In vitro Biosynthesis of the Nonproteinogenic Amino Acid Methoxyvinylglycine

Jon B. Patteson,[#] Zachary D. Dunn,[#] and Bo Li*

TABLE OF CONTENTS

Experimental procedures:

Supplementary Tables:

Supplementary Figures:

Experimental procedures:

Chemicals were purchased from commercial suppliers and used without further purification. L-glutamic acid, α-ketoglutaric acid (α-KG), (+)-sodium L-ascorbate, ethanol, β-nicotinamide adenine dinucleotide (reduced, NADH), ammonium chloride, β-mercaptoethanol, coenzyme A, adenosine triphosphate (ATP), sodium tetraborate, cysteamine hydrochloride, L-glutamic acid dehydrogenase (from bovine liver), and alcohol dehydrogenase (from *S. cerevisiae*) were purchased from Sigma-Aldrich. $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, HEPES, KH_2PO_4, $MgSO_4$, NaOH, Tris-HCl, acetonitrile, sodium chloride, glycerol, imidazole, and Luria broth media were purchased from Fisher Scientific. Isopropylthio-D-galactoside (IPTG), kanamycin, chloramphenicol, and ampicillin were purchased from Goldbio. L-alanine-L-glutamic acid was purchased from Chem-Impex International. 2,4,4-D_3-L-glutamic acid, 20% DCl in D_2O, D_2O, and DMSO-d_6 were purchased from Cambridge Isotope Laboratories. All restriction enzymes and *S*-adenosyl-methionine (SAM) were purchased from New England Biosciences. The stereochemistry of all amino acids used in this study are L, unless otherwise noted.

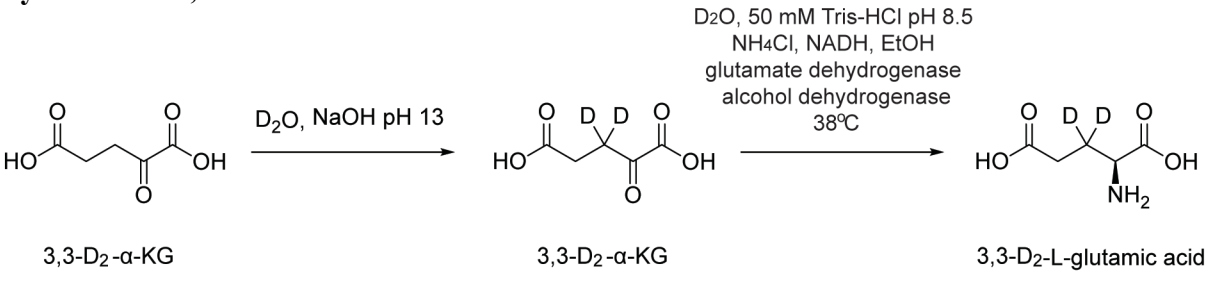
NMR spectra were recorded either on a Bruker 400 MHz or 600 MHz NMR spectrometer. NMR experiments are reported in δ units, parts per million (ppm), and were referenced to DMSO (δ 2.50 ppm) or D_2O (δ 4.79 ppm) as internal standards.

Synthesis of 4,4-D_2-L-Glu

Adapted from Ogrel *et al.*[1] A round bottom flask was oven dried and cooled in a desiccator. To this flask was added L-glutamic acid (16.8 mg). This flask was then evacuated by vacuum and flushed with N_2 three times. To this flask was added 20% DCl in D_2O (1 mL), and the reaction was stirred at 100°C for 10 days. Upon completion, the reaction was cooled to room temperature, neutralized to pH 7 using NaOH, and dried under reduced pressure. It is estimated that 95% of all Glu are 4,4-D_2-L-Glu based on peak integration. The resulting 4,4-D_2-L-Glu contains NaCl from the neutralization step. The mixture of 4,4-D_2-L-Glu and NaCl are used directly in enzymatic assays, because both 4,4-D_2-L-Glu and NaCl are highly polar and difficult to separate,.

^{1}H NMR (600 MHz, D_2O-d_6) δ: 2.37 (dd, J = 14.3, 6.4 Hz, 1H), 2.42 (dd, J = 14.7, 6.3 Hz, 1H), 4.13 (t, J = 6.4 Hz, 1H). HRMS (ESI) m/z $[M + H]^+$ calculated for $C_5H_8D_2NO_4$ 150.0730, observed 150.0749.

Synthesis of 3,3-D_2-L-Glu



Step 1: synthesis of 3,3-D_2-α-ketoglutarate

Adapted from Ogrel *et al.*[1] α-KG (24 mg) was added to a round bottom flask. The flask was evacuated by vacuum and flushed with nitrogen three times. A solution of 1 mL D_2O (pH 13 adjusted with NaOH) was added. The reaction was stirred at room temperature for 4 hours. Upon completion, the reaction was neutralized to pH 7 using HCl. The reaction was dried under reduced pressure. 92% of all α-KG are 3,3-D_2-α-KG by peak integration. The crude product was used directly in Step 2.

^{1}H NMR: ^{1}H NMR (400 MHz, D_2O) δ: 2.42 (s, 2H).

Step 2: enzymatic synthesis of 3,3-D_2-L-Glu from 3,3-D_2-α-ketoglutarate

Adapted from Ogrel *et al.*[1] To a microcentrifuge tube was added 0.5 mL D_2O, Tris-HCl pH 8.5 to a final concentration of 50 mM, 4 mg of crude 3,3-D_2-α-KG from Step 1, 5.6 mg NH_4Cl, 1 mg NADH, 8 μL ethanol, 4 U glutamate dehydrogenase, and 15 U alcohol dehydrogenase. Reaction was mixed slowly at 38°C for 2 hours. Proteins were precipitated with acetonitrile followed by centrifugation at 21,000 rcf. Crude reaction mixture was confirmed to contain 3,3-D_2-L-Glu by LC-HRMS and frozen at −20 °C for use directly in enzymatic reactions. It is estimated that 92% of all Glu are 3,3-D_2-L-Glu based on mass spectrometry analysis. $[M + H]^+$ calculated for $C_5H_8D_2NO_4$, *m/z* =150.0730, observed 150.0746.

Cloning of *ambB, C, D,* and *E*

The genes, *ambB*, *ambC, ambD,* and *ambE* were cloned using *P. aeruginosa* PAO1 genomic DNA as the template. Each gene was amplified using Phusion DNA polymerase (New England Biolabs) using primers listed in Table S1. The resulting PCR products were inserted into the pLIC-His vector using the ligation independent method described previously.[2] Each construct was confirmed by DNA sequencing.

AmbE-T_1 domain cloning

The boundaries for AmbE-T_1 domains was identified using Conserved Domain Database.[3] The DNA sequences for the T_1 domain was amplified using pLIC-His-AmbE as the template and using primers listed in Table S1. The amplified product and pET-30a vector were digested using KpnI and HindIII and ligated using T4 DNA ligase. *E. coli* DH5α cells were transformed with the resulting construct, yielding pET-30a-T_1. The construct was confirmed by DNA sequencing.

Site directed mutagenesis of AmbE

Overlap extension mutagenesis PCR was utilized for each desired point mutation, as described previously.[4] For AmbE-S1958A (thioesterase mutant), two PCR reactions were performed using primers JP00031 and JP00035, and JP00030 and JP00033. Both amplified products were used for overlap extension PCR. The PCR products were combined in equimolar amounts and amplified using primers JP00033 and JP00035. The PCR product and the plasmid pLIC-His-*ambE* were digested using NcoI and HindIII and ligated using T4 DNA ligase. *E. coli* DH5a cells were transformed with the resulting construct. The construct containing the desired mutation was verified by DNA sequencing.

For AmbE-S1819A (T_2 mutant), two PCR reactions were performed using primers JP00049 and JP00033, and JP00050 and JP00035. Both amplified products were used for

overlap extension PCR. The PCR products were combined in equimolar amounts and amplified using primers JP00033 and JP00035. The amplified product and the plasmid pLIC-His-*ambE* were digested using NcoI and HindIII and ligated using T4 DNA ligase. *E. coli* DH5a cells were transformed with the resulting construct. The construct containing the desired mutation was verified by DNA sequencing.

Protein expression and purification

For protein purification, *E. coli* BAP-1 were transformed with pLIC-His-*ambB*. *E. coli* BL-21 (DE3) were individually transformed with pLIC-His-*ambC,* pLIC-His-*ambE,* pET-30a-T_1, pLIC-His-*ambE*-S1958A, and pLIC-His-*ambE*-S1819A. *E. coli* BL21-CodonPlus (DE3) RIPL cells were transformed with pLIC-His-*ambD*. All transformations were performed by electroporation. *E. coli* were routinely cultured at 37 ºC.

Bacterial seed cultures of AmbB were started from a single colony and grown in LB medium supplemented with 50 μg/mL ampicillin. A sample of 2 mL seed culture was transferred to a 1 L LB medium supplemented with 50 μg/mL ampicillin grown at 37 °C until the cell density reached an OD_{600} of 0.4–0.6, when protein expression was induced with 0.1 mM IPTG. After growing at 16ºC for 16 hours, bacterial cells were harvested at 6,000 rcf and lysed by sonication. The resulting cell debris was removed by centrifugation at 21,000 rcf. The cells were resuspended for lysis in Buffer A (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol (BME), and 30 mM imidazole) and sonicated using a Fisher Scientific Sonic dismembrator model 500 (1.5 min total on time of cycles of 30% amplitude with 0.5 sec pulse on, and 1.5 sec pulse off). The lysed cells were centrifuged at 21,000 rcf for 30 min at 4 °C. The supernatant was filtered through a 0.45 μm filter before purification by FPLC. The filtered supernatant was purified by Ni affinity chromatography using a GE HisTrap HP Ni column. Buffer A was used as wash buffer, and Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM BME, and 300 mM imidazole) was used as elution buffer. Fractions containing AmbB were combined and concentrated to a final volume of 3.5 mL using a Millipore centrifugal filter with a 30 kDa molecular weight cut off. The combined fractions were then applied to a GE HiLoad 16/600 Superdex 200 pg column that had been equilibrated with Buffer C (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM BME, and 10% glycerol). After purification, fractions containing pure AmbB were pooled and again concentrated using a Millipore centrifugal filter with a 30 kDa molecular weight cut off to a final concentration of 50–200 μM. AmbB was then flash frozen in liquid nitrogen and stored at −80°C.

AmbE, T_1, AmbE-S1958A, and AmbE-S1819A were expressed and purified using the same methods as described above for AmbB. AmbC and AmbD were expressed and purified only by Ni affinity chromatography. Following Ni column purification, AmbC and AmbD were directly desalted using a GE PD-10 column equilibrated in Buffer C, concentrated using a Millipore centrifugal filter with a 10 kDa molecular weight cut off, and frozen in liquid nitrogen and stored at −80ºC.

***in vitro* reconstitution of Ala-AMB biosynthesis (one-pot assay)**

A sample of 7.5 μM AmbB and 7.5 μM AmbE were incubated with 0.5 μM Sfp (the promiscuous phosphopantetheinyl transferase), 100 μM coenzyme A, 8 mM $MgCl_2$, and 50 mM potassium phosphate (pH 8.0) at 28 °C for one hour. To this mixture were added 1 mM L-alanine, 1 mM L-glutamic acid, 1 mM α-KG, 1 mM sodium ascorbate, 1 mM SAM, 4 mM ATP, and 100 μM ammonium iron(II) sulfate hexahydrate. The reaction was initiated by the addition

of 6.5 μM AmbC and 8.5 μM AmbD, which brought the final volume to 100 μL. The assay was incubated at room temperature for 2 hours. To quench the reaction, 100 μL of acetonitrile was added and the assay was cooled to −20 °C for at least 20 min. The assays were then centrifuged at 14,000 rcf for 5 min to remove the protein precipitates. A sample of 10 μL of supernatant was used for analysis by liquid chromatography (LC) coupled electrospray ionization (ESI) high resolution mass spectrometry (HRMS). Samples were analyzed using an Agilent Technologies 6520 Accurate Mass Q-TOF LC-HRMS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 300 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 175 V, and skimmer 65 V. Samples were injected into a Kinetex C18 column (Phenomenex, 150 mm length, 2.6 μm particle size and 100 Å pore size) and separated using the follow method at a flow rate of 0.4 mL/min. Solvent A consisted of 0.1% formic acid in water (Fisher), and solvent B consisted of 0.1% formic acid in acetonitrile (Fisher). Mobile phase is held at 2% B for 10 min, increased from 2% to 20% B over 5 min in a linear gradient, increased from 20% to 95% B over 10 min in a second linear gradient, and then held at 95% B for 4 min before returning to 2% B over 1 min. Ala-AMB (**2**) was detected by HRMS ($[M + H]^+$ *m/z* calculated for $C_8H_{15}N_2O_4$ 203.1026, observed 203.1031).

***In vitro* biosynthesis and derivatization of Ala-AMB and isolation of Fmoc-Ala-AMB**

A 2 mg one-pot reaction was conducted for Ala-AMB for structural analysis by NMR. Ala-AMB was derivatized by Fmoc-HCl for separation from polar cofactors and substrates in the reaction. Fmoc derivatized product Fmoc-Ala-AMB and the remaining derivatized reactant Fmoc-Ala co-elute on reverse phase chromatography, thus reaction conditions were optimized to maximize the conversion of Ala to Ala-AMB (Ala concentration was reduced from 1 mM to 0.25 mM). A sample of 7.5 μM AmbB and a sample of 7.5 μM AmbE were incubated with 0.5 μM Sfp (phosphopantetheinyl transferase), 100 μM coenzyme A, 8 mM $MgCl_2$, and 50 mM potassium phosphate (pH 8.0) at 28 °C for 20 minutes. To this mixture were added 0.25 mM L-alanine, 1 mM L-glutamic acid, 1 mM α-KG, 1 mM sodium ascorbate, 1 mM SAM, 4 mM ATP, and 100 μM ammonium iron(II) sulfate hexahydrate. The reaction was initiated by the addition of 6.5 μM AmbC and 8.5 μM AmbD. The final volume was brought to 30 mL. After incubating the reaction at room temperature for 2 hour, the reaction was quenched with 30 mL of acetonitrile and centrifuged at 4000 rcf to remove protein precipitates. After protein removal, small molecules in the 60 mL of supernatant were derivatized by Fmoc-Cl by adding 15 mL of 400 mM (pH 10.4) sodium borate and 15 mL of 40 mM Fmoc-Cl dissolved in acetonitrile. The reaction was incubated at room temperature for 5 minutes before it was centrifuged for 20 minutes at 4000 rcf. The supernatant from this reaction was dried under reduced pressure. To prepare the sample for preparative HPLC purification, the crude reaction was resuspended in 10 mL of water and 10 mL DMSO. This mixture was centrifuged at 14,000 rcf for 10 minutes and the supernatant was separated by reverse-phase preparative HPLC. Column purification conditions are as follows: samples were injected on Phenomenex Luna C18 (10 μm, 250 x 21.20 mm) and separated using a gradient of 5–35% mobile phase B over 10 minutes, and 35-55% over 18 minutes (mobile phase A: water + 0.1% TFA; mobile phase B: acetonitrile + 0.1% TFA). Fmoc-Ala-AMB co-elute with some Fmoc impurities, and the resulting product was characterized by ^{1}H and (^{1}H,^{1}H) COSY NMR experiments (Table S2, Figures S5&6).

Fmoc-Ala-AMB (Fmoc-**2**) was detected by MS ($[M + H]^+$ $C_{23}H_{25}N_2O_6$, *m/z* calculated 425.1707, observed 425.1704).

AmbE-T_1 amino acid loading assay

To install the phosphopantetheine (Ppant) arm on the T domains in trans, 150 μM AmbE-T_1 was incubated with 0.5 μM Sfp, 1 mM coenzyme A, 4 mM $MgCl_2$, and 25 mM potassium phosphate (pH 7.5) at 28 °C for one hour. A 10 μL aliquot of this reaction mixture was added to 25 mM potassium phosphate buffer (pH 7.5), 2 mM ATP, 1 mM L-glutamic acid, and 2 μM AmbE-S1958A to a final volume of 50 μL. This reaction mixture was incubated at 28 °C for 2 hours. A 10 μL sample of the reaction was used for LC-HRMS analysis. Samples were analyzed using an Agilent Technologies 6520 Accurate Mass Q-TOF LC-HRMS/MS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 350 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 250 V, skimmer 65 V. Masses targeted for MS/MS were selected automatically based on abundance and charge state over 3, and masses were fragmented with a collision energy of 35 V. The protein samples were run on a Restek Viva C_4 column (150 mm length, 5 μm particle size). Solvents used for AmbE-T domain samples are the same as for small molecule samples, and samples were run at a flow rate of 0.2 mL/min. The gradient was 5% for 2 min, followed by a linear gradient to 35% over 6 min, and then to 60% B over 27 min. Then the gradient was increased to 95% over 5 min, held at 95% for 5 min before returning to 5% over 2 min. LC-HRMS data of intact protein were analyzed using Agilent MassHunter Qualitative Analysis B.03.01 protein deconvolution by maximum entropy method.

AmbC and AmbD modification of AmbE-T_1-bound Glu

To examine the activity of tailoring enzymes AmbC and AmbD, samples containing 50 μM AmbC and AmbD were separately incubated with 100 μM ammonium iron(II) sulfate hexahydrate in 25 mM potassium phosphate buffer (pH 7.5) for 10 min to reconstitute the iron center prior to activity assays. For each reaction, 5 μL of reconstituted AmbC or AmbD incubated with Fe(II) and 1 mM α-KG and 1 mM ascorbic acid were added to Ambe-T_1 loading assays (described above) to a final volume of 50 μL. The assay was incubated at 28 °C for 2 hours. A 10 μL sample of the reaction was used for LC-HRMS analysis. Samples were analyzed using an Agilent Technologies 6520 Accurate Mass Q-TOF LC-HRMS/MS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 350 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 250 V, skimmer 65 V. Masses targeted for MS/MS were selected automatically based on abundance and charge state over 3, and masses were fragmented with a collision energy of 35 V. The protein samples were run on a Restek Viva C_4 column (150 mm length, 5 μm particle size). Solvents used for AmbE-T domain samples are the same as for small molecule samples, and samples were run at a flow rate of 0.2 mL/min. The gradient was 5% for 2 min, followed by a linear gradient to 35% over 6 min, and then to 60% B over 27 min. Then the gradient was increased to 95% over 5 min, held at 95% for 5 min before returning to 5% over 2 min. LC-HRMS data of intact protein were analyzed using Agilent MassHunter Qualitative Analysis B.03.01 protein deconvolution by maximum entropy method.

Assays using deuterium labeled Glu as substrates

For isotopic labeling studies, 2,4,4-D_3-L-Glu was purchased from Cambridge Isotopes Laboratory. 3,3-D_2-L-Glu and 4,4-D_2-L-Glu were synthesized as described above. For *in vitro* reactions, the deteurium labeled Glu were used in place of unlabeled Glu at the same concentrations as described in each assay. Samples were quenched and analyzed as described in the section "*in vitro* reconstitution of Ala-AMB biosynthesis".

Fmoc derivatization and tandem MS analysis of *in vitro* synthesized Ala-AMB

The *in vitro* Ala-AMB reconstitution assay (one-pot) was quenched with acetonitrile, centrifuged, and the supernatant was separated from precipitated protein as described above. A 200 µL sample of the quenched supernatant was added to a 100 µL sample of 400 mM (pH 10.4) sodium borate. To this mixture, a 100 µL sample of 40 mM Fmoc-Cl dissolved in acetonitrile was added to the mixture. The reaction was incubated at room temperature for 5 minutes before it was centrifuged for 5 minutes at 13,000 rcf. A 10 µL sample was analyzed using an Agilent Technologies 6520 Accurate Mass Q-TOF LC-HRMS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 300 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 175 V, and skimmer 65 V. Samples were injected into a Kinetex C18 column (Phenomenex, 150 mm length, 2.6 µm particle size and 100 Å pore size) and separated using the following method at a flow rate of 0.4 mL/min. Solvent A consisted of 0.1% formic acid in water (Fisher), and solvent B consisted of 0.1% formic acid in acetonitrile (Fisher). Mobile phase is held at 2% B for 10 min, increased from 2% to 20% B over 5 min in a linear gradient, increased from 20% to 95% B over 10 min in a second linear gradient, and then held at 95% B for 4 min before returning to 2% B over 1 min. For MS/MS analysis, 15 V fragmentation energy was used, and m/z 425.1707 was selected for fragmentation. Fmoc-Ala-AMB was detected by HRMS ($[M + H]^+$ $C_{23}H_{25}N_2O_6$, m/z calcd 425.1707, observed 425.1713).

Chemical capture of NRPS-bound intermediates using cysteamine

The one-pot *in vitro* AMB reconstitution reaction was conducted using either AmbE-S1819A or AmbE-S1958A in place of wild-type AmbE to accumulate biosynthetic intermediates bound to AmbE. A sample of freshly prepared cysteamine hydrochloride was added to the one-pot assay (final concentration 50 mM) as the final component immediately after adding all other substrates and cofactors. Individual enzymes (AmbB, AmbC, or AmbD) or substrates (Ala or SAM) were excluded from the reaction mixture in separate reactions. Isotopically labeled glutamic acids were used in place of L-glutamic acid for corresponding reactions. The modified assays containing cysteamine were incubated at room temperature for 2 hours, and 100 µL of reaction was quenched with 200 µL acetonitrile. This questions reaction was derivatized with Fmoc-Cl by mixing a 25 µL sample of quenched reaction with 50 µL of 400 mM (pH 10.4) sodium borate, 90 µL of 18 mM Fmoc-Cl dissolved in acetonitrile, and 90 µL acetonitrile. The reaction was incubated at room temperature for 5 minutes followed by centrifugation for 20 minutes at 4000 rcf to remove any particulates. Samples were analyzed using an Agilent Technologies 6520 Accurate Mass Q-TOF LC-HRMS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 300 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 175 V, and skimmer 65 V. Derivatized samples were injected into a Kinetex C18 column (Phenomenex, 150 mm length, 2.6 µm particle size and 100 Å pore size) and separated using the following method at a flow rate of 0.4 mL/min. Solvent A consisted of 0.1% formic acid in water (Fisher), and solvent B consisted of 0.1% formic acid in

acetonitrile (Fisher). Mobile phase is held at 2% B for 10 min, increased from 2% to 20% B over 5 min in a linear gradient, increased from 20% to 95% B over 10 min in a second linear gradient, and then held at 95% B for 4 min before returning to 2% B over 1 min.

Phylogenetic analysis of AmbE-C*

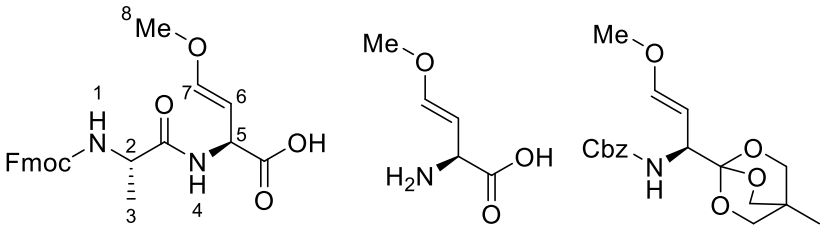
The AmbE-C* domain was aligned with all protein sequences from the MIBiG database,[5] and seven NRPS proteins were identified that contain C*-like domains homologous to AmbE-C* (AAG05690.1), including LgnD (AIZ66879.1), Zmn17 (CCM44337.1), DepE (ABP57749.1), HasO (CZT62784.1), NdaA (ATP76243.1), PuwF (AIW82283.1), and PuwG (AIW82284.1).[6,7,8,9,10,11] These C*-like domains, AmbE-C*, and the C*-like domains from McyA in microcystin biosynthesis and Bleom4 and Bleom7 from bleomycin biosynthesis were combined with the C domains from NaPDoS database,[12] and aligned with MUSCLE version 3.8.31.[13] A maximum likelihood tree containing bootstrap values was generated using RAxML version 8.2.4.[14] The tree was visualized using Interactive Tree Of Life (iTOL) version 3.5.4.[15]

Agar diffusion assay of *in vitro* synthesized Ala-AMB

Minimal media was prepared by dissolving 7 g potassium phosphate dibasic, 3 g potassium phosphate monobasic, 0.5 g sodium citrate, 0.1 g magnesium sulfate heptahydrate, and 1 g ammonium sulfate in 1 L of H_2O. After the media was autoclaved, 2 g of sterile glucose was added to the media. *E. coli* MG1655 was grown overnight at 37 °C in minimal media. The overnight culture was diluted 50 fold using minimal media and then a sample of 100 µL of diluted *E. coli* culture was applied to a minimal media plate containing 15 g/L agar. The plates were dried before paper discs were added. Ala-AMB was synthesized *in vitro* using the one-pot reconstitution reaction. Negative controls were performed without AmbE or lacking both Ala and Glu. Upon completion, either 5 or 10 µL of one-pot *in vitro* reconstitution reaction or control was added to a disc, and the plates were incubated for 16 hours at 37 °C before imaging.

Primer name	Sequence	Use
SJP1	TACTTCCAATCCAATGCGATGCAGGAGCGACATGGCTT	AmbB
SJP2	TTATCCACTTCCAATGCGCTATCAGGAAGCGTTGCAGCCCT	AmbB
SJP3	TACTTCCAATCCAATGCGATGAGTGCGTCAGAAGACCTGC	AmbC
SJP4	TTATCCACTTCCAATGCGCTATCAGGTTGCCAGGTTCGCCG	AmbC
SJP5	TACTTCCAATCCAATGCGATGGAACGACACGCTCCCAGTC	AmbD
SJP6	TTATCCACTTCCAATGCGCTAACGAGGCGCTCATGATGC GGCAC	AmbD
SJP8	TACTTCCAATCCAATGCGTCAACAATGAGCGCCTCGTTCAG CGC	AmbE
SJP9	TTATCCACTTCCAATGCGCTATCAGGGTTGGTGGACATGCA	AmbE
JP00013	GATACAGGTACCCAGACCCTGGAAGCGCCG	AmbE T_1
JP00014	GATACAAAGCTTAGAGCGCATCCGCCGGACT	AmbE T_1
JP00031	TAGGCGAGCATGCCGCCGAGG**GC**GGCGCCGATCAGGG CGAACG	AmbE-S1958A
JP00030	CGTTCGCCCTGATCGGCGCC**GC**CCTCGGCGGCATGCTC GCCTA	AmbE-S1958A
JP00049	CTTCGCCAGCGCCGGCGGGCAT**G**CGCTGCTGGGCGTGC AGGC	AmbE-S1819A
JP00050	GCCTGCACGCCCAGCAGCG**C**ATGCCCGCCGGCGCTGGC GAAG	AmbE-S1819A
JP00035	GATACACCATGGCTTCGAGGTGAGCATCCG	AmbE overlap extension PCR incorporation
JP00033	GATACAAAGCTTGTCGACGGAGCTCGAATTCGG	AmbE overlap extension PCR incorporation

Table S1. List of primers used in this study.



Fmoc-Ala-AMB (Fmoc-**2**) (3*E*)-AMB Protected (3*E*)-AMB

	δ_H (ppm) and mult. (*J* [Hz])			
Position	**Fmoc-2**	**Commercial Fmoc-L-Ala in DMSO-d_6**		**COSY**
1	7.52, d (7.9)	7.67, d (7.8)		2
2	4.04, m	4.01, t (7.5)		1, 3
3	1.22, m	1.29, d (7.4)		2
		Isolated (3*E*)-AMB in D_2O[16]	**Protected (3*E*)-AMB in $CDCl_3$**[17]	
4	8.23, d (6.9)			5
5	4.60, t (7.8)	4.67, d (10)	4.29, t (8.6)	4, 6
6	4.79, dd (12.6, 8.7)	5.45, dd (13, 10)	4.73, dd (12.7, 8.3)	5, 7
7	6.69, d (12.5)	7.33, d (13)	6.52, d (12.4)	6
8	3.50, s	4.11, s	3.53, s	

Table S2. NMR shifts of Fmoc-Ala-AMB (Fmoc-**2**) in DMSO-d_6 (600 MHz) isolated from *in vitro* reconstitution of Ala-AMB biosynthesis. Relevant NMR shifts are provided for comparison. Fmoc-Ala NMR shifts are obtained from Sigma-Aldrich website. The difference in ppm shifts for in vitro synthesized Ala-AMB and isolated or synthetic AMB is likely due to solvent effects.

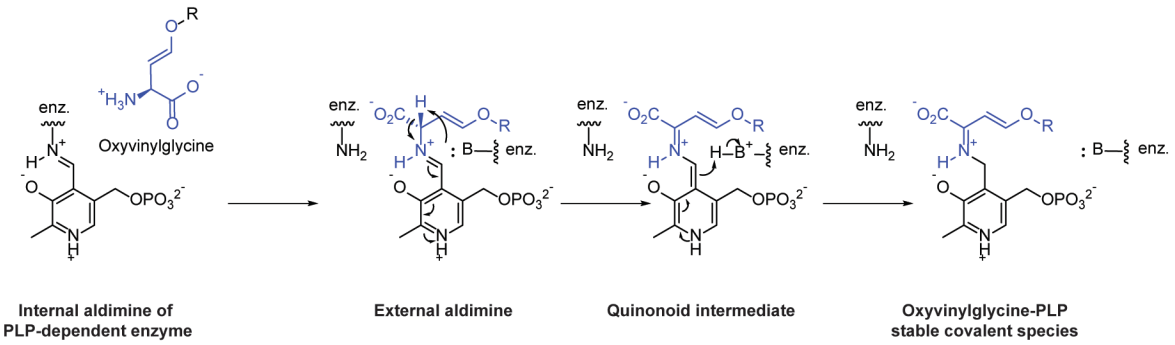


Figure S1. Proposed mechanism for oxyvinylglycine inhibition of the eliminase subgroup of PLP-dependent enzymes. Adapted from Berkowitz *et al.*[18]

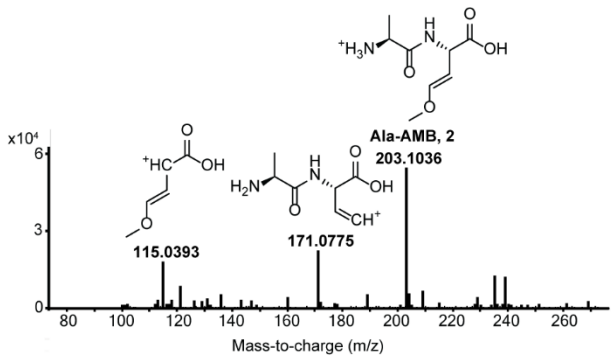


Figure S2. Tandem MS of Ala-AMB (**2**). Predicted structures are shown for major fragments.

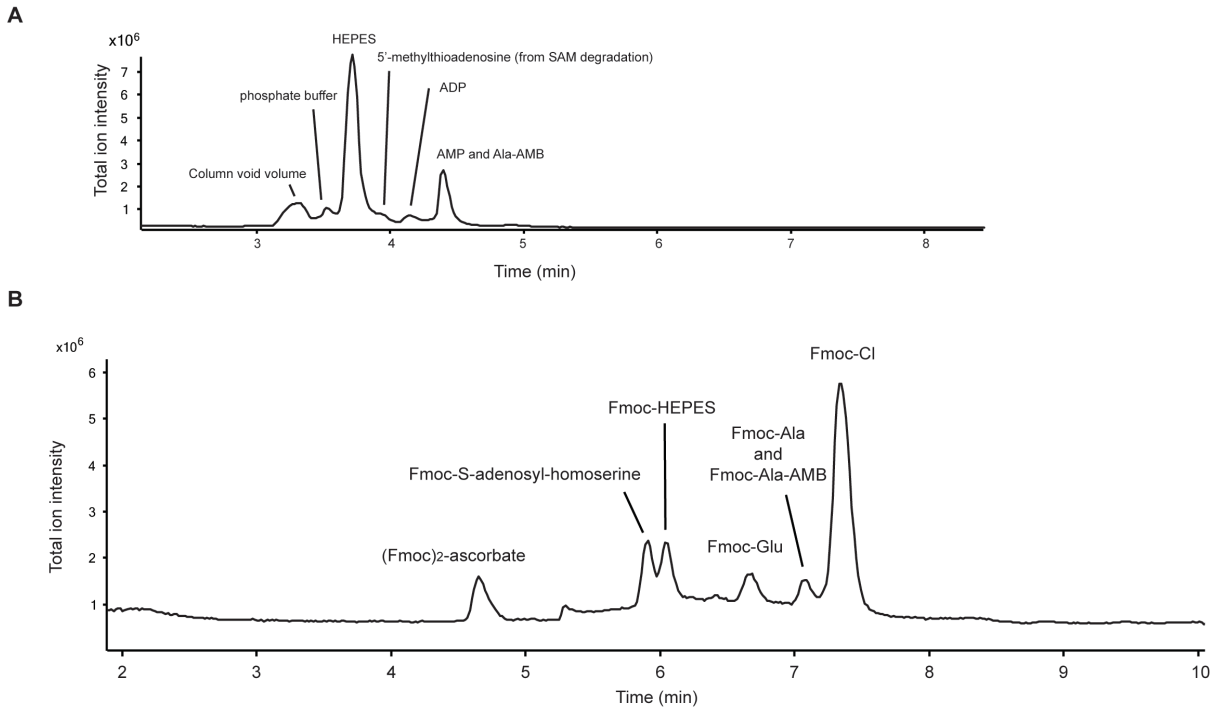


Figure S3. Total ion chromatograms (TIC) of one pot *in vitro* reconstitution reaction for Ala-AMB. (A) TIC for *in vitro* one-pot reaction without Fmoc-derivatization. (B) TIC for *in vitro* one-pot reaction after Fmoc-derivatization. The small peak in TIC at 5.4 minutes is due to infusion of reference masses. No Ala-AMB-Ala tripeptide or Fmoc-Ala-AMB-Ala was detected in either spectrum.

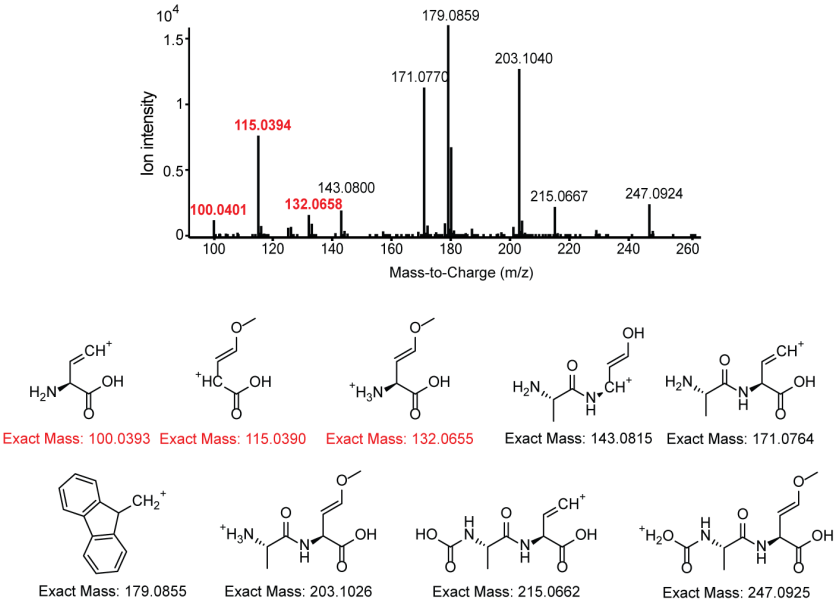


Figure S4. Full tandem MS of Fmoc-Ala-AMB (Fmoc-**2**) from Figure 2B. Predicted structures of all major fragments are shown. Mass signals colored red indicate that AMB is at the *C*-terminus of the Ala-AMB dipeptide.

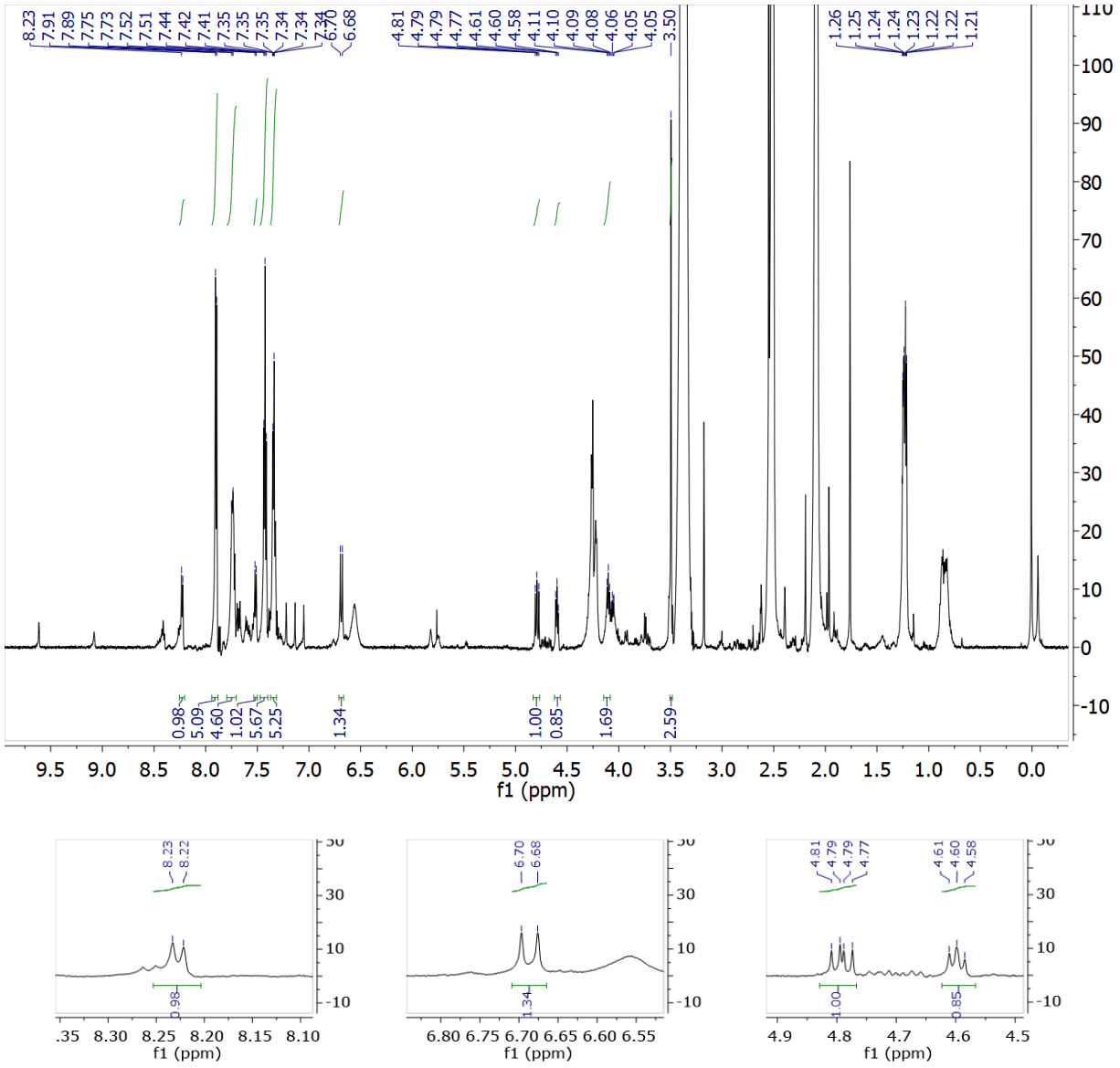


Figure S5. ^{1}H NMR of partially purified Fmoc-Ala-AMB generated from *in vitro* one-pot reaction in DMSO-d_6 (600 MHz). Key peaks corresponding to AMB protons are shown in more detail. The large integration of Fmoc peaks are due to Fmoc impurities that co-eluted with Fmoc-Ala-AMB.

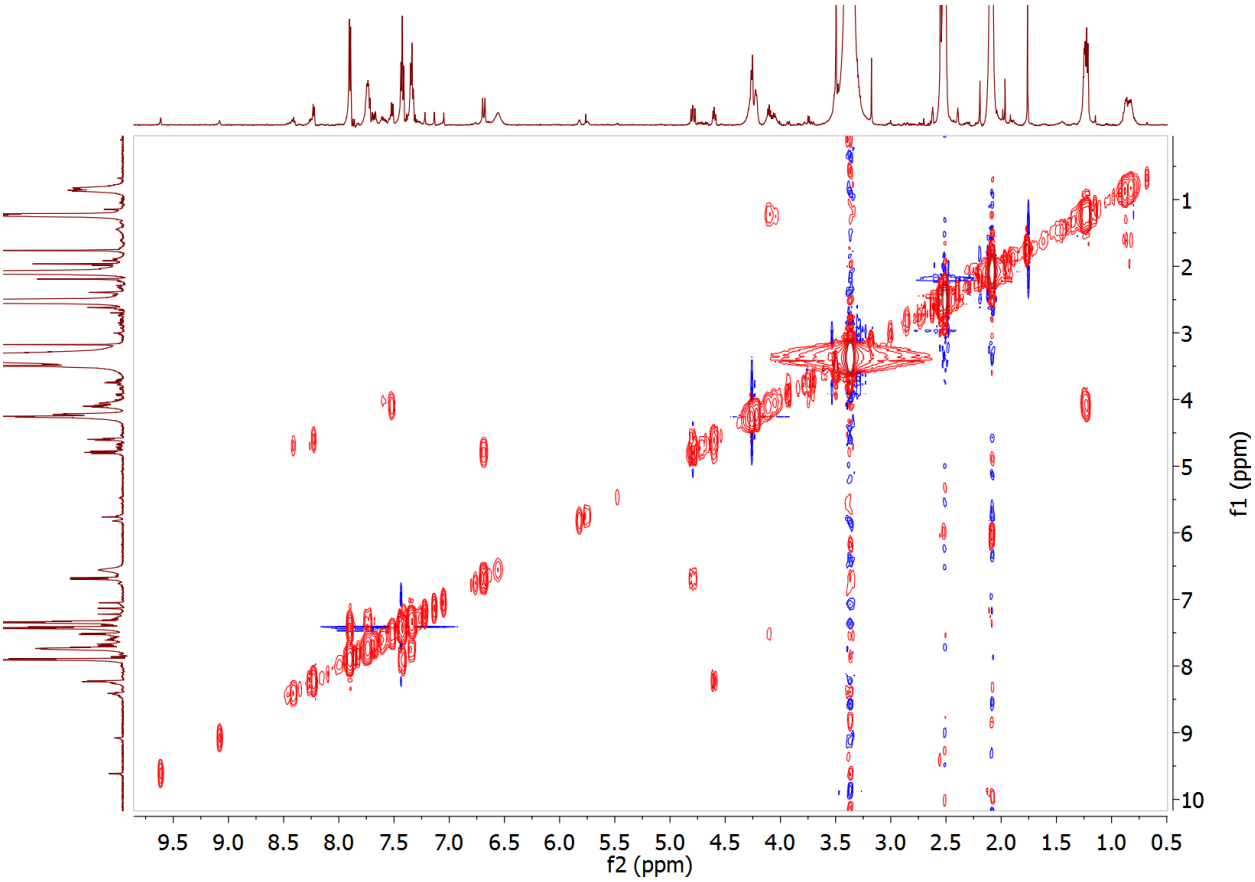


Figure S6. (^{1}H,^{1}H) COSY NMR of partially purified Fmoc-Ala-AMB generated from *in vitro* one-pot reaction in DMSO-d_6 (600 MHz).

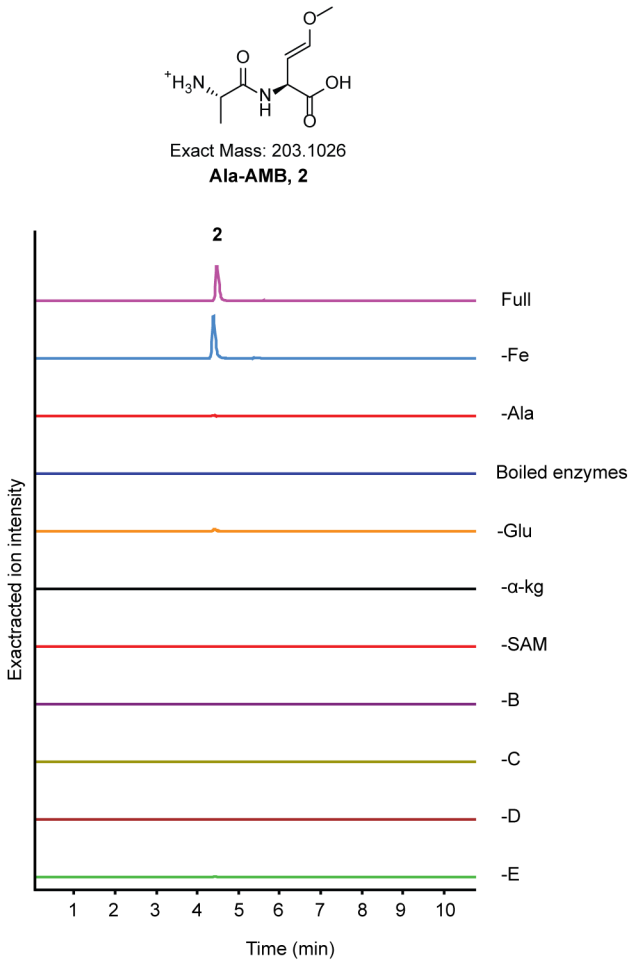


Figure S7. Negative controls for the *in vitro* reconstitution of Ala-AMB biosynthesis. Extracted ion chromatograms (EICs) of Ala-AMB (**2**, m/z 203.1026 $[M + H]^+$) are shown for reactions lacking the components indicated. B, C, D, E refers to enzyme AmbB, AmbC, AmbD, and AmbE, respectively. Each control was run at least three times, and representative chromatograms are shown.

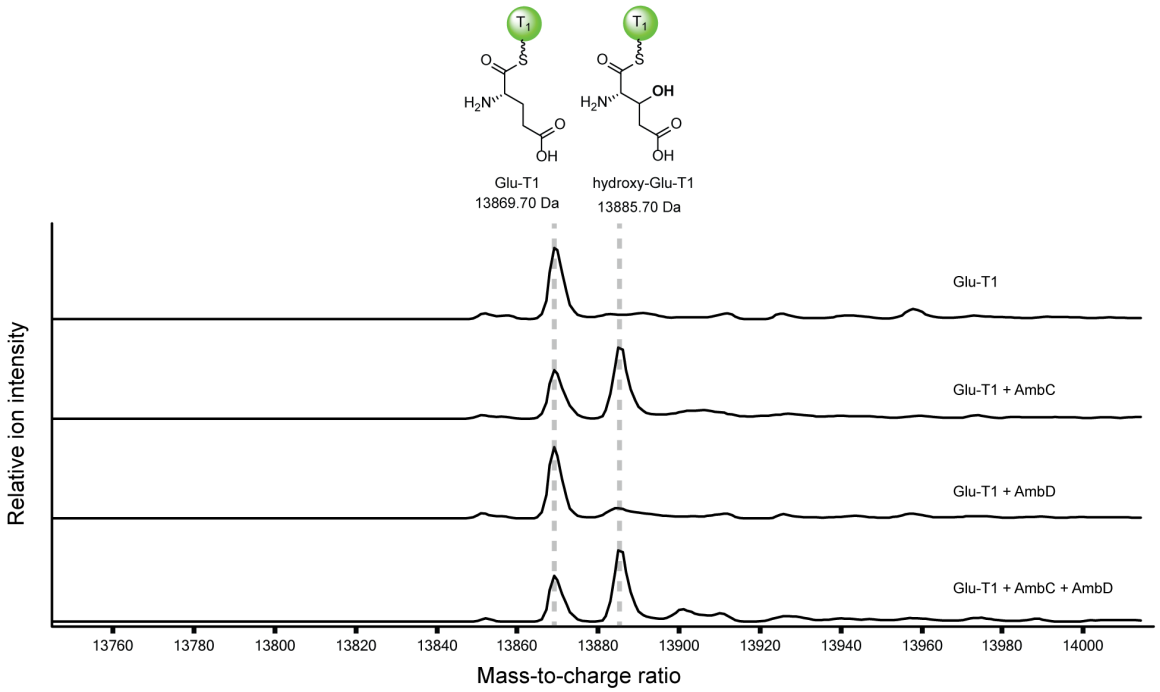


Figure S8. AmbD does not modify Glu-T_1 or hydroxyl-Glu-T_1. Deconvoluted protein mass spectra are shown. Glu-T_1 is oxygenated by AmbC. Neither Glu-T_1 nor the oxygenated product of AmbC is modified by AmbD.

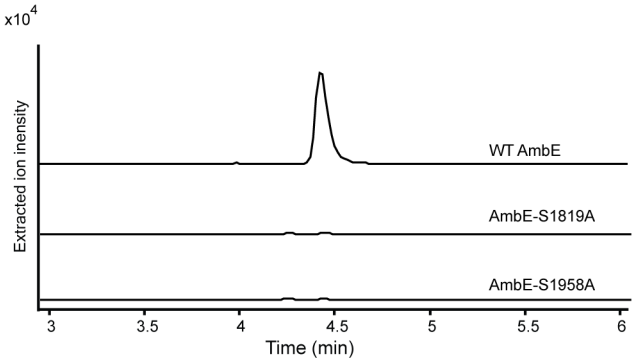


Figure S9. Biosynthesis of Ala-AMB requires the post-translationally modified T_2 domain, and the active TE domain of AmbE. One-pot *in vitro* reconstitution reactions were performed using AmbE wildtype (WT), AmbE S1819A (T_2 mutant), or AmbE S1958A (TE mutant) under the conditions described in experimental method "*in vitro* reconstitution of Ala-AMB biosynthesis (one-pot assay)" (page S6). AmbE S1819A and S1958A were used at the same concentration as AmbE WT (7.5 μM). EICs of **2** (*m/z* = 203.1026 $[M + H]^+$) are shown from each one-pot reaction.

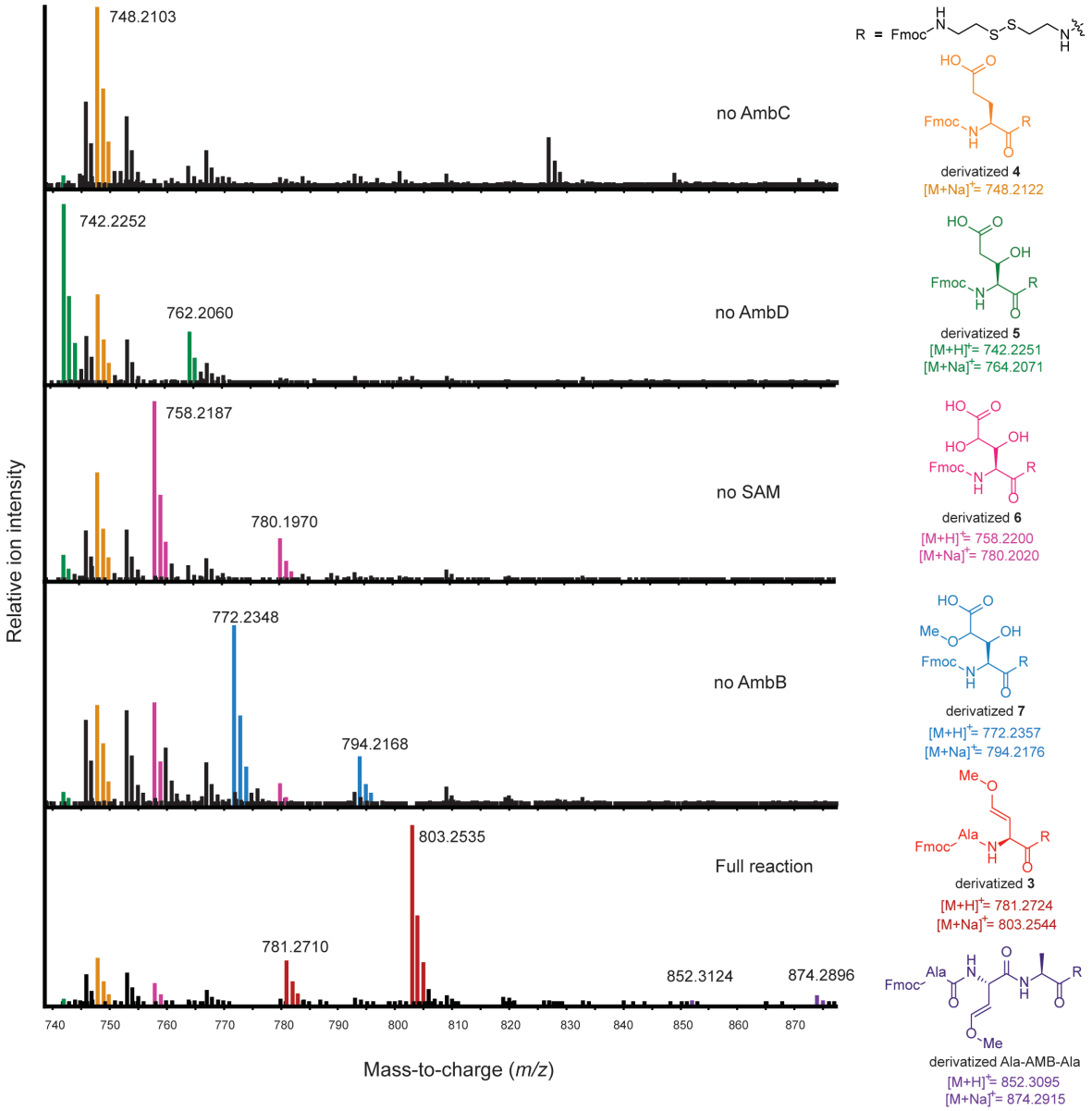


Figure S10. Mass spectra of chemical cleavage assay using AmbE S1958A (TE mutant) and L-Glu. Extended mass range for data from Figure 4 is shown. A small amount of derivatized Ala-AMB-Ala tripeptide was detected in the full assay (bottom panel, *m/z* =852.3095 for $[M+H]^+$ and 874.2915 for $[M+Na]^+$), suggesting formation of this tripeptide is a minor product under the assay conditions.

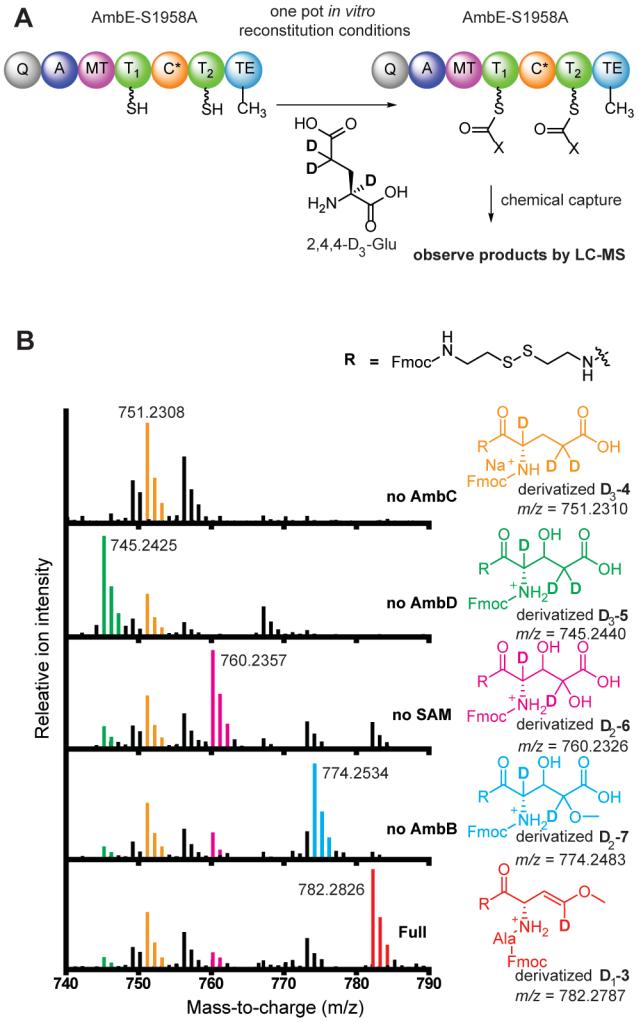


Figure S11. Chemical capture experiment using AmbE S1958A and 2,4,4-D_3-Glu. (A) Scheme illustrating chemical cleavage of intermediates in one-pot assay containing AmbE S1958A (TE mutant). (B) Mass spectra of intermediates captured by cysteamine in the one-pot assay excluding specific assay components. Structures of the captured intermediates are depicted in the same color as their mass spec signals.

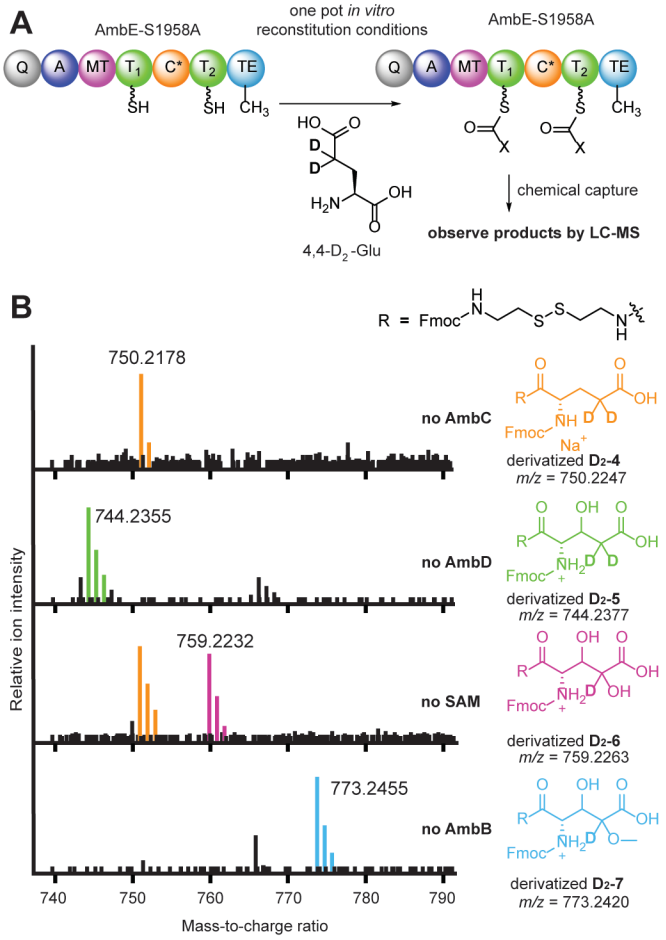


Figure S12. Chemical capture experiment using AmbE S1958A and 4,4-D_2-Glu. (A) Scheme illustrating chemical cleavage of intermediates in one-pot assay containing AmbE S1958A (TE mutant). (B) Mass spectra of intermediates captured by cysteamine in the one-pot assay excluding specific assay components. Structures of the captured intermediates are depicted in the same color as their mass spec signals. Low peak intensities are likely due to the additional salt (~50 mM NaCl) in 4,4-D_2-Glu substrate from the crude reaction, which may reduce enzyme activity (see supplemental methods).

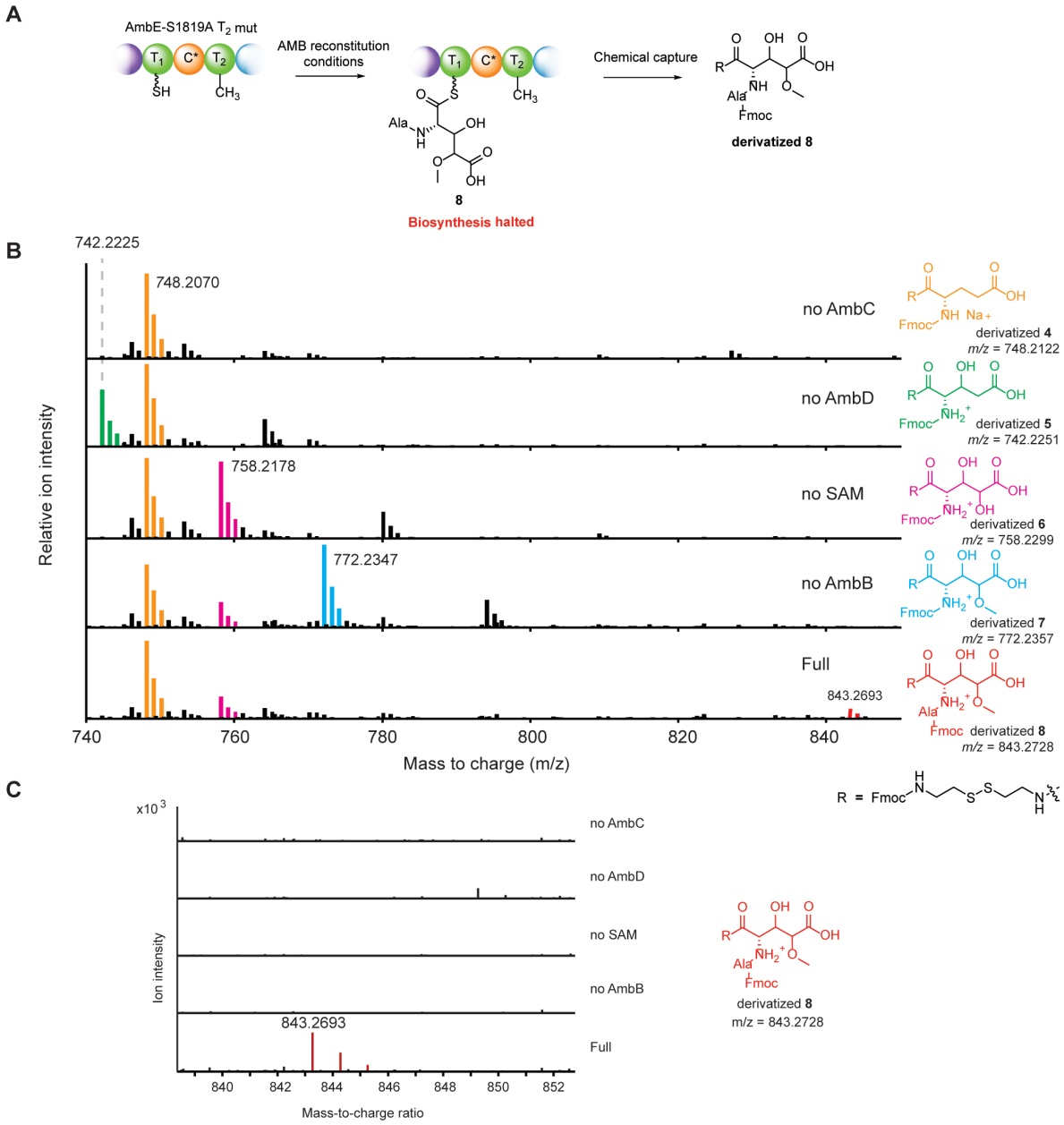


Figure S13. Chemical capture experiment with AmbE-S1819A (T_2 mutant) and Glu. (A) Mass spectra of intermediates captured in the full assay and assays excluding specific reaction components. Structures of the captured intermediates are depicted in the same color as their mass spec signals. (B) Expansion of mass spectra of derivatized **8**, which is only present in the full one-pot assay containing AmbE-S1819A.

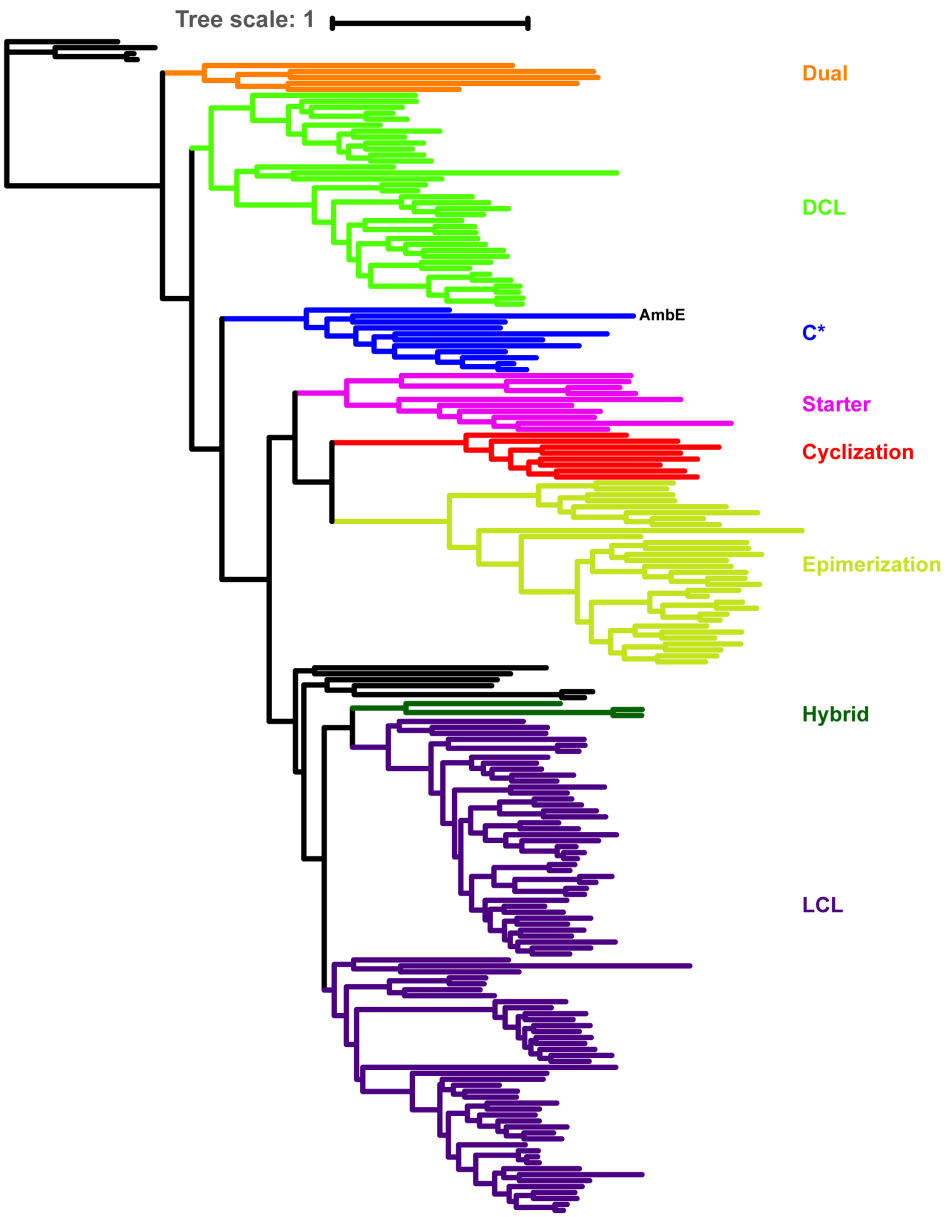


Figure S14. Bioinformatic analysis of AmbE-C* reveals that AmbE-C* clusters with a unique group of uncharacterized C domains. Proteins sequences include AmbE-C*, the C domains from the NaPDoS database, and seven C*-like domains from the MIBiG database.[5] From top to bottom in the C* group: DepE_C (ABP57749.1, FK228 biosynthesis), AmbE_C (AAG05690.1), Bleom7_C2 (AAG02359.1, bleomycin biosynthesis), Bleom4_C2 (AAG02355.1, bleomycin biosynthesis), LgnD_C (AIZ66879.1, legonmycin biosynthesis), HasO_C (CZT62784.1, hassallidin biosynthesis), Zmn17 (CCM44337.1, zeamine biosynthesis), NdaA (ATP76243.1, nodularin biosynthesis), McyA_C (BAA83992.1, microcystin biosynthesis), PuwF_C (AIW82283.1, puwainaphycin biosynthesis), and PuwG_C (AIW82284.1, puwainaphycin biosynthesis) . Abbreviations: LCL: condensation between two L amino acids, DCL: condensation between an L and D amino acid, Dual: condensation and epimerization, Starter:

acylation to variety of molecules, Hybrid: condensation of amino acid to polyketide. Tree scale represents average expected percentage (1= 100%) of amino acid substitutions per site.

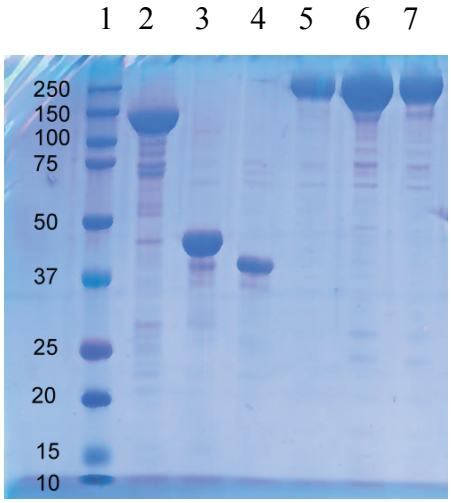


Figure S15. SDS-PAGE analysis of proteins used in this study. Lane 1, protein ladder; Lane 2, AmbB; Lane 3, AmbC; Lane 4, AmbD; Lane 5, full length AmbE; Lane 6, AmbE S1958A (TE mutant); Lane 7, AmbE S1819A (T_2 mutant).

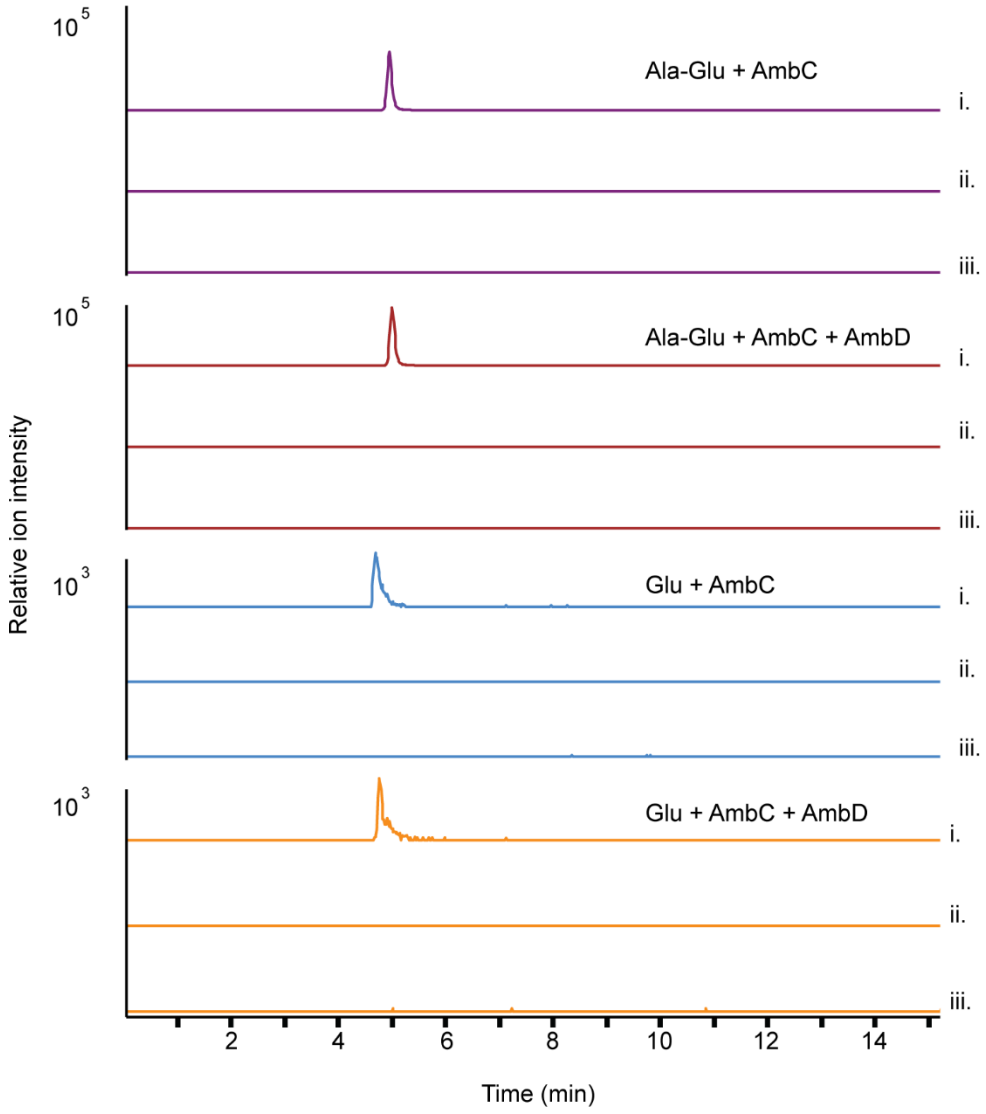


Figure S16. AmbC and AmbD do not modify free Ala-Glu or Ala. Free Ala-Glu or Ala were incubated with AmbC or both AmbC and AmbD in the presence of the necessary cofactors. Top two graphs: EICs of i) starting material (Ala-Glu, *m/z*=219.1975, $[M + H]^+$), ii) possible monohydroxylated product (hydroxy-Ala-Glu, *m/z*=235.0925, $[M + H]^+$), and iii) possible dihydroxylated product (dihydroxy-Ala-Glu, *m/z*=251.0874, $[M + H]^+$). Bottom two graphs: EICs of i) starting material (Glu, *m/z*=148.0604, $[M+H]^+$), ii) possible monohydroxylated product (hydroxy-Glu, *m/z*=164.0553, $[M+H]^+$), and iii) possible dihydroxylated product (dihydroxy-Glu, *m/z*=180.0503, $[M+H]^+$).

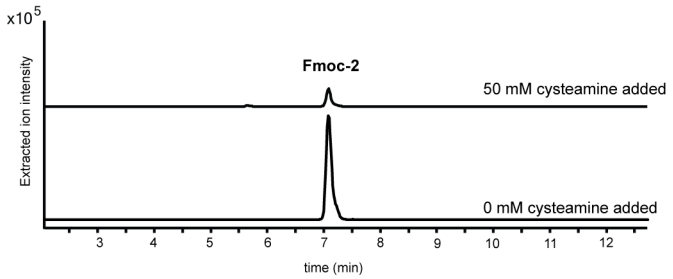


Figure S17. Ala-AMB is synthesized in the one-pot assay containing 50 mM cysteamine. One-pot assay was conducted in the presence or absence of 50 mM cysteamine and the resulting products were derivatized using Fmoc-Cl. EIC of Fmoc-Ala-AMB (Fmoc-**2**, *m/z*=425.1707, $[M + H]^+$) is shown, indicating that Ala-AMB is made in the presence of 50 mM cysteamine, albeit at a lower level than without cysteamine.

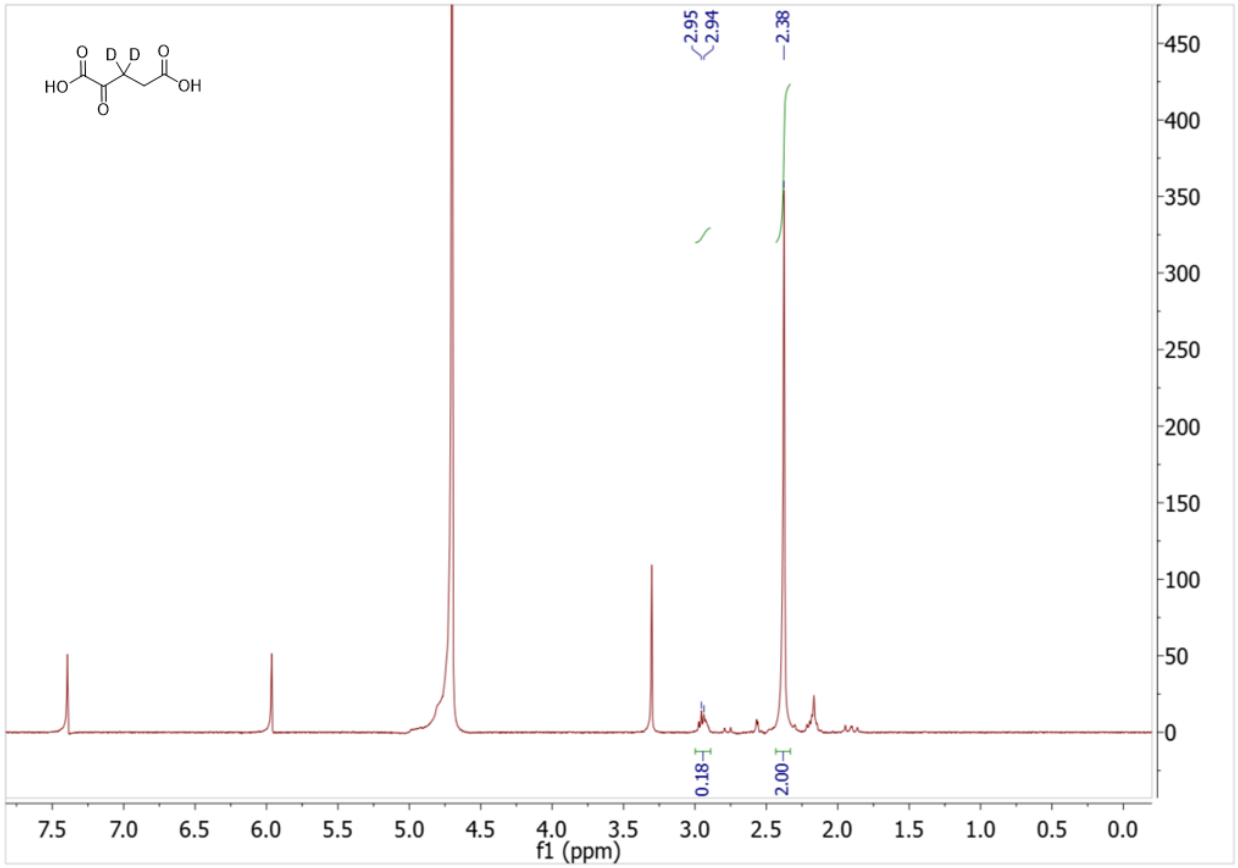


Figure S18. ^{1}H NMR spectrum of 3,3-D_2-α-ketoglutarate in D_2O (400 MHz). Peak integration indicates that 92% of α-KG are 3,3-D_2-α-KG.

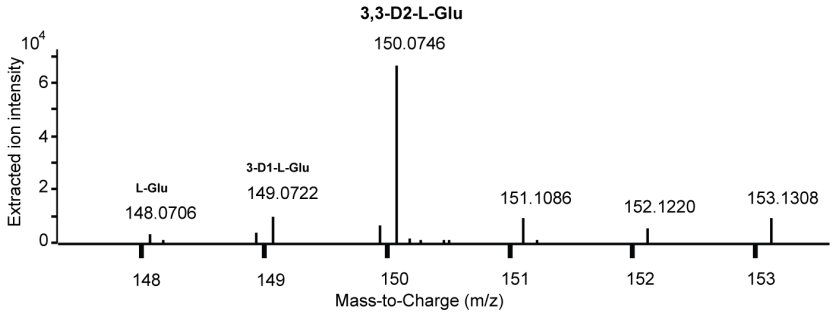


Figure S19. Mass spectrum of 3,3-D_2-L-Glu. Mass spec signals for D_0, D_1, and D_2-L-glutamic acid are shown. It is estimated that 92% of all Glu are 3,3-D_2-L-Glu.

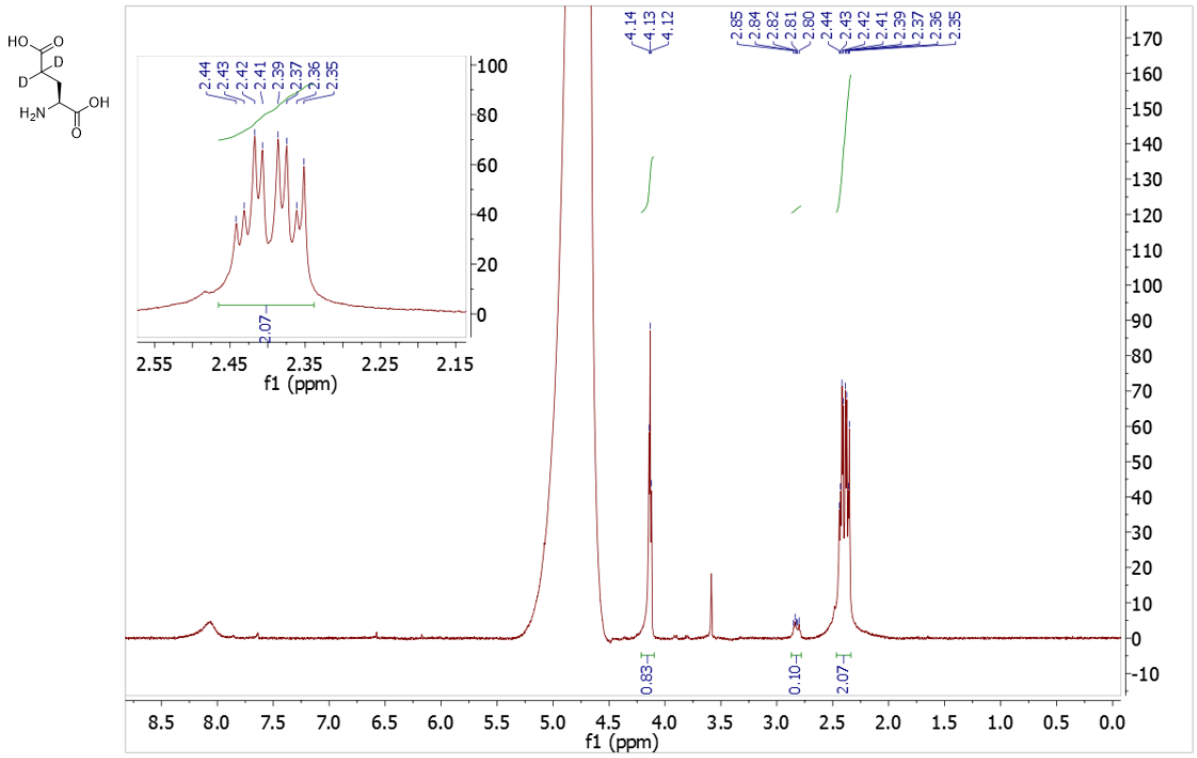


Figure S20. ^{1}H NMR of 4,4-D_2-L-Glu in D_2O (600 MHz). Peak integration indicates that 95% of all Glu are 4,4-D_2-Glu.

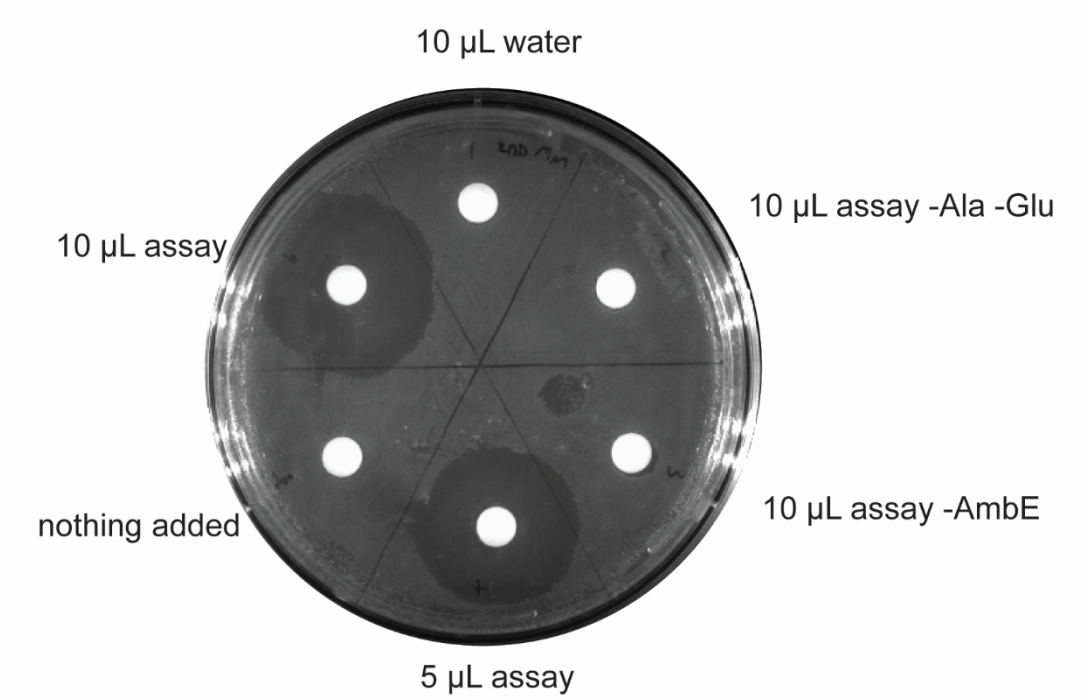


Figure S21. Ala-AMB synthesized *in vitro* inhibits growth of *E. coli* MG1655 in minimal media. Reconstitution reactions lacking enzyme AmbE or substrates (Ala, Glu) did not result in growth inhibition. Ala-AMB may act as a prodrug and take advantage of dipeptide permeases in *E. coli* for entry, and AMB may be released by an intracellular *E. coli* protease.[19]

References:

(1) Ogrel, A.; Vasilenko, I. A.; Lugtenburg, J.; Raap, J. *Recl. des Trav. Chim. des Pays-Bas* **1994**, *7*, 369–375.

(2) Stols, L.; Gu, M.; Dieckman, L.; Raffen, R.; Collart, F. R.; Donnelly, M. I. *Protein Expr. Purif.* **2002**, *1*, 8–15.

(3) Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C. J.; Lu, S.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z.; Yamashita, R. A.; Zhang, D.; Zheng, C.; Geer, L. Y.; Bryant, S. H. *Nucleic Acids Res.* **2017**, *45*, 200–203.

(4) Heckman, K. L.; Pease, L. R. *Nat. Protoc.* **2007**, *4*, 924–932.

(5) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Düsterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kötter, P.; Krug, D.; Masschelein, J.; Melnik, A. V; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nützmann, H.-W.; Pan, G.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X.; Yim, G.; Yu, F.; Xie, Y.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gómez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y.-Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K.-D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Höfte, M.; Jensen, S. E.; Ju, J.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H.-J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linquan, B.; Liu, X.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Méndez, C.; Metsä-Ketelä, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Müller, R.; Neilan, B. A.; Nett, M.; Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J.-L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Süssmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glöckner, F. O. *Nat. Chem. Biol.* **2015**, 625.

(6) Huang, S.; Tabudravu, J.; Elsayed, S. S.; Travert, J.; Peace, D.; Tong, M. H.; Kyeremeh, K.; Kelly, S. M.; Trembleau, L.; Ebel, R.; Jaspars, M.; Yu, Y.; Deng, H. *Angew. Chem., Int. Ed.* **2015**, *43*, 12697–12701.

(7) Masschelein, J.; Clauwers, C.; Awodi, U. R.; Stalmans, K.; Vermaelen, W.; Lescrinier, E.; Aertsen, A.; Michiels, C.; Challis, G. L.; Lavigne, R. *Chem. Sci.* **2015**, *2*, 923–929.

(8) Cheng, Y. Q.; Yang, M.; Matter, A. M. *Appl. Environ. Microbiol.* **2007**, *11*, 3460–3469.

(9) Pancrace, C.; Jokela, J.; Sassoon, N.; Ganneau, C.; Desnos-Ollivier, M.; Wahlsten, M.; Humisto, A.; Calteau, A.; Bay, S.; Fewer, D. P.; Sivonen, K.; Gugger, M. *ACS Chem. Biol.* **2017**, *7*, 1796–1804.

(10) Moffitt, M. C.; Neilan, B. a. *Appl. Environ. Microbiol.* **2004**, *11*, 6353–6362.

(11) Mareš, J.; Jek, J. H.; Urajová, P.; Kopecký, J.; Hrouzek, P. *PLoS One* **2014**, *11*, 1–11.
(12) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. *PLoS One* **2012**, *3*, 1–9.
(13) Edgar, R. C. *Nucleic Acids Res.* **2004**, *5*, 1792–1797.
(14) Stamatakis, A. *Bioinformatics* **2014**, *9*, 1312–1313.
(15) Letunic, I.; Bork, P. *Nucleic Acids Res.* **2016**, *1*, 242–245.
(16) Keith, D. D.; Tortora, J. a; Yang, R.; Hy, E. *J. Org. Chem.* **1978**, *19*, 3711–3713.
(17) Rose, N. G. .; Blaskovich, M. a; Wong, A.; Lajoie, G. a. *Tetrahedron* **2001**, *8*, 1497–1507.
(18) Berkowitz, D. B.; Charette, B. D.; Karukurichi, K. R.; McFadden, J. M. *Tetrahedron Asymmetry* **2006**, *6*, 869–882.
(19) Abouhamad, W. N.; Manson, M.; Gibson, M. M.; Higgins, C. F. *Mol. Microbiol.* **1991**, *5*, 1035–1047.