datasets were collected at 100K at the Advanced Photon Source at Argonne National Lab at beamlines 21-ID-F/G of the Life Sciences Collaborative Access Team (LS-CAT) or 23-ID-D/B of the General Medical Sciences/Cancer Collaborative Access Team (GM/CA-CAT). All diffraction images were processed with XDS[48,49] or the HKL2000 software package[50]. The initial phases were obtained from a SeMet-SznF crystal (condition 1) via single anomalous diffraction methods using datasets collected at the Se absorption peak (0.979 Å). ShelXC/D/E[51] and HKL2MAP[52] were used to identify the positions of six Se sites (out of seven predicted) per SznF protomer. An initial experimental map (FOM = 0.621) was obtained to 2.1 Å resolution and used as the input for automated model building via ARP/wARP[53] with manual modifications performed in COOT[54]. This initial model was used to phase native datasets by molecular replacement in PHASER MR[55]. Additional model building and refinement were carried out with Coot and Refmac5[56], the latter as part of the CCP4 software package[57]. TLS refinement was performed in Refmac5 with groups selected by the TLSMD server[58]. Geometric restraints for the **1** were generated with ProDRG[59]. All structural figures were generated with the PyMOL Molecular Graphics software package (Schrödinger, LLC).

Se-Met modified SznF was crystallized in condition 1 and contains four molecules in the asymmetric unit. Chain A contains residues 3 –151, 156 – 218, 220 – 314, and 319 – 471. Chain B contains residues 3 – 134, 136 – 151, 157 – 215, 223 – 312 and 319 – 471. Chain C contains residues 3 – 134, 136 – 152, 155 – 312, and 319 – 471. Chain D contains residues 1 – 151, 156 –

215, 221 – 313, and 319 – 471. Additionally, eight additional amino acids were modeled at the N-terminus of chain D that correspond to the His_6-affinity tag. The structural model also contains 1202 water molecules, four molecules of glycerol, five sulfate ions, two molecules of isopropanol and eight metal ions. In many of our datasets, His_6-tagged SznF exhibited *K*-edge x-ray absorption features consistent with significant incorporation of both Ni^{II} and Fe^{II} in the C-terminal domain. The Ni^{II} fraction is most likely an artifact of the purification procedure. In protein samples subjected to additional EDTA chelation and iron reconstitution, as described above, anomalous difference Fourier maps calculated from datasets collected above the Fe *K*-edge peak x-ray absorption wavelength (1.722 Å) consistently reveal a single anomalous scatterer at ~18 I/σ in the C-terminal domain. In certain datasets, a second anomalous scatterer can be found at weaker intensity (11 I/σ) in the central 7-helix bundle domain. The four metal ions found in the C-terminal domain of SznF were modeled as Fe ions at 100% occupancy, while the remaining four metal ions in the central domain were modeled as Fe ions at partial occupancy (ranging from 40 – 50%).

Intermediate **1** was cocrystallized anaerobically with untagged $Fe^{(II)}$-loaded SznF in condition 2. X-ray fluorescence and anomalous diffraction data collection on a second crystal from the same protein preparation revealed *K*-edge features for iron with very little contribution from other transition metals. An anomalous difference Fourier map calculated from datasets collected above the Fe *K*-edge peak absorption wavelength (1.74 Å) reveals a single anomalous scatterer at 8.8 I/σ in the C-terminal domain. The analysis does not reveal any peaks in the central domain. Molecule A contains residues 4 – 133, 137 – 150, 157 – 215, and 220 – 471. Chain B contains residues 3 –133, 137 – 151, 157 – 215 and 223 – 471. This structure additionally contains 980 water molecules, two metal ions (modeled as Fe), and two molecules

Formatted: Superscript

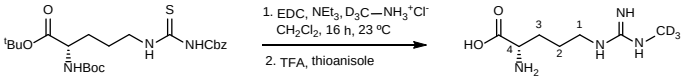
of **1**. A summary of the data collection and refinement statistics can be found in **Supplementary Table 7.**

Preparation of Compounds

All chemicals were purchased from Sigma Aldrich except unless otherwise noted. Solvents were obtained from Sigma Aldrich. Anhydrous reactions were performed using oven-dried or flame-dried glassware, which were then cooled under vacuum and purged with nitrogen (N_2) gas. 3 Å molecular sieves were oven-dried overnight and then cooled under high vacuum prior to use. Unless otherwise noted, all proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on Agilent DD2-600 (600 MHz, 150 MHz), Varian Inova-500 (500 MHz, 125 MHz), Varian Mercury 400 (400 MHz, 100 MHz), or JOEL 400 (400 MHz, 100 MHz) NMR spectrometers. Chemical shifts (δ) are reported in parts per million (ppm) downfield from tetramethylsilane using the solvent resonance as an internal standard for ^{1}H ($CDCl_3$ = 7.26 ppm, D_2O = 4.79 ppm, C_2D_6OS = 2.50 ppm, d_7-DMF = 8.03 ppm) and ^{13}C ($CDCl_3$ = 77.2 ppm, C_2D_6OS = 39.5 ppm, d_7-DMF = 163.2 ppm). Data are reported as follows: chemical shift, multiplicity (br = broad, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constants in Hertz (Hz), integration, and assignments. Some proton and carbon data were assigned using two-dimensional NMR spectroscopy techniques [^{1}H–^{1}H COSY (Correlation Spectroscopy); ^{1}H–^{13}C HSQC (Heteronuclear Single Quantum Coherence), and ^{1}H–^{13}C HMBC (Heteronuclear Multiple Bond Connectivity). All NMR solvents were purchased from Cambridge Isotope Laboratories unless

otherwise noted. NMR spectra were visualized using MestReNova, version 10.0.0-14411. Semi-preparative and preparative high-performance liquid chromatography (HPLC) was performed on a Dionex Ultimate 3000 instrument (Thermo Scientific). Some of the high-resolution mass spectral data for the synthetic compounds were obtained in the Magnetic Resonance Laboratory in the Chemistry and Chemical Biology department on a Bruker MicroQTOF-QII fitted with an electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV in the positive-ion mode, the drying gas temperature was maintained at 190 °C with a flow rate of 8 L/min, and a nebulizer pressure of 21.8 psi. Some spectra were obtained using an Agilent 6530 quadruple time-of-flight (qTOF) mass spectrometer with an electrospray ionization (ESI) source. The mass spectra data were recorded on positive or negative ionization mode with a mass range of 100 to 1700 m/z; spectra rate, 1 spectra/s; capillary voltage, 3500 V; nebulizer pressure, 35 psi; drying gas (N_2) flow, 8 L/min; temperature, 275 ºC.

Synthesis of N^{ω}-d_3-methyl-L-arginine (d_3-L-NMA)

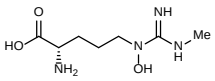


6 **d_3-L-NMA**

The carboxybenzyl (Cbz)-protected thiourea **6** was prepared according to Jansen Labby *et al.*[60] In a round bottom flask containing **6** (200 mg, 415 μmol, 1.00 equiv.) dissolved in 20 mL of anhydrous dichloromethane (CH_2Cl_2) at 0 ºC was added 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 88.8 mg, 457 μmol, 1.10 equiv.), triethylamine (NEt_3, 116 μL, 830 μmol, 2.00 equiv.), and d_3-methylamine hydrochloride ($D_3C–NH_3^+Cl^-$, Cambridge Isotope Inc.; 30.0 mg, 415 μmol, 1.00 equiv.) in one portion. After 16 h of stirring at room

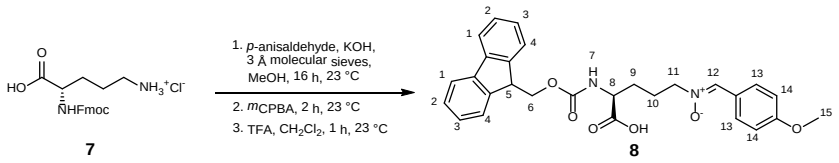
temperature, the reaction was washed with 1 x 10 mL of brine, and the collected organic layer was dried over anhydrous sodium sulfate (Na_2SO_4), filtered, and concentrated *in vacuo* to afford an oil. The crude oil was dissolved in 5 mL of trifluoroacetic acid (TFA) and 0.80 mL of thioanisole and stirred at room temperature for 16 h. The reaction was then concentrated *in vacuo* and 20 mL of toluene was added. The solution was concentrated *in vacuo,* and the residue was dissolved in 5 mL of water. The aqueous solution was then washed with 2 x 5 mL of diethyl ether to remove nonpolar impurities. The crude d_3-L-NMA in the aqueous layer was purified by reverse-phase semi-preparative HPLC. The LC column was Kromasil 100 C18 (250 x 10 μm, 5mm). The flow rate was 3 mL/min. The LC conditions were: a linear gradient from 5 to 10% solvent B in 6 min; 10 to 30% solvent B in 1 min; an isocratic hold at 30% solvent B for 1 min; a linear gradient from 30 B to 5% solvent B in 2 min; and an isocratic hold at 5% solvent B for 2 min. (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). d_3-L-NMA was monitored at 200 nm and eluted between 4 – 4.5 min. The collected fractions were lyophilized to afford 34 mg of d_3-L-NMA as a white solid (45% yield). ^{1}H NMR (600 MHz D_2O): δ (ppm) = 4.07 – 4.01 (m, 1H, H_4), 3.28 – 3.23 (m, 2H, H_1), 2.05 – 1.92 (m, 2H, H_2), 1.79 – 1.67 (m, 2H, H_3). HRMS: Calc'd for $C_7H_{14}D_3N_4O_2$ $[M+H]^+$ = 192.1534, found = 192.1526.

Synthesis of N^{δ}-hydroxy-N^{ω}-methyl-L-arginine (1)



1

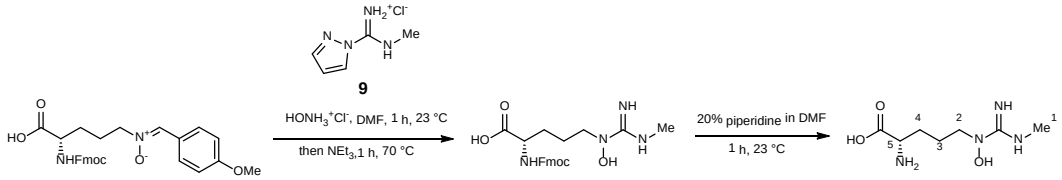
Step 1:



The nitrone **8** was prepared by adapting the procedure from Yun-Ming Lin *et al.*[61] An oven-dried flask was first added 3 Å activated molecular sieves followed by 12 mL of anhydrous methanol. *N*-alpha-(9-Fluorenylmethyloxycarbonyl)-L-ornithine hydrochloride (Fmoc-L-ornithine, **7**, Chem-Impex International, Inc.; 433 mg, 1.11 mmol, 1.00 equiv), *p*-anisaldehyde (149 μL, 1.22 mmol, 1.10 equiv.), and potassium hydroxide (KOH, 68.4 mg, 1.22 mmol, 1.10 equiv.) were added in one portion and stirred overnight at room temperature under N_2. Then the reaction was cooled to 0 ºC, and *m*-chloroperoxybenzoic acid (*m*CPBA, 191.5 mg, 1.11 mmol, 1.00 equiv.) was added to the stirring reaction mixture in two portions over 5 min. The reaction mixture was stirred at 0 ºC for 1 h before another equiv. of *m*CPBA was added to drive the reaction to completion. After 1 h of stirring at 0 ºC, the reaction was filtered to remove the molecular sieves and precipitated salts, and the sieves were thoroughly washed with methanol. The combined filtrate was concentrated to dryness *in vacuo*, and the residue was dissolved in 10 mL of CH_2Cl_2 and 1 mL of TFA. The reaction was allowed to stir at room temperature for 1 h. The reaction was then concentrated *in vacuo* to afford crude nitrone **8**, which was purified with a 10 g C18 Sep-Pak cartridge (Waters). The cartridge was first washed with 100 mL of water, then 100 mL of 30% acetonitrile in water to remove **7**, and the nitrone **8** was eluted with 200 mL of 50% acetonitrile in water. The desired fraction was concentrated *in vacuo* to remove the acetonitrile, affording an aqueous suspension containing **8**. The aqueous solution was then extracted with 2 x 50 mL of ethyl acetate. The combined organic layer was dried over Na_2SO_4, filtered, and evaporated to afford **8** as a light brown solid (160 mg, 30% yield). ^{1}H NMR (600

MHz $CDCl_3$): δ (ppm) = 8.18 (d, *J* = 8.5 Hz, 2H, H_1), 7.73 (d, *J* = 7.5 Hz, 2H, H_{13}), 7.58 (t, *J* = 8.4 Hz, 2H, H_4), 7.47 (s, 1H, H_{12}), 7.40 – 7.34 (m, 2H, H_2), 7.30 – 7.26 (m, 2H, H_3), 6.90 (d, *J* = 8.4 Hz, 2H, H_{14}), 5.94 (d, *J* = 6.9 Hz, 1H, NH_7), 4.41 – 4.31 (m, 3H, H_5 and H_6), 4.17 (t, *J* = 7.3 Hz, 1H, H_8), 4.13 – 3.93 (m, 2H, H_{11}), 3.81 (s, 3H, H_{15}), 2.20 – 2.01 (m, 2H, H_9), 1.98 – 1.88 (m, 2H, H_{10}). ^{13}C NMR (125 MHz $CDCl_3$): δ (ppm) = 173.4 (**C**), 162.4 (**C**), 155.9 (**C**), 143.8 (2 x **C**), 141.3 (**C**H), 139.3 (2 x **C**), 132.4 (2 x **C**H), 127.7 (2 x **C**H), 127.1 (2 x **C**H), 125.1 (2 x **C**H), 121.7 (**C**), 120.0 (2 x **C**H), 114.1 (2 x **C**H), 67.0 (**C**H_2), 65.2 (**C**H_2), 55.5 (**C**H_3), 53.2 (**C**H), 47.1 (**C**H), 29.6 (**C**H_2), 23.1 (**C**H_2). HRMS: Calc'd for $C_{28}H_{28}N_2O_6Na$ $[M+Na]^+$ = 511.1845, found = 511.1818.

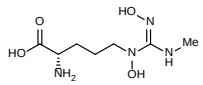
<u>Step 2 – 3:</u>



To nitrone **8** (100 mg, 0.205 mmol, 1.00 equiv.) in 20 mL of dimethylformamide (DMF) was added hydroxylamine hydrochloride ($HONH_3^+Cl^-$, 17.0 mg, 0.246 mmol, 1.20 equiv.). The reaction was stirred at room temperature for 1 h under N_2. Then, NEt_3 (57.0 μL, 0.410 mmol, 2.00 equiv.) and 1*H*-pyrazole-1-(*N*-methylcarboxamidine) hydrochloride (**9**, Alfa Aesar) (32.8 mg, 0.205 mmol, 1.00 equiv.) were added to the reaction in one portion, heated to 70 ºC with an oil bath, and stirred for 1 h under N_2. The reaction was then allowed to cool to room temperature, and 4 mL of piperidine were added to the reaction. After an additional 1 h of stirring at room temperature, the reaction was concentrated *in vacuo*, and the residue was dissolved in 2 mL of water. The aqueous solution was washed with 2 x 2 mL of ethyl acetate to remove nonpolar

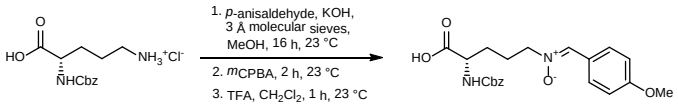
impurities, and the aqueous layer was concentrated *in vacuo* to remove residual organic solvents. Crude **1** in the aqueous solution was purified by preparative HPLC. The LC column was Thermo Hypersil Gold aQ C18 (250 × 20 mm, 5 μm, Thermo Fisher Scientific) The flow rate was 8 mL/min. The LC conditions were: a linear gradient from 5 to 95% solvent B in 15 min; an isocratic hold at 95% solvent B for 5 min; a linear gradient from 95 to 5% solvent B in 2 min; and an isocratic hold at 5% solvent B for 3 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). **1** was monitored at 200 nm and eluted between 7.2 – 8.2 min. The collected fractions were lyophilized to afford 12 mg of **1** as a white solid (28% yield). ^{1}H NMR (500 MHz D_2O): δ (ppm) = 4.06 (t, *J* = 6.2 Hz, 1H, H_5), 3.63 (t, *J* = 6.7 Hz, 2H, H_2), 2.91* (s, 2.3H, H_1), 2.89* (s, 0.6H, H_1), 2.09 – 1.78 (m, 4H, H_3 and H_4). ^{13}C NMR (125 MHz D_2O): δ (ppm) = 172.3 (**C**), 158.7 (**C**), 52.9 (**C**H), 51.5 (**C**H_2), 27.7 (**C**H_2), 27.1* (**C**H_3), 26.8* (**C**H_3), 21.5 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_3$ $[M-H]^-$ = 203.1150, found = 203.1153. *rotamers

Synthesis of N^{δ}-hydroxy-N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (2)

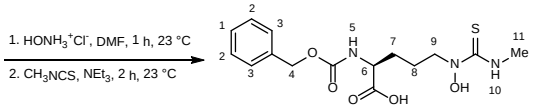


2

Steps 1 – 2:



11 **12**

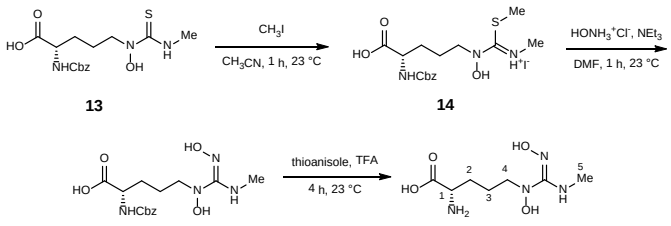


13

Cbz-protected nitrone **12** was prepared analogously to **8** with L-Cbz-ornithine hydrochloride (**11**, Chem-Impex International) as the starting material. To nitrone **12** (847 mg, 2.12 mmol, 1.00 equiv.) dissolved in 20 mL of DMF was added hydroxylamine hydrochloride (161 mg, 2.33 mmol, 1.10 equiv.) in one portion and stirred for 1 h under N_2. Then *N*-methyl isothiocyanate (CH_3NCS, 170 mg, 2.33 mmol, 1.10 equiv.) dissolved in 5 mL of anhydrous CH_2Cl_2 was added to the reaction in one portion, and NEt_3 (325 μL, 2.33 mmol, 1.10 equiv.) was added to the reaction mixture dropwise. After stirring at room temperature for 2 h, the reaction was quenched with 10 mL of 1 M HCl and extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with 10 mL of brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford crude thiourea **13** as a yellow oil. The compound was first purified with the CombiFlash Rf system to remove major impurities. The column was a 150 g C18 (RediSep Rf, Teledyne Isco). The flow rate was 75 mL/min. The LC conditions were: an isocratic hold at 5% solvent B for 2 min; a linear gradient from 5 to 100% solvent B in 11 min; an isocratic hold at 100% solvent B for 3 min; and an isocratic hold at 50% solvent B for 3 min (solvent A = 0.1% TFA in water, solvent B = 0.1% TFA in acetonitrile). **13** was monitored at 254 nm. The collected fractions containing **13** were combined and extracted with ethyl acetate (3 x 20 mL). The combined organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the an impure thiourea **13** (310 mg, 41% yield) for subsequent steps.

