An \(N\)-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin

Tai L. Ng1, Roman Rohac2, Andrew J. Mitchell1, Amie K. Boal2,3* & Emily P. Balskus1*

Small molecules containing the \(N\)-nitroso group, such as the bacterial natural product streptozotocin, are prominent carcinogens1,2 and important cancer chemotherapeutics3,4. Despite the considerable importance of this functional group to human health, enzymes dedicated to the assembly of the \(N\)-nitroso unit have not been identified. Here we show that SznE, a metalloenzyme from the biosynthesis of streptozotocin, catalyses an oxidative rearrangement of the guanidine group of \(N\)-methyl-\(L\)-arginine to generate an \(N\)-nitrosourea product. Structural characterization and mutagenesis of SznF reveal two separate active sites that promote distinct steps in this transformation using different iron-containing metallofactors. This biosynthetic reaction, which has little precedent in enzymology or organic synthesis, expands the catalytic capabilities of non-haem-iron-dependent enzymes to include \(N\)--\(N\) bond formation. We find that biosynthetic gene clusters that encode SznF homologues are widely distributed among bacteria—including environmental organisms, plant symbionts and human pathogens— which suggests an unexpectedly diverse and uncharacterized microbial reservoir of bioactive \(N\)-nitroso metabolites.

Streptozotocin (SZN, also known by the trade name Zanosar) is an \(N\)-nitrosourea natural product and approved cancer chemotherapeutic1,2 (Fig. 1a). SZN is also used to induce type I diabetes in animal models owing to its toxicity towards pancreatic beta cells5. Like other \(N\)-nitrosoureas, SZN exerts its activity in vivo by

---

1Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. 2Department of Chemistry, The Pennsylvania State University, University Park, PA, USA. 3Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA. *e-mail: albs20@psu.edu; balskus@chemistry.harvard.edu
generating electrophilic DNA-alkylating agents and nitric oxide, a precursor to reactive oxygen species\(^7,8\). As yet, 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)\(^9\) (Extended Data Fig. 1a). Although biosynthesis of the N-nitroso group was not investigated, the distal nitroso nitrogen atom was proposed to come from nitrite. All characterized routes for N-nitrosation in vivo use nitrite (Fig. 1b) and are not known to be catalysed by dedicated enzymes\(^1,10\). Notably, when we fed 15N-labelled nitrite, nitrate or ammonium salts to *S. achromogenes* 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)\(^9\) (Extended Data Fig. 1a). Although biosynthesis of the N-nitroso group was not investigated, the distal nitroso nitrogen atom was proposed to come from nitrite. All characterized routes for N-nitrosation in vivo use nitrite (Fig. 1b) and are not known to be catalysed by dedicated enzymes\(^1,10\). Notably, when we fed 15N-labelled nitrite, nitrate or ammonium salts to *S. achromogenes* NRRL 2697, we did not observe isotopically labelled SZN (Fig. 1c, Extended Data Fig. 1b). N-nitrosation in the assembly of SZN must therefore use a distinct biosynthetic strategy.

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

To validate the function of this gene cluster, we examined the N-methyltransferase SznE. Homology modelling of SznE revealed a resemblance to eukaryotic protein l-arginine N-methyltransferases\(^15\) (Extended Data Fig. 2b, c). Incubating purified SznE with l-arginine and S-adenosyl-l-methionine (SAM) resulted in the formation of N\(^\circ\)′-methyl-l-arginine (l-NMA) (Fig. 1e, Extended Data Fig. 2d), which suggests that l-NMA was a probable biosynthetic intermediate in the formation of SZN. Indeed, feeding \(d_1\)-l-NMA to *S. achromogenes*
To identify additional product(s) generated by SznF, we analysed assay mixtures by liquid chromatography coupled with mass spectrometry (LC–MS) and characterized four new compounds by comparison to synthetic standards (Fig. 2a, Extended Data Fig. 5c). SznF first hydroxylates l-NMA at the N\(^{-}\)position to afford N\(^{-}\)-hydroxy-N\(^{-}\)methyl-l-arginine (1). 1 is further hydroxylated to give N\(^{-}\)-hydroxy-N\(^{-}\)-hydroxy-N\(^{-}\)methyl-l-arginine (2). Subsequently, an oxidative rearrangement converts 2 to N\(^{-}\)-hydroxy-N\(^{-}\)-methyl-N\(^{-}\)-nitroso-l-citrulline (3). This proposed biosynthetic intermediate degrades in a non-enzymatic manner to the denitrosated product 4 and nitric oxide. The N\(^{-}\)-hydroxylation in this reaction sequence was unanticipated, as this nitrogen atom is not incorporated into SZN. SznF activity requires Fe\(^{ii}\), oxygen and external reductants (Fig. 2b), and l-arginine was not accepted as a substrate (Extended Data Fig. 6a).

