Supporting Information

Two-Enzyme Pathway Links L-Arginine to Nitric Oxide in *N*-Nitroso Biosynthesis

Hai-Yan He, Alyssa C. Henderson, Yi-Ling Du, and Katherine S. Ryan

Table of Contents

Methods

General methods. Primers were ordered from Integrated DNA Technologies. Plasmid DNA sequencing was carried out by the Nucleic Acids Protein Service (NAPS) unit at the University of British Columbia. Q5 High-Fidelity DNA polymerase, T4 DNA ligase, and restriction endonucleases were purchased from New England Biolabs and used according to the manufacturer's instructions. PCR and plasmid DNA purification kits were purchased from Qiagen. Nickel Sepharose High Performance histidine-tagged protein purification resin (Ni-iminodiacetic acid ligand, Ni-IDA) was purchased from GE Healthcare. Antibiotics ampicillin, chloramphenicol, kanamycin and apramycin were purchased from Bio Basic Inc. and used for selection of recombinant strains. ^{15}N-labelled reagents were purchased from Cambridge Isotopes Laboratories. ISP4 and TSB media were purchased from BD. The Griess reagent and the Fluorimetric Nitric Oxide Synthase Detection System were purchased from Sigma-Aldrich. Other chemicals, including streptozocin (**1**) and N^{ω}-monomethyl-L-arginine (**2**) chemical standards, were purchased from Sigma-Aldrich, Bio-Rad, or Bio Basic Inc. Inductively coupled plasma mass spectrometry (ICP-MS) was conducted by ALS Environmental. High-resolution NMR spectra were collected on BRUKER AVANCE 600 or 400dir spectrometers. HPLC analysis was performed on an Agilent 1260 HPLC apparatus. LC-MS was performed on an Agilent 6120 Quadruple LC-MS system in positive mode. HR-ESI-MS and LC-HR-MSMS analysis was performed on Bruker maXis Impact with an ultra-high resolution quadrupole time of flight (QTOF) detector connected to an Agilent 1200 SL LC System.

Bacteria strains, vectors and cultures. Strains and vectors used in this study are listed in Table S1. Recombinant *Escherichia coli* strains were grown in LB broth or on LB supplemented with 1.5 % (w/v) agar. Antibiotics ampicillin (100 μg/mL), chloramphenicol (25 μg/mL), kanamycin (50 μg/mL), or apramycin (100 μg/mL) were added, as required. *Streptomyces achromogenes* var. *streptozoticus* (NRRL 3125) and *Streptomyces achromogenes* subsp. *achromogenes* (DSMZ 40028) were cultivated on ISP4 medium[1] for sporulation and conjugation.

Genome sequencing. Genomic DNA of *S. achromogenes* var. *streptozoticus* NRRL 3125 was prepared using a salting out procedure[2]. In brief, 20 µL of a spore solution (10^8/mL) of NRRL 3125 was added to 50 mL of TSB medium (BD). After incubation for 30 h at 30 °C with shaking at 200 rpm in a 250 mL flask equipped with a

sterile spring, the mycelia were collected by centrifugation at 4000 rpm from 5 ml of cell suspension. Pellets were resuspended in 1 mL of STE buffer (75 mM NaCl, 25 mM EDTA pH 8.0, 20 mM Tris-HCl pH 7.5), then a lysozyme solution was added to 0.3 mg/mL. After lysis at 30 °C for 15 min, 0.1 mL of a 10 % (w/v) SDS solution was added and then mixed slowly by inversion and incubated 10 min at 55 °C. NaCl was added to 1.25 M and mixed thoroughly by inversion before one equivalent chloroform was added to precipitate proteins. The two-phase solution was mixed by inversion for 30 min at room temperature and centrifuged for 20 min at 6000 rpm. The upper aqueous phase was transferred to a new tube and 0.6 (v/v) equivalent isopropanol was added for salting out of DNA. The liquid was removed and 75 % (v/v) of ethanol was added to wash the DNA before dissolving the resulting pellet in TE buffer (pH 8.0). Sequencing was performed at the Chinese National Human Genome Center (Shanghai, China). Genomic DNA (1 μg) was fragmented by sonication and ~300 bp and ~500 bp fragments were separately recovered and used to construct libraries using TruSeqTM DNA Sample Prep Kit – Set A (Illumina, USA). The libraries were amplified by TruSeq PE Cluster Kit (Illumina, USA) and sequencing of two libraries was performed on the Illumina Hiseq2000. The 300 bp and 500 bp libraries yielded 6.80 and 5.20 Mbp, respectively, which after sequence assembly using Velvet 1.2.03[3] gave 8.79 Mbp of non-redundant sequence across 408 contigs (coverage is ~410×). Gene analysis and functional annotation were performed using Glimmer 3.02,[4] Genemark,[5] and ZCURVE,[6] combined with 2ndFind (http://biosyn.nih.go.jp/2ndfind/) and Blast P.[7]

Generation of mutants of *S. achromogenes* var. *streptozoticus* NRRL3125. For inactivation of the *stzE* gene, *stzE* was replaced with the *aac(3)IV-ermE** cassette from pYLD50, a plasmid that was constructed in turn by insertion of a 1.2-kbp HindIII/EcoRI fragment of pEH13[8] to pIJ778.[9] To construct the plasmid for conjugation, 2.8–3.0 kilobase pairs flanking each targeted gene was amplified from genomic DNA. For example, for the *stzE* deletion mutant, genomic DNA was amplified using primers KOstzE-L-for and KOstzE-L-rev (for the left arm) and KOstzE-R-for and KOstzE-R-rev (for the right arm) (Table S1). The fragment for *aac(3)IV-ermE** was amplified using KOstzE-Am-for and KOstzE-Am-rev as primers. The cloning vector pUC57/kana was amplified using primers KOstzE-Vct-for and KOstzE-Vct-rev. After purification, the four fragments (the left and right arms of *stzE*, the aac(3)IV-ermE* cassette, and pUC57) were ligated together by Gibson assembly[10] to obtain pUC57 with a ~7 kb insert, as confirmed by sequencing. This plasmid was digested by restriction endonuclease

HindIII and ligated with HindIII-digested pMDR400[11] to give new recombinant plasmid for intergeneric conjugation. The recombinant plasmid was transformed into the non-methylating *E. coli* ET12567/pUZ8002 by electroporation. Intergeneric conjugation between *E. coli* ET12567/pUZ8002 and *S. achromogenes* var. *streptozoticus* NRRL 3125 was carried out on ISP4 agar plates (supplemented with 10 mM of $MgCl_2$). Following incubation for 16 h at 30 °C, 50 μg/mL of nalidixic acid and 100 μg/mL of apramycin were overlaid on the plates to select for single-crossover mutants. Next, single-crossovers were sub-cultivated for three generations on antibiotic-free ISP4 agar. The new conjugants were transferred to apramycin ISP4 agar plates and the new colonies were confirmed by colony PCR using stzE-for and stzE-rev as primers to confirm double crossover mutants. Deletion mutants of *stzF* and *stzG* were constructed similarly, and the corresponding primers are listed in Table S1.

Fermentation of *S. achromogenes* wild type and gene inactivation strains and isolation of labelled compounds. *S. achromogenes* (wild-type and gene inactivation strains) were fermented as previously described,[12] in medium consisting of 40 g/L refined yellow cornmeal, 15 g/L soluble starch (BD), 3 g/L dextrose (Fisher Scientific), 3 g/L peptone (Fisher Scientific), and 10 g/L ammonium sulfate. A spore suspension (25 μL, ~10^8 spores/mL) was added to 50 mL of fermentation medium, and incubated for five days at 30 °C with shaking at 200 rpm before adding 50 mL of acetone followed by centrifugation at 6000 rpm for 10 min. The supernatant was then concentrated under vacuum to 10 mL and a further 40 mL of acetone was added to precipitate additional material. The supernatant was evaporated to 5 mL, and this solution was used for HPLC and LC-MS analysis on a C18 column (Phenomenex, Luna C18, 5 μm, 4.6 mm ID × 250 mm column) at a flow rate of 1 mL/min with UV detection at 230 nm with an isocratic elution with 98 % solution A (water with 0.1 % (v/v) formic acid) and 2 % solution B (acetonitrile with 0.1 % (v/v) formic acid) for 10 min. For each sample, 20 μL of the extracted solution was injected. Biosynthetic intermediates from wild type and deletion mutants of NRRL 3125 were eluted on a Syncronis™ HILIC column (5 μm, 4.6 mm ID × 250 mm) using 30:70 (v/v) with 10 mM of NH_4OAc and 0.15% (v/v) acetic acid in H_2O: acetonitrile at a flow rate of 1 mL/min.

For experiments investigating nitrogen incorporation, 0.5 mM of the precursors were supplemented three times to 50 mL of fermentation medium after incubation at 24 h, 48 h and 72 h. The culture was harvested at the fifth day and extracted as described above. Labelled **1** was purified by semi-preparative HPLC on a Phenomenex

Luna column (C18(2), 5 μm, 10 mm ID x 250 mm) using the same elution method as on the analytical HPLC with a flow rate of 3 mL/min.

To compare the production of metabolites (**4** and **1**) in wild type with or without arginine supplemented, the EIC peak area for compound **4** on LC-MS or the UV peak area at 230 nm for **1** was measured, respectively.

Expression and Purification of StzE, StzF, and StzF-Diiron domain. The genes encoding individual protein were amplified by PCR using primers StzE-for and StzE-rev, StzF-for and StzF-rev, and StzF-for and StzF-Diiron-rev, respectively. PCR products were recovered and ligated into the NdeI/XhoI sites in pET28a before transformation into *E. coli* DH5α. The recombinant plasmid was confirmed free of errors by sequencing before transformation into *E. coli* BL21(DE3) for expression.

To prepare starting cultures, cells harboring pET28a-*stzE*, pET28a-*stzF*, or pET28a-*stzF-diiron* were grown overnight in LB medium supplemented with 50 μg/mL of kanamycin, and 5 mL of cell suspension was inoculated into 750 mL of LB medium with kanamycin. Cells were incubated at 37 °C with shaking at 200 rpm to an OD_{600} of 0.4 and then cooled to 20 °C before adding 30 μM of IPTG along with 1 mM of $FeSO_4$ (to cells expressing StzF and StzF-Diiron). After 18 h, the cells were harvested by centrifugation and the cell pellet was washed and resuspended in 50 mM MOPS or Tris pH 8, 500 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol. Next, cells were lysed by sonication (30 cycles of 10 s of sonication followed by 20 s of rest) and separated by centrifugation at 15,000 rpm for 40 min. Soluble His-tagged protein was purified using Ni-IDA affinity resin and dialyzed against 50 mM MOPS or Tris (pH 7.5), 50 mM NaCl, and 10 % glycerol, and stored at -80 °C for future use. The concentrations of proteins were measured by Coomassie (Bradford) Protein Assay Kit. ICP-MS data were generated by ALS Vancouver – Environmental in Burnaby, BC, Canada. The purified proteins were diluted to 100 μM and treated with hydrochloric acid and nitrate acid to release all metal ions before testing.

To prepare lysates of StzE and StzF, induced cells from 750 mL of medium were harvested and suspended in dialysis buffer (50 mM MOPS, pH 7.5, 50 mM NaCl, and 10% glycerol) and then lysed by sonication followed by removal of precipitates by high-speed centrifugation. Clarified lysates were stored at -80 °C for future use.

***In vitro* StzE reactions**. *In vitro* assays of StzE lysates were performed in 100 μL solutions consisting of 10 μL StzE lysate, 1 mM of SAM, and 0.5 mM of L-arginine or L-citrulline in 50 mM MOPS buffer pH 7.0 and were

incubated at 30 °C for 2 h. Parallel control reactions were performed without SAM. For analysis by HPLC, the reaction was derivatized with dansyl chloride (DNS-Cl).[13] To 25 µL of the reaction assay, 25 µL of 1.5 mg/ml DNS-Cl dissolved in CH_3CN and 20 µL of 80 mM Li_2CO_3 (pH 10) was added and incubated at room temperature for 1 h. Excess DNS-Cl was reacted with 10 uL of a 2 % (v/v) of ethylamine solution for another 30 min, and the mixture was analyzed by HPLC and LC-MS on the C18 column at a flow rate of 1 mL/min with UV detection at 290 nm with the following gradient of solutions A and B, described above, at room temperature: constant 90 % A:10 % B (v/v) for 0 to 3 min, followed by a linear gradient to 60 % A:40 % B over 3 to 20 min. All the samples were centrifugated before injection and 20 µL of the mixture was loaded onto the HPLC.

Reactions with purified StzE were assayed in solutions of 50 mM MOPS buffer pH 7.0, 0.5 mM L-arginine, 1 mM SAM and 10 µM StzE. Parallel control samples omitted SAM or used boiled StzE. Derivatization and analysis were carried out identically to the StzE lysate reactions. For intact MS detection the reactions were instead quenched by adding of two volumes of acetonitrile after 2 h incubation. All the samples were centrifugated before loading 30 µL of the mixture onto LC-MS using the HILIC column and eluted with 30:70 (v/v) of 10 mM of NH_4OAc and 0.15% (v/v) formic acid in H_2O: acetonitrile at a flow rate of 1 mL/min. The signals of *m/z* at 175 and 189 were extracted.

Kinetic analysis for StzE catalyzed reaction. A series of reactions of 150 µL in 50 mM MOPS buffer pH 7.0, containing 1 mM of SAM, 1 µM of StzE, and L-arginine (50 µM, 100 µM, 150 µM, 200 µM, 300 µM, 500 µM, 1 mM and 2 mM), were incubated at 30 °C. Reactions were quenched with one volume of acetonitrile after 5, 10 and 15 min for each concentration of arginine. Next, these were derivatized with DNS-Cl and analyzed by HPLC, as described above. The reaction rate was calculated by the peak integration of **2**-DNS relative to Arg-DNS. Each condition (concentration, time) was repeated three times. To calculate k_{cat} and the K_M values, the data were fitted to the Michaelis-Menten model by nonlinear regression using OriginPro (OriginLab, Northampton, MA).

***In vitro* StzF reactions**. *In vitro* assays of StzF lysates were performed in 100 µL solutions consisting of 10 µL StzF lysate, 50 µM PMS, 1.5 mM NADPH, 1.5 mM NADH, 50 µM $FeSO_4$, 0.5 mM of L-arginine or L-citrulline in 50 mM MOPS buffer pH 7.0 and were incubated at 30 °C for 2 h. Parallel control reactions were performed without NADPH and NADH. Derivatization of DNS-Cl and analysis by HPLC are the same as the StzE lysate

reactions. Reactions of StzF with compound **2** were performed in 100 μL solutions of 50 mM MOPS pH 7.0, 50 μM phenazine methosulfate (PMS), 1.5 mM NADPH, 1.5 mM NADH, 50 μM $FeSO_4$, 0.5 mM **2** and either 10 μL StzF lysate or 10 μM purified StzF. Reactions were incubated at 30 °C for 2 h, and derivatized with DNS-Cl as described above. HPLC analysis was carried out using constant 90 % A:10 % B (v/v) for 0 to 3 min followed by a linear gradient to 40 % A :60 % B from 3 to 20 min. The assays of StzF for intact MS detection were carried out in 100 μL solutions of 50 mM Tris pH 5.0, 0.5 mM of **2**, 10 μM PMS, 3 mM NADH, and 10 μM StzF. Negative controls omitted PMS or NADH, or used boiled StzF. When using ferredoxin and ferredoxin reductase to replace PMS, the reaction was performed in 100 μL solutions of 50 mM Tris pH 5.0, 0.5 mM **2**, 0.05 mg/mL of spinach ferredoxin, spinach ferredoxin-$NADP^+$ reductase, 3 mM NADPH, and 10 μM StzF. To determine the iron dependence of the StzF reaction, StzF was treated with 10 mM of EDTA (pH 8.0) for 30 min before adding it to *in vitro* assays. To determine different metal dependence of the StzF, the enzyme purified from *E. coli* BL21(DE3) was dialyzed into a buffer containing 1 mM EDTA (50 mM Tris 7.5, 50 mM NaCl, 10 % glycerol, 1 mM EDTA). The reaction was carried out in 100 μL solutions of 50 mM Tris pH 7.0, 0.5 mM **2**, 50 μM PMS, 3 mM NADH, 10 μM StzF and 5 mM metal salts ($FeSO_4$, $MnCl_2$, $CuSO_4$, $NiSO_4$, $CoCl_2$ or $ZnSO_4$). *In vitro* reactions of StzF at varying pH values were carried out in phosphate buffer in 100 μL solutions of 50 mM phosphate, 0.5 mM **2**, 50 μM PMS, 3 mM NADH, 200 μM $FeSO_4$ and 10 μM StzF. *In vitro* reactions of StzF-Diiron domain and other StzF variants were performed in 100 μL solutions of 50 mM Tris pH 5.0, 0.5 mM of **2**, 10 μM PMS, 3 mM NADH, and 10 μM purified StzF-diiron, with StzF or boiled StzF-diiron added as controls. For reactions analyzed on a HILIC column, the elution method for HPLC condition was identical to that described above for analysis of the StzE reactions. The signals of *m/z* at 175, 177, 179, 187, 188, 189, 190, 191, 192, 193, 205, 206, 207, 208, 209, 217, 218, 219, 220, 221, 231, 232, 233, 234, 235, 236, 237, 238, 239 for potential products were selected to extract.

***In vitro* StzE-StzF coupled reactions.** The assays were performed in 100 μL solutions of 50 mM MOPS pH 7.0, 0.6 mM SAM, 50 μM PMS, 3 mM NADH, 50 μM $FeSO_4$, 10 μM StzE, and 20 μM StzF, with arginine, $^{15}N_2$-L-arginine or $^{15}N_4$-L-arginine as the substrate. Reactions were incubated at 30 °C for 2 h and analyzed by LC-MS and LC-HR-MSMS on HILIC columns. LC-MS method is the same as described above. LC-HR-MSMS was

performed on a Silica HILIC column (100Å, 3 μm, 2.1 mm X 100 mm) with the same elution buffers with a flow rate at 0.3 mL/min.

Purification of compounds produced in StzF reaction. A 50-mL reaction containing 500 μM of **2**, 20 μM of PMS, 2 mM NADH, 50 μM of $FeSO_4$ and 10 μM of StzF was incubated for 4 h at 30 °C and quenched with 100 mL acetonitrile. Precipitated material was removed by centrifugation, and then acetonitrile was removed from the supernatant under Speed-Vac. Finally, water was removed via lyophilization. The resulting residue was dissolved in 3 ml of water and remaining solids were removed by centrifugation. Compounds **3** and **4** were purified from this solution on the HILIC column using HPLC using an isocratic elution of 30%: 70% (v/v) of H_2O containing 10 mM of NH_4OAc and 0.15% formic acid: CH_3CN.

To purify **4**-DNS and **5**-DNS, the same reaction was carried out. After incubation for 4 hours at 30 °C the reaction was brought to room temperature before adding 50 mL of 1 mg/ml DNS-Cl dissolved in CH_3CN and 30 mL of 80 mM Li_2CO_3 (pH 10). After stirring for 1 h, 10 mL of 2 % ethylamine solution was added to react with excess DNS-Cl for another 30 min. Precipitate was removed by centrifugation and acetonitrile was removed under vacuum followed by lyophilization to remove water. The resulting residue was dissolved in 5 mL methanol and undissolved solids were removed by centrifugation. Compound **4**-DNS and **5**-DNS was separated from the supernatant by using semi-prep HPLC on the C18 column described above using an isocratic elution of 90% A/10% B to 3 min and a gradient to 60% A/40% B at 3 to 20 min with a flow rate of 3 mL/min.

***In vitro* StzF reactions using purified 3 and 4 as substrates.** To identify whether **3** and **4** are intermediates, purified **3** or **4** as substrates were assayed with StzF. The reaction was performed in 50 μL solutions consisting of 10 μM StzF, 50 μM PMS, 1.5 mM NADH, 50 μM $FeSO_4$, and 200 μM **3** or **4** in 50 mM Tris buffer at pH 5.0 and incubated at 30°C for 2 h.

Construction, expression of mutants and activity test of StzF. Q5® Site-Directed Mutagenesis Kit (New England Biolabs) was used to generate plasmids. First, plasmid DNA was amplified by PCR using primers in Table S1 and pET28a-*stzF* as the template. The PCR products were purified and treated with Kinase-Ligase-DpnI (KLD) enzyme and transformed into *E. coli* DH5. The plasmids extracted from transformants were validated by sequencing. Expression and purification of those mutants are the same as the wild type of StzF. *In vitro* reactions were carried out as described above for wild-type StzF reactions.

Detection of nitrite and nitric oxide from the StzF reaction and fermentation extracts. The Griess reagent[14] was used for detection of nitrite formation. For qualitative observations, the StzF reaction (50 mM Tris pH 5.0, 0.5 mM of **2**, 10 μM PMS, 3 mM NADH, 100 μM $FeSO_4$ and 10 μM StzF) or controls (50 μL) were diluted in 4 volumes water and mixed with 250 μL Griess reagent, with formation of a pink solution indicating formation of nitrite. To determine the concentration of nitrite from the StzF reaction, 50 μL of reactions were diluted in water to 500 μL before mixing with equal volumes of Griess reagent. To prepare the samples of fermentation extracts, the cells of *S. achromogenes* var. *streptozoticus* NRRL3125 and gene inactivation mutants were grown as described above. Arginine feeding to mutants and the wild type was also performed as described above. After incubation for 5 d, 3 mL of broth was treated with equal volumes of acetone and the precipitates were removed by centrifugation. Acetone was removed by vacuum on Speed-Vac and the aqueous residue was diluted back to 3 mL using water. A 500 μL solution of this mixture was mixed with Griess reagent for nitrite detection. The UV absorbance at 540 nm was read after 20 min at room temperature. A standard curve was made using 0.5 mL of $NaNO_2$ solution in H_2O (2 μM, 3 μM, 5 μM, 10 μM, 20 μM, 50 μM, 100 μM) mixed with equal volumes of Griess reagent.

For nitric oxide release, we employed the Fluorimetric Nitric Oxide Synthase Detection System.[15] The system utilizes a cell-permeable diacetate derivative of 4,5-diaminofluorescein (DAF-2 DA), which is hydrolyzed by an intracellular esterase to DAF-2 that, in turn, reacts with NO produced *in vivo* to form a fluorescent triazolofluorescein. To use it in *in vitro* reactions, *E. coli* lysates (0.5 μL) were added to release the fluorimetric indicator DAF-2. Each assay (50 mM MOPS, pH 7.0) contained 100 μM of **2**, 1.2 mM NADH, 2 μM PMS, 20 μM $FeSO_4$, 10 μM StzF and 5 μM of DAF-2 DA. Negative controls omitted PMS or NADH, or used boiled StzF. NO donor 2-(*N*, *N*-diethylamino)-diazenolate 2-oxide (DEANO) (10 μM) was used as a positive control for NO release. The fluorescent product was measured using an excitation wavelength of 485 nm and an emission wavelength of 535 nm.

Table S1. Bacterial strains, vectors, and oligonucleotide primers used in this study.

Strains	Description	Source
Streptomyces achromogenes var. *streptozoticus* NRRL 3125	wild type producer of streptozocin	NRRL
Streptomyces achromogenes subsp. *achromogenes* NRRL B-2120 (DSMZ 40028)	non-streptozocin production strain	DSMZ
E. coli DH5α	general cloning host	Laboratory stock
E. coli BL21(DE3)	protein expression host	Laboratory stock
E. coli ET12567/pUZ8002	methylation-deficient strain for conjugation (kanamycin- and chloramphenicol-resistant)	Ref 2

Vectors	Description	Source
pSP72	general cloning vector (ampicillin-resistant)	Laboratory stock
pUC57/kana	general cloning vector (kanamycin-resistant)	Laboratory stock
pMDR400	*E. coli-Streptomyces* shuttle vector, derivative of pJTU870, based on pIJ101 replicon	Ref 11
pET28a	protein expression vector	Novagen
pYLD50	Derivative of pEH13, with *aac(3)IV-ermE** intergrated	Laboratory stock

Primers	Sequence (5′-3′)	Description
KOstzE-L-for	GAGCAGATTGTACTGAGAGTGCACAAGCTTCCAACGTCTGGCGCAACCACCTG	HindIII
KOstzE-L-rev	*GCAGGCTGGCGGGAACTGAATTC*GAGCGACTGCAACAGCGCCGCTC	
KOstzE-Am-for	GAGCGGCGCTGTTGCAGTCGCTC*GAATTCAGTTCCCGCCAGCCTGC*	
KOstzE-Am-rev	ACCTGGTTCAGCTTCAGCACCTC*TTACCAACCGGCACGATTGTGCCC*	
KOstzE-R-for	*GGGCACAATCGTGCCGGTTGGTAA*GAGGTGCTGAAGCTGAACCAGGT	
KOstzE-R-rev	TATGACCATGATTACGCCAAGCTTGTCGACGTACTCCTCGACCAGCA	HindIII
KOstzE-Vct-for	TGCTGGTCGAGGAGTACGTCGACAAGCTTGGCGTAATCATGGTCATA	HindIII
KOstzE-Vct-rev	CAGGTGGTTGCGCCAGACGTTGGAAGCTTGTGCACTCTCAGTACAATCTGCTC	HindIII
KOstzF-L-for	GAGCAGATTGTACTGAGAGTGCACAAGCTTTGCAAGAGCTGCCGGCCTCTCAC	HindIII
KOstzF-L-rev	*GCAGGCTGGCGGGAACTGAATTC*GCACGATCGCGTCGCGCAGTTCC	
KOstzF-Am-for	GGAACTGCGCGACGCGATCGTGC*GAATTCAGTTCCCGCCAGCCTGC*	
KOstzF-Am-rev	ACCAGCGGCTTGATGATCTTCTC*TTACCAACCGGCACGATTGTGCCC*	

KOstzF-R-for	*GGGCACAATCGTGCCGGTTGGTAA*GAGAAGATCATCAAGCCGCTGGT	
KOstzF-R-rev	TATGACCATGATTACGCCAAGCTTGCGACAGCAGGCTGGAGTAGAAG	HindIII
KOstzF-Vct-for	CTTCTACTCCAGCCTGCTGTCGCAAGCTTGGCGTAATCATGGTCATA	HindIII
KOstzF-Vct-rev	GTGAGAGGCCGGCAGCTCTTGCAAAGCTTGTGCACTCTCAGTACAATCTGCTC	HindIII
KOStzG-L-for	GAGCAGATTGTACTGAGAGTGCACAAGCTTCTGATCTTCACCGTCCACCACCG T	HindIII
KOStzG-L-rev	*GCAGGCTGGCGGGAACTGAATTC*CCCCGCGATGACCACGCAGTTGTC	
KOStzG-Am-for	GACAACTGCGTGGTCATCGCGGGG*GAATTCAGTTCCCGCCAGCCTGC*	
KOStzG-Am-rev	CGTTCCGGGAAGACGGCGGTGACC*TTACCAACCGGCACGATTGTGCCC*	
KOStzG-R-for	*GGGCACAATCGTGCCGGTTGGTAA*GGTCACCGCCGTCTTCCCGGAACG	
KOStzG-R-rev	TATGACCATGATTACGCCAAGCTTCGTTCGACGGGGTAGGTGTAGAAG	HindIII
KOStzG-Vct-for	CTTCTACACCTACCCCGTCGAACGAAGCTTGGCGTAATCATGGTCATA	HindIII
KOStzG-Vct-rev	ACGGTGGTGGACGGTGAAGATCAGAAGCTTGTGCACTCTCAGTACAATCTGCT C	HindIII
StzE-for	AGCAGCCATATGCCGGCGAACGGCTCGCCC	NdeI
StzE-rev	AGCAGCCTCGAG*TTA*GATGCTCACGCGGAGGGTGAA	XhoI
StzF-for	AGCAGCCATATGAGTCACGTCCCCCCGCAC	NdeI
StzF-rev	AGCAGCCTCGAG*TTA*CAGGCACTTTCGGTAGTCC	XhoI
StzF-Diiron-rev	AGCAGC*CTCGAGTTA*CAGCTCGCCGCGCGGTTCGACC	XhoI
StzG-for	AGCAGCCATATGACCTTCTCCGCCGAGGGGC	NdeI
StzG-rev	AGCAGCCTCGAG*TTA*GCCGGCCCTGGACGGACGGC	XhoI
StzF-DE214/5AA-for	ATCGCCGCGTACGGCTACGGCGTCCAC	
StzF-DE214/5AA-rev	GACGACCTTGAACCACTCGGAC	
StzF-H311A-for	GAGGCCATCCACATCGACCAGCACC	
StzF-H311A-rev	GGTGAAGTAGGTGGTGTCCACC	
StzF-H318A-for	CACGCCGGCCGGATGGCCCGCGAGAAG	
StzF-H318A-rev	CTGGTCGATGTGGATGTGCTCG	

Table S2. High-resolution mass spectrometry (HRMS) detection of purified small compounds

Compound	Molecular formula	Theoretical weight (Da)	Observed weight (Da)	Error (ppm)
1 standard	$[M+Na]^+$: $C_8H_{15}N_3NaO_7$	288.080221	288.079973	0.9
1-^{15}N (Purified from ^{15}N-Acetyl-Glucosamine-fed extracts)	$[M+Na]^+$: $C_8H_{15}N_3NaO_7$	288.080221	288.081115	-3.1
	$[M+Na]^+$: $C_8H_{15}N_2{}^{15}NNaO_7$	289.077255	289.078878	-5.6
1-$^{15}N_2$-I (Purified from $^{15}N_2$-Arg-fed extracts)	$[M+Na]^+$: $C_8H_{15}N_3NaO_7$	288.080221	288.080317	0.3
	$[M+Na]^+$: $C_8H_{15}N^{15}N_2NaO_7$	290.074290	290.074505	0.7
1-$^{15}N_2$-II (Purified from $^{15}N_2$-Arg-fed extracts)	$[M+Na]^+$: $C_8H_{15}N_3NaO_7$	288.080221	288.080715	1.7
	$[M+Na]^+$: $C_8H_{15}N^{15}N_2NaO_7$	290.074290	290.075061	2.7
3	$[M+H]^+$: $C_7H_{17}N_4O_3$	205.129517	205.129218	-1.5
	$[M+Na]^+$:$C_7H_{16}N_4NaO_3$	227.111461	227.110725	-3.2
4	$[M+H]^+$: $C_7H_{16}N_3O_4$	206.113532	206.113885	-1.7
	$[M+Na]^+$: $C_7H_{15}N_3NaO_4$	228.095477	228.095323	-0.7
5	$[M+H]^+$: $C_7H_{14}N_3O_3$	188.102968	188.102600	-2.0
4-DNS	$[M+H]^+$: $C_{19}H_{27}N_4O_6S$	439.164582	439.166226	-3.7
	$[M+Na]^+$:$C_{19}H_{26}N_4NaO_6S$	461.146526	461.147787	2.7
5-DNS	$[M+H]^+$: $C_{19}H_{25}N_4O_5S$	421.154017	421.155627	3.8
	$[M+Na]^+$: $C_{19}H_{24}N_4NaO_5S$	443.135962	443.137753	-4.0

Table S3. Analysis of HRMS of **1** purified from extracts of precursor feeding. The percentage incorporation is listed, which is calculated using peak area of mass signals based on unlabeled **1**, which is designated as 100 %.

