Published in final edited form as:

Nature. 2019 February ; 566(7742): 94–99. doi:10.1038/s41586-019-0894-z.

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

Tai L. Ng[1], **Roman Rohac**[2], **Andrew 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

Abstract

N-nitroso-containing small molecules, such as the bacterial natural product streptozotocin, are prominent carcinogens[1,2] and important cancer chemotherapeutics[3,4]. Despite this functional group's significant impact on human health, dedicated enzymes involved in *N*-nitroso assembly have not been identified. Here, we describe a metalloenzyme from streptozotocin biosynthesis (SznF) that catalyzes an oxidative rearrangement of the guanidine group of N^{ω}-methyl-L-arginine to generate an *N*-nitrosourea product. Structural characterization and mutagenesis of SznF uncovered two separate active sites that promote distinct steps in this transformation using different iron-containing metallocofactors. The discovery of this biosynthetic reaction, which has little precedent in enzymology or organic synthesis, expands the catalytic capabilities of non-heme iron-dependent enzymes to include N–N bond formation. We find biosynthetic gene clusters encoding SznF homologs are widely distributed among bacteria, including environmental organisms, plant symbionts, and human pathogens, suggesting an unexpectedly diverse and uncharacterized microbial reservoir of bioactive *N*-nitroso metabolites.

Streptozotocin (SZN) (Zanosar®) is an *N*-nitrosourea natural product and approved cancer chemotherapeutic (Fig. 1a)[3,5]. SZN is also used to induce type I diabetes in animal models due to its toxicity towards pancreatic beta cells (> 28,500 PubMed references)[6]. Like other *N*-nitrosoureas, SZN exerts its activity *in vivo* by generating electrophilic DNA alkylating

*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).

Author Contribution:

T.L.N. and E.P.B. initiated the study. T.L.N. performed bioinformatics analyses and located the gene cluster, carried out the *in vivo* gene knockout and feeding experiments, biochemical characterization of SznE and SznF, chemical syntheses of substrates and standards, liquid chromatography and mass spectrometry analyses, and site-directed mutagenesis experiments. A.K.B. and A.J.M designed the structure determination component of the study. R.R and A.J.M. performed all crystallography experiments with assistance from A.K.B in data analysis. All authors analyzed and discussed the results and prepared the manuscript.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

The authors declare no competing financial interests.

agents and nitric oxide, a precursor to reactive oxygen species[7,8]. To date, the SZN biosynthetic pathway, including the enzyme(s) that install the *N*-nitrosourea pharmacophore, is unknown.

Previous feeding experiments with *Streptomyces achromogenes var. streptozoticus* NRRL 2697 revealed the origins of the sugar scaffold (D-glucosamine), the ureido linkage (L-arginine or L-citrulline), and the *N*-methyl group (L-methionine) (Extended Data Fig. 1a)[9]. Though *N*-nitroso group biosynthesis was not investigated, the distal nitroso nitrogen atom was hypothesized to come from nitrite. All characterized routes for *N*-nitrosation *in vivo* employ nitrite (Fig. 1b) and are not known to be catalyzed by dedicated enzymes.[1,10] Surprisingly, when we fed ^{15}N-labeled nitrite, nitrate, or ammonium salts to *S. achromogenes*, we did not observe isotopically labeled SZN. (Fig. 1c, Extended Data Fig. 1b). *N*-nitrosation in SZN assembly must therefore employ a distinct biosynthetic logic.

To identify the *szn* biosynthetic gene cluster, we sequenced the *S. achromogenes var. streptozoticus* NRRL 2697 genome and searched for genes that might confer resistance to the DNA alkylating activity of SZN[11]. This strategy revealed a 14 kb gene cluster encoding homologs of the DNA repair enzymes AlkB and AGT (SznBCD) (Fig. 1d, Supplementary Table 1). This gene cluster, which is absent in a non-producing strain of *S. achromogenes* (Extended Data Fig. 2a), encodes an *N*-methyltransferase (SznE), two C–N bond-forming enzymes (ATP-grasp enzymes SznH and SznK), and an esterase (SznJ)[12]. It also encodes SznF, an enzyme of unknown function predicted to contain a C-terminal cupin domain and a central domain initially assigned as a ferritin-like motif. As both of these protein scaffolds can catalyze *N*-oxidation[13,14], we hypothesized SznF might participate in *N*-nitrosourea formation.

To validate the function of this gene cluster, we examined the *N*-methyltransferase SznE. Homology modeling of SznE revealed a resemblance to eukaryotic protein L-arginine *N*-methyltransferases (Extended Data Fig. 2b,c)[15]. Incubating purified SznE with L-arginine and *S*-adenosyl-L-methionine (SAM) resulted in N^{ω}-methyl-L-arginine (L-NMA) formation (Fig. 1e, Extended Data Fig. 2d), suggesting that L-NMA was a likely biosynthetic intermediate. Indeed, feeding d_3-L-NMA to *S. achromogenes* produced d_3-SZN (Extended Data Fig. 2e). SZN production was also completely abolished in an *sznE* mutant (Fig. 1f), while biosynthesis was restored with L-NMA (Extended Data Fig. 2f).

To establish which of the *N*-nitroso nitrogen atoms in SZN were derived from L-arginine, we fed [$^{13}C_6{}^{15}N_4$]-L-arginine to *S. achromogenes*. Unexpectedly, we observed only two product isotopologues, [$^{15}N_2{}^{13}C$]-SZN (40%) and [$^{15}N^{13}C$]-SZN (6%) (Fig. 1g, Extended Data Fig. 3a). MS/MS and degradation experiments revealed the two *N*-nitroso nitrogen atoms were ^{15}N labeled in the major isotopologue (Extended Data Fig. 3b,c). Notably, we did not observe any [$^{15}N^{12}C$]-SZN, which would result from *N*-nitrosation of an unlabeled urea with a separate labeled nitrosating species (Fig. 1g). The retention of ^{13}C in all ^{15}N labeled SZN strongly suggested that the two *N*-nitroso nitrogen atoms and the urea carbon atom originate from the intact guanidine side chain of L-arginine. The small amount of [$^{15}N^{13}C$]-SZN likely arises from loss of one ^{15}N label during the interconversion of L-arginine and L-citrulline in cells. Overall, this experiment indicated that *N*-nitrosation

involves an intramolecular oxidative rearrangement of an L-arginine-derived guanidine group, a transformation distinct from known N–N bond forming reactions in biosynthesis and chemical synthesis.

We next characterized the predicted iron-containing enzyme SznF. Genetic knockout of *sznF* abolished SZN production (Fig. 1f). In comparison to wild-type (WT), the *sznF* mutant accumulated L-NMA, suggesting it could be the substrate for this enzyme (Extended Data Fig. 4). SznF expressed in *E. coli* and purified aerobically contained ~0.9 Fe/monomer. As iron-dependent oxygenases typically require initial reduction to the Fe^{II} state to add dioxygen, we incubated purified SznF in air with L-NMA, phenazine methosulfate (PMS), and NADH as a reductant (Extended Data Fig. 5a)[16]. We observed SznF-dependent production of nitrite and nitric oxide from L-NMA, confirming this enzyme catalyzed *N*-oxygenation (Extended Data Fig. 5b).

To identify additional product(s) generated by SznF, we analyzed assay mixtures by LC-MS and characterized four new compounds via comparison to synthetic standards (Fig. 2a, Extended Data Fig. 5c). SznF first hydroxylates L-NMA at the N^{δ}-position to afford N^{δ}-hydroxy-N^{ω}-methyl-L-arginine (**1**). **1** is further hydroxylated to give N^{δ}-hydroxy-N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (**2**). Subsequently, an oxidative rearrangement converts **2** to N^{δ}-hydroxy-N^{ω}-methyl-N^{ω}-nitroso-L-citrulline (**3**). This proposed biosynthetic intermediate degrades non-enzymatically to the denitrosated product **4** and nitric oxide. The N^{δ}-hydroxylation in this reaction sequence was unanticipated, as this N-atom is not incorporated into SZN. SznF activity requires Fe^{II}, O_2, and external reductants (Fig. 2b). L-arginine was not accepted as a substrate (Extended Data Fig. 6a).

We confirmed that **3** is derived from an intact guanidine group using a crossover experiment. We observed no scrambling of the product nitrogen atoms in assays containing both unlabeled and ^{15}N-labeled L-NMA (Fig. 2c, Extended Data Fig. 6b). Adding various metal ions to *apo*-SznF showed that Fe(II) is necessary and sufficient for activity (Extended Data Fig. 6c). SznF purified from *E. coli* does not exhibit heme absorption. We also co-expressed SznF with Rieske ferredoxin SznG. Though these proteins co-purified (Extended Data Fig. 5a), SznG did not improve product conversion *in vitro*. Together, these data show that SznF catalyzes the oxidative rearrangement of a guanidine group to an *N*-nitrosourea in a non-heme iron- and oxygen-dependent manner.

Next, we further defined the role of oxygen in this transformation. Adding L-NMA to SznF assay mixtures accelerated O_2 consumption (Extended Data Fig. 6d). All three *N*-nitrosourea oxygen atoms in **3** were labeled by $^{18}O_2$ (Fig. 2d, Extended Data Fig. 6e), while reactions performed in $H_2{}^{18}O$ did not yield any enrichment (Extended Data Fig. 6f). Finally, the addition of catalase or superoxide dismutase did not affect SznF activity (Extended Data Fig. 6g). These experiments establish that SznF is a monooxygenase. The two new O-atoms in **2** arise from sequential *N*-hydroxylations, each involving one molecule of O_2, with a third equivalent of O_2 providing the ureido oxygen in **3**. Elements of this reaction resemble heme-dependent nitric oxide synthase (NOS), which oxidizes the guanidine group of L-arginine to generate nitric oxide and L-citrulline[17]. NOS also hydroxylates a guanidine nitrogen, and the

ureido oxygen atom of L-citrulline is derived from O_2. However, the involvement of non-heme iron and catalysis of N–N bond formation are key distinguishing features of SznF.

To connect SznF's *in vitro* activity to SZN production *in vivo*, we fed intermediate **1** to Δ*sznE and* Δ*sznF* mutant strains. Feeding **1** to the Δ*sznE* strain restored SZN production, confirming this compound is an on-pathway intermediate (Fig. 3a). However, SZN was not detected when **1** was fed to the Δ*sznF* mutant, indicating that additional processing of this intermediate by SznF is required. We also generated mutant strains lacking predicted downstream biosynthetic enzymes (Δ*sznH*, Δ*sznJ*, Δ*sznK*). Deleting *sznK* completely abolished SZN production, and only trace amounts of SZN were observed in both *sznH* and *sznJ* mutants (Fig. 1f). Though the unstable *N*-nitrosourea **3** did not accumulate in these strains, we observed elevated levels of decomposition product **4** in mutant extracts compared to WT (Extended Data Fig. 4). Together, these results implicate SznF's activity in SZN biosynthesis and suggest SznH, J, and K may transfer the *N*-nitrosourea group of **3** to the D-glucosamine scaffold (Fig. 3b).

To gain further insights into the unusual chemistry mediated by SznF, we characterized its x-ray structure using SeMet-SznF. A 2.08-Å-resolution dataset revealed a dimer with three distinct domains (Fig. 4a, Extended Fig. 7a). N-terminal helical motifs are flanked by C-terminal β-barrel cupin domains, each of which features a 3-His Fe^{II}-binding site (Fig. 4b, Extended Data Fig. 7b). The central domain is a seven-helix bundle, devoid of full occupancy metal ions in our crystals, but similar to dinuclear metalloproteins related by fold to heme oxygenase (HO)[19] (Fig. 4c, Extended Data Fig. 7c). This structural family includes the *Pseudomonas aeruginosa* fatty acid decarboxylase UndA[16,20], and a protein of unknown function from *Chlamydia trachomatis* termed CADD[21] (Extended Data Fig. 8a-c). X-ray structures of CADD[21] and UndA[16] contain di- and mono-nuclear iron centers, respectively, but the structure and oxidation state of the active metallocofactor remains undefined.[20,21] Selected datasets of SznF exhibit a single metal ion in the central domain (Extended Data Fig. 7c), but extensive optimization failed to improve occupancy, suggesting an additional factor is likely needed to assemble or stabilize this metallocofactor for crystallographic characterization.[22] Nevertheless, the central domain of SznF conserves all of the metal-binding His/Asp/Glu ligands present in CADD[21] (Fig. 4c, Extended Data Fig. 8d,e). Substituting any one of these residues (E215, H225, E281, H311, D315, or H318) with Ala abolished all SznF activity (Extended Data Fig. 8f), consistent with their involvement in binding an essential multi-nuclear metallocofactor.

To investigate the cupin domain, we substituted its three Fe^{II}-binding His residues with Ala. In reactions with L-NMA or **1** as the substrate, the H407A/H409A/H448A variant accumulated intermediate **2** but did not generate **3** (Fig. 5a,b). Conversely, an SznF variant (E215A) with an intact cupin domain and disrupted central domain converted synthetic intermediate **2** to *N*-nitrosourea **3** (Fig. 5c). The two metallocofactors of SznF are therefore functionally distinct, with the central domain catalyzing two sequential *N*-hydroxylations of L-NMA and the cupin domain enabling oxidative rearrangement and N–N bond formation to yield the *N*-nitrosourea product (Fig. 5d). Both of these activities expand the chemistry of their respective structural superfamilies and predicted iron cofactors.

While crystallization of L-NMA with SznF yielded no occupancy in either active site, a 1.6 Å resolution dataset of intermediate **1** co-crystallized with SznF (Fig. 4b, Extended Data Fig. 9a,b) reveals the modified L-Arg side chain coordinated to the cupin Fe^{II} via the unmethylated N^{ω} and newly installed N^{δ}-OH. Though **1** is not processed by the cupin domain, this coordination mode suggests the N^{δ}-hydroxyl group is important for Fe^{II} binding. Indeed, assays with N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (**5**) indicate the N^{δ}-OH is required for the oxidative rearrangement. (Extended Data Fig. 9c).

Though the central domain of SznF requires NADH, the cupin domain completely converted dihydroxylated intermediate **2** to *N*-nitrosourea **3** without an external reductant (Extended Data Fig. 9d). This result indicates that the 4-electron oxidation necessary for the final rearrangement is fully coupled to a single 4-electron reduction of O_2. Using our experimental results and the mechanisms of Fe^{II}/2-(oxo)-glutarate and Fe^{II} aromatic ring-cleaving dioxygenases[23] as starting points, we propose two pathways for *N*-nitrosourea formation in SznF that invoke either a transient Fe^{IV}-NO complex or an Fe^{II}-tethered diaziridine as a key intermediate (Supplementary Discussion). Precedent exists for similar reactivity in heme enzyme-catalyzed N–N bond formation[24], including NO reductase, which utilizes non-heme Fe^{II} and heme cofactors to generate N_2O. However, the role of each site in N–N bond formation is unknown[25]. Importantly, the SznF rearrangement reaction does not resemble any known cupin enzyme-mediated transformation.