Next, we further defined the role of oxygen in this transformation. The addition of l-NMA to SznF assay mixtures accelerated the consumption of oxygen (Extended Data Fig. 6d). All three N-nitrosourea oxygen atoms in 3 were labelled by \(^{18}\)O\(_{2}\) (Fig. 2d, Extended Data Fig. 6e), whereas 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 monoxygenase. The two new oxygen 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 that of haem-dependent nitric oxide synthase, which oxidizes the guanidine group of l-arginine to generate nitric oxide and l-citrulline\(^{17}\). Nitric oxide synthase also hydroxylates a guanidine nitrogen, and the ureido oxygen atom of l-citrulline is derived from O\(_{2}\). However, the involvement of non-haem iron and oxygen.

To connect the in vitro activity of SznF to SZN production in vivo, we fed intermediate 1 to ΔsznE and ΔsznF mutant strains. Feeding 1 to the ΔsznE strain restored SZN production, confirming that this compound is an on-pathway intermediate (Fig. 3a). However, SznF was not detected when 1 was fed to the ΔsznF mutant, which indicates that additional processing of this intermediate by SznF is required. We also generated mutant strains that lacked predicted downstream bio-enzymes (ΔsznH, ΔsznJ and ΔsznK). The deletion of sznK completely abolished the production of SZN, and only trace amounts of SZN were observed with both ΔsznH and ΔsznJ mutants (Fig. 1f). Although the unstable N-nitrosourea 3 did not accumulate in these strains, we observed increased levels of decomposition product 4 in extracts from mutants compared to the wild type (Extended Data Fig. 4). Together, these results implicate the activity of SznF in the biosynthesis of SZN, and suggest that SznH, SznJ and SznK may transfer Ω-methyl-l-arginine (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 SznF–SeMet. A dataset at a resolution of 2.08 Å revealed a dimer with three distinct domains (Fig. 4a, Extended Data 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 produced Δ3-SZN (Extended Data Fig. 2e). SZN production was also completely abolished in a mutant in which sznE was deleted (ΔsznE, Fig. 1f), whereas biosynthesis was restored with l-NMA (Extended Data Fig. 2f). To connect the in vitro activity of SznF to SZN production in vivo, we fed intermediate 1 to ΔsznE and ΔsznF mutant strains. Feeding 1 to the ΔsznE strain restored SZN production, confirming that this compound is an on-pathway intermediate (Fig. 3a). However, SznF was not detected when 1 was fed to the ΔsznF mutant, which indicates that additional processing of this intermediate by SznF is required. We also generated mutant strains that lacked predicted downstream bio-enzymes (ΔsznH, ΔsznJ and ΔsznK). The deletion of sznK completely abolished the production of SZN, and only trace amounts of SZN were observed with both ΔsznH and ΔsznJ mutants (Fig. 1f). Although the unstable N-nitrosourea 3 did not accumulate in these strains, we observed increased levels of decomposition product 4 in extracts from mutants compared to the wild type (Extended Data Fig. 4). Together, these results implicate the activity of SznF in the biosynthesis of SZN, and suggest that SznH, SznJ and SznK may transfer Ω-methyl-l-arginine (3) to the d-glucosamine scaffold (Fig. 3b).
haem oxygenase\(^1\) (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 mononuclear iron centres, respectively, but the structure and oxidation state of the active metallocofactor remains undefined\(^20,21\). Selected datasets of SznF exhibit a single full-occupancy metal ion in the crystals, but a series of histidine and carboxylate ligands that are found in related diiron proteins are conserved.

Whereas 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) shows that the modified l-Arg side chain is coordinated to the cupin Fe\(^{II}\) via the unmethylated N\(^\delta\) and the newly installed N\(^\omega\)-OH. Although 1 is not processed by the cupin domain, this coordination mode suggests that the N\(^\delta\)-hydroxy group is important for Fe\(^{II}\) binding. Indeed, assays with N\(^\delta\)-hydroxy-N\(^\omega\)-methyl-l-arginine (5) indicate that the N\(^\delta\)-OH is required for the oxidative rearrangement (Extended Data Fig. 9c).

Although the central domain of SznF requires NADH, the cupin domain catalyses 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.