	$[M+Na]^+$	**$[M+Na+1]^+$**	**$[M+Na+2]^+$**
1 Standard	100%	9.15 %	1.63 %
Feeding ^{15}N-Acetyl-Glucosamine	100%	56.5 %	7.42 %
Feeding $^{15}N_2$-Arg	100%	15.4 %	36.8 %
Feeding $^{15}N_2$-Arg	100%	15.5 %	37.6 %

Table S4. Annotation of genes in the biosynthetic cluster of streptozocin (**1**)

Gene	**Length (basepairs)**	**Putative protein function**	**Source of homolog with the highest identity**
orf-3	1695	Hypothetical protein	*Streptomyces peucetius* subsp. *caesius* ATCC 27952
orf-2	207	PIN domain nuclease	*Streptomyces phaeoluteigriseus*
orf-1	537	Radical SAM protein	*Streptomyces* sp. FXJ1.172
stzA	594	RNA polymerase sigma factor	*Streptomyces* sp. di188
stzB	801	Cysteine S-methyltransferase	*Streptomyces* sp. di188
stzC	546	Cysteine S-methyltransferase	*Streptomyces* sp. 351MFTsu5.1
stzD	768	2-Oxoglutarate-dependent dioxygenase	*Pseudonocardia dioxanivorans* CB1190
stzE	1059	Protein arginine *N*-methyltransferase	*Stigmatella aurantiaca* DW4/3-1
stzF	1426	Cupin domain-containing diiron protein	*Stigmatella aurantiaca* DW4/3-1
stzG	333	2Fe-2S protein	*Cyanothece* sp. PCC 7822
stzH	1251	Carbamoyl-phosphate synthase/ATP-binding protein	*Streptomyces violaceusniger* Tu 4113
stzI	1296	Major facilitator superfamily protein	*Kribbella flavida* DSM 17836
stzJ	957	Alpha/beta hydrolase	*Bacillus cereus* BDRD-ST24
stzK	1275	ATP-binding protein	*Bacillus thuringiensis* BMB171
stzL	633	homoserine/threonine efflux protein	*Bacillus thuringiensis serovar pakistani* str. T13001
stzM	2037	regulatory protein AfsR	*Streptomyces roseosporus* NRRL 15998
orf+1	1263	amidohydrolase	*Streptomyces achromogenes* subsp. *achromogenes*
orf+2	1428	MFS transporter	*Streptomyces achromogenes* subsp. *achromogenes*
orf+3	1308	Serine/threonine protein kinase	*Streptomyces hygroscopicus* subsp. *limoneus*
orf+4	744	ANTAR domain-containing protein	*Streptomyces achromogenes* subsp. *achromogenes*

Table S5. Genes encoding Rieske and ferredoxins in the NRRL 3125 genome.

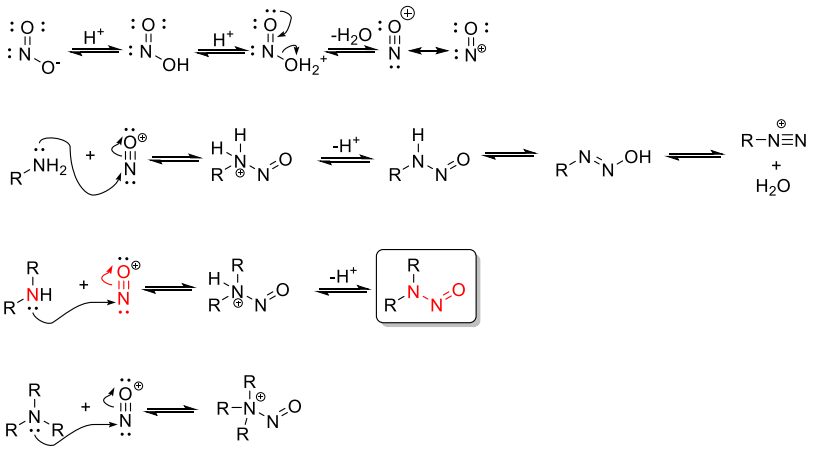
Gene	**Length (basepairs)**	**Protein function**	**Description**
orf7768	867	Rieske Fe-S membrane protein	Containing a Rieske domain in C-terminal
orf2842	297	Ferredoxin	[4Fe-4S] single cluster
orf4103	324	Ferredoxin	[4Fe-4S] dicluster
orf5049	318	Ferredoxin	[4Fe-4S] dicluster

Table S6. ICP-MS analysis of StzF and its variants to identify potential bound metal ions. The calculated concentration (µmol per 100 µmol of protein) is calculated from raw mg/L data.

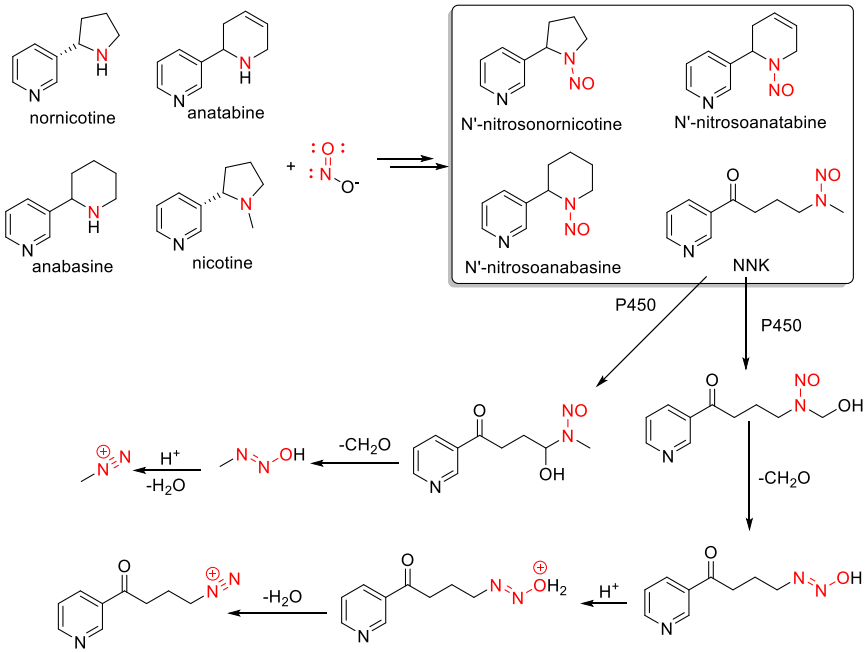
	Wild type of StzF		Diiron domain of StzF		StzF-D214A-E215A		StzF-H311A		StzF-H318A	
Analyzed metal ions	Obs mg/L	Calc µmol /100 µmol	Obs mg/L	Calc µmol /100 µmol	Obs mg/L	Calc µmol /100 µmol	Obs mg/L	Calc µmol /100 µmol	Obs mg/L	Calc µmol /100 µmol
Cobalt (Co)	0.035	1.3	<0.0020	<0.034	<0.0020	<0.034	<0.0020	<0.034	<0.0020	<0.034
Copper (Cu)	0.160	0.0025	0.093	1.46	0.120	1.89	0.120	1.89	0.100	1.57
Iron (Fe)	5.23	93.4	<0.20	<3.57	2.65	47.3	2.65	47.3	1.88	33.6
Magnesium (Mg)	<0.05	<2.08	<0.10	<4.16	<0.10	<4.16	<0.10	<4.16	<0.10	<4.16
Manganese (Mn)	0.161	2.92	0.0025	0.045	0.054	0.97	0.0535	0.97	0.267	0.49
Molybdenum (Mo)	<0.0005	<0.005	0.0013	0.014	0.0012	0.013	0.0012	0.013	0.0012	0.013
Nickel (Ni)	2.64	44.7	7.15	121	8.00	136	8.00	136	4.72	80
Zinc (Zn)	1.43	22.0	0.528	8.12	1.13	17.4	17.4	14.6	0.952	14.6
Aluminum (Al)	0.035	1.3								
Cadmium (Cd)	0.000217	0.002								
Calcium (Ca)	<0.5	<12.5								
Potassium (K)	<0.5	12.8								
Lithium (Li)	<0.01	1.43								
Chromium (Cr)	<0.001	<0.019								

Figure S1

a



b



c

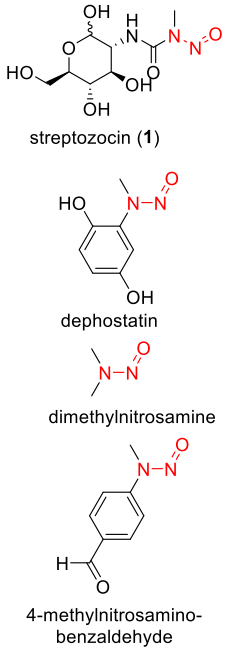
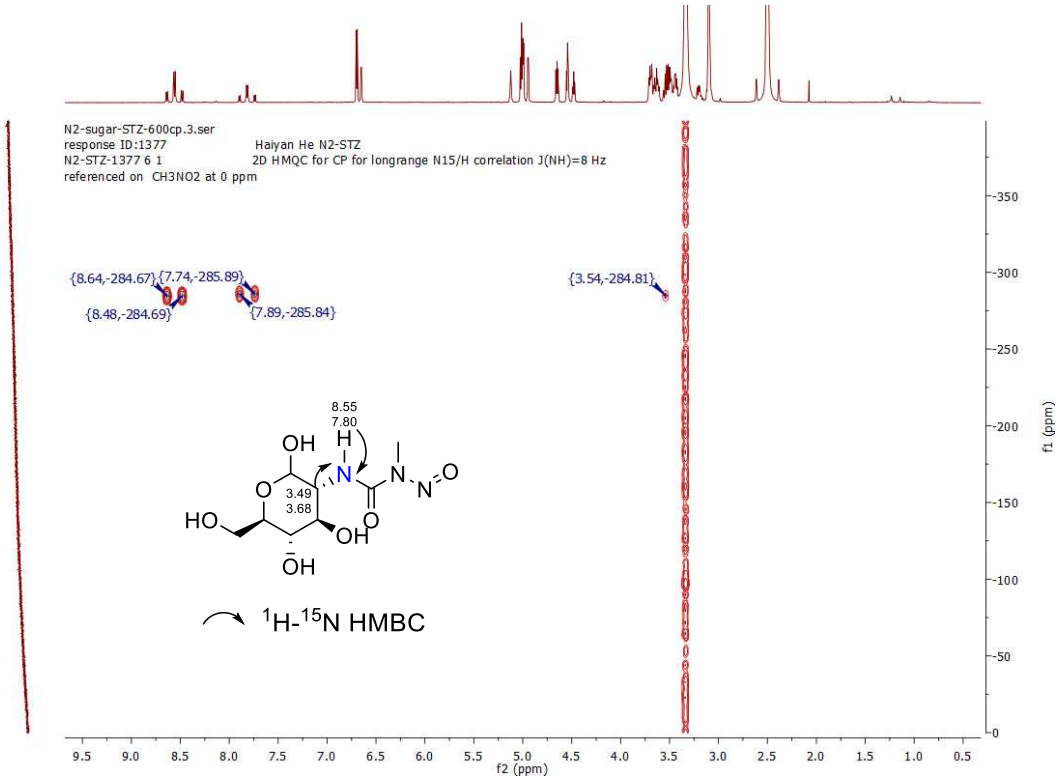


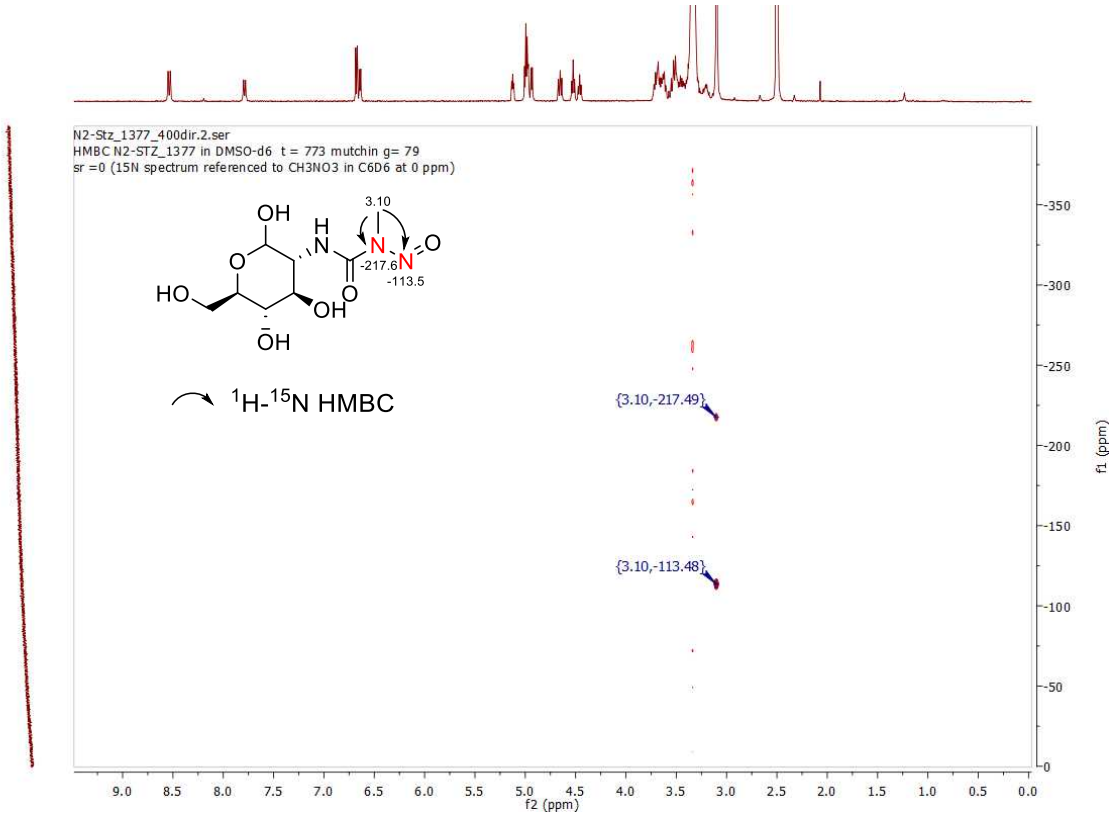
Figure S1. *N*-nitroso compounds. **a**, The major pathway to the *N*-nitroso is a non-enzymatic formation of the nitrosonium ion from nitrite, followed by reaction with secondary amines, amides, or urea-containing molecules.[16] Primary or tertiary amines, however, are not susceptible to *N*-nitroso formation: primary amines instead spontaneously form alkyl diazonium ions, and tertiary amines lack a proton that can be removed to give a stable *N*-nitroso. **b**, Industrial processing of nicotine- and nornicotine-containing tobacco can give tobacco-specific *N*-nitroso compounds.[17] Modification of these *N*-nitroso compounds by cytochrome P450 enzymes in humans can give reactive alkyl diazonium ions.[18] NNK: 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. **c**, A small number of *Streptomyces*-derived natural products contain the *N*-nitroso group. These are streptozocin (**1**) and dephostatin.[19] Additionally, dimethylnitrosamine was detected from the urine of humans infected by *Proteus mirabilis*[20] and 4-methylnitrosaminobenzaldehyde was first isolated from *Clitocybe suaveolens*.[21]

Figure S2

a


N2-sugar-STZ-600cp.3.ser
response ID:1377
Haiyan He N2-STZ
N2-STZ-1377 6 1
2D HMQC for CP for longrange N15/H correlation J(NH)=8 Hz
referenced on CH3NO2 at 0 ppm
{8.64,-284.67}
{7.74,-285.89}
{8.48,-284.69}
{7.89,-285.84}
{3.54,-284.81}
8.55
7.80
3.49
3.68
^{1}H-^{15}N HMBC
f2 (ppm)
f1 (ppm)

b


N2-Stz_1377_400dir.2.ser
HMBC N2-STZ_1377 in DMSO-d6 t = 773 mutchin g= 79
sr =0 (15N spectrum referenced to CH3NO3 in C6D6 at 0 ppm)
3.10
-217.6
-113.5
^{1}H-^{15}N HMBC
{3.10,-217.49}
{3.10,-113.48}
f2 (ppm)
f1 (ppm)

c

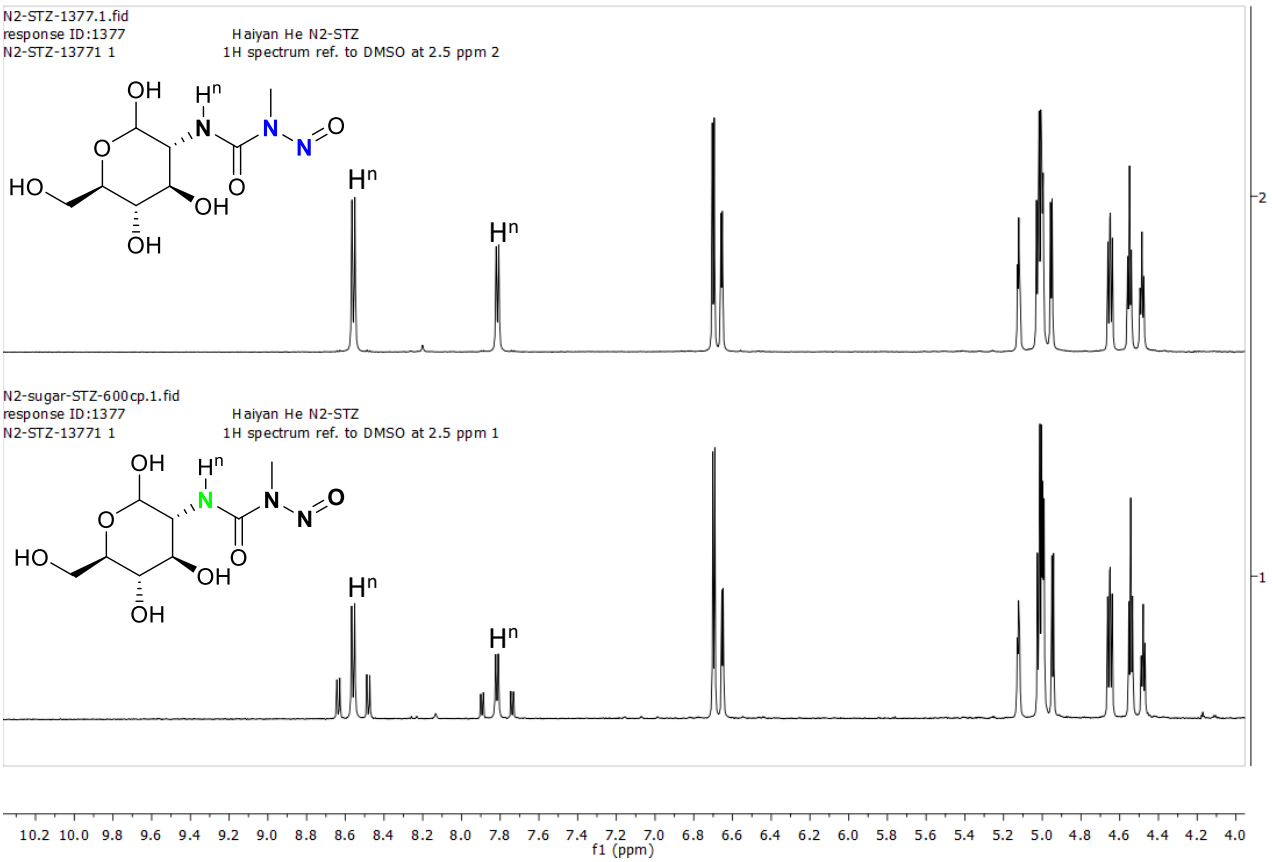
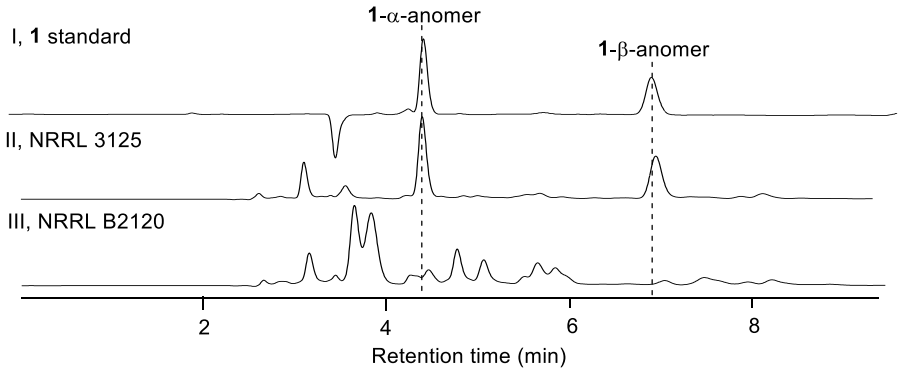


Figure S2. NMR spectra in DMSO-d_6 of **1** purified from medium supplemented by ^{15}N-labelled precursors. **a**, ^{1}H-^{15}N correlation of **1** purified from ^{15}N-acetylglucosamine supplemented extract. **b**, ^{1}H-^{15}N correlation of **1** purified from L-arginine-[guanidino-$^{15}N_2$] incorporated extract. **c**, Proton spectra comparison of **1** from different precursor-fed medium. Top, **1** purified from L-arginine-[guanidino-$^{15}N_2$] incorporated extract; lower, **1** purified from ^{15}N-acetylglucosamine supplemented extract. In the lower spectrum, the H^n splitting by nitrogen was observed. ^{1}H-^{15}N correlation data for **1** purified from L-arginine-[guanidino-$^{15}N_2$] incorporated extract were collected from 400 MHz dir spectrometer (referenced to CH_3NO_2 in C_6D_6 at 0 ppm) and ^{1}H-^{15}N correlation data for **1** purified from ^{15}N-acetylglucosamine supplemented extract and 1D proton data were collected from 600 MHz spectrometer. Note that the analyzed material contains a mixture of labelled and unlabeled products.

Figure S3

a



b

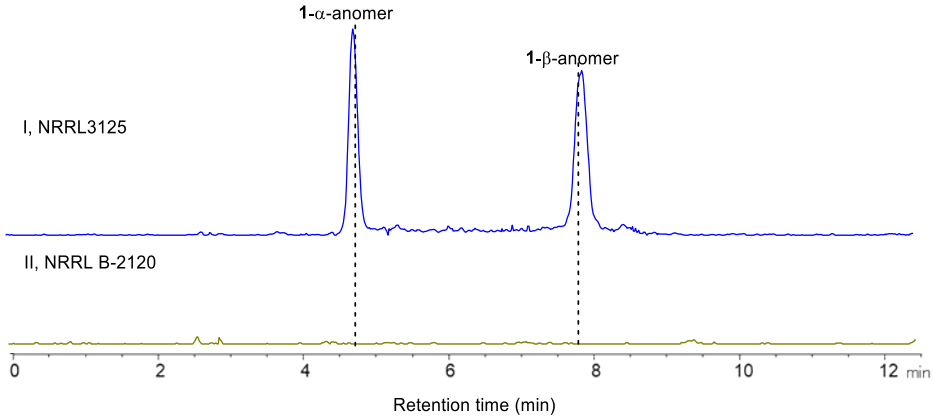


Figure S3. Analysis of **1** production from NRRL 3125 and NRRL B-2120. **a**, HPLC analysis at 230 nm of fermentation of *S. achromogenes* var. *streptozoticus* NRRL3125 and *Streptomyces achromogenes* subsp. *achromogenes* NRRL B-2120. **b**, LC-MS analysis of fermentation of *S. achromogenes* var. *streptozoticus* NRRL3125 and *Streptomyces achromogenes* subsp. *achromogenes* NRRL B-2120. EIC traces of *m/z* at 248 ($C_8H_{14}N_3O_6$: $[M\text{-}H_2O+H^+]$) are shown. The identification of α- and β-anomers is based on the relative abundance in the chemical standard (≥75 % α-anomer).