The novelty of SznF motivated us to assess its distribution in microbial genomes. BLASTP searches uncovered SznF homologs containing both the central and cupin domains encoded in numerous cryptic biosynthetic gene clusters (e-value < 1E-5) (Extended Data Fig. 10, Supplementary Tables 2, 3). These gene clusters are widely distributed across bacterial phyla and habitats (Extended Data Fig. 10a, b), including soil and rhizosphere *Burkholderia* and *Paraburkholderia* strains, the plant symbiont *Frankia*, and the human pathogen *Legionella pneumophila*. This analysis suggests diverse bacteria produce *N*-nitroso-containing compounds, highlighting potential new biological roles for these metabolites and enabling genome mining-based natural product discovery.

In studying the biosynthesis of the pancreatic cancer drug SZN, we have uncovered the *N*-nitrosating metalloenzyme SznF. This discovery reshapes our view of the origins and roles of *N*-nitroso groups and other reactive functionality found in biologically active metabolites. Such structural features are typically considered to derive from non-enzymatic transformations in living systems. Our work demonstrates that enzymes have evolved to synthesize *N*-nitroso compounds. The unexpectedly wide distribution of SznF homologs in bacteria suggests that *N*-nitroso metabolites have underappreciated biological significance, including roles in symbioses and pathogen-host interactions.

Methods

Materials and general methods

Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA) and Sigma-Aldrich (Atlanta, GA). Recombinant plasmid DNA was purified with a Qiaprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up

were performed using Zymoclean™ Gel DNA Recovery Kit and DNA Clean & Concentrator kit from Zymo Research (Irvine, CA). DNA sequencing was performed by Genewiz (Boston, MA) and Eton Bioscience (Charleston, MA). Nickel-nitrilotriacetic acid agarose (Ni-NTA) resin was purchased from Qiagen and Thermo Scientific. SDS-PAGE gels were purchased from BioRad. Protein concentrations were determined by measuring absorption at 280 nm and using ExPASy ProtParam (http://web.expasy.org/protparam/) to calculate the extinction coefficients. Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm.

Analytical and preparative HPLC were performed on a Dionex Ultimate 3000 instrument (Thermo Fisher Scientific). High-resolution mass spectral data for the synthetic compounds were obtained on a Bruker MicroQTOF-QII mass spectrometer fitted with a dual-spray electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV and the end plate offset to −500 V, the drying gas temperature was maintained at 190 ºC with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. The liquid chromatography was performed using an Agilent Technologies 1100 series LC. Isopropanol, methanol, and water used for LC-ESI-MS were B & J Brand High Purity Solvents (Honeywell Burdick & Jackson).

High resolution mass spectrometry data were obtained using an Agilent 1200 series LC system coupled to an Agilent 6530 quadrupole time-of-flight (qTOF) mass spectrometer with an ESI source. The mass spectra data were recorded in either 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. For MS2 settings, mass range, 100 to 1700 *m/z*; collision energy, 10 eV; isolation width MS/MS, ~1.3 amu; spectra rate, 1 spectra/s; nebulizer pressure, 35 psi; drying gas (N_2) flow, 8 L/min; precursor threshold, 200 counts; temperature, 275 ºC. A mass window of 5 ppm was used to extract the ion counts for generating extracted ion chromatograms with Agilent ChemStation software. The specific LC methods for separating streptozotocin (SZN) and amino acids are described below.

1. Cultivation of *Streptomyces achromogenes subsp. streptozoticus* NRRL 2697 and detection of streptozotocin.

Streptomyces achromogenes subsp. streptozoticus NRRL 2697 was obtained from the Agricultural Research Service (ARS) Culture Collection. The organism was grown on mannitol-soy (MS) agar plates (20 g/L soy flour, 20 g/L mannitol, and 20 g/mL Bacto™ agar (Difco™)) at 30 ºC for five days. The spores were scraped and inoculated into 5 mL of tryptic soy broth (TSB) medium (Difco™) and allowed to shake at 30 ºC until saturation. 3 mL of this starter culture was inoculated into 100 mL production medium in a 250 mL baffled flask (40 g/L yellow corn meal (Arrowhead Mills), 15 g/L potato starch (Bob's Red Mill), 1 g/L glucose, 3 g/L peptone (Bacto™), and 14 g/L ammonium sulfate). The fermentation was carried out for 5 days at 30 ºC shaking at 200 rpm. 750 μL of the culture was centrifuged at 13,000 rpm for 1 min, and an equivalent volume of acetone was added to the supernatant to further precipitate proteins and starches. After vortexing, the precipitate

was removed by centrifugation at 13,000 rpm, and the acetone was removed under vacuum. The samples were analyzed with LC-HRMS.

Streptozotocin (SZN) was detected as the molecular ions $[M + H]^+$ or $[M – H_2O + H]^+$. Mass spectra data were collected only between 1.4 and 5 min, with the LC stream diverted to waste before 1.4 min and after 5 min. The LC column was an Acclaim^{Tm} PolarAdvantage II C18 column (3 μm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 mL/min. The LC conditions were: 95% solvent A, hold for 2 min; 95 to 50% solvent A in 6 min; 50% solvent A, hold for 1 min; 50 to 95% solvent A in 3 min; and 4 min equilibration at 95% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile).

Genome sequencing of *Streptomyces achromogenes var. streptozoticus* NRRL 2697 and genomic DNA library construction—Genomic DNA was purified using an UltraClean® Microbial DNA Isolation Kit (MoBio). Library construction from genomic DNA, sequencing, and assembly were performed by Cofactor Genomics (St. Louis, MO). Next-generation sequencing used HiSeq Illumina reads of two short-insert paired-end libraries (300 bp insert and 500 bp insert) and a long-insert mate-pair library (1–2 kb insert). Assembly of the reads resulted in 8.6 MB of non-redundant sequence distributed over 24 contigs. Annotations were carried out using PROKKA's database of other *Streptomyces* species[26]. The assembled data were converted into a local BLAST database using Geneious Pro Version 7.1.6 (Biomatters). The *Streptomyces achromogenes var. streptozoticus* NRRL 2697 fosmid library was prepared using the CopyControl HTP Fosmid Library Production Kit (Epicentre) following the manufacturer's protocol. A library of 2,000 clones was picked into 96-well plates and stored at –80 ºC as 50% glycerol stocks.

Feeding experiments with inorganic nitrogen salts—*S. achromogenes var. streptozoticus* NRRL 2697 was grown in 25 mL of fermentation medium in 250 mL baffled flasks while shaking at 30 ºC as described above. After 16 h, stock solutions of [^{15}N]-calcium nitrate, [^{15}N]-sodium nitrite, and [^{15}N]-ammonium chloride (Cambridge Isotope Inc.) in water were sterilized by passing through a 0.22 μm filter membrane. Each of these nitrogen sources was added to the fermentation cultures to 1 mM final concentration. After 4 more days of fermentation at 30 ºC, the presence of labeled and unlabeled SZN was determined by LC-HRMS after removing cell debris as described above. These feeding experiments were performed at least twice on different days and the same result was obtained.

Identification of the putative streptozotocin (*szn*) biosynthetic gene cluster—A BLAST search using *E. coli* DNA repair enzymes AlkA (annotation: DNA-3-methyladenine glycosylase 2, NCBI Accession: AUG16819.1) revealed 3 homologs in the *S. achromogenes var. streptozoticus* genome. Comparison of the *S. achromogenes var. streptozoticus* and *S. achromogenes var. achromogenes* NRRL B-2120 (Genbank Assembly Accession Number: GCA_000720835.1) genomes was performed using Mauve alignment software[27]. The region encoding DNA repair enzyme homologs and putative biosynthetic enzymes that is absent from the non-producing strain is depicted in Extended Data Figure 2a. Annotations of the open reading frames (ORFs) in the *szn gene* cluster are found in Supplementary Table 1.

2. Homology modeling and sequence alignment of SznE

Structural homologs of SznE were identified using HHPred (http://toolkit.tuebingen.mpg.de/hhpred)[28]. The closest structural homologs' sequences were aligned with SznE using ClustalW[29]. The SznE homology model was generated with MODELLER (Max Planck Institute for Developmental Biology)[30] using the structure of TbPRMT7, a protein arginine methyltransferase that exclusively monomethylates arginine residues, as a template (PDB code: 4M37)[15]. The resulting file was aligned with the TbPRMT7 PDB file using Pymol Molecular Graphics system, Version 1.8 (Shrödinger, LLC).

Cloning, overexpression, and purification of SznE—PCR reaction mixtures contained 22.75 μL of water, 0.125 μL of forward and reverse primers (Supplementary Table 4), 0.5 μL of genomic DNA, 1.5 μL of DMSO, and 25 μL of Q5® High-Fidelity 2x Master Mix (New England Biolabs) in a final volume of 50 μL. Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) using the following parameters: denaturation for 1 min at 98 ºC; 35 cycles of 0.5 min at 98 ºC, 0.5 min at 70 ºC, 2 min at 72 ºC; and a final extension time of 10 min at 72 ºC. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide or SYBR safe staining, pooled, and purified. Amplified fragments were digested with Fastdigest NdeI and Fastdigest HindIII (Thermo Fisher Scientific) for 2 h at 37 ºC. Digests contained 16.5 μL purified PCR product, 2 μL of FastDigest Buffer (10x), 0.4 μL of NdeI (20 U/μL), and 0.4 μL of HindIII (20 U/μL). Restriction digests for PCR were purified directly using Zymo DNA Clean & Concentrator™. To prepare a linearized pET28a vector, 50 ng of the vector was digested with NdeI and HindIII as described with the exception of addition of 1 μL calf-intestinal alkaline phosphatase in the reaction after 2 h and further incubation for 1 h. The linearized DNA was purified first by agarose gel electrophoresis and further purified with Zymoclean™ Gel DNA Recovery Kit. The insert digest was ligated into linearized expression vector pET28a using T4 DNA ligase (New England Biolabs). Ligations were run overnight at 16 ºC and contained 1 μL T4 Ligase Buffer (10x), 1.5 μL digested vector, 5.5 μL digested insert DNA, and 2 μL T4 DNA Ligase (400 U/μL). 5 μL of each ligation was used to transform a single tube of *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pET28a-SznE construct was confirmed by sequencing of purified plasmid DNA. These constructs were transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) and stored at –80 °C as frozen glycerol stocks.

A 50 mL starter culture of pET28a-SznE BL21 *E. coli* was inoculated from a single colony and grown overnight at 37 °C in LB medium supplemented with 50 μg/ml kanamycin. 7 mL of the saturated starter culture was inoculated into 700 mL of LB. The cultures were incubated at 37 ºC with shaking for ~ 2.5 h, induced with 250 μM IPTG at OD600 ~ 0.5, and then incubated at 15 ºC for 16 h. Cells were pelleted by centrifugation (6,000 rpm x 10 min) and resuspended in 40 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM $MgCl_2$, pH 8). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 10,000 psi, and the lysate was clarified by centrifugation (10,800 rpm x 40 min). The supernatant was incubated with 2 mL of Ni-NTA resin for 1 h at 4 ºC. The mixture was then loaded into a glass column and washed with 20 mL of wash buffer (50 mM HEPES, 20 mM imidazole, 10 mM $MgCl_2$, pH 8.0). The protein was eluted with approximately 20 mL

of elution buffer (50 mM HEPES, 200 mM imidazole, 10 mM $MgCl_2$, pH 8.0). SDS–PAGE analysis (4 – 15% Tris-HCl gel) confirmed the presence and purity of SznE. The eluted protein was concentrated to ~1 mL using a Corning® Spin-X® UF 20 mL Centrifugal Concentrator (30,000 MWCO Membrane) after centrifugation at 4000 rpm. 12 mL of exchange buffer (20 mM HEPES, 50 mM NaCl, 10% glycerol, pH 8.0) was added, and the sample was concentrated again to 1 mL. This process was repeated one more time before the concentrated, desalted solution containing purified SznE was frozen in liquid N_2 and stored at –80 ºC.

^{1}H NMR assay of SznE activity—In a 500 μL reaction mixture, 50 mM potassium phosphate pH 8.0, 10 mM $MgCl_2$, 2 mM L-arginine, 4 mM *S*-(5′-Adenosyl)-L-methionine (SAM), and 5 μM of SznE were mixed and incubated overnight at room temperature. Negative control assays omitting SAM, L-arginine, or SznE in the reaction were also performed. The reaction mixtures were flash frozen with liquid N_2 and lyophilized. The residues were resuspended in D_2O (Cambridge Isotope Inc.) and analyzed with proton nuclear magnetic resonance (^{1}H NMR) spectroscopy using an Agilent DD2–600 NMR spectrometer (600 MHz). Chemical shifts (δ) are reported in parts per million (ppm) downfield from tetramethylsilane using the solvent resonance as an internal standard for ^{1}H (D_2O = 4.79 ppm). This experiment was performed in triplicate.

Feeding experiments with labeled arginine and degradation studies— *Streptomyces achromogenes var. streptozoticus* NRRL 2697 was grown in 10 mL of fermentation medium in 50 mL baffled flasks while shaking at 30 ºC as described in section 1. Stock solutions of [$^{15}N_4{}^{13}C_6$]-L-arginine, [$^{15}N_2$]-L-arginine, or [d_3]-L-N^{ω}-methyl-L-arginine in water were sterilized by passing through a 0.22 μm filter membrane and added to the fermentation culture 16 h after inoculation to give 1 mM final concentration. After 2 more days of fermentation at 30 ºC, the presence of labeled and unlabeled SZN was determined by LC-HRMS as described in section 1. The positions of the labeled nitrogen and carbon atoms were elucidated using degradation experiments as described in Extended Data Figure 3.

3. Gene disruption and chemical complementation experiments

Gene inactivation in *Streptomyces achromogenes var. streptozoticus* NRRL 2697 was performed according to standard protocols[31]. Briefly, primers used to amplify *sznE* were used to screen the fosmid library to obtain a fosmid containing the *szn* gene cluster. The fosmid was transformed into *E. coli* BW25113/pkD46 by electroporation. The *aac(3)IV*-oriT cassette was amplified by PCR from pIJ773 using the primers listed in Supplementary Table 5. Each of the biosynthetic genes was replaced with the *aac(3)IV*-oriT cassette PCR product using PCR targeting and λ-red-mediated recombination. The mutant fosmid was transformed into *E. coli* WM6026 for conjugation with *S. achromogenes subsp. streptozoticus* NRRL 2697. Double crossover mutants were selected with apramycin resistance on mannitol-soy agar, and the exconjugants were grown in TSB with apramycin added (50 μg/mL final concentration). After the seed culture reached saturation, the genomic DNA was isolated, and insertion of the resistance marker was confirmed by PCR (Extended Data Fig. 4, Supplementary Table 4). The mutants were tested for SZN production under the

same fermentation conditions as described in section 1 for the wild type strain with the exception of added antibiotics. To confirm the gene deletions for *sznF*, primers pIJ773-F and sznF-R and pIJ773-R and sznF-F were used. For *sznF*, an internal primer specific to the resistance cassette was also used due to the similar band sizes of the wild type and inserted gene (both expected 1.5 kb).

For the chemical complementation experiments, wild type and mutant strains were grown in 10 mL of fermentation medium in 50 mL baffled flasks while shaking at 30 ºC as described in section 1. Stock solutions of L-NMA or **1** in water were sterilized by passing through a 0.22 μm filter membrane and added to the fermentation to 1 mM final concentration after 16 h. After 2 more days of fermentation at 30 ºC, the presence of SZN was determined by LC-HRMS as described in section 1.