Fig. 4 | The X-ray crystal structure of SznF. a, The structure of the SznF homodimer, coloured by domain, reveals two candidate active sites. Disordered regions are shown as black dashed lines. The C-terminal Fe\(^{II}\) site is shown as an orange sphere and the bound intermediate l-HMA (1) is shown in stick format. Predicted metal-binding residues are highlighted on the cartoon schematic at the 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 grey mesh and contoured at 3.0 \(\sigma\). c, The central domain lacks full-occupancy metal ions in the crystals, but a series of histidine and carboxylate ligands that are found in related diiron proteins are conserved. See also Extended Data Figs. 7, 8.
diazo line as a key intermediate (Supplementary Discussion). There is precedent for similar reactivity in N–N bond formation catalysed by haem enzymes\textsuperscript{24,25} including NO reductase, which utilizes non-haem Fe\textsuperscript{2+} and haem cofactors to generate N\textsubscript{2}O. However, the role of each site in N–N bond formation is unknown\textsuperscript{25}. Notably, 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 revealed SznF homologues containing both the central and cupin domains encoded in numerous cryptic biosynthetic gene clusters (\(E < 1 \times 10^{-5}\)) (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 \textit{Burkholderia} and \textit{Paraburkholderia} strains, organisms known to make \(N\)--\(N\) nitroso natural products\textsuperscript{26}. The plant symbiont \textit{Frankia} and the human pathogen \textit{Legionella pneumophila} also possess SznF-encoding gene clusters (Extended Data Fig. 10c). This analysis suggests that diverse bacteria produce \(N\)--\(N\) nitroso-containing compounds, highlighting potential new biological roles for these metabolites and enabling the discovery of natural products based on genome mining.

In studying the biosynthesis of the pancreatic cancer drug SZN, we have revealed 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-enzyme transformations in living systems. Our work demonstrates that enzymes have evolved to synthesize \(N\)-nitroso compounds. The unexpectedly widespread distribution of SznF homologues in bacteria suggests that the biological importance of \(N\)-nitroso metabolites, including their roles in symbioses and pathogen–host interactions, is underestimated.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0894-z.

Received: 17 August 2018; Accepted: 12 December 2018; Published online 6 February 2019.
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.* **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).

24. Caranto, J. D., Vilbert, A. C. & Lancaster, K. M. Nitrosomonas europaea cytochrome P460 is a direct link between nitrification and nitrous oxide emission. *Proc. Natl Acad. Sci. USA* **113**, 14704–14709 (2016).

25. Molène-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).

26. Hermeau, R. et al. Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nat. Chem. Biol.* **14**, 841–843 (2018).

Acknowledgements We thank J. Wang for assistance with LC–MS method development and analysis, M. Wilson for assistance with chemical synthesis and NMR characterization, J. Bergman for assistance with crystallography experiments, and L. Rajakovich for assistance with crystallography data analysis and reading the manuscript. We thank J. M. Bollinger Jr. and C. Krebs for discussions of proposed mechanisms and W. Zhang 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 US 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).

Author contributions 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 analysed and discussed the results and prepared the manuscript.

Competing interests The authors declare no competing interests.

Additional information Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0894-z.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0894-z.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.K.B. or E.P.B.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Materials and general methods. Oligonucleotide primers were synthesized by Integrated DNA Technologies and Sigma-Aldrich. Recombinant plasmid DNA was purified with a Quickprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean-up were performed using Zymo Research gel DNA Recovery Kit and DNA Clean & Concentrator kit from Zymo Research. DNA sequencing was performed by Genewiz and Eton Bioscience. Nickel-nitrotriacetic 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 absorbance 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 °C with a flow rate of 8 l min$^{-1}$ 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 were recorded in either positive or negative ionization mode with a mass range of 100 to 1,700 m/z; spectrometer, 1 spectrum per second; capillary voltage, 3,500 V; nebulizer pressure, 35 psi; drying gas (N$_2$) flow, 8 l min$^{-1}$; temperature, 275 °C. For MS2 settings: mass range, 100 to 1,700 m/z; collision energy, 10 eV; isolation width MS/MS, around 0.1% m/z; positive or negative ionization mode with a mass range of 100 to 1,700 m/z; spectra rate, 1 spectrum per second; temperature, 275 °C. A mass window of 5 p.p.m. was used to extract the ion counts for generating extracted ion chromatograms with Agilent ChemStation software. The specific LC methods for separating SNZ and amino acids are described below.

**Cultivation of Streptomyces achromogenes var. streptozoticus NRRL 2697**

**Detection of streptozotocin.** Streptomyces achromogenes var. 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$^{-1}$ soy flour, 20 g l$^{-1}$ mannitol and 2 g l$^{-1}$ Bacto agar (Difco)) at 30 °C for five days. The spores were scraped and inoculated into 5 ml of tryptic soy broth medium (TSB) at 30°C under shaking (150 rpm). This starter culture was inoculated into 100 ml production medium in a 250 ml baffled flask (40 g l$^{-1}$ yellow corn meal (Arrowhead Mills), 15 g l$^{-1}$ soy flour, 20 g l$^{-1}$ potato starch (Bob's Red Mill), 1 g l$^{-1}$ peptone (Bacto) and 14 g l$^{-1}$ N$_2$O$_4$. The fermentation was carried out for 5 days at 30 °C with shaking at 200 r.p.m. 1 g l$^{-1}$ calf-intestinal alkaline phosphatase was added to the fermentation medium. At the end of fermentation, the presence of Streptozotocin was detected as the molecular ions [M$^+$H]$^+$ or [M–H$_2$O$+H]^+$.