Figure S4

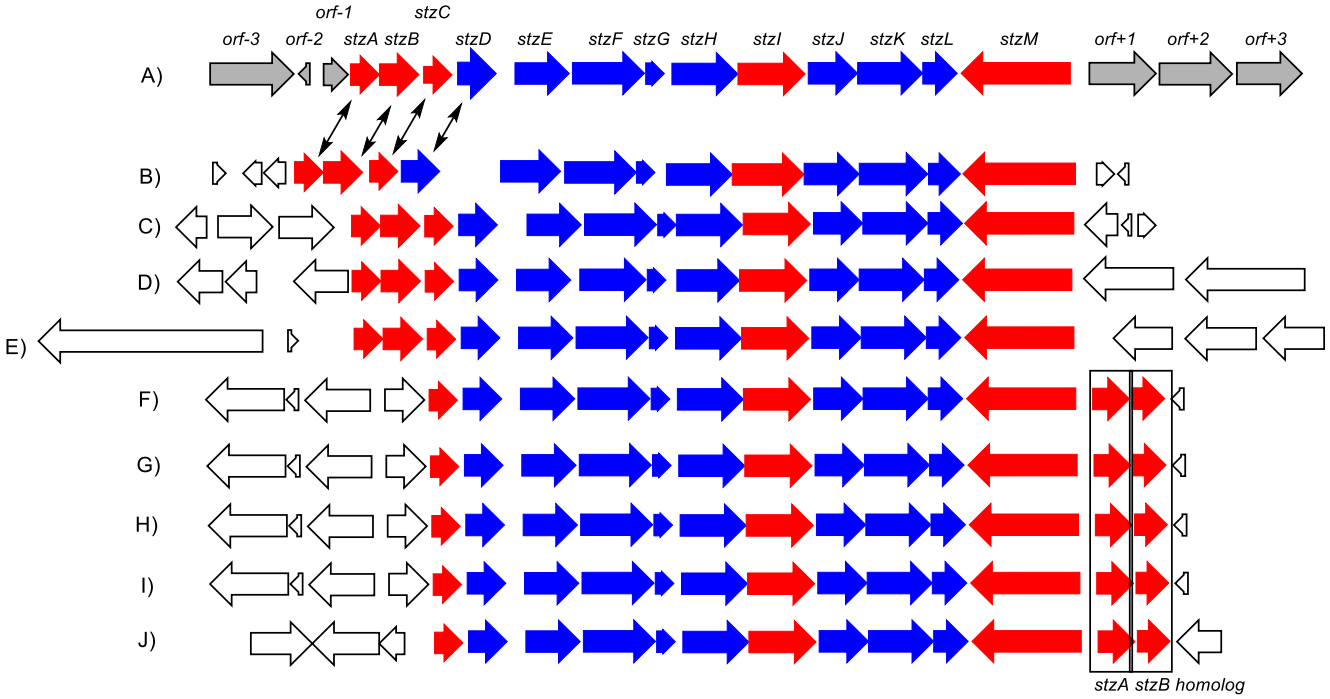


Figure S4. Putative gene cluster for **1** from *Streptomyces* bacteria. **a**, **1** producer *Streptomyces achromogenes* var. *streptozoticus* NRRL 3125; **b**, *Streptomyces* sp. di50b (GenBank accession: FMCK01000266); **c**, *Streptomyces* sp. SM1 (GenBank accession: NZ_NEUB01000615); **d**, *Streptomyces thermolilacinus* SPC6 (GeneBank accession: NZ_ASHX02000001); **e**, *Streptomyces* sp. 351MFTsu5.1 (GenBank accession: NZ_KB911794); **f**, *Streptomyces* sp. NRRL S-515 (GenBank accession: NZ_JODE01000003); **g**, *Streptomyces* sp. CG 926 (GenBank accession: NZ_JODE01000003); **h**, *Streptomyces* sp. NRRL S-237 (GenBank accession: NZ_JODA01000003); **i**, *Streptomyces virginiae* strain NRRL B-8091(GenBank accession: NZ_JNYC01000002); **j**, *Streptomyces* sp. NRRL F-2580 (GenBank accession: NZ_JOIR01000004). Blue, putative biosynthetic genes; red, regulator, transporter, and resistance genes; grey, genes found in NRRL 3125. white, genes unique to other *Streptomyces* species. Note: for gene clusters shown in **f-j**, *stzA* and *stzB* homologs are positioned after the *stzM* homolog.

Figure S5

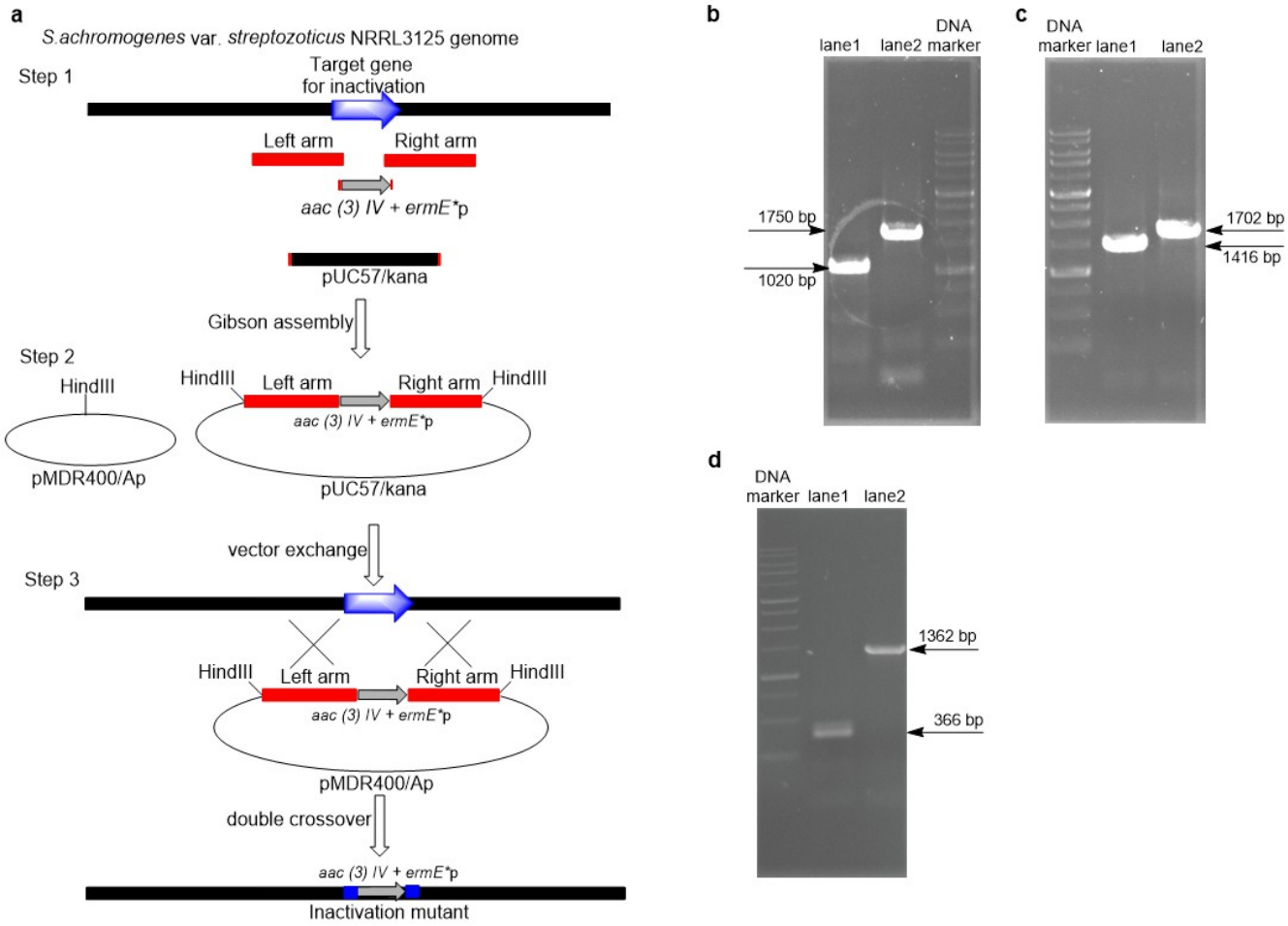
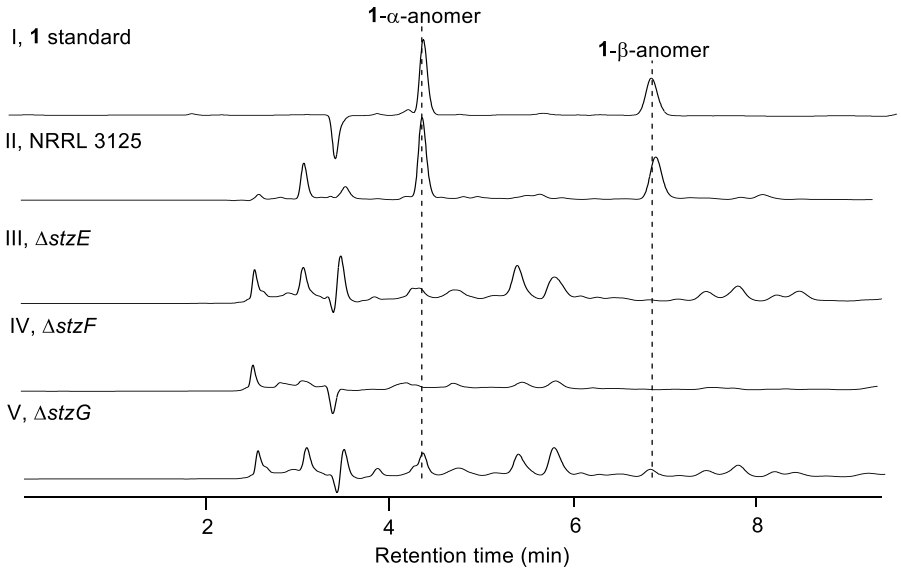


Figure S5. Inactivation of *stzE*, *stzF* and *stzG*. **a**, Schematic procedure of construction of mutants; b-d, PCR verification of each mutant, DNA ladder (top to bottom, kbp): 10, 8, 6, 5, 4, 3, 2, 1.5, 1, 0.75, 0.5, 0.25. **b**, PCR reaction using StzE-for and StzE-rev as primers, 458 bp of fragment from *stzE* replaced by 1188 bp cassette consisting of apramycin resistance gene *aac(3)IV* and constitutive promoter *ermE**p, lane 1, wild type and lane 2, inactivation mutant of s*tzE*; **c**, PCR reaction using StzF-for and StzF-rev as primers, 902 bp of fragment from *stzF* replaced by apramycin resistance marker, lane 1, wild type and lane 2, inactivation mutant of s*tzF*; **d**, PCR reaction using StzG-for and StzG-rev as primers, 192 bp of fragment from *stzG* replaced by apramycin resistance marker, lane 1, wild type and lane 2, inactivation mutant of *stzG*.

Figure S6

a



b

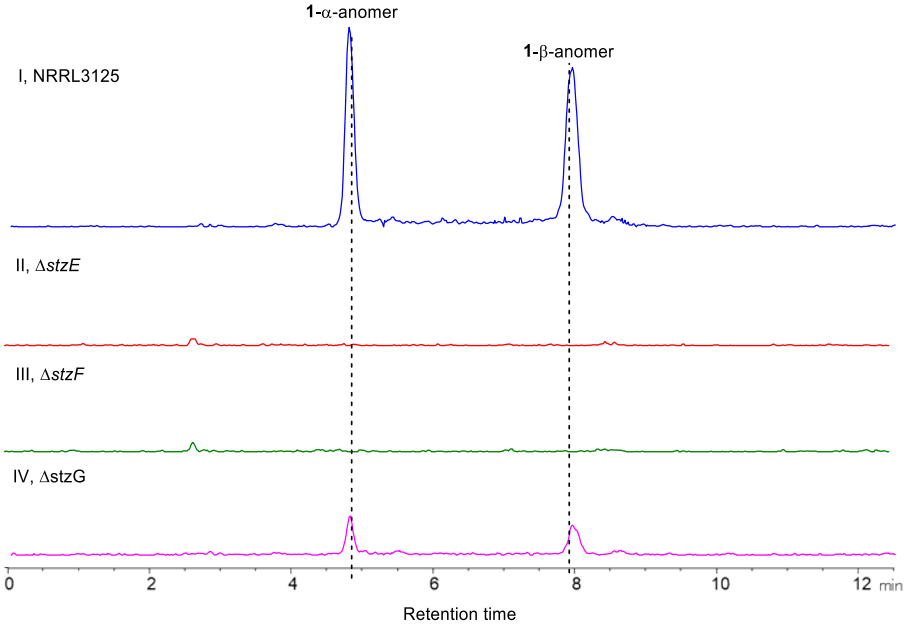
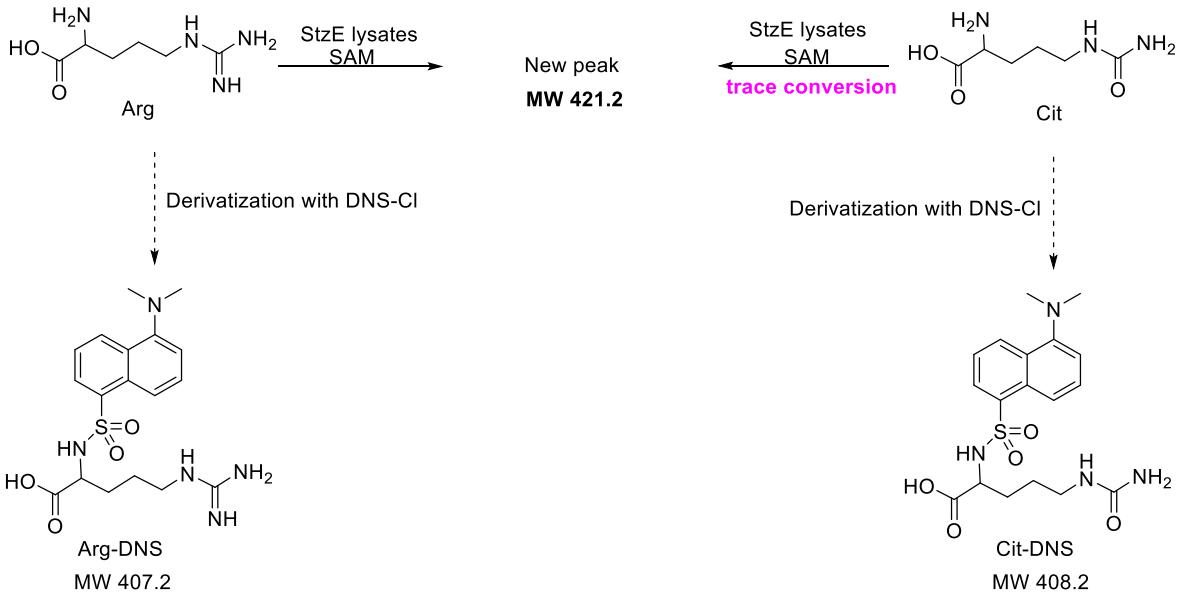


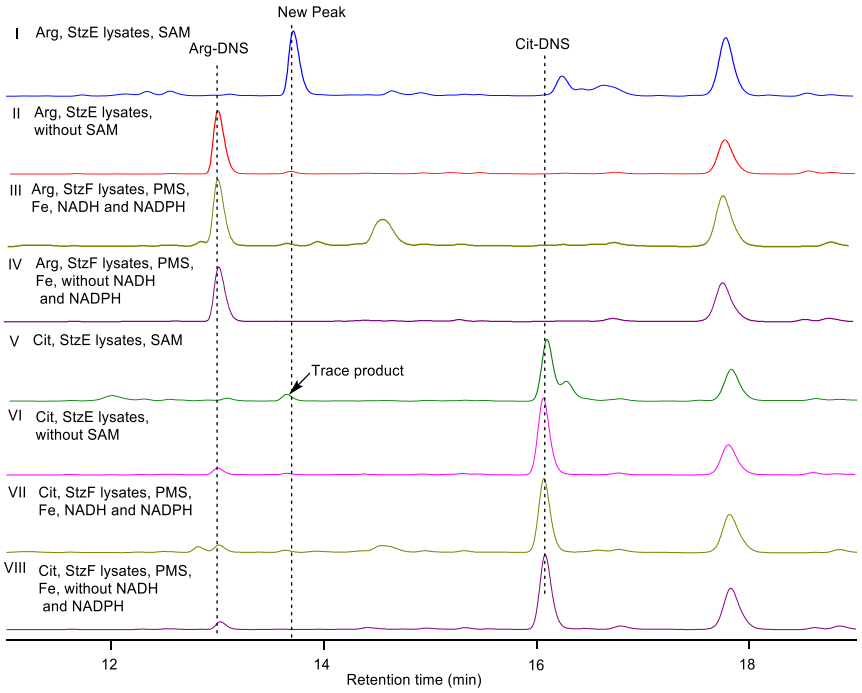
Figure 6. Analysis of wild-type and gene inactivation mutants. **a**, HPLC analysis at 230 nm of fermentation extracts of *Streptomyces achromogenes* NRRL 3125 and B2120 and the NRRL 3125 mutants Δ*stzE*, Δ*stzF* and Δ*stzG*. **b**, LC-MS analysis of crude fermentation extracts of NRRL 3125 wild-type and gene inactivation mutants. EIC traces of *m/z* at 248 ($C_8H_{14}N_3O_6$: [M-H_2O+H^+]) are shown.

Figure S7

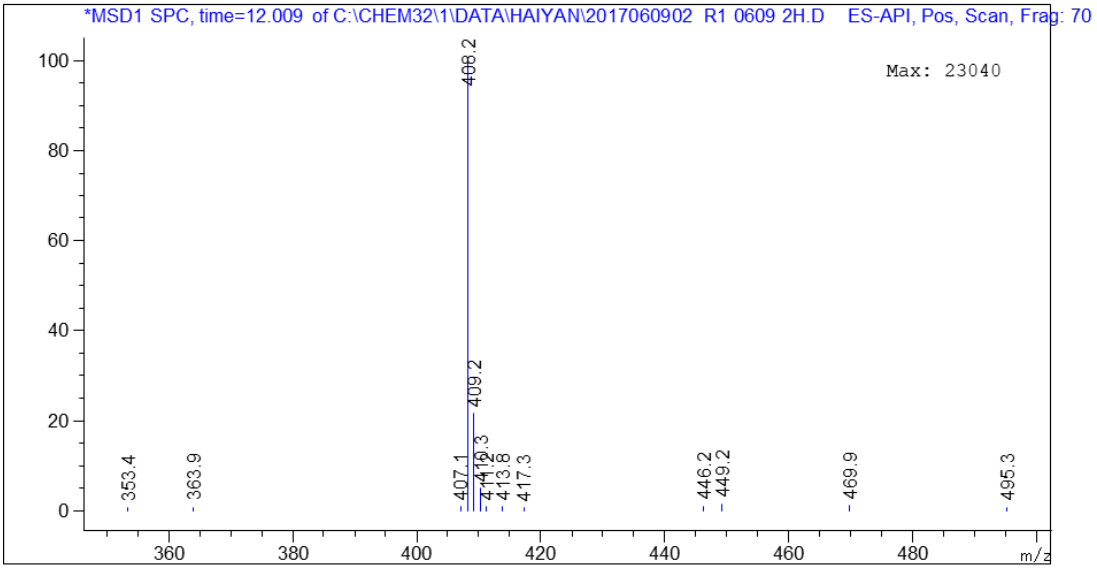
a


Arg
StzE lysates
SAM
New peak
MW 421.2
StzE lysates
SAM
trace conversion
Cit
Derivatization with DNS-Cl
Derivatization with DNS-Cl
Arg-DNS
MW 407.2
Cit-DNS
MW 408.2

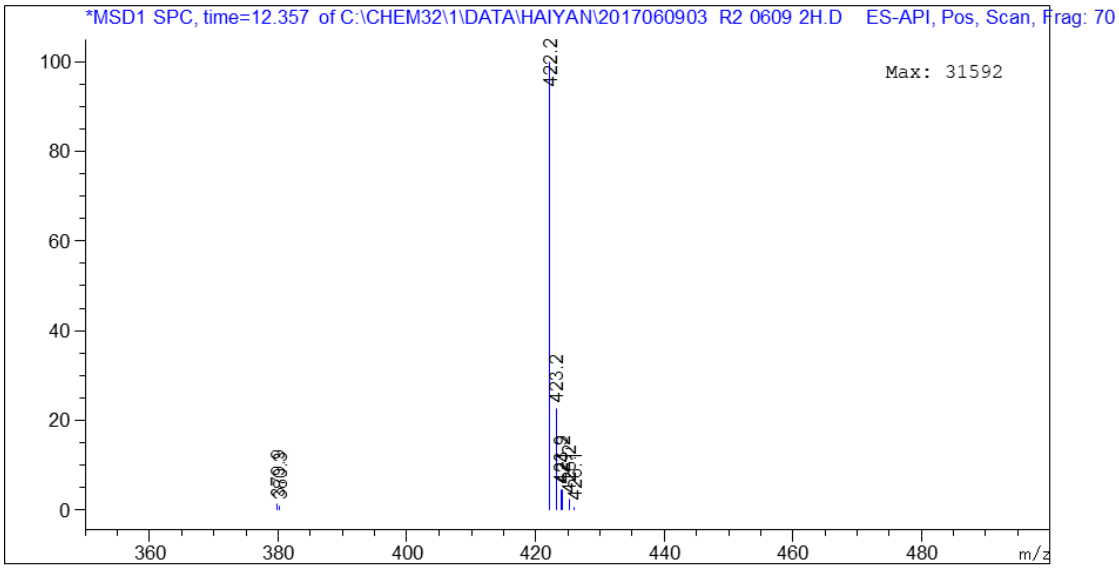
b


New Peak
Arg-DNS
Cit-DNS
I Arg, StzE lysates, SAM
II Arg, StzE lysates, without SAM
III Arg, StzF lysates, PMS, Fe, NADH and NADPH
IV Arg, StzF lysates, PMS, Fe, without NADH and NADPH
V Cit, StzE lysates, SAM
Trace product
VI Cit, StzE lysates, without SAM
VII Cit, StzF lysates, PMS, Fe, NADH and NADPH
VIII Cit, StzF lysates, PMS, Fe, without NADH and NADPH
12
14
16
18
Retention time (min)

c


*MSD1 SPC, time=12.009 of C:\CHEM32\1\DATA\HAIYAN\2017060902 R1 0609 2H.D ES-API, Pos, Scan, Frag: 70
Max: 23040
408.2
409.2
353.4
363.9
407.1
410.3
413.8
417.3
446.2
449.2
469.9
495.3
m/z

d


*MSD1 SPC, time=12.357 of C:\CHEM32\1\DATA\HAIYAN\2017060903 R2 0609 2H.D ES-API, Pos, Scan, Frag: 70
Max: 31592
422.2
423.2
m/z

e

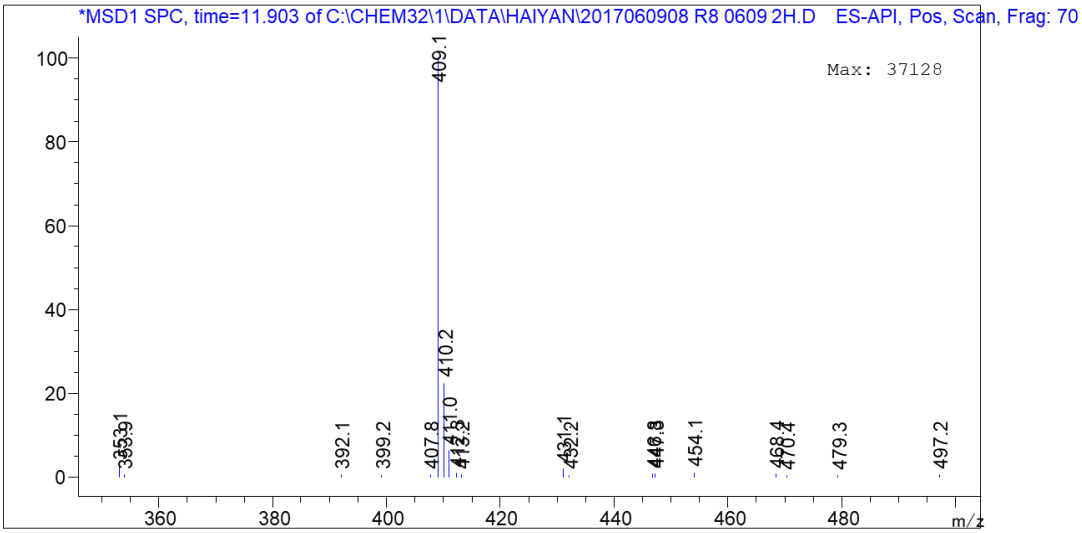


Figure S7. Analysis of reactions of StzE or StzF lysates with L-arginine or L-citrulline followed by derivatization with DNS-Cl. **a** Schematic of reactions observed. **b**. HPLC analysis of reactions. Detection wavelength = 290 nm. For trace V, the new peak at 13.7 min and the substrate peak at 16.2 min were integrated. The new peak is 2.4 % of the substrate peak. **c**, MS spectra of the peak at 13.0 min in part **b**, trace I, which is consistent with arginine derivatization by DNS-Cl; **d**, MS of the new peak at 13.7 min in part **b**, trace II; **e**, MS spectra of the peak at 16.2 min in part **b**, trace V.

Figure S8

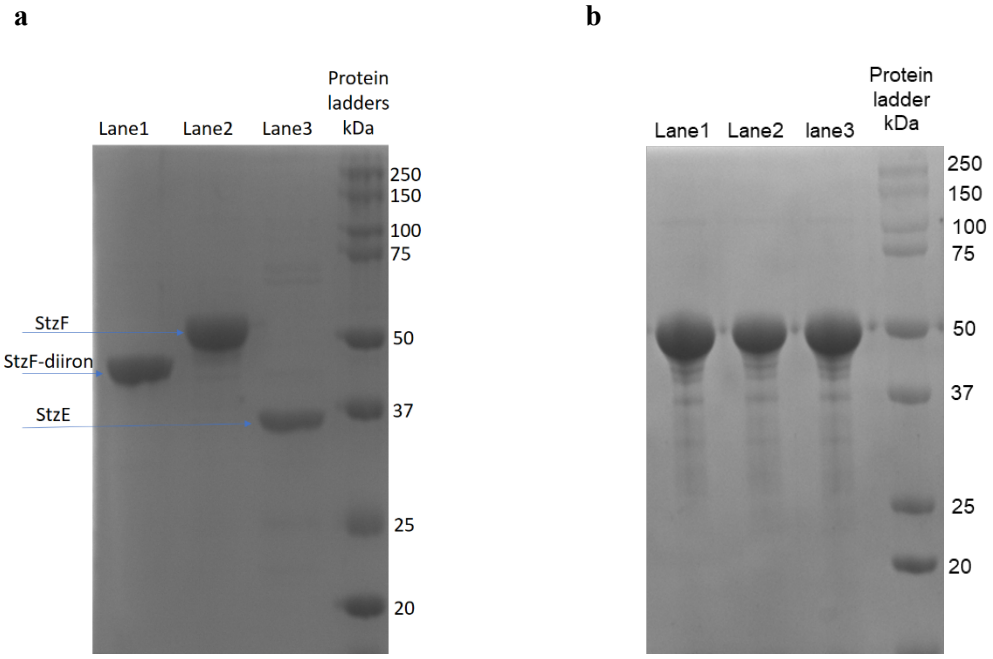


Figure S8. SDS–polyacrylamide-gel electrophoresis of purified enzymes. **a**) Lane1, StzF-diiron domain, 47.8 kDa; Lane2, StzF, 55.2 kDa; Lane3, StzE, 37.4 kDa; Lane4 250 kDa ladder. **b**) Lane 1, StzF-D214A-E215A, 55.2 kDa; Lane 2, StzF-H311A, 55.2 kDa; Lane 3, StzF-H318A, 55.2 kDa; Lane 4, 250 kDa protein ladder.

Figure S9

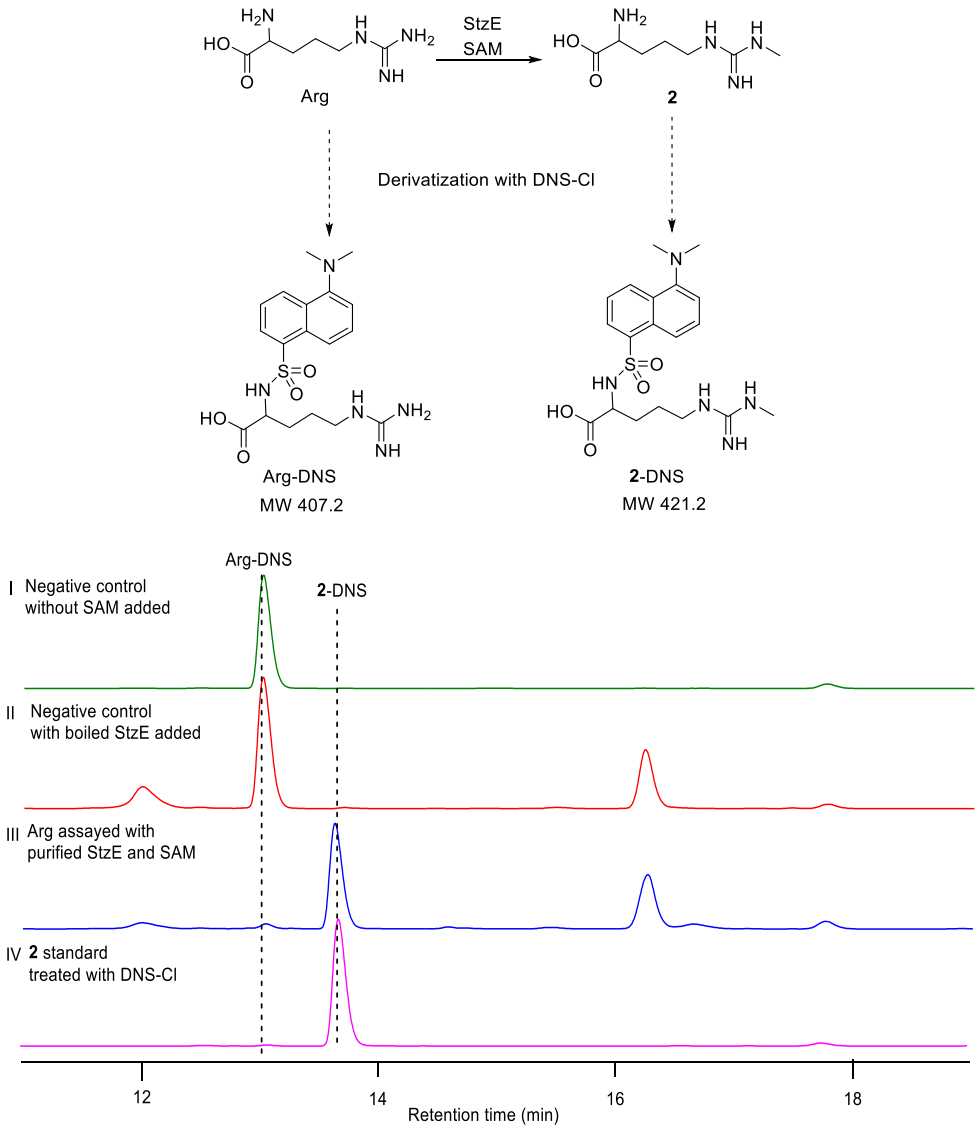


Figure S9. HPLC analysis of StzE reactions derivatized by DNS-Cl. Detection wavelength = 290 nm.