To obtain an analytically pure sample for spectral characterization, a few miligrams of **13** were further purified using preparative HPLC. The LC column was Thermo Hypersil Gold aQ C18 (250 × 20 mm, 5 μm, Thermo Fisher Scientific) The flow rate was 8 mL/min. The LC conditions were: a linear gradient from 50 to 95% solvent B in 15 min; an isocratic hold at 95% solvent B for 5 min; a linear gradient from 95 to 50% solvent B in 4 min; and an isocratic hold at 50% solvent B for 4 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). **13** was monitored at 264 nm and eluted between 10.9 – 11.3 min. The collected fractions were lyophilized to afford **13** as a white powder. ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) = 7.41 – 7.30 (m, 5H, $H_{1\text{-}3}$), 7.16 (br s, 1H, NH_{10}), 5.85 (d, *J* = 8.2 Hz, 1H, NH_5), 5.16 – 5.04 (m, 2H, H_4), 4.75 – 4.61 (m, 1H, H_9), 4.41 (t, *J* = 9.3 Hz, 1H, H_6), 3.88 – 3.79 (m, 1H, H_9), 3.10* (s, 2.5H, H_{11}), 3.01* (s, 0.5H, H_{11}), 2.24 –1.63 (m, 4H, H_7 and H_8). ^{13}C NMR (100 MHz $CDCl_3$): δ (ppm) = 181.0 (**C**), 174.1 (**C**), 157.4 (**C**), 135.6 (**C**), 128.7 (2 x **C**H), 128.6 (**C**H), 128.1 (2 x **C**H), 67.8 (**C**H_2), 54.2 (**C**H), 53.6 (**C**H), 32.2 (**C**H_2), 31.7 (**C**H_3), 22.1 (**C**H_2). HRMS: Calc'd for $C_{15}H_{21}N_3O_5SNa$ $[M+Na]^+$ = 378.1100, found = 378.1092. *rotamers

Step 3 – 5:

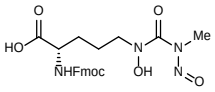


Thiourea **13** (297 mg, 836 μmol, 1.00 equiv.) was dissolved in 10 mL of acetonitrile (CH_3CN), and methyl iodide (CH_3I, 57.3 μL, 917 μmol, 1.10 equiv.) was added dropwise to the stirring solution. The reaction was stirred for 1 h under N_2 at room temperature. The reaction was

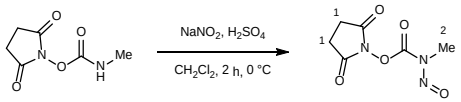
then concentrated *in vacuo* and resuspended in 10 mL of acetonitrile. Hydroxylamine hydrochloride (63.9 mg, 917 μmol, 1.10 equiv.) was added to the solution in one portion followed by NEt_3 (128 μL, 917 μmol, 1.10 equiv.) and 2 mL of DMF. This was stirred under N_2 for 2 h until conversion to the guanidine **15** was confirmed by mass spectrometry. The reaction was acidified with 100 μL of TFA and concentrated *in vacuo*. The residue was applied directly on a C18 Sep-Pak cartridge. After washing the cartridge with 100 mL of water + 0.1% formic acid and 100 mL of 5% acetonitrile in water + 0.1% formic acid, **15** was eluted with 200 mL of 15% acetonitrile in water + 0.1% formic acid. The fractions were combined, concentrated *in vacuo* to remove acetonitrile, and the resulting aqueous layer was lyophilized to dryness to afford **15** as a white solid (13 mg, 5% yield), which was used immediately without characterization.

The Cbz group was removed by dissolving **15** in 2 mL of 10% (*v*/*v*) thioanisole in TFA. After stirring for 4 h at room temperature, 5 mL of water was added to the reaction. The aqueous solution containing **2** was first washed with 2 x 5 mL of diethyl ether and then applied to preparative HPLC with the same gradient method as purifying amino acid **1** to afford 4 mg of amino acid **2** (49% yield from **15**) as a white solid. ^{1}H NMR (500 MHz D_2O): δ (ppm) = 3.80 (t, J = 6.0 Hz, 1H, H_1), 3.53 (t, J = 6.9 Hz, 2H, H_4), 2.98 (s, 3H, H_5), 2.03 – 1.74 (m, 4H, H_2 and H_3).^{13}C NMR (150 MHz D_2O): δ (ppm) = 174.3 (**C**), 162.3 (**C**), 54.5 (**C**H), 54.4 (**C**H_2), 29.5 (**C**H_2), 27.3 (**C**H_3), 21.7 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_4$ $[M-H]^-$ =219.1099, found = 219.1101.

Synthesis of Fmoc-N^{δ}-hydroxy-N^{ω}-methyl-N^{ω}-nitroso-L-citrulline (18)



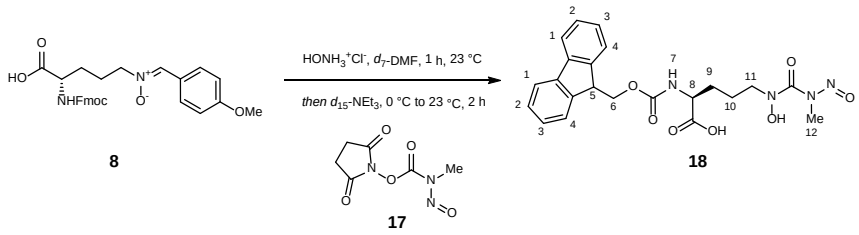
Step 1:



16 **17**

To *N*-succinimidyl *N*-methylcarbamate (**16**, 3.31 mmol, 570 mg, 1.00 equiv.) in 25 mL of CH_2Cl_2 was added 200 μL of concentrated sulfuric acid (H_2SO_4). The flask was cooled to 0 ºC, and sodium nitrite ($NaNO_2$) dissolved (343 mg, 4.97 mmol, 1.50 equiv.) in 5 mL of water was added dropwise over 5 min into the stirring solution. After 1 h, another equivalent of $NaNO_2$ and H_2SO_4 were added, and stirring was continued at 0 ºC for 1 h. 10 mL of water was then added to the reaction, and N_2 was bubbled through the solution for 5 min to remove NO_2. The biphasic solution was extracted with 3 x 50 mL of CH_2Cl_2. The combined organic layers were dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford crude **17**, which was purified using the CombiFlash Rf system. The column was a 40 g silica column (RediSep Rf Gold, Teledyne Isco). The flow rate was 40 mL/min. The LC conditions were: an isocratic hold at 0% solvent B for 2 min; a linear gradient from 0 to 100% solvent B in 13 min; an isocratic hold at 100% solvent B for 3 min (solvent A = hexanes, solvent B = ethyl acetate). **17** was monitored at 254 nm. The fractions containing **17** were pooled and concentrated *in vacuo* to obtain 32 mg of **17** as a light yellow powder (22% yield). ^{1}H NMR (600 MHz $CDCl_3$): δ (ppm) = 3.22 (s, 3H, H_2), 2.94 (s, 4H, H_1). ^{13}C NMR (125 MHz $CDCl_3$): δ (ppm) = 168.3 (**C**), 151.0 (**C**), 28.5 (**C**H_3), 25.6 (2 x **C**H_2).

Step 2:

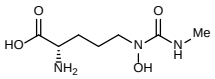


While we can synthesize **18** from **8** and **17**, we have been unsuccessful in isolating analytically pure **18** for characterization. Despite numerous attempts to isolate **18** by semi-preparative HPLC under different purification conditions, **18** readily degraded to the corresponding denitrosated product **19** and other uncharacterized products. We have obtained exact mass, a diagnostic MS/MS spectrum, and used labeled materials to confirm the structure of **18** from biochemical assays of SznF (**Extended Data Fig. 5**). We have also obtained NMR data for **18**, by performing the reaction in an NMR tube using deuterated base and solvent.

To a solution of nitrone **8** (6.00 mg, 12.3 μmol, 1.00 equiv.) dissolved in 450 μL of d_7-DMF (Sigma Aldrich) was added 6.66 μL of a 1 mM stock solution of hydroxylamine hydrochloride in d_7-DMF (1.03 mg, 14.8 μmol, 1.20 equiv.). After incubating the reaction in a 5 mm NMR tube for 1 h at room temperature, d_{15}-NEt_3 (Cambridge Isotope Inc.; 2.07 μL,14.8 μmol, 1.20 equiv.) was added, followed by **17** (2.47 mg, 12.3 μmol, 1.00 equiv.) dissolved in 50 μL of d_7-DMF. The reaction mixture was incubated for 30 min, and the conversion from **17** to **18** was analyzed by ^{1}H NMR. The integration cannot be accurately obtained since the conversion was not complete, and the overlapping peaks from **8**, **18**, and other intermediates made integration assignments difficult. ^{1}H NMR of **18** (400 MHz d_7-DMF): δ (ppm) = 7.96 – 7.91 (m, H_1), 7.82 – 7.75 (m), 7.49 – 7.32 (m), 4.32 – 4.01 (m, H_5, H_6, and H_8), 3.74 (t, J = 5.5 Hz, H_{11}), 3.15 (s, H_{12}), 2.02 – 1.59 (m, H_9 and H_{10}). ^{13}C NMR of **18** (100 MHz d_7-DMF): δ (ppm) = 175.5

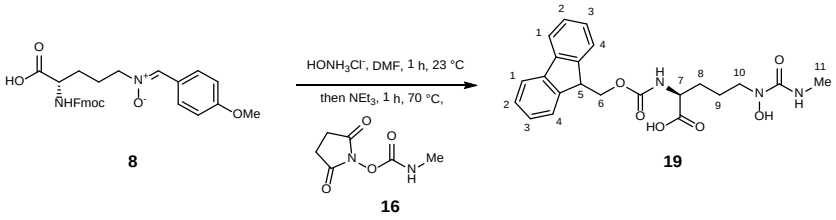
(**C**), 163.5 (**C**), 155.5 (**C**), 144.4 (2 x **C**), 141.3 (2 x **C**), 127.8 (2 x **C**H), 127.3 (2 x **C**H), 126.0 (2 x **C**H), 120.3 (2 x **C**H), 66.4 (**C**H_2), 55.1 (**C**H), 50.7 (**C**H_2), 47.3 (**C**H), 30.5 (**C**H_3), 29.6 (**C**H_2), 23.4 (**C**H_2). HRMS for **18**: Calc'd for $C_{22}H_{23}N_4O_7$ $[M-H]^-$ = 455.1572, found = 455.1588.

Synthesis of N^{δ}-hydroxy-N^{ω}-methyl-L-citrulline (4)



4

Step 1:

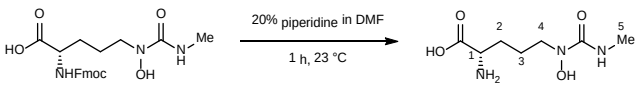


8 **16** **19**

To nitrone **8** (100 mg, 0.205 mmol, 1.00 equiv.) dissolved in 20 mL of DMF was added hydroxylamine hydrochloride (17.0 mg, 0.246 mmol, 1.20 equiv.) in one portion and stirred for 1 h under N_2. Then, *N*-succinimidyl-*N*-methylcarbamate (**16**, 38.7 mg, 2.26 μmol, 1.10 equiv.) and NEt_3 (57.0 μL, 0.410 mmol, 2.00 equiv.) were added in one portion, and stirring was continued for 16 h at room temperature. 50 mL of ethyl acetate were added to the reaction, the layers were separated, and the organic layer was washed with 10 mL of 1 M HCl, 10 mL of water, and 10 mL of brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to dryness. The residue was dissolved in methanol and crudely purified with a C18 Sep-pak cartridge. After an initial wash with 100 mL of water + 0.1% formic acid to remove salts, crude

19 was eluted with 200 mL of 30% acetonitrile + 0.1% formic acid. The desired fraction was concentrated *in vacuo* to remove the acetonitrile, affording an aqueous solution containing **19**. The aqueous solution was then extracted with 2 x 50 mL of ethyl acetate. The combined organic layer was dried over Na_2SO_4, filtered, and evaporated to afford a mixture of **19** and *p*-anisaldeyhde (75:25), which was carried forward without further purification. ^{1}H NMR for **19** (600 MHz d_7-DMF): δ (ppm) = 7.93 (d, J = 7.5 Hz, 2H, H_1), 7.61 – 7.58 (m, 2H, H_4), 7.44 (t, J = 7.4 Hz, 2H, H_2), 7.38 – 7.33 (m, 2H, H_3), 4.35 – 4.27 (m, 3H, H_5 and H_6), 4.24 – 4.19 (m, 1H, H_7), 3.51 (t, J = 6.4 Hz, 2H, H_{10}), 2.75 (d, J = 4.7 Hz, 3H, H_{11}), 2.00 – 1.72 (m, 4H, H_8 and H_9). ^{13}C NMR for **19** (600 MHz d_7-DMF): δ (ppm) = 174.2 (**C**), 160.8 (**C**), 156.7 (**C**), 144.4 (2 x **C**), 144.3 (2 x **C**), 128.1 (2 x **C**H), 127.8 (2 x **C**H), 127.3 (2 x **C**H), 120.3 (2 x **C**H), 66.5 (**C**H_2), 55.1 (**C**H_3), 54.59 (**C**H), 50.4 (**C**H_2), 47.2 (**C**H), 29.1 (**C**H_2), 23.9 (**C**H_2). HRMS for **19**: Calc'd for $C_{22}H_{24}N_3O_6$ $[M\text{-}H]^-$ = 426.1671, found = 426.1679.

Step 2:

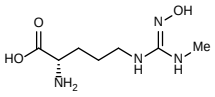


19 **4**

Crude urea **19** from previous step was dissolved in 10 mL of 20% (*v/v*) piperidine in DMF and stirred for 4 h at room temperature. The solvent was removed *in vacuo*, and the solids were dissolved in 5 mL of water. The aqueous solution containing crude **4** was washed with 2 x 5 mL of diethyl ether to remove nonpolar impurities and then applied to reverse phase semi-preparative HPLC. The LC column was Kromasil 100 C18 (250 x 10 μm, 5mm). The flow rate was 3 mL/min. The LC conditions were: 5 to 12% solvent B in 10 min; a linear gradient from 12

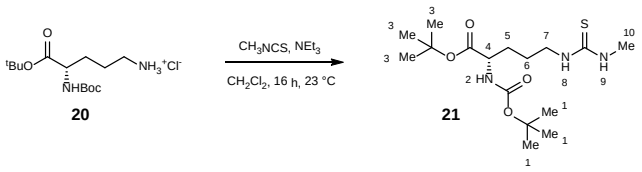
to 30% solvent B in 2.5 min; a linear gradient from 30 to 5% solvent B in 1.5 min; and an isocratic hold at 5% solvent B for 3 min (solvent A = 0.1 % formic acid in water, solvent B = 0.1% formic acid in acetonitrile). **4** was monitored at 200 nm and eluted between 4.6 – 5.2 min. The collected fractions were lyophilized to afford 14 mg of **4** (33% yield from **8**). ^{1}H NMR (600 MHz D_2O): 3.97 (t, *J* = 6.3 Hz, 1H, H_1), 3.52 (t, *J* = 6.7 Hz, 2H, H_4), 2.74 (s, 3H, H_5), 2.02 – 1.88 (m, 2H, H_2), 1.78 – 1.67 (m, 2H, H_3). ^{13}C NMR (125 MHz DMSO): δ (ppm) = 171.4 (**C**), 161.8 (**C**), 52.3 (**C**H), 50.1 (**C**H_2), 27.9 (**C**H_2), 26.9 (**C**H_3), 22.4 (**C**H_2). HRMS: Calc'd for $C_7H_{14}N_3O_4$ $[M-H]^-$ = 204.0990, found = 204.0991.

Synthesis of N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (5)



5

Step 1:

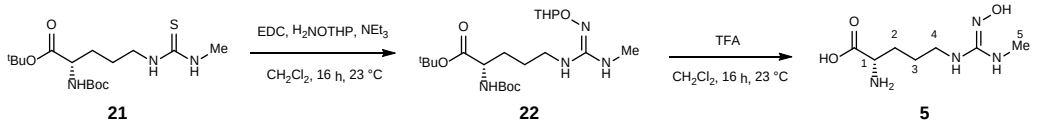


20 **21**

Compound **5** was prepared by adapting the procedures of Jansen Labby *et al.*[60] The protected amino acid **20** (Bachem, 200 mg, 617 μmol, 1.00 equiv.) was dissolved in 50 mL of anhydrous CH_2Cl_2 at 0 ºC, and NEt_3 (172 μL, 1.23 mmol, 2.00 equiv.) was added to the solution in one portion. Then methyl isothiocyanate (CH_3NCS, 45.1 mg, 617 μmol, 1.00 equiv.) dissolved

in 5 mL of CH_2Cl_2 was added dropwise to the reaction mixture, and the reaction was allowed to warm to room temperature while stirring overnight. The reaction mixture was washed with 1 x 10 mL of 1M HCl, 1 x 10 mL of saturated sodium bicarbonate, 1 x 10 mL of water, and 1 x 10 mL of brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford crude **21**. The compound was purified with silica gel chromatography (70:30 hexanes: ethyl acetate) to afford **21** (155 mg, 69.5% yield). ^{1}H NMR (600 MHz $CDCl_3$): δ 6.40 (br s, 2H, NH_8 and NH_9), 5.24 (d, J = 7.9 Hz, 1H, NH_2), 4.17 – 4.08 (m, 1H, H_4), 3.57 – 3.43 (m, 2H, H_7), 3.02 (s, 3H, H_{10}), 1.85 – 1.60 (m, 4H, H_5 and H_6), 1.46 (d, J = 1.9 Hz, 9H), 1.43 (d, J = 2.0 Hz, 9H). Calc'd for $C_{16}H_{32}N_3O_4S$ $[M+H]^+$ = 362.2108, found = 362.2086.

Step 2 – 3:



21 **22** **5**

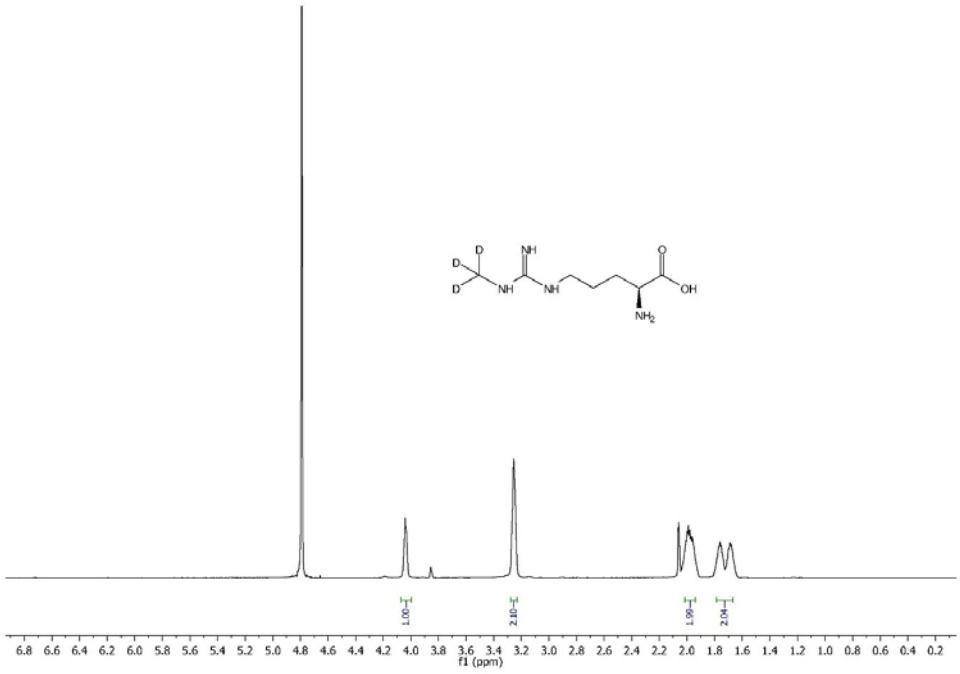
To thiourea **21** (96.0 mg, 169 μmol, 1.00 equiv.) dissolved in 20 mL of CH_2Cl_2 was added EDC (28.9 mg, 186 μmol 1.10 equiv.), NEt_3 (47.1 μL, 338 μmol, 2.00 equiv.), and 4-(aminooxy)tetrahydro-2*H*-pyran (H_2NOTHP, Santa Cruz Biotechnology; 19.8 mg, 169 μmol, 1.00 equiv.) in one portion. After 16 h of stirring at room temperature, the reaction mixture was washed with 1 x 10 mL of brine. The layers were separated, and the organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford a light yellow oil. The crude oil was dissolved in 20 mL of CH_2Cl_2 and 5 mL of TFA and stirred at room temperature for 16 h. The reaction mixture was then concentrated *in vacuo*. The residue was dissolved in 5 mL of water and the aqueous solution was washed with 2 x 5 mL of diethyl ether to remove nonpolar

impurities. The crude product in the aqueous layer was purified by preparative HPLC using the same column and method as described for purifying **1**. Lyophilizing the collected fractions afforded **5** as a white solid (11 mg, 32% yield). ^{1}H NMR (500 MHz, D_2O): δ 3.78 (t, *J* = 6.1 Hz, 1H, H_1), 3.26 (t, *J* = 6.9 Hz, 2H, H_4), 2.84 (s, 3H, H_5), 1.96 – 1.87 (m, 2H, H_2), 1.80 – 1.62 (m, 2H, H_3). ^{13}C NMR (100 MHz, D_2O): δ 174.4 (**C**), 157.8 (**C**), 54.3 (**C**H), 40.1 (**C**H_2), 27.6 (**C**H_2), 27.0 (**C**H_3), 23.8 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_3$ $[M-H]^-$ = 203.1150, found = 203.1159.