**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. 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 Phylostomys Genome Browser. 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 ControlCopy HTTP Fosmid Library Production Kit (Epigenet) following the manufacturer's protocol. A library of 2,000 clones was picked into 96-well plates and stored at −80 °C 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 °C as described above. After 16 h, stock solutions of [15$^N$]calcium nitrate, [15$^N$]sodium nitrate and [15$^N$]ammonium chloride (Cambridge Isotope Laboratories) in water were sterilized by passing through a 0.22-μm filter membrane. Each nitrogen source was added at the concentration of 50 μM to a series of cultures to a final concentration of 1 mM. After four more days of fermentation at 30°C, the presence of labelled and unlabelled SNZ 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 (SNZ) biosynthetic gene cluster.** A BLAST search using *E. coli* DNA repair enzymes AlkA (annotation: DNA-3-methylenediacetylalumylene glycosylase 2, National Center for Biotechnology Information (NCBI) GenBank accession: AUY16819.1) revealed three homologues in the *S. achromogenes var. streptozoticus* genome. Comparison of the *S. achromogenes var. streptozoticus* and *S. achromogenes var. NRRL B-2120* (GenBank Assembly accession number: GCA_000720835.1) genomes was performed using Mauve alignment software. The region encoding DNA repair enzyme homologues and putative biosynthetic enzymes that is absent from the non-producing strain is depicted in Extended Data Fig. 2a. Annotations of the open reading frames in the *snz* gene cluster are found in Supplementary Table 1.

**Homology modelling and sequence alignment of SznE.** Structural homologues of SznE were identified using HHPred (http://toolkit.tuebingen.mpg.de/HHpred) (26). 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 were recorded in either positive or negative ionization mode with a mass range of 100 to 1,700 m/z; spectrometer, 1 spectrum per second; capillary voltage, 3,500 V; nebulizer pressure, 35 psi; drying gas (N$_2$) flow, 8 l min$^{-1}$; precursor threshold, 200 counts; temperature, 275 °C. A mass window of 5 p.p.m. was used to extract the ion counts for generating extracted ion chromatograms with Agilent ChemStation software. The specific LC methods for separating SNZ and amino acids are described below.

**Cultivation of Streptomyces achromogenes var. streptozoticus NRRL 2697 and detection of streptozotocin.** Streptomyces achromogenes var. 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$^{-1}$ soy flour, 20 g l$^{-1}$ mannitol and 2 g l$^{-1}$ Bacto agar (Difco)) at 30 °C for five days. The spores were scraped and inoculated into 5 ml of tryptic soy broth medium (TSB) at 30°C under shaking (150 rpm). This starter culture was inoculated into 100 ml production medium in a 250 ml baffled flask (40 g l$^{-1}$ yellow corn meal (Arrowhead Mills), 15 g l$^{-1}$ potato starch (Bob's Red Mill), 1 g l$^{-1}$ glucose, 3 g l$^{-1}$ peptide (Bacto) and 14 g l$^{-1}$ ammonium sulfate). The fermentation was carried out for 5 days at 30 °C with shaking at 200 r.p.m. 750 μl of the culture was centrifuged at 16,100 g 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 16,100g, and the acetone was removed under vacuum. The samples were analysed with liquid chromatography coupled with high-resolution mass spectrometry (LC–HRMS).

Streptozotocin was detected as the molecular ions [M+H]$^+$ or [M–H$_2$O+H]$^+$.

Mass spectra data were collected only between 1.4 and 5 min, with the LC stream diverted to waste between 1.4 min and after 5 min. The LC column was an Acclaim Polar Advantage II C18 column (3 μm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 ml min$^{-1}$. 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 elution at 95% solvent A (solvent A – 0.1% formic acid in water; solvent B – 0.1% formic acid in acetonitrile).
HEPES, 50 mM NaCl, 10% glycerol, pH 8.0) was added, and the sample was concentrated again to 1 ml. This process was repeated once more before the concentrated, desalted solution containing purified SznE was frozen in liquid N2 and stored at −80 °C.

1H NMR assay of SznE activity. In a 500 μl reaction mixture, 50 mM potassium phosphate pH 8.0, 10 mM MgCl2, 2 mM L-arginine, 4 mM SAM and 5 μM of SznE were mixed and incubated at room temperature overnight. Negative control assays omitting SAM, L-arginine or SznE in the reaction were also performed. The reaction mixtures were flash-frozen with liquid N2 and lyophilized. The residues were resuspended in D2O (Cambridge Isotope Laboratories) and analysed with 1H NMR spectroscopy using an Agilent DD2-600 NMR spectrometer (600 MHz). Chemical shifts (δ) are reported in parts per million (p.p.m.) downfield from tetramethylsilane using the solvent resonance as an internal standard for 1H (D2O = 4.79 p.p.m.). This experiment was performed in triplicate.