Figure S10

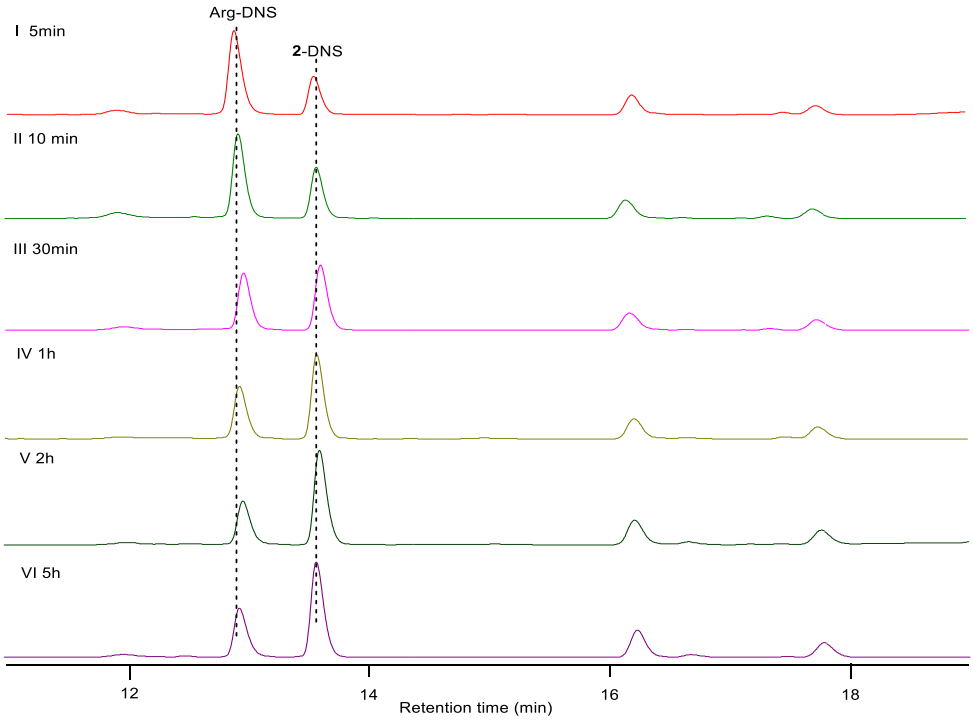


Figure S10. Time-course of arginine methylation catalyzed by StzE. Before reaction with DNS-Cl, each sample was quenched by two volumes of acetonitrile. Detection wavelength = 290 nm.

Figure S11

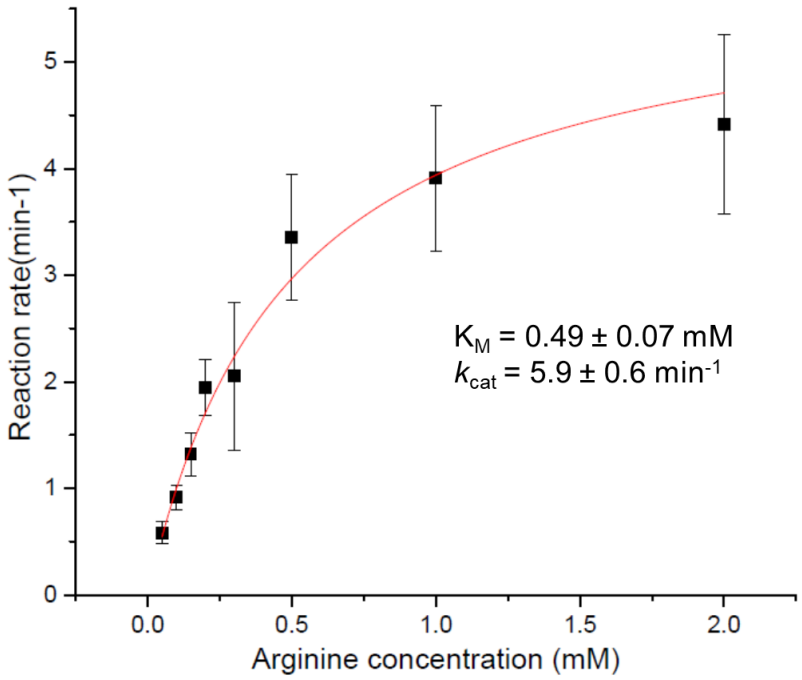


Figure S11. Kinetic analysis of StzE catalyzed reaction. Conversion amount was calculated by the peak ratio of **2**-DNS and Arg-DNS as described in the Methods.

Figure S12

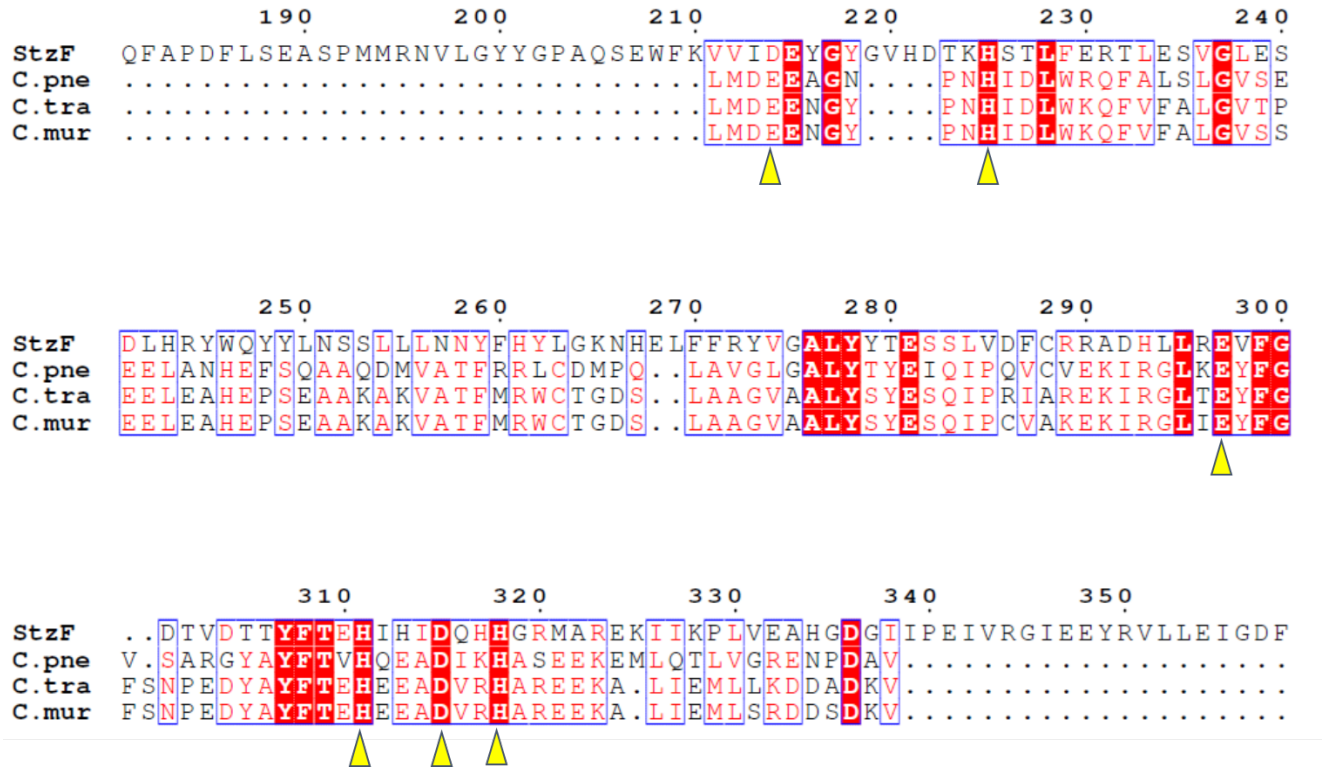


Figure S12. Sequence alignment of the diiron domain of StzF with *Chlamydia* protein CADD (Chlamydia protein associating with death domains).[22] Multisequence alignment was created with Clustal Omega[23] and displayed with ESPript 3.0.[24] Diiron-coordinating active site residues of *Chlamydia trachomatis* CADD are marked by yellow triangles. Note: The variant StzF-D214A-E215A was generated to account for the possibility that E215 is iron-binding, even though D214 aligns with the iron-binding sites in CADD.

Figure S13

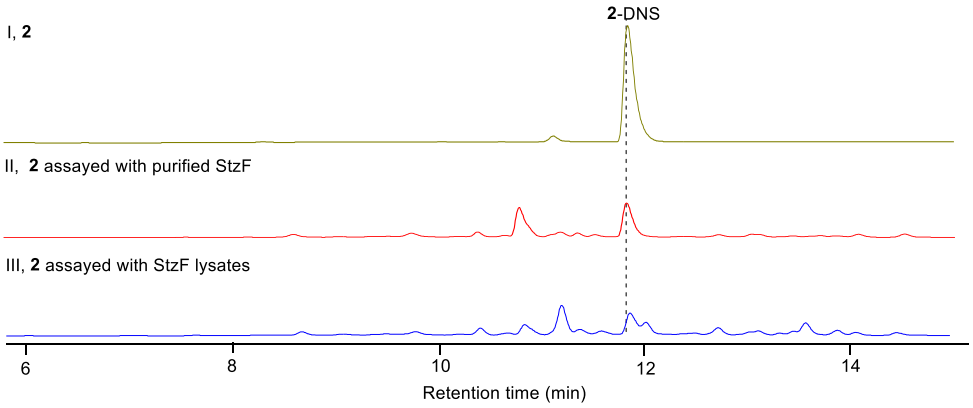
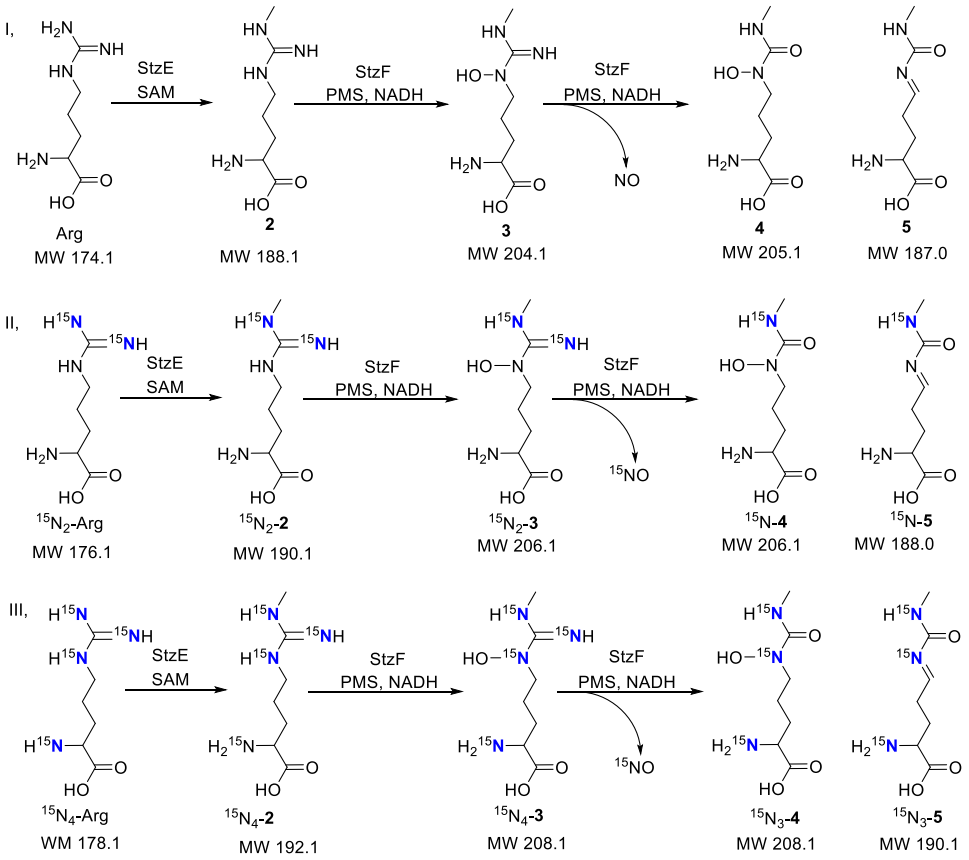


Figure S13. HPLC analysis of StzF reaction with **2** as substrate derivatized by DNS-Cl, detection wavelength = 290 nm. Trace I, derivatization of **2** standard compound; Trace II, the reaction with purified StzF added; Trace III, the reaction with StzF lysates added.

Figure S14

a


I,
StzE
SAM
StzF
PMS, NADH
StzF
PMS, NADH
NO
Arg
MW 174.1
2
MW 188.1
3
MW 204.1
4
MW 205.1
5
MW 187.0
II,
StzE
SAM
StzF
PMS, NADH
StzF
PMS, NADH
15NO
15N2-Arg
MW 176.1
15N2-2
MW 190.1
15N2-3
MW 206.1
15N-4
MW 206.1
15N-5
MW 188.0
III,
StzE
SAM
StzF
PMS, NADH
StzF
PMS, NADH
15NO
15N4-Arg
WM 178.1
15N4-2
MW 192.1
15N4-3
MW 208.1
15N3-4
MW 208.1
15N3-5
MW 190.1

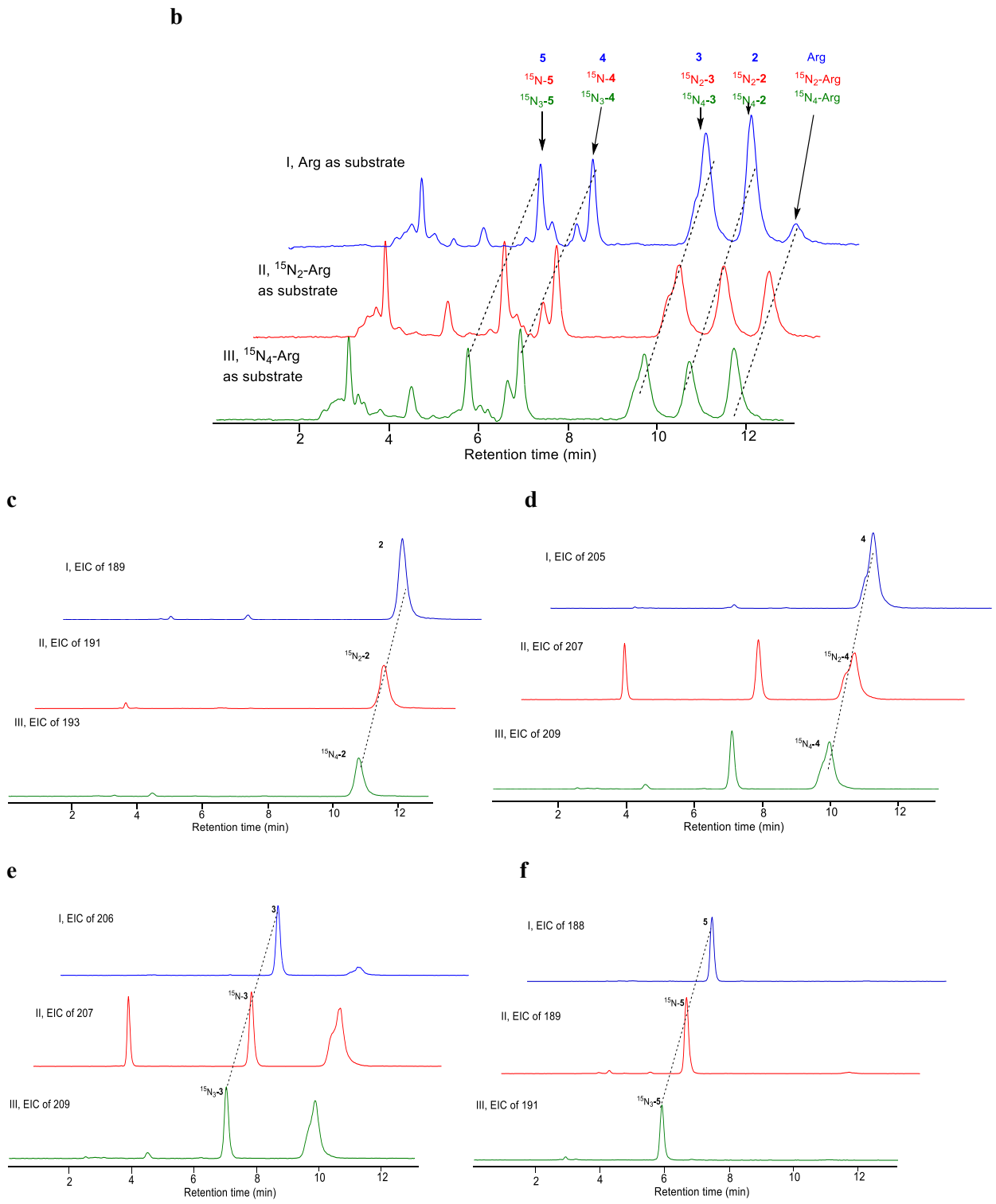


Figure S14. LC-MS analysis of StzE coupled with StzF reactions using different ^{15}N-labeled arginines. **a**, StzE coupled with StzF reactions using different ^{15}N-labeled arginines; **b**, EIC of each reaction: I, arginine added; II, $^{15}N_2$-arginine added; III, $^{15}N_4$-arginine added as substrate; **c-f**, EIC of peaks originated from arginine or labeled arginines: I, arginine added; II, $^{15}N_2$-arginine added; III, $^{15}N_4$-arginine added as substrate.

Figure S15

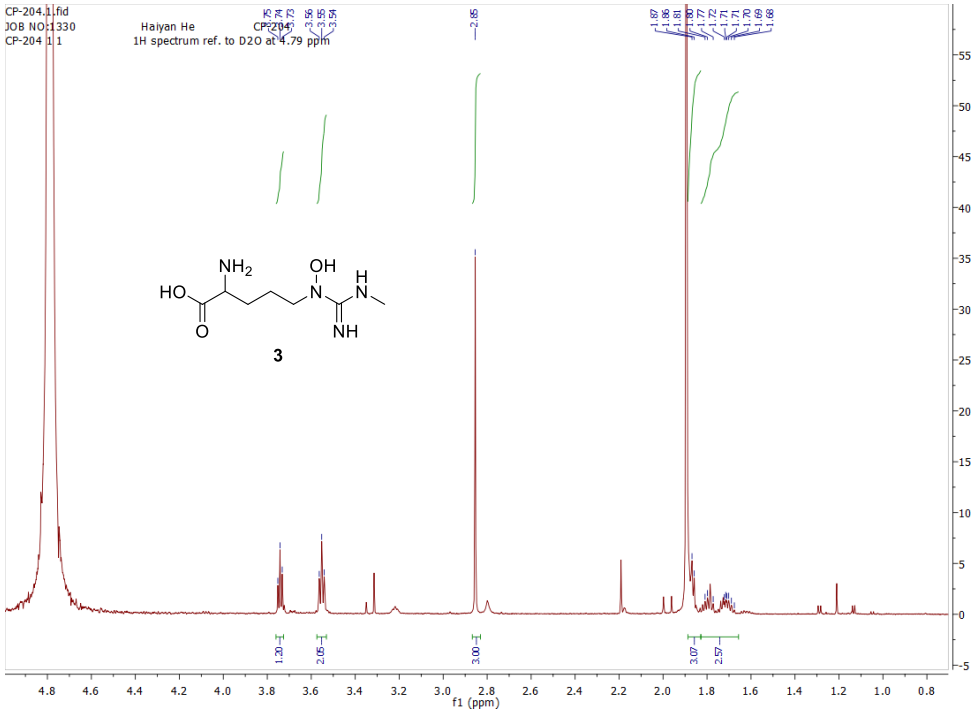


Figure S15a. ^{1}H NMR spectrum of compound **3** in D_2O. ^{1}H NMR (600 MHz, D2O) δ 3.74 (t, *J* = 6.0 Hz, 1H), 3.55 (t, *J* = 6.8 Hz, 2H), 2.85 (s, 3H), 1.86 (d, *J* = 6.0 Hz, 2H), 1.83 – 1.66 (m, 2H).

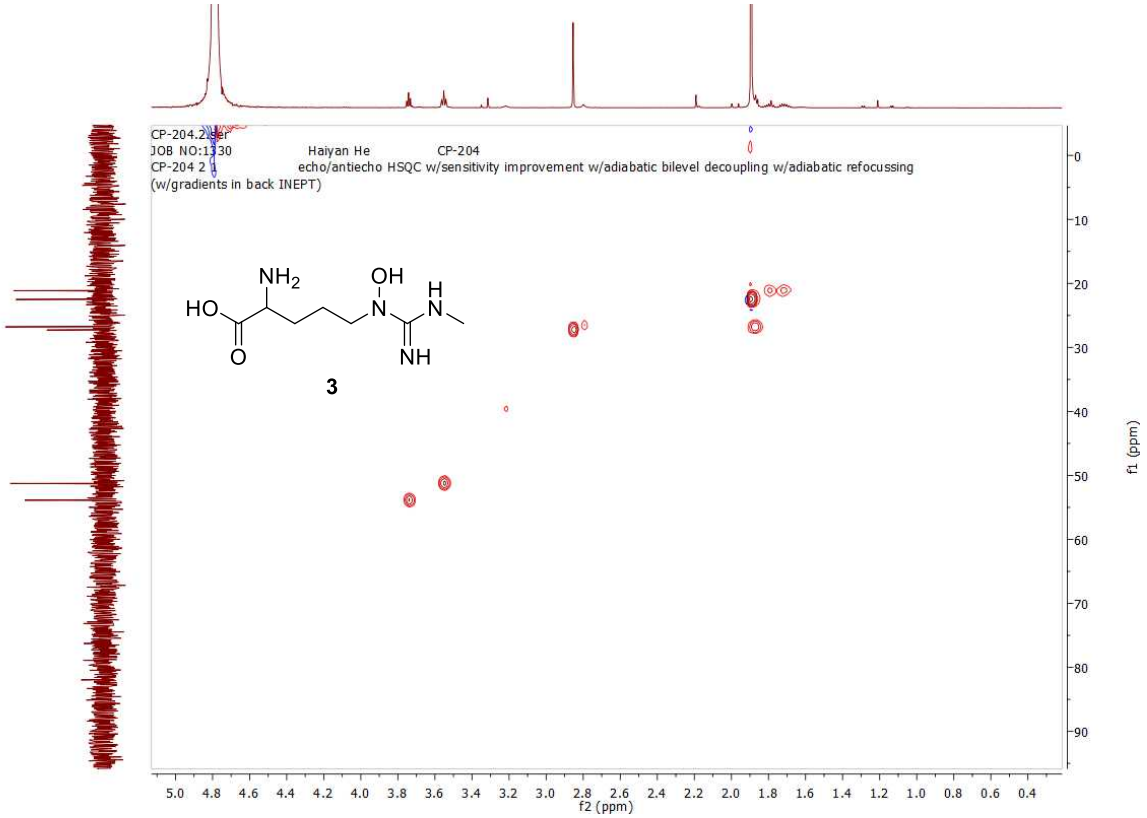


Figure S15b. 2D HSQC NMR spectrum of compound **3** in D_2O.

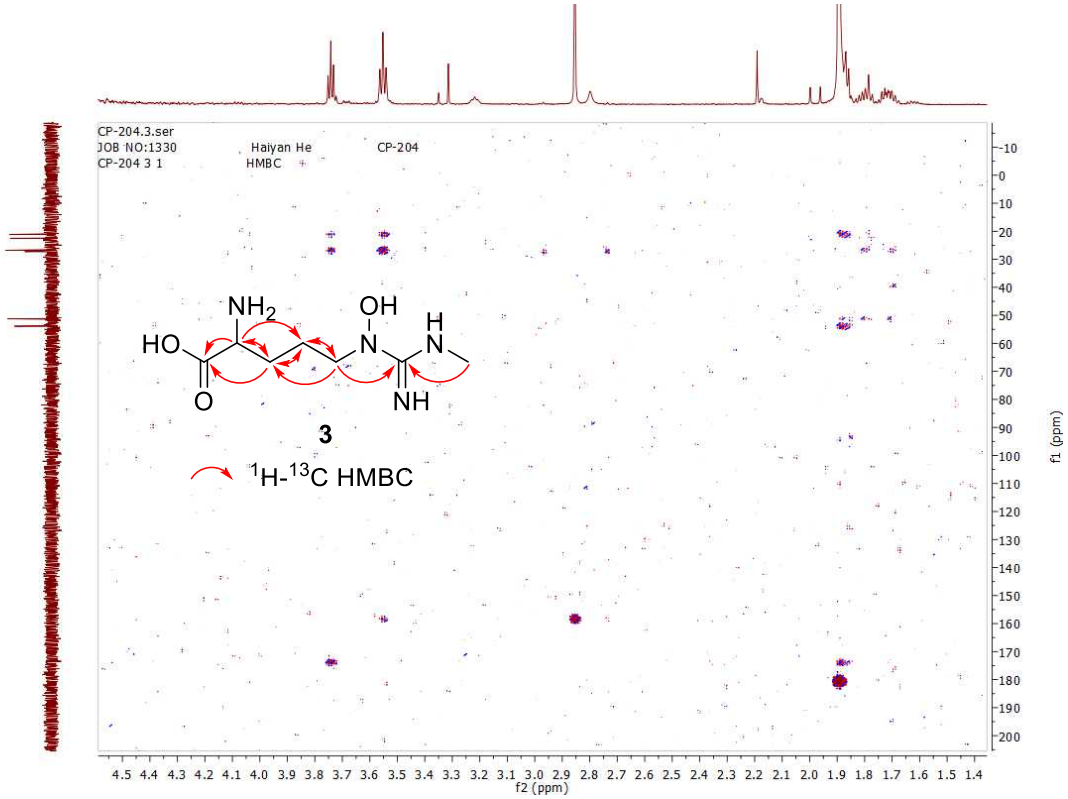


Figure S15c. 2D HMBC NMR spectrum of compound **3** in D_2O.