For LC-HRMS detection of L-NMA, **1**, and **2** from culture supernatant extracts, negative ion mode was used. The LC column was a Cogent Diamond Hydride column (4 μm, 100 Å, 3 × 150 mm, Microsolv Technology Corp.). The flow rate was 0.5 mL/min. The LC conditions were: 10% solvent A, hold for 1 min; 10 to 70% solvent A in 19 min; 70% solvent A, hold for 1 min; 70 to 10% solvent A in 4 min; and 4 min equilibration at 10% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). The extracted ion chromatograms were generated using the $[M-H]^-$ molecular ions with a 5 ppm window.

4. *In vitro* characterization of SznF

Overexpression and purification of SznF—The *sznF* gene was amplified from the cosmid encoding the *szn* gene cluster using the primers in Supplementary Table 4. The amplified gene was digested with NdeI and HindIII and ligated into the pET28a vector as described above for pET28a-SznE. The plasmid was transformed into *E. coli* BL21 (DE3) cells (Invitrogen), and the protein was expressed and purified using the same procedure as described for SznE with the exception that the exchange buffer contained 20 mM MOPS pH 7.5 instead of 20 mM HEPES pH 8.

LC-MS assay of SznF activity—In a total volume of 50 μL, 50 mM MOPS pH 7.5, 20 μM phenazine methosulfate (PMS), 80 μM of SznF, and 1 mM of amino acid substrate were mixed. NADH (Merck Millipore) was added at a final concentration of 5 mM to initiate the reaction. Reaction mixtures were quenched by adding 50 μL of LC-MS grade methanol. After 10 min on ice, the mixtures were centrifuged to remove precipitated protein and analyzed using normal phase LC-HRMS as described in section 3 in which L-NMA, **1**, and **2** were detected from cell extracts.

For analysis of Fmoc derivatives, a 10 mM 9-Fluorenylmethyl *N*-succinimidyl carbonate (Fmoc-OSu) acetonitrile solution was first prepared. An additional 50 μL of acetonitrile was added to the quenched reaction mixtures, followed by 5 μL of the 10 mM Fmoc-OSu solution. This mixture was incubated for 1 h, centrifuged to remove precipitates, and analyzed by reversed phase LC-HRMS. The LC column was an Acclaim^{Tm} PolarAdvantage II C18 column (3 μm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 mL/min. The LC conditions were: 95 to 5% solvent A in 15 min; 5% solvent A, hold for

5 min; 5 to 95% solvent A in 4 min; and 5 min equilibration at 95% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile).

To access labeled L-NMA, a large-scale SznE assay was prepared. The assay mixture contained 50 mM potassium phosphate buffer pH 8, 10 mM $MgCl_2$, 10 mM L-arginine, 40 mM SAM, and 10 μM SznE in a final volume of 1 mL. The reaction mixture was allowed to proceed overnight at room temperature and was quenched with 2 mL of MeOH. After centrifugation to remove precipitates, the enzyme reaction was evaporated to dryness, resuspended in 1 mL of water, and labeled L-NMA was purified using semi-preparative scale HPLC. The LC column was Kromasil 100 C18 (5 mm, 100 Å, 10 × 250 μm). The flow rate was 3 mL/min. The LC conditions were: a linear gradient from 5 to 50% solvent B in 12 min; an isocratic hold at 50% solvent B for 2 min; a linear gradient from 50 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 dryness, and the residue was resuspended to make a 100 mM stock solution, which was kept at –20 ºC.

Testing the metal dependence of SznF—*apo*-SznF was prepared by inoculating 1 L of M9 medium with 1% of a saturated LB culture of BL21 harboring pET28a-SznF. At an OD600 ~ 0.4, 0.2 mM of IPTG was added, and the culture was incubated at 15 ºC overnight. The protein purification procedure was identical to that described above for SznF, except that after elution from the Ni-NTA column, a 100 mM aqueous solution of EDTA pH 7 was added to a final concentration of 0.5 mM to remove any residual Ni bound to SznF. This solution was concentrated to ~ 400 μL and diluted into 12 mL of exchange buffer containing 0.5 mM EDTA. The protein solution was concentrated to ~ 400 μL and diluted again into 12 mL of exchange buffer without EDTA. The protein sample was concentrated once more to obtain a desalted EDTA-treated SznF sample. An aliquot of this solution was submitted to the University of Georgia Center for Applied Isotope Studies for inductive-coupled plasma-mass spectrometry (ICP-MS) analysis. This SznF sample had < 0.04 ppm ^{56}Fe, ^{65}Cu, ^{59}Co, ^{55}Mn by ICP-MS analysis was referred to as *apo*-SznF.

To screen the effects of different metals on the SznF-mediated *N*-oxygenation reactions, 100 mM stock solutions of $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, $MnCl_2 \cdot 4H_2O$, $CaCl_2 \cdot 2H_2O$, $ZnCl_2$, $CuCl_2 \cdot 2H_2O$, $CoCl_2 \cdot 6H_2O$, $Na_2MoO_4 \cdot 2H_2O$, and $NiSO_4 \cdot 6H_2O$ were prepared in 1 M aqueous HCl. Assay mixtures contained 50 mM MOPS pH 7.5, 1 mM L-NMA, 200 μM metal, 20 μM PMS, 5 mM NADH, and 100 μM *apo*-SznF in a final volume of 30 μL. Assay mixtures were incubated for 4 h before quenching with methanol and analyzing with LC-HRMS.

Griess assay for nitrite production and nitric oxide detection by EPR—To characterize nitrite liberated from the SznF-mediated *N*-nitrosation, a solution of 0.5% sulfanilic acid in 30% acetic acid (Solution A) and 0.1% (Napthyl)ethylenediamine dihydrochloride in 30% acetic acid (Solution B) were prepared. 25 μL of Solution A, 25 μL of Solution B, 10 μL 6 M HCl, and 10 μL of the enzyme assay mixtures were combined and heated at 60 ºC for 3 min. The samples were centrifuged to remove precipitated proteins, and the absorbance at 548 nm was recorded with a BioTek Gen5 Microplate Reader.

Nitric oxide detection was carried out by adapting the work of Yong Xia *et al.*[32] In brief, a 10x stock aqueous solution containing 50 mM iron ammonium sulfate and 10 mM *N*-methyl-D-glucamine dithiocarbamate (MGD) was freshly prepared. 25 μL of this $(MGD)_2Fe(II)$ solution was then added to 225 μL of an SznF assay mixture. After 30 min at room temperature, the reaction mixture was transferred to a quartz capillary and analyzed with a Bruker ElexSysE500 EPR instrument with a 100 K–600 K Digital Temperature Control at room temperature. The following parameters were used: 9.75 GHz microwave frequency; 100 kHz modulation frequency; 10 mW microwave power, 3.81 G modulation amplitude; 3418 G center field; 82 ms time constant. An enzyme-free sample containing 1 mM sodium 2-(*N*,*N*-diethylamino)-diazenolate-2-oxide (DEANO) was used as a positive control. To demonstrate that the nitric oxide was generated from the terminal nitrogen atoms of L-NMA, 1 mM of $[^{15}N_2]$-L-arginine (Cambridge Isotope Inc.) was first converted to $[^{15}N_2]$-L-NMA with SznE as described in section 2. After 1 h, SznF, PMS, and NADH were added directly to the reaction mixture to the same final concentrations as described in section 4. $(MGD)_2Fe(II)$ was then added, and the SznEF assay mixture was analyzed with EPR spectroscopy.

Addition of superoxide dismutase and catalase to SznF reaction mixtures— Stock solutions of superoxide dismutase (from bovine, recombinant from *E. coli*, Sigma-Aldrich) and catalase (from bovine liver, Sigma Aldrich) were prepared by adding 100 μL of buffer (20 mM MOPS pH 8, 50 mM NaCl) to the lyophilized powders. After centrifugation to remove particulates, the enzyme solutions were added to the SznF reaction mixtures at a final concentration of 1U / 50 μL prior to addition of NADH. The reactions were allowed to proceed for 1 h before activity was measured using the Griess assay. The experiments were performed in triplicate on three different days.

Oxygen consumption assay—Dissolved oxygen was monitored using a MultiFrequency phase fluorometer equipped with an FOXY optode (Ocean Optics). The optode was calibrated using a two-point calibration method with air-saturated 20 mM MOPS buffer pH 7.7 (280 μM) and the same solution with dithionite (0 μM). 500 μL of an SznF reaction mixture was prepared in a sealed pear-shaped flask with stirring as described above. The data was analyzed using NeoFox Viewer version 2.40. The different experimental conditions tested are described in Extended Data Figure 6d. The experiments were performed at least twice on different days.

$^{18}O_2$ and $H_2{}^{18}O$ labeling experiment—Solutions of buffer, $MgCl_2$, and water were degassed by sparging with argon for 1 h. SznF was made anaerobic by repeated cycles of vacuum and purging with argon. The enzyme, degassed buffer, degassed solution of $MgCl_2$, degassed water, PMS, and NADH were brought into an MBraun glove box (Stratham, NH) under an atmosphere consisting of 99.997% nitrogen (N_2) with less than 0.1 ppm O_2. 50 mM HEPES pH 8.0, 10 mM $MgCl_2$, 1 mM L-NMA or 1 mM **1**, 200 μM PMS, and 100 μM of SznF were mixed in a vial in a total volume of 100 μL. The reaction mixture and a degassed, aqueous solution of NADH were brought into a glovebag purged with nitrogen. The reaction in the vial was charged with 5 mL of $^{18}O_2$ using a gas-tight syringe, and NADH was added to a final concentration of 5 mM. After 1 h, methanol briefly sparged with

argon was drawn into a gas-tight syringe previously filled with N_2 inside the glovebag and added to the reaction mixture. The reactions were Fmoc derivatized and analyzed with HR-LCMS and MS/MS. For the reaction performed in $H_2{}^{18}O$, the reaction mixture was prepared as above except $H_2{}^{18}O$ was used to a final volume of 50 μL instead of 100 μL. The reaction mixture was quenched, derivatized, and analyzed with LC-HRMS.

Site-directed mutagenesis—The SznF point mutants were constructed by adapting the Quikchange protocol[33]. In brief, each PCR reaction contained 41.65 μL H_2O, 1.5 μL DMSO, 0.2 μL dNTP (50 mM stock), 0.4 μL of pET28a-SznF (50 ng/μL stock), 0.125 μL of each mutagenesis primer (200 μM stock, nucleotide sequences in Supplementary Table 6), 1 μL of Pfu Turbo AD and 5 μL of Buffer (Turbo AD). Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) using the following parameters: denaturation for 2 min at 95 ºC; 18 cycles of 0.5 min at 95 ºC, 1 min at 71 ºC, and an extension time of 13 min at 72 ºC. 40 μL of the PCR reaction mixture was digested with 2 μL of DpnI for 2 h at 37 ºC. Afterwards, the reaction was transformed into chemically competent *E. coli* Top10 cells, and the constructs containing the desired point mutants were confirmed by sequencing. The mutant SznF enzymes were expressed, purified, and assays using the procedures described above for wild type SznF.

Phylogeny of SznF and gene cluster analysis—The maximum likelihood phylogenetic tree was generated using sequences of SznF that contain both the central helix domain and cupin domain using the NCBI non-redundant (nr) sequences Database (2018, e-value cutoff < 1E-50). The tree was calculated with MEGA-X with 50 bootstrap replications, LG substitution model, and Nearest Neighbor-Interchange heuristic model[34]. Genome neighborhood analysis of SznF homologs were performed with the software tools available at Integrated Microbial Genomes and Microbiome system (Joint Genomic Institute)[35].

SznF crystallization and x-ray structure determination—The materials and methods for purifying SznF for crystallographic characterization, SznF crystallization, and x-ray structure determination are described in the Supplementary Information.

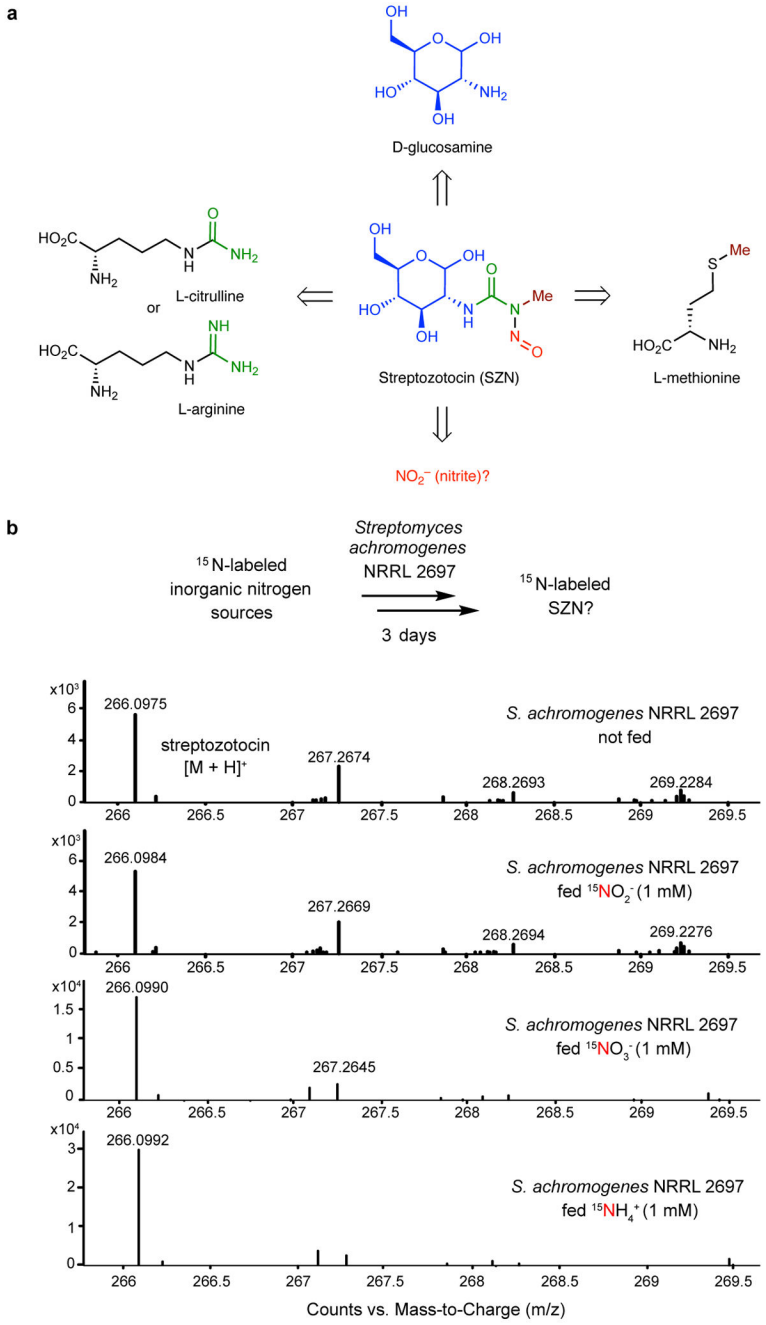
Chemical synthesis and compound characterization—The procedures for preparing the synthetic compounds used in this study and spectral characterization are reported in Supplementary Information.

Data Availability—The nucleotide sequences for the *szn* biosynthetic gene cluster and individual genes have been deposited into NCBI (accession numbers: TBD). The structure factor and coordinate of SznF have been deposited in the Protein Data Bank (PDB codes: 6M9R, 6M9S). Additional data that support the conclusions of the paper can be requested from the corresponding authors.