Feeding experiments with labelled arginine and degradation studies. Streptomyces achromogenes var. streptociticus. NRRRL 2697 was grown in 10 ml of fermentation medium in 50 ml baffled flasks while shaking at 30 °C as described above. Stock solutions of [15N2]l-NMA, [15N3]l,N3C13]-l-arginine, [15N2]l-alanine or [d15]-N2- methyl-l-arginine were sterilized by passing through a 0.22-μm filter membrane and added to the fermentation culture 16 h after inoculation to give a final concentration of 1 mM. After two more days of fermentation at 30 °C, the presence of labelled and unlabelled Szn was determined by LC–HRMS as described above. The positions of the labelled nitrogen and carbon atoms were elucidated using degradation experiments as described in Extended Data Fig. 3.

Gene disruption and chemical complementation experiments. Gene inactivation in Streptomyces achromogenes var. streptociticus. NRRRL 2697 was performed according to standard protocols in brief, primers used to amplify SznE were used to screen the fosmid library to obtain a fosmid containing the szn cluster. The fosmid was transformed into E. coli BW25113/pKD46 (Coli Genetic Stock Collection) by electroporation. The aac(3)IV-orf1 cassette was amplified by PCR from pJ773 using the primers listed in Supplementary Table 5. Each of the biosynthetic genes was replaced with the aac(3)IV-orf1 cassette PCR product using PCR targeting and λ-red-mediated recombination. The mutant fosmid was used to transform into E. coli WM6026 for conjugation with S. achromogenes var. achromogenes NRRRL 2697. Double crossover mutants were selected with apramycin resistance on mannitol-soy agar, and the exconjugants were grown in tryptic soy broth for 3–5 days at 30 °C as described above. Stock solutions of EDTA pH 7 was added to a final concentration of 0.5 mM to remove any residual Ni bound to Szn. This solution was concentrated to about 400 μl and diluted into 12 ml of exchange buffer containing 0.5 mM EDTA. The protein solution was concentrated once more to obtain a desalted EDTA-treated Szn sample. An aliquot of this solution was submitted to the University of Georgia Center for Applied Isotope Studies for inductively-coupled plasma-mass spectrometry (ICP-MS) analysis. This Szn sample had <0.04 p.p.m. Fe3+, Cu2+ and Co2+ 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 (NH4)2Fe(SO4)2·6H2O, MnCl2·4H2O, CaCl2·2H2O, ZnCl2, CuCl2·2H2O, CoCl2·6H2O, Na2MoO4·2H2O and NiSO4·6H2O were prepared in 1 M aqueous HCl. Assay mixtures contained 50 mM MOPS pH 7.5, 1 mM t-NMA, 200 μM metal, 200 μM PMS, 5 mM NADH and 100 μM SznE in a final volume of 0.1 ml was incubated for 4 h before quenching with methanol and analysing with LC–HRMS.

Griess assay for nitrite production and nitric oxide detection by electron paramagnetic resonance spectroscopy. 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% (naphthyl)ethylenediamine dihydrochloride in 30% acetic acid (solution B) were prepared. 25 μl of solution A, 25 μl of solution B, 10 μl of HCl and 10 μl of the enzyme assay mixtures were combined and heated at 60 °C 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 ref. 33. In brief, a 10× 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)Fe2+ 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 cuvette and analysed with a Bruker ElexsysE580 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; 3.418 G centre field; 82 ms time constant. An enzyme-free sample containing 1 mM sodium 2-(N,N-diethylamino)-diazeneolate-2-oxide (DEANO) was used as a positive control. To demonstrate that the nitric oxide was generated from the terminal nitrogen atoms of t-NMA, 1 mM of [15N2]l-arginine (Cambridge Isotope Laboratories) was first converted to [15N2]l-NMA with SznE as described in the section ‘1H NMR assay of SznE activity’. After 1 h, SznF, PMS and NADH were added directly to the reaction mixture to the same final concentrations as described in the section ‘LC–MS assay of SznF activity’. (MGD)Fe2+ was then added, and the SznE assay mixture was analysed by the same procedure to remeasure label-free nitric oxide production.

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 per 50 μl before 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 a 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 saturated with sodium 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 analysed using NeoFox Viewer version 2.40. The different experimental conditions tested are described in Extended Data Fig. 6d. The experiments were performed at least twice on different days.