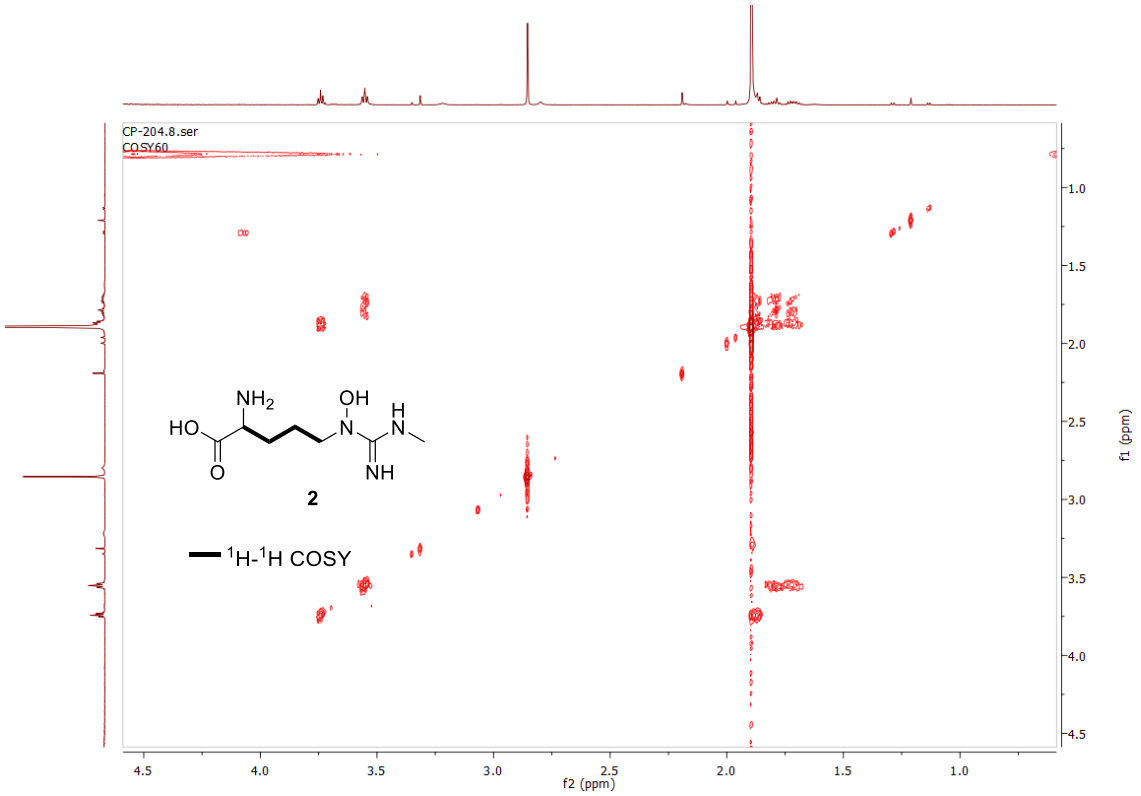


Figure S15d. 2D COSY NMR spectrum of compound **3** in D_2O.

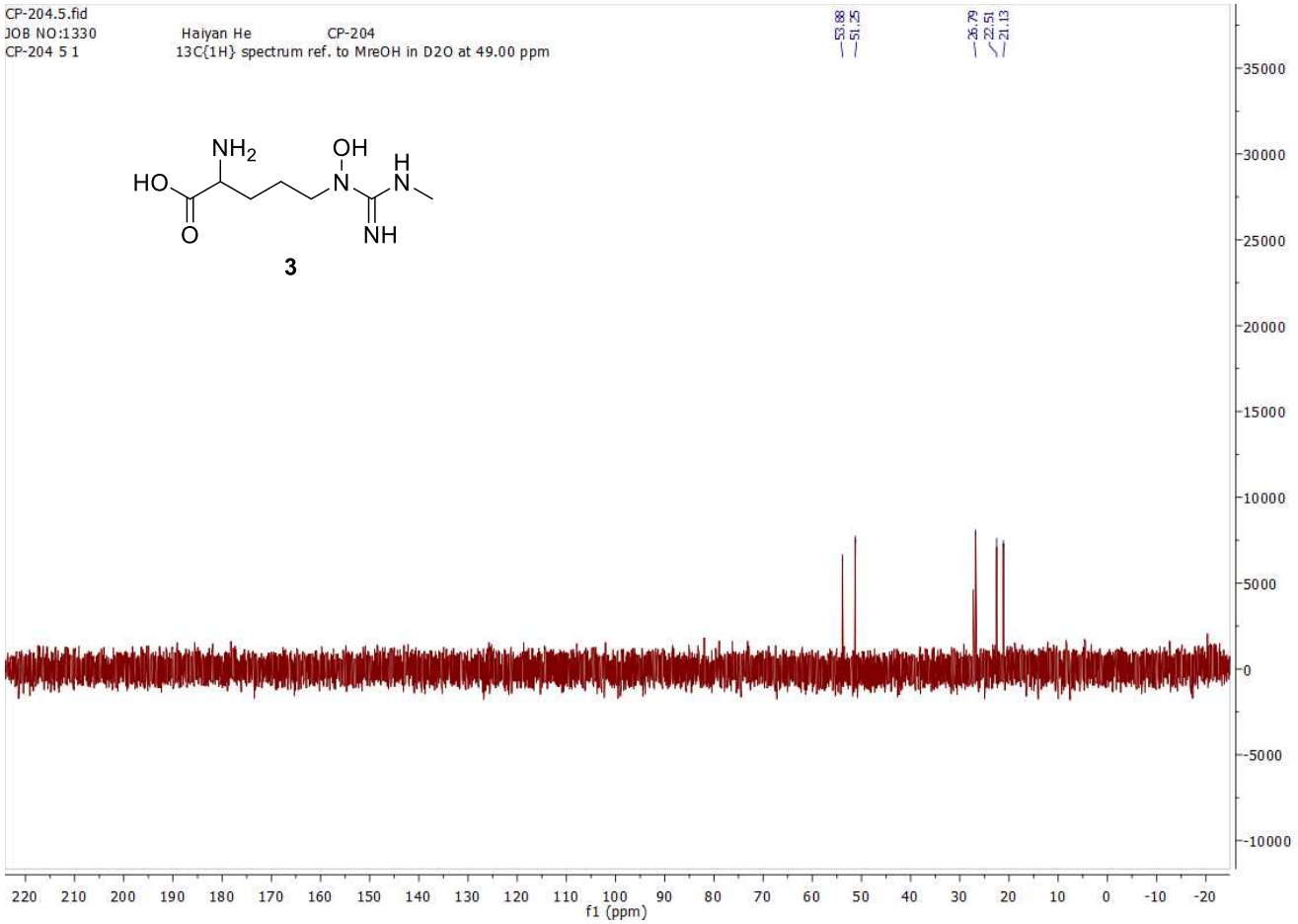


Figure S15e. ^{13}C NMR spectrum of compound **3** in D_2O (^{13}C NMR (150 MHz, Deuterium Oxide) δ 173.9, 158.5, 53.88, 51.25, 26.79, 22.51, 21.13). Note that the listed shifts at 173.9 and 158.5 were obtained from the HMBC data shown in Figure S15c.

Figure S16

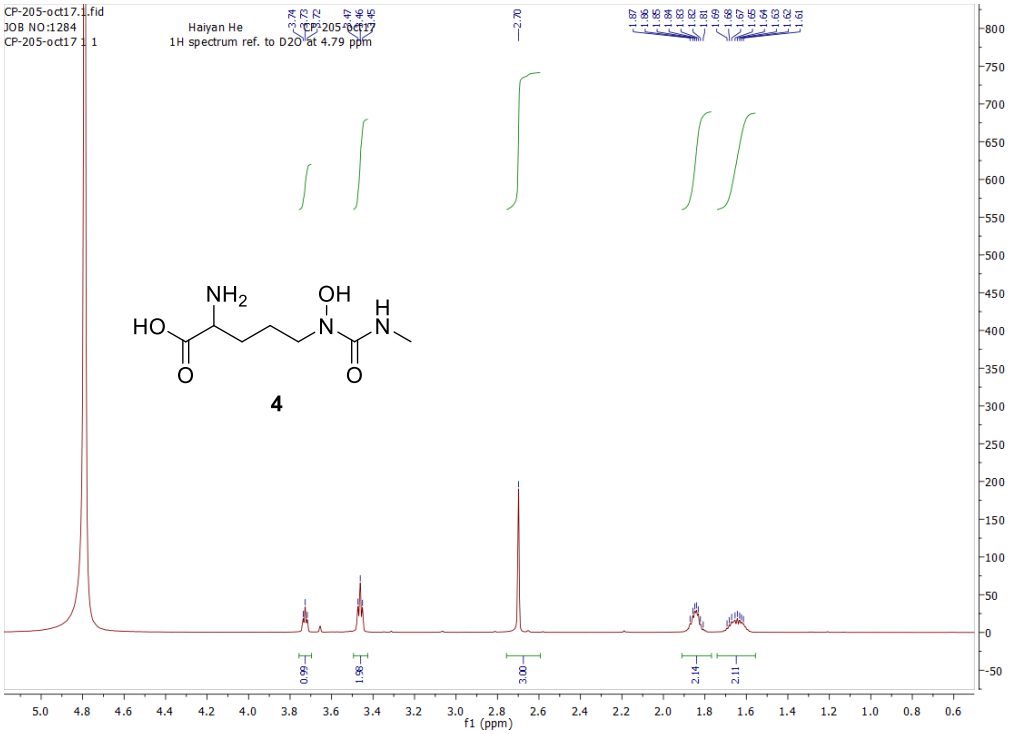


Figure S16a. ^{1}H NMR spectrum of compound **4** in D_2O. ^{1}H NMR (600 MHz, Deuterium Oxide) δ 3.73 (t, J = 5.9 Hz, 1H), 3.46 (t, J = 6.5 Hz, 2H), 2.70 (s, 3H), 1.84 (m, 2H), 1.65 (m, 2H).

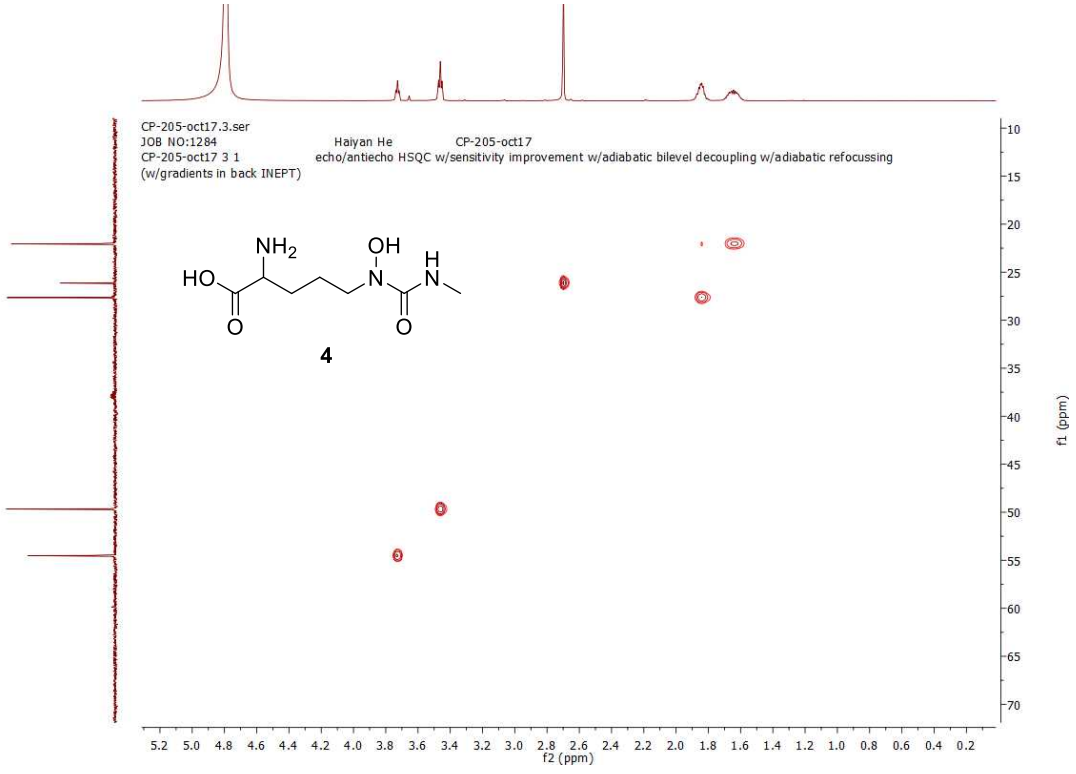


Figure S16b. 2D HSQC NMR spectrum of compound **4** in D_2O.

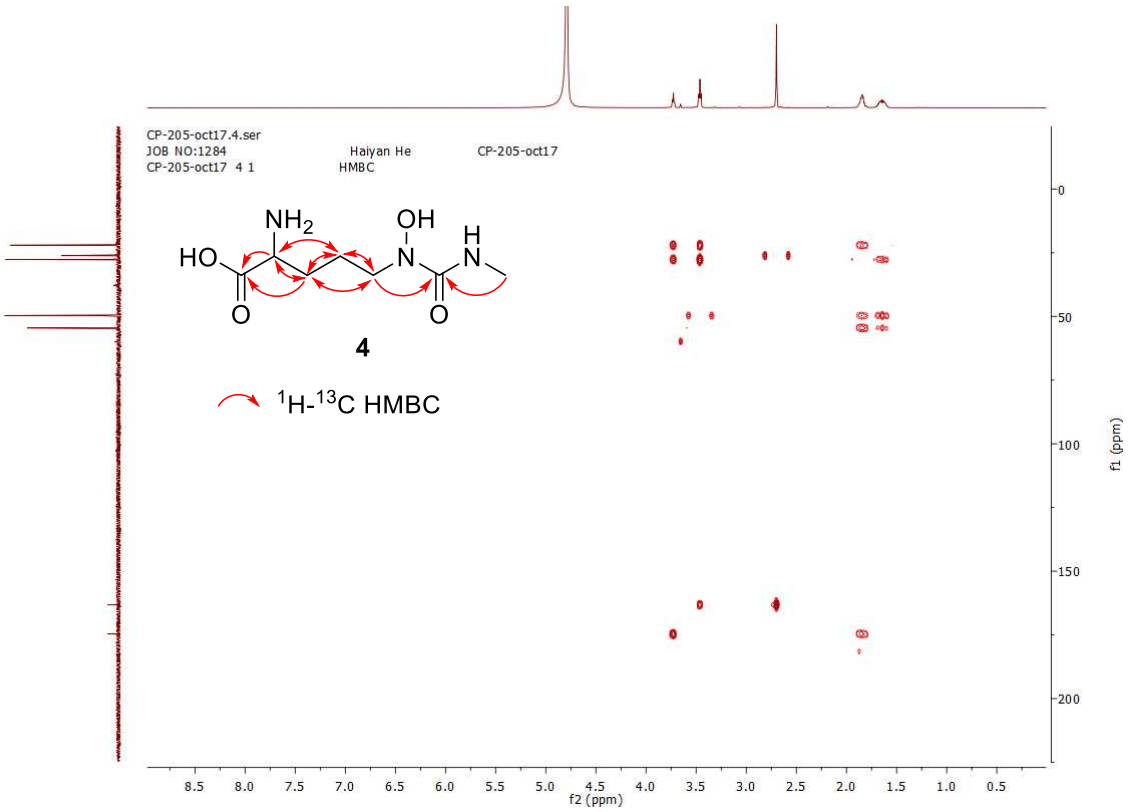


Figure S16c. 2D HMBC NMR spectrum of compound **4** in D_2O.

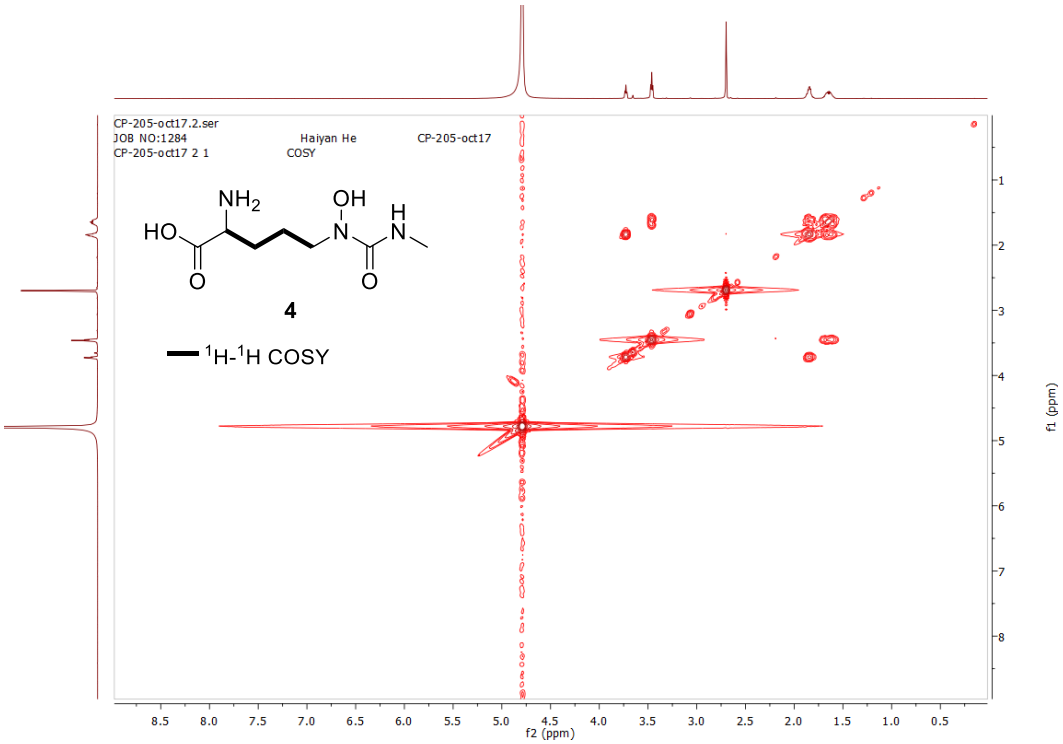


Figure S16d. 2D COSY NMR spectrum of compound **4** in D_2O.

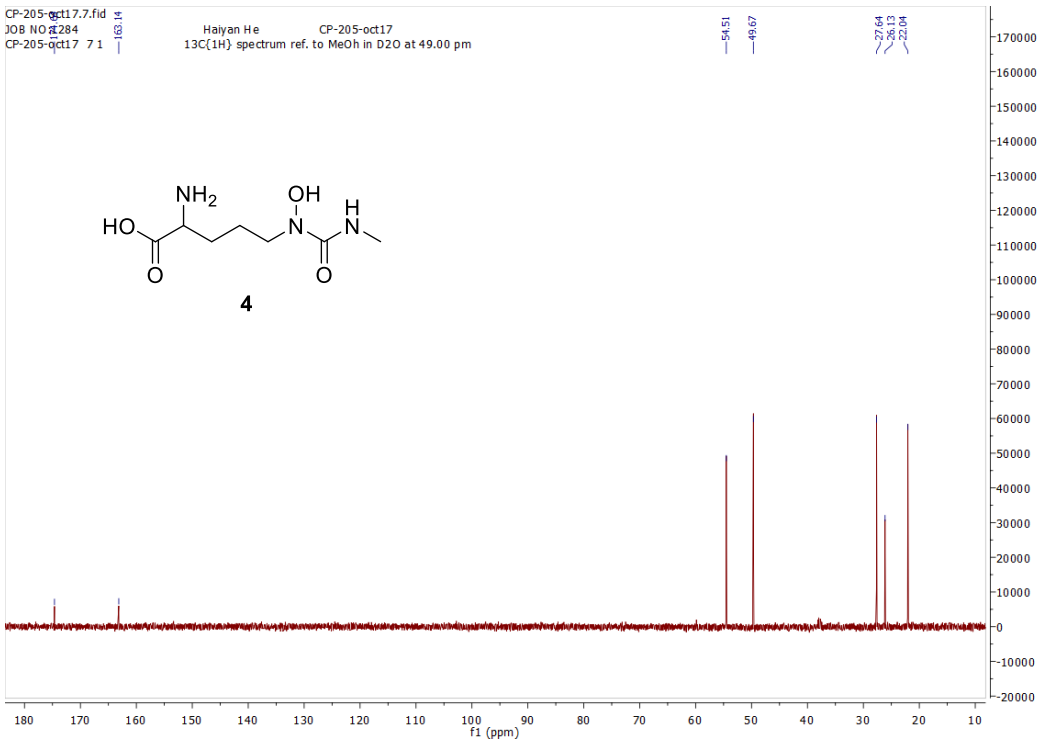


Figure S16e. ^{13}C NMR spectrum of compound **4** in D_2O. ^{13}C NMR (150 MHz, Deuterium Oxide): δ 174.6, 163.1, 54.51, 49.67, 27.64, 26.13, 22.04.

Figure S17

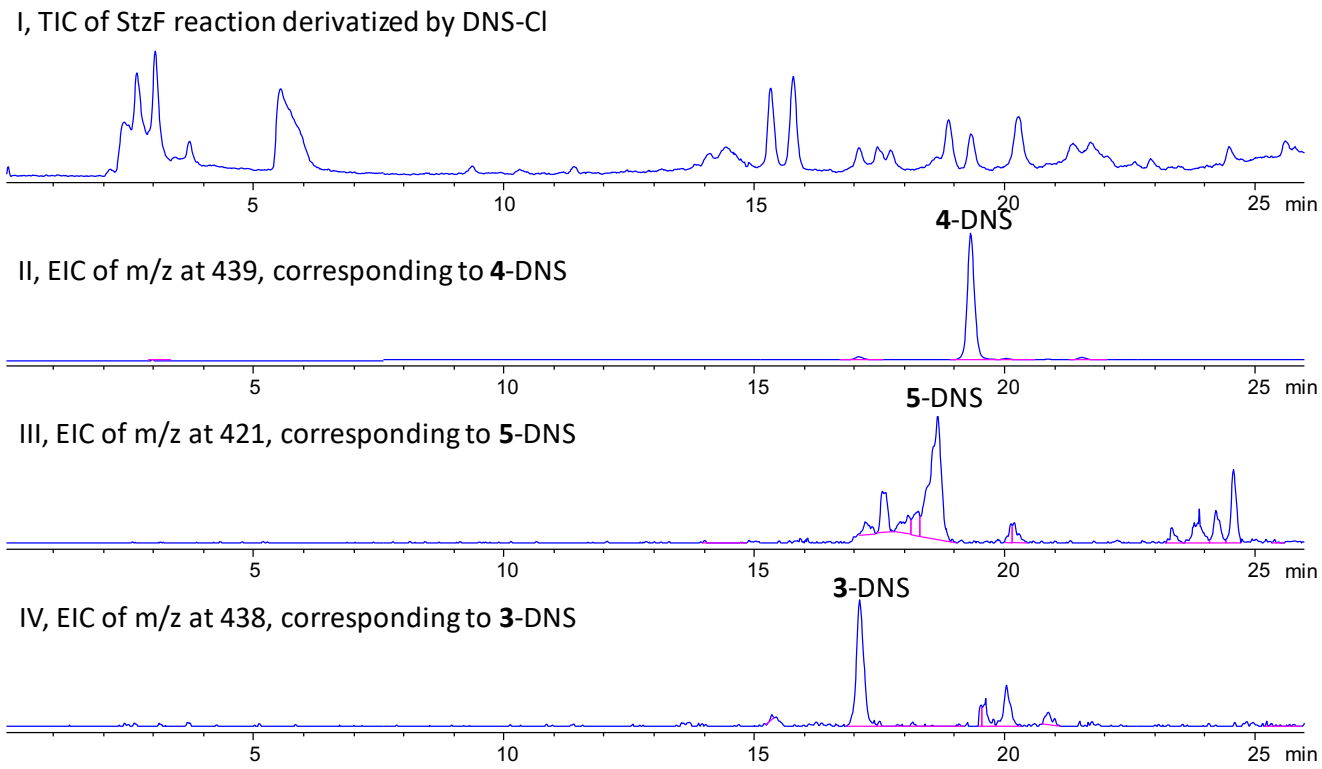


Figure S17. LC-MS analysis of StzF reaction products derivatized by DNS-Cl. Products catalyzed by StzF **3**, **4** and **5** all could be derivatized. Derivatives **4**-DNS was purified by semi-preparation of HPLC and its structure was elucidated by nuclear magnetic resonance spectroscopy (NMR). Derivatives **5**-DNS was purified and analyzed by HR-MS-MS.

Figure S18

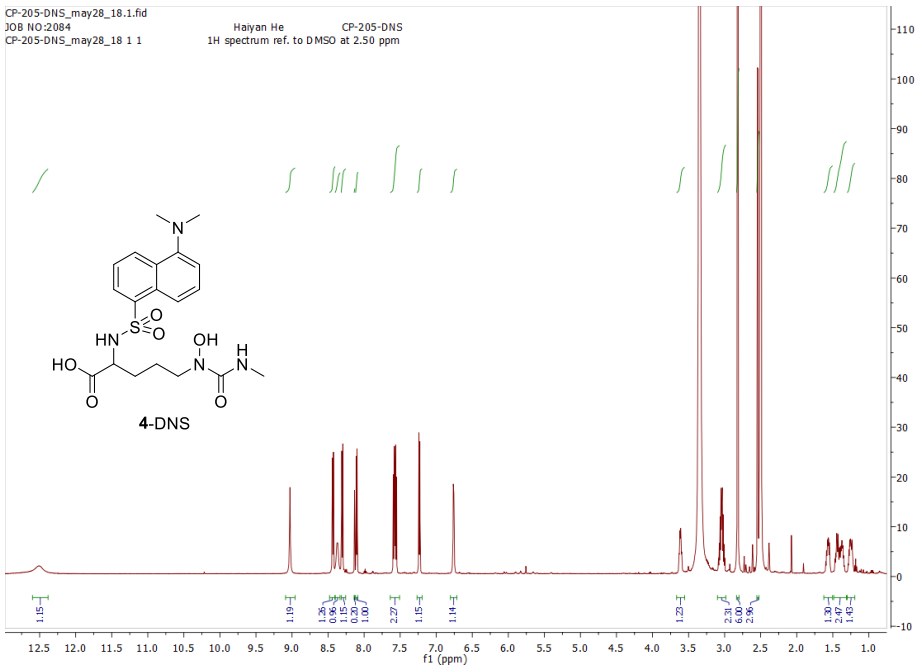


Figure S18a. ^{1}H NMR spectrum of compound **4**-DNS in DMSO-d$_6$. ^{1}H-NMR (600 MHz, DMSO-*d*$_6$): δ 12.51 (s, 1H), 9.03 (s, 1H), 8.43 (dt, *J* = 8.6, 1.2 Hz, 1H), 8.37 (d, *J* = 8.6 Hz, 1H), 8.30 (d, *J* = 8.6 Hz, 1H), 8.10 (dd, *J* = 7.3, 1.2 Hz, 1H), 7.58 (dd, *J* = 8.6, 7.3 Hz, 1H), 7.56 (dd, *J* = 8.6, 7.3 Hz, 1H), 7.23 (d, *J* = 7.3 Hz, 1H), 6.76 (q, *J* = 4.6 Hz, 1H), 3.61 (m, 1H), 3.04 (dq, 2H), 2.82 (s, 6H), 2.54 (d, *J* = 4.6 Hz, 3H), 1.60 – 1.20 (m, 4H).

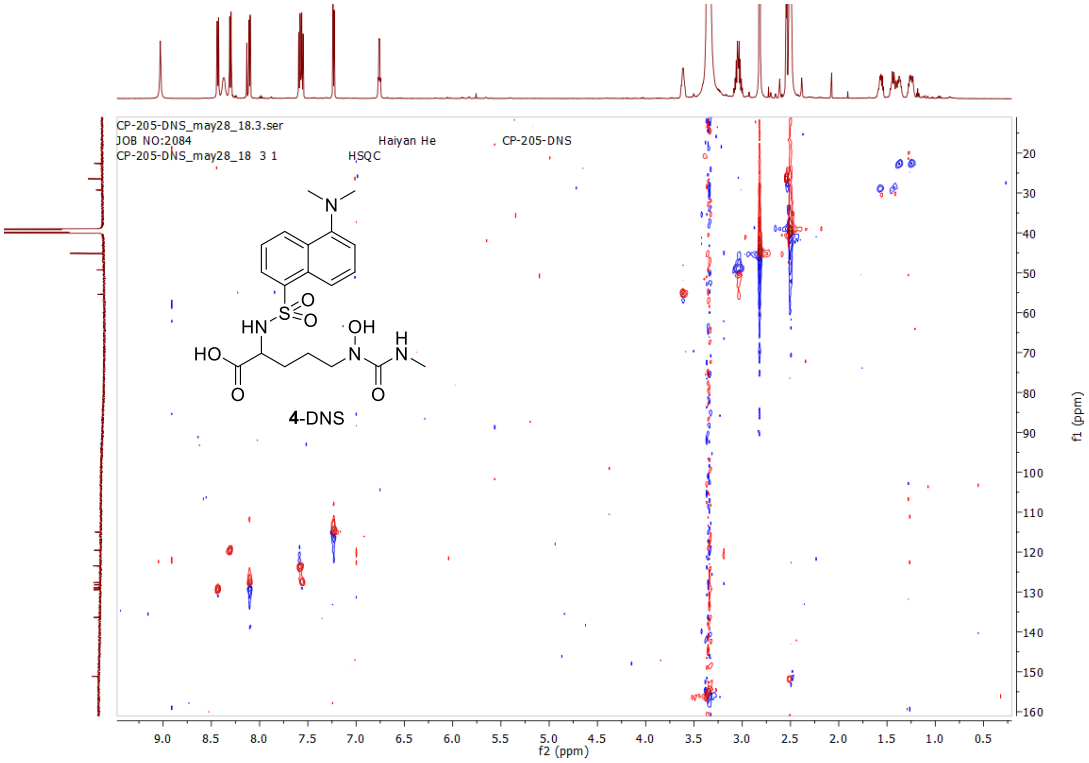


Figure S18b. 2D HSQC NMR spectrum of compound **4**-DNS in DMSO-d$_6$.

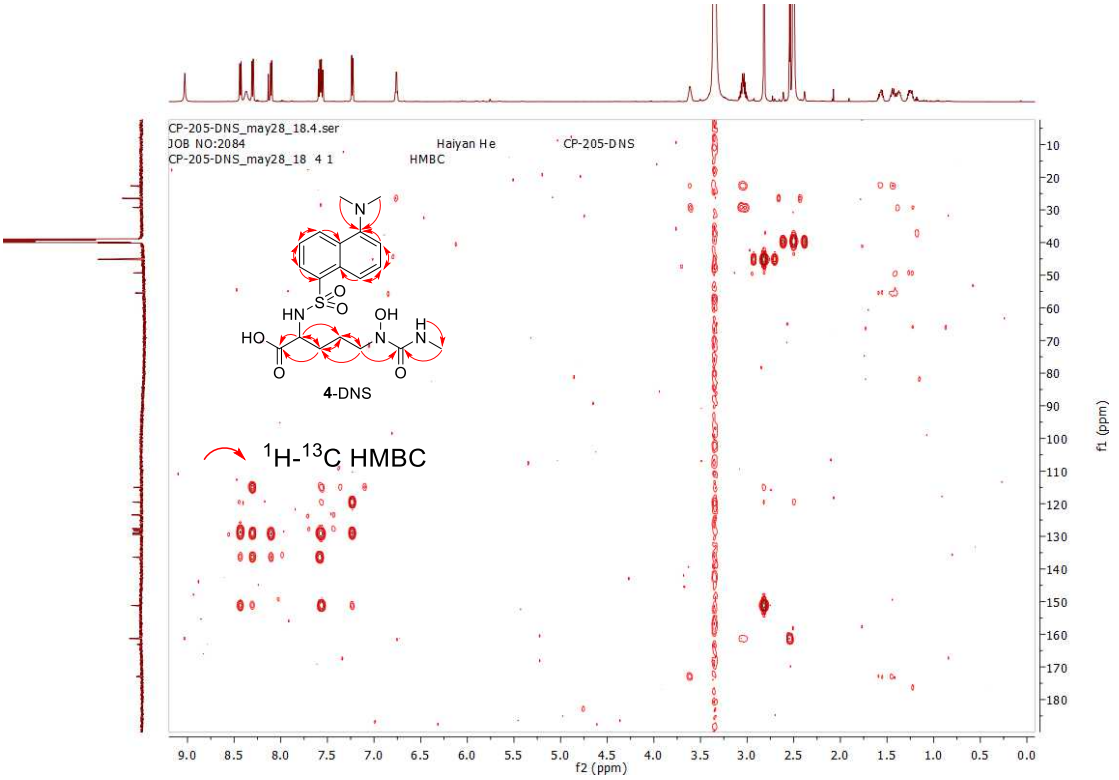


Figure S18c. 2D HMBC NMR spectrum of compound **4**-DNS in DMSO-d_6.

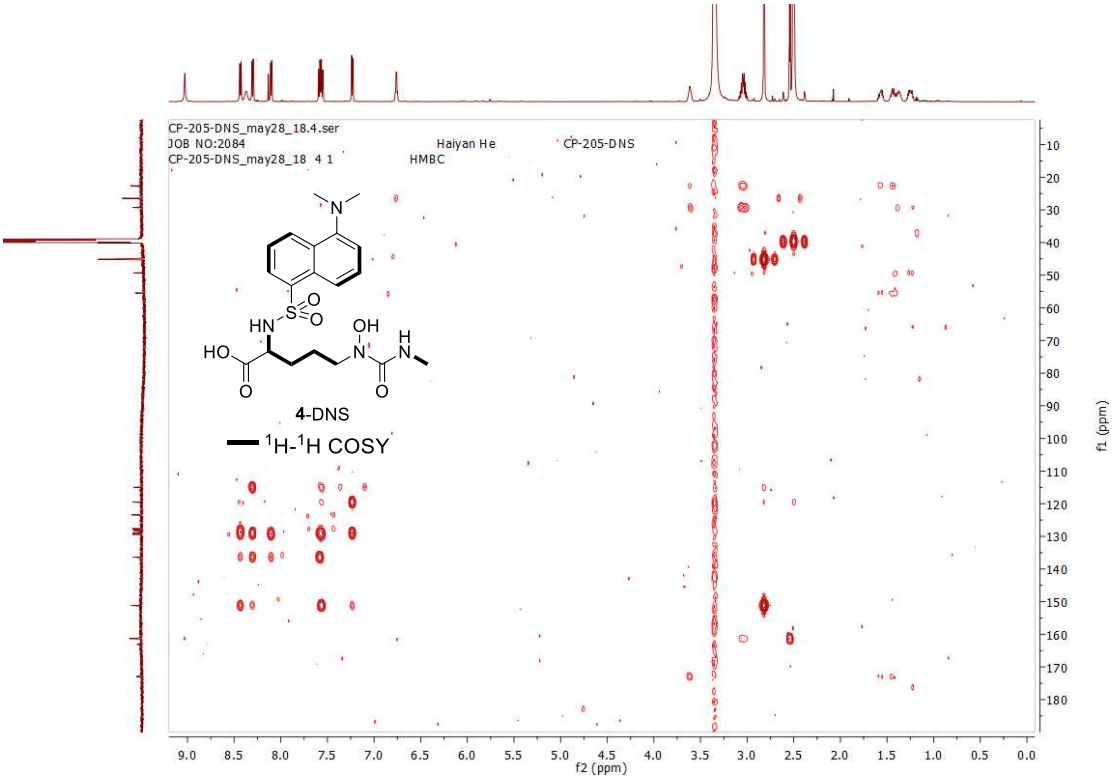


Figure S18d. 2D COSY NMR spectrum of compound **4**-DNS in DMSO-d_6.

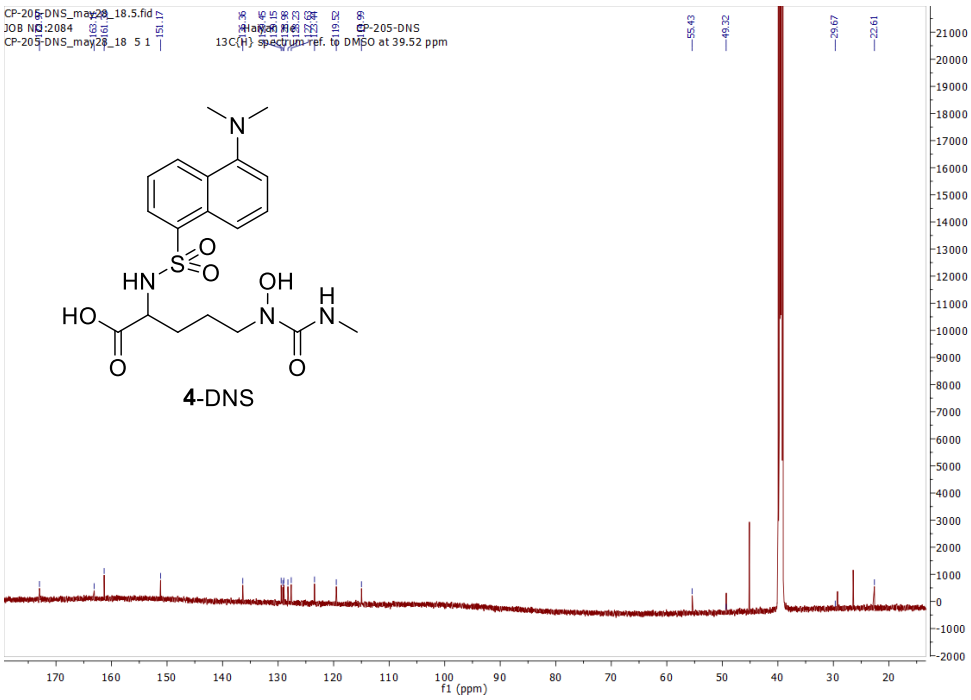
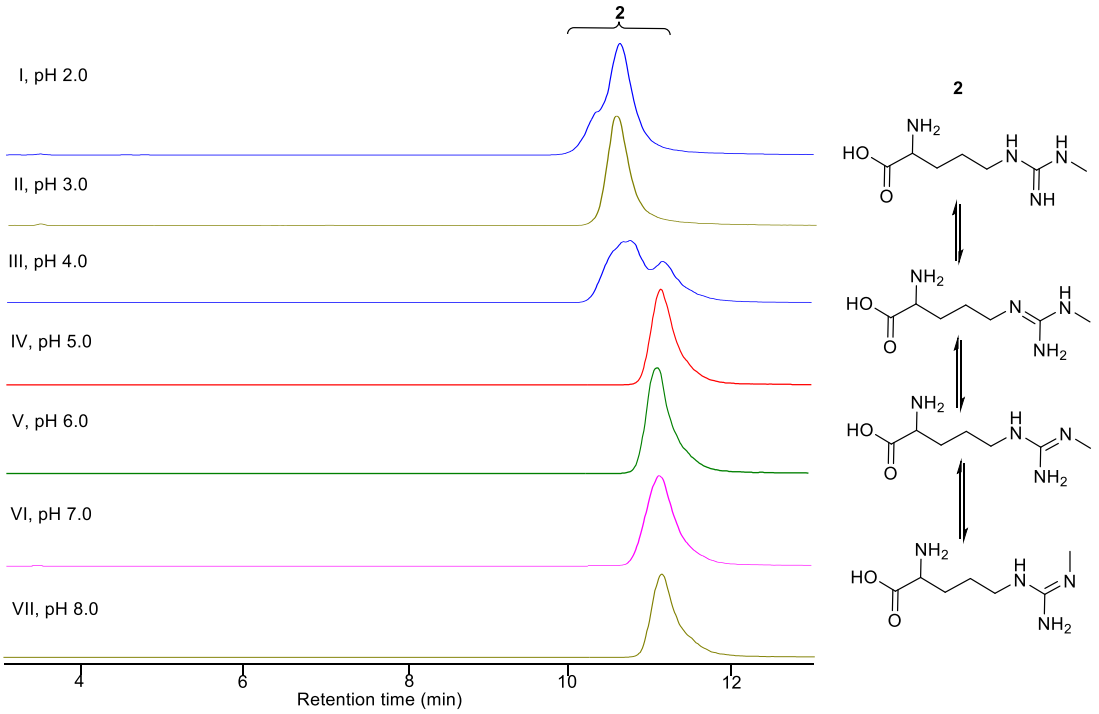


Figure S18e. ^{13}C NMR spectrum of compound **4**-DNS in DMSO-d$_6$. ^{13}C NMR (150 MHz, DMSO-*d*$_6$): δ 172.97, 161.33, 151.17, 136.36, 129.45, 129.15, 128.98, 128.23, 127.63, 123.44, 119.52, 114.99, 55.43, 49.46, 45.13, 29.26, 26.73, 22.70.

Figure S19

a



b

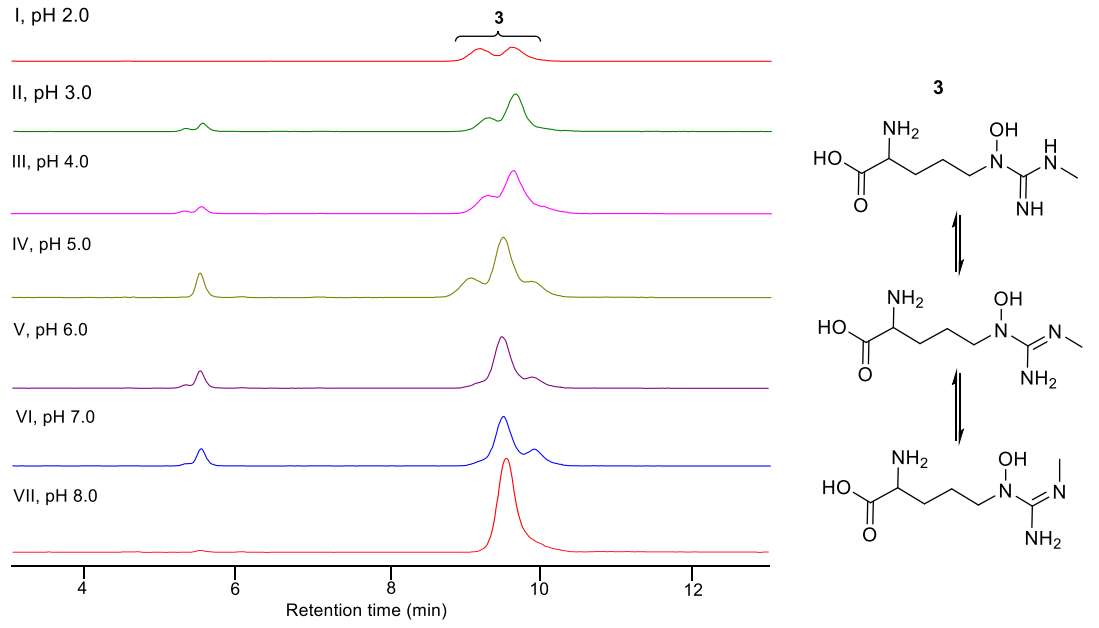
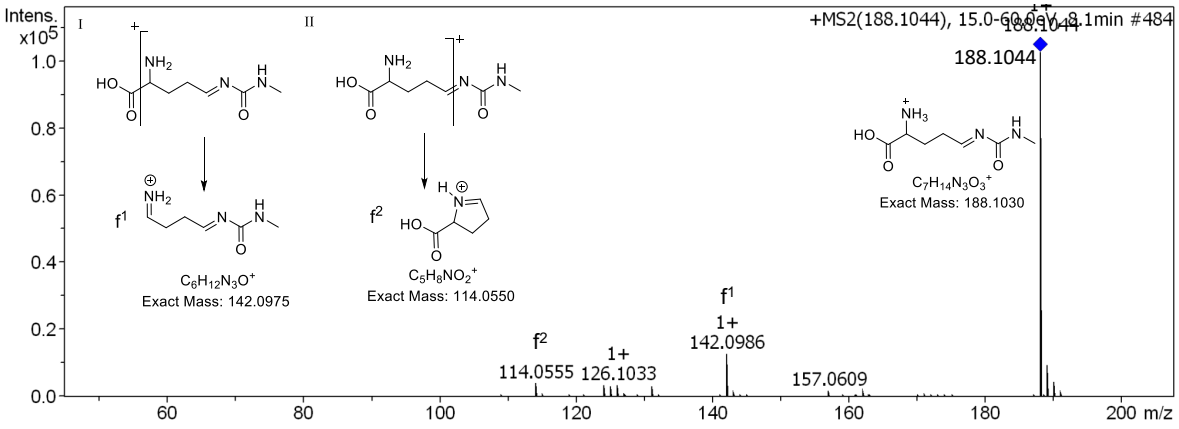


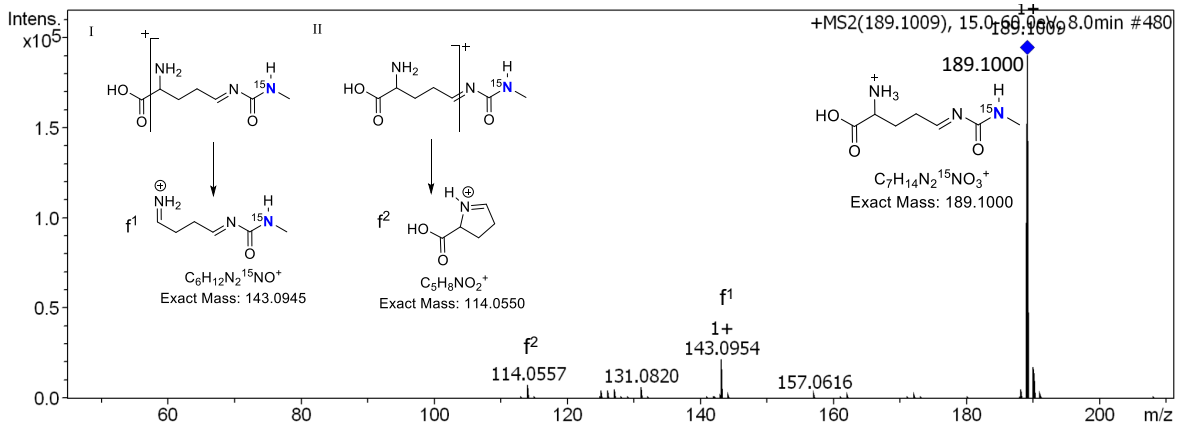
Figure S19. LC-MS analysis of compound **2** and **3** at different pH values. 500 uM of compound was dissolved in 50 mM Tris buffer adjusted to different pH values (2.0 ~ 8.0). Left, Analysis results by LC-MS; Right, possible tautomerization of guanidino group in the two compounds. **a**) EIC of specific *m/z* at 189 for compound **2**; **b**) EIC of specific *m/z* at 205 for compound **3**.

Figure S20

a



b



c

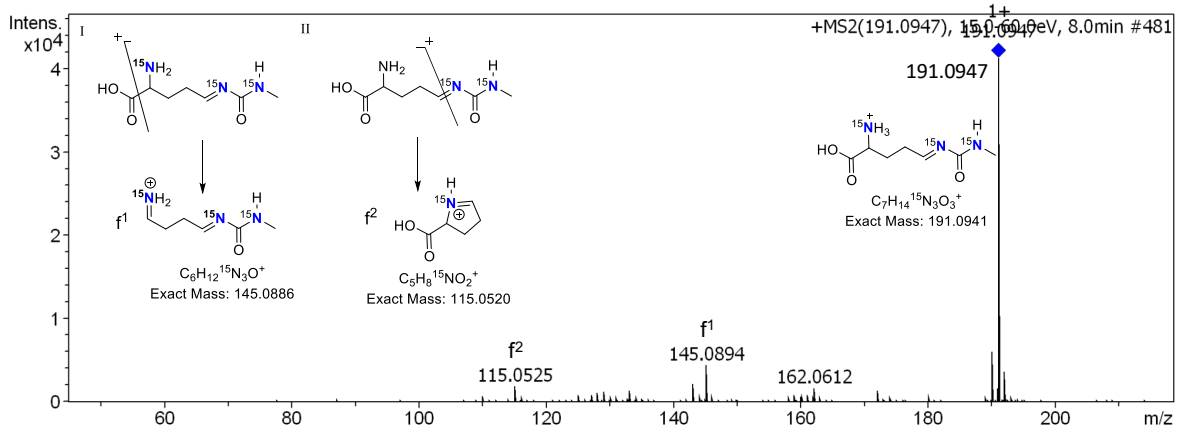


Figure S20. LC-MS/MS analysis of compound **5**. Compound **5** was generated from StzE/StzF assays using different labelled arginines as the substrate. **a**, unlabelled arginine assay; **b**, $^{15}N_2$-arginine assay; **c**, $^{15}N_4$-arginine assay.

Figure S21

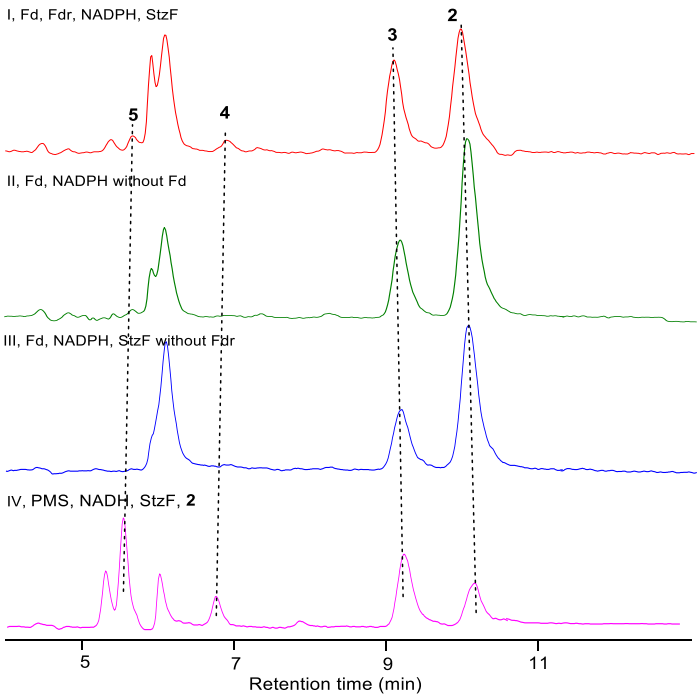


Figure S21. LC-MS analysis of StzF reactions using spinach ferredoxin (Fd) and spinach ferredoxin reductase (Fdr) as the electron transfer system. I, Reaction containing Fd, Fdr, StzF, $FeSO_4$ and NADPH; II, control reaction without Fd added; III, control reaction without Fdr added; IV, positive control of PMS and NADH added.

Figure S22

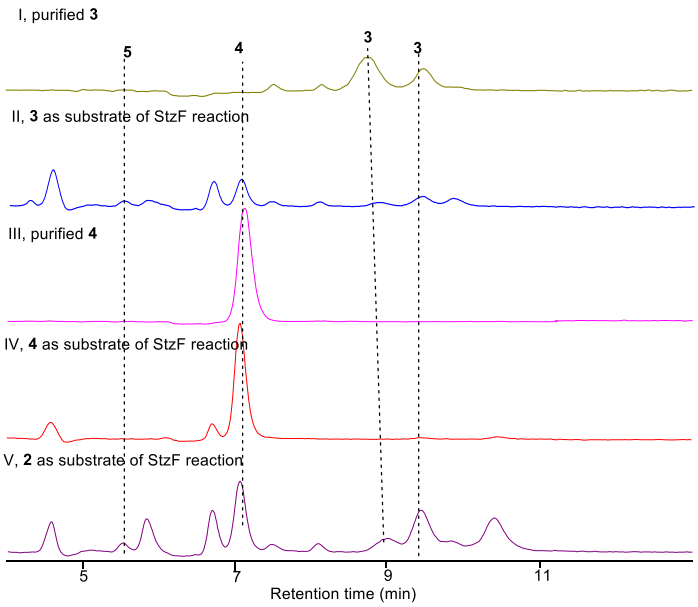
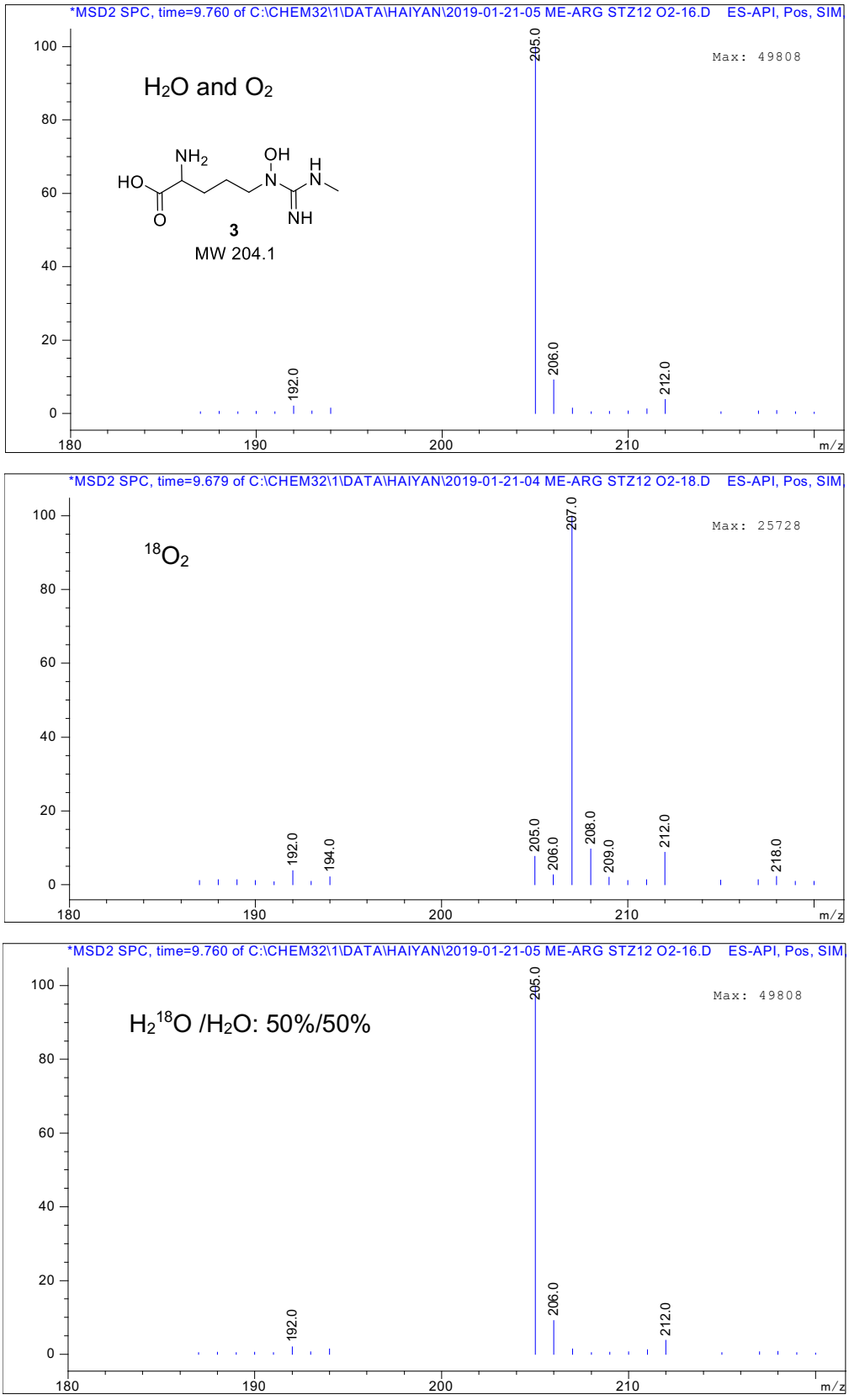


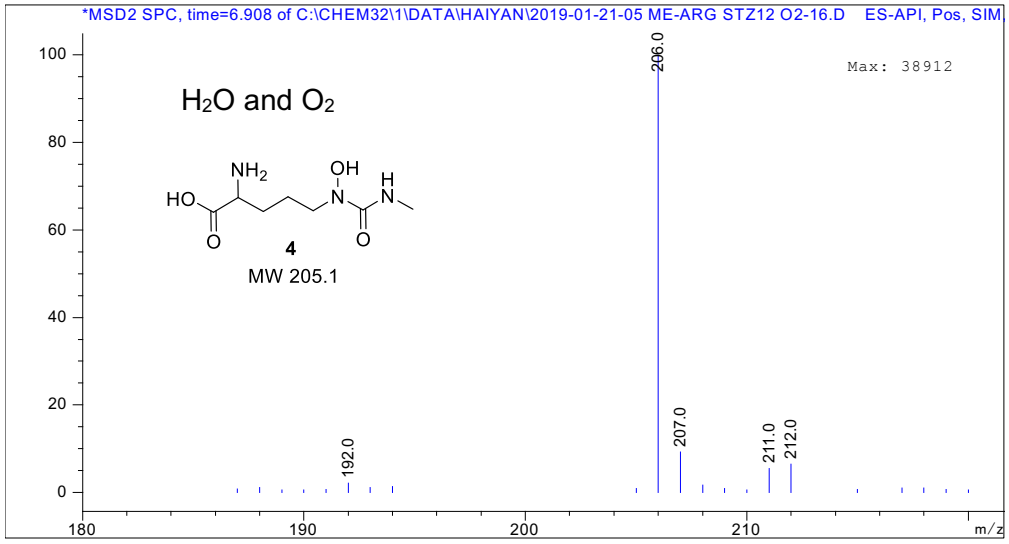
Figure S22. LC-MS analysis of StzF reactions using purified **3** and **4** as substrates. I, purified **3** in water; II, **3** assayed with StzF and PMS, NADH; III, purified **4** in water; IV, **4** assayed with StzF and PMS, NADH; V, **2** assayed with StzF and PMS, NADH, as a positive control. All reactions were incubated at 30 ºC for 1 h before analysis.

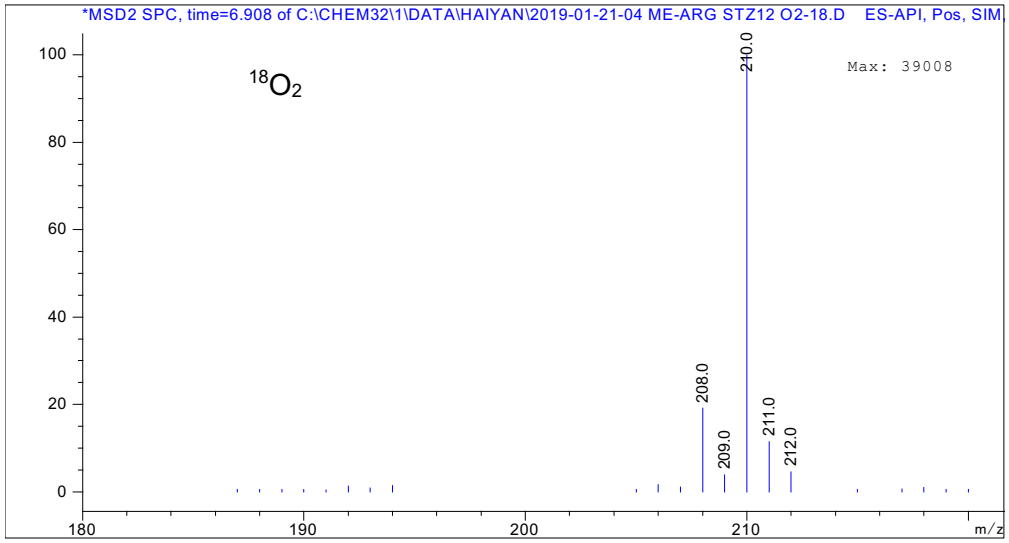
Figure S23

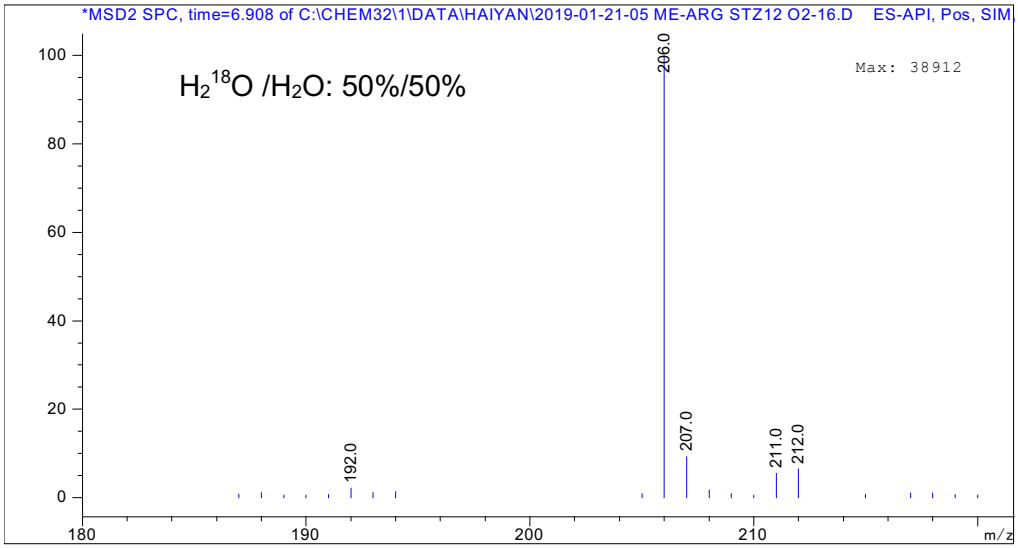
a


*MSD2 SPC, time=9.760 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-05 ME-ARG STZ12 O2-16.D ES-API, Pos, SIM
Max: 49808
H2O and O2
NH2
OH
H
HO
N
N
O
NH
3
MW 204.1
205.0
206.0
212.0
192.0
*MSD2 SPC, time=9.679 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-04 ME-ARG STZ12 O2-18.D ES-API, Pos, SIM
Max: 25728
18O2
207.0
205.0
206.0
208.0
209.0
212.0
218.0
192.0
194.0
*MSD2 SPC, time=9.760 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-05 ME-ARG STZ12 O2-16.D ES-API, Pos, SIM
Max: 49808
H2 18O /H2O: 50%/50%
205.0
206.0
212.0
192.0
100
80
60
40
20
0
180
190
200
210
m/z

b


*MSD2 SPC, time=6.908 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-05 ME-ARG STZ12 O2-16.D ES-API, Pos, SIM
Max: 38912
H2O and O2
NH2
OH
HO
H
N
N
O
O
4
MW 205.1
206.0
207.0
192.0
211.0
212.0
100
80
60
40
20
0
180
190
200
210
m/z


*MSD2 SPC, time=6.908 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-04 ME-ARG STZ12 O2-18.D ES-API, Pos, SIM
Max: 39008
18O2
210.0
208.0
209.0
211.0
212.0
100
80
60
40
20
0
180
190
200
210
m/z


*MSD2 SPC, time=6.908 of C:\CHEM32\1\DATA\HAIYAN\2019-01-21-05 ME-ARG STZ12 O2-16.D ES-API, Pos, SIM
Max: 38912
H2 18O /H2O: 50%/50%
206.0
207.0
192.0
211.0
212.0
100
80
60
40
20
0
180
190
200
210
m/z

c

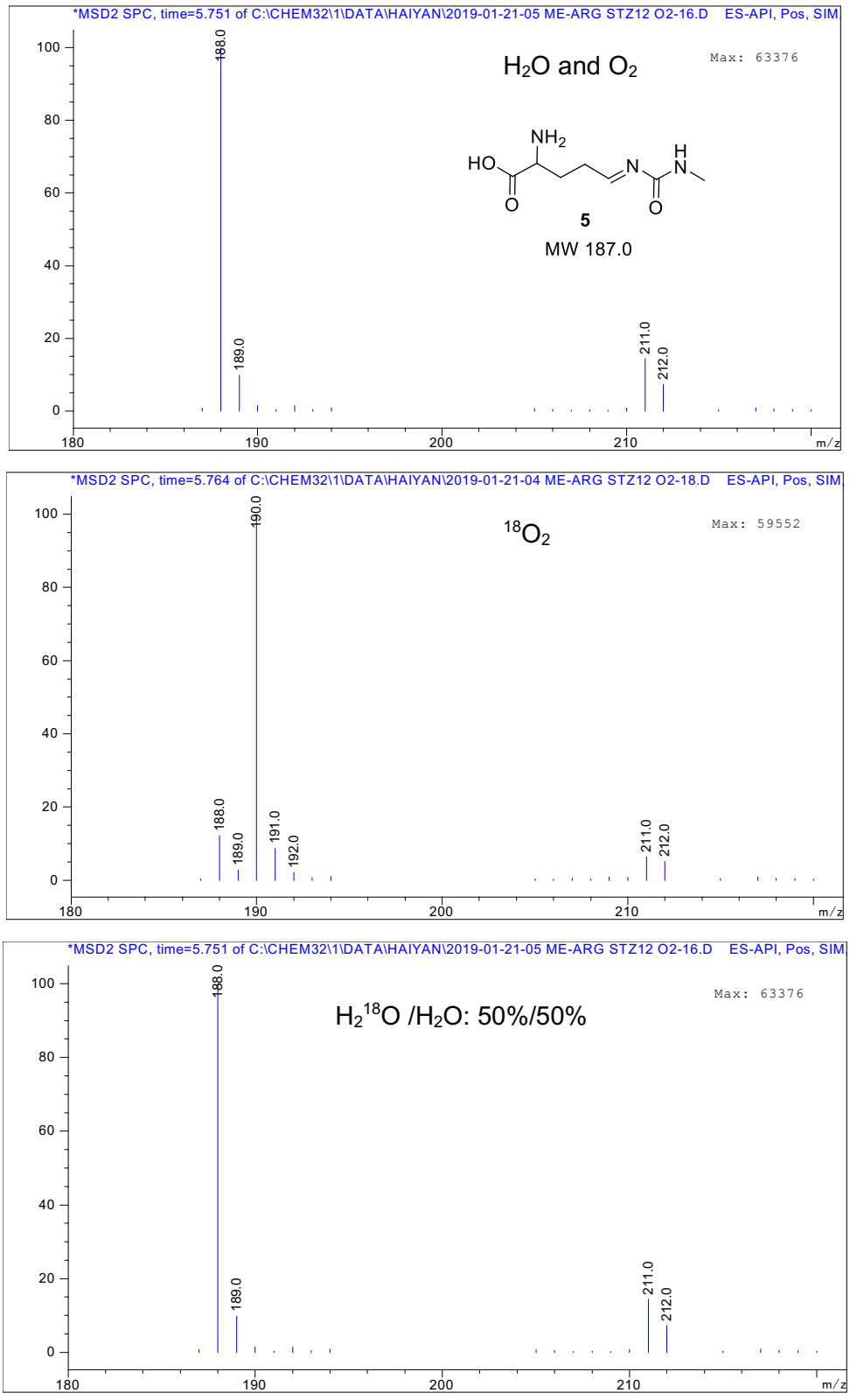


Figure S23. MS spectra of products from StzF reactions run in water, $^{18}O_2$, or 50/50 $H_2{}^{18}O/H_2O$. **a**, MS spectra of product **3**; **b**, MS spectra of product of **4**; **c**, MS spectra of product **5**. Structures of each compound are shown with the predicted molecular weight.

Figure S24

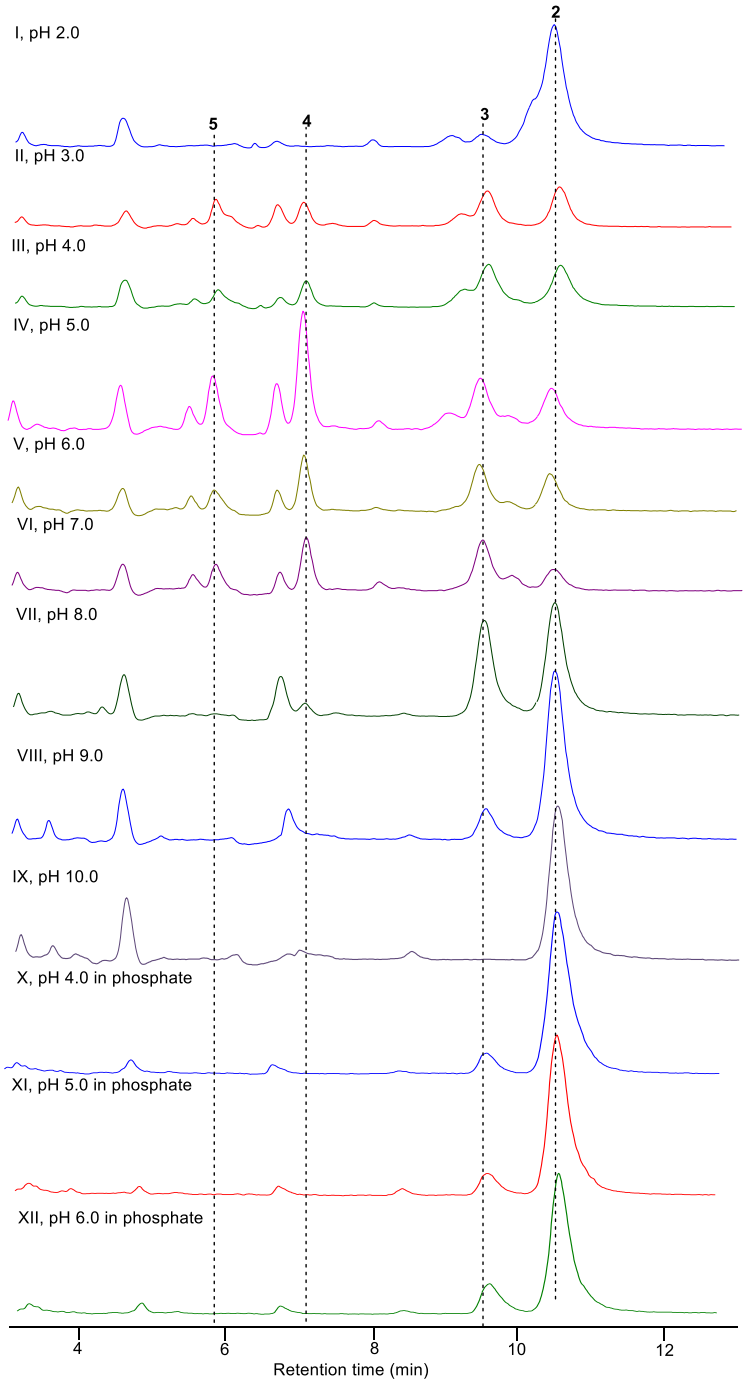
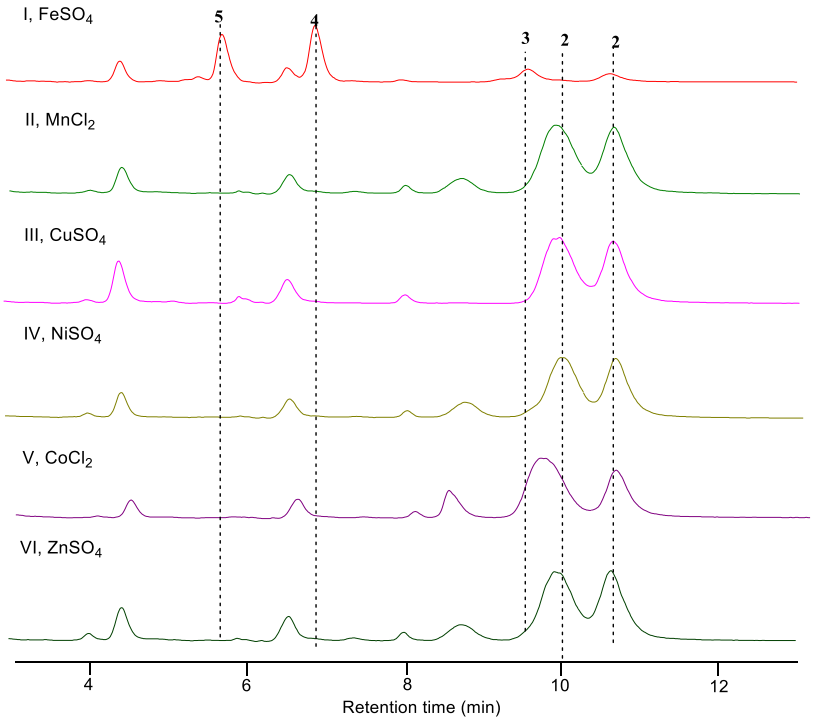


Figure S24. Test of StzF activity at different pH values. Each assay was detected on HILIC column by LC-MS. I to IX refer to EIC of reactions in Tris buffer (50 mM) at pH from 2.0 to 10.0, respectively. X to XII refer to EIC of reactions in sodium phosphate buffer at pH 4, 5 and 6. StzF is stable at pH 3.0~10.0, however at pH 2.0, it precipitates immediately and has no activity.

Figure S25

a



b

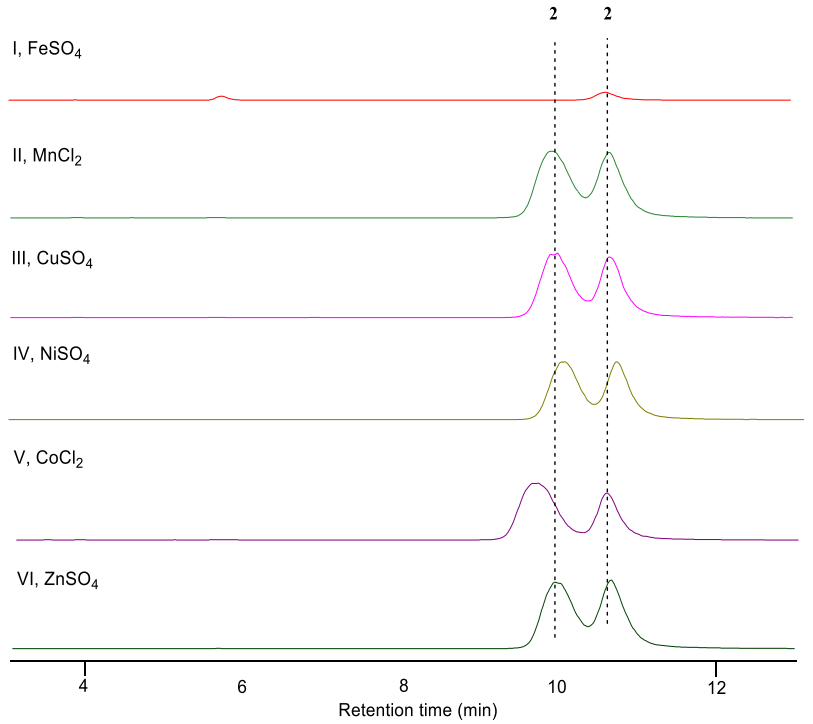
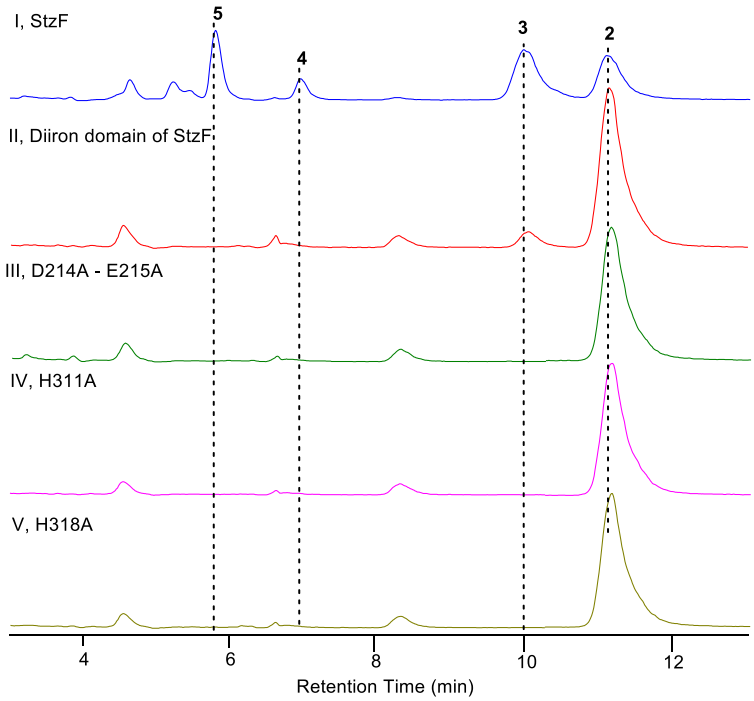


Figure S25. Test of StzF activity in the present of different metal ions. **a**, EIC of LC-MS detection of the reactions; **b**, EIC of specific *m/z* at 189, corresponding to compound **2**, the substrate of StzF. Note: we observe a retention time shift for compound **2** in Trace V, which may be due to the high concentration of $CoCl_2$.

Figure S26.

a



b

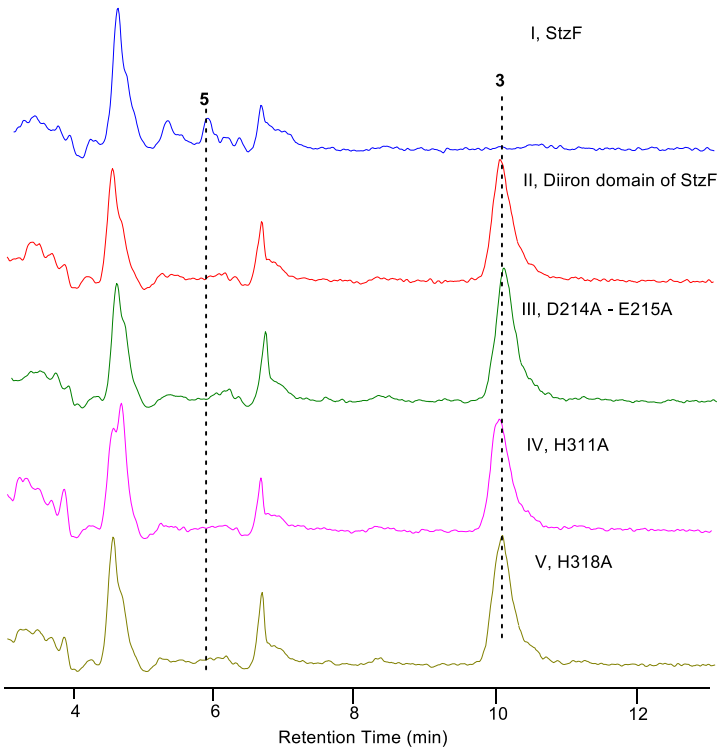
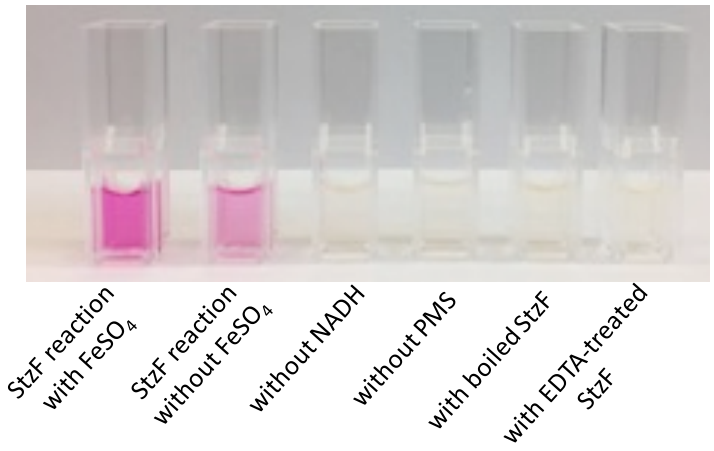


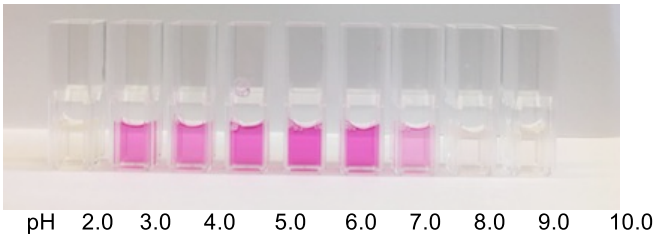
Figure S26. LC-MS analysis of in vitro reactions catalyzed by StzF mutants. **a**) Compound **2** (0.5 mM) as the substrate. **b**) Purified compound **3** (0.1 mM) as the substrate. All reactions were incubated at 30 ºC for 4 h before analysis.

Figure S27

a



b



c

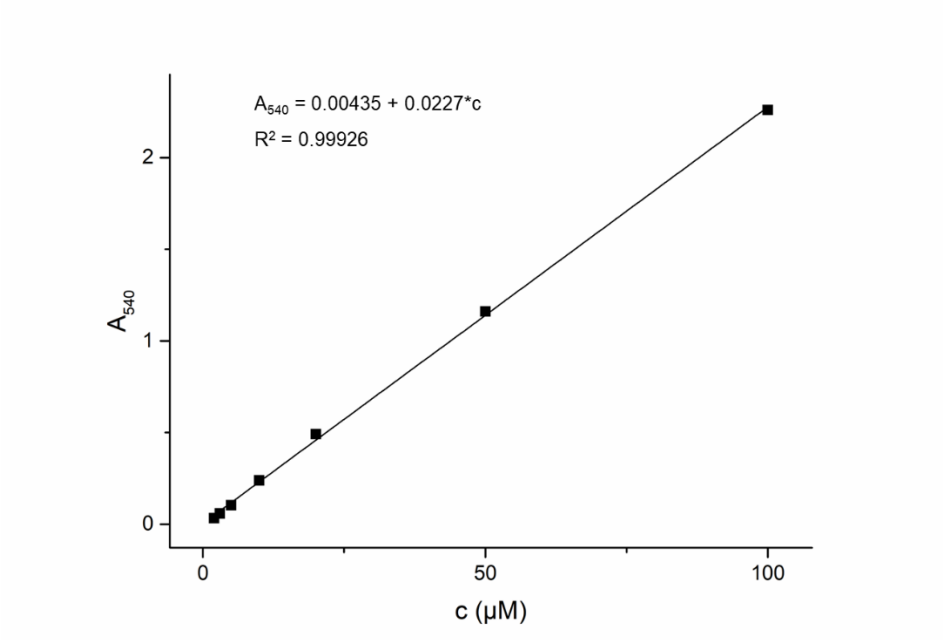


Figure S27. Nitrite detection of StzF reaction using Griess Reagent. **a**, Detection of the whole StzF reaction and controls. **b**, Detection of StzF reactions at different pH values. **c,** Standard curve of Griess Reagent for nitrite detection. $NaNO_2$ (2 μM, 3 μM, 5 μM, 10 μM, 20 μM, 50 μM, 100 μM) was used as the standard.

Figure S28

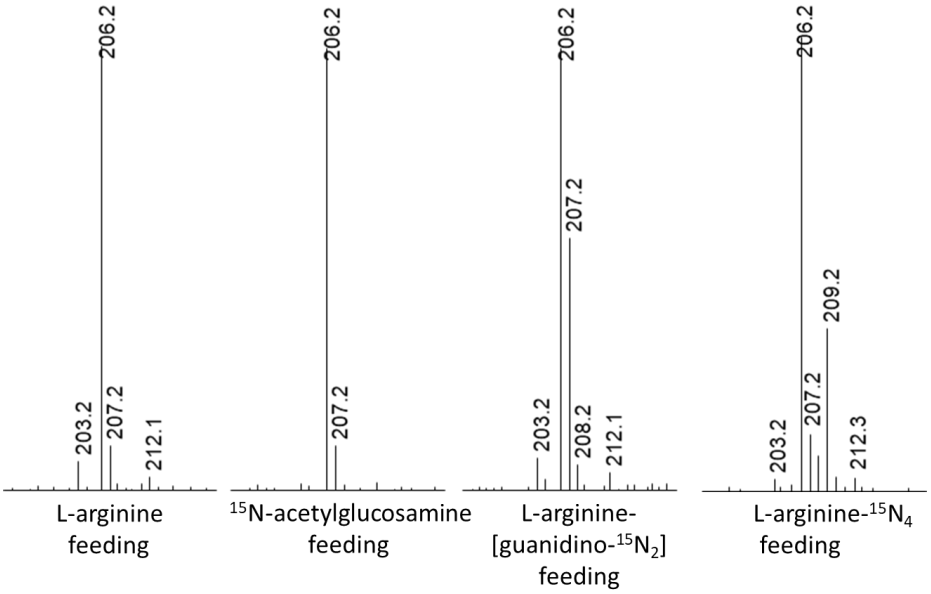
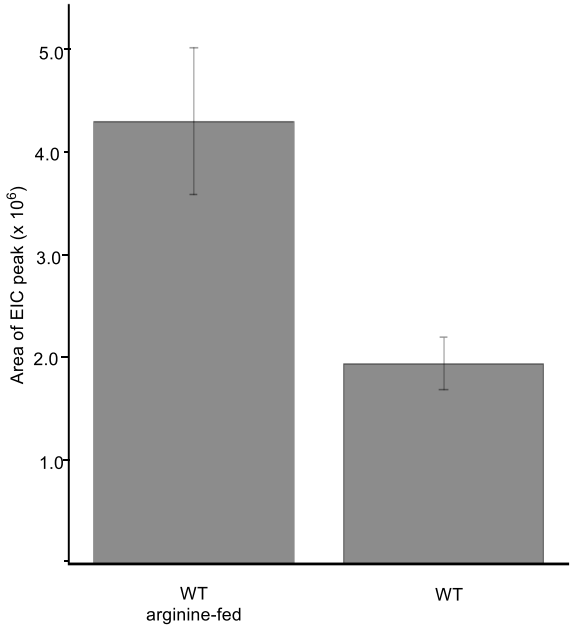


Figure S28. LC-MS detection of signals consistent with compound **4** from extracts supplemented with ^{15}N-labeled precursors.

Figure S29

a



b

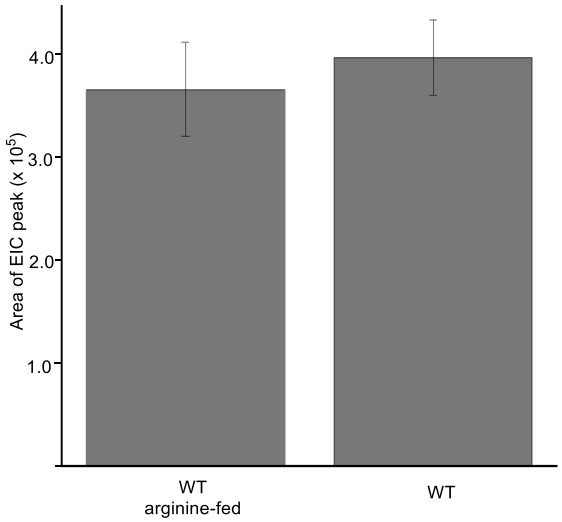
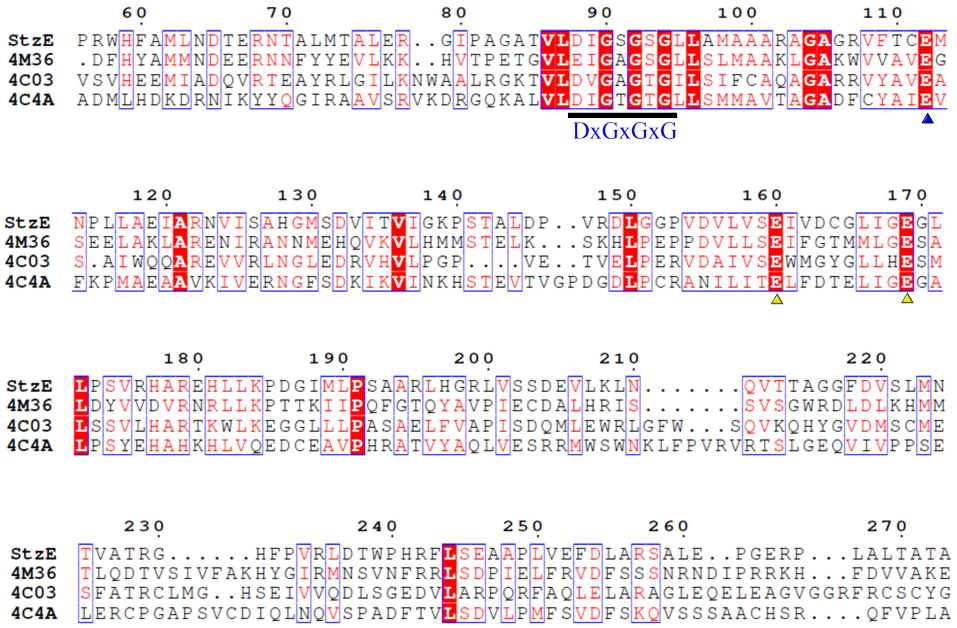


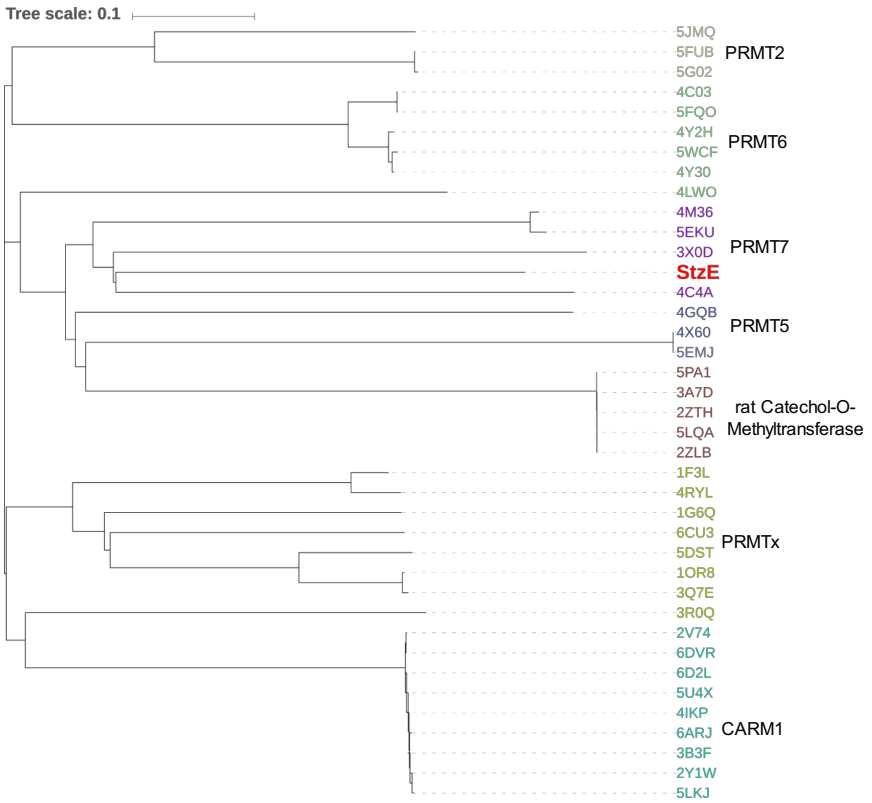
Figure 29. *In vivo* production of compound **4** (**a**) and compound **1** (**b**). The peak integration of the EIC (*m/z* at 206 for **4** and 248 for **1**) was used for quantification.

Figure S30

a



b



c

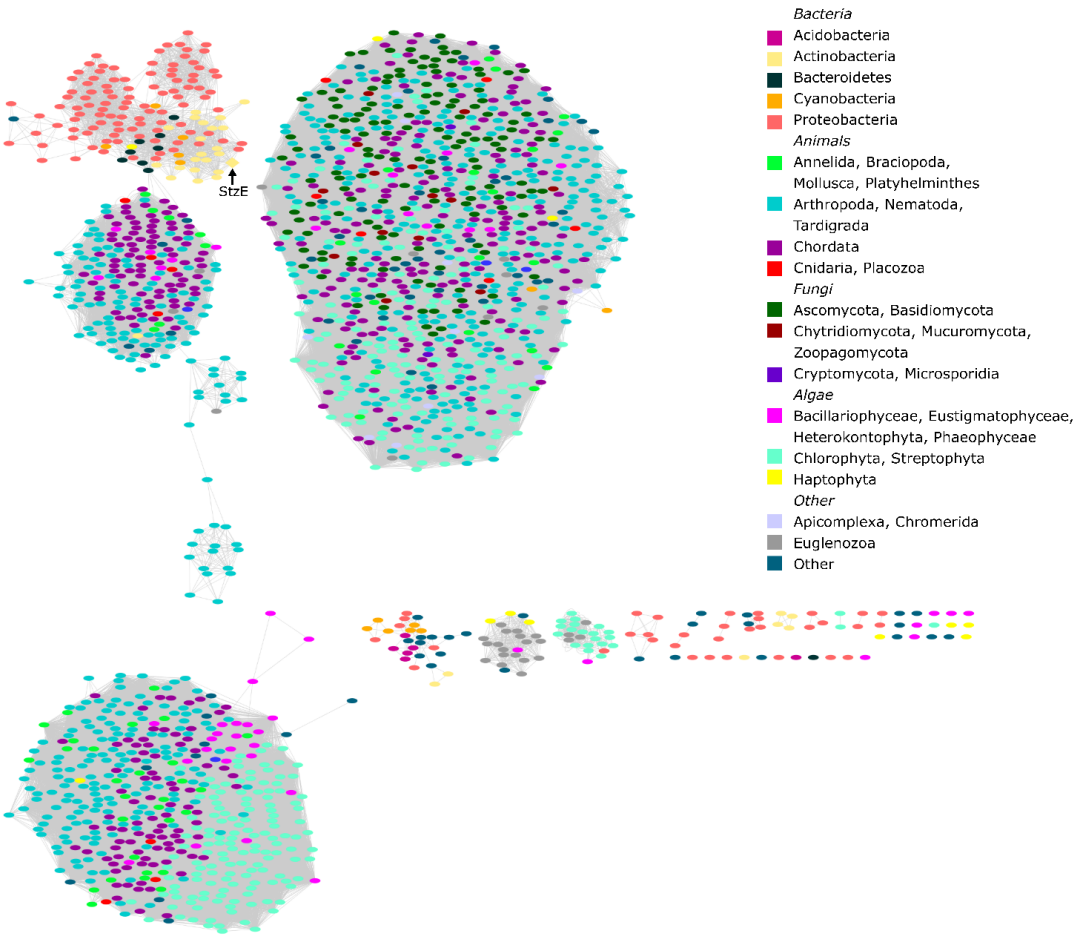


Figure S30. Bioinformatic analysis of StzE with protein arginine methylthansferases (PRMTs). **a**, Multisequence alignment was created with Clustal Omega[23] and displayed with ESPript 3.0.[24] DxGxGxG motif and Glu residue marked by the blue triangle are critical for SAM binding; two Glu residues marked by yellow triangles are guanidino binding sites. Structures are labelled with their PDB IDs: 4M36, a PRMT7 from *Trypanosoma brucei*; 4C03, a PRMT6 from *Mus musculus*; and 4C4A, a PRMT7 from *Mus musculus*. **b**, Phylogenetic analysis of StzE with PRMTs. To obtain homologs of StzE, an Entity-based search was carried out in Protein Data Bank (PDB). The entities sharing the same protein sequence from 145 structures in total were filtered and 38 proteins were remaining. The phylogenetic tree was built with MUSCLE[25] and displayed with iTOL.[26] PRMTx represents PRMT3 (1F3L and 4RYL), PRMT1 (1OR8 and 3Q7E), PRMT8 (5DST), one yeast arginine transferase (1G6Q) and one PRMT from *Naegleria fowleri* (6CU3). CARM1 (coactivator-associated arginine methyltransferase 1) is also named PRMT4. **c**, Sequence similarity network of StzE with protein arginine methylthansferases (PRMTs). 3000 most closely related sequences were selected and computed using EFI-EST[27] to make the network. The network was visualised by Cytoscape.[28] Nodes are colored according to the organism's phylum and sequences with identity that exceeds 80% are represented by a single node. Edges connect sequences that align with an E-value threshold of 1.0×10^{-54}. The yFiles organic layout[29] was used to visualise the network, meaning that the length of the edges does not precisely correlate to the similarity of the connected sequences.

Figure S31

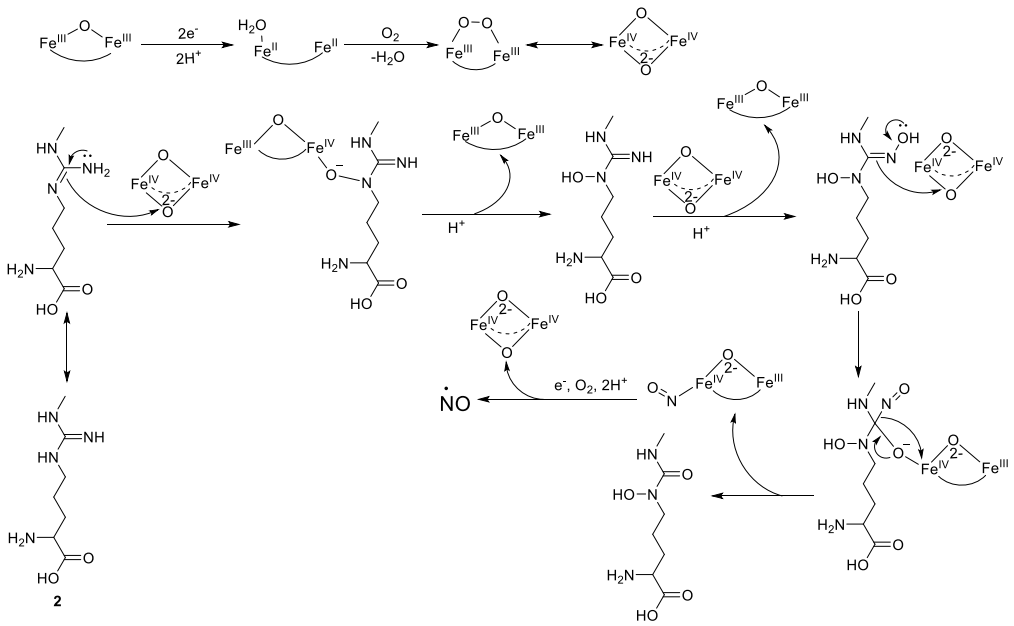


Figure S31. Proposed mechanism for diiron-catalyzed oxidation of **2**. Whether the cupin domain participates the oxidation directly or not is not known. Here we just show one possibility of activation of oxygens and oxidation induced by the diiron centre of StzF.

Figure S32

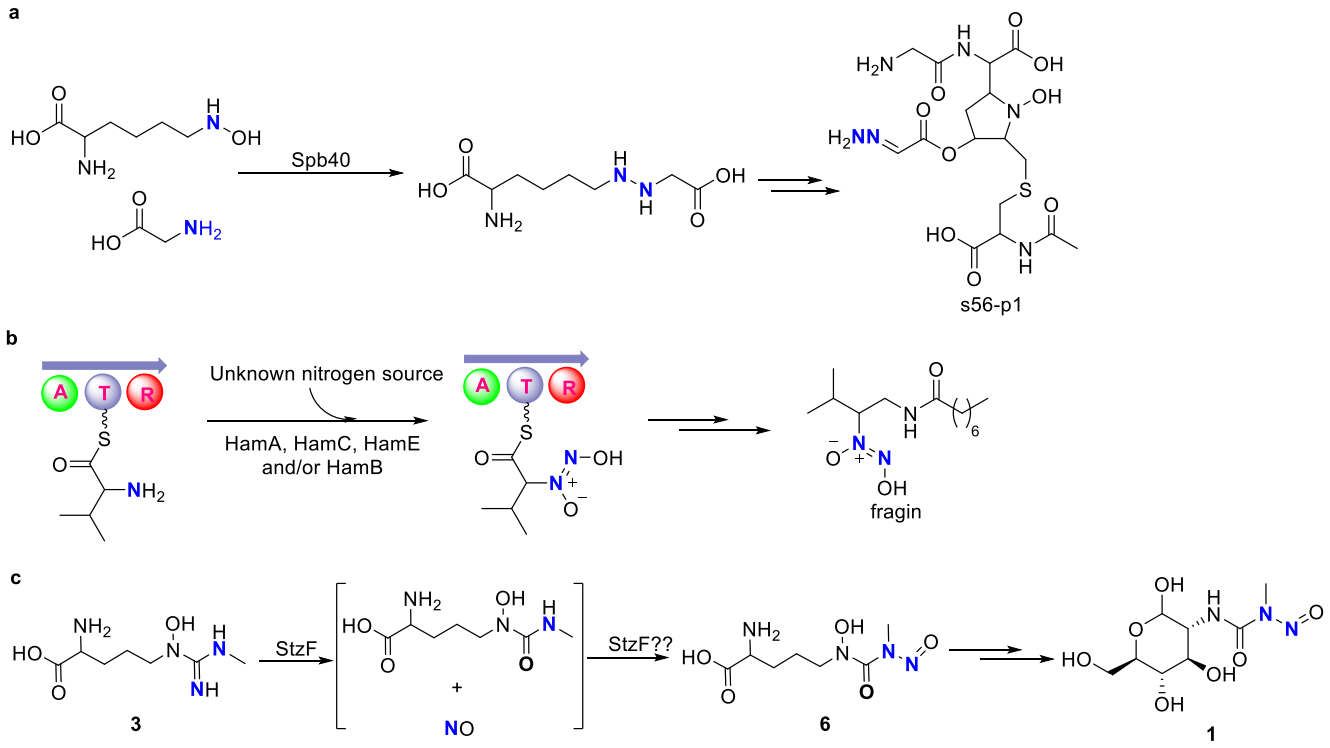


Figure S32. Natural products with N-N bonds linked to gene clusters encoding cupin domain-containing enzymes. **a**, The reaction catalyzed by Spb40, a cupin and MetRS-like didomain protein. **b**, Biosynthetic pathway of fragin. HamB, a free standing cupin protein, is encoded in the gene cluster. **c**, Proposed *N*-nitroso formation catalyzed by StzF, a diiron and cupin didomain protein, in streptozocin biosynthesis.

Figure S33

Biosynthetic gene cluster of streptozocin from NRRL3125 (only biosynthetic genes included)

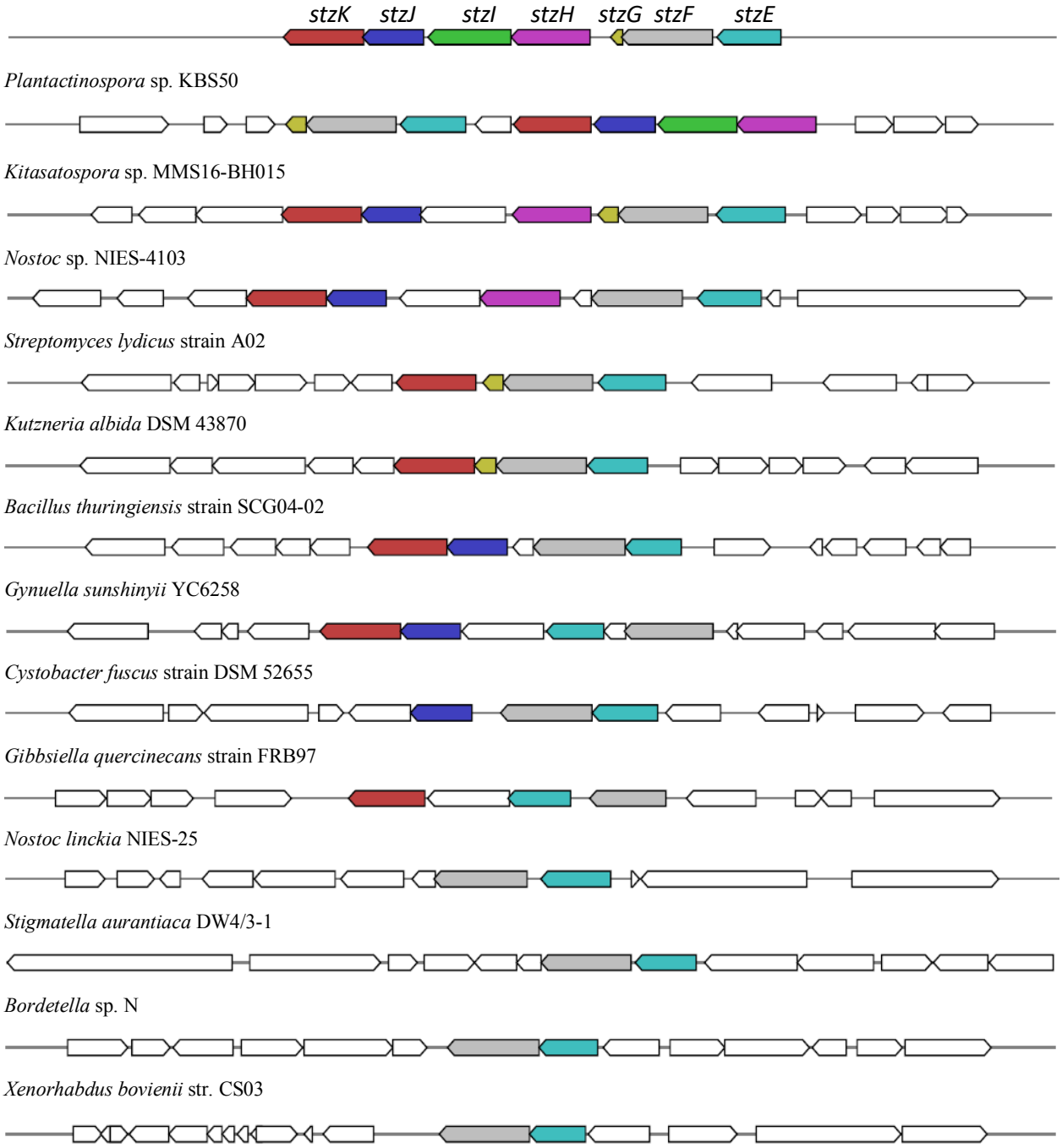


Figure S33. Analysis of distribution of *stzE-stzF* gene pairs by MutiGeneBlast.[30] Homologous genes from different strains are shaded by the same color.

Figure S34

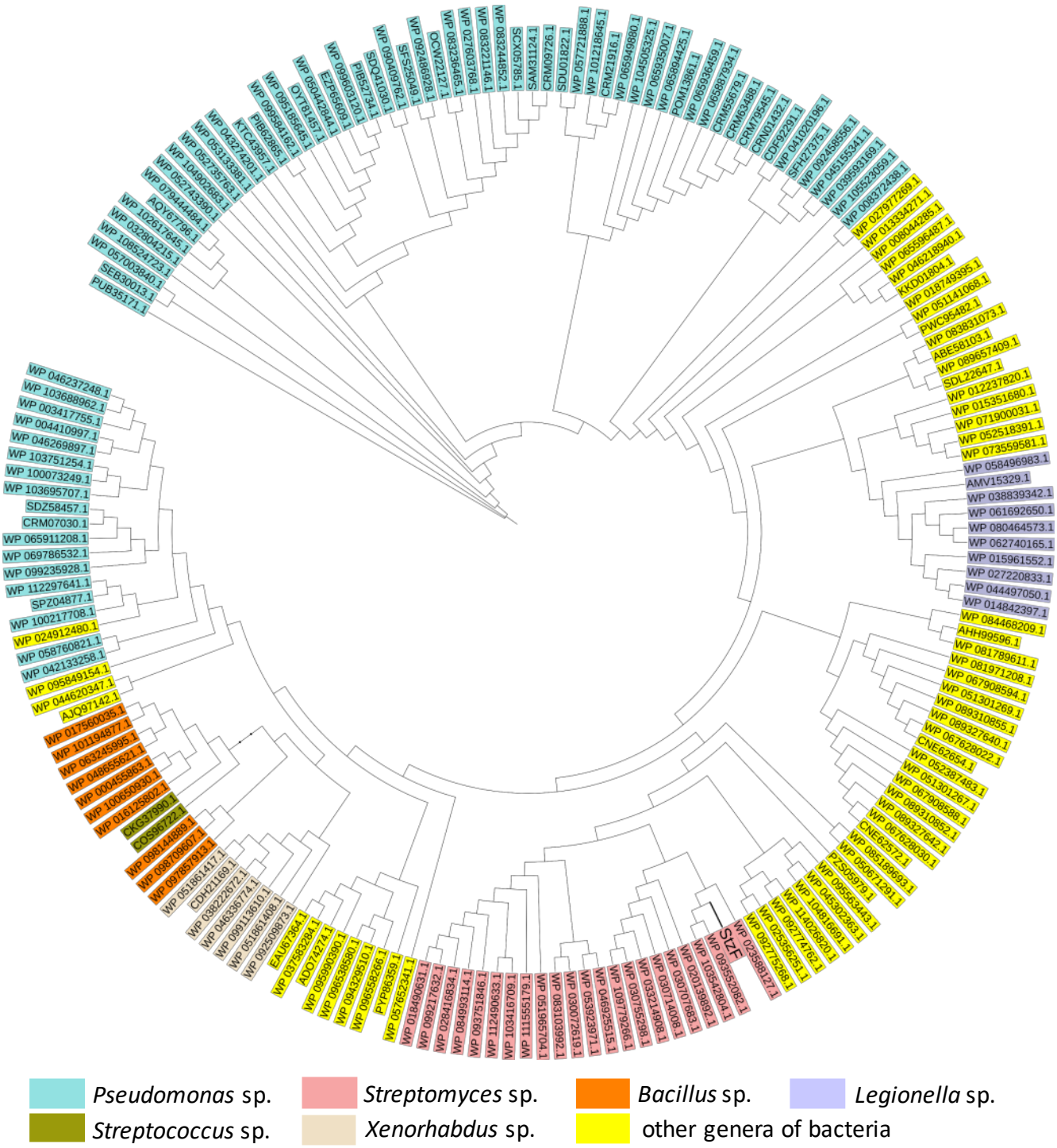


Figure S34. Cladogram analysis of StzF and homologous proteins from bacteria. Search of homologous protein was performed on Blast P[7] and sequences sharing >39% identities with StzF were selected to download for analysis. The neighbour-joining tree was built using online tool Clustal Omega[23] and represented with iTOL.[26]

References

1. Shirling, E. B.; Gottlieb, D., Methods for characterization of *Streptomyces* species. *Int. J. Syst. Evol. Microbiol.* **1966**, *16*, 313-340.
2. Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A., *Practical Streptomyces Genetics* John Innes Foundation: Norwich, UK, 2000.
3. Zerbino, D. R.; Birney, E., Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821-829.
4. Delcher, A. L.; Harmon, D.; Kasif, S.; White, O.; Salzberg, S. L., Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **1999**, *27*, 4636-4641.
5. Besemer, J.; Lomsadze, A.; Borodovsky, M., GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607-2618.
6. Guo, F.-B.; Ou, H.-Y.; Zhang, C.-T., ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.* **2003**, *31*, 1780-1789.
7. McGinnis, S.; Madden, T. L., BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **2004**, *32*, W20-W25.
8. Heinzelmann, E.; Kienzlen, G.; Kaspar, S.; Recktenwald, J.; Wohlleben, W.; Schwartz, D., The phosphinomethylmalate isomerase gene *pmi*, encoding an aconitase-like enzyme, is involved in the synthesis of phosphinothricin tripeptide in *Streptomyces viridochromogenes*. *Appl. Environ. Microbiol.* **2001**, *67*, 3603-3609.
9. Gust, B.; Challis, 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 geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1541-1546.
10. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison III, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343-345.
11. Du, Y.-L.; Shen, X.-L.; Yu, P.; Bai, L.-Q.; Li, Y.-Q., Gamma-butyrolactone regulatory system of *Streptomyces chattanoogensis* links nutrient utilization, metabolism and development. *Appl. Environ. Microbiol.* **2011**, *77*, 8415-8426.
12. Singaram, S.; Lawrence, R. S.; Hornemann, U., Studies on the biosynthesis of the antibiotic streptozotocin (streptozocin) by *Streptomyces achromogenes* var. *streptozoticus*. *J Antibiot* **1979**, *32*, 379-385.
13. Hiratsuka, T.; Koketsu, K.; Minami, A.; Kaneko, S.; Yamazaki, C.; Watanabe, K.; Oguri, H.; Oikawa, H., Core assembly mechanism of quinocarcin/SF-1739: bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chem. Biol.* **2013**, *20*, 1523-1535.
14. Green, L. C.; Wagner, D. A.; Glogowski, J.; Skipper, P. L.; Wishnok, J. S.; Tannenbaum, S. R., Analysis of nitrate, nitrite, and [^{15}N] nitrate in biological fluids. *Anal. Biochem.* **1982**, *126*, 131-138.
15. Kojima, H.; Sakurai, K.; Kikuchi, K.; Kawahara, S.; Kirino, Y.; Nagoshi, H.; Hirata, Y.; Nagano, T., Development of a fluorescent indicator for nitric oxide based on the fluorescein chromophore. *Chem. Pharm. Bull.* **1998**, *46*, 373-375.
16. Digenis, G. A.; Issidorides, C. H., Some biochemical aspects of *N*-nitroso compounds. *Bioorg. Chem.* **1979**, *8*, 97-137.
17. Andra, S. S.; Makris, K. C., Tobacco-specific nitrosamines in water: an unexplored environmental health risk. *Environ. Int.* **2011**, *37*, 412-417.
18. Xue, J.; Yang, S.; Seng, S., Mechanisms of cancer induction by tobacco-specific NNK and NNN. *Cancers* **2014**, *6*, 1138-1156.
19. Imoto, M.; Kakeya, H.; Sawa, T.; Hayashi, C.; Hamada, M.; Takeuchi, T.; Umezawa, K., Dephostatin, a novel protein tyrosine phosphatase inhibitor produced by *Streptomyces*. *J. Antibiot.* **1993**, *46*, 1342-1346.
20. Brooks, J. B.; Cherry, W. B.; Thacker, L.; Alley, C. C., Analysis by gas chromatography of amines and nitrosamines produced *in vivo* and i*n vitro* by *Proteus mirabilis*. *J. Infect. Dis.* **1972**, *126*, 143-153.
21. Herrmann, H., Identification of a metabolite of *Clitocybe suaveolens* as 4-methylnitrosaminobenzaldehyde. *Hoppe Seylers Z Physiol. Chem.* **1961**, *326*, 13-16.
22. Schwarzenbacher, R.; Stenner-Liewen, F.; Liewen, H.; Robinson, H.; Yuan, H.; Bossy-Wetzel, E.; Reed, J. C.; Liddington, R. C., Structure of the *Chlamydia* protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **2004**, *279*, 29320-29324.

23. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D.G, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
24. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320-W324.
25. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792-1797.
26. Letunic, I.; Bork, P., Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242-W245.
27. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019-1037.
28. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498-2504.
29. Wiese, R.; Eiglsperger, M.; Kaufmann, M., *yFiles*—Visualization and automatic layout of graphs. In *Graph Drawing Software*; Jünger, M. Mutzel, P. Eds.; Springer: 2004; pp. 173-191.
30. Medema, M. H.; Takano, E.; Breitling, R., Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30*, 1218-1223.