Supplementary Figures

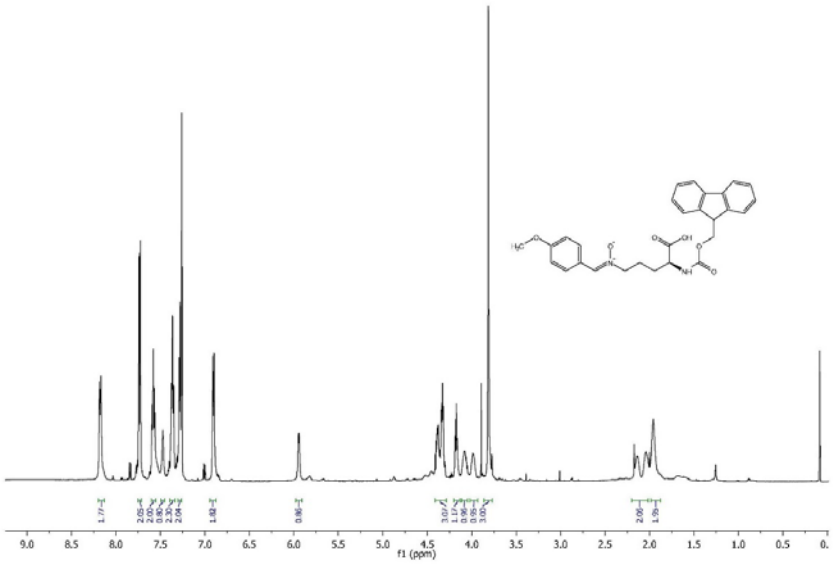
Supplementary Figure 1 | NMR spectra of d_3-L-NMA. **a,** ^{1}H-NMR of d_3-L-NMA (recorded in D_2O at 600 MHz).

a

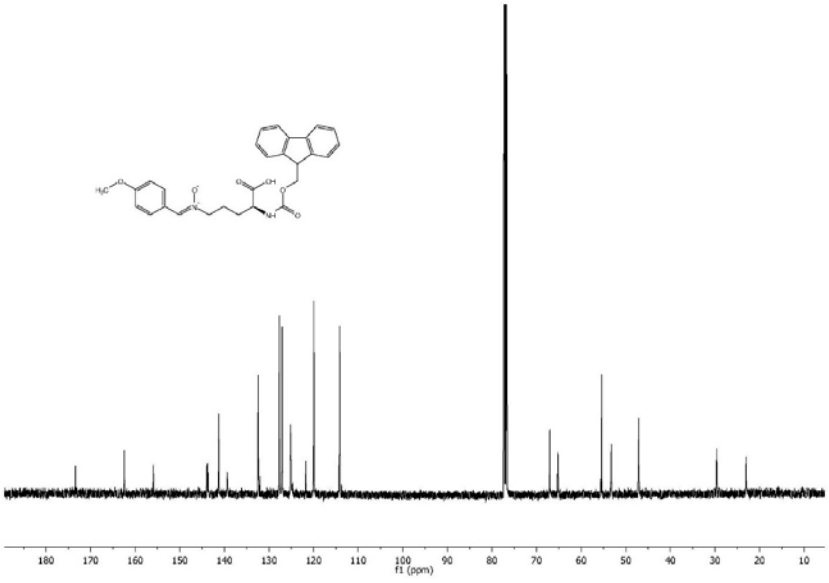


Supplementary Figure 2 | NMR spectra of **8**. **a**, ^{1}H-spectrum of **8** (recorded in $CDCl_3$ at 600 MHz). **b**, ^{13}C-spectrum of **8** (recorded in $CDCl_3$ at 125 MHz).

a

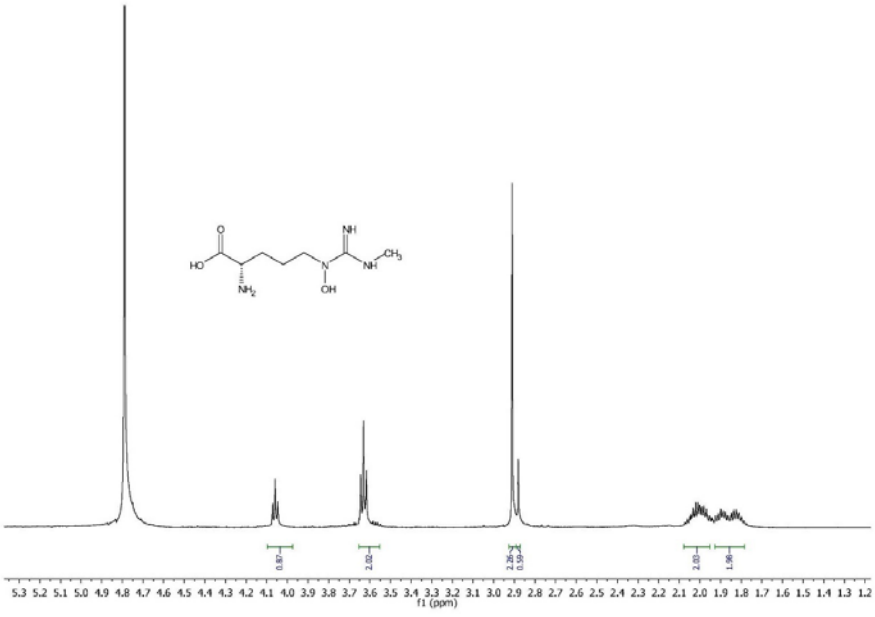


b

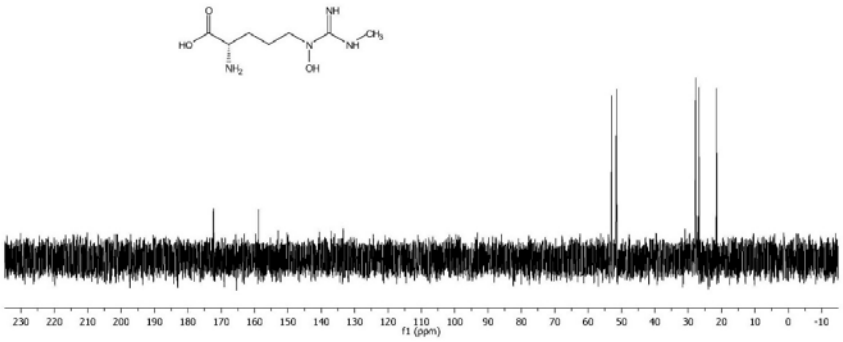


Supplementary Figure 3 | NMR spectra of **1**. **a**, ^{1}H-spectrum of **1** (recorded in D_2O at 500 MHz). **b**, ^{13}C-spectrum of **1** (recorded in D_2O at 125 MHz).

a

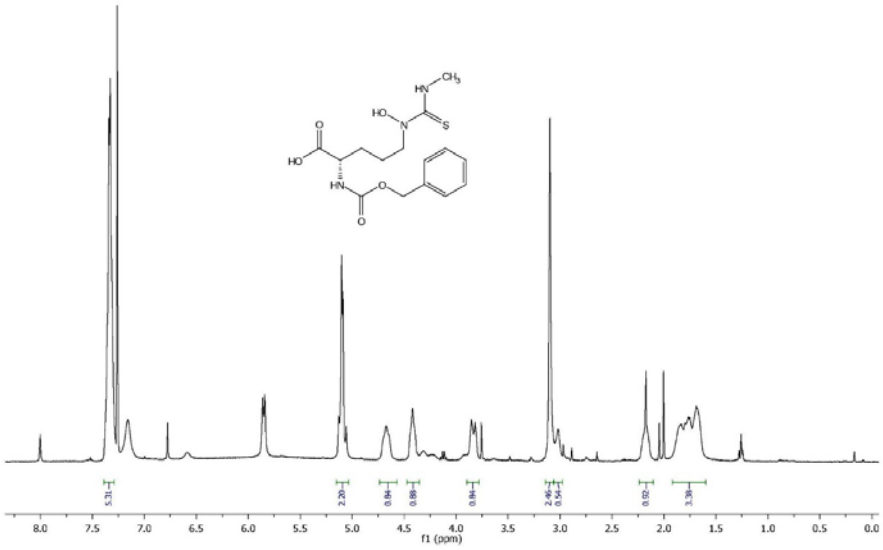


b

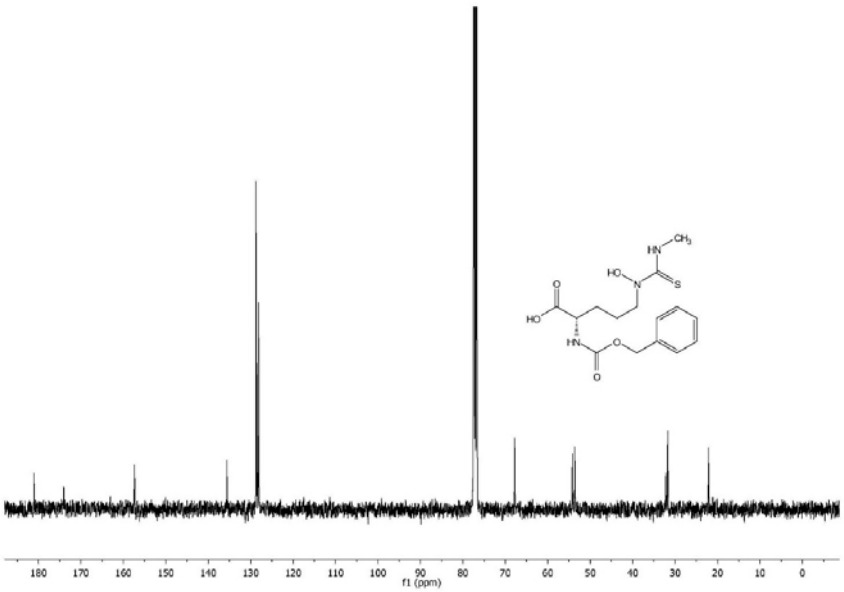


Supplementary Figure 4 | NMR spectra of **13**. **a**, ^{1}H-spectrum of **13** (recorded in $CDCl_3$ at 400 MHz). **b**, ^{13}C-spectrum of **13** (recorded in $CDCl_3$ at 100 MHz). **c**, ^{1}H–^{13}C-HSQC spectrum of **13** (recorded in $CDCl_3$ at 400 MHz and 100 MHz).

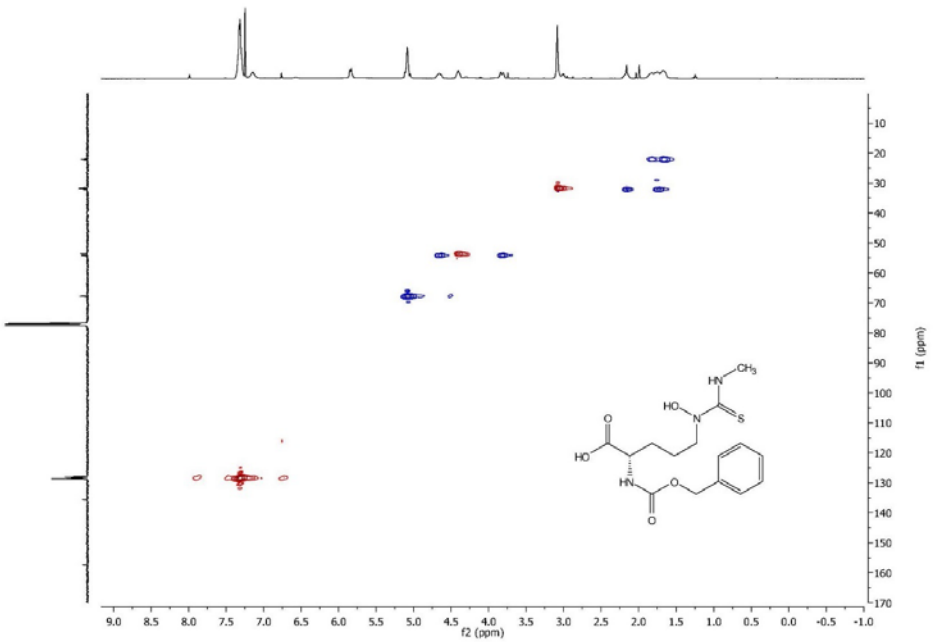
a



b

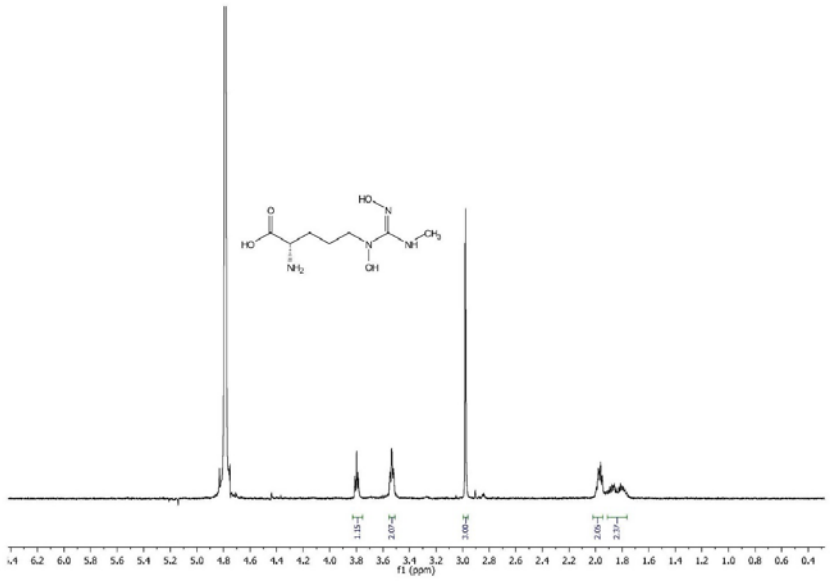


c

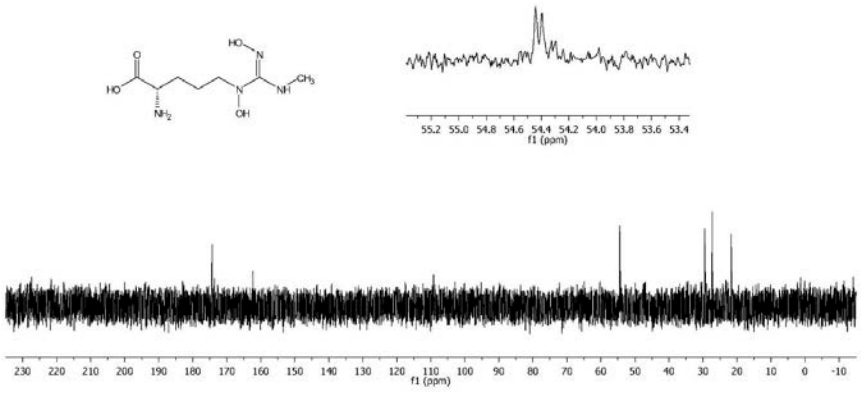


Supplementary Figure 5 | NMR spectra of **2**. **a,** ^{1}H-spectrum of **2** (recorded in D_2O at 500 MHz). **b**, ^{13}C-spectrum of **2** (recorded in D_2O at 150 MHz).

a

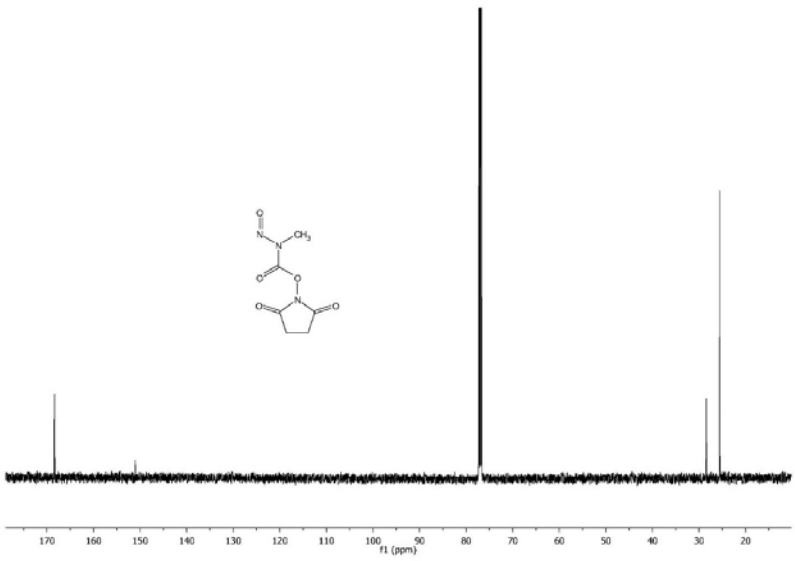


b

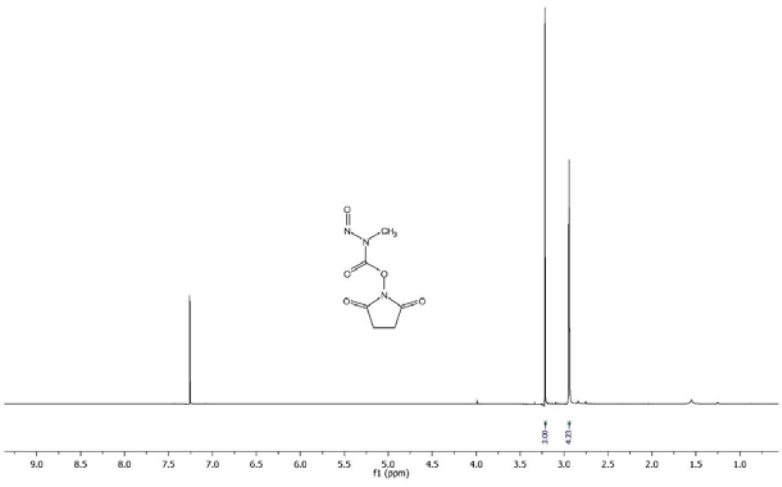


Supplementary Figure 6 | NMR spectra of **17**. **a**, ^{1}H-spectrum of **17** (recorded in $CDCl_3$ at 600 MHz). **b**, ^{13}C-spectrum of **17** (recorded in $CDCl_3$ at 125 MHz).

a

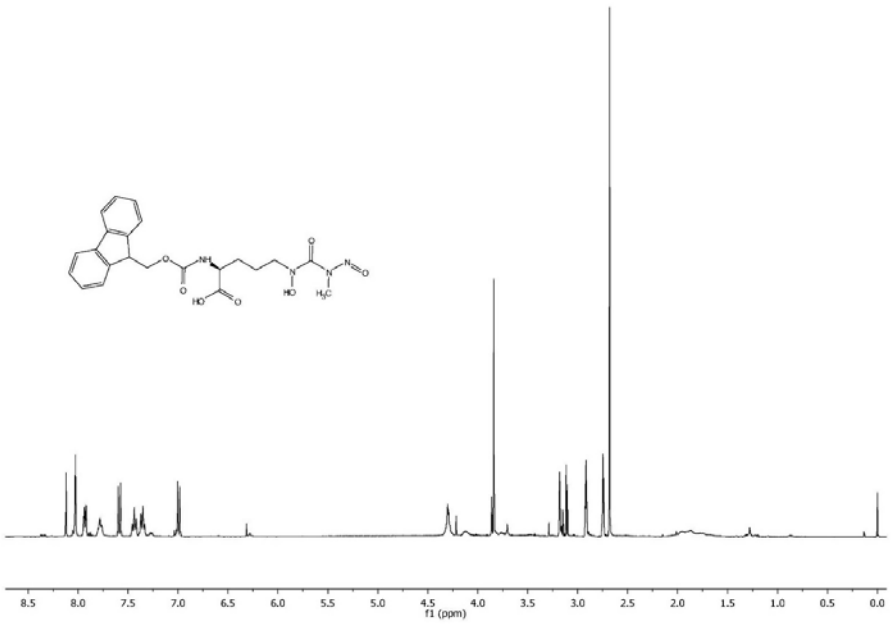


b

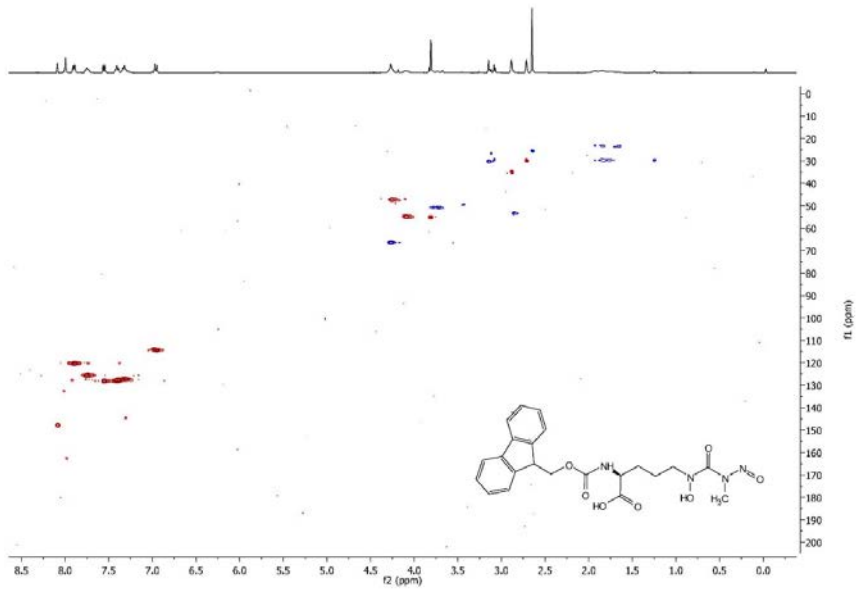


Supplementary Figure 7 | NMR spectra of **18**. **a**, ^{1}H-spectrum of **18** (recorded in d_7-DMF at 400 MHz). **b**, ^{1}H–^{13}C-HSQC spectrum of **18** (recorded in d_7-DMF at 400 MHz and 100 MHz). **c**, ^{1}H–^{13}C-HMBC spectrum of **18** (recorded in d_7-DMF at 400 MHz and 100 MHz). **d**, ^{1}H–^{1}H-COSY spectrum of **18** (recorded in d_7-DMF at 400 MHz).

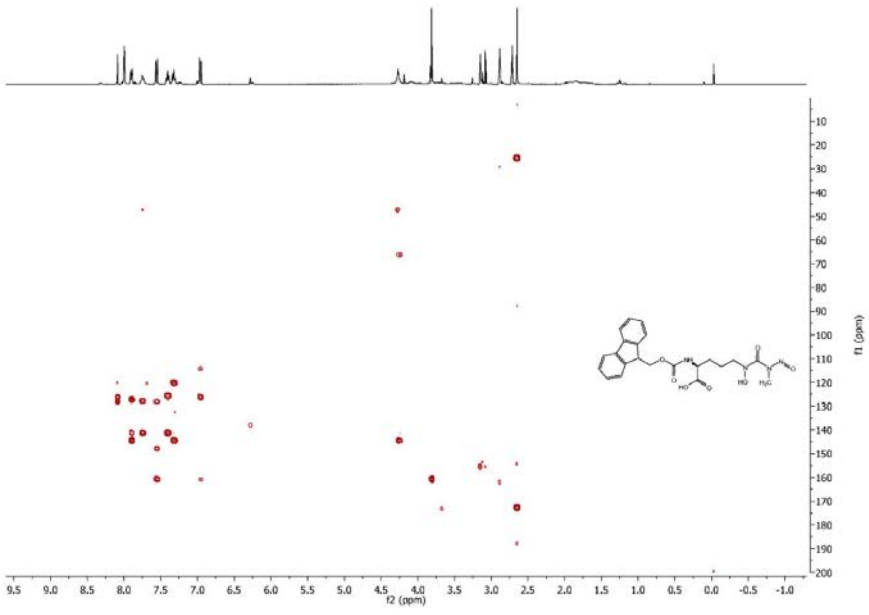
a



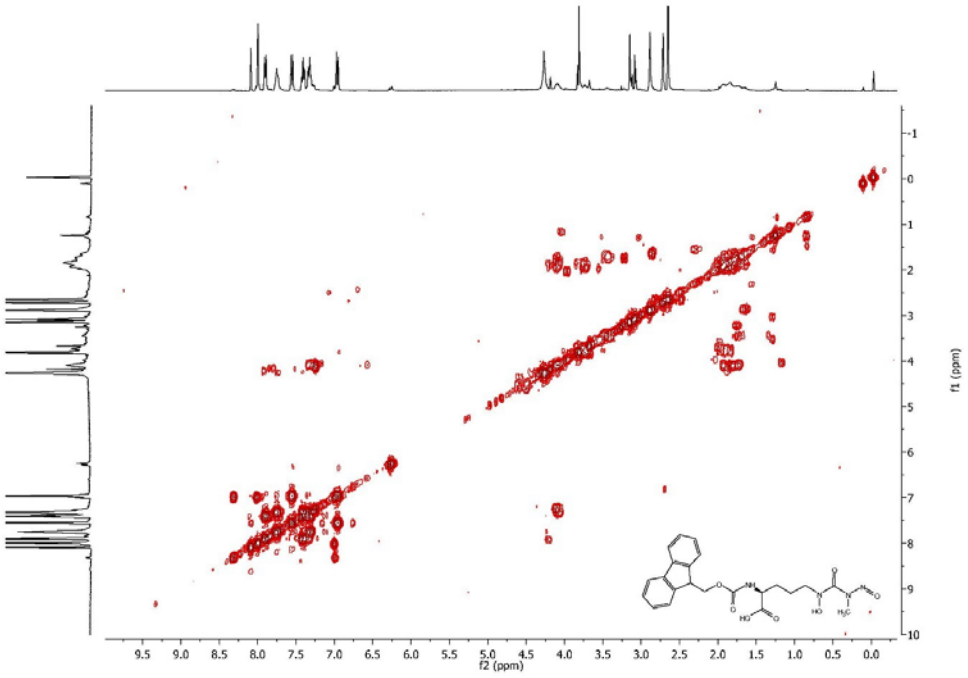
b



c

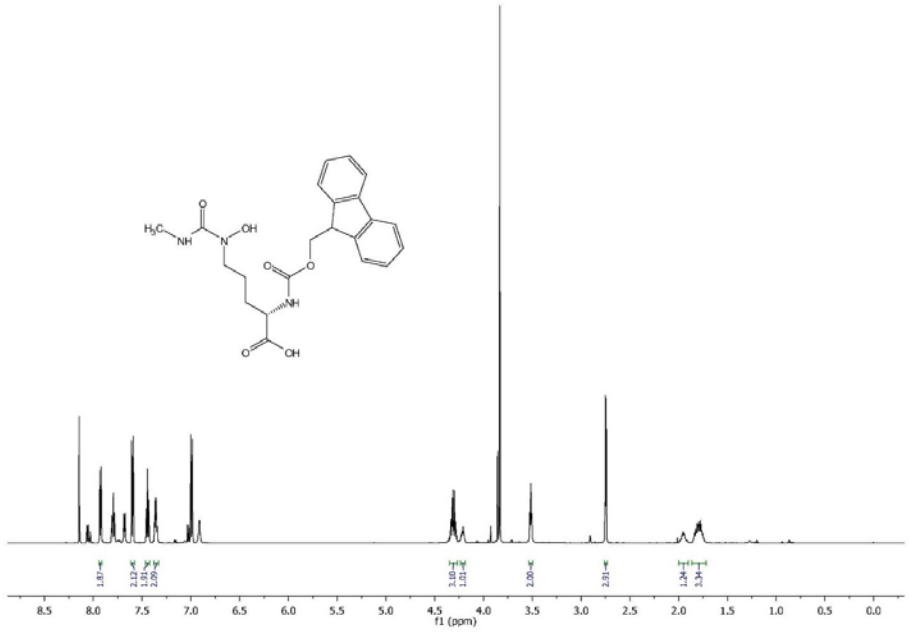


d

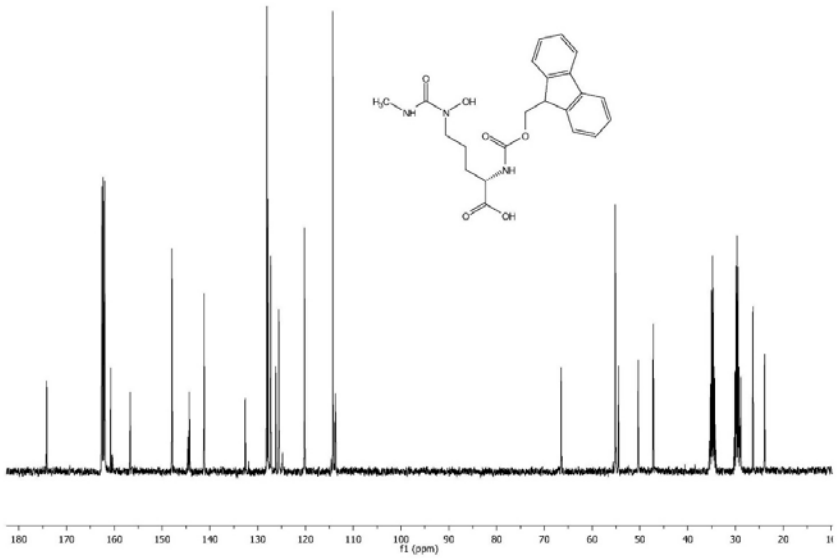


Supplementary Figure 8 | NMR spectra of **19**. **a**, ^{1}H-spectrum of **19** (recorded in d_7-DMF at 600 MHz). **b**, ^{13}C-spectrum of **19** (recorded in d_7-DMF at 100 MHz). **c**, ^{1}H–^{13}C- HSQC spectrum of **19** (recorded in d_7-DMF at 400 MHz and 100 MHz).

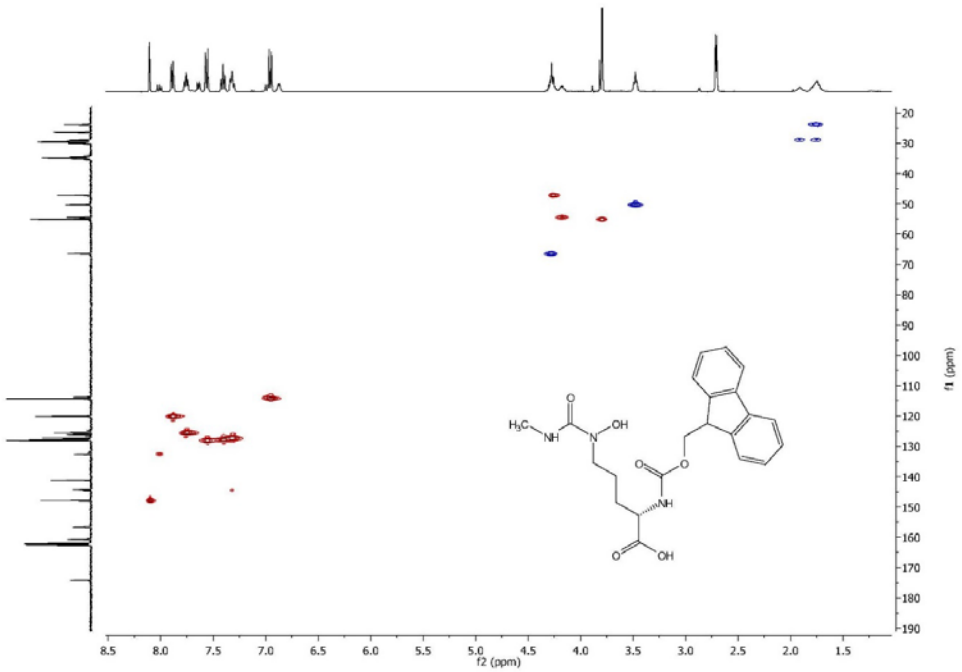
a



b

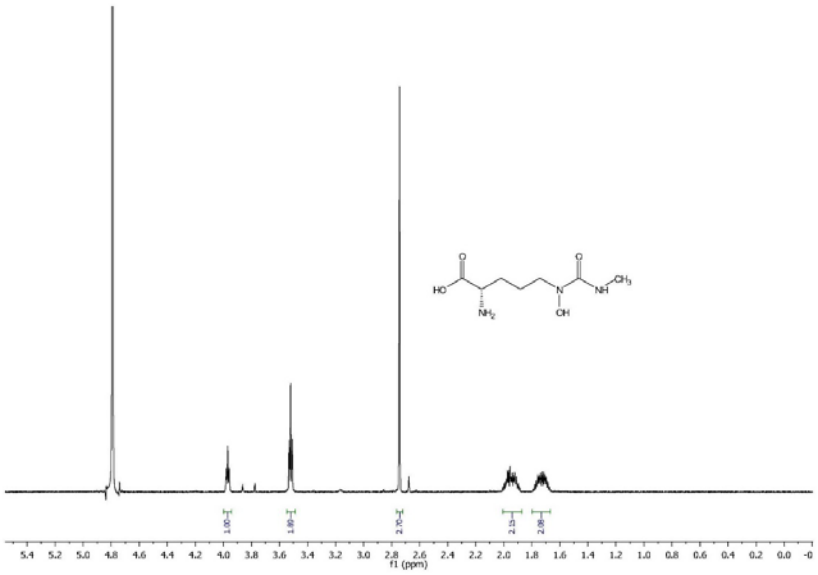


c

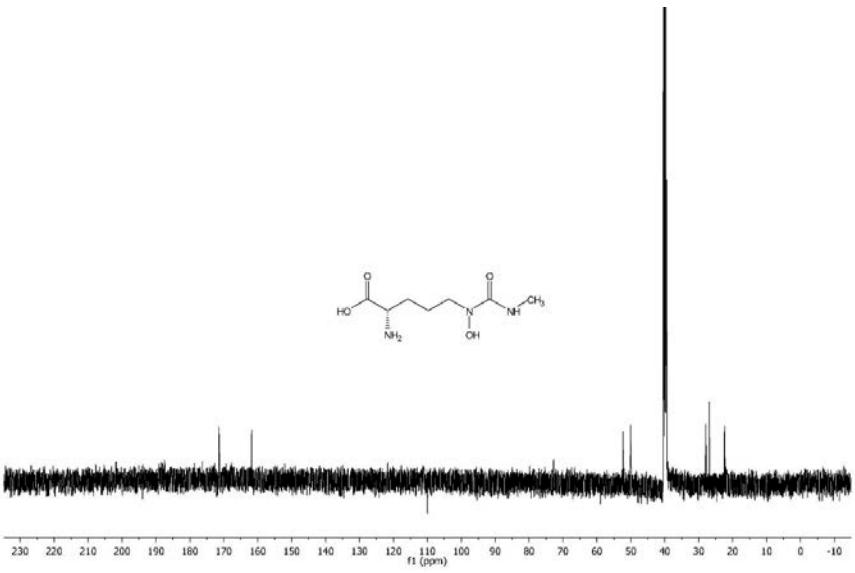


Supplementary Figure 9 | NMR spectra of **4**. **a**, ^{1}H-spectrum of **4** (recorded in D_2O at 600 MHz). **b**, ^{13}C-spectrum of **4** (recorded in d_6-DMSO at 125 MHz).

a

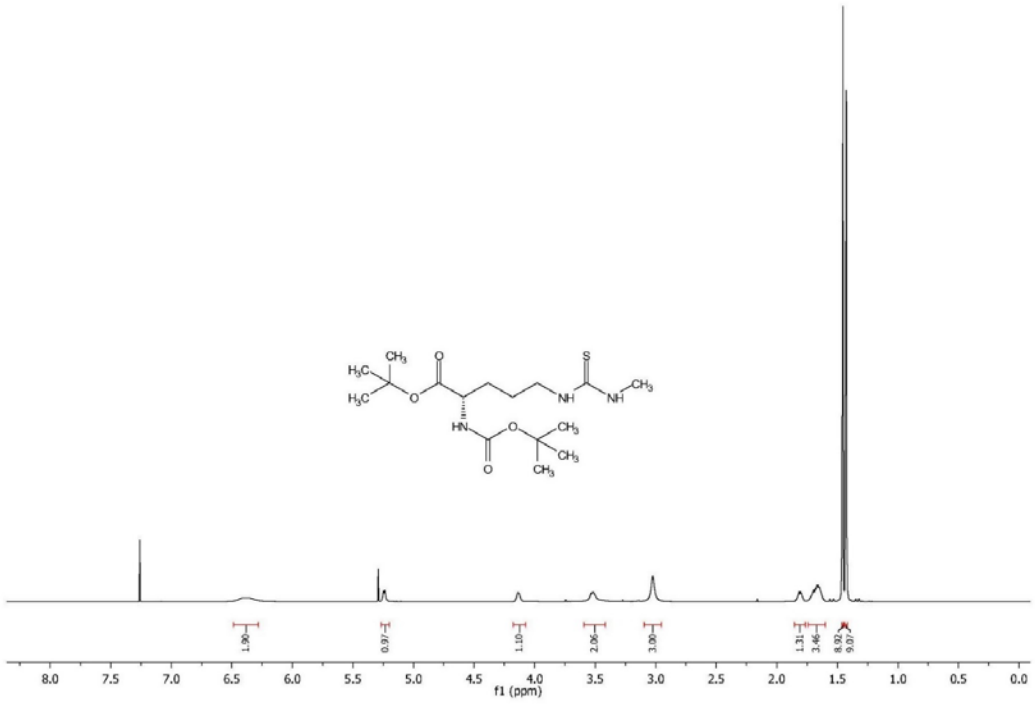


b



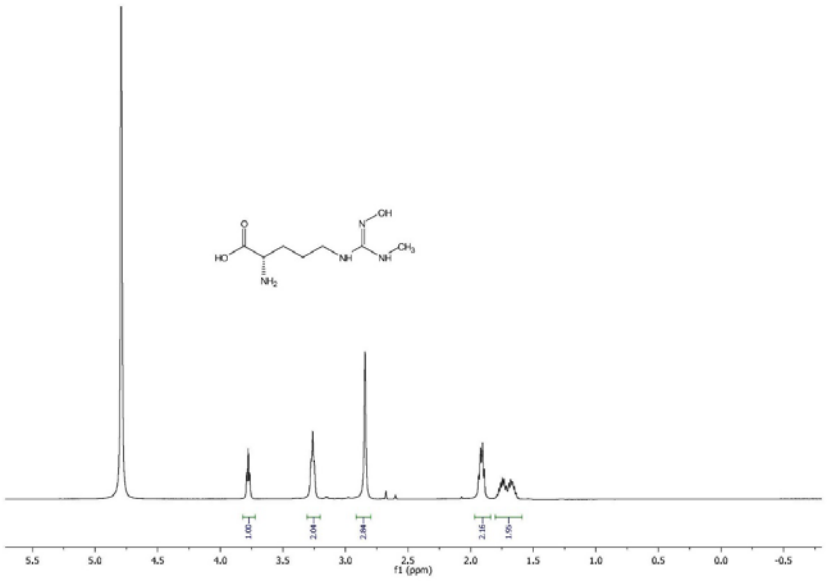
Supplementary Figure 10 | NMR spectra of **21**. **a**, ^{1}H-spectrum of **21** (recorded in $CDCl_3$ at 400 MHz).

a

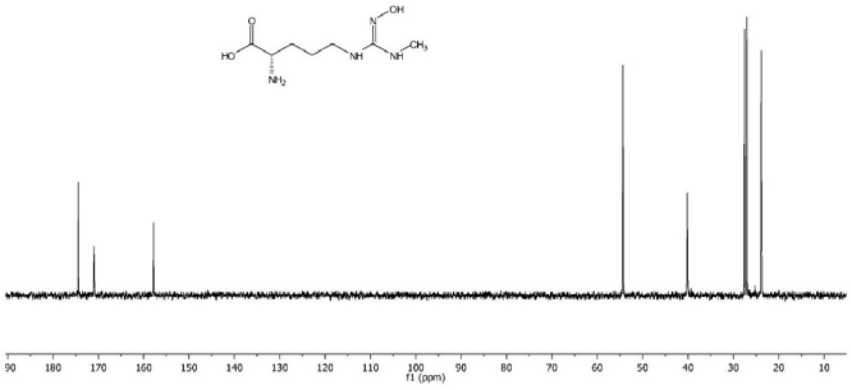


Supplementary Figure 11 | NMR spectra of **5**. **a**, ^{1}H-spectrum of **5** (recorded in D_2O at 500 MHz). ^{13}C-spectrum of **5** (recorded in D_2O at 100 MHz).

a



b



2. Supplementary Tables

Supplementary Table 1 | Annotations of the *szn* biosynthetic genes

Protein	Size, aa	Predicted function	Closest homolog	Accession number of homolog	% Identity/Similarity
SznA	197	RNA polymerase sigma-70 factor	*Streptomyces sp. di50b*	SCD88966.1	85/90
SznB	266	O6-methylguanine-DNA--protein-cysteine methyltransferase (MGMT)	*Streptomyces sp. di50b*	SCD88943.1	82/87
SznC	181	Methylated-DNA-[protein]-cysteine S-methyltransferase	*Streptomyces sp. di50b*	SCD88925.1	81/85
SznD	255	Proline hydroxylase (Fe(II)-2-oxoglutarate-dependent oxygenase family, AlkB homolog)	*Streptomyces sp. NRRL F-2580*	WP_078892722.1	76/81
SznE	361	methyltransferase domain-containing protein	*Streptomyces virginiae*	WP_078908984.1	72/82
SznF	471	Iron-containing redox enzyme	*Streptomyces sp. di50b*	SCD88894.1	94/97
SznG	124	Rieske [2Fe-2S] domain-containing protein	*Streptomyces sp. di50b*	SCD88886	82/87
SznH	416	ATP-grasp domain-containing protein	*Streptomyces sp. SM1*	WP_103542806.1	86/92
SznI	431	MFS transporter	*Streptomyces sp. SM1*	WP_103542807.1	86/90
SznJ	318	Alpha/beta hydrolase	*Streptomyces sp. SM1*	WP_103542817.1	89/92
SznK	424	ATP-grasp domain-containing protein	*Streptomyces sp. SM1*	WP_103542808.1	87/91
SznL	210	LysE family translocator	*Streptomyces sp. SM1*	WP_103542809.1	75/82

Supplementary Table 2 | SznF homologs containing the central and cupin domains obtained from BLAST searches (NCBI non-redundant (nr) sequences Database 2018) and used for maximum likelihood phylogenetic analysis (**Extended Data Fig. 10a**)

Accession	Description	e- value	Organism	% Pairwise Identity	Query coverage
WP_093552082	MULTISPECIES: cupin	0	*Streptomyces*	93.60%	100.00%
WP_103542804	cupin domain-containing protein	0	*Streptomyces sp.*	92.10%	100.00%
WP_023588127	cupin domain-containing protein	0	*Streptomyces thermolilacinus*	89.40%	100.00%
WP_020139892	iron-containing redox enzyme family protein	0	*Streptomyces sp.*	88.30%	100.00%
WP_030714008	iron-containing redox enzyme family protein	0	*Streptomyces sp.*	87.90%	100.00%
WP_033214908	iron-containing redox enzyme family protein	0	*Streptomyces virginiae*	87.70%	100.00%
WP_030755298	MULTISPECIES: iron-containing redox enzyme family protein	0	*Streptomyces*	87.70%	100.00%
WP_030008798	iron-containing redox enzyme family protein	0	*Streptomyces lavendulae*	87.30%	100.00%
WP_030707683	iron-containing redox enzyme family protein	0	*Streptomyces sp.*	86.40%	100.00%
WP_053923971	cupin	0	*Streptomyces chattanoogensis*	75.90%	100.00%
WP_046925515	cupin	0	*Streptomyces lydicus*	75.90%	100.00%
WP_083103992	cupin	0	*Streptomyces gilvosporeus*	75.10%	100.00%
WP_030072619	cupin	0	*Streptomyces natalensis*	75.10%	100.00%
WP_095563443	cupin	0	*Plantactinospora sp.*	73.10%	100.00%
WP_084993114	cupin	0	*Streptomyces sp.*	71.80%	100.00%
WP_051965704	cupin	0	*Kitasatospora mediocidica*	71.60%	100.00%
WP_028416834	cupin	0	*Streptomyces sp.*	71.60%	100.00%
WP_103416709	cupin	0	*Streptomyces sp.*	71.60%	100.00%
WP_018490631	hypothetical protein	0	*Streptomyces sp.*	71.60%	100.00%
WP_092775268	iron-containing redox enzyme family protein	0	*Actinokineospora terrae*	71.50%	100.00%
WP_099217632	cupin	0	*Streptomyces sp.*	71.40%	100.00%
WP_093751846	cupin	0	*Streptomyces sp.*	71.40%	100.00%
WP_025356251	iron-containing redox enzyme family protein	0	*Kutzneria albida*	71.00%	100.00%
WP_045302363	iron-containing redox enzyme family protein	0	*Saccharothrix sp.*	68.40%	100.00%
WP_104816691	iron-containing redox enzyme family protein	0	*Kitasatospora sp.*	67.40%	100.00%
WP_092774762	iron-containing redox enzyme family protein	0	*Actinokineospora terrae*	66.00%	100.00%
WP_085189693	iron-containing redox enzyme family protein	2.13E-157	*Mycolicibacterium doricum*	52.10%	95.75%
WP_096558266	cupin	3.15E-172	*Nostoc sp.*	51.10%	96.60%
WP_095990390	cupin	1.03E-163	*Cystobacter fuscus*	50.40%	100.00%
WP_094329510	cupin	2.54E-169	*Nostoc sp.*	50.00%	96.60%
WP_096538580	cupin	1.13E-168	*Nostoc linckia*	49.60%	96.60%
WP_037583284	cupin	1.96E-162	*Stigmatella aurantiaca*	49.60%	97.03%
ADO74274	Cupin domain protein	1.97E-162	*Stigmatella aurantiaca*	49.60%	97.03%
EAU67364	Cupin domain protein	6.00E-162	*Stigmatella aurantiaca*	49.60%	97.03%
WP_050671291	iron-containing redox enzyme family protein	6.45E-149	*Luteipulveratus halotolerans*	49.30%	95.97%
ABE58103	Cupin 2 protein	3.54E-120	*Chromohalobacter salexigens*	44.80%	94.06%
WP_083831073	cupin domain-containing protein	3.01E-120	*Chromohalobacter salexigens*	44.80%	94.06%
PWC95482	hypothetical protein	1.68E-121	*Azospirillum sp.*	44.80%	92.99%

WP_057652341	iron-containing redox enzyme family protein	8.62E-143	*Bordetella sp.*	44.70%	99.36%
WP_051141068	cupin domain-containing protein	1.02E-122	*Azospirillum brasilense*	44.40%	95.75%
COS96722	Uncharacterised protein	2.91E-112	*Streptococcus pneumoniae*	44.30%	86.20%
WP_057721888	cupin domain-containing protein	8.22E-96	*Pseudomonas orientalis*	43.90%	76.65%
WP_101218645	cupin domain-containing protein	2.22E-95	*Pseudomonas fluorescens*	43.60%	76.65%
WP_065949880	cupin domain-containing protein	5.69E-96	*Pseudomonas sp.*	43.60%	76.65%
WP_098144889	cupin	1.11E-122	*Bacillus toyonensis*	43.50%	96.39%
WP_098709607	cupin	3.27E-123	*Bacillus toyonensis*	43.50%	96.39%
WP_097857913	cupin	3.60E-123	*Bacillus wiedmannii*	43.50%	96.39%
WP_104505325	cupin domain-containing protein	8.76E-95	*Pseudomonas orientalis*	43.50%	76.65%
WP_008044285	cupin	6.74E-118	*Reinekea blandensis*	43.40%	95.54%
WP_027977269	cupin	6.23E-116	*gamma proteobacterium*	43.20%	94.48%
POM13861	cupin	7.85E-94	*Pseudomonas sp.*	43.20%	76.65%
WP_065935007	cupin domain-containing protein	9.74E-94	*Pseudomonas sp.*	43.20%	76.65%
WP_065936459	cupin domain-containing protein	7.52E-94	*Pseudomonas sp.*	43.20%	76.65%
WP_065894425	cupin domain-containing protein	7.12E-94	*Pseudomonas sp.*	43.20%	76.65%
WP_018749395	iron-containing redox enzyme family protein	1.55E-114	*Chitiniphilus shinanonensis*	43.20%	94.27%
WP_065887934	MULTISPECIES: cupin domain-containing protein	5.10E-94	*Pseudomonas*	43.20%	76.65%
WP_017560035	cupin	1.76E-123	*Bacillus cereus*	43.10%	96.39%
WP_063245995	cupin	1.45E-123	*Bacillus cereus*	43.10%	96.39%
WP_103688962	cupin	5.54E-99	*Pseudomonas syringae*	43.10%	83.86%
WP_048655621	cupin	5.99E-123	*Bacillus cereus*	42.90%	96.39%
WP_101194877	cupin	1.41E-122	*Bacillus sp.*	42.90%	96.39%
WP_003417755	cupin	1.12E-97	*Pseudomonas syringae*	42.90%	83.86%
WP_004410997	cupin	4.21E-97	*Pseudomonas syringae*	42.90%	83.86%
WP_046269897	cupin	1.42E-97	*Pseudomonas syringae*	42.90%	83.86%
WP_046237248	cupin	8.80E-98	*Pseudomonas syringae*	42.90%	83.86%
WP_043274201	cupin domain-containing protein	1.88E-92	*Pseudomonas sp.*	42.90%	77.28%
WP_092458556	cupin domain-containing protein	5.63E-96	*Pseudomonas sp.*	42.90%	77.71%
WP_016125802	hypothetical protein	5.68E-123	*Bacillus cereus*	42.90%	96.39%
WP_000455863	MULTISPECIES: cupin	6.61E-123	*Bacillus*	42.90%	96.39%
WP_100650930	cupin	6.79E-122	*Bacillus cereus*	42.70%	96.39%
WP_099584162	cupin domain-containing protein	2.55E-90	*Pseudomonas sp.*	42.50%	77.28%
WP_100073249	cupin	4.45E-97	*Pseudomonas syringae*	42.30%	83.86%
WP_103695707	cupin	2.89E-97	*Pseudomonas syringae*	42.30%	83.86%
WP_053133381	cupin domain-containing protein	3.09E-93	*Pseudomonas sp.*	42.30%	77.28%
WP_083236465	cupin domain-containing protein	3.30E-93	*Pseudomonas sp.*	42.30%	77.28%
WP_090409762	cupin domain-containing protein	2.57E-92	*Pseudomonas grimontii*	42.20%	77.28%
WP_083244852	cupin domain-containing protein	1.58E-92	*Pseudomonas sp.*	42.20%	77.28%
WP_027603768	MULTISPECIES: cupin domain-containing protein	1.47E-92	*Pseudomonas*	42.20%	77.28%
WP_083221146	MULTISPECIES: cupin domain-containing protein	1.70E-92	*Pseudomonas*	42.20%	77.28%
WP_045155341	cupin domain-containing protein	1.42E-94	*Pseudomonas fluorescens*	42.10%	77.71%

CKG37990	Uncharacterised protein	9.22E-117	*Streptococcus pneumoniae*	42.10%	96.39%
WP_103751254	cupin	5.05E-96	*Pseudomonas syringae*	42.00%	83.86%
WP_057003840	MULTISPECIES: cupin domain-containing protein	2.68E-92	*Pseudomonas*	42.00%	77.28%
WP_099235928	cupin	2.11E-92	*Pseudomonas sp.*	41.90%	83.86%
WP_092486928	cupin domain-containing protein	1.06E-93	*Pseudomonas sp.*	41.90%	77.28%
WP_069786532	cupin	4.49E-92	*Pseudomonas salomonii*	41.80%	84.08%
WP_065911208	cupin	5.28E-92	*Pseudomonas sp.*	41.80%	84.08%
WP_039593169	cupin domain-containing protein	5.10E-94	*Pseudomonas frederiksbergensis*	41.80%	77.71%
SDU01822	cupin domain-containing protein	2.42E-106	*Pseudomonas orientalis*	41.80%	89.60%
WP_108524723	cupin domain-containing protein	7.40E-92	*Pseudomonas sp.*	41.80%	77.28%
WP_104902683	cupin domain-containing protein	5.53E-92	*Pseudomonas sp.*	41.80%	77.28%
WP_008372438	cupin domain-containing protein	2.23E-116	*Pseudomonas sp.*	41.80%	95.12%
WP_041020196	cupin domain-containing protein	1.48E-93	*Pseudomonas sp.*	41.80%	77.71%
WP_052735763	cupin domain-containing protein	5.18E-92	*Pseudomonas veronii*	41.80%	77.28%
WP_052743390	cupin domain-containing protein	1.23E-91	*Pseudomonas veronii*	41.80%	77.28%
WP_079444484	cupin domain-containing protein	5.53E-92	*Pseudomonas veronii*	41.80%	77.28%
WP_102617645	MULTISPECIES: cupin domain-containing protein	6.43E-92	*Pseudomonas*	41.80%	77.28%
KKD01804	cupin	8.26E-101	*Photobacterium halotolerans*	41.70%	90.23%
WP_046218940	cupin domain-containing protein	3.56E-101	*Photobacterium halotolerans*	41.70%	90.23%
WP_044620347	iron-containing redox enzyme family protein	2.31E-102	*Gynuella sunshinyii*	41.70%	84.08%
CRM21916	Cupin domain protein	1.42E-106	*Pseudomonas sp.*	41.60%	89.60%
OCW22127	cupin	1.17E-105	*Pseudomonas sp.*	41.50%	90.02%
WP_012237820	cupin domain-containing protein	3.34E-103	*Sorangium cellulosum*	41.50%	97.66%
SFH27375	Mannose-6-phosphate isomerase, cupin superfamily	6.82E-106	*Pseudomonas sp.*	41.50%	90.02%
WP_024912480	cupin	5.94E-106	*Chania multitudinisentens*	41.40%	93.63%
SCX05785	cupin domain-containing protein	3.92E-105	*Pseudomonas sp.*	41.40%	90.02%
SAM31124	Cupin domain protein	4.28E-105	*Pseudomonas sp.*	41.40%	90.02%
CRM09726	Cupin domain protein	4.71E-105	*Pseudomonas sp.*	41.40%	90.02%
CRM07030	hypothetical protein	1.90E-105	*Pseudomonas sp.*	41.40%	93.63%
SDZ58457	Iron-containing redox enzyme	2.26E-105	*Pseudomonas salomonii*	41.40%	93.63%
WP_100217708	cupin	3.49E-110	*Pseudomonas luteola*	41.30%	93.63%
WP_042133258	cupin	8.85E-97	*Pseudomonas oryzihabitans*	41.30%	84.08%
KTC43957	cupin	8.42E-92	*Pseudomonas fluorescens*	41.20%	77.28%
PIB62865	cupin	2.38E-101	*Pseudomonas sp.*	41.20%	90.02%
SDQ41030	cupin domain-containing protein	3.91E-104	*Pseudomonas grimontii*	41.20%	90.02%
CRM55679	Cupin domain protein	1.50E-104	*Pseudomonas sp.*	41.20%	89.60%
CRM79545	Cupin domain protein	7.74E-105	*Pseudomonas sp.*	41.20%	89.60%
CRN01432	Cupin domain protein	9.41E-105	*Pseudomonas sp.*	41.20%	89.60%
CRM63488	Cupin domain protein	1.03E-104	*Pseudomonas sp.*	41.20%	89.60%
WP_065596487	cupin domain-containing protein	2.46E-108	*Aliivibrio fischeri*	41.10%	90.45%
OYT81457	cupin	4.66E-91	*Pseudomonas sp.*	41.00%	79.83%
SEB30013	cupin domain-containing protein	4.45E-104	*Pseudomonas panacis*	40.90%	90.23%

WP_032804215	cupin domain-containing protein	7.24E-104	*Pseudomonas veronii*	40.90%	90.23%
WP_058760821	cupin	6.54E-92	*Pseudomonas psychrotolerans*	40.80%	84.08%
CDF92291	FIG00980710: hypothetical protein	2.16E-103	*Pseudomonas sp.*	40.80%	90.02%
AJQ97142	hypothetical protein	3.31E-113	*Gynuella sunshinyii*	40.80%	94.48%
AQY67796	cupin	9.28E-104	*Pseudomonas veronii*	40.70%	90.23%
WP_099603120	cupin domain-containing protein	2.53E-90	*Pseudomonas sp.*	40.70%	79.83%
WP_095185645	cupin domain-containing protein	1.24E-90	*Pseudomonas sp.*	40.70%	79.83%
WP_050442844	cupin domain-containing protein	1.43E-90	*Pseudomonas sp.*	40.70%	79.83%
PUB35171	heme oxygenase-like protein	1.10E-103	*Pseudomonas sp.*	40.70%	90.23%
WP_084468209	iron-containing redox enzyme family protein	1.92E-98	*Actinokineospora inagensis*	40.50%	92.99%
WP_046336774	cupin	5.13E-108	*Xenorhabdus bovienii*	40.20%	95.97%
WP_105523059	MULTISPECIES: cupin	7.04E-107	*Pseudomonas*	40.20%	96.60%
PIB52734	cupin	3.57E-100	*Pseudomonas sp.*	40.10%	90.02%
WP_092509873	cupin	3.09E-113	*Xenorhabdus mauleonii*	40.10%	97.66%
EZP65609	Cupin domain protein	1.87E-100	*Pseudomonas sp.*	40.10%	90.02%
WP_038222672	cupin	8.52E-108	*Xenorhabdus bovienii*	40.00%	95.97%
SFS25049	cupin domain-containing protein	1.13E-106	*Pseudomonas sp.*	40.00%	94.69%
WP_099113610	cupin	1.78E-107	*Xenorhabdus miraniensis*	39.60%	95.97%
WP_013334271	cupin domain-containing protein	1.62E-105	*Cyanothece sp.*	39.60%	94.90%
WP_015351680	cupin domain-containing protein	1.17E-87	*Myxococcus stipitatus*	39.30%	89.81%
WP_089327642	iron-containing redox enzyme family protein	1.05E-87	*Actinomadura meyerae*	39.30%	90.02%
WP_071900031	hypothetical protein	4.22E-88	*Cystobacter ferrugineus*	39.20%	89.81%
WP_095849154	iron-containing redox enzyme family protein	1.93E-91	*Gibbsiella quercinecans*	39.00%	84.08%
WP_051861408	cupin	5.44E-103	*Xenorhabdus bovienii*	38.90%	95.33%
CNE62654	Uncharacterised protein	3.81E-92	*Mycobacterium tuberculosis*	38.80%	97.24%
WP_067628022	iron-containing redox enzyme family protein	1.61E-99	*Actinomadura latina*	38.70%	96.39%
WP_061692650	iron-containing redox enzyme family protein	4.78E-96	*Legionella pneumophila*	38.50%	93.63%
WP_080464573	iron-containing redox enzyme family protein, partial	2.60E-95	*Legionella pneumophila*	38.30%	93.63%
WP_051301267	iron-containing redox enzyme family protein	5.96E-83	*Actinomadura rifamycini*	38.30%	92.36%
WP_014842397	iron-containing redox enzyme family protein	4.53E-96	*Legionella pneumophila*	38.30%	93.63%
WP_015961552	iron-containing redox enzyme family protein	5.74E-96	*Legionella pneumophila*	38.30%	93.63%
WP_027220833	iron-containing redox enzyme family protein	6.39E-96	*Legionella pneumophila*	38.30%	93.63%
WP_038839342	iron-containing redox enzyme family protein	8.74E-96	*Legionella pneumophila*	38.30%	93.63%
WP_062740165	iron-containing redox enzyme family protein	2.65E-95	*Legionella pneumophila*	38.30%	93.63%
WP_089310852	iron-containing redox enzyme family protein	1.88E-86	*Actinomadura mexicana*	38.20%	93.63%
WP_089310855	iron-containing redox enzyme family protein	7.46E-94	*Actinomadura mexicana*	38.00%	95.97%
AHH99596	hypothetical protein	4.43E-88	*Kutzneria albida*	37.70%	93.21%
WP_081789611	iron-containing redox enzyme family protein	7.28E-88	*Kutzneria albida*	37.70%	93.21%
WP_052518391	iron-containing redox enzyme family protein	1.37E-81	*Archangium violaceum*	37.60%	90.87%
WP_081971208	iron-containing redox enzyme family protein	8.91E-83	*Dactylosporangium aurantiacum*	37.40%	91.08%
WP_067628030	iron-containing redox enzyme family protein	1.03E-86	*Actinomadura latina*	37.30%	94.90%

WP_044497050	iron-containing redox enzyme family protein	5.67E-97	*Legionella pneumophila*	37.30%	97.24%
WP_067908588	MULTISPECIES: iron-containing redox enzyme family protein	1.85E-82	*Actinomadura*	37.20%	94.69%
WP_058496983	iron-containing redox enzyme family protein	5.50E-92	*Legionella drozanskii*	37.10%	97.24%
WP_089327640	iron-containing redox enzyme family protein	1.05E-93	*Actinomadura meyerae*	36.80%	94.48%
CNE62572	Uncharacterised protein	9.36E-84	*Mycobacterium tuberculosis*	36.70%	94.90%
WP_073559581	iron-containing redox enzyme family protein	2.99E-81	*Archangium sp.*	36.60%	90.87%
WP_051301269	iron-containing redox enzyme family protein	4.07E-86	*Actinomadura rifamycini*	36.20%	95.97%
WP_052387483	iron-containing redox enzyme family protein	1.72E-78	*Dactylosporangium aurantiacum*	35.70%	91.93%
WP_051341416	peptide synthetase	1.37E-52	*Pseudonocardia spinosispora*	29.10%	93.63%

Supplementary Table 3 | SznF homologs containing the central and cupin domains obtained from BLAST searches (IMG/JGI "All Isolate" Database) and used to tabulate phylogenetic distribution (**Extended Data Fig. 10b**)

IMG/JGI Gene ID	Gene Product Name	Organism	Sequence Identity (%, aa)
2660109213	Iron-containing redox enzyme	*Streptomyces sp. di188*	93.4
2660059956	Iron-containing redox enzyme	*Streptomyces sp. di50b*	93.4
2513741850	Iron-containing redox enzyme	*Streptomyces sp. AmelKG-F2B*	93.2
2554257953	Cupin domain-containing protein	*Streptomyces violaceusniger SPC6*	89.2
2521695431	Iron-containing redox enzyme	*Streptomyces sp. 351MFTsu5.1*	88.1
2676412626	heme oxygenase-like protein	*Streptomyces sp. CG 926*	87.7
2766235519	heme oxygenase-like protein	*Streptomyces sp. NRRL F-2580*	86.4
2652743222	Iron-containing redox enzyme	*Streptomyces lydicus A02*	75.1
2638325290	Iron-containing redox enzyme	*Streptomyces natalensis ATCC 27448*	74.3
2735483371	heme oxygenase-like protein	*Actinokineospora cianjurensis DSM 45657*	71.7
2616783557	Iron-containing redox enzyme	*Actinokineospora terrae DSM 44260*	71.1
2524979716	heme oxygenase-like protein	*Kitasatospora sp. SolWspMP-SS2h*	71
2566151749	Iron-containing redox enzyme	*Kutzneria albida DSM 43870*	70.9
2574545549	Iron-containing redox enzyme	*Kitasatospora mediocidica KCTC 9733*	70.8
2513739293	Iron-containing redox enzyme	*Streptomyces sp. AmelKG-F2B*	70.8
2515530376	Iron-containing redox enzyme	*Streptomyces sp. CcalMP-8W*	70.8
2515900704	Iron-containing redox enzyme	*Streptomyces sp. SolWspMP-sol2th*	70.8
2597099370	Iron-containing redox enzyme	*Actinokineospora globicatena DSM 44256*	70.7
2520515127	Iron-containing redox enzyme	*Streptomyces sp. AmelKG-D3*	70.6
2660099747	Iron-containing redox enzyme	*Streptomyces sp. DvalAA-19*	70.6
2753098478	heme oxygenase-like protein	*Saccharothrix sp. ST-888*	67.8
2597098434	Iron-containing redox enzyme	*Actinokineospora globicatena DSM 44256*	66
2735485269	heme oxygenase-like protein	*Actinokineospora cianjurensis DSM 45657*	65.3
2616783061	Iron-containing redox enzyme	*Actinokineospora terrae DSM 44260*	65.3
2596791491	Iron-containing redox enzyme	*Actinokineospora diospyrosa DSM 44255*	65.1
2648247982	Iron-containing redox enzyme	*Luteipulveratus halotolerans C296001*	47.7
649691428	cupin domain-containing protein	*Stigmatella aurantiaca DW4/3-1*	46.6
639859690	cupin domain protein	*Stigmatella aurantiaca DW4/3-1*	46.4
2687102037	Iron-containing redox enzyme	*Bordetella sp. N*	42.5
637965912	Cupin 2 protein	*Chromohalobacter salexigens 1H11, DSM 3043*	42.1
2597304787	Cupin domain-containing protein	*Azospirillum brasilense DSM 1690*	42.1
2553282870	Cupin domain-containing protein	*Azospirillum brasilense FP2*	42.1
2599103261	heme oxygenase-like protein	*Azospirillum brasilense sp7*	42.1
2536470909	Iron-containing redox enzyme	*Pseudomonas syringae pv. avellanae ISPaVe013*	41.5
2536475340	Iron-containing redox enzyme	*Pseudomonas syringae pv. avellanae ISPaVe037*	41.5
638998432	hypothetical protein	*Reinekea blandensis MED297*	41.1
2517867191	Iron-containing redox enzyme	*Chitiniphilus shinanonensis DSM 23277*	40.5
2625014454	Iron-containing redox enzyme	*Bacillus cereus B4087*	40.5
2729121694	Iron-containing redox enzyme	*Bacillus cereus FSL W8-0275*	40.5
2650567063	Iron-containing redox enzyme	*Bacillus cereus HWW 274-2*	40.5

2553013294	Iron-containing redox enzyme	*Bacillus sp. WBUNB001_A*	40.5
2744198776	heme oxygenase-like protein	*Bacillus cereus B4081*	40.2
2744252089	heme oxygenase-like protein	*Bacillus cereus B4120*	40.2
644508593	Cupin domain protein	*Bacillus cereus BDRD-ST24*	40.2
2689008177	Iron-containing redox enzyme	*Bacillus cereus FORC_013*	40.2
2729160631	Iron-containing redox enzyme	*Bacillus cereus FSL W8-0824*	40.2
2729161078	Iron-containing redox enzyme	*Bacillus cereus FSL W8-0824*	40.2
2533210737	Iron-containing redox enzyme	*Bacillus cereus VD196*	40.2
2533229114	Iron-containing redox enzyme	*Bacillus cereus VD200*	40.2
2639856652	Iron-containing redox enzyme	*Bacillus sp. CL115*	40.2
2640250127	Iron-containing redox enzyme	*Bacillus sp. CL25*	40.2
2639862486	Iron-containing redox enzyme	*Bacillus sp. CL96*	40.2
646824459	Cupin domain-containing protein	*Bacillus thuringiensis BMB171*	40.2
2575274618	Iron-containing redox enzyme	*Clostridium drakei SL1*	40.2
2586221513	Iron-containing redox enzyme	*gamma proteobacterium L18*	40.1
2637085032	Iron-containing redox enzyme	*Bacillus cereus LK9*	40
2710916546	Cupin domain-containing protein	*Gammaproteobacteria bacterium RIFCSPHIGHO2_12_FULL_37_14*	39.8
2532511254	Cupin domain-containing protein	*Pseudomonas sp. M47T1*	39.5
2654655309	Cupin domain-containing protein	*Pseudomonas veronii 1YB2*	38.9
2628651429	Cupin domain-containing protein	*Photobacterium halotolerans MELD1*	38.9
2558420645	Iron-containing redox enzyme	*Alteromonadaceae bacterium 1120W.S.0a.04*	38.6
2667984620	heme oxygenase-like protein	*Dongia mobilis CGMCC 1.7660*	38.4
2586184112	Iron-containing redox enzyme	*Pseudomonas oryzihabitans NBRC 102199*	38.2
2667521427	Iron-containing redox enzyme	*Pseudomonas salomonii LMG 22120*	37.9
2685871948	Iron-containing redox enzyme	*Xenorhabdus mauleonii DSM 17908*	37.8
2552016557	Iron-containing redox enzyme	*Pseudomonas luteola XLDN4-9*	37.8
2505604652	Iron-containing redox enzyme	*Pseudomonas syringae PavISPaVe013*	37.8
2505610011	Iron-containing redox enzyme	*Pseudomonas syringae PavISPaVe037*	37.8
2678149665	Iron-containing redox enzyme	*Pseudomonas syringae pv. coryli ICMP 17001*	37.8
2628633274	Iron-containing redox enzyme	*Pseudomonas syringae pv. coryli str. NCPPB 4273*	37.8
2631876890	Iron-containing redox enzyme	*Pseudomonas syringae pv. syringae CRAFRU12*	37.8
2559355874	Iron-containing redox enzyme	*Chania multitudinisentens RB-25*	37.8
2577717775	Iron-containing redox enzyme	*Chania multitudinisentens RB-25*	37.8
2752705853	heme oxygenase-like protein	*Rhizobiales bacterium Alpha_04*	37.5
2656766121	Iron-containing redox enzyme	*Xenorhabdus bovienii CS03*	37.3
648199426	cupin 2 barrel domain-containing protein	*Cyanothece sp. PCC 7822*	37.3
2648838232	Mannose-6-phosphate isomerase, cupin superfamily	*Pseudomonas fluorescens NT0133*	37.2
2640004715	Cupin domain-containing protein	*Pseudomonas marginalis NCPPB 667*	37.2
2664867782	Cupin domain-containing protein	*Pseudomonas orientalis DSM 17489*	37.2
2639741299	Cupin domain-containing protein	*Pseudomonas orientalis LMG 23660*	37.2
2606198233	heme oxygenase-like protein	*Pseudomonas poae A2-S9*	37.2
2738895712	heme oxygenase-like protein	*Pseudomonas sp. GV047*	37.2

2738808352	heme oxygenase-like protein	*Pseudomonas sp. GV087*	37.2
2521890718	Cupin domain-containing protein	*Pseudomonas sp. LAMO17WK12:I1*	37.2
2523544124	Cupin domain-containing protein	*Pseudomonas sp. LAMO17WK12:I2*	37.2
2599878563	Cupin domain-containing protein	*Pseudomonas sp. NFACC25*	37.2
2599503955	Mannose-6-phosphate isomerase, cupin superfamily	*Pseudomonas sp. NFACC45*	37.2
2619069127	Iron-containing redox enzyme	*Xenorhabdus bovienii feltiae Moldova*	37.1
2632359851	Iron-containing redox enzyme	*Gynuella sunshinyii YC6258*	37.1
2580714530	Cupin domain-containing protein	*Pseudomonas sp. CHM02*	36.9
2716163515	Cupin domain-containing protein	*Pseudomonas sp. MIACH*	36.9
2599947766	Cupin domain-containing protein	*Pseudomonas sp. NFACC42-2*	36.9
2733722306	Iron-containing redox enzyme	*Pseudomonas psychrotolerans NS2*	36.8
641349487	Cupin domain hypothetical protein	*Sorangium cellulosum So ce 56*	36.8
2529258052	Iron-containing redox enzyme	*Actinokineospora inagensis DSM 44258*	36.8
2632688552	Mannose-6-phosphate isomerase, cupin superfamily	*Pseudomonas frederiksbergensis SI8*	36.7
2640350741	Cupin domain-containing protein	*Pseudomonas grimontii DSM 17515*	36.7
2630807721	Mannose-6-phosphate isomerase, cupin superfamily	*Pseudomonas sp. SHC52*	36.7
2554657740	Cupin domain-containing protein	*Pseudomonas veronii 1YdBTEX2*	36.7
2554661358	Cupin domain-containing protein	*Pseudomonas veronii 1YdBTEX2*	36.7
2702921226	Cupin domain-containing protein	*Pseudomonas veronii DSM 11331*	36.7
2642479420	Cupin domain-containing protein	*Pseudomonas veronii LMG 17761*	36.7
2703158269	Cupin domain-containing protein	*Pseudomonas fluorescens ABAC62*	36.5
2738749783	heme oxygenase-like protein	*Pseudomonas sp. GV058*	36.5
2757784095	heme oxygenase-like protein	*Pseudomonas sp. GV062*	36.5
2738671389	heme oxygenase-like protein	*Pseudomonas sp. GV077*	36.5
2757646757	heme oxygenase-like protein	*Pseudomonas sp. GV082*	36.5
2757842299	heme oxygenase-like protein	*Pseudomonas sp. GV086*	36.5
2757806997	heme oxygenase-like protein	*Pseudomonas sp. GV101*	36.5
2738858823	heme oxygenase-like protein	*Pseudomonas sp. GV105*	36.5
2511477145	Cupin domain-containing protein	*Pseudomonas sp. HCN1*	36.5
2621298228	Cupin domain-containing protein	*Pseudomonas veronii R4 Genome sequencing*	36.5
2602274084	Iron-containing redox enzyme	*Legionella pneumophila Leg01/16*	36.4
2725410661	Iron-containing redox enzyme	*Legionella pneumophila NY23*	35.9
2619177689	Iron-containing redox enzyme	*Xenorhabdus bovienii kraussei Quebec*	35.9
2580305754	Cupin domain-containing protein	*Pseudomonas sp. RIT357*	35.9
2722797161	Iron-containing redox enzyme	*Legionella pneumophila D-7630*	35.9
2719574304	Iron-containing redox enzyme	*Legionella pneumophila D-7631*	35.9
2719938967	Iron-containing redox enzyme	*Legionella pneumophila D-7632*	35.9
2702157734	Iron-containing redox enzyme	*Mycobacterium tuberculosis 401416*	35.8
2657780125	Iron-containing redox enzyme	*Legionella drozanskii ATCC 700990*	35.6
2571123665	Iron-containing redox enzyme	*Actinomadura glauciflava DSM 44770*	35.6
2732249157	Iron-containing redox enzyme	*Actinomadura latina NBRC 106108*	35.6
2674046301	Iron-containing redox enzyme	*Legionella pneumophila EUL25*	35.4

2703794573	Iron-containing redox enzyme	*Legionella pneumophila EUL48*	35.4
2703814023	Iron-containing redox enzyme	*Legionella pneumophila EUL56*	35.4
2664673310	Iron-containing redox enzyme	*Legionella pneumophila HO4020049*	35.4
2720183559	Iron-containing redox enzyme	*Legionella pneumophila pneumophila FFI104*	35.4
2721650733	Iron-containing redox enzyme	*Legionella pneumophila pneumophila FFI105*	35.4
2720180486	Iron-containing redox enzyme	*Legionella pneumophila pneumophila FFI337*	35.4
2659631767	Iron-containing redox enzyme	*Legionella pneumophila 12_3965*	35.2
2545538199	Iron-containing redox enzyme	*Legionella pneumophila 121004*	35.2
2707405485	Iron-containing redox enzyme	*Legionella pneumophila Bnt314*	35.2
2719571200	Iron-containing redox enzyme	*Legionella pneumophila C9_S*	35.2
2706865242	Iron-containing redox enzyme	*Legionella pneumophila EUL1*	35.2
2706881304	Iron-containing redox enzyme	*Legionella pneumophila EUL10*	35.2
2706901759	Iron-containing redox enzyme	*Legionella pneumophila EUL100*	35.2
2706895429	Iron-containing redox enzyme	*Legionella pneumophila EUL101*	35.2
2706898743	Iron-containing redox enzyme	*Legionella pneumophila EUL102*	35.2
2706903562	Iron-containing redox enzyme	*Legionella pneumophila EUL103*	35.2
2699628372	Iron-containing redox enzyme	*Legionella pneumophila EUL104*	35.2
2706908442	Iron-containing redox enzyme	*Legionella pneumophila EUL105*	35.2
2703885378	Iron-containing redox enzyme	*Legionella pneumophila EUL109*	35.2
2672904998	Iron-containing redox enzyme	*Legionella pneumophila EUL110*	35.2
2673802143	Iron-containing redox enzyme	*Legionella pneumophila EUL114*	35.2
2675868303	Iron-containing redox enzyme	*Legionella pneumophila EUL116*	35.2
2706918199	Iron-containing redox enzyme	*Legionella pneumophila EUL117*	35.2
2706921942	Iron-containing redox enzyme	*Legionella pneumophila EUL119*	35.2
2674934680	Iron-containing redox enzyme	*Legionella pneumophila EUL120*	35.2
2706926825	Iron-containing redox enzyme	*Legionella pneumophila EUL121*	35.2
2699953112	Iron-containing redox enzyme	*Legionella pneumophila EUL122*	35.2
2674453521	Iron-containing redox enzyme	*Legionella pneumophila EUL123*	35.2
2699612183	Iron-containing redox enzyme	*Legionella pneumophila EUL124*	35.2
2706963325	Iron-containing redox enzyme	*Legionella pneumophila EUL125*	35.2
2699031318	Iron-containing redox enzyme	*Legionella pneumophila EUL13*	35.2
2699977704	Iron-containing redox enzyme	*Legionella pneumophila EUL14*	35.2
2703778694	Iron-containing redox enzyme	*Legionella pneumophila EUL16*	35.2
2677338664	Iron-containing redox enzyme	*Legionella pneumophila EUL168*	35.2
2706966745	Iron-containing redox enzyme	*Legionella pneumophila EUL169*	35.2
2698906298	Iron-containing redox enzyme	*Legionella pneumophila EUL17*	35.2
2706970016	Iron-containing redox enzyme	*Legionella pneumophila EUL170*	35.2
2706972975	Iron-containing redox enzyme	*Legionella pneumophila EUL27*	35.2
2706862435	Iron-containing redox enzyme	*Legionella pneumophila EUL3*	35.2
2706873385	Iron-containing redox enzyme	*Legionella pneumophila EUL31*	35.2
2677812415	Iron-containing redox enzyme	*Legionella pneumophila EUL37*	35.2
2673378469	Iron-containing redox enzyme	*Legionella pneumophila EUL38*	35.2

2675034746	Iron-containing redox enzyme	*Legionella pneumophila EUL39*	35.2
2675533128	Iron-containing redox enzyme	*Legionella pneumophila EUL42*	35.2
2677787209	Iron-containing redox enzyme	*Legionella pneumophila EUL43*	35.2
2703800833	Iron-containing redox enzyme	*Legionella pneumophila EUL44*	35.2
2678616275	Iron-containing redox enzyme	*Legionella pneumophila EUL45*	35.2
2672829520	Iron-containing redox enzyme	*Legionella pneumophila EUL46*	35.2
2699426781	Iron-containing redox enzyme	*Legionella pneumophila EUL50*	35.2
2703807944	Iron-containing redox enzyme	*Legionella pneumophila EUL53*	35.2
2678909702	Iron-containing redox enzyme	*Legionella pneumophila EUL54*	35.2
2703810865	Iron-containing redox enzyme	*Legionella pneumophila EUL55*	35.2
2699960442	Iron-containing redox enzyme	*Legionella pneumophila EUL57*	35.2
2703820425	Iron-containing redox enzyme	*Legionella pneumophila EUL58*	35.2
2703784593	Iron-containing redox enzyme	*Legionella pneumophila EUL6*	35.2
2699444267	Iron-containing redox enzyme	*Legionella pneumophila EUL60*	35.2
2703830291	Iron-containing redox enzyme	*Legionella pneumophila EUL67*	35.2
2703834581	Iron-containing redox enzyme	*Legionella pneumophila EUL68*	35.2
2706852316	Iron-containing redox enzyme	*Legionella pneumophila EUL7*	35.2
2703840843	Iron-containing redox enzyme	*Legionella pneumophila EUL70*	35.2
2672572768	Iron-containing redox enzyme	*Legionella pneumophila EUL71*	35.2
2683360059	Iron-containing redox enzyme	*Legionella pneumophila EUL75*	35.2
2699601987	Iron-containing redox enzyme	*Legionella pneumophila EUL76*	35.2
2699707876	Iron-containing redox enzyme	*Legionella pneumophila EUL77*	35.2
2677382728	Iron-containing redox enzyme	*Legionella pneumophila EUL82*	35.2
2703858034	Iron-containing redox enzyme	*Legionella pneumophila EUL84*	35.2
2703861665	Iron-containing redox enzyme	*Legionella pneumophila EUL85*	35.2
2703869185	Iron-containing redox enzyme	*Legionella pneumophila EUL86*	35.2
2699142887	Iron-containing redox enzyme	*Legionella pneumophila EUL88*	35.2
2699373899	Iron-containing redox enzyme	*Legionella pneumophila EUL9*	35.2
2703870973	Iron-containing redox enzyme	*Legionella pneumophila EUL93*	35.2
2703874626	Iron-containing redox enzyme	*Legionella pneumophila EUL94*	35.2
2674382774	Iron-containing redox enzyme	*Legionella pneumophila EUL95*	35.2
2602364437	Iron-containing redox enzyme	*Legionella pneumophila Leg01/11*	35.2
2602360015	Iron-containing redox enzyme	*Legionella pneumophila Leg01/20*	35.2
2602353642	Iron-containing redox enzyme	*Legionella pneumophila Leg01/53*	35.2
2706981620	Iron-containing redox enzyme	*Legionella pneumophila LP_617*	35.2
2678672708	Iron-containing redox enzyme	*Legionella pneumophila NY1*	35.2
2725380601	Iron-containing redox enzyme	*Legionella pneumophila NY10*	35.2
2725385934	Iron-containing redox enzyme	*Legionella pneumophila NY12*	35.2
2678590683	Iron-containing redox enzyme	*Legionella pneumophila NY14*	35.2
2725393978	Iron-containing redox enzyme	*Legionella pneumophila NY15*	35.2
2725396252	Iron-containing redox enzyme	*Legionella pneumophila NY16*	35.2
2725359458	Iron-containing redox enzyme	*Legionella pneumophila NY2*	35.2
2725413209	Iron-containing redox enzyme	*Legionella pneumophila NY24*	35.2

2725360481	Iron-containing redox enzyme	*Legionella pneumophila NY3*	35.2
2729074104	Iron-containing redox enzyme	*Legionella pneumophila NY4*	35.2
2725377925	Iron-containing redox enzyme	*Legionella pneumophila NY9*	35.2
2672828332	Iron-containing redox enzyme	*Legionella pneumophila Ofk308*	35.2
2719841545	Iron-containing redox enzyme	*Legionella pneumophila OLDA*	35.2
2703887150	Iron-containing redox enzyme	*Legionella pneumophila OLDA_1*	35.2
637581130	hypothetical protein	*Legionella pneumophila Paris*	35.2
2668790292	Iron-containing redox enzyme	*Legionella pneumophila pneumophila ATCC 35096*	35.2
2670362061	Iron-containing redox enzyme	*Legionella pneumophila pneumophila ATCC 35289*	35.2
2562283913	Iron-containing redox enzyme	*Legionella pneumophila pneumophila Lorraine*	35.2
2719850951	Iron-containing redox enzyme	*Legionella pneumophila Pontiac*	35.2
2706975887	Iron-containing redox enzyme	*Legionella pneumophila Pontiac-1*	35.2
2619014694	Iron-containing redox enzyme	*Legionella pneumophila serogroup 1 Quebec*	35.2
650624580	hypothetical protein	*Legionella pneumophila sv. 1 130b*	35.2
2582394549	Iron-containing redox enzyme	*Legionella pneumophila sv. 1 TUM13948*	35.2
2579962445	Iron-containing redox enzyme	*Legionella pneumophila W1046*	35.2
2699470071	Iron-containing redox enzyme	*Legionella pneumophila Ymg289*	35.2
2727833483	Iron-containing redox enzyme	*Actinomadura mexicana DSM 44485*	34.5
2732265394	Iron-containing redox enzyme	*Actinomadura rubrobrunea NBRC 15275*	34.3
2730677285	heme oxygenase-like protein	*Actinomadura viridilutea DSM 44433*	34.3
2732249160	Iron-containing redox enzyme	*Actinomadura latina NBRC 106108*	34
2523680377	Iron-containing redox enzyme	*Actinomadura rifamycini DSM 43936*	33.9
2521988398	Iron-containing redox enzyme	*Myxococcus stipitatus DSM 14675*	33.8
2571123668	Iron-containing redox enzyme	*Actinomadura glauciflava DSM 44770*	33.7
2566155247	Iron-containing redox enzyme	*Kutzneria albida DSM 43870*	33.6
2732265396	Rieske [2Fe-2S] domain-containing protein	*Actinomadura rubrobrunea NBRC 15275*	33.6
2730677287	Rieske-like 2Fe-2S protein	*Actinomadura viridilutea DSM 44433*	33.6
2727833480	Iron-containing redox enzyme	*Actinomadura mexicana DSM 44485*	33.5
2702157731	Iron-containing redox enzyme	*Mycobacterium tuberculosis 401416*	33.3
2523680380	Iron-containing redox enzyme	*Actinomadura rifamycini DSM 43936*	33.1
2525751418	hypothetical protein	*Pseudonocardia spinosispora DSM 44797*	27
2519022431	hypothetical protein	*Streptomyces sulphureus DSM 40104*	25.9
2664127102	hypothetical protein	*Lentzea flaviverrucosa CGMCC 4.578*	25.6
2668006505	hypothetical protein	*Streptomyces alni CGMCC 4.3510*	25.4
2646533367	hypothetical protein	*Kibdelosporangium sp. MJ126-NF4*	25.1
2656756700	hypothetical protein	*Streptomyces anulatus ATCC 11523*	24.6
2633981733	hypothetical protein	*Streptomyces sp. MNU77*	24.6
2644405294	hypothetical protein	*Streptomyces sp. Root1295*	24.6
2643900484	hypothetical protein	*Streptomyces sp. Root63*	24.6
2558994424	hypothetical protein	*Micromonosporaceae bacterium URHE0070*	24.2
2517759022	hypothetical protein	*Frankia sp. DC12*	24
2626638909	amino acid adenylation domain-containing protein	*Frankia sp AvcI.1*	24

2689971921	hypothetical protein	*Frankia sp. EUN1h*	23.9
638100845	Putative non-ribosomal peptide synthetase	*Frankia alni ACN14a*	23.8
2508674662	Non-ribosomal peptide synthetase modules and related proteins	*Frankia sp. CN3*	23.6
2712186985	hypothetical protein	*Kitasatospora sp. MBT63*	23.5
2689960864	hypothetical protein	*Frankia sp. BMG5.36*	23.3
2619857042	hypothetical protein	*Frankia sp. ACN1ag*	23.1
2686891540	hypothetical protein	*Pandoraea norimbergensis DSM 11628*	23.1
2620352777	hypothetical protein	*Frankia sp. CpI1-P*	22.9
2579856648	hypothetical protein	*Frankia sp. CpI1-S*	22.9
2554001268	hypothetical protein	*Nocardiopsis baichengensis YIM 90130*	22.9
2656930539	hypothetical protein	*Paraburkholderia sediminicola LMG 24238*	22.7
2656937390	hypothetical protein	*Paraburkholderia terricola LMG 20594*	22.7
2754959761	hypothetical protein	*Burkholderia sp. BDU5*	22.6
2754846917	hypothetical protein	*Burkholderia sp. BDU6*	22.6
649751577	non-ribosomal peptide synthetase	*Frankia inefficax EuI1c*	22.6
2682362425	hypothetical protein	*Herbaspirillum rubrisubalbicans NBRC 102523*	22.3
2511247803	hypothetical protein	*Herbaspirillum sp. CF444*	22.2
2586329116	hypothetical protein	*Robbsia andropogonis Ba3549*	22.2
2629174494	hypothetical protein	*Robbsia andropogonis ICMP2807*	22.2
2653812110	hypothetical protein	*Burkholderia palmae JS23*	22.2
2655479416	hypothetical protein	*Herbaspirillum rhizosphaerae UMS-37*	22.1
2728394001	hypothetical protein	*Acetobacter senegalensis LMG 23690*	22
2688770459	hypothetical protein	*Burkholderia sp. HB1*	21.8
2509122343	hypothetical protein	*Paraburkholderia dilworthii WSM3556*	21.8
642425482	putative non-ribosomal peptide synthetase	*Paraburkholderia graminis C4D1M*	21.8
2738878383	hypothetical protein	*Paraburkholderia sp. GV052*	21.8
2739224976	hypothetical protein	*Paraburkholderia sp. GV060*	21.8
2738836852	hypothetical protein	*Paraburkholderia sp. GV068*	21.8
2739190075	hypothetical protein	*Paraburkholderia sp. GV072*	21.8
2738824443	hypothetical protein	*Paraburkholderia sp. GV073*	21.8
2687101073	hypothetical protein	*Bordetella sp. N*	21.8
2546377957	non-ribosomal peptide synthetase	*Amycolatopsis vancoresmycina DSM 44592*	21.8
2710236110	hypothetical protein	*Paraburkholderia bryophila LMG 23644*	21.6
2671603205	hypothetical protein	*Paraburkholderia megapolitana LMG 23650*	21.6
2653831125	hypothetical protein	*Paraburkholderia caballeronis LMG 26416*	21.6
2766825189	hypothetical protein	*Paraburkholderia caballeronis NZu-6*	21.6
2766764778	hypothetical protein	*Paraburkholderia caballeronis TNe-835*	21.6
2694833467	hypothetical protein	*Paraburkholderia caballeronis TNe-841t*	21.6
2766730814	hypothetical protein	*Paraburkholderia caballeronis TNe-878*	21.6
2617851852	hypothetical protein	*Candidimonas bauzanensis CGMCC 1.10190*	21.5
2580609262	hypothetical protein	*Paraburkholderia bannensis NBRC 103871*	21.5
2519046361	hypothetical protein	*Streptomyces sp. AmelKG-E11A*	21.4

2527080764	hypothetical protein	*Burkholderia sp. JPY366*	21.4
2522930572	hypothetical protein	*Burkholderia sp. URHA0054*	21.4
2611856519	hypothetical protein	*Caballeronia sordidicola S170*	21.4
2554977344	hypothetical protein	*Paraburkholderia phenoliruptrix AC1100*	21.4
2766751987	hypothetical protein	*Paraburkholderia caballeronis TNe-834*	21.4
2766741361	hypothetical protein	*Paraburkholderia caballeronis TNe-8641*	21.4
2766754826	hypothetical protein	*Paraburkholderia caballeronis TNe-8682*	21.4
2766739603	hypothetical protein	*Paraburkholderia caballeronis TNe-8683*	21.4
2690674141	AraC-like ligand binding domain-containing protein	*Paraburkholderia fungorum GAS106B*	21.4
2518901180	hypothetical protein	*Paraburkholderia phenoliruptrix BR3459*	21.3
2519490135	putative non-ribosomal peptide synthetase	*Tistrella mobilis KA081020-065*	21.3
2744700014	hypothetical protein	*Tistrella mobilis MCCC 1A02139*	21.3
2583267073	hypothetical protein	*Paraburkholderia ferrariae NBRC 106233*	21.3
2546981049	hypothetical protein	*Burkholderia MF376*	21.2
2521673915	hypothetical protein	*Paraburkholderia bryophila 376MFSha3.1*	21.2
2766716536	hypothetical protein	*Paraburkholderia tropica SIr-6529*	21.2
2766771005	hypothetical protein	*Paraburkholderia tropica SIr-6563*	21.2
2671593524	hypothetical protein	*Paraburkholderia aspalathi LMG 27731*	21.2
2710166187	hypothetical protein	*Paraburkholderia eburnea JCM 18070*	21.2
2710159837	hypothetical protein	*Paraburkholderia eburnea LMG 29537*	21.2
2553403762	hypothetical protein	*Paraburkholderia kururiensis M130*	21.1
2600813973	hypothetical protein	*Paraburkholderia kururiensis thiooxydans NBRC 107107*	21.1
2730309699	hypothetical protein	*Bordetella flabilis AU10664*	21.1
2566169591	hypothetical protein	*Burkholderia glumae 336gr-1*	21
2568571001	hypothetical protein	*Burkholderia glumae 336gr-1*	21
2648806547	hypothetical protein	*Burkholderia glumae ATCC 33617*	21
643877861	Cupin 2 protein	*Burkholderia glumae BGR1*	21
2582266435	hypothetical protein	*Paraburkholderia caledonica NBRC 102488*	21
2733418605	hypothetical protein	*Paraburkholderia kirstenboschensis RAU2d2*	21
2656188737	hypothetical protein	*Paraburkholderia tropica LMG 22274*	21
2758089148	hypothetical protein	*Paraburkholderia sp. PDC91*	20.8
2583068335	hypothetical protein	*Burkholderia gladioli pv. alliicola*	20.8
648207443	cupin 2 protein	*Burkholderia sp. CCGE1003*	20.8
2509131496	hypothetical protein	*Burkholderia sp. WSM2232*	20.8
2757734790	hypothetical protein	*Paraburkholderia sp. OV446*	20.8
2765693608	hypothetical protein	*Bordetella genomosp. 10 AU16122*	20.7
2553308861	hypothetical protein	*Burkholderia gladioli 3848s-5*	20.6
650774265	Cupin 2 protein	*Burkholderia gladioli BSR3*	20.6
2654669209	hypothetical protein	*Burkholderia gladioli NGJ1*	20.6
2629981428	hypothetical protein	*Burkholderia gladioli SN82F6*	20.6
2651297271	hypothetical protein	*Burkholderia gladioli UCD-UG_CHAPALOTE*	20.6
2583099876	hypothetical protein	*Burkholderia sp. 9120*	20.6

2709116570	hypothetical protein	*Paraburkholderia caryophylli Ballard 720*	20.5
2641525822	hypothetical protein	*Burkholderia gladioli ATCC 10248*	20.4
2618422036	hypothetical protein	*Burkholderia gladioli ATCC 25417*	20.4
2600761664	hypothetical protein	*Burkholderia gladioli NBRC 13700*	20.4
2568581744	hypothetical protein	*Burkholderia glumae PG1*	20.4
2689304171	hypothetical protein	*Burkholderia plantarii ATCC 43733*	20.4
2577228585	hypothetical protein	*Burkholderia sp. A1*	20.4
2710177026	hypothetical protein	*Paraburkholderia sp. BL21I4N1*	20.4
2732338316	hypothetical protein	*Millisia brevis NBRC 105863*	20.2
2513963676	hypothetical protein	*Burkholderia sp. WSM2230*	20.2
2695457666	hypothetical protein	*Paraburkholderia phenazinium GAS95*	20.2
2698503406	hypothetical protein	*Paraburkholderia phenazinium GAS86*	20
2654005411	hypothetical protein	*Paraburkholderia phenazinium LMG 2247*	20
2715154792	hypothetical protein	*Caballeronia sordidicola LMG 22029*	19.8
2565385086	hypothetical protein	*Nocardia jiangxiensis NBRC 101359*	19.3
2733021929	hypothetical protein	*Nocardia miyunensis NBRC 108239*	19.1
2682380164	hypothetical protein	*Nocardia vaccinii NBRC 15922*	18.9

Supplementary Table 4 | Primers used for cloning for protein expression and verifying genotype of various *szn* mutants

Oligo	Nucleotide Sequence (5' to 3')	Targeted gene
sznE-F	TATCATATGCGGCACGTACAAGAGGCACG	*sznE*
sznE-R	ATAAAGCTTTCAGATGCTCACGCGGAGGG	*sznE*
sznF-F	TATCATATGAGTCACGTCCCCCCGCACGT	*sznF*
sznF-R	ATAAAGCTTCTACAGGCACTTTCGGTAGT	*sznF*
sznH-F	TATCATATGGGCAGGGACGGGGACAAGCC	*sznH*
sznH-R	ATAAAGCTTTCATCGGGGCGGCTCCTGCG	*sznH*
sznJ-F	TATCATATGGCCTACGCAGAAGGCAT	*sznJ*
sznJ-R	ATAAAGCTTTCAGCGCCCGCCATCGGCGG	*sznJ*
sznK-F	TATCATATGGCGGGCGCTGACCCGGG	*sznK*
sznK-R	ATAAAGCTTTCATGGCCCGGTGATGACCA	*sznK*
pET29b-sznF-F	ACGAAT**CATATG**AGTCACGTCCCCCC	
pET29b-sznF-R	GCGTCTG**AATTCC**TTACAGGCACTTTCGGTAGTCCC	
pIJ773-F	AATGCAGCTCACGGTAACTG	
pIJ773-R	TCCAAGTGGCCCATCTTCGA	

Supplementary Table 5 | Primers used for generating knockout cassettes from pIJ773

Oligo	Nucleotide Sequence (5' to 3')	Targeted gene
sznEKO-F	CGGCGCTGTTGCAGTCGCTCGCGCAGGAGATGAACGCGCATTCCGGGGATCCGTCGACC	*sznE*
sznEKO-R	TGCATCCAGTGCGAGCGGGGGTTGTCCGGCGGGTTGGACTGTAGGCTGGAGCTGCTTC	*sznE*
sznFKO-F	TCGACTGGCTGAACCACGACAACCCCTACCGCCGGCAGCATTCCGGGGATCCGTCGACC	*sznF*
sznFKO-R	CCCTCGCCGGCCTGGAGGGTCAGCGAGGAGCCCAGGCCG TGTAGGCTGGAGCTGCTTC	*sznF*
sznHKO-F	TCTGGCTGCTCCAGCCGGCACCGGTGACCTGGGAGGAGCATTCCGGGGATCCGTCGACC	*sznH*
sznHKO-R	ATGACCCGGGCGTACCGGGAGATGTAGCCGCGCGGCGGCTGTAGGCTGGAGCTGCTTC	*sznH*
sznJKO-F	TACGACTCGCTCACCGAGGTCCTGCCCATCCCCGTCCCCATTCCGGGGATCCGTCGACC	*sznJ*
sznJKO-R	GCGCACCGGGGCGTGCACCATCCCCTCCTCGCGCACATATGTAGGCTGGAGCTGCTTC	*sznJ*
sznKKO-F	CGGGGCGTCCCGGTCGCGGCGGTGGTGAGCAGTCCGCGCATTCCGGGGATCCGTCGACC	*sznK*
sznKKO-R	GAAGTCCAGGCTGTCGACGAGGTCCTTGGTGACGGGGACTGTAGGCTGGAGCTGCTTC	*sznK*

Supplementary Table 6. Primers used for generating SznF variants

Oligo	Nucleotide Sequence (5' to 3')
E215A-F	CATCGAC**GCC**TACGGCTACGGCGTCCACGACACC
E215A-R	CCGTA**GGC**GTCGATGACGACCTTGAACCACTCGG
H225A-F	GTCCACGACACCAAG**GCC**AGCACCCTGTTCGAG
H225A-R	CTCGAACAGGGTGCT**GGC**CTTGGTGTCGTGGAC
E281A-F	GCGCTGTACTACACG**GCC**AGCTCCCTGGTCGAC
E281A-R	GTCGACCAGGGAGCT**GGC**CGTGTAGTACAGCGC
H311A-F	CACCTACTTCACCGAG**GCC**ATCCACATCGACCAG
H311A-R	CTGGTCGATGTGGAT**GGC**CTCGGTGAAGTAGGTG
D315A-F	CATC**GCC**CAGCACCACGGCCGGATGGCCCGCG
D315A-R	GCTG**GGC**GATGTGGATGTGCTCGGTGAAGTAGG
H318A-F	GCAC**GCC**GGCCGGATGGCCCGCGAGAAGATCATC
H318A-R	CGGCC**GGC**GTGCTGGTCGATGTGGATGTGCTCG
H407A-F,	GAGCTGTCCAACACC**GCC**TGCCACGACGGTGAC
H407A-R	GTCACCGTCGTGGCA**GGC**GGTGTTGGACAGCTC
H409A-F	GTCCAACACCCACTGC**GCC**GACGGTGACGAGCTG
H409A-R	CAGCTCGTCACCGTC**GGC**GCAGTGGGTGTTGGAC
H448A-F	CAAGCGCAACCGGCTG**GCC**GGCGCGAACATCGAG
H448A-R	CTCGATGTTCGCGCC**GGC**CAGCCGGTTGCGCTTG
H408A_triple_F1	CTGC**GCC**GACGGTGACGAGCTGTGCCACATCG
H408A_triple_R1	CGTC**GGC**GCAGGCGGTGTTGGACAGCTCGCCG

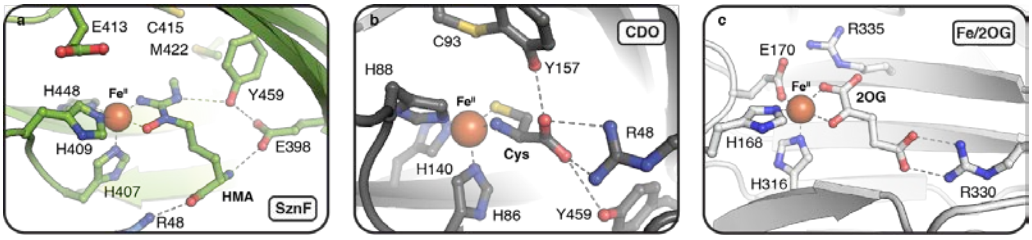
Supplementary Table 7 | Data collection and refinement statistics for SznF x-ray structure determination

	SeMet apo His$_6$-SznF	**apo SznF (Fe ano)**	**HMA•Fe•SznF**	**HMA•Fe•SznF (Fe ano)**
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	61.50, 109.46, 365.98	57.90, 107.73, 150.19	57.95, 107.73, 150.54	57.95, 107.76, 150.65
Resolution (Å)	50.00-2.08 (2.20-2.08)	50.00-2.34 (2.48-2.34)	50.00-1.69 (1.79-1.69)	50.00-2.67 (2.74-2.67)
R_{merge}	0.092 (0.991)	0.089 (0.313)	0.09 (0.952)	0.139 (0.959)
$I / \sigma I$	17.4 (1.9)	11.8 (2.9)	14.3 (2.3)	10.6 (1.9)
CC(1/2)	1.00 (0.73)	1.00 (0.90)	1.00 (0.82)	1.00 (0.73)
Completeness (%)	100 (99.1)	97.3 (84.8)	100 (99.7)	100 (100)
Redundancy	10.1 (7.8)	3.8 (2.3)	8.2 (8.2)	7.8 (7.8)
Refinement				
Resolution (Å)	50.00-2.08		50.00-1.69	
No. reflections	141876		101695	
R_{work} / R_{free}	0.1755/0.1879		0.1757/0.1950	
No. atoms	16328		8548	
Protein	15061		7538	
Ligand/ion	65		30	
Water	1202		980	
B-factors				
Protein	32.02		20.59	
Ligand/ion	58.56		28.95	
Water	48.46		40.29	
R.m.s. deviations				
Bond lengths (Å)	0.007		0.007	
Bond angles (°)	1.223		1.224	
PDB accession code	6M9S		6M9R	

*Values in parentheses are for highest-resolution shell.

3. Supplementary Discussion

Crystallographic characterization of the **1**•$Fe^{(II)}$•SznF complex in the cupin domain active site revealed an unusual bidentate coordination mode for the modified guanidine functional group. We also find that certain mechanistic features of the SznF-catalyzed oxidative rearrangement resemble reactions performed by characterized oxygenases. A lack of a requirement for either an additional co-substrate or source of external electrons is observed for SznF and other cupin enzymes, including cysteine dioxygenase (CDO)[62,63]. CDO shares the unusual tri-His Fe^{II}-binding site with SznF but transfers both oxygen atoms of O_2 to the thiol of a coordinated L-cysteine substrate (**Supplementary Figure 12**). The oxidative rearrangement catalyzed by the SznF cupin domain instead involves transfer of a single oxygen atom to the distal C^{ε} of **2**, similar to Fe^{II}/2-(oxo)-glutarate (Fe/2OG)-dependent oxygenases[64]. These structural comparisons, together with the observations that the last step of the SznF reaction does not require input of reducing equivalents and that all of the guanidine nitrogen atoms from L-NMA are retained in the *N*-nitrosourea product **3**, provide a basis to formulate two mechanistic hypotheses for the novel N–N bond formation reaction performed by the enzyme.



Supplementary Figure 12 | Comparison of the active sites of the cupin domain of SznF (**a**) to cysteine dioxygenase (CDO) (**b**) and a representative Fe/2OG oxygenase, VioC (**c**). The active site of human cysteine dioxygenase (CDO) contains a similar tri-His-coordinated Fe^{II} center with a bidentate amino acid substrate ligand (PDB: ~~accession code~~ 2IC1)[63]. Interestingly, CDO also conserves a second-sphere Cys-Tyr motif, used in this system as a post-translationally crosslinked cofactor that enhances turnover. We do not observe evidence of cross-link formation in SznF. Despite the structural similarities, the chemistry facilitated by these two enzymes is likely quite different as preliminary analysis suggests SznF operates as a monooxygenase (**Figure 2, Extended Data Figure 6**). The observed SznF activity is perhaps more similar to oxo-acid cosubstrate processing in Fe^{II}/2-oxo-glutarate (2OG) oxygenases (PDB: ~~accession code~~ 6ALM). These enzymes use a slightly different metal binding motif (H-X-D/E-X_n-H) to fragment 2OG to succinate and CO_2. This transformation results in incorporation of a single dioxygen-derived O atom into succinate and formation of an Fe^{IV}=O (ferryl) intermediate.

*Possible binding mode of intermediate **2** to SznF*

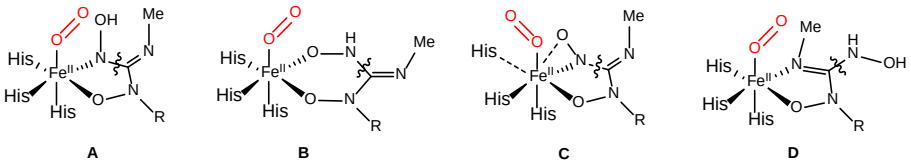
We do not have an experimentally determined view of the true rearrangement substrate, **2**, bound to the cupin $Fe^{(II)}$ cofactor. However, analog studies and *in vitro* biochemical characterization support N^{δ}-OH chelation of the iron center (**Extended Data Figure 9c**), as observed crystallographically in **1**•$Fe^{(II)}$• SznF. Shown below are several possible orientations in

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

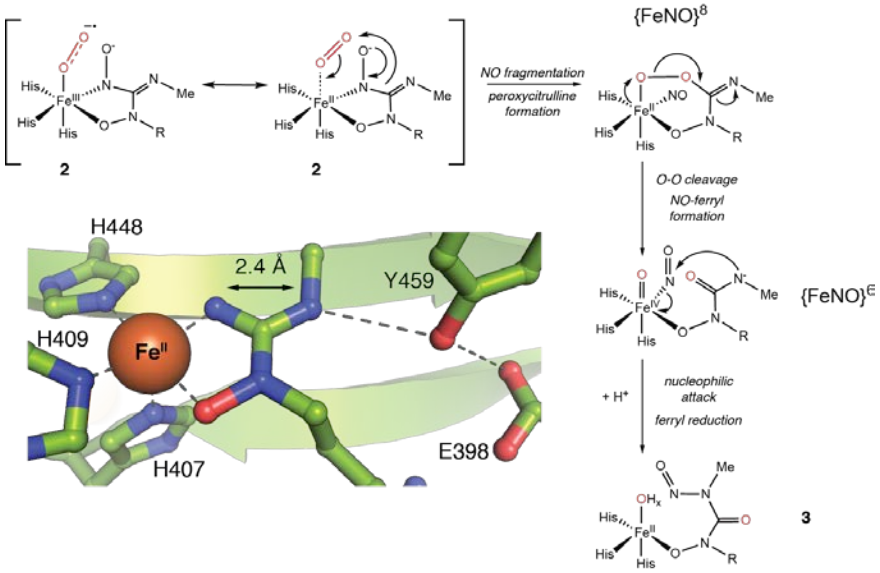
which **2** could bind (**Supplementary Fig. 13**). The C–N bond cleaved during the reaction is indicated by a wavy line.



Supplementary Figure 13 | Possible binding modes of amino acid **2** to the cupin iron cofactor.

In the proposals below, we have used orientation A as the starting point because it yields a five-membered chelate similar to reactant complexes in iron(II)- and 2-(oxo)-glutarate (Fe/2OG) oxygenases[64] and extradiol ring-cleaving Fe^{II} dioxygenases[65]. In orientation B, the C^{ε} of **2** could be too far from the distal oxygen atom of the metal-O_2 adduct for the attack required to form the citrulline product. Another possibility is a side-on coordination of the hydroxylamine moiety of **2** to the metal center (orientation C)[66]. This binding mode would require a dissociation of one of the histidine ligands. Orientation D is disfavored because the methylated N^{ω} of **2** acts as a nucleophile in both mechanistic routes, and it will be more reactive if it is not coordinated to the Fe^{II} cofactor.

Mechanistic proposal 1:



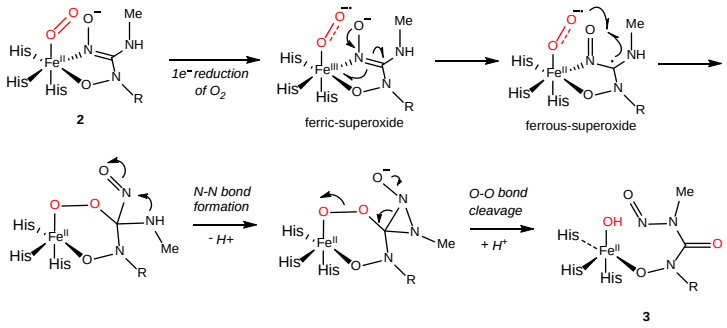
Supplementary Figure 14 | A mechanistic proposal for oxidative rearrangement of 2 involving generation of transient {FeNO}[6] complex. Oxygen addition to the SznF cupin Fe^{II} site induces fragmentation of **2** to yield an

{FeNO}^{8} complex that drives subsequent formation of a peroxide adduct (top path). Cleavage of the O-O bond would generate an NO-bound ferryl intermediate and an amidate nucleophile (right path). Attack by the latter group upon the {FeNO}^{6} is the proposed route for N–N bond formation. The inset shows second-sphere groups that could facilitate proton transfers and the proximity of the N-atoms involved in the product N–N bond.

We propose initial Fe^{II} coordination of substrate **2** (via mode **A**) in a fully deprotonated state to achieve charge balance with a neutral protein ligand set (**Supplementary Fig. 14**). Loss of the proton from the methylated N^{ω} atom may be facilitated by the Y459/E398 pair in the second-sphere of the SznF cupin domain active site. The initial stages of this proposed mechanism resemble the first steps in well-established scission of 2-(oxo)-glutarate to a ferryl-succinate complex and CO_2 by mononuclear iron enzymes in the Fe/2OG superfamily[64,67]. We propose that Fe^{II}-binding of **2** triggers addition of O_2 to yield an end-on six-coordinate adduct. Fragmentation of **2** at the N^{ω}-C^{ε} bond generates both an {FeNO}^{8} complex and a cyclic peroxycitrulline intermediate. The latter complex is analogous to the peroxysuccinate intermediate proposed computationally[67] and detected crystallographically[68] in Fe/2OG enzymes. In those systems, O–O bond cleavage yields the first accumulating intermediate in solution, a high-spin Fe^{IV}=O complex[64]. In SznF, we propose that a similar transformation generates a distinctive anionic NO coordinated ferryl, defined here in Enemark-Feltham notation[69] as an {FeNO}^{6} complex. Formation of the carbonyl at C^{ε} also yields an amidate nucleophile at Me-N^{ω}.

At this stage, the SznF reaction pathway may diverge from that of Fe/2OG enzymes. We propose that that the Me-N^{ω} amidate nitrogen attacks the coordinated nitrogen of the {FeNO}^{6} unit in SznF. Importantly, this nitrosyl transfer event returns the catalyst to the Fe^{II} state, consistent with experimental observation (see **Extended Data Fig. 9d**) of full coupling of the oxidizing equivalents in co-substrate O_2 to the four-electron oxidation of substrate **2**. We believe that the SznF {FeNO}^{6} intermediate is suited for such reactivity based on structural considerations and comparison to enzymatic heme and iron model-complex NO reactivity. In the **1**•Fe^{II}•SznF complex, the two N^{ω} atoms involved in N–N bond formation are located 2.4 Å apart. This proximity is optimal for nucleophilic attack and is likely maintained in the NO-coordinated intermediate via substrate coordination at the O-N^{δ} position. Many iron-nitrosyl complexes are susceptible to nucleophilic attack at the Fe-coordinated nitrogen[70] and, in a particularly relevant biological reaction, a heme {FeNO}^{6} complex in a cytochrome P460 implicated in nitrogen cycle chemistry undergoes N–N bond formation to generate N_2O via reaction with hydroxylamine[24]. Together, these examples provide important precedent for the novel SznF-dependent N–N bond formation in a distinctive Arg-derived substrate. Interestingly, the citrulline derivative and Fe-NO complex generated transiently in this reaction pathway are similar to the products of the second reaction of heme-dependent nitric oxide synthase (NOS), in which a hydroxylated L-Arg intermediate is fragmented at the C^{ε}–N^{ω} bond to serve as the biological source of NO[71]. The cofactor requirements of NOS, a 5-coordinate His-ligated Fe-porphyrin complex, preclude coordination of the substrate L-Arg and the hydroxylated L-Arg intermediate. In SznF, we speculate that a non-heme iron site with fewer requirements for tethering ligands allows for direct interaction with substrate and subsequent evolution of new N–N coupling activity.

Mechanistic proposal 2:



Supplementary Figure 15 | A mechanistic proposal for oxidative rearrangement of 2 without involving transient NO generation. The second proposal follows a route in which the iron cofactor remains in a ferrous state throughout the key N–N bond forming step.

In order for proposal 1 to be consistent with our crossover experiment (**Figure 2c**), the transient {FeNO}^{6} complex cannot exchange prior to N–N bond formation. This phenomenon could be promoted by the close proximity of the functional groups enforced by initial bidentate Fe$^{(II)}$ coordination of the substrate. However, with the result of the crossover experiment in mind, we have also proposed a second mechanism in which transient NO formation is not invoked (**Supplementary Fig. 15**). Bidentate coordination of a redox-active substrate by SznF is a feature shared in common with non-heme iron catechol dioxygenases[65]. In the second proposal, binding of substrate **2** triggers addition of O_2 to yield an initial Fe$^{(III)}$-superoxo adduct[72]. But, as in the proposed mechanism of ring-cleaving catechol dioxygenases, direct coordination of substrate promotes electron transfer to generate an Fe$^{(II)}$-superoxo and a substrate-centered radical[65]. This radical would be stabilized by the adjacent nitrogen atoms. Recombination between the superoxide radical and the C$^{\varepsilon}$• of **2** forms an Fe$^{(II)}$-peroxo bridge at C$^{\varepsilon}$ of **2**[67,68]. In the catechol dioxygenases, a similar structure undergoes a Criegee rearrangement and hydrolysis to ultimately incorporate both O atoms of cosubstrate O_2 into the ring-opened product. Here we propose that the alkylperoxide of SznF breaks down following N–N bond formation to generate a diaziridine intermediate. Rearrangement of the diaziridine could result in O–O bond cleavage, generating *N*-nitrosourea **3** and water as a co-product.

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

The change in oxidation state of the iron cofactor during catalysis is a key point of difference in these proposals and provides an important basis for experimental validation. If the

SznF rearrangement reaction follows Proposal 1, the enzyme could accumulate a high-valent NO-coordinated ferryl intermediate. While similar non-heme {FeNO}^{6} states have been detected in model complexes[73], they have not yet been trapped and characterized as reactive intermediates in an enzyme. In proposal 2, the cofactor would remain in the Fe$^{(II)}$ state throughout the key N–N bond forming step, similar to the aforementioned extradiol aromatic dioxygenases. In the latter systems, x-ray crystallographic characterization of intermediates has been important for corroboration of mechanism[65].

Formatted: Superscript

The mechanism of SznF-mediated oxidative rearrangement is most likely distinct from the reactions catalyzed by KtzT and Spb40, two enzymes recently implicated in N–N bond formation[74,75]. KtzT is a heme-dependent enzyme that performs an intramolecular cyclization of *N*-hydroxy-L-ornithine to afford L-piperazic acid. Although the purified enzyme has not been characterized in detail, proposed coordination of the substrate N-OH to the heme iron metallocofactor in KtzT could enhance the electrophilicity of the N-atom to promote nucleophilic attack by the α-amine[74]. Unlike the SznF reaction, KtzT-mediated cyclization is seemingly oxygen-independent. Other details of the reaction pathway await further experimental investigation. Spb40 is a cupin-adenylation domain fusion enzyme recently identified in a hydrazone-containing dipeptide natural product biosynthetic pathway. This enzyme is proposed to couple *N*-hydroxy-L-lysine and glycine to yield a hydrazine product. Annotation of Spb40 as a cupin-fusion protein suggests structural similarity with the C-terminal domain of SznF. However, the role of metal ions in the Spb40 reaction remains unknown. Additionally, the structures of the putative reactants/products indicate a redox-neutral transformation that is fundamentally distinct from the oxidative reaction mediated by the cupin domain in SznF. The SznF rearrangement reaction requires O_2 to as a source of oxidizing equivalents for *N*-nitrosation and to facilitate O-atom incorporation into the product. These aspects of the reaction mandate a mechanism that is distinct among the increasingly diverse known strategies of N–N bond formation.

Supplementary References

45 Miller, J. H. *Experiments in Molecular Genetics*. (Cold Spring Harbor Laboratory Press, 1972).

46 Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).

47 Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–603 (2012).

48 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta. Crystallogr. D. Biol. Crystallogr.* **D66**, 133–144 (2010).

49 Kabsch, W. XDS. *Acta. Crystallogr. D. Biol. Crystallogr.* **D66**, 125–132 (2010).

50 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

51 Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta. Crystallogr. D. Biol. Crystallogr.* **D66**, 479–485 (2010).

52 Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004).
53 Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
54 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D. Biol. Crystallogr.* **D60**, 2126–2132 (2004).
55 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
56 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. D. Biol. Crystallogr.* **D53**, 240–255 (1997).
57 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D. Biol. Crystallogr.* **D67**, 235–242 (2011).
58 Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta. Crystallogr. D. Biol. Crystallogr.* **D62**, 439–450 (2006).
59 Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta. Crystallogr. D. Biol. Crystallogr.* **D60**, 1355–1363 (2004).
60 Jansen Labby, K. *et al.* Methylated Nω-Hydroxy-l-arginine analogues as mechanistic probes for the second step of the nitric oxide synthase-catalyzed reaction. *Biochemistry* **52**, 3062–3073 (2013).
61 Lin, Y.-M. & Miller, M. J. Practical synthesis of hydroxamate-derived siderophore components by an indirect oxidation method and syntheses of a DIG−siderophore conjugate and a biotin−siderophore conjugate. *J. Org. Chem.* **64**, 7451–7458 (1999).
62 McCoy, J. G. *et al.* Structure and mechanism of mouse cysteine dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3084–3089 (2006).
63 Ye, S. *et al.* An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. *J. Biol. Chem*. **282**, 3391-3402 (2007).
64 Bollinger, J. M., Jr. *et al.* in *2-oxoglutarate-dependent oxygenases* (eds R. P. Hausinger & C. J. Schofield) 95–122 (The Royal Society of Chemistry, 2015).
65 Kovaleva, E. G. & Lipscomb, J. D. Crystal structures of Fe(II) dioxygenase superoxo, alkylperoxo, and bound product intermediates. *Science* **316**, 453–457 (2007).
66 Wieghardt, K. Mechanistic studies involving hydroxylamine. *Adv. Inorg. Bioinorg. Mech.* **3**, 213–274 (1984).
67 Ye, S. *et al.* Electronic structure analysis of the oxygen-activation mechanism by Fe(II)- and alpha-ketoglutarate-dependent dioxygenases. *Chemistry* **18**, 6555–6567 (2012).
68 Mitchell, A. J. *et al.* Visualizing the reaction cycle in an iron(II)- and 2-(oxo)-glutarate-dependent hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
69 Enemark, J. H. & Feltham, R. D. Principles of structure, bonding, and reactivity for metal nitrosyl complexes. *Coord. Chem. Rev.* **13**, 339–406 (1974).
70 Richter-Addo, G. B. & Legzdins, P. *Metal Nitrosyls*. 281–285 (Oxford University Press, 1992).
71 Li, H. & Poulos, T.L. Structure-function studies on nitric oxide synthases. *J. Inorg. Biochem.* **99**, 293–305 (2005).

72 Tamanaha, E. et al. Spectroscopic Evidence for the Two C–H-cleaving intermediates of *Aspergillus nidulans* isopenicillin *N* synthase. *J. Am. Chem. Soc.* **138**, 8862–8874 (2016).
73 Speelman, A.L., White, C.J., Zhang, B., Alp, E.E., Zhao, J., Hu, M., Krebs, C., Penner-Hahn, J., & Lehnert, N. Non-heme high-spin {FeNO}6-8 complexes: one ligand platform can do it all. *J. Am. Chem. Soc.* **140**, 11341–11359 (2018).
74 Du, Y., He, H., Higgins, M. A., & Ryan, K. S. A heme-dependent enzyme forms the nitrogen-nitrogen bond in piperazate. *Nat. Chem. Biol.* **13**, 836–838 (2017).
75 Matsuda, K. *et al.* Discovery of unprecedented hydrazine-forming machinery in bacteria. *J. Am. Chem. Soc.* **140**, 9083–9086 (2018).