**H2O2 and H18O18O labelling experiment.** Solutions of buffer, MgCl2 and water were degassed 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 MgCl2, degassed water, PMS and NADH were brought into an M Braun glove box under an atmosphere consisting of 99.997% nitrogen (N2) with less than 0.1 p.p.m. O2. 50 mM MOPS pH 8.0, 10 mM MgCl2, 1 mM L-NMA or 1 mM L-arginine, which generated the corresponding labelled products (data not shown). LC–MS analysis in Fig. 1f is representative of more than three individual experiments.

**Supplementary Figs. 1–11.** The labelling studies and chemical complementation experiments shown in Extended Data Fig. 2e, f were repeated independently two years apart. The experiment shown in Extended Data Fig. 3a, b were repeated twice, and the analysis for Extended Data Fig. 3c was performed once. At least two con jugants were picked for each mutant for the PCR analysis in Extended Data Fig. 4a, and the metabolite analysis has been performed at least three times independently with similar results. Purification of SznF and SznFPG has been replicated several times by the research groups of both corresponding authors (Extended Data Fig. 5a).

The detection of NO by EPR was performed in duplicate on the same day. The LC–MS analysis of the labelled products 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 Fmo-3, and twice for 4 using positive and negative ion mode. The experiments for Extended Data Fig. 6a, c, e, f were performed once, and the O2 consumption assay (Extended Data Fig. 6d) was performed twice. The activity assay shown in Extended Data Fig. 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 Fig. 9c was repeated once, and the activity assay depicted in Extended Data Fig. 9d was repeated on different days with similar results. The NMR spectra supplied in the Supplementary Figures were recorded once.

**Data availability.** The nucleotide sequences for the znf biosynthetic gene cluster and individual genes have been deposited into the NCBI (GenBank accession number for the znf gene cluster: MK365372; Genbank accession numbers for SznA–SznL: MK291255–MK291266). Structural factors and coordinates of SznF have been deposited in the Protein Data Bank (PDB: 6H9R, 6H9S). Additional data that support the conclusions of the paper can be requested from the corresponding authors.

28. Darling, A. C., Mau, B., Blatter, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403 (2004).

29. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive protein server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W249 (2005).

30. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948 (2007).

31. Eswar, N. et al. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinform. 15, 5.6.1–5.6.30 (2006).

32. Gust, B., Chaliss, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geomisin. Proc. Natl Acad. Sci. USA 100, 1541–1546 (2003).

33. Xia, Y. & Zwiebler, J. L. Direct measurement of nitric oxide generation from nitric oxide synthase. Proc. Natl Acad. Sci. USA 94, 12705–12710 (1997).

34. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl Acad. Sci. USA 82, 488–492 (1985).

35. 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).

36. Markowitz, V. M. et al. IMG: The Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 40, D115–D122 (2012).

37. Winter, J. M., Jansma, A. L., Handel, T. M. & Moore, B. S. Formation of the pyrroloquinoline quinone synthase. Mol. Biol. Evol. 26, 1541–1546 (2009).

38. Sugishita, M. et al. Crystal structure of dimeric heme oxygenase-2 from Synechosystis sp. PCC 6803 in complex with hemin. Biochemistry 44, 4257–4266 (2005).

39. Adams, N. E. et al. Promiscuous and adaptable enzymes fill “holes” in the tetrahydrofolate pathway in Chlamydia tetrahydrofolate pathways. Nucleic Acids Res. 35, W649–W652 (2007).

40. Magnuson, D. T. et al. Quinone biosynthesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc. Natl Acad. Sci. USA 101, 7913–7918 (2004).

41. Adams, N. E. et al. Promiscuous and adaptable enzymes fill “holes” in the tetrahydrofolate pathway in Chlamydia tetrahydrofolate pathways. Nucleic Acids Res. 35, W649–W652 (2007).

42. Magoc, S. et al. Quinone biosynthesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc. Natl Acad. Sci. USA 101, 7913–7918 (2004).