Statistics and Reproducibility—The *in vitro* assay for SznE activity in Figure 1e was repeated twice on different days with different preparations of SznE. The activity was also confirmed via HPLC after Fmoc-derivatization (data not shown). The LC-MS analysis in Figure 1f is representative of more than three individual experiments from different starter cultures over the course of two years. The labeling studies in Figure 1g were repeated with

lower concentrations of labeled arginine, and the same labeling pattern was observed. The experiment was also repeated with [$^{15}N_2$]-L-arginine which generated the corresponding labeled products (data not shown). The LC-MS time course in Figure 2a was repeated independently once using a different enzyme preparation. The control assays in Figure 2b were performed a total of two times. The end-point activity assay of SznF has been performed numerous (>10) times, with the metabolites **1**,**2**,**3**, and **4** consistently produced every time. The labeling study with $^{18}O_2$ in Figure 2d was performed once. The chemical complementation experiment in Figure 3a was performed twice with different starter and fermentation cultures. LC-MS traces shown in Fig. 5abc are representative of at least two individual experiments performed on different days with different enzyme preparations. These experiments were also repeated during submission and revision of the manuscript. The feeding experiment in Extended Data Figure 1 was performed in triplicate for each of the labeled salts tested. The enzyme SznE was purified at least twice with purity similar to that shown in Extended Data Figure 2d. The labeling study and chemical complementation experiments shown in Extended Data Figure 2ef were repeated independently two years apart. The experiment shown in Extended Data Figure 3ab were repeated twice, and the analysis for Extended Data Figure 3c was performed once. At least two exconjugants were picked for each mutant for the PCR analysis in Extended Data Figure 4a, and the metabolite analysis has been performed at least three times independently with similar results. Purification of SznF and SznFG has been replicated multiple times by both of the corresponding authors' research groups (Extended Data Figure 5a). The detection of NO by EPR was performed in duplicate on the same day. The LC-MS analysis of each SznF product and standard was performed more than three times on different days, and the LC-MS/MS analysis was performed twice for **1** using positive and negative ion mode, once for **2**, three times for Fmoc-**3**, and twice for **4** using positive and negative ion mode. The experiments for Extended Data Figure 6acef were performed once, and the O_2 consumption assay (Extended Data Figure 6d) was performed twice. The activity assay shown in Extended Figure 8f was performed once except for E215A, which was performed more than four times independently with different enzyme preparations. The experiment shown in Extended Data Figure 9c was repeated once, and the activity assay depicted in Extended Data Figure 9d was repeated on different days with similar results. The NMR spectra supplied in the Supplementary Figures were recorded once.

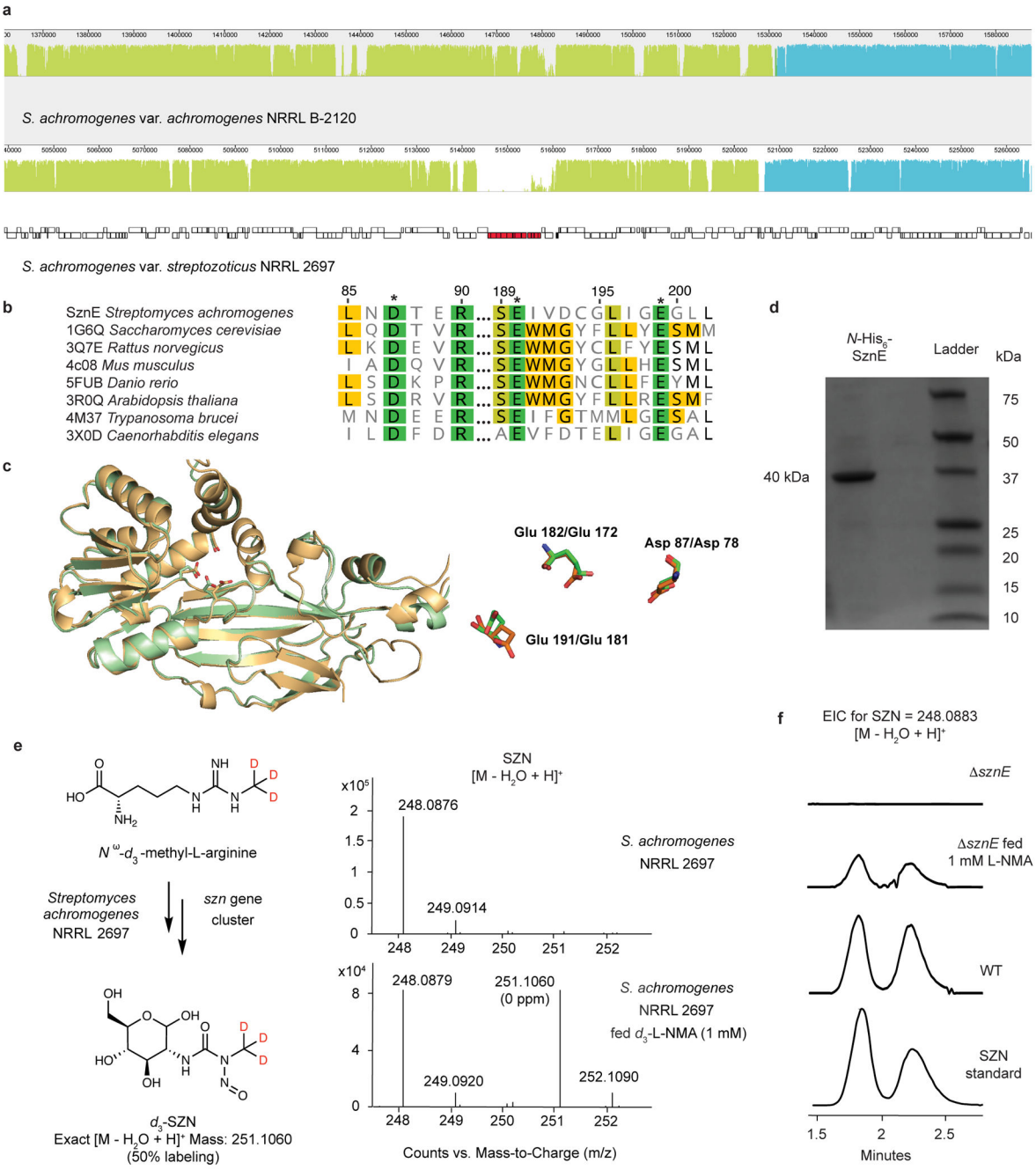
Extended Data



Extended Data Figure 1 |. Inorganic nitrogen sources are not precursors to SZN.

a, Previous feeding experiments suggested that D-glucosamine, L-citrulline or L-arginine, and L-methionine-derived *S*-adenosylmethionine (SAM) would be building blocks for SZN biosynthesis[9]. **b**, Mass spectra of culture extracts in which ^{15}N-nitrate, ^{15}N-nitrite, or ^{15}N-ammonium chloride were fed to *S. achromogenes var. streptozoticus* NRRL 2697. The expected masses ($[M+H]^+$) for SZN, ^{15}N-SZN, $[^{15}N_2]$-SZN, and $[^{15}N_3]$-SZN are 266.0983, 267.0953, 268.0923, and 269.0894, respectively. These results contrast sharply with the

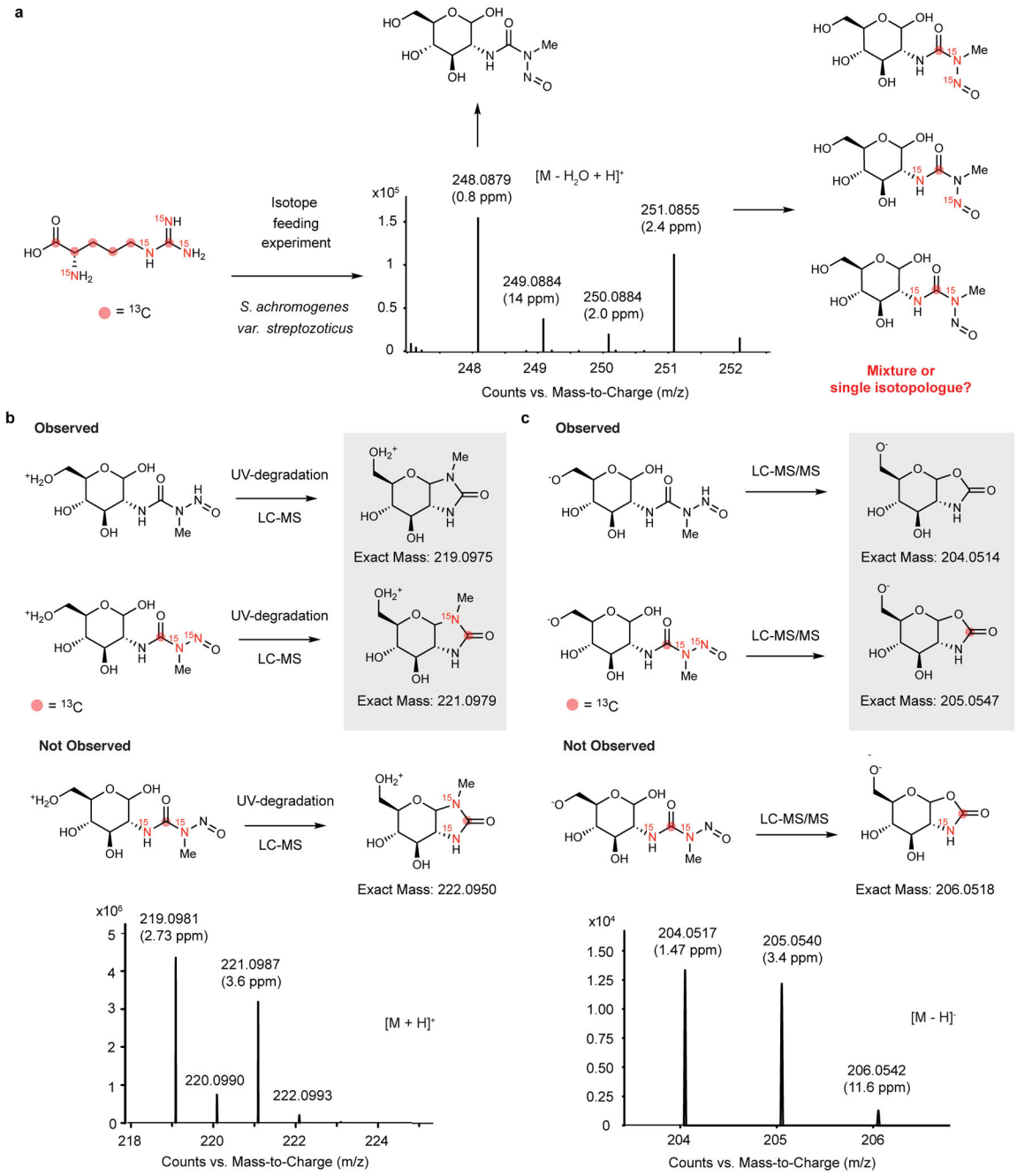
strong labeling (>75%) observed in studies of pathways that use nitrite for diazo biosynthesis[36,37].



Extended Data Figure 2 |. N^{ω}-methyl-L-arginine (L-NMA) is an on-pathway intermediate in SZN biosynthesis.

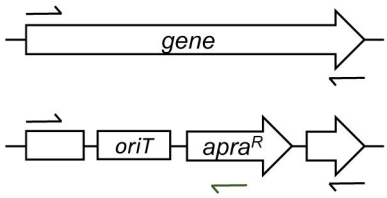
a, Comparative genomic analysis of *S. achromogenes var. achromogenes* NRRL B-2120, a nonproducer of SZN and *S. achromogenes var. streptozoticus* NRRL 2697 using Mauve[27]. The *szn* gene cluster is colored in red. Gene annotations are tabulated in Supplementary Table 1. **b**, Multiple sequence alignment of SznE with structurally characterized protein

arginine methyl transferases (PRMT) from eukaryotes. Conserved residues involved in binding L-arginine are marked with an asterisk. **c**, Overlay of a homology model of SznE (green) with the crystal structure of PRMT7 from *T. brucei* (PDB accession code 4M37) (orange). The highlighted carboxylate residues are involved in binding of the basic guanidine group[15]. **d**, SDS-PAGE of purified SznE. The expected molecular weight is 40 kDa. Ladder = Precision Plus Protein All Blue Standards (BioRad). **e**, Mass spectra of SZN produced when feeding *S. achromogenes var. streptozoticus* NRRL with d_3-L-NMA. The expected masses $[M\text{-}H_2O+H]^+$ for SZN and d_3-SZN are 248.0993 and 251.1060, respectively. **f**, LC-MS traces demonstrating restoration of SZN production by the Δ*sznE* mutant upon chemical complementation with L-NMA. The extracted ion chromatograms (EIC) are generated within a 5 ppm window.

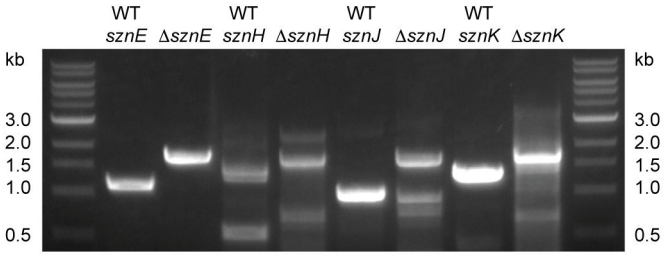


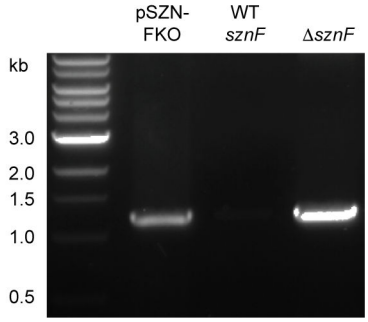
Extended Data Figure 3 |. The *N*-nitrosourea of SZN is derived from an intact guanidine group of L-arginine.

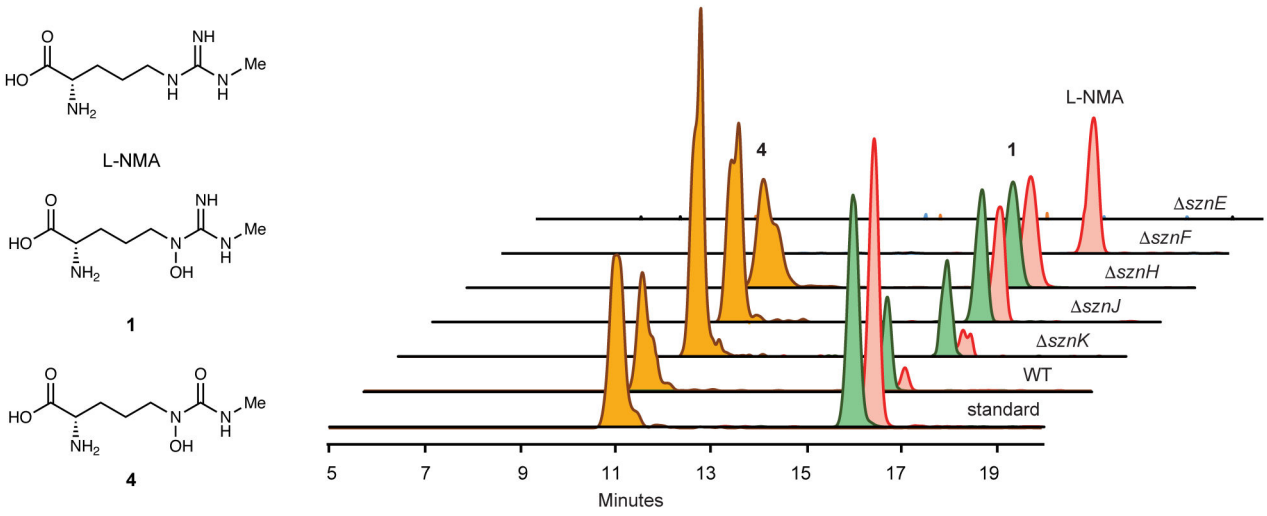
a, The mass spectrum of SZN $[M\text{-}H_2O+H]^+$ when 1 mM of $[^{15}N_4{}^{13}C_6]$-L-arginine was added to fermentation culture. To determine whether the labeled SZN was a single or a mixture of isotopologues, degradation (panel **b**) and MS/MS (panel **c**) experiments were performed. **b**, Exposing SZN to UV-light generated a one-carbon and one-nitrogen labeled cyclic urea previously reported to be a denitrosated SZN product, indicating that the distal nitroso nitrogen is labeled[9]. **c**, MS/MS fragmentation of SZN revealed a one-carbon labeled cyclic carbamate fragment, indicating that both the *N*-nitroso nitrogens are labeled.



Gene	Expected WT band size (kb)	Expected band size from mutants (kb)
sznE	1.2	1.5
sznF	No band	1.2
sznH	1.3	1.5
sznJ	1.0	1.5
sznK	1.3	1.5

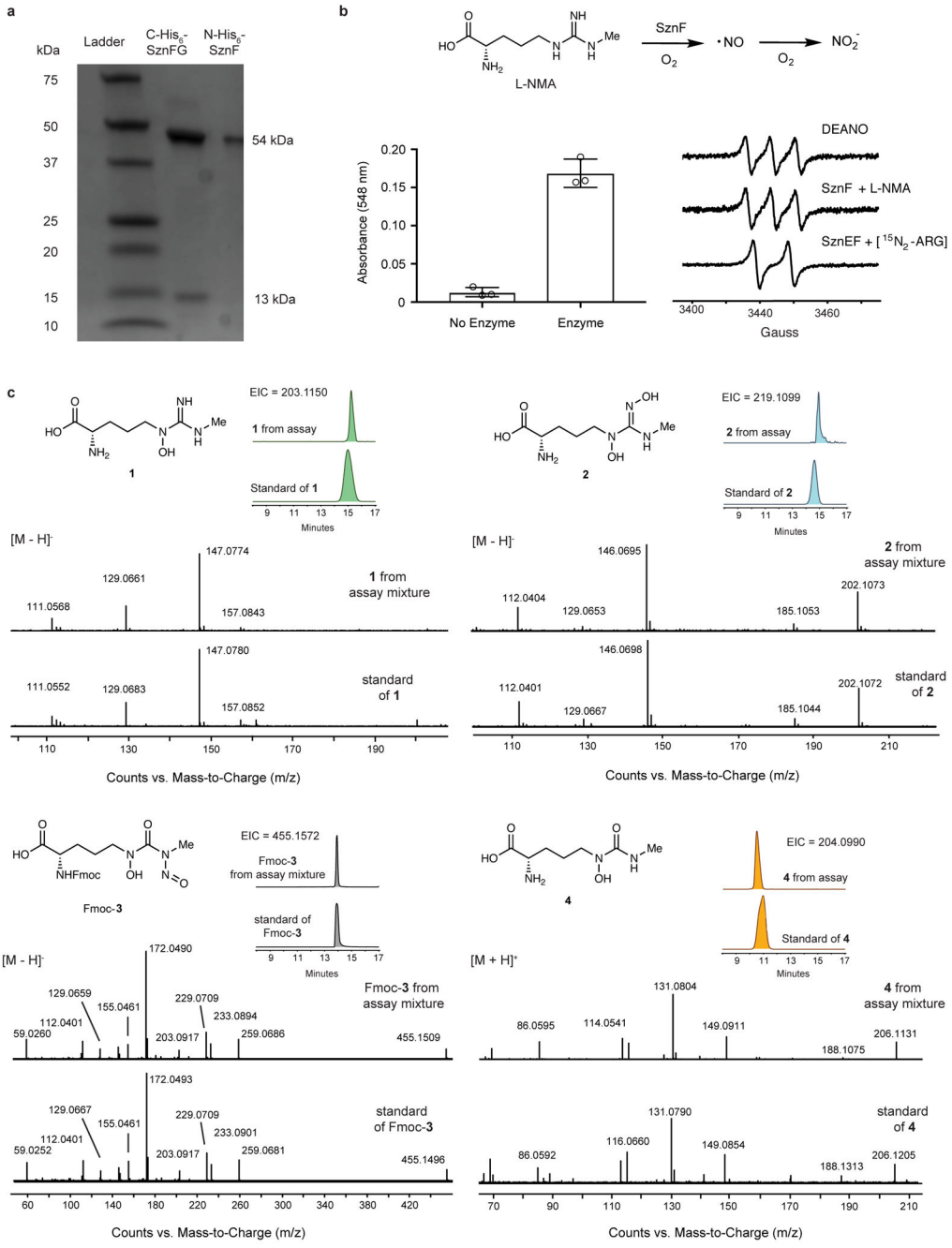






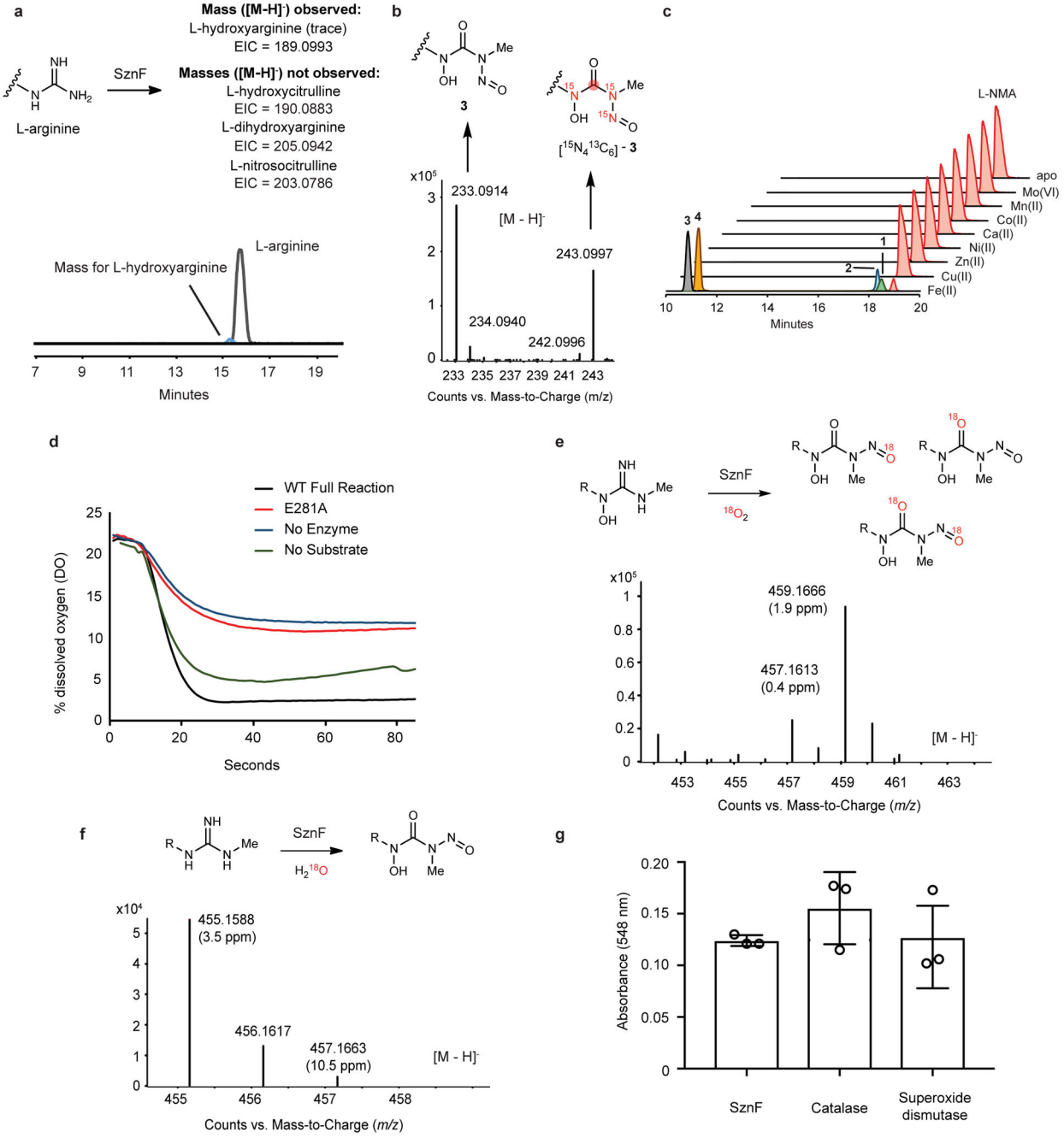
Extended Data Figure 4 |. Analysis of metabolite production by *szn* mutants.

Insertions of the antibiotic cassette into each of the *szn* biosynthetic genes were confirmed by PCR. Culture supernatant extracts from each mutant were analyzed with LC-HRMS. Extracted ion chromatograms for the amino acids ($[M+H]^+$) were generated with a 5 ppm window.



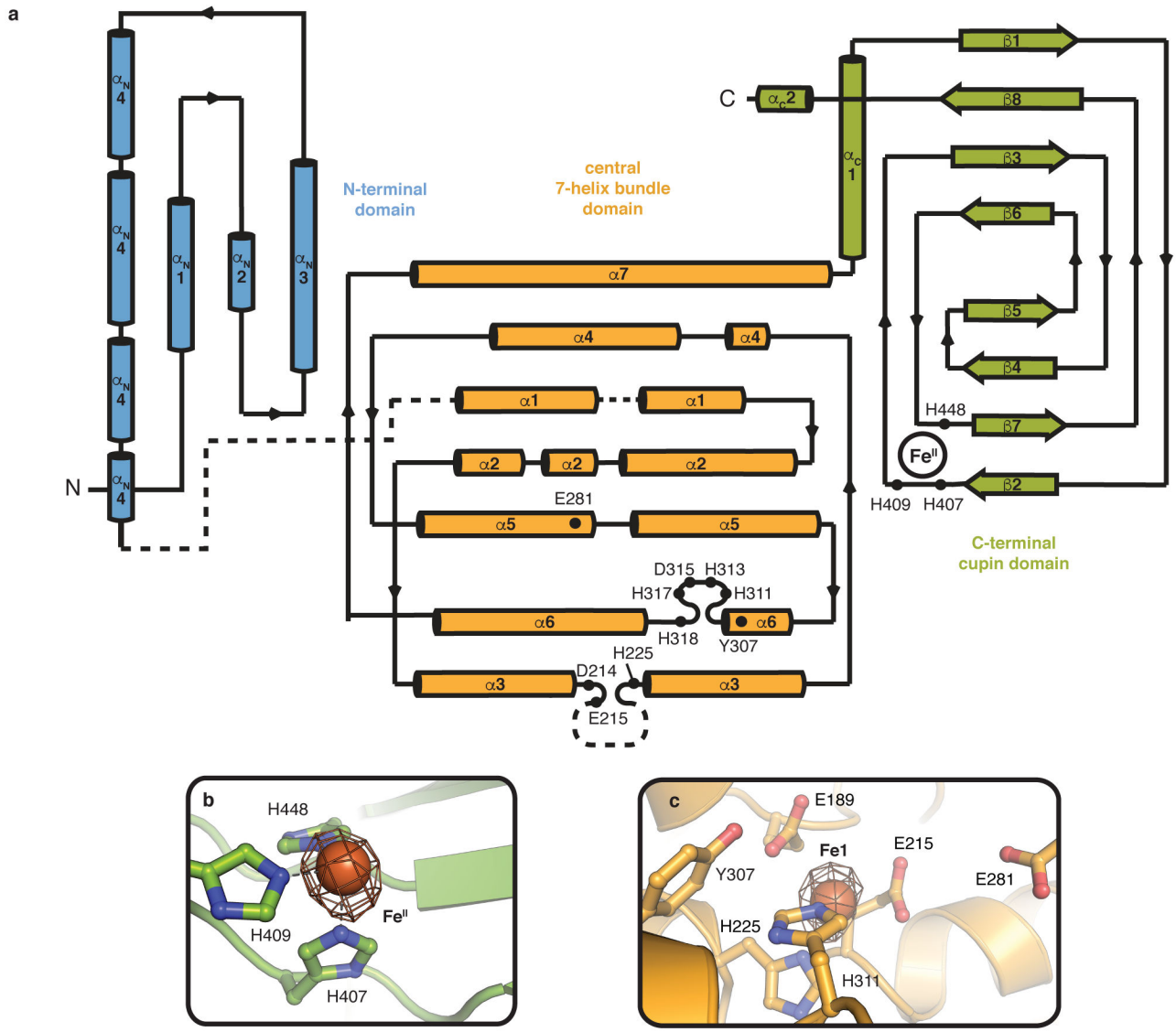
Extended Data Figure 5 | SznF generates an *N*-nitrosourea-containing amino acid.
a, SDS-PAGE of purified SznF and SznFG. The molecular weights of SznF and SznG are 54 kDa and 13 kDa, respectively. Ladder = Precision Plus Protein All Blue Standards (BioRad). **b**, Nitrite and nitric oxide (NO) were detected when Fe(II)-SznF and L-NMA were incubated together. Nitrite was detected with the Griess reagent, and absorbance was measured at 548 nm[38]. Error bars represent standard deviation of the mean of three replicates. NO was trapped with Fe(II)-*N*-methyl-D-glucamine dithiocarbamate (MGD) and analyzed by electron paramagnetic resonance (EPR) spectroscopy at room temperature[32]. Sodium 2-(*N, N*-diethylamino)-diazenolate-2-oxide (DEANO) was used as a positive

control for NO detection. Assays using [guanidino-$^{15}N_2$]-L-NMA as a substrate revealed changes in hyperfine splitting by EPR spectroscopy, indicating that NO is derived from the terminal guanidine group of L-NMA. We hypothesize that the NO detected derives from the degradation of **3** or is generated as part of the N–N bond formation step. **c**, Comparison of retention times and MS/MS fragmentation patterns of **1**, **2**, Fmoc-**3**, and **4** generated in SznF assay mixtures with the corresponding synthetic standards. NMR characterization and synthetic procedures are reported in the Supplementary Information.



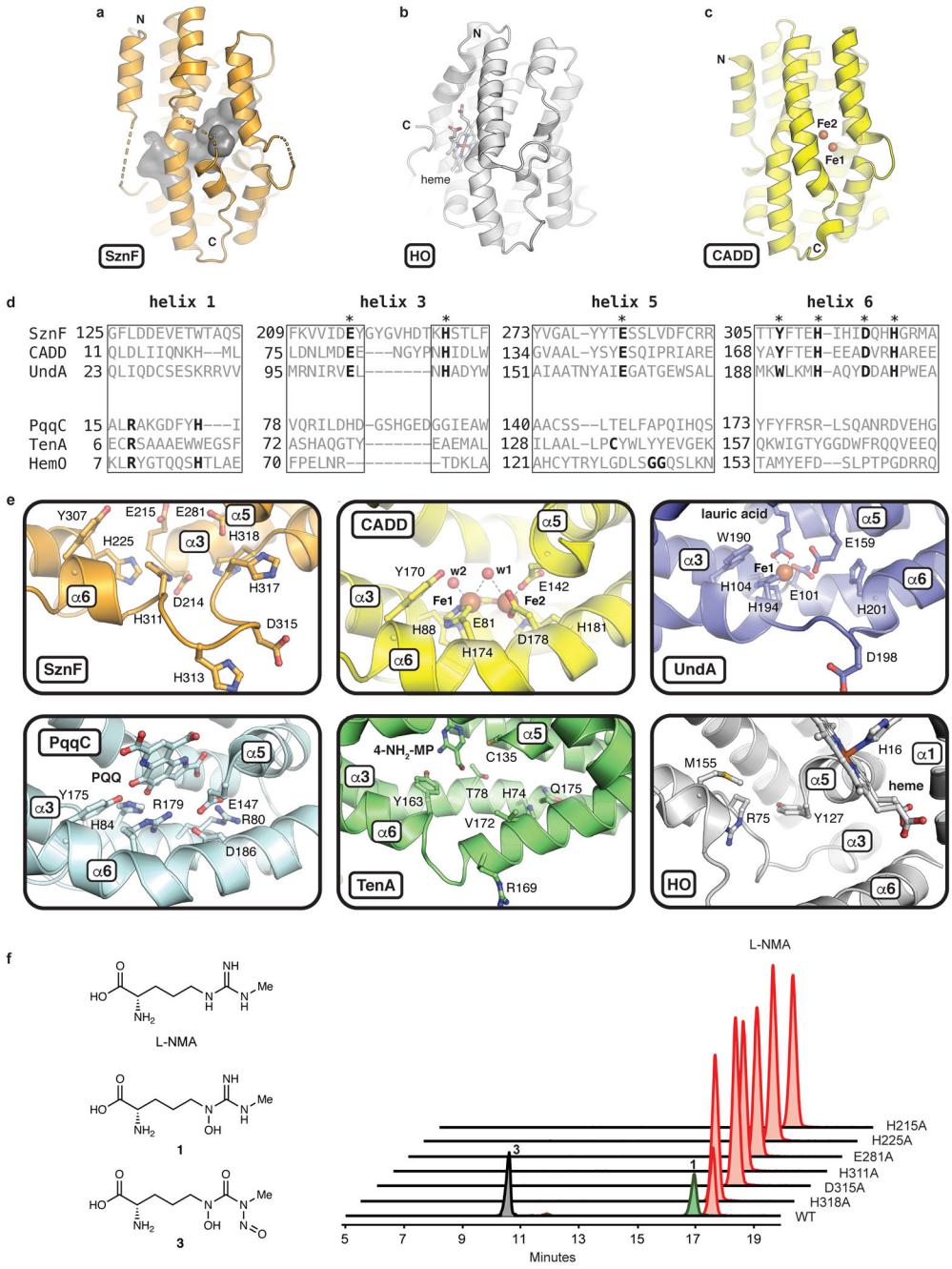
Extended Data Figure 6 |. SznF is an iron-dependent monooxygenase.

a, When 1 mM L-arginine, 80 μM SznF, 20 μM phenazine methosulfate (PMS), and 5 mM NADH were incubated at room temperature for one hour, only trace amounts of a mass corresponding to L-hydroxyarginine (EIC ($[M-H]^-$) = 189.0993) was observed. No masses corresponding to L-hydroxycitrulline (EIC ($[M-H]^-$) = 190.0883), L-dihydroxyarginine (EIC ($[M-H]^-$) = 205.0942), or the L-nitrosocitrulline (EIC($[M-H]$)$^-$ = 203.0786) were observed. **b**, The $[M-H]^-$ mass spectrum of **3** generated when [$^{15}N_4{}^{13}C_6$]-L-NMA and unlabeled L-NMA were mixed in the same SznF reaction mixture. **c**, Testing the metal dependence of SznF. 80 μM of *apo*-SznF was incubated with 200 μM of various divalent metals, 20 μM PMS, 1 mM L-NMA, and 5 mM NADH for one hour at room temperature. The EIC traces were generated with a 5 ppm window. **d**, O_2 was rapidly consumed in the presence of L-NMA and SznF as measured by an optode. SznF E281A, which lacks a key predicted iron-binding residue in the central domain, failed to consume O_2 above background. The background consumption of O_2 arises from the nonenzymatic reduction of PMS by NADH. **e**, Incubating $^{18}O_2$, 1 mM L-HMA (**1**), and 80 μM SznF at room temperature for one hour resulted in labeling of two of the *N*-nitrosourea oxygens. MS/MS analysis revealed retention of the N^{δ}-OH (data not shown). **f**, Addition of $H_2{}^{18}O$ to an SznF assay mixture did not label the *N*-nitrosourea group. The expected $[M-H]^-$ masses for Fmoc-**3**, [^{18}O]-Fmoc-**3**, [$^{18}O_2$]-Fmoc-**3**, and [$^{18}O_3$]-Fmoc-**3** are 455.1572, 457.1615, 459.1657, and 461.1700, respectively. **g**, Addition of catalase or superoxide dismutase to the assay mixtures did not affect SznF-catalyzed *N*-oxygenation as measured by the Griess assay. Error bars represent standard deviation of the mean of three replicates.



Extended Data Figure 7 | Topology diagram and Fe anomalous difference maps for putative SznF catalytic domains.

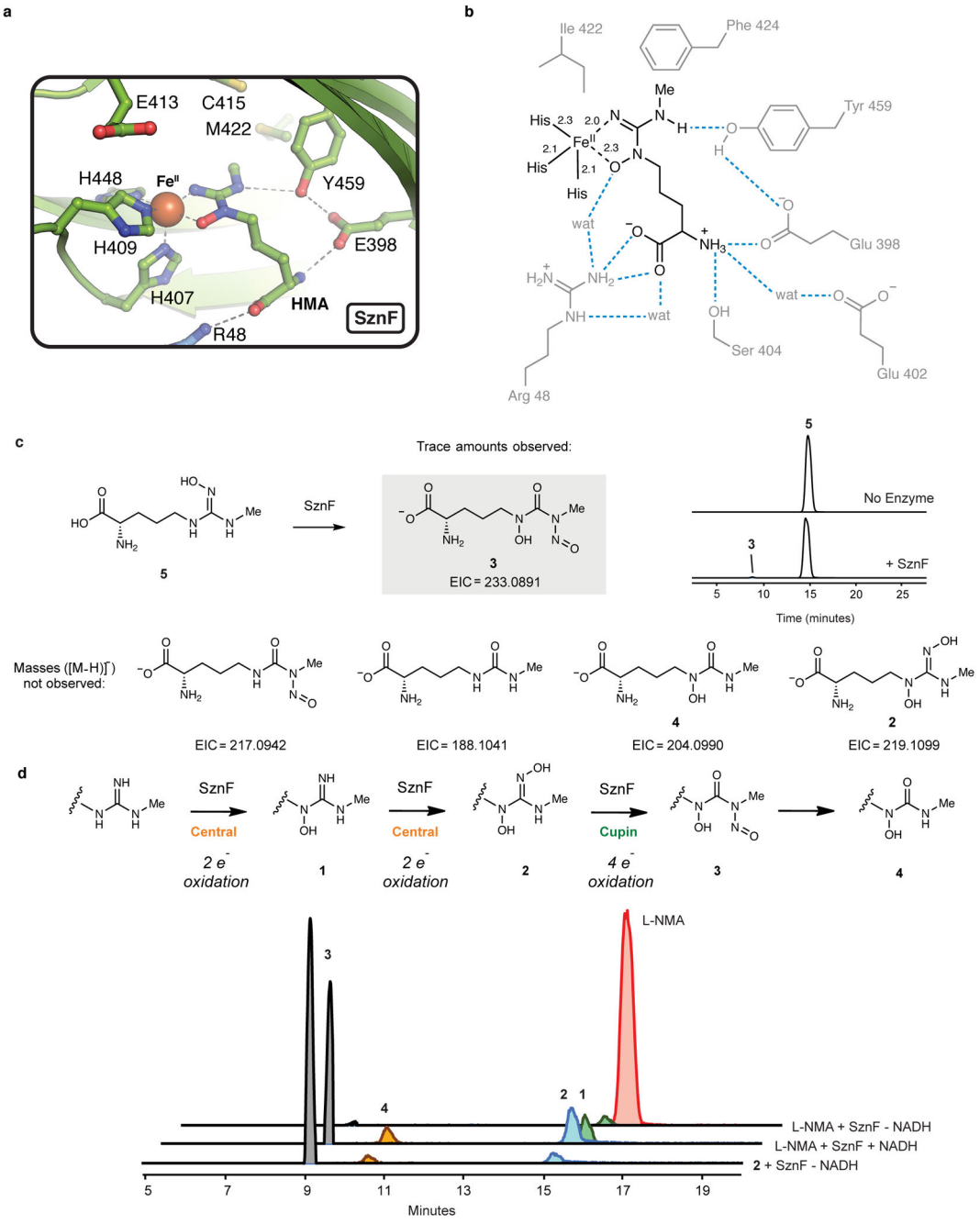
a, A diagram of the secondary structures found in the N-terminal domain (blue), central helical bundle domain (orange), and C-terminal cupin domain (green) of SznF. Fe^{II} ligands and proposed active site residues are indicated as black dots. Disordered regions are shown as dashed lines. The cupin Fe^{II}-binding site is depicted as a circle. **b**, An Fe anomalous difference map (orange mesh, contoured at 5.0σ) is shown for the fully-occupied mononuclear His-coordinated Fe^{II} site (orange sphere) in the cupin domain. Selected amino acids are shown in stick format. **c**, The central domain contains a partially occupied (~50%) Fe-binding site in selected crystals, with a smaller peak in the Fe anomalous difference map (orange mesh, contoured at 3.0σ).



Extended Data Figure 8 |. Comparison of the SznF central domain to heme/diiron structural homologs.

a, SznF contains a large cavity (gray surface, 1.9 Å probe radius) in the middle of its central helical bundle domain (orange). Additionally, most of the secondary structures in this domain contain loop disruptions and disordered regions, suggesting considerable refolding upon binding/release of the L-NMA substrate and/or assembly of the iron-based cofactor. **b**, The central domain of SznF is similar in topology to heme oxygenase (HO), compared here to HO-2 from *Synechocystis* sp. PCC 6803 (PDB accession code 1WOW)[39]. SznF contains an open pocket near the heme binding site in HO-2 but lacks conserved cofactor ligation and

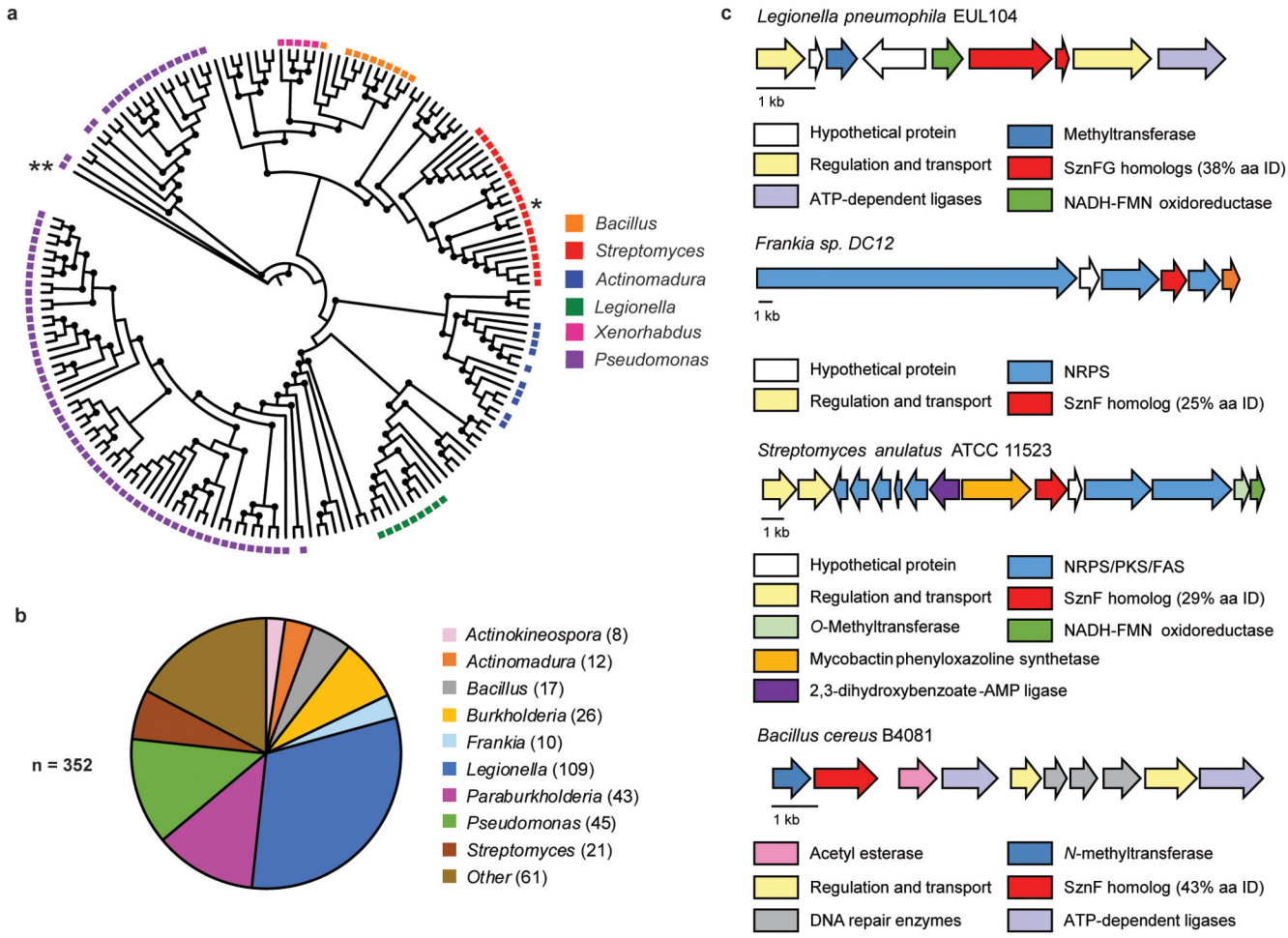
H-bonding motifs (**panel d**). **c**, SznF instead more closely resembles a *Chlamydia trachomatis* dinuclear iron protein in this structural superfamily (CADD)[21] implicated in *para*-aminobenzoic acid biosynthesis[40]. SznF conserves all of the metal-binding residues but fails to stably incorporate iron in this domain in the current preparations. All three systems share a propensity for distorted secondary structure motifs that perhaps enable complex formation with large and polar substrates for oxidative transformations. **d**, A structure-based sequence alignment[41] of six HO-like enzymes in selected regions relevant to substrate/cofactor interaction and catalysis. SznF conserves all six His/carboxylate residues used to coordinate a dinuclear iron cluster in the active form of fatty acid oxidative decarboxylase UndA (PDB accession code 4WWZ)[16] and in the uncharacterized CADD protein (PDB accession code 1RCW)[21]. Prior to our discovery and characterization of SznF, UndA was the only HO-like enzyme with a defined substrate and activity. The published crystal structure of UndA contains only a single iron ion and a mechanism was initially proposed utilizing a mononuclear cofactor (located in site 1)[16]. However, recent spectroscopic studies by Rajakovich et al.[20] show that this enzyme uses a dinuclear non-heme iron cofactor and corresponding alternative reaction pathway. To date, the dinuclear form of UndA has remained refractory to crystallographic characterization due to a propensity for disorder in the helix containing the site 2 metal ion ligands. As in SznF, mutagenesis of any of the six predicted ligands to the recently characterized dinuclear site in UndA completely abolishes activity.[20] As a consequence, we propose that all HO-like non-heme-iron proteins (including SznF) assemble a multinuclear cofactor but require a second protein or other factor to stabilize the active form in high yield and at high concentration. **e**, Comparative views of the cofactor site and/or substrate binding site in (left to right, top to bottom) SznF, *C. trachomatis* CADD, *Pseudomonas fluorescens* UndA, *Klebsiella pneumoniae* pyrroloquinoline quinone (PQQ) synthase PqqC (PDB accession code 1OTW)[42], *Synechocystis* sp. PCC 6803 heme oxygenase (HO) 2 (PDB accession code 1WOW)[39], and *Bacillus subtilis* thiamin synthase TenA (PDB accession code 1YAK)[43]. Substrates, products, and selected side chains are shown in stick format. Iron ions and water molecules are shown as orange and red spheres, respectively. **f**, Additional mutation of the predicted iron-binding residues in the SznF central bundle helix domain abolished *N*-oxygenation activity. Assay mixtures contained 1 mM L-NMA, 80 μM SznF or variant, 20 μM PMS, and 5 mM NADH and were incubated at room temperature for one hour. The EIC traces were generated with a 5 ppm window using the $[M-H]^-$ masses.



Extended Data Figure 9 |. The binding mode of 1 in the SznF C-terminal cupin domain and assays with a constitutional isomer suggest the N^{δ}-OH group is critical for the oxidative rearrangement.

a, An extended water-mediated H-bonding network tethers the non-metabolizable ligand **1** (green sticks, black lines) in the active site via its Me-N^{ω}, N^{δ}-O(H), and backbone amine/carboxylate functional groups. Selected SznF amino acids are shown in stick format in panel **a** and in gray lines in panel **b**. H-bonding/ionic interactions are shown as gray (panel **a**) or blue (panel **b**) dashed lines. Analysis of the network suggests a mechanism for deprotonation of **1** Me-N^{ω} via Y459 and E98. The cupin active site also contains an open

hydrophobic pocket near the unmethylated N^{ω} position. Apart from the aforementioned Y459 interaction, there are no H-bonds between the substrate functional groups and residues in the active site. **b**, Ligand interaction map showing Fe^{II}-coordination interactions (distances in Å) and H-bonding interactions with selected sidechains and water molecules. **c**, Use of the substrate analog N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (**5**) at 1 mM final concentration with 80 μM SznF, 20 μM PMS, and 5 mM NADH resulted in production of only trace amounts of **3** after incubation for one hour at room temperature. No $[M-H]^-$ masses corresponding to **2**, **3** without an N^{δ}-OH (EIC = 217.0942), **4**, or **4** without an N^{δ}-OH (EIC = 188.1041) were observed. **d**, The reaction catalyzed by the cupin domain does not require an external reductant. The conversion from **2** to **3** proceeded when 1 mM of **2** was incubated with 80 μM SznF without NADH at room temperature for one hour. The EIC traces were generated with a 5 ppm window.



Extended Data Figure 10 | Distribution of SznF homologs in microbial genomes.
a, Maximum-likelihood phylogenetic tree inferred from 50 replicates showing the relationship between selected SznF homologs containing both the central domain and cupin domain (NCBI "non-redundant protein sequences" database, 2018) (e-value < 1E-50). The branch corresponding to *S. achromogenes* SznF is highlighted by a single asterisk. Bootstrap

confidence values of >50 are indicated by black circles on the nodes. The amino acid sequence of UndA is used as an outgroup (highlighted as **). The sequences used to generate this tree are tabulated in Supplementary Table 2. **b**, Distribution of 352 SznF homologs that contain both a central domain and a cupin domain in different bacterial genera (IMG/JGI “all isolates” database, 2018) (e-value < 1E-5). See also Supplementary Table 3. **c**, Selected biosynthetic gene clusters encoding homologs of SznF.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement:

We acknowledge Jennifer X. Wang (Harvard University) for assistance with liquid chromatography-mass spectrometry method development and analysis, Matthew R. Wilson (Harvard University) for assistance with chemical synthesis and NMR characterization, Jonathan A. Bergman (Penn State University) for assistance with crystallography experiments, Lauren J. Rajakovich (Penn State University) for assistance with crystallography data analysis and reading the manuscript. We thank Prof. J. Martin Bollinger Jr. (Penn State University) and Prof. Carsten Krebs (Penn State University) for helpful discussions of proposed mechanisms and Prof. Wenjun Zhang (UC Berkeley) for providing *E. coli* WM6026. We acknowledge support from the National Institutes of Health (DP2 GM105434 to E.P.B. and GM119707 to A.K.B.), a Cottrell Scholar Award (to E.P.B.), a Camille Dreyfus Teacher-Scholar Award (to E.P.B.), the Searle Scholars Program (to A.K.B.), and Harvard University. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02–06CH11357. The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289–01A1). Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

References

1. Lundberg JO, Weitzberg E & Gladwin MT The nitrate-nitrite-nitric oxide pathway in physiology and therapeutics. Nat. Rev. Drug Discov 7, 156–167 (2008). [PubMed: 18167491]
2. Lundberg JO, Weitzberg E, Cole JA & Benjamin N Nitrate, bacteria and human health. Nat. Rev. Microbiol 2, 593–602 (2004). [PubMed: 15197394]
3. Krug S et al. Streptozocin-based chemotherapy in patients with advanced neuroendocrine neoplasms–predictive and prognostic markers for treatment stratification. PLoS One 10, e0143822, doi:10.1371/journal.pone.0143822 (2015). [PubMed: 26630134]
4. Weiss RB & Issell BF The nitrosoureas: carmustine (BCNU) and lomustine (CCNU). Cancer Treat. Rev 9, 313–330 (1982). [PubMed: 6762924]
5. Vavra JJ, Deboer C, Dietz A, Hanka LJ & Sokolski WT Streptozotocin, a new antibacterial antibiotic. Antibiot. Annu 7, 230–235 (1959). [PubMed: 13841501]
6. Wu KK & Huan Y Streptozotocin-induced diabetic models in mice and rats. Curr. Protoc. Pharmacol Chapter 5, Unit 5.47, doi:10.1002/0471141755.ph0547s40 (2008).
7. Pathak S, Dorfmueller HC, Borodkin VS & van Aalten DM Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death. Chem. Biol 15, 799–807 (2008). [PubMed: 18721751]
8. Bolzan AD & Bianchi MS Genotoxicity of streptozotocin. Mutat. Res 512, 121–134 (2002). [PubMed: 12464347]
9. Singaram S, Lawrence RS & Hornemann U Studies on the biosynthesis of the antibiotic streptozotocin (streptozocin) by Streptomyces achromogenes var. streptozoticus. Feeding experiments with carbon-14 and tritium labelled precursors. J. Antibiot. (Tokyo) 32, 379–385 (1979). [PubMed: 157345]

10. Vitturi DA et al. Convergence of biological nitration and nitrosation via symmetrical nitrous anhydride. Nat. Chem. Biol 11, 504–510 (2015). [PubMed: 26006011]
11. Fu D, Calvo JA & Samson LD Balancing repair and tolerance of DNA damage caused by alkylating agents. Nat. Rev. Cancer 12, 104–120 (2012). [PubMed: 22237395]
12. Bergy ME et al. Streptozotocin and its production. U.S. Patent 3027300A (1962).
13. Komor AJ, Jasniewski AJ, Que L & Lipscomb JD Diiron monooxygenases in natural product biosynthesis. Nat. Prod. Rep 35, 646–659 (2018). [PubMed: 29552683]
14. Lee J, Simurdiak M & Zhao H Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation. J. Biol. Chem 280, 36719–36727 (2005). [PubMed: 16150698]
15. Wang C et al. Structural determinants for the strict monomethylation activity by trypanosoma brucei protein arginine methyltransferase 7. Structure 22, 756–768 (2014). [PubMed: 24726341]
16. Rui Z et al. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. Proc. Natl. Acad. Sci. U. S. A 111, 18237–18242 (2014). [PubMed: 25489112]
17. Crane BR, Sudhamsu J & Patel BA Bacterial nitric oxide synthases. Annu. Rev. Biochem 79, 445–470 (2010). [PubMed: 20370423]
18. Liebschner D et al. Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr. D. Struct. Biol 73, 148–157 (2017). [PubMed: 28177311]
19. Poulos TL Heme enzyme structure and function. Chem. Rev 114, 3919–3962 (2014). [PubMed: 24400737]
20. Rajakovich LJ *Exploring the functional and mechanistic diversity of diiron oxidases and oxygenases* Ph.D. thesis, The Pennsylvania State University, (2017).
21. Schwarzenbacher R et al. Structure of the *Chlamydia* protein CADD reveals a redox enzyme that modulates host cell apoptosis. J. Biol. Chem. 279, 29320–29324 (2004). [PubMed: 15087448]
22. Merkx M et al. Dioxygen Activation and Methane Hydroxylation by Soluble Methane Monooxygenase: A Tale of Two Irons and Three Proteins. Angew. Chem. Int. Ed. Engl 40, 2782–2807 (2001).
23. Kal S, Que L Dioxygen activation by nonheme iron enzymes with the 2−His−1−carboxylate facial triad that generate high−valent oxoiron oxidants. J. Biol. Inorg. Chem 22, 339–365 (2017). [PubMed: 28074299]
24. Caranto JD, Vilbert AC & Lancaster KM *Nitrosomonas europaea* cytochrome P460 is a direct link between nitrification and nitrous oxide emission. Proc. Natl. Acad. Sci. U. S. A 113, 14704–14709 (2016). [PubMed: 27856762]
25. Möenne-Loccoz P Spectroscopic characterization of heme iron-nitrosyl species and their role in NO reductase mechanisms in diiron proteins. Nat. Prod. Rep 24, 610–620 (2007). [PubMed: 17534533]
26. Seemann T Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014). [PubMed: 24642063]
27. Darling AC, Mau B, Blattner FR & Perna NT Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403 (2004). [PubMed: 15231754]
28. Soding J, Biegert A & Lupas AN The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–248 (2005). [PubMed: 15980461]
29. Larkin MA et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948 (2007). [PubMed: 17846036]
30. Eswar N et al. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics Chapter 5, Unit-5.6, doi:10.1002/0471250953.bi0506s15 (2006).
31. Gust B, Challis GL, Fowler K, Kieser T & Chater KF PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. U. S. A 100, 1541–1546 (2003). [PubMed: 12563033]
32. Xia Y & Zweier JL Direct measurement of nitric oxide generation from nitric oxide synthase. Proc. Natl. Acad. Sci. U. S. A 94, 12705–12710 (1997). [PubMed: 9356514]
33. Kunkel TA Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. U. S. A 82, 488–492 (1985). [PubMed: 3881765]

34. Kumar S, Stecher G, Li M, Knyaz C & Tamura K MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol 35, 1547–1549 (2018). [PubMed: 29722887]

35. Markowitz VM et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40, D115–D122 (2012). [PubMed: 22194640]

36. Winter JM, Jansma AL, Handel TM & Moore BS Formation of the pyridazine natural product azamerone by biosynthetic rearrangement of an aryl diazoketone. Angew. Chem. Int. Ed. Engl 48, 767–770 (2009). [PubMed: 19072974]

37. Wang K-KA et al. Glutamic acid is a carrier for hydrazine during the biosyntheses of fosfazinomycin and kinamycin. Nat. Commun 9, 3687, doi:10.1038/s41467-018-06083-7 (2018). [PubMed: 30206228]

38. Forist AA Spectrophotometric determination of streptozotocin. Anal. Chem 36, 1338–1339 (1964).

39. Sugishima M et al. Crystal structure of dimeric heme oxygenase-2 from Synechocystis sp. PCC 6803 in complex with heme. Biochemistry 44, 4257–4266 (2005). [PubMed: 15766254]

40. Adams NE et al. Promiscuous and adaptable enzymes fill “holes” in the tetrahydrofolate pathway in Chlamydia species. MBio 5, e01378–01314 (2014). [PubMed: 25006229]

41. Pei J, Kim BH, Tang M & Grishin NV PROMALS web server for accurate multiple protein sequence alignments. Nucleic Acids Res. 35, W649–652 (2007). [PubMed: 17452345]

42. Magnusson OT et al. Quinone biogenesis: structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc. Natl. Acad. Sci. U. S. A 101, 7913–7918 (2004). [PubMed: 15148379]

43. Toms AV, Haas AL, Park JH, Begley TP & Ealick SE Structural characterization of the regulatory proteins TenA and TenI from Bacillus subtilis and identification of TenA as a thiaminase II. Biochemistry 44, 2319–2329 (2005). [PubMed: 15709744]

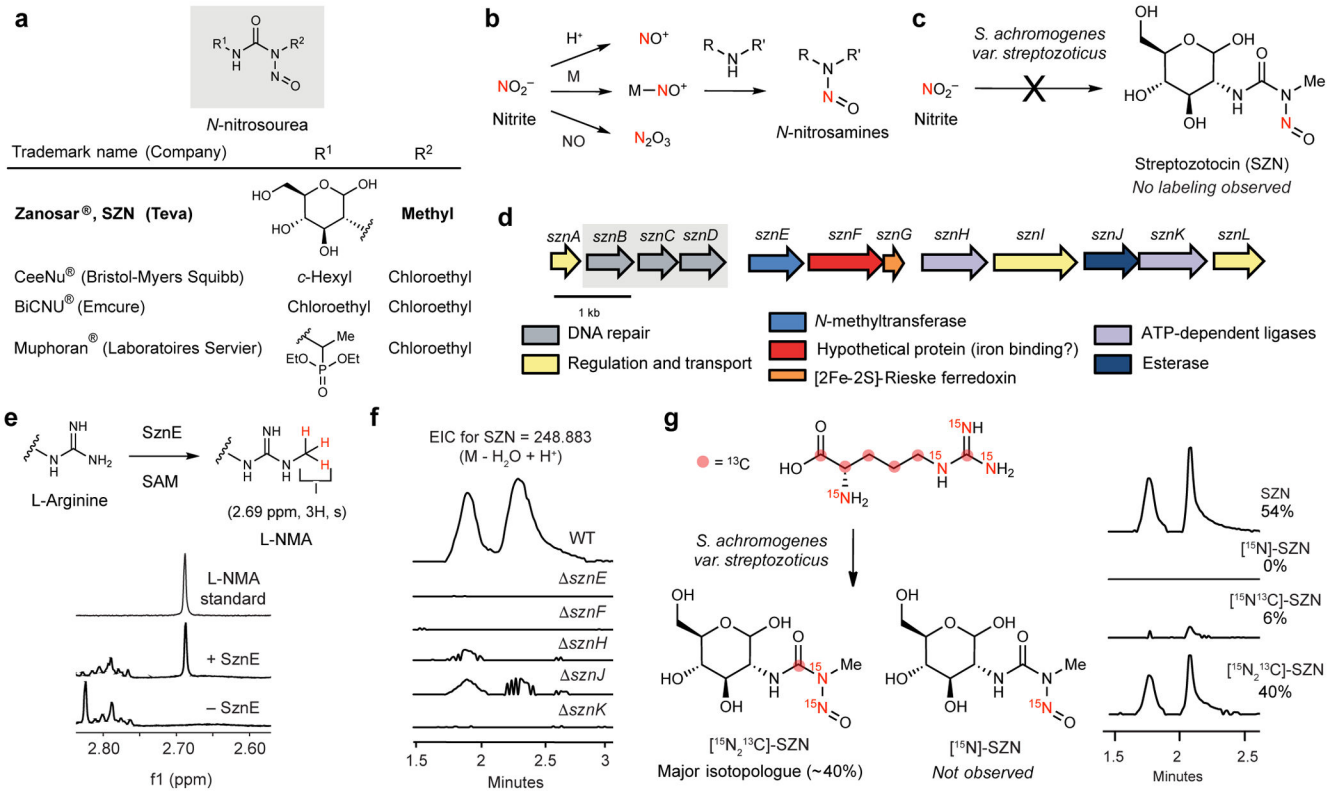


Figure 1: The streptozotocin biosynthetic gene cluster encodes a unique pathway for *N*-nitrosation.

a, The structures of selected FDA-approved *N*-nitrosourea-containing drugs. **b**, Known strategies for *N*-nitrosation in biological systems. M = metalloenzyme. **c**, Feeding ^{15}N-nitrite to the SZN producer did not result in labeled SZN (**Mass spectra in** Extended Data Fig. 1b). **d**, The *szn* biosynthetic gene cluster. Genes encoding DNA repair enzymes are highlighted in grey. **e**, ^{1}H NMR assay shows that SznE uses *S*-adenosyl-L-methionine (SAM) as a methyl donor to convert L-arginine to N^{ω}-methyl-L-arginine (L-NMA) *in vitro*. **f**, Gene inactivation studies of the *szn* gene cluster. Extracted ion chromatograms (EICs) for SZN. The two peaks correspond to the α- and β-anomers. **g**, Feeding labeled L-arginine to the producing organism suggests the transfer of the intact guanidine group to the *N*-nitrosourea. EIC ($[M\text{-}H_2O+H]^+$) for SZN = 248.0877, [^{15}N]-SZN = 249.0847, [^{13}C^{15}N]-SZN = 250.0877, and [^{13}C^{15}N$_2$]-SZN = 251.0847. See Extended Data Fig. 3 for assignment of the positions of carbon and nitrogen labels.

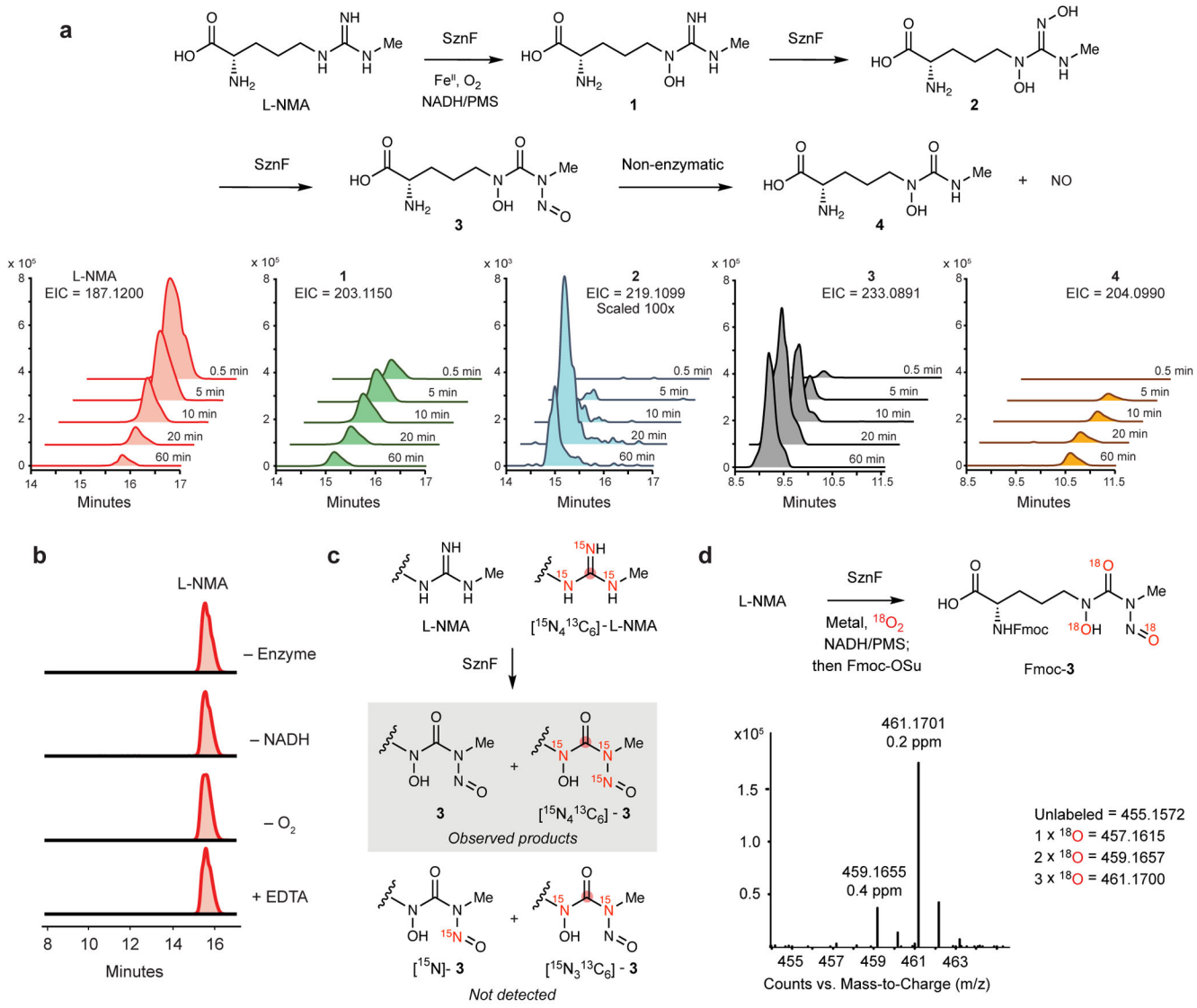


Figure 2 |. SznF catalyzes the oxidative rearrangement of L-NMA to generate an *N*-nitrosourea. **a**, LC-MS time course of the SznF-mediated transformation of L-NMA to *N*-nitrosourea **3**. EICs are shown. The structures of **1**–**4** were confirmed by comparison to synthetic standards (Extended Data Fig. 5c). **b**, LC-MS traces of SznF *in vitro* assays defining the requirements for activity. EICs are shown. **c**, Crossover experiment showed that SznF catalyzes rearrangement of an intact guanidine group. See Extended Data Fig. 6b for mass spectrum. **d**, ^{18}O labeling experiment revealed molecular O_2 as the source of the three oxygen atoms in *N*-nitrosourea **3**.

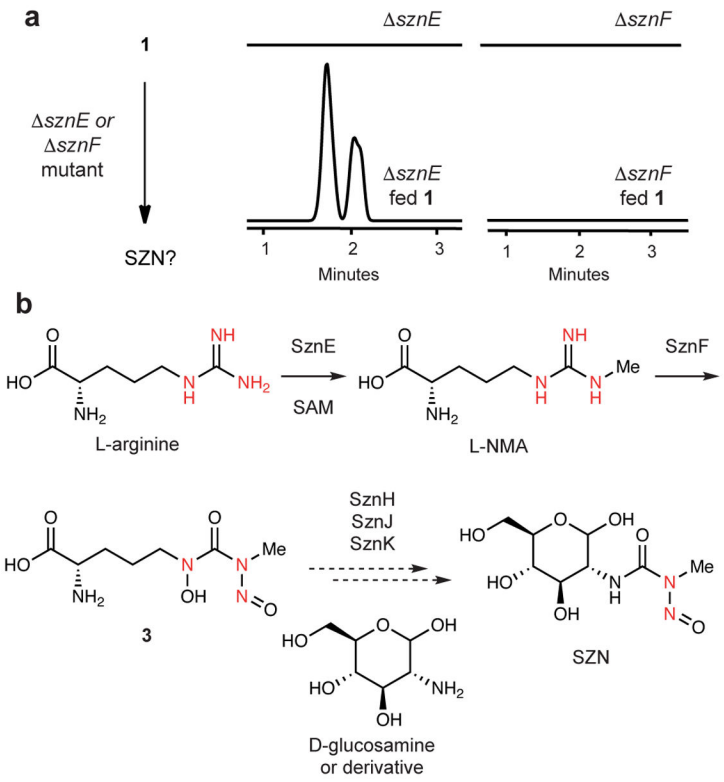


Figure 3 |. The reaction catalyzed by SznF is essential for SZN production *in vivo*.

a, SznF-derived intermediate **1** complements the *sznE* mutant but not the *sznF* mutant. EICs for SZN ($[M\text{-}H_2O+H]^+$ = 248.0883) are shown. **b**, Proposed pathway for SZN biosynthesis.

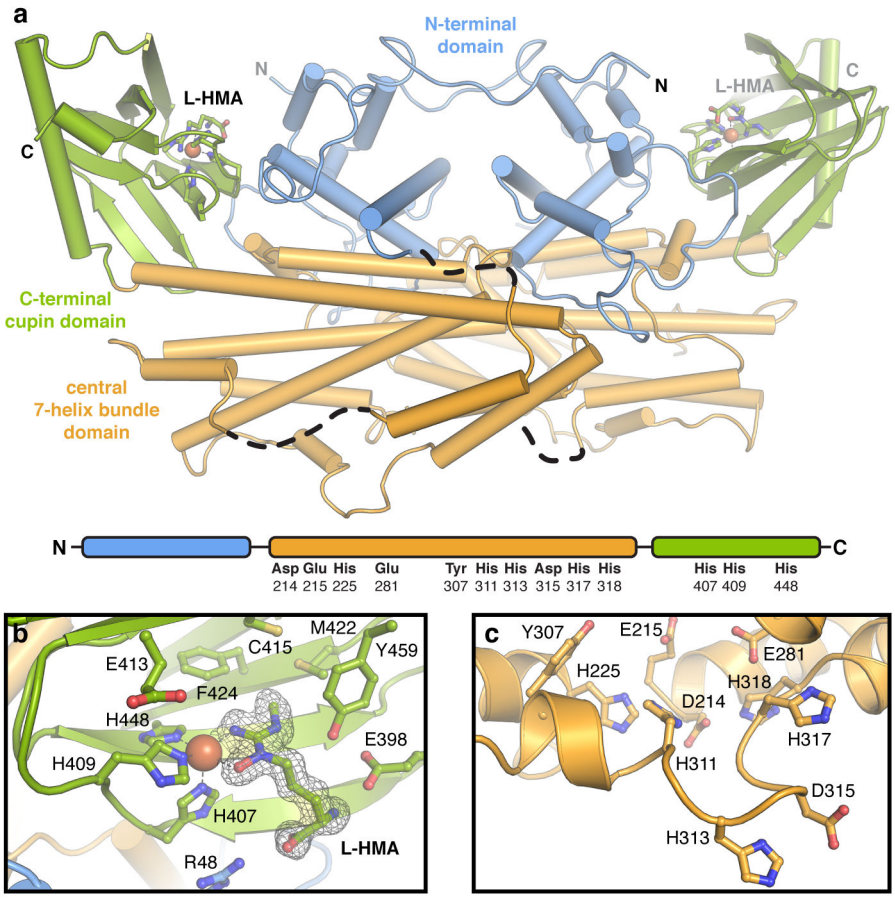


Figure 4 |. The x-ray crystal structure of SznF.

a, The structure of the SznF homodimer, colored by domain, reveals two candidate active sites. Disordered regions are shown as black dashed lines. C-terminal Fe^{II} site is shown as an orange sphere and bound intermediate L-HMA (**1**) is shown in stick format. Predicted metal-binding residues are highlighted on the cartoon schematic at bottom. **b**, A mononuclear His-coordinated Fe^{II} site (orange sphere) in the cupin domain recruits intermediate L-HMA (green sticks) as a bidentate ligand. Selected amino acids are shown in stick format and a polder omit electron density map is shown in gray mesh and contoured at 3.0σ[18]. **c**, The central domain lacks full occupancy metal ions in the crystals but conserves a series of His/carboxylate ligands found in related diiron proteins. See also Extended Data Fig. 7, 8.

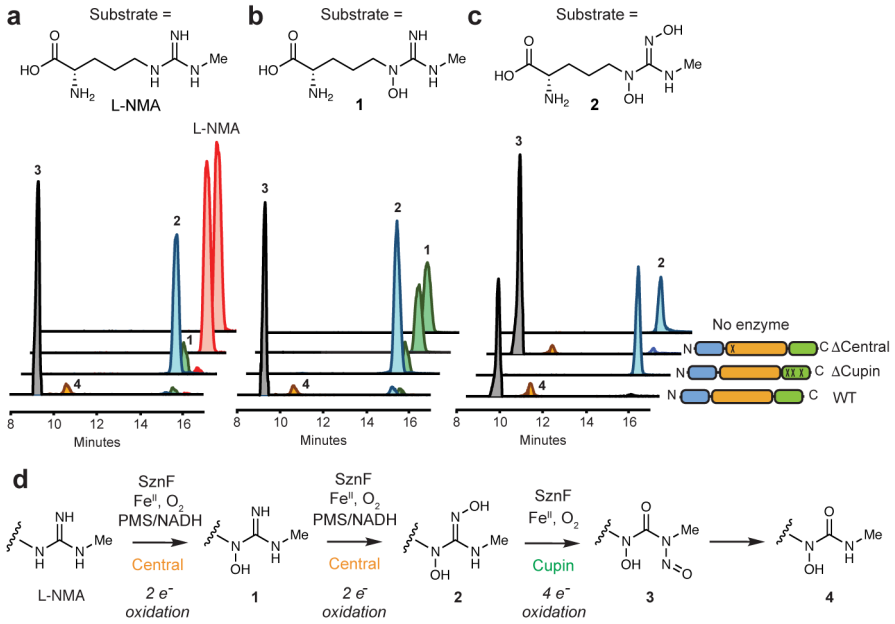


Figure 5 | The two metallocofactor sites in SznF are functionally distinct.

To test the reactivity of the separate domains, L-NMA (panel **a**), **1** (panel **b**), or **2** (panel **c**) were incubated with no enzyme, an SznF mutant with an inactive central domain (E215A), an SznF mutant with an inactive cupin domain (H407A H409A H448A), or WT SznF. EICs were generated within 5 ppm. **d**, The reactions catalyzed by the two metallocofactor binding sites of SznF.