43. Tomis, A. V., Haas, A. L., Park, J. H., Begley, T. P. & Ealick, S. E. Structural LC–MS analysis of the protein product of TenA and TenB from Bacillus subtilis and identification of TenA as a thiamine I. Biochemistry 44, 2319–2329 (2005).
Extended Data Fig. 1 | Inorganic nitrogen sources are not precursors to SZN. a. The results of previous feeding experiments suggested that D-glucosamine, L-citrulline or L-arginine, and L-methionine-derived 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 labelling (>75%) observed in studies of pathways that use nitrite for diazo biosynthesis\(^{37,38}\).
Extended Data Fig. 2 | l-NMA is an on-pathway intermediate in SZN biosynthesis. a, Comparative genomic analysis of S. achromogenes var. achromogenes NRRL B-2120, a non-producer of SZN and S. achromogenes var. streptozoticus NRRL 2697 using Mauve28. The szn gene cluster is coloured 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: 4M37) (orange). The highlighted carboxylate residues are involved in binding of the basic guanidine group15. 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 d1-l-NMA. The expected masses [M − H2O + H]+ for SZN and d1-SZN are 248.0883 and 251.1060, respectively. f, LC–MS traces demonstrating the restoration of SZN production by the ΔsznE mutant upon chemical complementation with l-NMA. The EICs are generated within a 5 p.p.m. window.
Extended Data Fig. 3 | The N-nitrosourea of SZN is derived from an intact guanidine group of L-arginine. a, The mass spectrum of SZN \([M - H_2O + H]^{+}\) when 1 mM of \([^{15}N_4^{13}C_6]\)L-arginine was added to the fermentation culture. To determine whether the labelled SZN was a single isotopologue or a mixture, degradation (b) and MS/MS (c) experiments were performed. b, Exposure of SZN to ultraviolet light generated a one-carbon- and a one-nitrogen-labelled cyclic urea that was previously reported to be a denitrosated SZN product, indicating that the distal nitroso nitrogen is labelled\(^9\). c, MS/MS fragmentation of SZN revealed a one-carbon-labelled cyclic carbamate fragment, indicating that both of the N-nitroso nitrogens are labelled.
Extended Data Fig. 4 | Analysis of metabolite production by \( \text{szn} \) mutants. Insertions of the antibiotic cassette into each of the \( \text{szn} \) biosynthetic genes were confirmed by PCR. Culture supernatant extracts from each mutant were analysed with LC–HRMS. Extracted ion chromatograms for the amino acids ([M + H]+) were generated with a 5 p.p.m. window.
Extended Data Fig. 5  See next page for caption.
Extended Data Fig. 5 | SznF generates an N-nitrosocontaining amino acid. a, SDS–PAGE of purified SznF and SznF–SznG complex. 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 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$^{39}$. Data are mean ± s.d. of three replicates. NO was trapped with MGD and analysed by EPR spectroscopy at room temperature$^{33}$. 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 propose that the NO detected is derived 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 Supplementary Information.
**Extended Data Fig. 6 | SznF is an iron-dependent monoxygenase.**

a, When 1 mM l-arginine, 80 μM SznF, 20 μM PMS and 5 mM NADH were incubated at room temperature for 1 h, only trace amounts of a mass corresponding to l-hydroxyarginine (EIC ([M-H]−) = 189.0993) were 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 [15N4,13C6]l-NMA and unlabelled 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 1 h at room temperature. The EIC traces were generated with a 5 p.p.m. window. d, Oxygen 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 oxygen above background. The background consumption of oxygen arises from the non-enzymatic reduction of PMS by NADH. e, Incubating 18O2, 1 mM l-HMA (1) and 80 μM SznF at room temperature for 1 h resulted in labelling of two of the N-nitrosourea oxygens. MS/MS analysis revealed retention of the Nδ-OH (data not shown). f, Addition of H218O to an SznF assay mixture did not label the N-nitrosourea group. The expected [M-H]− masses for Fmoc-3, [18O]Fmoc-3, [18O2]Fmoc-3 and [18O3]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-catalysed N-oxygenation as measured by the Griess assay. Data are mean ± s.d. of three replicates.
Extended Data Fig. 7 | Topology diagram and iron 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 histidine-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 (around 50%) iron-binding site in selected crystals, with a smaller peak in the iron anomalous difference map (orange mesh, contoured at 3.0σ).
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Comparison of the SznF central domain to haem and diiron structural homologues. a, SznF contains a large cavity (grey 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 or 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 haem oxygenase (HO), compared here to HO-2 from *Synechocystis* sp. PCC 6803 (PDB: 1WOW)\(^4\). SznF contains an open pocket near the haem-binding site in HO-2 but lacks conserved cofactor ligation and hydrogen-bonding motifs (d). c, SznF instead more closely resembles a *C. trachomatis* dinuclear iron protein in this structural superfamily (CADD)\(^2\) implicated in the biosynthesis of *para*-aminobenzoic acid\(^4\). 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. 

The published crystal structure of UndA contains only a single iron ion and a mechanism was initially proposed using a mononuclear cofactor (located in site 1)\(^1\). However, recent spectroscopic studies\(^20\) show that this enzyme uses a dinuclear non-haem iron cofactor and corresponding alternative reaction pathway. So far, the dinuclear form of UndA has remained refractory to crystallographic characterization owing 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-haem-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: 1OTW)\(^4\), *Synechocystis* sp. PCC 6803 haem oxygenase (HO) 2 (PDB: 1WOW)\(^4\), and *Bacillus subtilis* thiamin synthase TenA (PDB: 1YAK)\(^4\). 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 1 h. The EIC traces were generated with a 5 p.p.m. window using the [M − H]\(^−\) masses.
Extended Data Fig. 9 | See next page for caption.
The binding mode of 1 in the SznF C-terminal cupin domain and assays with a constitutional isomer suggest that the $N^\delta$–OH group is critical for the oxidative rearrangement. a, An extended water-mediated hydrogen-bonding network tethers the non-metabolizable ligand 1 (green sticks, black lines) in the active site via its Me–$N^\omega$, $N^\delta$–O(H), and backbone amine and carboxylate functional groups. Selected SznF amino acids are shown in stick format in a and in grey lines in b. Hydrogen-bonding and ionic interactions are shown as grey (a) or blue (b) dashed lines. Analysis of the network suggests a mechanism for deprotonation of 1 Me-$N^\omega$ via Y459 and E98. The cupin active site also contains an open hydrophobic pocket near the unmethylated $N^\omega$ position. Apart from the aforementioned Y459 interaction, there are no hydrogen bonds between the substrate functional groups directly involved in the rearrangement reaction and residues in the active site. b, Ligand-interaction map showing Fe$^{II}$-coordination interactions (distances in Å) and hydrogen-bonding interactions with selected side chains and water molecules. c, Use of the substrate analogue $N^\omega$-hydroxy-$N^\omega$-methyl-$l$-arginine (5) at 1 mM final concentration with 80 $\mu$M SznF; 20 $\mu$M PMS and 5 mM NADH resulted in production of only trace amounts of 3 after incubation for 1 h at room temperature. No [M − H]$^-$ masses corresponding to 2, 3 without an $N^\delta$–OH (EIC = 217.0942), 4, or 4 without an $N^\delta$–OH (EIC = 188.1041) were observed. d, The reaction catalysed 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 $\mu$M SznF without NADH at room temperature for 1 h. The EIC traces were generated with a 5 p.p.m. window.
Extended Data Fig. 10 | Distribution of SznF homologues in microbial genomes. **a**, Maximum-likelihood phylogenetic tree inferred from 50 replicates showing the relationship between selected SznF homologues containing both the central domain and cupin domain (NCBI Non-Redundant Protein Sequences database, 2018) ($E < 1 \times 10^{-50}$). 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 homologues that contain both a central domain and a cupin domain in different bacterial genera (IMG/JGI 'all isolates' database, 2018) ($E < 1 \times 10^{-5}$). See also Supplementary Table 3. **c**, Selected biosynthetic gene clusters that encode homologues of SznF.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- n/a | Confirmed
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
- Clearly defined error bars
  State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
- Geneious Pro 467 Version 7.1.6 (Biomatters) was used for pBLAST searches and sequence alignments.
- MEGA-X was used to construct phylogenetic tree.
- MODELLER (Max Planck Institute for Developmental Biology) was used for building homology models.
- ProDRG program was used to build the HMA ligand
- Data collection software for x-ray crystallography was Blu-Ice
- Genome sequencing and annotations were performed by Era7 Bioinformatics
- Structural homologs of SznE were identified using HHpred (HHSuite v. 3)
- The closest structural homologs’ sequences were aligned with SznE using clustalo (version 1.2.4).

Data analysis
- Pymol Molecular Graphics system, Version 1.8 (Shrödinger, LLC) was used to view PDB files and for generating structural figures
- Geneious Pro 467 Version 7.1.6 (Biomatters) and MEGA-X were used to view sequence alignments
- NMR spectra were visualized and analyzed using MestReNova, version 10.0.0-14411
- Data analysis XDS, HKL2000, ShelX, HKL2MAP, ARP/wARP, PHASER, Refmac5, COOT, TLSMD server for crystallographic characterization of SznF
- Microsoft Excel was used to construct bar graphs, Supplementary tables, and pie charts
- Adobe Illustrator CC2015 was used to construct figures
- Extracted ion chromatograms were generated with Agilent ChemStation software
- Mauve was used for comparative genomics analysis
- O2 consumption assay was analyzed using NeoFox Viewer version 2.40.
Genome neighborhood analysis of SznF homologs were performed with the software tools available at Integrated Microbial Genomes and Microbiome system (Joint Genomic Institute).
Extinction coefficients for determining protein concentrations were performed by protparam.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
No sample size calculations were completed, since the report explores biosynthetic functions of two enzymes. For analysis of metabolite production by various mutants, at least triplicates performed on different days were performed.

Data exclusions
No data were excluded from analyses.

Replication
The number of times an experiment was repeated is detailed in the Statistics and Reproducibility section of the manuscript. All attempted replications yielded similar results. Measures taken to ensure reproducibility include performing biochemical assays on different days using different enzyme preparations and picking different monoclonal colonies for in vivo knockout analysis.

Randomization
Randomization is irrelevant since our study does not involve group allocation.

Blinding
Blinding is irrelevant since our study does not involve group allocation.

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | Involved in the study |
| ☒ | Unique biological materials |
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| n/a | Involved in the study |
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |