Supporting Information

Comparative Metabolomics and Structural Characterizations Illuminate Colibactin Pathway-Dependent Small Molecules

Maria I. Vizcaino[†,‡], Philipp Engel[†,‡,⊥], Eric Trautman[†,‡], Jason M. Crawford[*,†,‡,∥]

[†]Department of Chemistry, Yale University, New Haven, CT 06520, USA
[‡]Chemical Biology Institute, Yale University, West Haven, CT 06516, USA
[∥]Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, CT 06510, USA

*correspondence to jason.crawford@yale.edu

Table of Contents

1. Experimental Methods

Bacterial strains and plasmids. Strains and plasmids used in this study are listed in Table S1. All gene deletion mutants were generated using the Red recombinase system of bacteriophage lambda.[1] For *clbP* and *clbL* gene deletions in pBAC *clb*+, the FRT-flanked apramycin (AP) resistance cassette of the plasmid pIJ773[2] was PCR amplified using primers with short sequence extensions homologous to the flanking regions of the gene to be deleted (Table S1). Purified PCR products were digested for 1 h with *Dpn*I (NEB), desalted by dialysis, and transformed into *E. coli* DH10B carrying plasmids pBAC *clb*+ and pKD46. After 1.5 h of phenotypic expression at 30°C, bacteria were plated on Lysogeny broth (LB) agar supplemented with 50 μg/mL AP and 12.5 μg/mL chloramphenicol (CAM) and incubated at 30°C for about 20 h. Colonies were analyzed with overspanning PCR to confirm replacement of the gene of interest with the FRT-flanked AP resistance cassette. The cassette was subsequently flipped out by transforming with the plasmid pKD20 encoding the Flp recombinase.[1] Successful deletion of the resistance cassette was confirmed by replica plating colonies resistant to ampicillin and CAM on LB agar supplemented with 50 μg/mL AP and 12.5 μg/mL CAM and by overspanning PCR. For experimental purposes, all pBAC constructs were purified and re-transformed into *E. coli* DH10B. Two *ΔclbP* mutants were created from the pBAC construct, one that completely deleted *clbP* but also affected the first two amino acids of *clbQ* (pBAC *ΔclbP**) due to gene overlap. To avoid polar effects, a second mutant (pBAC *ΔclbP*) was generated by partial gene deletion of *clbP*, leaving 27 bp at the 5' end region of the *clbP* gene that overlaps with the ribosomal binding site and 3' end region of *clbQ* (Table S1). Deletion mutants of *E. coli* Nissle1917 (EcN) were constructed as described above. However, instead of amplifying the AP resistance cassette of pIJ773, the spectinomycin (SP) resistance cassette of pIJ778 was amplified and used for transformation. Additionally, streptomycin (50 μg/mL) was used for selection of transformants instead of AP. EcN *ΔclbP* mutant was created based on partial deletion of *clbP*. To generate the complete deletion of the *clb* locus in EcN (EcN *Δclb*), the upstream region of the *clb* island was replaced with a SP resistance cassette and flipped out as described above. Then, the downstream region of the *clb* island was replaced with a SP resistance cassette. Flipping out this second replacement resulted in the complete deletion of the entire region between the upstream FRT site and the two downstream FRT sites. Deletions were confirmed by PCR and sequencing.

Permanent stock cultures of all *E. coli* strains (Table S1) used in the study were stored in 25% glycerol at -80°C.

HeLa cell assays. Megalocytosis phenotypes previously reported[3] were confirmed by HeLa cell assays for pBAC strains (Figure S2). HeLa cells were cultured in Dulbecco's modified eagle medium (DMEM, GIBCO®) supplemented with 5% fetal calf serum (FCS, GIBCO®) at 37°C in 5% CO_2. For the megalocytosis assays, ca. $2x10^4$ HeLa cells per well were seeded in a 24-well plate and incubated for 14 h. The cultured HeLa cells were exposed to the different *E. coli* strains for 4 h. To this end, stationary overnight cultures of *E. coli* strains grown in LB were spun down and re-suspended in DMEM supplemented with 5% FCS. Optical densities (OD) of bacterial suspensions were measured at a wavelength of 600 nm (OD_{600}) and cell number adjusted to $2x10^7$ bacteria per mL of DMEM supplemented with 5% FCS (based on an 1.0 OD_{600} equal to 10^9 cells). To infect the HeLa cells with a multiplicity of infection (MOI) of 100, 100 µL of the diluted suspension was added to each well of cultured HeLa cells containing 900 µL of DMEM supplemented with 5% FCS. After 4 h of bacterial exposure, the HeLa cells were washed at least four times with DMEM supplemented with 5% FCS. HeLa cells were then incubated in DMEM supplemented with 5% FCS and 200 µg/mL of gentamicin to suppress further bacterial growth. HeLa cell experiments for testing the activities of different compounds were carried out in the same way. Small molecules were dissolved in DMSO and pre-diluted in DMEM supplemented with 5% FCS. The DMSO concentration did not exceed 1% (v/v) when added to HeLa cells. The experiments were stopped after 4 h by washing as described above.

Analysis of megalocytosis phenotypes. For Giemsa staining, HeLa cells were washed once with 1 mL 1x PBS, exposed to a 1:1 mixture of 1x PBS and methanol for 2 min, exposed to pure methanol for 10 min, washed once with pure methanol, and finally stained with 300 µL Giemsa (Sigma) for 10 min. Cells were then washed 3x with 1 mL of ddH_2O. For methylene blue staining and protein quantification, HeLa cells were fixed with 4% paraformaldehyde in PBS for 30 min at room temperature, washed once with 1x PBS, and then stained with 150 µL methylene blue (1% w/v in Tris- HCl 0.01M) for 20 min. Cells were washed three times with 1 mL 1x PBS for 5 min on a shaking platform. Then, methylene blue was extracted with 1 mL 0.1 N HCl and

the absorbance at a wavelength of 660 nm measured. If necessary, extracts were diluted to measure OD_{660} between 0.05 and 1.00.

Organic extractions for metabolomics analysis. Single colonies of pBAC *clb*-, pBAC *clb*+, pBAC Δ*clbP*, pBAC Δ*clbL*, EcN *wt*, EcN *ΔclbP* and EcN *Δclb* were used to inoculate 5 mL of LB, supplemented with 12.5 µg/mL CAM when needed, for a total of five biological replicates per strain, and incubated at 37°C on a shaker (250 rpm) for 16 h. The next day, the LB seed cultures were all brought up to OD_{600} of 3.45 and 25 µL of this suspension was used to inoculate 5 mL of Difco M9 minimal medium supplemented with casamino acids (5 g/L), 20% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and CAM when needed (12.5 µg/mL), and 1 g/L of the following amino acids: Gly, Cys, Ser, Asn, Ala, Val. The inoculated cultures were grown at 37°C (250 rpm) until an OD_{600} of 0.5-0.6, at which point induction was performed on the pBAC cultures with 1 µL of 1.0 M isopropyl β-D-l-thiogalactopyranoside (IPTG) solution. All pBAC and EcN cultures were then incubated at 25°C while shaking (250 rpm). After 48 h, each whole culture was extracted with 6 mL ethyl acetate (EtOAc) by shaking vigorously for 30 seconds. The two layers were separated by centrifugation (3000 rpm x 15 min), and 4 mL of EtOAc was carefully collected, dried, and stored at -20°C until analysis.

Metabolomics data acquisition. All high-resolution mass (HRM) spectrometry was performed with an electrospray ionization (ESI) source on an Agilent (Santa Clara, CA, USA) iFunnel 6550 Quadrupole time-of-flight (QTOF) coupled to an Agilent Infinity 1290 high-performance liquid chromatogram (HPLC). Metabolites were analyzed on a Phenomenex Kinetex 1.7 µ C18 100Å column (100 x 2.10 mm) with a water:acetonitrile (ACN) gradient solvent system containing 0.1 % formic acid (FA): 0-2 min, 5% ACN; 2-26 min, 5 to 98% ACN; hold for 10 min, 98% ACN. Column temperature was set at 25°C and flow at 0.3 mL/min. Mass spectra were acquired using Dual Agilent Jet Stream (AJS) ESI in positive mode scanning from 25-1700 m/z at 1.00 spectra/second. The capillary and nozzle voltage were set at 3500V and 1000V, respectively. The source parameters were set with a gas temperature at 225°C and flow at 12 L/min, nebulizer at 50 psig, sheath gas temperature at 275°C and flow at 12 L/min. MS data were acquired with MassHunter Workstation Data Acquisition (Version B.05.01, Agilent Technologies) and processed with MassHunter Qualitative Analysis (Version B.05.00, Agilent Technologies).

Immediately prior to analysis, each sample extract was dissolved in 500 µL MeOH and 5 µL of a 1:5 dilution was injected. For the pBAC *clb*- sample, a non-diluted injection was also performed to increase identification of the molecular features (MOFs) from the control. During data acquisition, a MeOH blank was run prior to the first sample injection and a MeOH wash was performed after each sample set. The centroid MS data was processed with MassHunter Qualitative Analysis set to extract molecular features using the 'common organic molecules' model. The extracted MS data, set at an intensity cut-off of 1.0 raw count abundance, was statistically analyzed using MassHunter Mass Profiler Professional (MPP, version B.12.01, Agilent Technologies). A conservative unique ion list for each colibactin-containing sample was built by removing the MOFs found in at least one of the control replicates (no gene cluster) and keeping those that were present in all five biological replicates (Figure S3).

Tandem MS (MS^2) networking. Tandem mass spectrometry (MS^2) was performed using an untargeted auto-MS^2 mode and a targeted auto-MS^2 mode selecting only for the *clb*-pathway dependent MOFs present in the preferred unique ion list acquired for each sample. For the untargeted approach, an unbiased isotope model was used to obtain optimal fragmentation coverage. To maximize MS^2 fragmentation coverage of unique ions in the targeted approach, a peptide and unbiased isotope model method was run for each sample. MS^2 data was acquired im the positive mode from 25-1700 m/z for both approaches at a ms/ms scan rate of 5 spectra/sec. The untargeted approach had a fixed collision energy of 10, while the targeted approach had fixed collision energies of 10 and 30. The same LC method and column was used as described above. The MS^2 data files were used to build mass spectral networking files as described previously[4] using the open source software platform Cytoscape version 3.1.0 (http://www.cytoscape.org). Clusters were built based on cosine cutoff of 0.5, which dictates the connectivity strength between the ion masses. The untargeted auto approach led to a network of 511 nodes that had limited coverage of colibactin pathway-dependent unique ions (Figure S4). In our analysis, the untargeted approach missed ions of potential interest amongst nonentity metabolites and limited our ability to focus on colibactin pathway-dependent metabolites. Our next step of targeting specific colibactin MOFs allowed for a more tightly packed network

(Figure 2), which contained two layers of statistical analysis (MPP and networking) and allowed us to preselect metabolites of interest for further structural characterization.

General conditions for large-scale growth and LC-MS analysis. A single colony of the corresponding *E. coli* DH10B strain, grown overnight at 37°C on an LB agar plate supplemented with 12.5 µg/mL CAM, was used to seed 5 mL LB with antibiotic. This seed culture was grown overnight at 37°C with shaking (250 rpm). The final growth medium, Difco M9 minimal medium supplemented with casamino acids (5 g/L), 20% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and CAM (12.5 µg/mL), was inoculated with the seed culture at a ratio of 1:200. Each culture was grown at 37°C with shaking (250 rpm) until the OD_{600} measured between 0.5-0.6. Growth cultures were cooled at 4°C for approximately 10 min prior to induction with IPTG (0.2 mM final concentration). Growth was continued at 25°C with shaking (250 rpm) for a total of 48 h. Monitoring of desired masses was performed on an Agilent 6120 Infinity series quadrupole mass spectrometer coupled to an Agilent 1260 series LC system consisting of a quaternary pump, autosampler, thermostat column compartment, vacuum degasser, and diode array detector. The mass spectrometer was operated with an atmospheric pressure electrospray ionization (API-ES) source in positive ion mode. LC-MS chromatographic analysis was performed over a Polaris 5 C18-A HPLC column (250 x 4.6 mm, 180Å, 5 µm particle size, Agilent) with a water:ACN gradient containing 0.1% FA at 0.5 mL/min: 0-25 min, 5 to 98% ACN; 25-35 min, hold at 98% ACN.

Isolation of metabolites 1 and 2. A 9 L culture of pBAC *clb*+ was centrifuged at 4000 rpm for 20 min. The decanted supernatant was extracted with 4.5 L of EtOAc (Acros Organics). The dried EtOAc layer (282.6 mg) was purified with a SEP-PAK C18 column (Thermo Scientific) and fractions eluted with increasing concentrations of MeOH. Fractions 7 and 8 (eluted with 70% and 80% MeOH, respectively), containing both *m/z* 343 and *m/z* 341 positive ions, were combined (41.2 mg) and further purified by semi-preparative HPLC. A Polaris 5 C18-A (Agilent, 250 x 10 mm) column was used with a water:ACN gradient containing 0.1% FA at 2 mL/min: 0-5 min, 20% ACN; 5-40 min, 20-100% ACN , 41-45 min, 100% ACN. Fractions containing the desired masses, monitored by LC-MS, were combined and purified on the same

semi-prep HPLC system using the following gradient method: 0-5 min, 50% ACN, 5-40 min, 50-62% ACN; 41-45 min, 100% ACN. Fractions containing the desired mass, monitored by LC-MS, were combined and characterized by NMR (metabolite **1**, Figures S5-S8; metabolite **2**, Figures S11-S14). HRMS was acquired for **1** and **2**, and the molecular formula for each was confirmed based on exact mass and MS^2 fragmentation patterns (Figures S9 and S10, respectively**)**.

Isolation of metabolite 3. A 4 L culture of pBAC *ΔclbP** was grown for 48 h and the whole culture was extracted with 2 L of EtOAc (Acros Organics). The dried EtOAc layer (193.0 mg) was purified with a SEP-PAK C18 column (Thermo Scientific) and fractions were eluted with increasing concentrations of MeOH. Fraction 5 (eluted with 100% MeOH), containing *m/z* 440 positive ion as monitored by LC-MS, was further purified by semi-prep HPLC on a Polaris 5 C18-A (Agilent, 250 x 10 mm) column with a water:ACN gradient containing 0.1% FA at 2 mL/min: 0-5 min, 50% ACN; 5-35 min, 50-100%; 35-45 min, 100% ACN. Fractions containing the desired mass, as monitored by LC-MS, were combined and characterized by NMR (Figures S19-S22). HRMS was acquired for **3**, and the molecular formula was confirmed based on exact mass and MS^2 fragmentation patterns (Figure S18).

NMR analyses of small molecules isolated from bacterial extracts. One and two-dimensional (1D and 2D, respectively) NMR spectra were recorded on an Agilent DD2 400 MHz (compound **1**) or a 600 MHz NMR (compounds **2** and **3**) equipped with a cold-probe (VnmrJ 3.2 Datastation). All three compounds were dissolved in dimethyl-sulfoxide-$_{d6}$ (DMSO-$_{d6}$) and a series of 1D (^{1}H) and 2D (gCOSY, gHSQCAC and gHMBCAD) data were collected and calibrated to the residual solvent peak DMSO-$_{d6}$ at δ2.50 ppm. Data are recorded in Table S3, S4, and S5 for isolated metabolites **1**, **2**, and **3** respectively. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q= quartet, qu = quintet, dd = doublet of doublets, m = multiplet.

Determination of the absolute configuration of isolated compounds 1-3. Compounds **1-3** were acid hydrolyzed, derivatized, and analyzed by the Marfey's method.[5] Briefly, ca. 0.3 mg of the respective compound was hydrolyzed in 0.5 mL of 6N HCl for 1.5 h at 100°C. The acid was

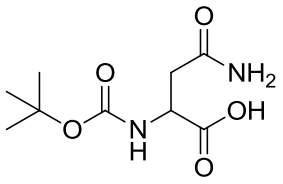
removed by rotary evaporation, re-suspended in deionized-water (0.5 mL) and re-dried by rotary evaporation. This was repeated three more times to remove residual acid. The hydrolyzed compounds were dissolved in 100 µL of 1N $NaHCO_3$. Amino acid standards L-Asn, D-Asn, L-Asp and D-Asp were individually dissolved in 100 µL of 1N $NaHCO_3$ to 1 mg/mL. To the 1N $NaHCO_3$ solutions, 50 µL (10 mg/mL) of *N*α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA, Marfey's reagent) in acetone was added and the reaction incubated in a water bath for 15 min at 55°C. After the allotted time, the reactions were quenched by 50 µL of 2N HCl. The resulting samples were diluted with ACN (1/5 ratio) and analyzed by LC-MS. LC-MS chromatographic separation was performed over a Phenomenex C18-A HPLC column (250 x 4.6 mm, 180Å, 5 µm particle size, Agilent) with a water:ACN gradient containing 0.1% FA at 0.5 mL/min: 0-2 min, 5% ACN, 2-25 min, 5-98% ACN. Analysis with both Asn and Asp amino acids indicates the conversion of Asn to Asp during the acid hydrolysis step as suggested by an observed *m/z* 386 positive ion for **1**, **2**, and **3** as opposed to the expected *m/z* 385 positive ion. Retention times: L-Asn, 16.53 min; D-Asn, 16.49 min; L-Asp, 17.21 min; D-Asp, 17.36 min; **1**, 17.36 min; **2**, 17.36; **3**, 17.36 min.

Ozonolysis of isolated compound 2. Compound **2** (ca. 1 mg) was subject to ozonolysis, reductive workup, and analyzed by HRMS QTOF to determine the location of the cis-double bond. Briefly, **2** was dissolved in 1:1 v/v of dichloromethane:methanol (2 mL total) and cooled with a dry ice-acetone mixture. The solution was then treated with ozone-containing oxygen stream for ca. 5 min. Once complete, the reaction was sparged with O_2 and then N_2. The resulting ozonide compounds were treated with dimethyl sulfide and allowed to stir for 20 min. After the allotted time, the solution was dried down under a N_2 stream before HRMS QTOF analysis.

Determination of absolute configuration for isolated compound 3. Synthetic diasteromers (prepared as described below) were separated using a Phenomenex Kinetex C18 column (250 x 4.6 mm, 100Å, 5 µm particle size). An isocratic separation was used with 35% ACN in water at a flow rate of 1 mL/min. Products synthesized from racemic and entantiopure propylene oxide were compared to the isolated natural product (Figure S23).

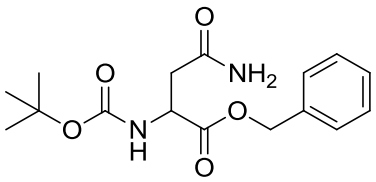
General information for synthesis reactions. All non-aqueous reactions were performed with oven-dried glassware under N_2 unless otherwise stated. Commercial reagents were used as received. *N*-myristoyl-D-asparagine (**1**) and its enantiomer, *N*-myristoyl-L-asparagine (**12**), were synthesized from the corresponding enantiopure amino acid. (*S)*-L-asparagine was purchased from Sigma-Aldrich, (*R*)-D-asparagine from Alfa Aesar. Low-resolution mass spectra (LRMS) were acquired using an Agilent LC-MS 6120 instrument with an atmospheric pressure chemical ionization (APCI) source. NMR spectra were measured on an Agilent 400 MHz spectrometer in deuterated solvents calibrated to the solvent residual peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q= quartet, qu = quintet, dd = doublet of doublets, m = multiplet.

Synthesis of *N*-myristoyl-asparagine (1 and 12) (Scheme S1)



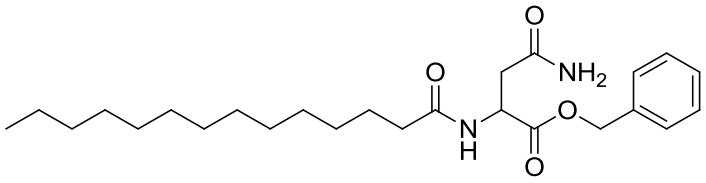
S1

L- or D-Asparagine (1.0 g, 1 equivalent), di-tert-butyl dicarbonate (1.93 g, 1.2 equivalents), and K_2CO_3 (1.05 g, 1 equivalent) were dissolved in 40 mL of a 1:1 dioxane: water solution in a round bottom flask. The solution was stirred at room temperature for 4 h and then acidified with HCl to a pH of 2 until a precipitate formed. The precipitate was collected by vacuum filtration and washed with water, yielding 1.134 g (64% yield) of *N*-boc-asparagine. ^{1}H NMR (DMSO-*$_{d6}$*, 400 MHz): δ 12.51 (s, 1H), δ 7.29 (s, 1H), δ 6.88 (s, 1H), δ 6.84 (d, 1H), δ 4.20 (m, 1H), δ 2.44 (m, 2H), δ 13.5 (s, 1H). LRMS (APCI): Calculated for $C_9H_{16}N_2O_5$ [M-H] 231.11, observed [M-H] 231.1.



S2

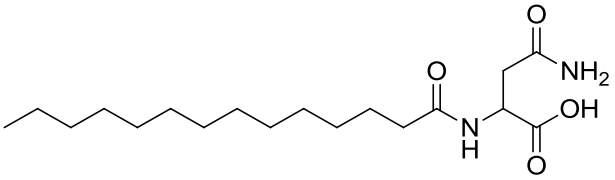
N-boc-asparagine (1 g, 1 equivalent) was dissolved in 50 mL MeOH. A 20% (w/v) solution of Cs_2CO_3 was added until the pH was between 7 and 8, and then stirred for 1 h. The solvent was removed by rotary distillation and re-dissolved in 20 mL *N,N'*-dimethylformamide (DMF). Benzyl bromide (770 µL, 1.5 equivalents) was added, and the solution was stirred at room temperature overnight. The solution was concentrated and water was added to form a precipitate. The precipitate was collected by vacuum filtration, dissolved in 100 mL EtOAc, and washed with 50 mL of water. The aqueous layer was extracted twice with 100 mL EtOAc. The organic layers were combined and dried over $MgSO_4$. The solvent was removed and *N*-boc-*O*-benzyl asparagine was recrystallized from EtOAc and hexanes, resulting in a white solid (1.003 g, 75% yield). ^{1}H NMR (acetone-$d6$, 400 MHz): δ 7.36 (m, 5H), δ 6.93 (s, 1H), δ 6.36 (s, 1H), δ 6.24 (d, 1H), δ 5.15 (s, 2H), δ 4.54 (m, 1H), δ 2.85 (dd, 1H), δ 2.72 (dd, 1H), δ 1.40 (s, 9H). LRMS (APCI): Calculated for $C_{16}H_{22}N_2O_5$ [M+H] 323.15, observed $[M+Na]^+$ 345.1.



S3

N-Boc-*O*-benzyl asparagine (800 mg, 1 equivalent) was dissolved in 20 mL dichloromethane (DCM). Trifluoroacetic acid (1.9 mL, 10 equivalents) was added, and the reaction was stirred at room temperature for 30 min. The solvent was removed and the crude material, triethylamine (1.4 mL, 4 equivalents), and 4-dimethylaminopyridine (DMAP, 30.3 mg, 0.1 equivalents) were dissolved in 10 mL DCM at 0°C. Myristic acid was dissolved in 20 mL DCM, to which 3 drops of DMF and oxalyl chloride (2 M in THF) (1.7 mL, 1.3 equivalents) were added. The solution was stirred for 1 h and then the solvent was removed. The brown liquid acid chloride was dissolved in 20 mL DCM and added drop wise to the solution containing *O*-benzyl asparagine over 20 min. The solution was stirred for 1 h, and the precipitate was collected by vacuum filtration. The crude material was washed with DCM and dried, yielding 480 mg (45% yield over two steps) of the benzyl ester of N-myristoyl asparagine. ^{1}H NMR (MeOD, 400 MHz): δ 7.36 (m, 5H), δ 4.86(s, 2H), δ 4.48 (dd, 1H), δ 2.79 (dd, 1H), δ 2.74 (dd, 1H), δ 2.21 (t, 2H), δ 1.57

(quin, 2H), δ 1.28 (20H), δ 0.90 (t, 3H); LRMS (APCI): Calculated for $C_{25}H_{40}N_2O_4$ [M+H] 433.3, observed [M+H] 433.2.



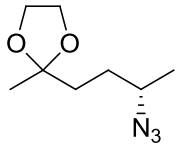
1 or 12

Palladium on carbon (10% w/w), 20 mL EtOH, and the benzyl ester of either *N*-myristoyl-D-asparagine or *N*-myristoyl-L-asparagine (367 mg) were combined in a round bottom flask. Hydrogen balloons were attached, and the mixture was stirred overnight. For the *S* enantiomer, the mixture was filtered, and the filtrate was dried to give the pure product (55.8 mg of *N*-myristoyl-L-asparagine (**12**), 20% yield). For purification of the *R* enantiomer, an additional step was performed to increase the yield. The filter cake was dissolved in EtOH and filtered to remove Pd/C. The product was recrystallized from hot EtOH, giving a total of 217 mg of *N*-myristoyl-D-asparagine (**1**) as a white solid (70% yield). ^{1}H NMR (MeOD, 400 MHz): δ 4.71 (dd, 1H), δ 2.78 (dd, 1H), δ 2.72 (dd, 1H), δ 2.23 (t, 2H), δ 1.60 (qu, 2H), δ 1.31 (m, 20H), δ 0.90 (t, 3H). NMR spectra were also recorded in DMSO-$_{d6}$ for comparison with the isolated natural product (see Figure S28 and Table S8). ^{13}C spectrum for synthetic compound is shown in Figure S29. LRMS (APCI): Calculated for $C_{18}H_{34}N_2O_4$ [M+H] 343.25, observed [M+H] 343.3.

Synthesis of 3 and 2,5-dimethylpyrroline (11) (Scheme S2)

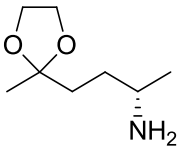
Synthesis of 2-(3-chlorobutyl)-2-methyl-1,3-dioxolane was carried out based on the synthesis reported by Warren.[6] Attempts to synthesize 5-amino-2-hexanone (**10**), either via acid catalyzed deprotection of the carbonyl or direct reduction from 5-azido-2-hexanone via a Staudinger reduction, resulted in the formation of 2,5-dimethylpyrroline (**11**). Recovery of pure product by distillation was not achieved on the small scale attempted, thus, **11** was synthesized as reported by Evans.[7] The protected amino intermediate was coupled to **1** followed by deprotection to produce the 5-amino-2-hexanone-linked *N*-myristoyl-asparagine (**3**) for bioassays and stereochemical verification. For synthesis of the enantioenriched compound, enantiopure (*S*)-

propylene oxide was used. Two inversions of configuration during the synthesis gave the enantioenriched (*S*)-amine.



S5

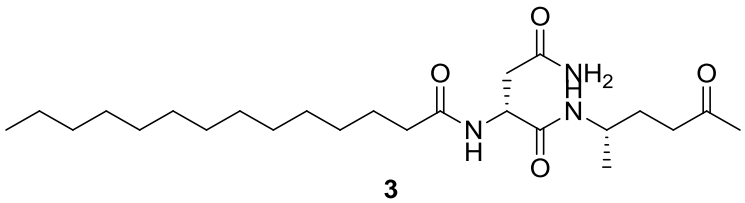
Sodium azide (1.64 g, 25.3 mmol, 1.3 equivalents) and 2-(3-chlorobutyl)-2-methyl-1,3-dioxolane (3 g, 16.9 mmol, 1 equivalent) were dissolved in DMF and heated at 90°C overnight. The reaction mixture was diluted in 100 mL water and extracted three times with 50 mL of DCM. The organic extracts were combined and washed with 50 mL of water, followed by 50 mL of saturated brine. The organic layer was dried over anhydrous Na_2SO_4 and the EtOAc was removed to recover the azide as an oil in 88% yield (2.9 g). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.95 (m, 4H), δ 3.46 (m, 1H), δ 1.77 (m, 1H), δ 1.69 (m, 1H), δ 1.59 (m, 2H), δ 1.32 (s, 3H), δ 1.27 (d, 2H). LRMS (APCI): Calculated for $C_8H_{15}N_3O_2$ [M+H] 186.1, [M+H] was not observed by APCI.



S6

Palladium on carbon (10% w/w) was added to a flame-dried flask under N_2 along with ethyl acetate, 2-(3-azidobutyl)-2-methyl-1,3-dioxolane (1 equivalent) and acetic acid (3 equivalents). The solvent was purged with N_2, and two hydrogen balloons were attached. The reaction was stirred overnight. The flask was flushed with N_2 before adding 50 mL of 1 M HCl. The organic layer was separated from the aqueous layer and the aqueous layer was washed one more time with 50 mL EtOAc. The aqueous layer was basified with 1 M NaOH and extracted with 3 portions of 50 mL EtOAc. The organic layer was dried over anhydrous Na_2SO_4 and the EtOAc was removed to give the amine in 52% yield. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.95 (m, 4H), δ

2.87 (sextet, 1H), δ 2.47 (m, 1H), δ 1.69 (m, 1H), δ 1.59 (m, 2H), δ 1.32 (s, 3H), δ 1.27 (d, 2H). LRMS (APCI): Calculated for $C_8H_{17}NO_2$ [M+H] 160.1, observed [M+H] 160.2.



3

2-(3-aminobutyl)-2-methyl-1,3-dioxolane (48 mg, 1.3 equivalents), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (57 mg, 1.3 equivalents), 1-hydroxybenzotriazole hydrate (47 mg, 1.5 equivalents), and triethylamine (0.1 mL, 3 equivalents) were dissolved in 8 mL DMF. *N*-myristoyl-*(S)*-asparagine (80 mg, 1 equivalent) was added as solution in 2 mL DMF. The reaction was stirred overnight. The reaction mixture was diluted in 25 mL 1 M HCl and extracted with 100 mL EtOAc. The organic layer was washed with 50 mL of saturated $NaHCO_3$, followed by 50 mL of brine. The organic layer was dried over anhydrous Na_2SO_4. The product was dissolved in a 1:1 mixture of THF:1 M HCl at 0^oC. The reaction was stirred overnight, warming to room temperature. The solution was extracted with three portions of 25 mL EtOAc. The organic layer was washed with brine and dried over anhydrous Na_2SO_4 to give 81 mg (62% yield) of a white solid. ^1H NMR (DMSO-*$_{d6}$*, 400 MHz): δ 7.87 (t, 1H), δ 7.44 (t, 1H), δ 7.24 (s, 1H), δ 6.83 (s, 1H), δ 4.44, (m, 1H), δ 3.68 (m, 1H), δ 2.40 (m, 4H), δ 2.08 (t, 2H), δ 2.03 (s, 3H), δ 1.57 (m, 1H), δ 1.46 (m, 3H), δ 1.24 (m, 20H), δ 1.00 (d, 3H), δ 0.85 (t, 3H). LRMS (APCI): Calculated for $C_{24}H_{45}N_3O_4$ [M+H] 440.3, observed [M+H] 440.3. Comparison of NMR spectra with the isolated natural product are shown in Figure S30 (^1H NMR) and tabulated in Table S7. ^{13}C spectrum for synthetic compound is shown in Figure S31.

LC-MS analysis for the *in vivo* cleavage of synthetic 3 by ClbP. The plasmid pClbP (pPEB018) (Table S1) was constructed by amplifying the *clbP* gene from DH10B carrying the plasmid pBAC *clb*+ with primers encoding Sac*I* (prPE236) and Xma*I* (prPE237) restriction sites at their 5' ends. prPE236 also encoded a ribosomal binding site sequence. The resulting PCR product was purified and digested with enzymes Sac*I* and Xma*I* for 1 h. After gel purification, the digested PCR product was ligated into the plasmid backbone pBAD18 which was also cut open with *Sac*I and *Xma*I and purified over an agarose gel. This ligation resulted in plasmid

pPEB018, which was transformed into *E. coli* BL21 or DH10B cells. The correct construction of pClbP was confirmed by sequencing over the ligated *clbP* gene.

Starter cultures of *E. coli* BL21 and DH10B (5 mL), harboring the empty vector pBAD18[8] or pClbP (pPEB018), were inoculated from frozen stocks and grown overnight at 37°C in LB medium supplemented with 100 µg/mL ampicillin. The next morning, 500 µL of these saturated cultures were used to inoculate 25 mL of LB medium containg 100 µg/mL ampicillin. Cultures were incubated at 37°C with shacking at 250 rpm. At an OD_{600} of 0.4-0.5, cultures were induced with L-arabinose and left to grow for another 30 min. Then, the synthetic **3** substrate (10 mM stock solution in DMSO) or DMSO (vector control) was added to each culture to a final concentration of 100 µM, and incubated at 37°C until desired time points. After 4 h and 7 h, 2.5 mL aliquots were removed and immediately extracted with 6 mL of EtOAc. The cultures aliquot were vigorously shaken for 30 sec, centrifuged at 3000 rpm for 15 min, and 4 mL of EtOAc layer recovered. The recovered fractions were dried and then brought up in 500 µL MeOH. LC-MS chromatographic analysis was performed over a Polaris 5 C18-A column (150 x 4.6 mm, 180Å, 5 µm particle size, Agilent) with a water:ACN gradient containing 0.1% FA at 0.5 mL/min: 1-6 min, 5 to 100% ACN; hold for 2 min, 100% ACN.

Biosynthesis of compound 1. To determine physiological concentrations of **1** in our growth conditions, the synthesized compound was used to generate a calibration curve (Figure S27). First, known concentrations (0.097 µg/mL, 0.195 µg/mL, 0.781 µg/mL, 1.56 µg/mL, 3.125 µg/mL, 6.25 µg/mL, 12.5 µg/mL and 25 µg/mL) of the synthesized compound **1** were added to 5 mL of the supplemented M9 media (in duplicates). The spiked media were extracted with 6 mL of EtOAc, shaken vigorously for 20 sec, centrifuged at 3000 rpm for 15 min, and then 4 mL of EtOAc layer was recovered and dried down. Samples were stored at -20°C until analysis. In the meantime, 5 mL cultures of pBAC *clb*+ and pBAC *clb*- were grown (in triplicate) and extracted at four different time points (6 h, 11 h, 25 h, and 48 h). At each time point, the samples were centrifuged and the supernatant was filter-sterilized. Both the filter-sterilized supernatant and the cell pellet were extracted independently with 6 mL of EtOAc as described above. The recovered EtOAc extract (4 mL) was dried for ca. 1 h and then re-dissolved in 500 µL of MeOH. All fractions from the synthetic (dissolved in 500 µL MeOH) and bacterial extracts were analyzed by

LC-MS using a water:ACN solvent system with 0.1% FA: 0-1 min, 10% ACN; 1-6 min, 10-100% ACN; hold for 2 min at 100% ACN. Column used: Eclipse Plus C18, 3.5 µM, 2.1 x 100 mm. Flow rate set at 0.5 mL/min.

Growth inhibition of *Bacillus subtilis* NCIB 3610. Growth inhibition was determined for synthetic compound **1** in comparison to its L-enantiomer (**12**), myristic acid and D-tyrosine. An overnight culture of *B.subtilis* was diluted in fresh LB to an OD_{600} of 0.05. From this overnight culture, 99 µL was added to 1 µL of compound dissolved in DMSO or water in a 96-well plate. The plates were covered with an AirPore Tape Sheet (Qiagen), placed in a humidified bag, and incubated at 37^{o}C while shaking (250 rpm). OD_{600} measurements were taken at four hours post inoculation (Figure S26) using a Perkin Elmer EnVision plate reader.

PDSP functional assay. Primary assays and K_i determinations of synthetic compounds were generously provided by the National Institute of Mental Health's Psychoactive Drug Screening Program (NIMH PDSP), contract # HHSN-271-2008-025C. The NIMH PDSP is directed by Bryan L. Roth MD, PhD at the University of North Carolina at Chapel Hill and Project Officer Jamie Driscoll at NIMH, Bethesda, MD (USA). Synthetic compounds (**1**, **3**, **11**, and **12**) and Tuamine were subjected to primary (1°) assays at 10 µM concentration. Binding activity represents mean % inhibition (N = 4 determinations) for each compound tested against different receptor subtypes. If any compound exhibited >50%, which was considered significant, they were further subjected to a secondary (2°) assay. The Ki (nM) value was determined from non-linear regression of radioligand competition binding isotherms. The Ki values were calculated from best-fit IC_{50} values using the Cheng-Prusoff equation. Detailed experimental protocols for the radioligand and functional receptor assays are available on the NIMH PDSP website (http://pdsp.med.unc.edu/UNC-CH%20Protocol%20Book.pdf). Data is reported in Table S7.

2. Supplemental Tables

Table S1. Strains, plasmids, and primers used in study.

Strains	**Description**	**Reference**
E. coli DH10B	*F- mcrA _(mcrBC-hsdRMS-mrr) [_80d_lacZ_M15] _lacX74 deoR recA1 endA1 araD139_(ara,leu)7697 galU galK _- rpsL nupG*	Invitrogen
E. coli Nissle1917 (EcN *wt*)	wildtype	ArdeyPharm
E. coli BL21	*F- dcm ompT hsdS(rB- mB-) gal [malB+]K-12(λS)*	Invitrogen
E. coli Nissle1917 *ΔclbP* (EcN *ΔclbP*)	*clbP::FRT*, gene deletion leaving 27 bp at the end of *clbP* overlapping with *clbQ*	This study
E. coli Nissle1917 *Δclb* (EcN *Δclb*)	*clb::FRT*, complete deletion of *clb* locus	This study
Plasmids		
pBAC-empty (*clb*-)	pBeloBAC11 without insert	(3)
pBAC *clb*+	genomic fragment of *E. coli* IHE3034 encoding the complete *clb* island cloned into pBeloBAC11	(3)
pBAC *ΔclbP**	pBAC *clbP::FRT*, complete gene deletion of *clbP* with polar effect on *clbQ*	This study
pBAC *ΔclbP*	pBAC *clbP::FRT*, gene deletion leaving 27 bp at the end of *clbP* overlapping with *clbQ*	This study
pBAC *ΔclbL*	pBAC *clbL::FRT*, complete gene deletion of *clbL*	This study
pBAD18	Expression vector with *araBAD* promoter for induction with arabinose	(8)
pClbP (pPEB018)	*clbP* gene amplified from pBAC *clb*+ and cloned into pBAD18	This study
Primers	**Sequence**	**Purpose**
prPE126	CGTAAACGTGCCCAACAGGGAGAAAAGGGCGATGAGT GAATTCCGGGGATCCGTCGACC	forward primer for *clbL* deletion, *ΔclbL*
prPE127	GCGCCGCACCTTTATTGGCGCTGTCACCGATATCCGCC TTGTAGGCTGGAGCTGCTTC	reverse primer for *clbL* deletion, *ΔclbL*
prPE139	CGCGAGAATCTGGCGTTG	forward primer for overspanning PCR, *clbL* deletion
prPE140	CGTCATACCTGCTCCTTATTC	reverse primer for overspanning PCR, *clbL* deletion

prPE128	TCCGAATGTGTTGATGGTATTTTATGAGGTGTTCACAG GATTCCGGGGATCCGTCGACC	forward primer for *clbP* deletions, *ΔclbP** (polar effect) and *ΔclbP* (no polar effect)
prPE129	GCAGAACCACCTGAATATGGCAAACAATACAAACTGA TATGTAGGCTGGAGCTGCTTC	reverse primer for *clbP* deletion, *ΔclbP** (polar effect)
prPE251	ATACAAACTGATATTACTCATCGTCCCACTCCTTGTTGT GTGTAGGCTGGAGCTGCTTC	reverse primer for *clbP* deletion, *ΔclbP* (no polar effect)
prPE141	GTCCGACAGCACGCTACGC	forward primer for overspanning PCR, all *clbP* deletions
prPE142	CAAATCAGCGACACTGAATAC	reverse primer for overspanning PCR, all *clbP* deletions
prPE241	GGTTTGCCGCGATCCTGTCGAGTACACGGACCACTCGC TGATTCCGGGGATCCGTCGACC	forward primer for deletion of upstream region of *clb* locus in *E. coli* Nissle1917, *clb* deletion
prPE242	GTCGATAATATTGATTTTCATATTCCGTCGGTGGTGTAA GTGTAGGCTGGAGCTGCTTC	reverse primer for deletion of upstream region of *clb* locus in *E. coli* Nissle1917, *clb* deletion
prPE124	GGTAGAAATAGTTGTGTAACTATACAAGGAGCAATAG ATATTCCGGGGATCCGTCGACC	forward primer for deletion of downstream region of *clb* locus in *E. coli* Nissle1917, *clb* deletion
prPE125	GAAAAATCATCAGTGGGGAGGCAAACGGTAAGCACCC CGTGTAGGCTGGAGCTGCTTC	reverse primer for deletion of downstream region of *clb* locus in *E. coli* Nissle1917, *clb* deletion
prPE247	GCTCAATGAAAGCGTTCTGC	forward primer for overspanning PCR, deletion of upstream region of *clb* locus in *E. coli* Nissle1917
prPE248	CATTTCTCGCATTAGCCTCTC	reverse primer for overspanning PCR, deletion of upstream region of *clb* locus in *E. coli* Nissle1917
prPE138	CCAGTCCACACTCATCGCTG	reverse primer for overspanning PCR, deletion of complete *clb* locus in *E. coli* Nissle1917

prPE236	GACTGAGAGCTCAAGAAGGAGATATACATGACAATAATGGAACACGTTAG	forward primer encoding a RBS sequence and *Sac*I restriction site for amplification of the *clbP* gene and cloning into pBAD18.
prPE237	TCAGTCCCCGGGTTACTCATCGTCCCACTCC	reverse primer encoding a *Xma*I restriction site for amplification of the *clbP* gene and cloning into pBAD18.

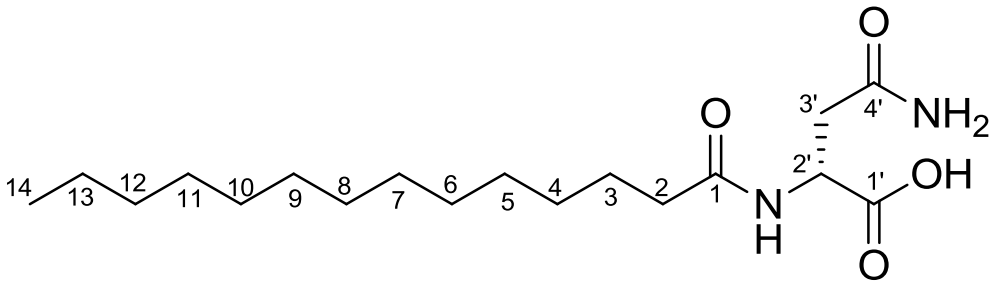
Table S2. Colibactin pathway-dependent molecular features (MOFs). Calculated mass [M], positively charged mass [M+H], retention time (RT), and raw abundance are included. The ten ions whose structures are proposed here based on NMR characterization and network analysis are highlighted in red.

			Raw Abundance[Δ]					
[M]	**[M+H]**	**RT (min)**	**pBAC *clb*-**	**pBAC *clb*+**	**pBAC Δ*clbP***	**EcN Δ*clb***	**EcN *wt***	**EcN Δ*clbP***
97.0892	98.0965	0.92	1.0E+00	4.2E+05	1.0E+00	1.0E+00	(2/5)	1.0E+00
179.0407	180.048	10.40	1.0E+00	1.0E+00	1.0E+00	1.0E+00	3.5E+05	4.1E+05
204.1267	205.1336	17.29	1.0E+00	1.0E+00	1.9E+05	1.0E+00	1.0E+00	1.0E+00
207.0566	208.064	1.18	1.0E+00	4.0E+05	5.5E+05	1.0E+00	(1/5)	1.0E+00
210.1266	211.1341	14.49	1.0E+00	(4/5)	5.8E+05	1.0E+00	1.0E+00	1.0E+00
213.1731	214.1805	12.44	1.0E+00	(4/5)	1.0E+06	1.0E+00	8.4E+05	5.7E+05
215.1887	216.196	11.33	1.0E+00	(4/5)	3.6E+05	1.0E+00	(4/5)	3.1E+05
222.1373	223.1444	0.94	1.0E+00	3.3E+06	1.0E+00	1.0E+00	6.9E+05	1.0E+00
231.1266	232.1339	16.67	1.0E+00	1.1E+06	1.0E+00	1.0E+00	4.9E+05	1.0E+00
231.1837	232.1908	10.47	1.0E+00	(4/5)	2.1E+05	1.0E+00	(3/5)	(4/5)
239.1888	240.1959	13.86	1.0E+00	(4/5)	6.5E+05	1.0E+00	4.9E+05	(4/5)
240.0613	241.0686	7.78	(1/6)	(2/5)	(2/5)	1.0E+00	(1/5)	1.4E+05
241.2045	242.2118	14.99	1.0E+00	(4/5)	(4/5)	1.0E+00	2.7E+05	(4/5)
243.2201	244.2274	13.29	1.0E+00	(4/5)	4.3E+05	1.0E+00	3.4E+05	2.9E+05
257.1992	258.2065	11.57	1.0E+00	(4/5)	1.9E+05	1.0E+00	(3/5)	(3/5)
258.122	259.1292	15.34	1.0E+00	3.4E+05	1.0E+00	1.0E+00	(3/5)	1.0E+00
259.1579	260.1653	17.55	1.0E+00	1.0E+00	5.0E+05	1.0E+00	1.0E+00	1.0E+00
259.215	260.2223	12.51	1.0E+00	5.1E+05	6.3E+05	1.0E+00	3.6E+05	4.1E+05
269.1343	270.1418	16.44	1.0E+00	1.0E+00	3.5E+05	1.0E+00	1.0E+00	(1/5)
269.142	270.1498	17.55	1.0E+00	1.0E+00	4.5E+05	1.0E+00	1.0E+00	(1/5)
271.2493	272.2572	15.14	1.0E+00	(2/5)	1.0E+05	1.0E+00	(1/5)	(1/5)
273.1894	569.3678	17.29	1.0E+00	1.0E+00	3.1E+06	1.0E+00	1.0E+00	5.0E+05
285.2307	286.2381	13.31	1.0E+00	(4/5)	2.1E+05	1.0E+00	(3/5)	(3/5)
287.2463	288.2536	14.37	1.0E+00	2.4E+05	2.8E+05	1.0E+00	(4/5)	2.0E+05
290.0185	291.0258	12.33	1.0E+00	1.0E+00	1.0E+00	1.0E+00	5.5E+05	8.5E+05
301.1687	302.1761	16.72	1.0E+00	1.0E+00	1.9E+05	1.0E+00	1.0E+00	(1/5)
308.0865	309.094	1.79	1.0E+00	1.0E+00	1.0E+00	1.0E+00	(1/5)	1.4E+05
310.063	311.0695	10.23	1.0E+00	(1/5)	8.4E+04	1.0E+00	1.0E+00	1.0E+00
314.2209	315.2281	14.44	1.0E+00	1.9E+06	1.0E+00	1.0E+00	7.9E+05	1.0E+00
316.0344	317.0416	17.74	1.0E+00	1.0E+00	1.0E+00	1.0E+00	2.7E+05	2.7E+05
322.0086	323.0159	13.67	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.5E+05	1.6E+05
324.0241	325.0314	8.56	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.9E+05	1.6E+05
324.0784	347.0661	11.09	(3/6)	5.4E+05	5.5E+05	1.0E+00	1.0E+00	1.3E+05

324.9871	325.9951	12.33	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.8E+05	(4/5)
325.2255	326.2328	16.67	1.0E+00	5.0E+05	1.0E+00	1.0E+00	2.3E+05	1.0E+00
331.2182	332.2256	20.03	1.0E+00	1.0E+00	1.0E+00	1.0E+00	7.2E+04	9.4E+04
340.2366	341.244	15.34	1.0E+00	2.7E+06	1.0E+00	1.0E+00	1.1E+06	1.0E+00
342.2518	343.2592	16.67	1.0E+00	1.0E+08	5.2E+05	1.0E+00	3.9E+07	(4/5)
355.1741	356.1816	15.13	1.0E+00	1.0E+00	9.9E+04	1.0E+00	1.0E+00	1.0E+00
355.2113	378.2012	16.22	1.0E+00	1.0E+00	7.1E+05	1.0E+00	1.0E+00	(2/5)
356.1817	357.1891	16.16	1.0E+00	1.0E+00	3.5E+06	1.0E+00	1.0E+00	1.1E+06
361.2262	362.2334	16.67	1.0E+00	5.6E+05	1.0E+00	1.0E+00	(4/5)	1.0E+00
363.1893	364.1967	16.03	1.0E+00	1.0E+00	8.9E+04	1.0E+00	1.0E+00	5.7E+04
368.2677	369.2748	17.37	1.0E+00	4.6E+05	1.0E+00	1.0E+00	4.4E+05	1.0E+00
370.2826	371.29	18.95	1.0E+00	3.5E+05	1.0E+00	1.0E+00	(3/5)	1.0E+00
374.1918	749.3906	15.73	1.0E+00	1.0E+00	2.9E+05	1.0E+00	1.0E+00	(2/5)
375.1561	376.1628	16.16	1.0E+00	1.0E+00	1.3E+05	1.0E+00	1.0E+00	(2/5)
384.1509	385.1593	16.16	1.0E+00	1.0E+00	2.3E+05	1.0E+00	1.0E+00	1.0E+00
385.2579	386.265	14.13	1.0E+00	1.0E+00	3.4E+05	1.0E+00	1.0E+00	2.9E+05
392.1109	393.1186	6.68	1.0E+00	3.0E+05	3.1E+05	1.8E+05	1.8E+05	(4/5)
398.0791	399.0865	13.21	1.0E+00	1.0E+00	1.0E+00	1.0E+00	(4/5)	1.4E+05
403.0724	404.0798	17.13	1.0E+00	1.0E+00	1.0E+00	1.0E+00	5.5E+04	(2/5)
411.2734	412.2806	15.00	1.0E+00	1.0E+00	1.4E+05	1.0E+00	1.0E+00	1.2E+05
411.3098	434.2991	15.34	1.0E+00	3.4E+05	2.2E+06	1.0E+00	(2/5)	1.9E+05
411.3098	412.3168	17.31	1.0E+00	1.0E+00	6.9E+04	1.0E+00	1.0E+00	(2/5)
413.289	414.296	14.38	1.0E+00	9.3E+04	3.5E+05	1.0E+00	(1/5)	(4/5)
413.2891	414.2964	16.22	1.0E+00	1.7E+05	1.0E+07	1.0E+00	(3/5)	5.0E+06
422.318	423.325	17.55	1.0E+00	1.0E+00	6.7E+04	1.0E+00	1.0E+00	1.0E+00
437.3257	438.3326	16.22	1.0E+00	1.0E+00	2.1E+06	1.0E+00	(2/5)	1.5E+05
439.305	440.3122	15.23	1.0E+00	1.0E+00	2.2E+05	1.0E+00	1.0E+00	(2/5)
439.3411	440.3485	17.55	1.0E+00	1.2E+06	4.4E+07	1.0E+00	2.6E+05	2.8E+06
441.3207	442.328	16.44	1.0E+00	4.0E+05	8.8E+06	1.0E+00	2.6E+05	3.4E+06
443.1872	444.1942	15.83	1.0E+00	1.0E+00	1.0E+05	1.0E+00	1.0E+00	1.0E+00
458.3163	459.3229	17.55	1.0E+00	1.0E+00	1.4E+05	1.0E+00	1.0E+00	1.0E+00
465.3578	488.3479	18.23	1.0E+00	1.0E+00	1.8E+05	1.0E+00	1.0E+00	(1/5)
467.1883	468.196	14.20	1.0E+00	1.0E+00	3.3E+05	1.0E+00	1.0E+00	7.9E+04
467.3719	490.3618	18.47	1.0E+00	1.0E+00	3.4E+05	1.0E+00	1.0E+00	1.0E+00
467.6902	468.6972	14.20	1.0E+00	1.0E+00	1.4E+05	1.0E+00	1.0E+00	(1/5)
479.2999	480.3069	14.52	1.0E+00	1.0E+00	1.2E+05	1.0E+00	(1/5)	(2/5)
483.3312	484.3388	16.53	1.0E+00	8.7E+04	1.0E+00	1.0E+00	(3/5)	3.0E+05
485.3466	486.3538	15.71	1.0E+00	1.0E+00	2.5E+05	1.0E+00	1.0E+00	(1/5)
497.3943	498.4015	17.55	1.0E+00	1.0E+00	1.3E+05	1.0E+00	1.0E+00	1.0E+00
501.3105	524.3002	17.56	1.0E+00	1.0E+00	1.0E+05	1.0E+00	1.0E+00	1.0E+00

505.3161	506.3224	15.34	1.0E+00	1.0E+00	7.1E+04	1.0E+00	(1/5)	1.0E+00
507.331	508.3382	16.49	1.0E+00	1.0E+00	(3/5)	1.0E+00	(2/5)	1.1E+05
511.3621	534.3514	18.68	1.0E+00	1.0E+00	2.4E+05	1.0E+00	1.0E+00	1.0E+00
518.3473	519.3539	15.26	1.0E+00	1.0E+00	9.0E+04	1.0E+00	1.0E+00	(2/5)
519.0621	520.0697	1.72	1.0E+00	1.4E+05	1.7E+05	1.0E+00	1.0E+00	1.0E+00
521.3465	522.3538	17.76	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.1E+06
532.3516	533.3591	16.67	1.0E+00	2.6E+05	1.0E+00	1.0E+00	(3/5)	1.0E+00
533.1277	534.1357	8.57	1.0E+00	(4/5)	(3/5)	1.0E+00	1.0E+00	5.2E+04
539.4042	540.4113	17.55	1.0E+00	1.0E+00	1.1E+05	1.0E+00	1.0E+00	1.0E+00
544.3624	545.3696	16.08	1.0E+00	1.0E+00	9.5E+04	1.0E+00	1.0E+00	(1/5)
546.3791	547.3859	17.29	1.0E+00	1.0E+00	2.5E+06	1.0E+00	1.0E+00	7.0E+05
552.3998	553.4061	27.07	1.0E+00	6.3E+04	1.0E+00	1.0E+00	1.0E+00	1.0E+00
552.7239	553.7311	15.02	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.3E+05
562.3724	585.3618	15.53	1.0E+00	1.0E+00	1.1E+05	1.0E+00	1.0E+00	(1/5)
564.3892	587.3783	16.72	1.0E+00	1.0E+00	4.6E+06	1.0E+00	1.0E+00	1.3E+06
566.3571	567.3644	16.10	1.0E+00	1.0E+00	9.2E+04	1.0E+00	1.0E+00	1.0E+00
566.3681	589.3571	16.53	1.0E+00	1.0E+00	1.4E+05	1.0E+00	1.0E+00	(1/5)
572.3576	573.3646	17.07	1.0E+00	1.0E+00	3.7E+05	1.0E+00	1.0E+00	1.1E+05
576.3885	599.3781	17.26	1.0E+00	1.0E+00	1.1E+05	1.0E+00	1.0E+00	1.0E+00
608.3787	609.3862	16.25	1.0E+00	1.0E+00	2.8E+06	1.0E+00	1.0E+00	1.2E+06
632.439	633.4459	27.67	1.0E+00	1.4E+05	(4/5)	1.0E+00	1.0E+00	1.0E+00
636.4109	659.3995	17.90	1.0E+00	1.0E+00	1.3E+05	1.0E+00	1.0E+00	1.0E+00
647.2357	648.2436	20.09	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	9.3E+04
664.9382	665.9456	12.32	1.0E+00	1.0E+00	1.0E+00	1.0E+00	(4/5)	4.9E+05
676.4651	677.4715	27.46	1.0E+00	1.2E+05	(1/5)	1.0E+00	1.0E+00	1.0E+00
728.4683	729.4747	16.67	1.0E+00	5.8E+04	1.0E+00	1.0E+00	1.0E+00	1.0E+00
738.4219	739.4293	16.67	1.0E+00	1.4E+05	1.0E+00	1.0E+00	(2/5)	1.0E+00
815.3701	816.378	16.60	1.0E+00	1.0E+00	3.5E+05	1.0E+00	1.0E+00	1.1E+06

[Δ]Raw abundance is the average of the five replicates when data was available. If a specific mass was observed in <5 replicates, its presence is denoted by a fraction depicting the number of samples it was observed over the total number of samples (example, if observed in only three samples, data recorded as 3/5).



1

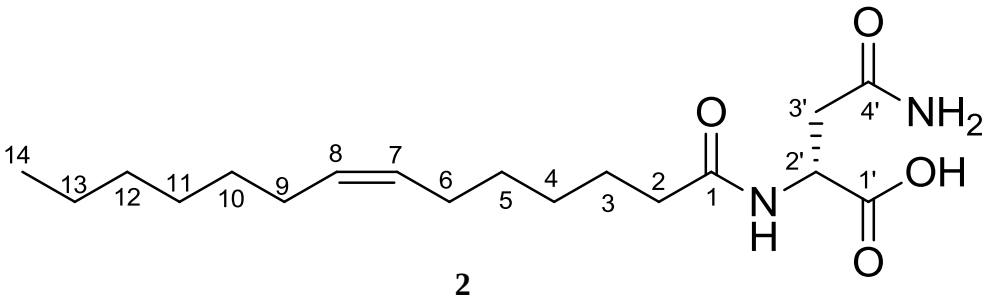
Table S3. ^{1}H and ^{13}C NMR chemical shifts of isolated metabolite **1** in DMSO-*d6*

	C	δC[±] (mult.)	δH	mult (*J* in Hz)	gCOSY	gHMBC
Myristoyl	1	172.60	-	-	-	-
	2	35.80	2.07	t (7.4)	3	1, 3
	3	25.76	1.45	qu (6.4)	2, 4	4-12
	4-12[ǂ]	28- 32	1.23	m	3	4-12
	13	22.48	1.23	br	14	14, 12
	14	14.39	0.85	t (6.8)	13	13, 4-12
D-Asn	1'	175.20	-	-	-	-
	2'	48.60	4.45	q (6.7)	3', 1-NH	1’
	3'	37.88	2.40, 2.48	m	2'	2’, 4’
	4'	171.80	-	-	-	-
	1-NH	-	7.93	d (7.8)	2'	1
	4'-NH_2	-	7.44, 6.86	s	-	3', 4’
	OH	-	-	-	-	-

[±]Determined from gHSQC/gHMBC NMR data

[ǂ]Carbon atoms from fatty acid; carbon and proton signals were not distinguishable from each other

*Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q= quartet, qu = quintet, m = multiplet; br = broad



2

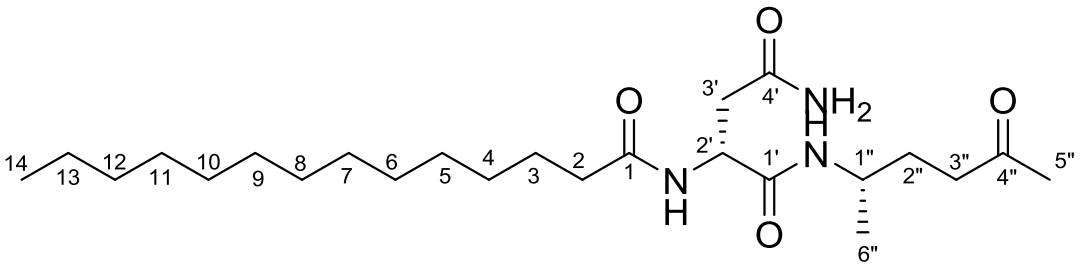
Table S4. ^{1}H and ^{13}C NMR chemical shifts of isolated metabolite **2** in DMSO-*d6*

	C	δC[±] (mult.)	δH	mult (*J* in Hz)	gCOSY	gHMBC
cis-7-tetradecenoyl	1	171.27	-	-	-	-
	2	37.05	2.06	dd (15.58)	3	1, 2, 4-5
	3	25.99	1.45	m	4-5, 2	1, 2, 4-5
	4-5[‡]	28 – 29	1.26 - 1.29	m	4-5, 6	4-5, 6, 7-8
	6	27.51	1.98	m	4-5, 7	7-8, 4-5
	7-8	130.30	5.32	td (7.29)	7-8	4-5, 10-12
	9	27.51	1.98	m	10-12, 8	7-8, 10-12
	10-12[‡]	28 – 29	1.26 - 1.29	m	10-12, 9	9, 10-12, 13
	13	31.79	1.25	m	14	14, 10-12
	14	14.72	0.85	t (6.96)	13	13
D-Asn	1'	173.53	-	-	-	-
	2'	51.73	3.93	m	1-NH, 3'	1', 3'
	3'	41.34	2.38	m	2'	2', 1', 4'
			2.20	dd (15.47)	2'	2', 1', 4'
	4'	173.37	-	-	-	-
	1-NH	-	7.29	d (6.81)	2'	1
	OH	-	-	-	-	-
	4'-NH_2	-	8.55, 6.63	s	-	-

[±]Determined from gHSQC/gHMBC NMR data

[‡]Carbon atoms from fatty acid; carbon and proton signals were not distinguishable from each other

*Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, m = multiplet.



3

Table S5. ^{1}H and ^{13}C NMR chemical shifts of isolated metabolite **3** in DMSO-*d6*

	C	δC[±] (mult.)	δH	mult (*J* in Hz)	gCOSY	gHMBC
Myristoyl	1	172.60	-	-	-	-
	2	35.80	2.07	td (7.08)	3	3, 4-11, 1
	3	25.70	1.44	m	2, 4-11	4-11
	4-11[‡]	29-30	1.21	m	3	2, 3, 4-11
	12	31.80	1.22	m	-	4-11, 13
	13	22.60	1.24	m	14	12, 14
	14	14.70	0.84	t (7.03)	13	12, 13
D-Asn	1'	170.90	-	-	-	-
	2'	50.70	4.41	m	3', 1-NH	3', 1'
	3'	38.00	2.39, 2.34	m	2'	2', 4'
	4'	171.80	-	-	-	-
	1-NH	-	7.93	s	2'	1, 2'
	4'-NH_2	-	7.25, 6.82	-	-	4'
5-amino-2-hexanone	1"	44.25	3.66	m	6", 1'-NH	-
	2"	30.39	1.56, 1.44	m	1", 3"	3", 6", 4"
	3"	39.88	2.39	m	2"	1", 5", 4"
	4"	208.58	-	-	-	-
	5"	30.67	2.01	s	-	3", 4"
	6"	21.50	0.98	d (6.48)	1"	2", 1"
	1'-NH	-	7.44	s	1"	1', 1"

[±]Determined from gHSQC/gHMBC NMR data

[‡]Carbon atoms from fatty acid; carbon and proton signals were not distinguishable from each other

*Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, m = multiplet.

Table S6. LC-MS data for the *in vivo* peptidase activity of arabinose-inducible ClbP with synthetic **3**. EIC positive peaks correspond to starting material (**3**, *m/z* 440) and the cleavage product *N*-myristoyl-D-Asn (**1**, *m/z* 343) at two different time points (4 h and 7 h) using two different *E. coli* strains, BL21 and DH10B**.** pBAD18 is the empty plasmid backbone. "-" indicates that no quantifiable peak was detected.

***E. coli* strain**	**Time point**	**EIC peak area *m/z* 440**	**EIC peak area *m/z* 343**
BL21 + pBAD18	4 h	1.5E+05	-
BL21 + pClbP	4 h	1.2E+05	1.1E+05
DH10B + pBAD18	4 h	1.3E+05	7.9E+03
DH10B + pClbP	4 h	9.5E+04	1.6E+05
BL21 + pBAD18	7 h	1.6E+05	-
BL21 + pClbP	7 h	8.8E+04	8.7E+04
DH10B + pBAD18	7 h	7.7E+04	-
DH10B + pClbP	7 h	4.3E+04	9.6E+04

Table S7. Psychoactive Drug Screening Program (PDSP) activity summary[Δ]

Compound	**5-HT$_{2C}$**		**5-HT7**		**D5**		**DAT**		**M3**	
	1° Assay (10μM)	**2° Assay Ki (nM)**	**1° Assay (10μM)**	**2° Assay Ki (nM)**	**1° Assay (10μM)**	**2° Assay Ki (nM)**	**1° Assay (10μM)**	**2° Assay Ki (nM)**	**1° Assay (10μM)**	**2° Assay Ki (nM)**
1	13.6		57.5	>10,000	73.5	>10,000	45.5		-15.1	
12	86.5	936	13.7		19		66.1	4,319	-13.1	
3	30.1		2.6		14.5		31.3		13.7	
11	-1		-15.4		10.1		33		-0.5	
Tuamine	4.4		0.8		20.5		29.1		50.3	>10,000

[Δ]5-HT$_{2C}$, 5-hydroxytryptamine$_{2C}$ receptor; 5-HT7, 5-hydroxytryptamine 7 transporter; D5, dopamine 5 transporter; DAT, dopamine transporter; M3, muscarinic acetylcholine receptor M3. Activities below the PDSP significance threshold are not shown. Significant activities are highlighted.

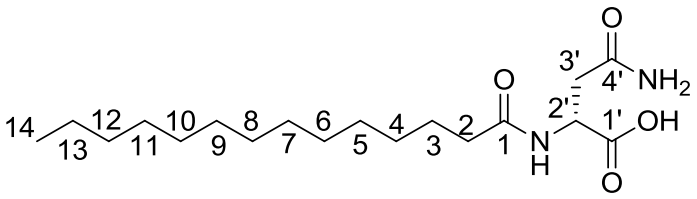


Table S8. ^{1}H and ^{13}C NMR chemical shifts of synthetic (red) vs natural (blue) **1** in DMSO-*d6* (See accompanying Figures S28 and S29).

		Synthetic		Natural	
	C	δC (mult.)	δH	δC (mult.)	δH
Myristoyl	1	171.97	-	172.60	-
	2	35.09	2.07	35.80	2.07
	3	25.23	1.45	25.76	1.45
	4 to 12	28 to 32	1.24	28 to 32	1.23
	13	22.1	-	22.48	1.23
	14	13.95	0.85	14.39	0.85
D-Asn	1'	172.99	-	175.20	-
	2'	48.72	4.46	48.60	4.45
	3'	36.83	2.41, 2.54	37.88	2.40, 2.48
	4'	171.24	-	171.80	-
	NH	-	7.94	-	7.93
	NH_2	-	7.33, 6.87	-	7.44, 6.86
	OH	-	12.51	-	-

[Δ] Carbon shifts determined from ^{13}C NMR spectrum for synthetic compound
[ǂ] Carbon shifts determined from gHSQC and gHMBC for natural compound

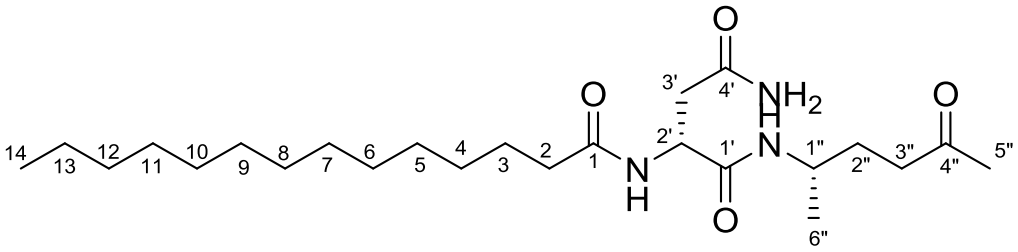


Table S9. ^{1}H and ^{13}C NMR chemical shifts of synthetic (red) vs natural (blue) **3** in DMSO-*d6* (See accompanying Figures S30 and S31).

		Synthetic		Natural	
	C	δC^{Δ}	δH	$\delta C^{\ddagger}$	δH
Myristoyl	1	172.51	-	172.60	-
	2	35.66	2.08	35.80	2.07
	3	25.66	1.46	25.70	1.44
	4-11	29-31	1.24	29-30	1.21
	12	29-31	1.24	31.80	1.22
	13	22.54	1.24	22.60	1.24
	14	14.39	0.85	14.70	0.84
D-Asn	1'	170.85	-	170.90	-
	2'	50.34	4.44	50.70	4.41
	3'	37.74	2.40	38.00	2.39, 2.34
	4'	171.85	-	171.80	-
	NH	-	7.89	-	7.93
	NH_2	-	7.25, 6.86	-	7.25, 6.82
5-amino-2-hexanone	1"	44.03	3.68	44.25	3.66
	2"	30.38	1.57, 1.46	30.39	1.56, 1.44
	3"	37.96	2.40	39.88	2.39
	4"	208.52	-	208.58	-
	5"	31.73	2.03	30.67	2.01
	6"	21.10	0.98	21.50	0.98
	NH	-	7.45	-	7.44

$^{\Delta}$Carbon shifts determined from ^{13}C NMR spectrum for synthetic compound
‡Carbon shifts determined from HSQC and HMBC for natural compound

3. Supplemental Figures

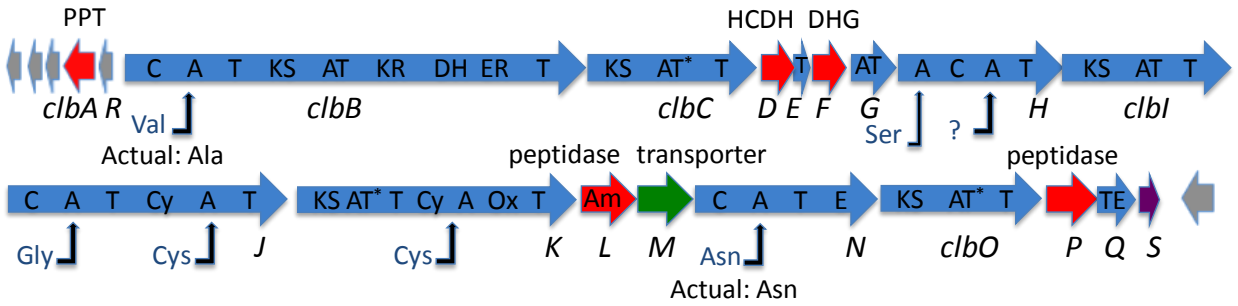


Figure S1. Revised domain architecture of the colibactin (*clb*) gene cluster. Genes encoding hybrid polyketide synthase (PKS)-nonribosomal peptide synthetase (NRPS) enzymes are shown in blue, genes encoding tailoring and accessory enzymes are shown in red, the putative transporter-encoding gene *clbM* is shown in green, the expressed gene *clbS* encoding a protein with a currently undefined role is shown in purple, and peripheral transposase/integrase-encoding genes or gene remnants are shown in grey. Predicted amino acid A-domain specificities are depicted using the following abbreviations: C, condensation; A, adenylation; T, thiolation sequence of acyl- and peptidyl-carrier proteins; KS, ketosynthase; AT, acyl-transferase; KR, ketoreductase; DH, dehydratase; ER, enoyl-reductase; HCDH, hydroxyacyl-CoA dehydrogenase; DHG, dehydrogenase; Cy, cyclase; Ox, oxidase; Am, amidase; E, epimerase; TE, thioesterase; PPT, phosphopantetheinyl-transferase; *, protein fold predicted via structural topology. Structural topology predictions suggest that AT* domains are inactivated AT ancestral relics. We suggest that this feature supports that the "trans-AT" proteins in the *clb* pathway could be evolutionary products of currently undefined "cis-AT" ancestors, which is in contrast to current evolutionary models for other "trans-AT" polyketide synthases.[9] Additionally, a DH domain can be detected in ClbB, which was missed in the original domain analysis.[3] Structural characterization of **3** here also supports a functional DH domain in ClbB.

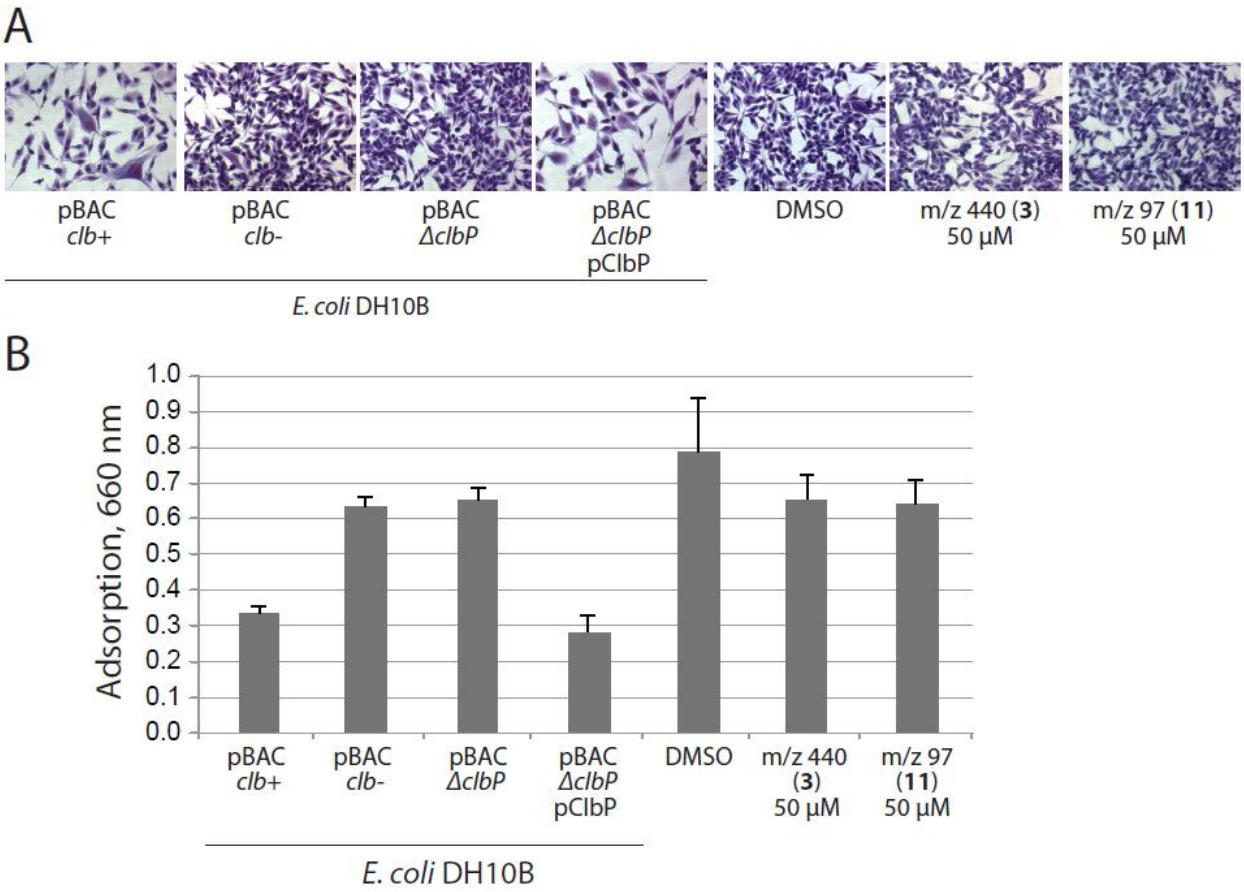


Figure S2. Megalocytosis activity of *E. coli* DH10B harboring different pBAC constructs as well as synthesized compounds. (A) Giemsa stained HeLa cells 48h after exposure to different *E. coli* strains or the synthesized compounds **3** and **11**. Megalocytosis of HeLa cells is only observed when cells have been exposed to pBAC *clb+* or when pBAC *ΔclbP* was complemented with pClbP (non-induced). No activity was found when HeLa cells were exposed to the synthesized compounds **3** and **11**. Images were taken using bright-field microscopy (20x magnification). (B) Quantification of megalocytosis activity based on protein content per well using methylene blue staining followed by methylene blue extraction and OD_{660} measurements. For each condition, three independent wells were quantified and the average and standard deviation are shown.

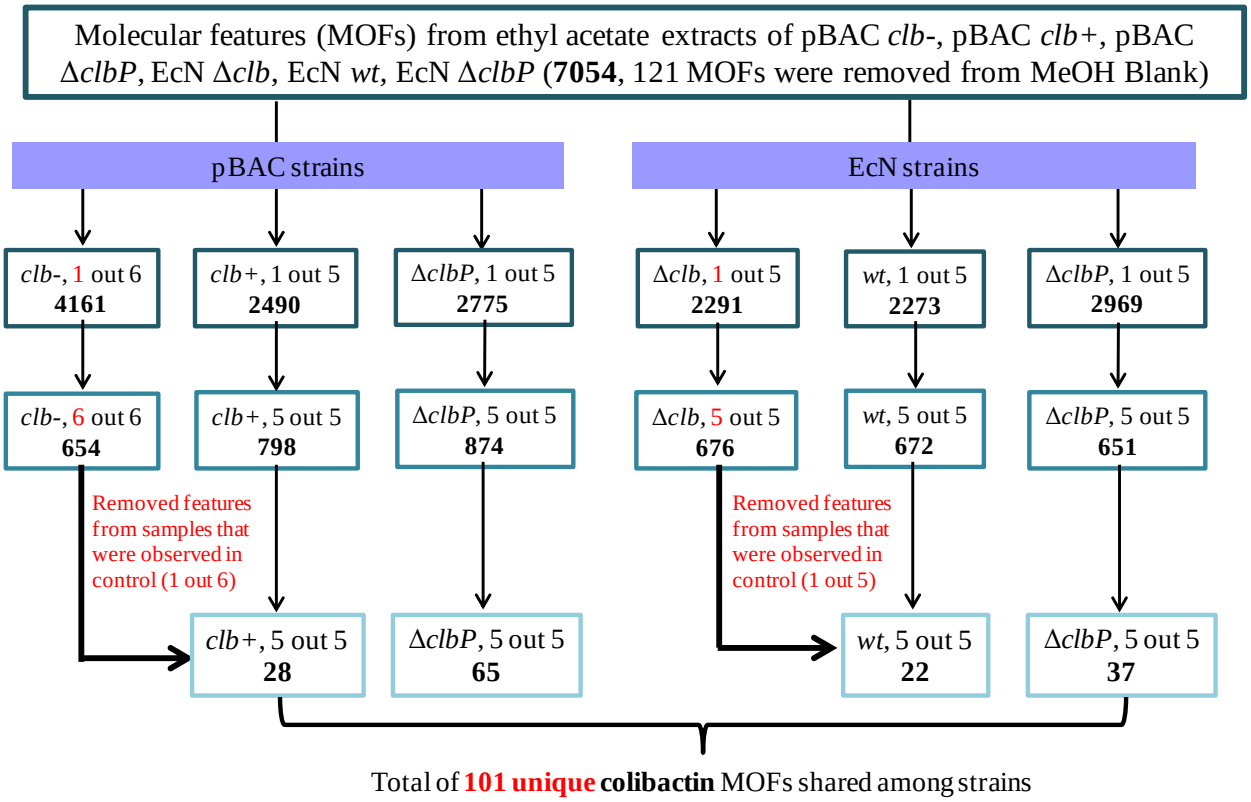


Figure S3. Systematic approach for conservatively identifying *clb*-specific MOFs. Mass Profiler Professional (Agilent Technologies) was used to extract statistically significant molecular features (MOFs) from each sample. Each interpretation (the 5 replicates were used as one interpretation as dictated by rectangle) was "Filtered by Abundance" based on Raw Data (signal intensity values) with 1.0 lower cut-off. MOFs present in any of the control samples were maintained in the control interpretation, while only those MOFs found in all five *clb*-containing samples were used in the sample interpretations. Any MOFs found in control were removed from sample interpretations, which provided a set of *clb*-pathway dependent MOFs. We noted that small molecule complexation sometimes led to separate MOFs for the same molecule, but ultimately, conservative feature extraction led to a much smaller number of unique MOFs dependent on the *clb* pathway.

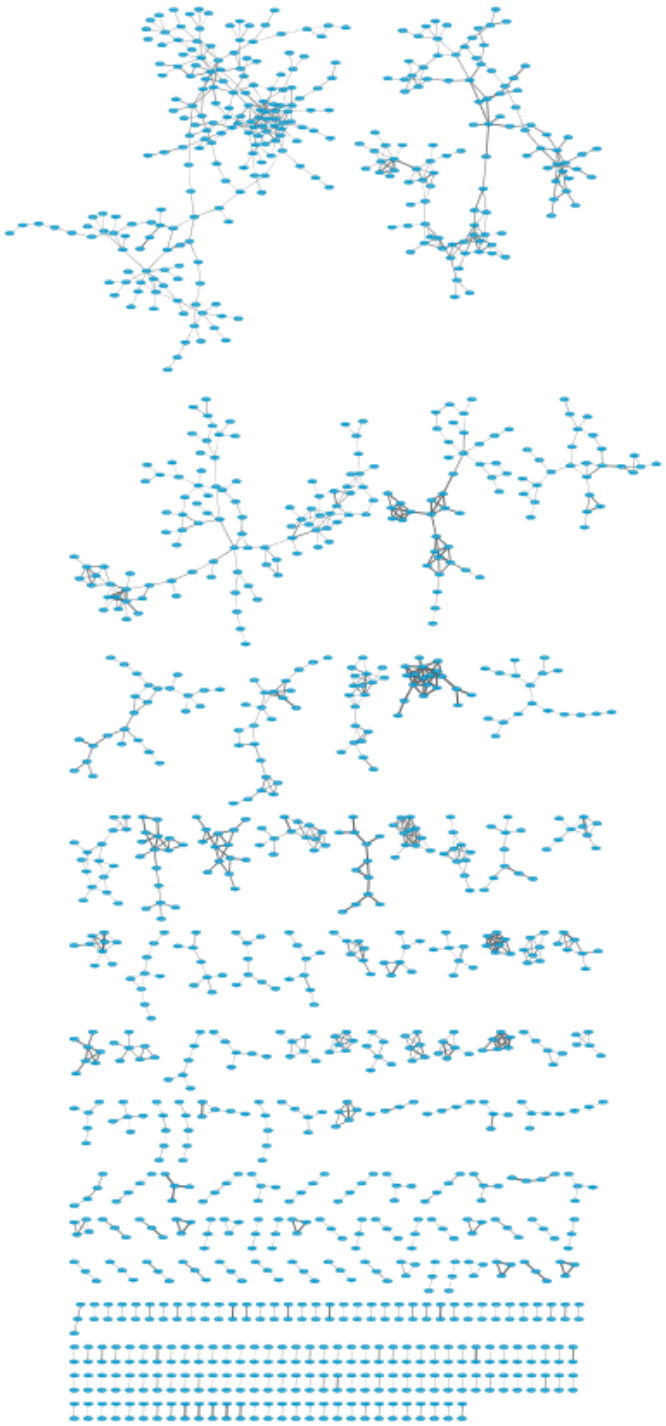


Figure S4. Untargeted MS^2 metabolomics network. While we noted fragmentation of abundant *clb* pathway-dependent metabolites in the raw data (signal intensity values), there was much less network coverage of pathway-dependent ions when using untargeted analysis. Consequently we switched to a targeted network analysis in this study.

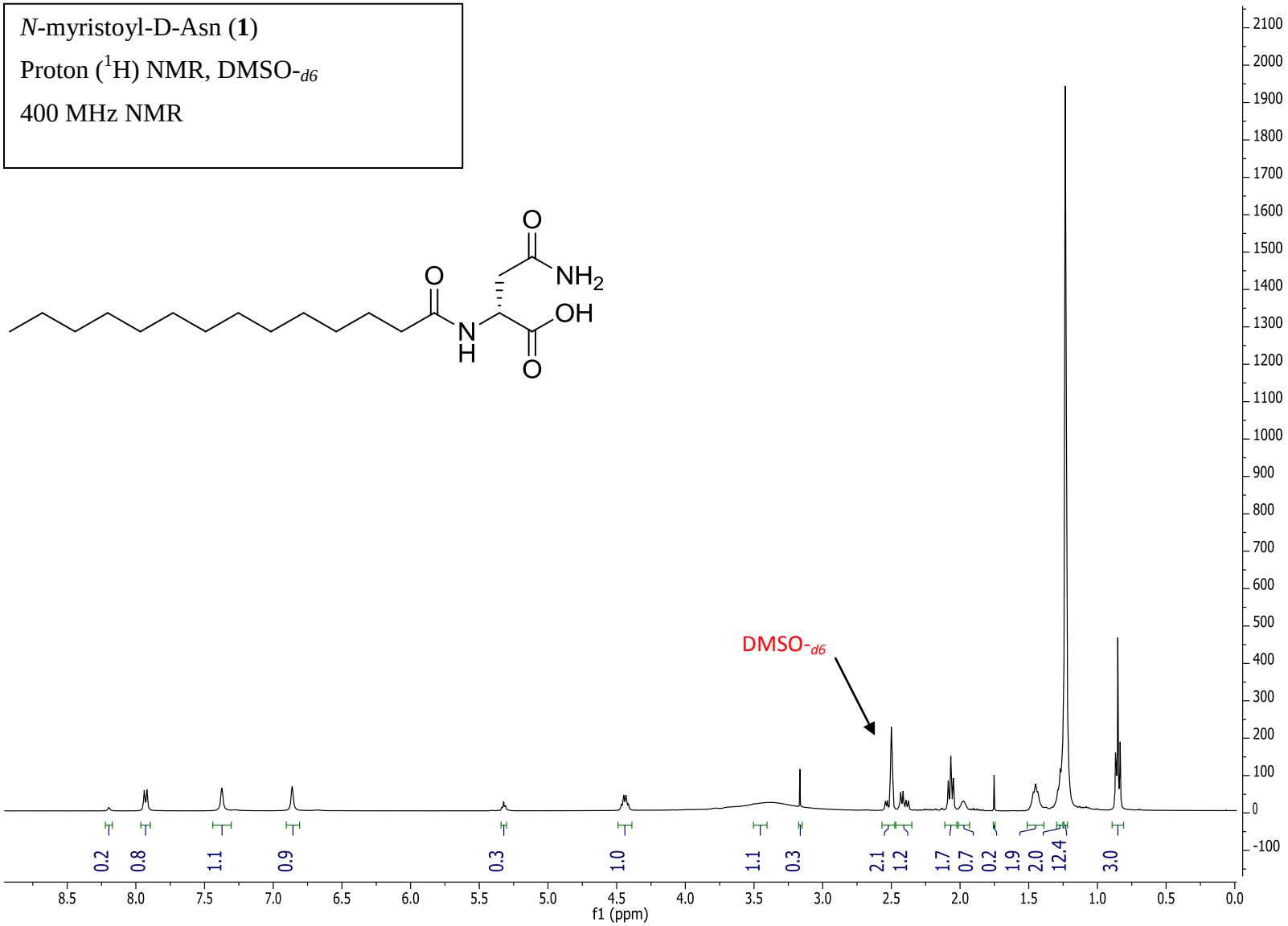


Figure S5. ^{1}H NMR of isolated *N*-myristoyl-D-Asn (**1**). Recorded with 400 MHz NMR in DMSO-*d6*.

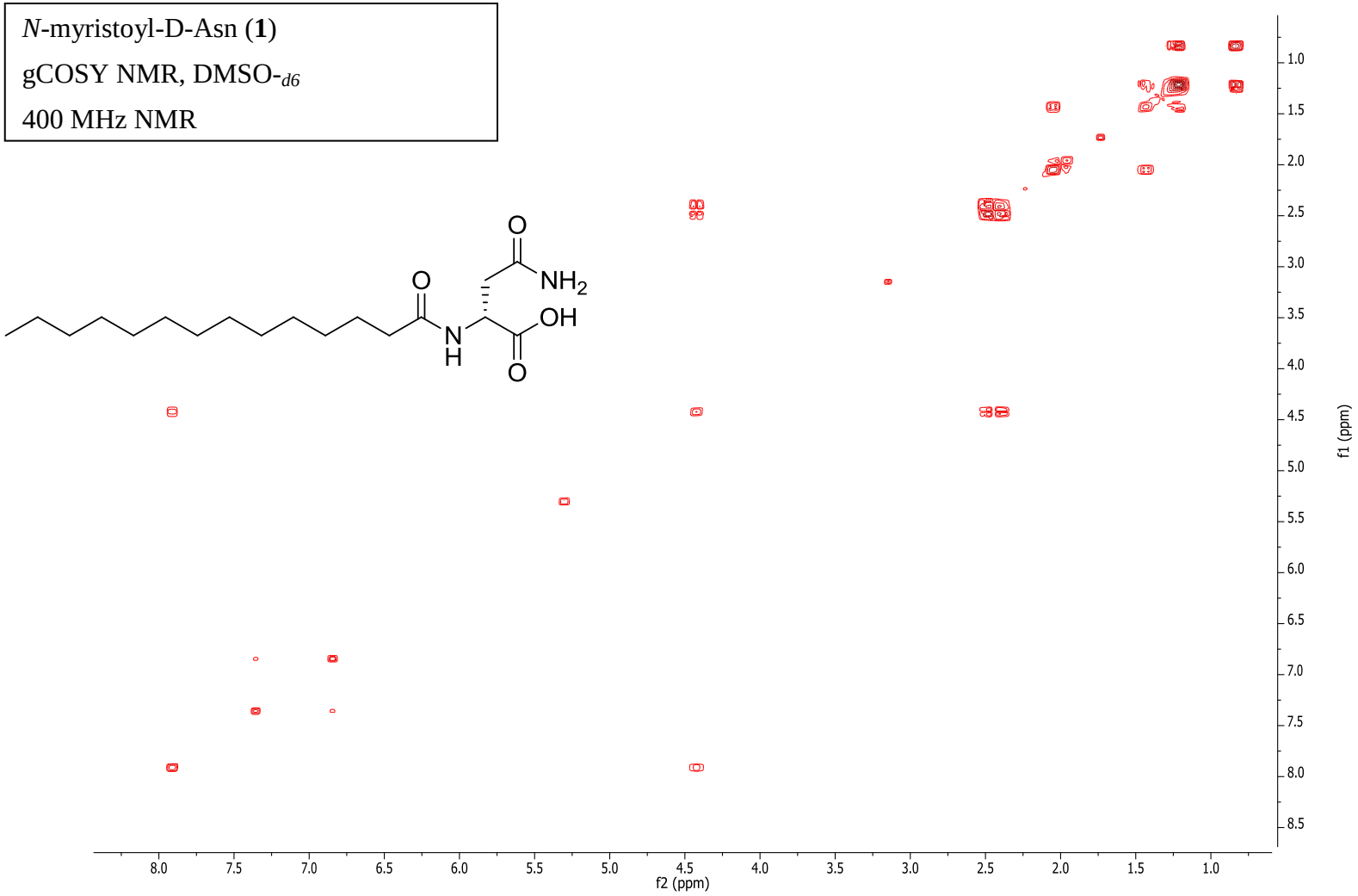


Figure S6. gCOSY NMR of isolated *N*-myristoyl-D-Asn (**1**)**.** Recorded with 400 MHz NMR in DMSO-*d6*.

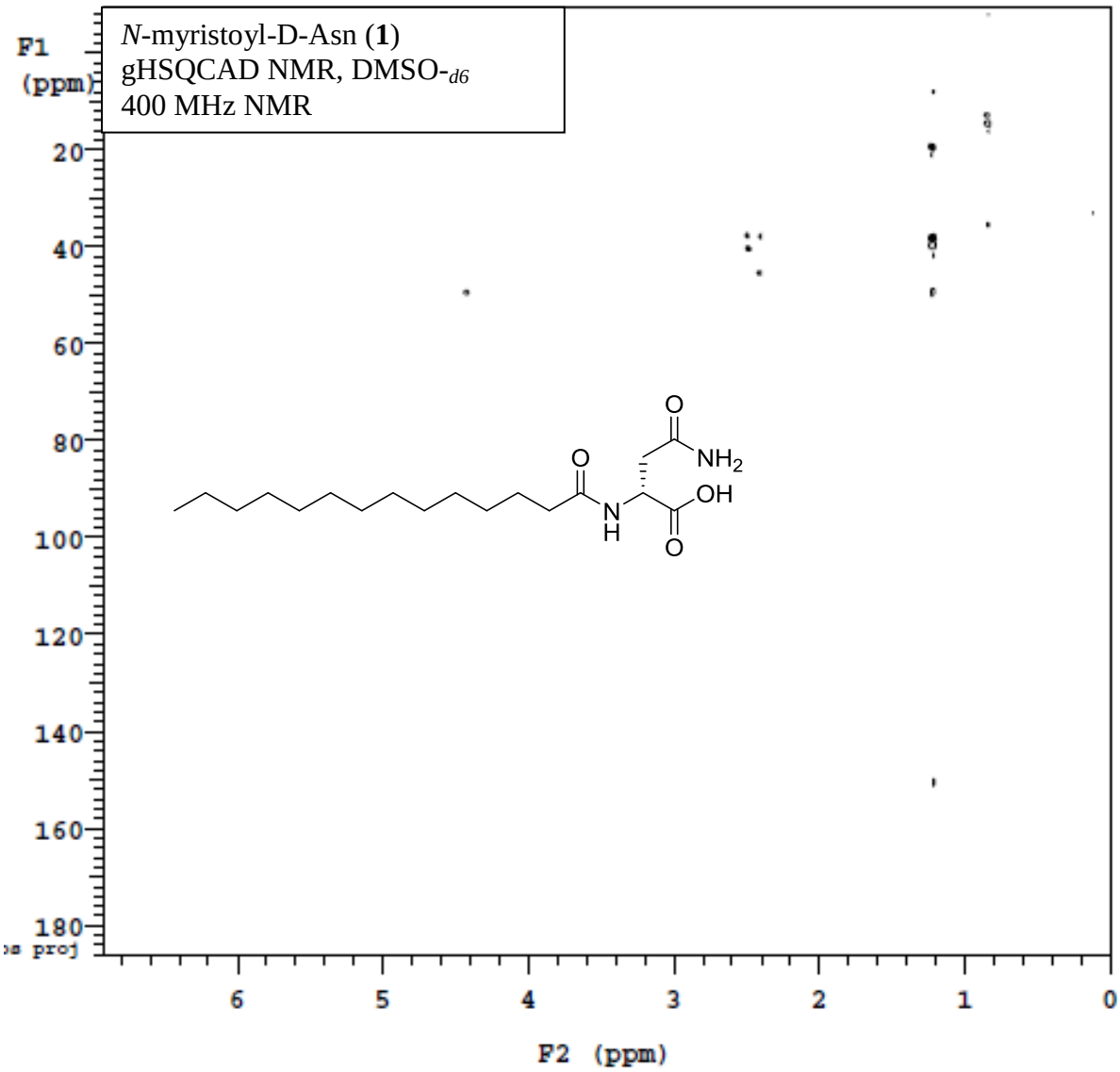


Figure S7. gHSQCAD NMR of isolated *N*-myristoyl-D-Asn (**1).** Recorded with 400 MHz NMR in DMSO-*d6*.

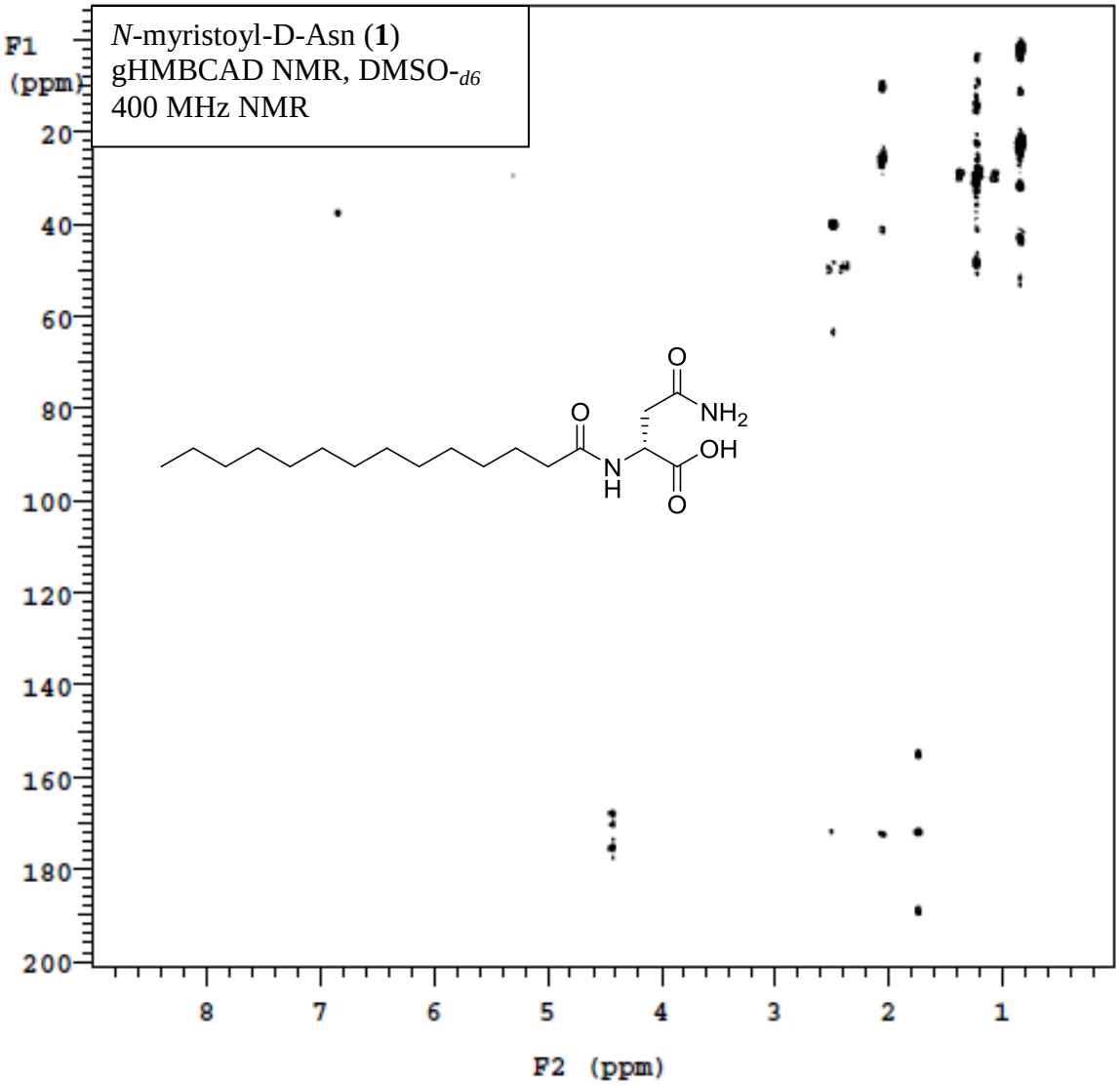
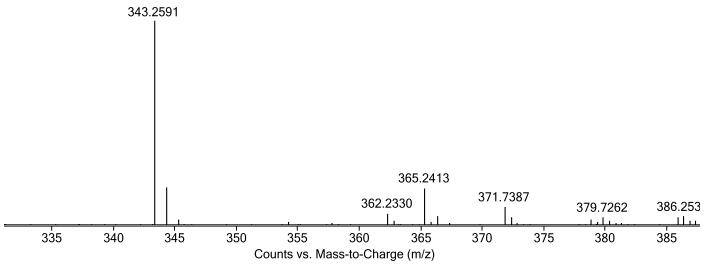


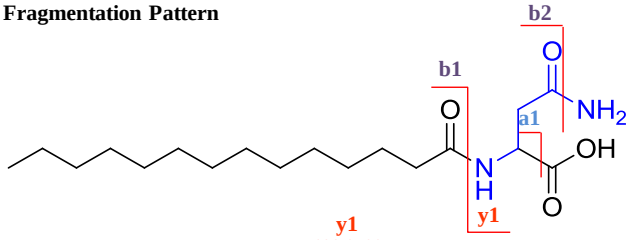
Figure S8. gHMBCAD NMR of isolated *N*-myristoyl-D-Asn (**1**)**.** Recorded with 400 MHz NMR in DMSO-*d6*.

ESI-QTOF-HRMS



[M+H]+
$C_{18}H_{35}N_2O_4$
Obs: 343.2591
Calc: 343.2597
Error (ppm):-1.75

MS^2 Fragmentation Pattern



Species	Obs. Mass	Calc mass	Error (ppm)
a1	87.0550	87.0553	-3.4
b1	211.2047	211.2056	-4.3
b2	326.2301	326.2326	-7.7
y1	133.0599	133.0608	-6.8

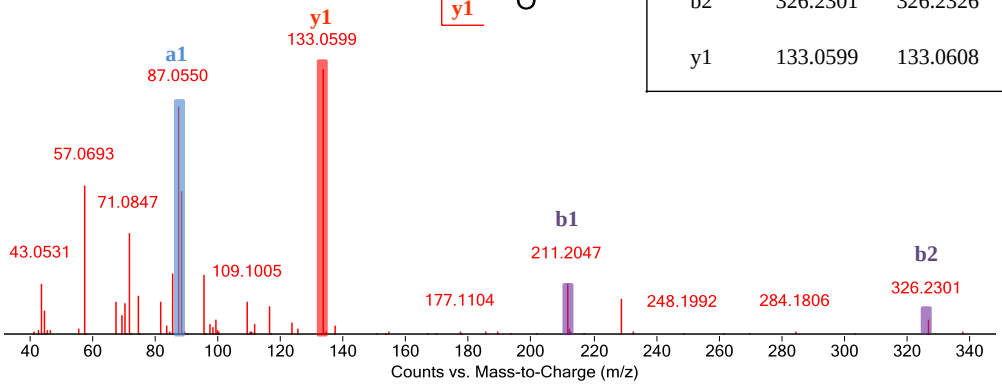
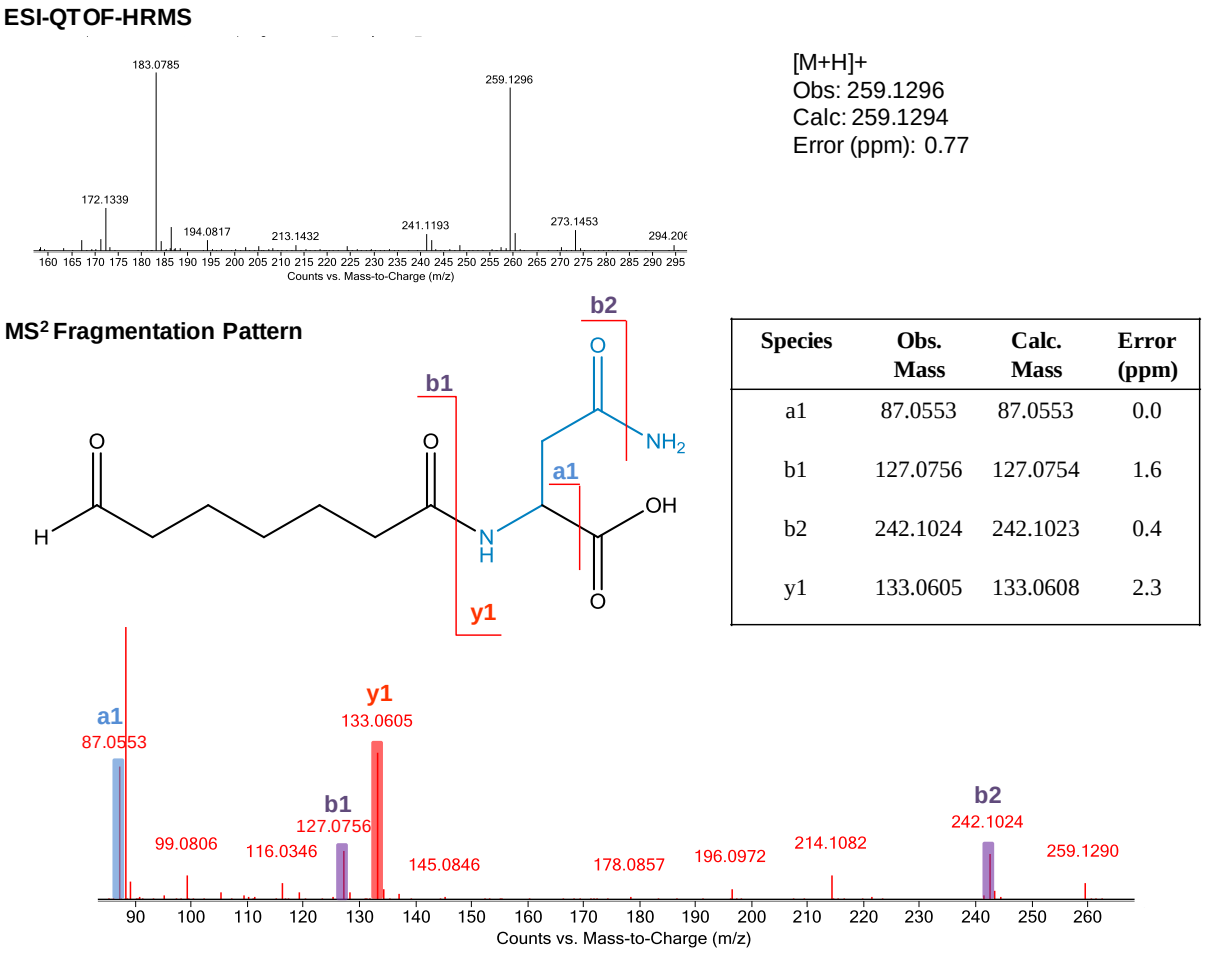


Figure S9. ESI-QTOF-HRMS analysis (top) and MS^2 fragmentation pattern (bottom) of isolated *N*-myristoyl-D-Asn (**1**).



Species	Obs. Mass	Calc. Mass	Error (ppm)
a1	87.0553	87.0553	0.0
b1	127.0756	127.0754	1.6
b2	242.1024	242.1023	0.4
y1	133.0605	133.0608	2.3

Figure S10. Ozonolysis product confirmation of isolated cis-7-tetradecenoyl-D-Asn (**2**) by ESI-QTOF-HRMS and MS^2 fragmentation pattern.

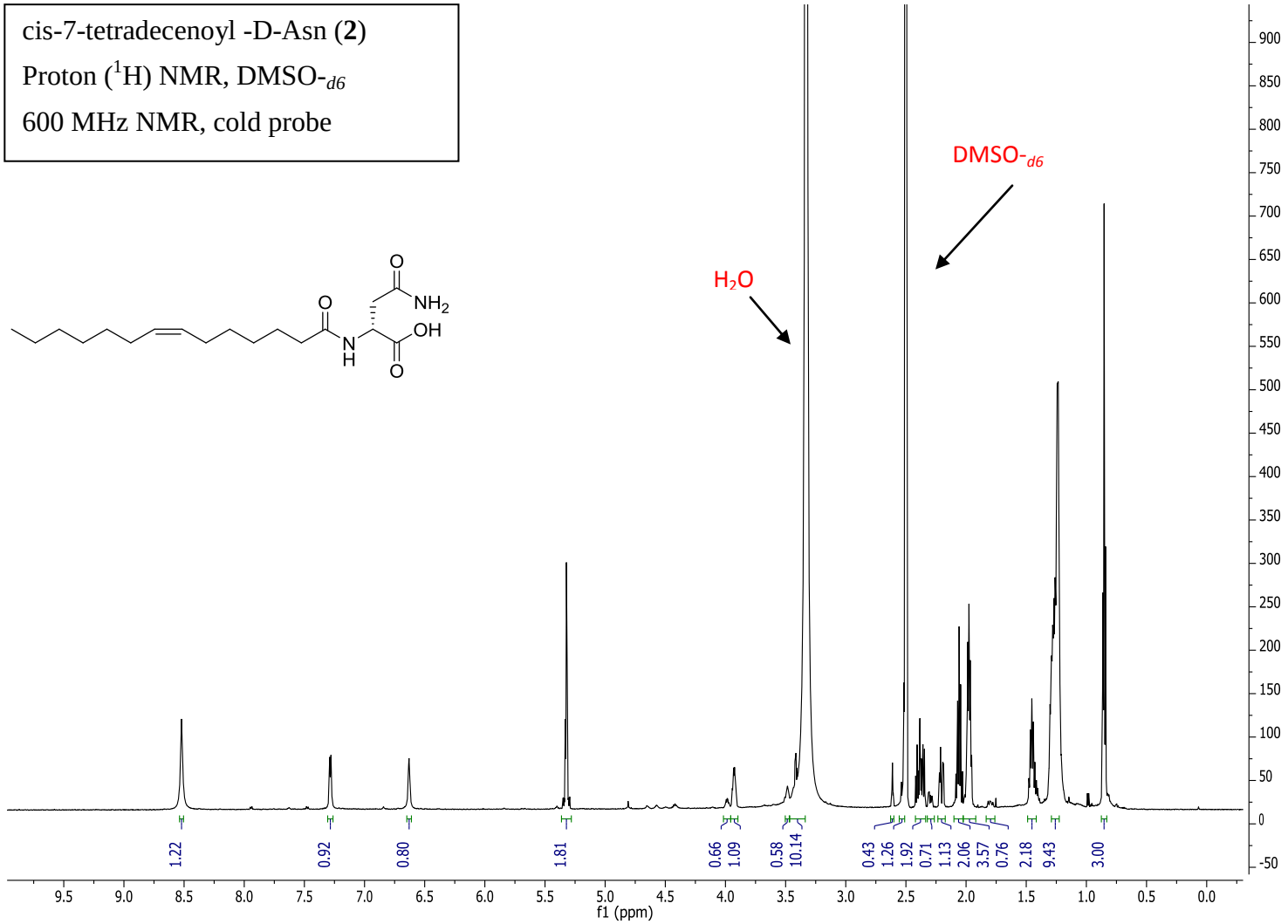


Figure S11. ^{1}H NMR of isolated cis-7-tetradecenoyl-D-Asn (**2**). Recorded with 600 MHz NMR, cold probe, in DMSO-*d6*.

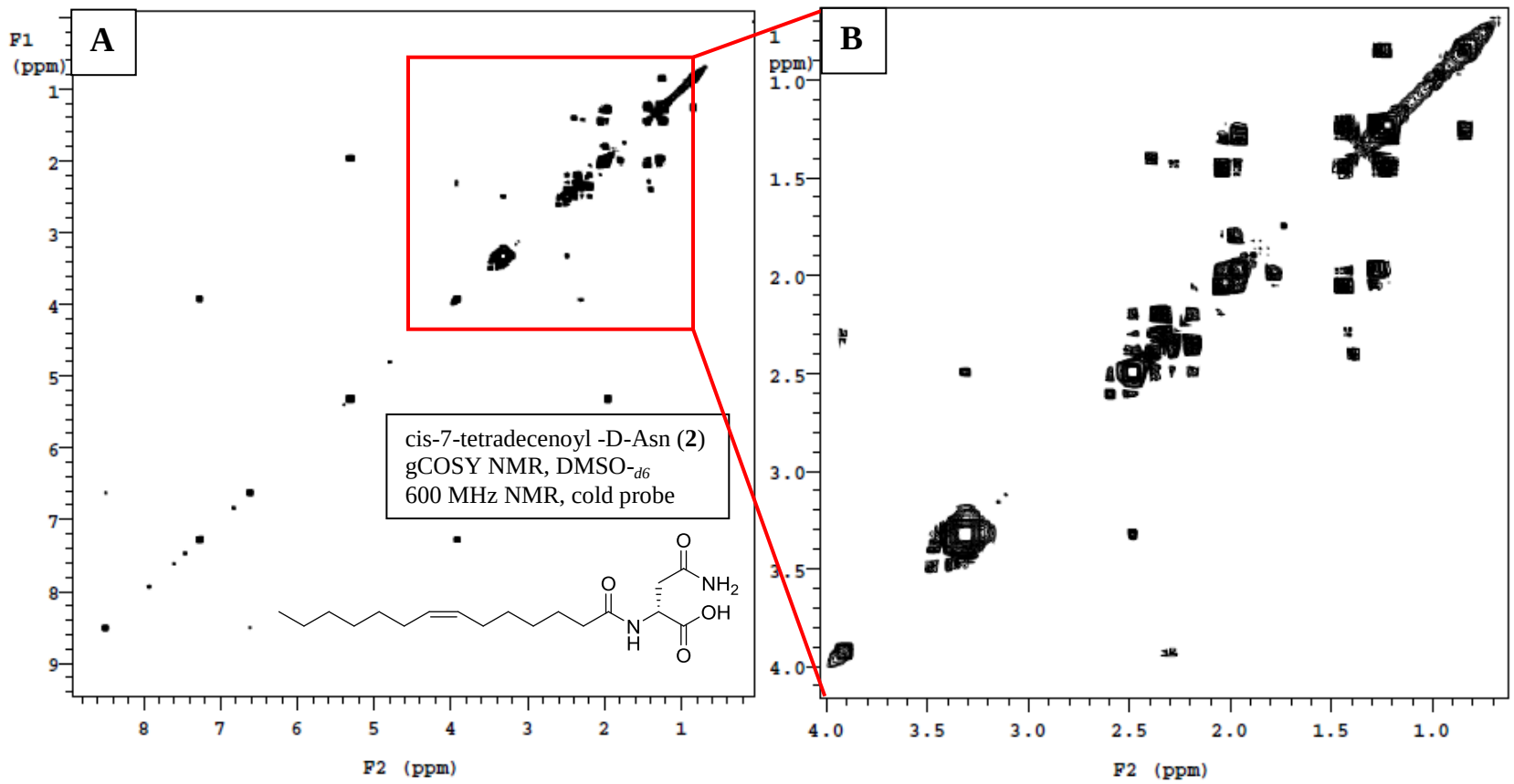


Figure S12. gCOSY NMR of isolated cis-7-tetradecenoyl-D-Asn (**2**). Full spectrum is shown in (A), and close-up of section highlighted (red square) is shown in (B). Recorded with 600 MHz NMR, cold probe, in DMSO-$_{d6}$.

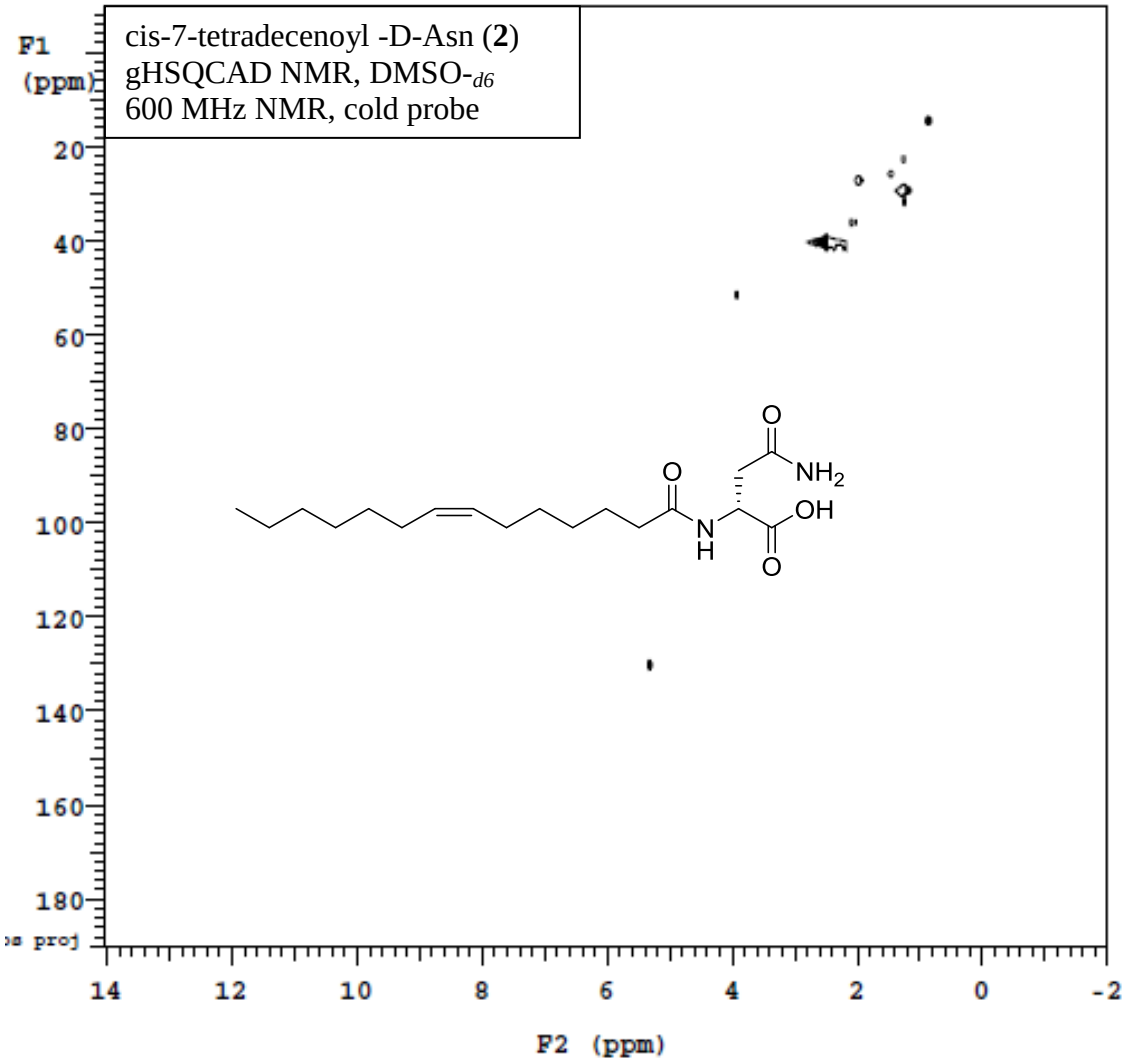


Figure S13. gHSQCAD NMR of isolated cis-7-tetradecenoyl-D-Asn (**2**). Recorded with 600 MHz NMR, cold probe, in DMSO-*d6*.

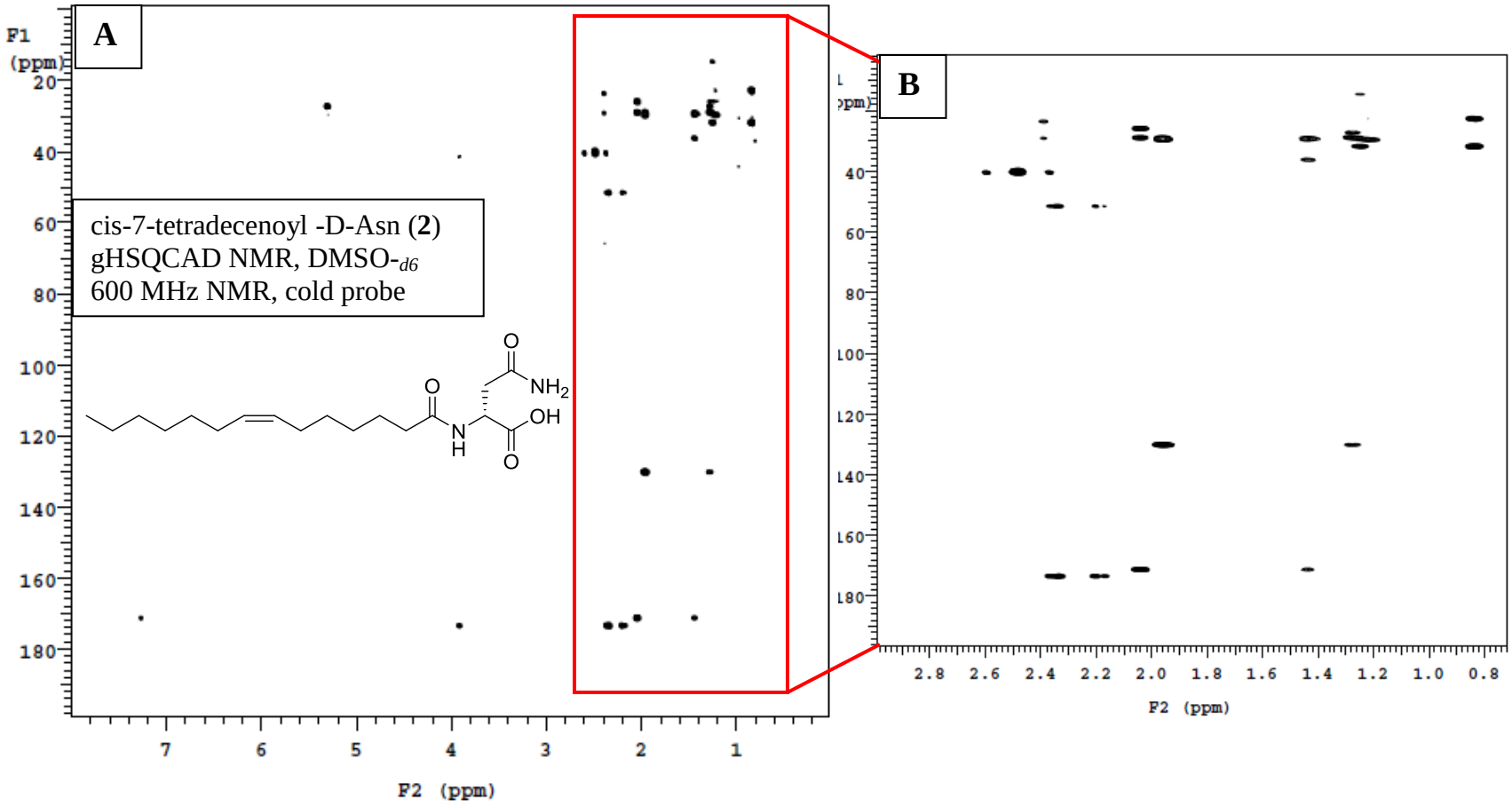
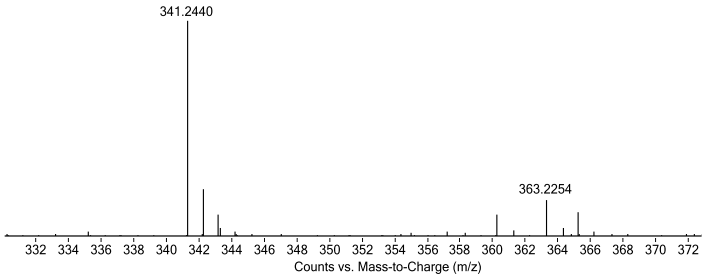


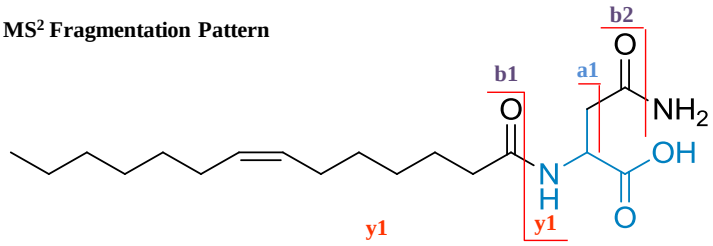
Figure S14. gHMBCAD NMR of isolated cis-7-tetradecenoyl-D-Asn (**2**). Full spectrum is shown in (A), and expanded highlighted region (red square) is shown in (B). Recorded with 600 MHz NMR, cold probe, in DMSO-*d6*.

ESI-QTOF-HRMS



[M+H]+
$C_{18}H_{32}N_2O_4$
Obs: 341.2440
Calc: 341.2440
Error (ppm): 0.0

MS^2 Fragmentation Pattern



Species	Obs. Mass	Calc mass	Error (ppm)
a1	88.0389	88.0393	4.5
b1	209.1906	209.1900	2.9
b2	324.2172	324.2169	0.9
y1	133.0604	133.0608	3.0

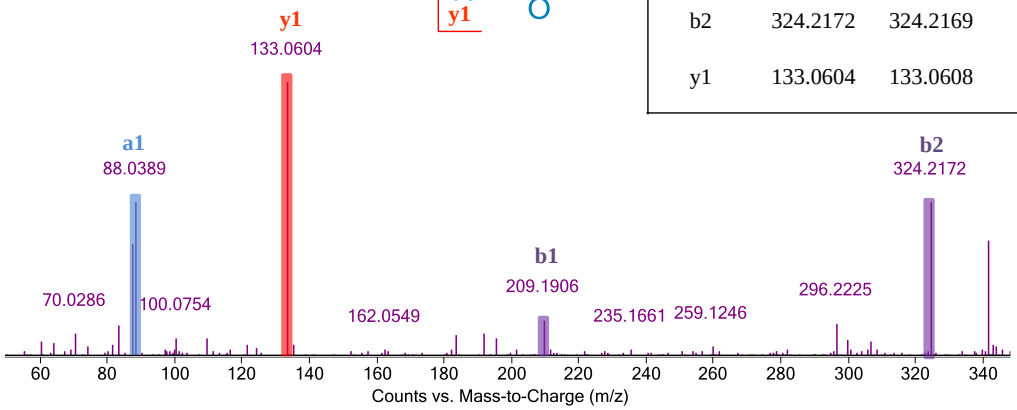
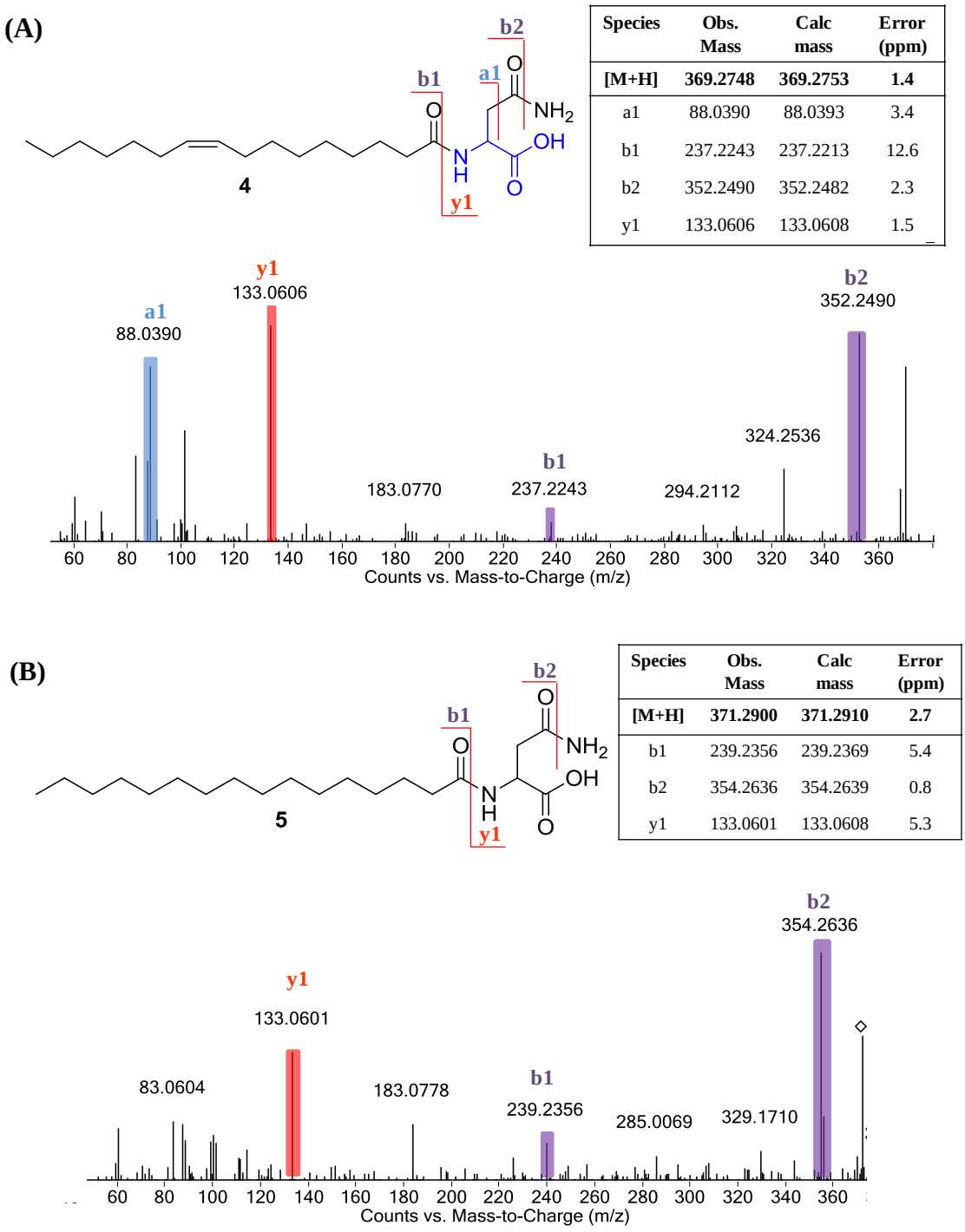


Figure S15. ESI-QTOF-HRMS analysis (top) and MS^2 fragmentation pattern (bottom) of isolated cis-7-tetradecenoyl-D-Asn (**2**).

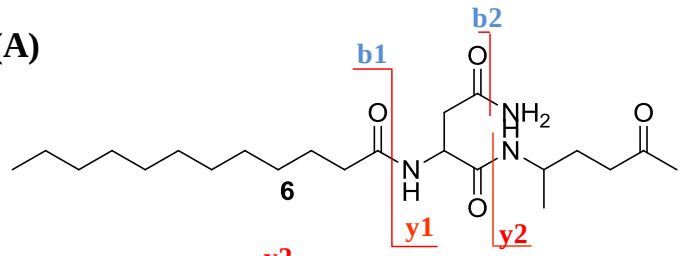


Species	Obs. Mass	Calc mass	Error (ppm)
[M+H]	**369.2748**	**369.2753**	**1.4**
a1	88.0390	88.0393	3.4
b1	237.2243	237.2213	12.6
b2	352.2490	352.2482	2.3
y1	133.0606	133.0608	1.5

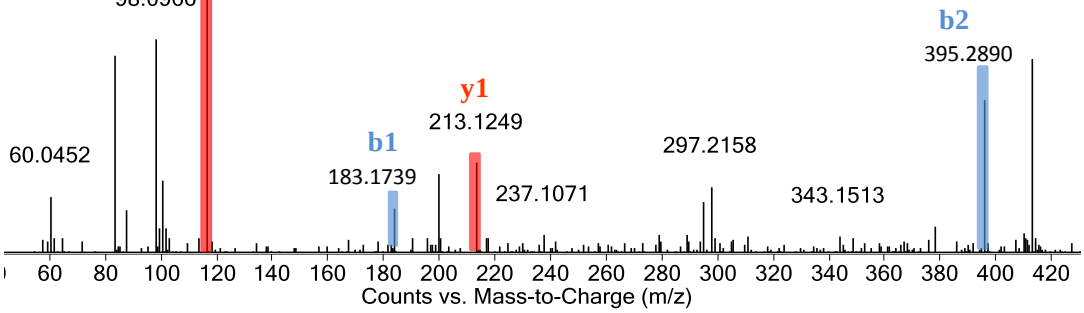
Species	Obs. Mass	Calc mass	Error (ppm)
[M+H]	**371.2900**	**371.2910**	**2.7**
b1	239.2356	239.2369	5.4
b2	354.2636	354.2639	0.8
y1	133.0601	133.0608	5.3

Figure S16. MS^2 fragmentation of proposed compounds (A) *m/z* 369 (**4**) and (B) *m/z* 371 (**5**). Location of the double bond in **4** has not been experimentally determined.

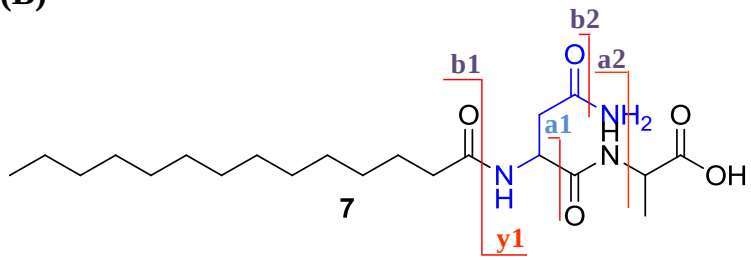
(A)



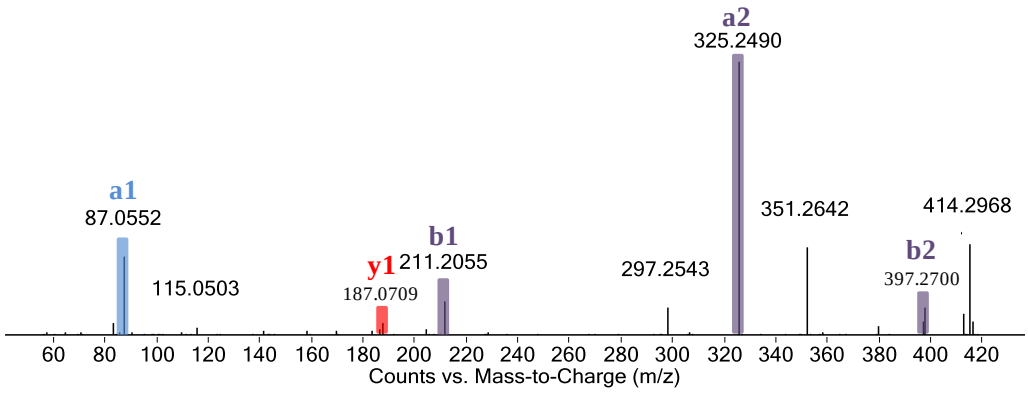
Species	Obs. Mass	Calc. mass	Error (ppm)
[M+H]	**412.3170**	**412.3175**	**1.2**
b1	183.1739	183.1743	2.2
b2	395.2890	395.2904	3.5
y1	213.1249	213.1234	7.0
y2	116.1069	116.1070	0.9

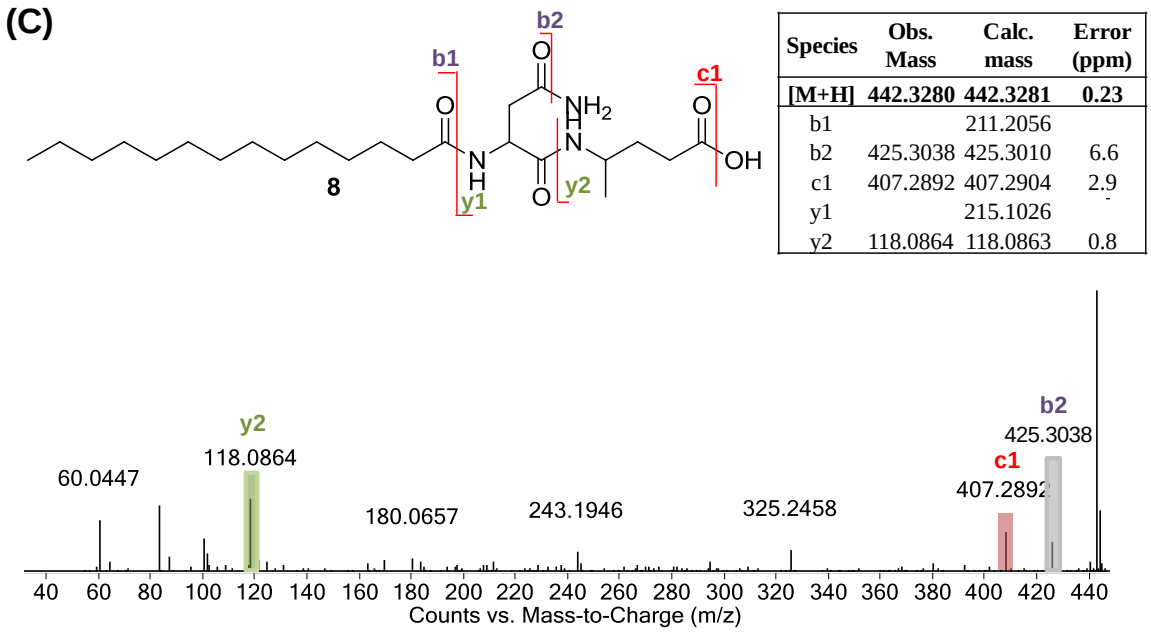


(B)



Species	Obs. Mass	Calc. mass	Error (ppm)
[M+H]	**414.2968**	**414.2968**	**1.0**
a1	87.0552	87.0553	1.0
a2	325.2490	325.2486	1.2
b1	211.2055	211.2056	0.5
b2	397.2700	397.2697	0.8
y1	187.0709	187.0713	2.1

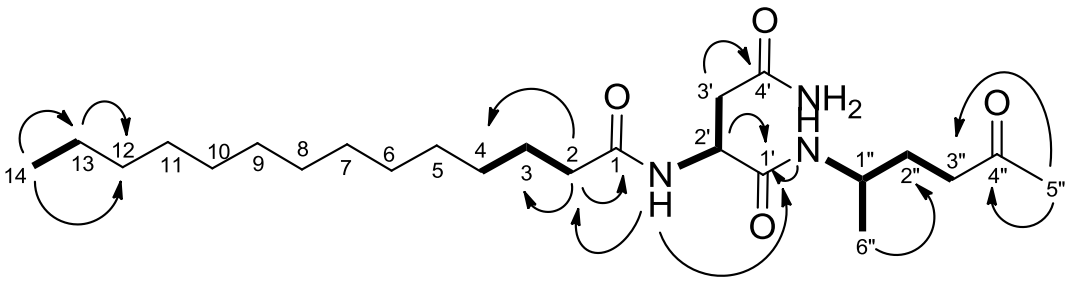




Species	Obs. Mass	Calc. mass	Error (ppm)
[M+H]	**442.3280**	**442.3281**	**0.23**
b1		211.2056	
b2	425.3038	425.3010	6.6
c1	407.2892	407.2904	2.9
y1		215.1026	
y2	118.0864	118.0863	0.8

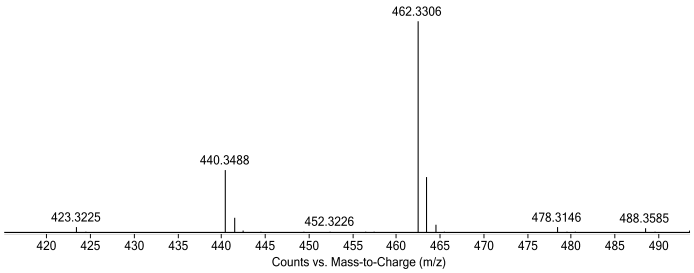
Figure S17. MS^2 fragmentation of proposed compounds (A) *m/z* 412 (**6**), (B) *m/z* 414 (**7**), and (C) *m/z* 442 (**8**).

(A)



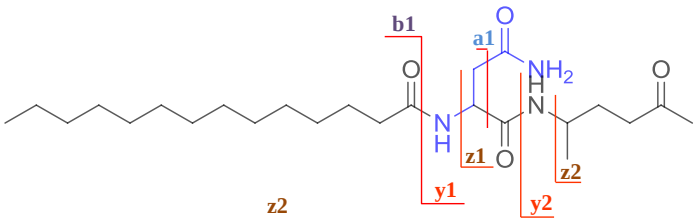
(B)

ESI-QTOF-HRMS



[M+Na]+
$C_{24}H_{45}N_3O_4Na$
Obs: 462.3306
Calc: 462.3308
Error (ppm): 0

MS² Fragmentation Pattern



Species	Obs. Mass	Calc. mass	Error (ppm)
a1	87.0550	87.0553	3.4
b1		211.2056	
y1		230.1499	
y2	116.1066	116.1070	3.4
z1	213.1232	213.1234	0.9
z2	99.0803	99.0804	1.0

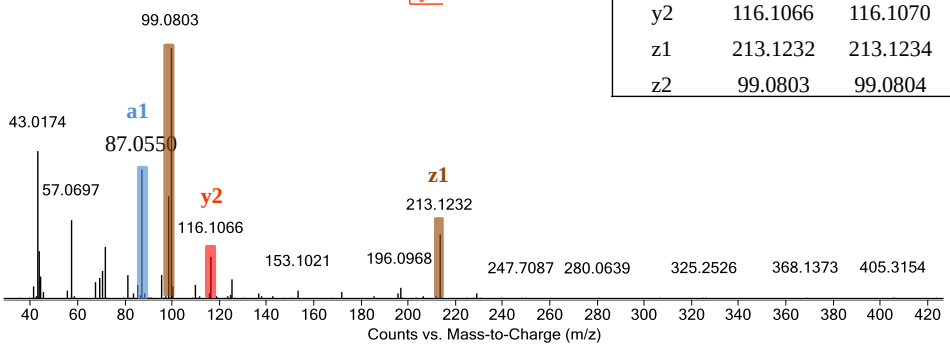


Figure S18. Structural characterization of isolated metabolite **3**. (A) Key 2D NMR correlations, where COSY correlations are bolded and HMBC correlations are denoted by arrows. (B) ESI-QTOF-HRMS analysis (top) and MS^2 fragmentation pattern (bottom) of **3**.

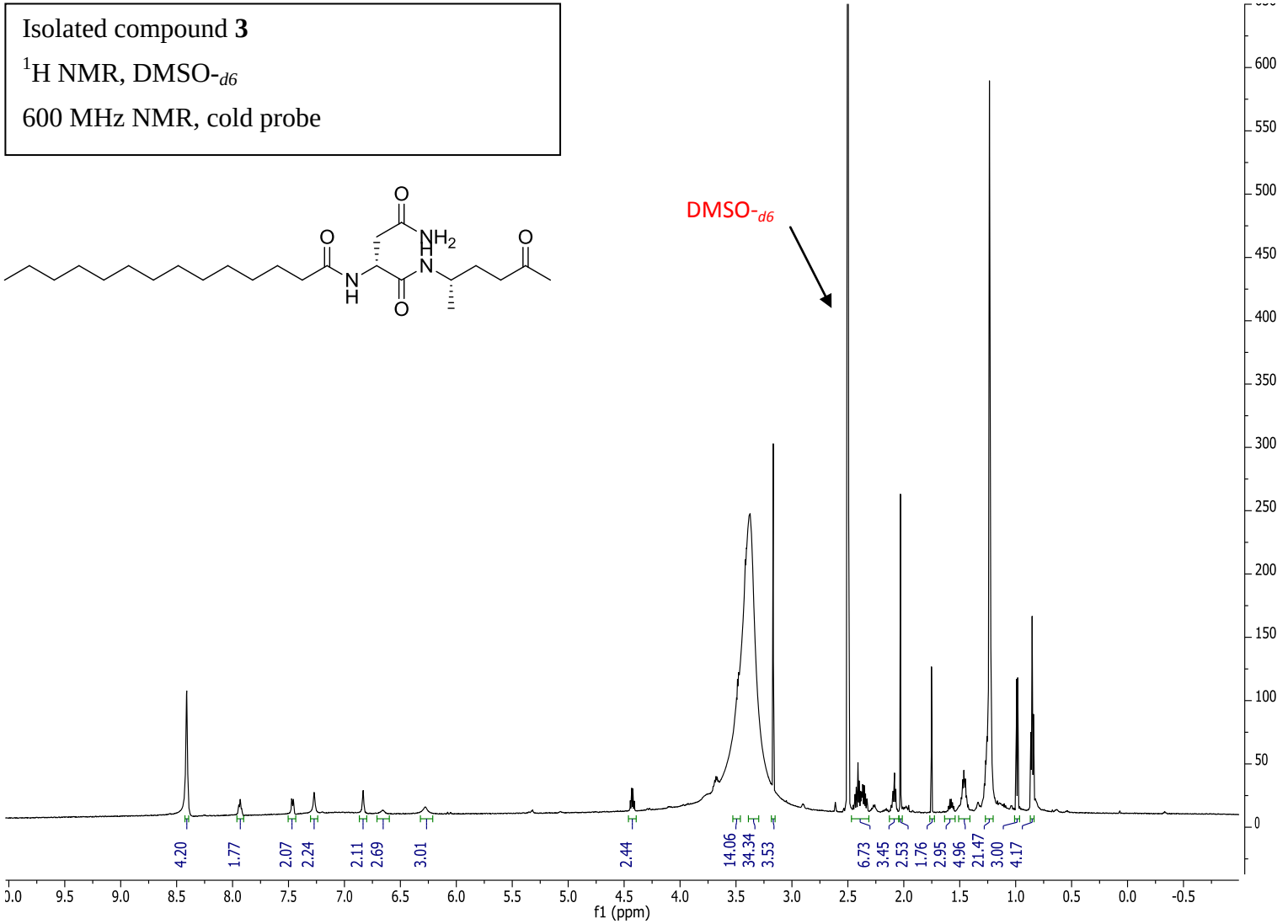


Figure S19. ^{1}H NMR of isolated **3**. Recorded with 600 MHz NMR, cold probe, in DMSO-*d6*.

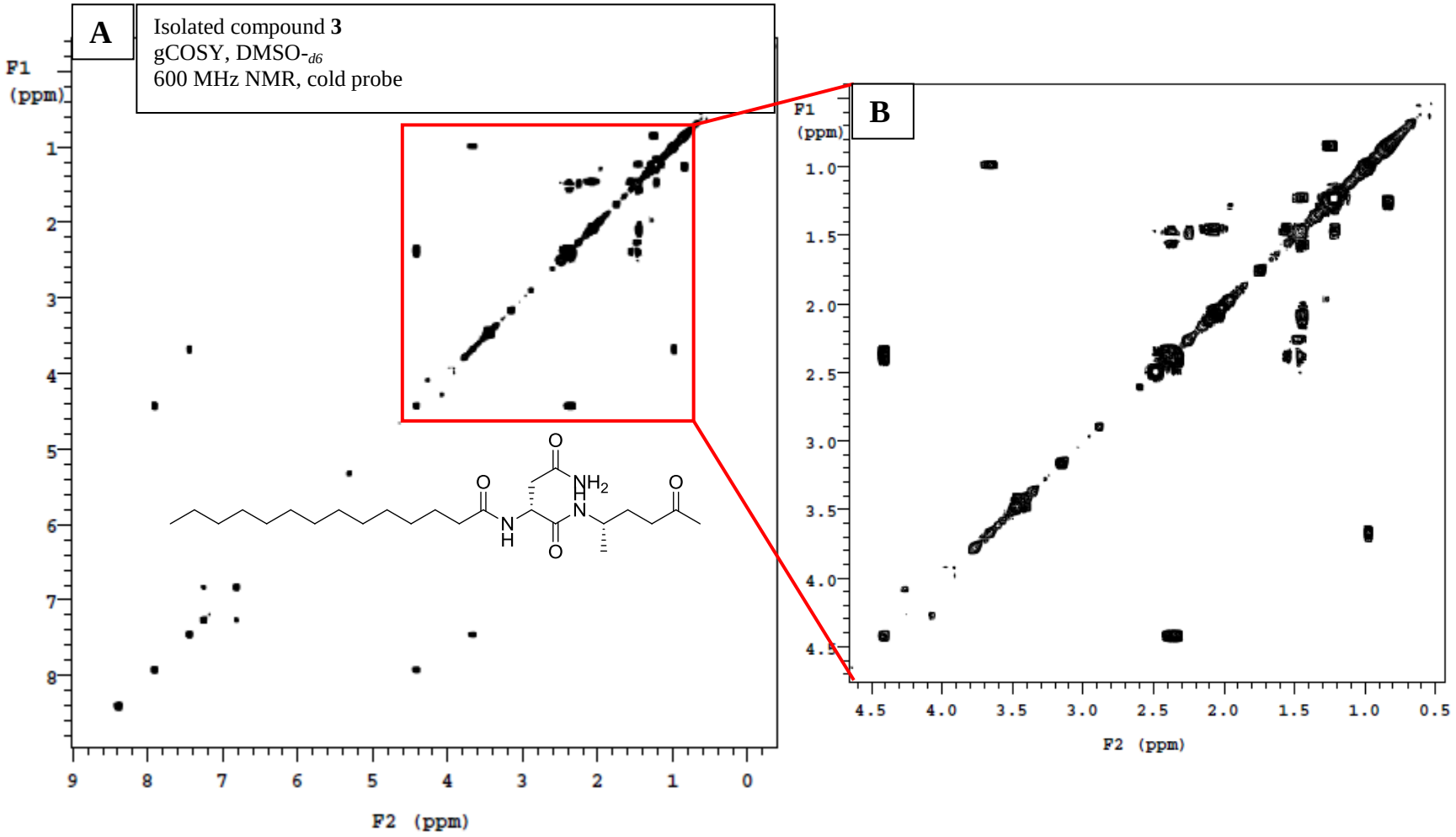


Figure S20. gCOSY NMR of isolated **3**. Full spectrum is shown in (A) and expanded highlighted region (red square) is shown in (B). Recorded with 600 MHz NMR, cold probe, in DMSO-*d6*.

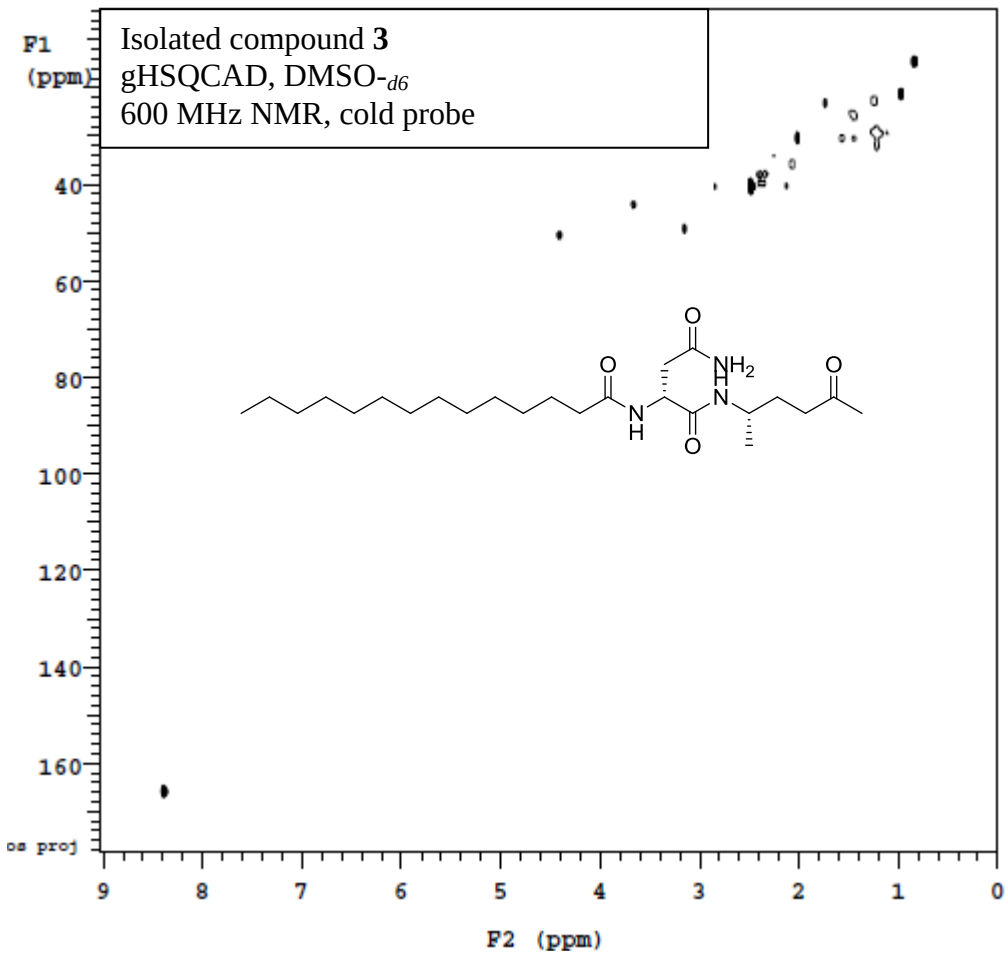


Figure S21. gHSQCAD NMR of isolated **3**. Recorded with 600 MHz NMR, cold probe, in DMSO-$_{d6}$.

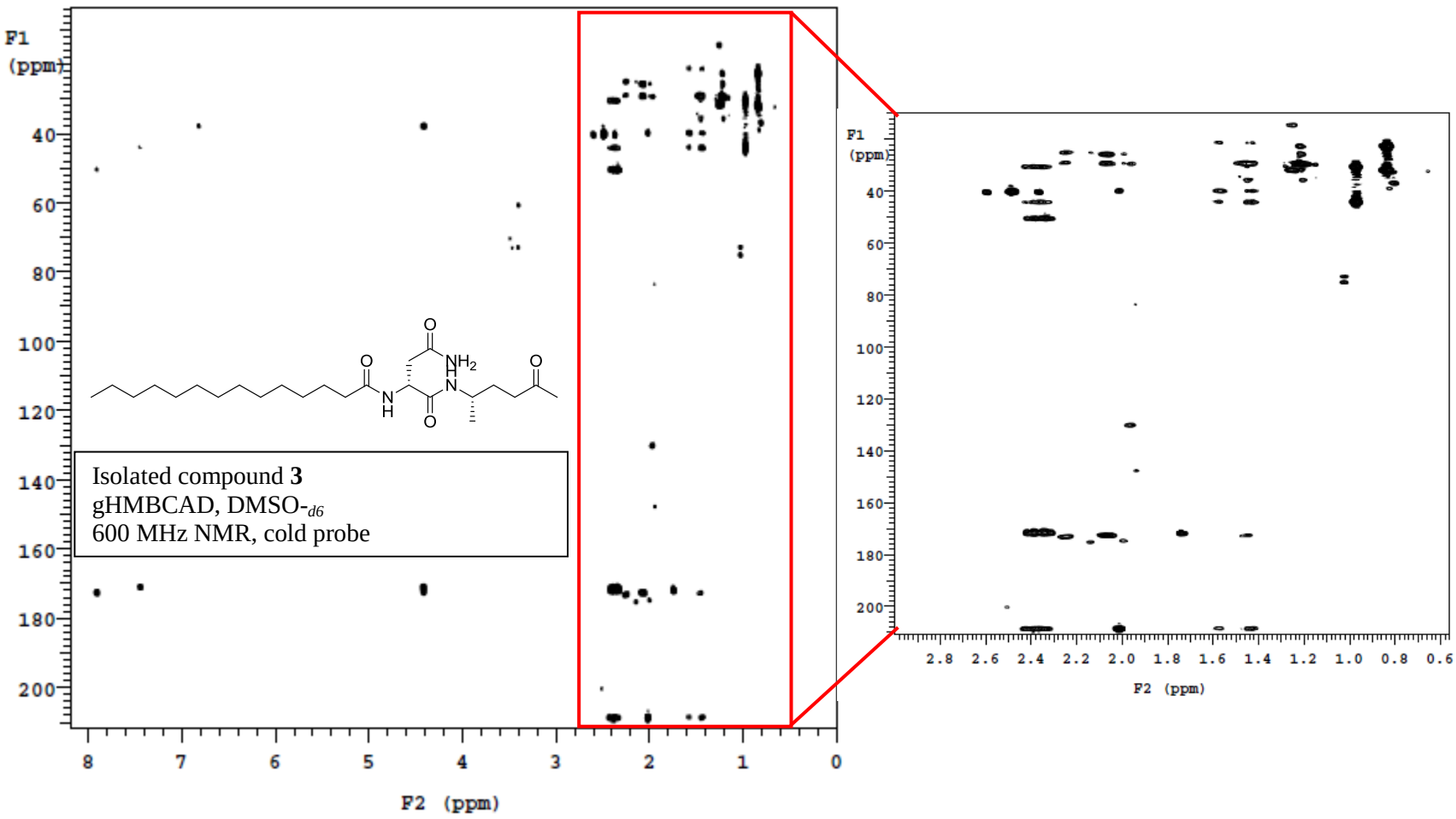


Figure S22. gHMBCAD NMR of isolated **3**. Full spectrum is shown in (A) and expanded highlighted region (red square) is shown in (B). Recorded with 600 MHz NMR, cold probe, in DMSO-$_{d6}$.

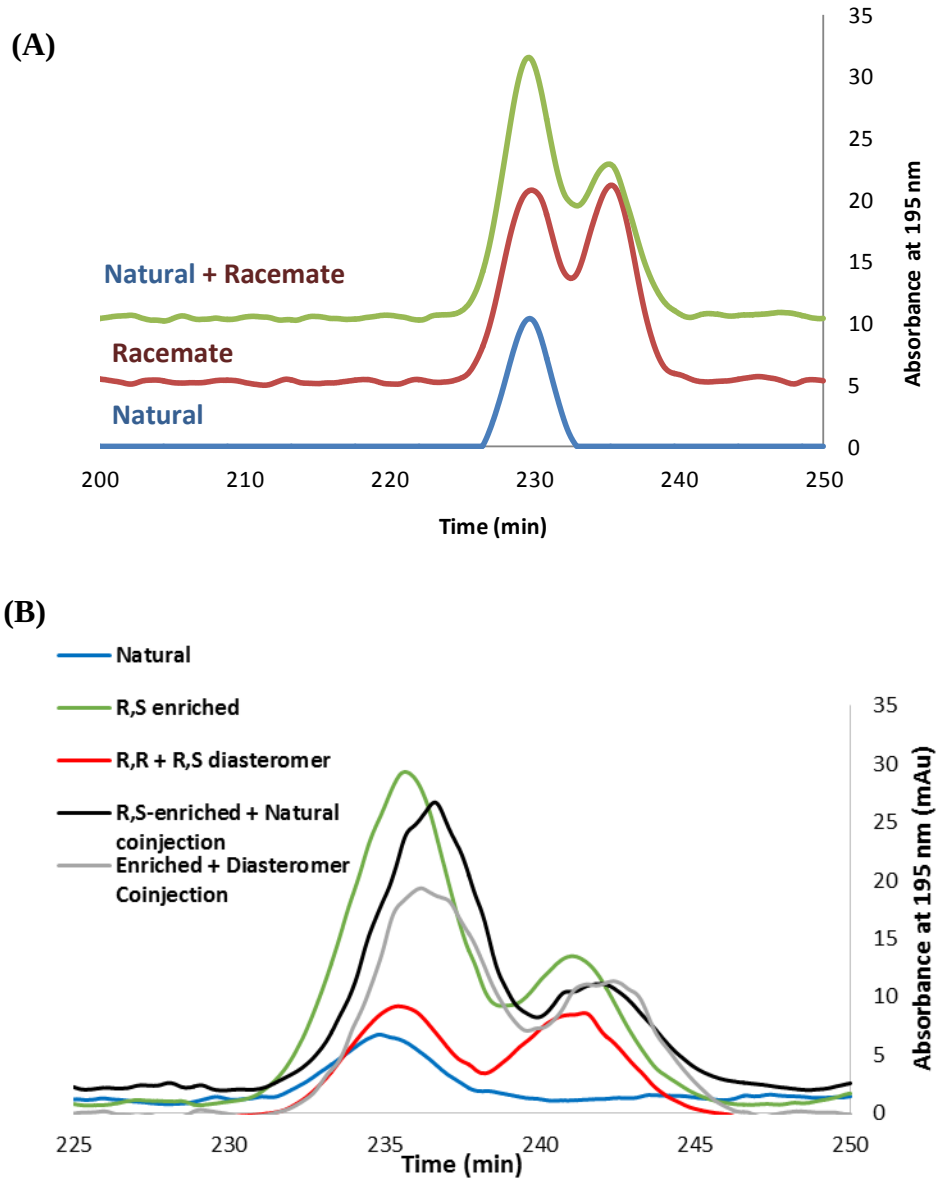
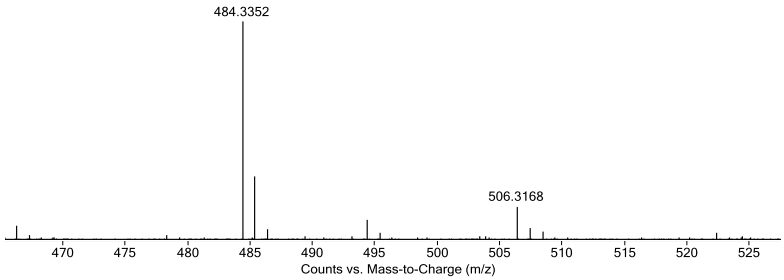


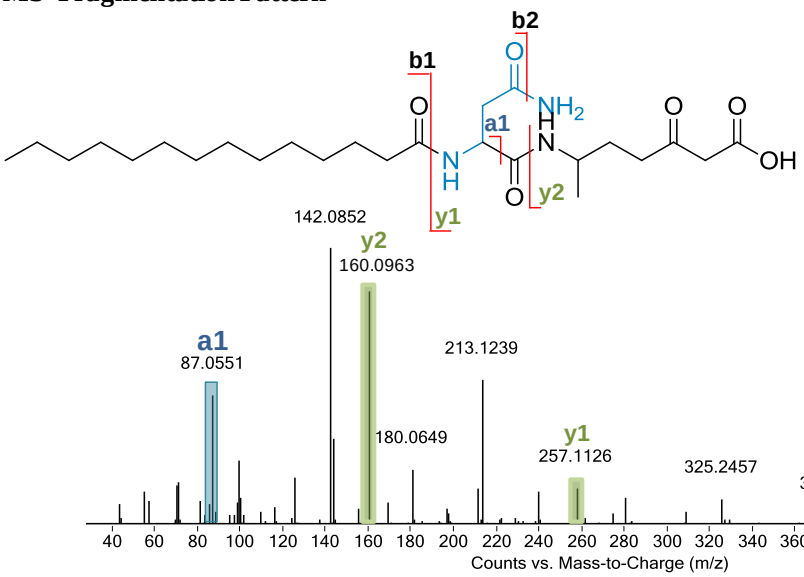
Figure S23. Absolute configuration determination of isolated **3** by diastereomeric resolution. (A) *R, S* and *R, R* diasteromers of **3** were separated by HPLC. Injection of isolated natural product (blue) compared to a synthetic standard from racemic starting material (red) show the natural product is a single diastereomer. This is confirmed by a coinjection of the two samples (green). (B) Absolute configuration of **3** was determined by synthesis. Synthesis starting from *(S)*-propylene oxide and **1** enriches for the *R, S* diastereomer (blue). Retention time comparison (red) and co-injection (black) reveal the natural product has the same configuration.

ESI-QTOF-HRMS



[M+H]+
$C_{25}H_{46}N_3O_6$
Obs: 484.3352
Calc: 484.3387
Error (ppm): -7.23

MS^2 Fragmentation Pattern



Species	Obs. Mass	Calc. mass	Error (ppm)
a1	87.0551	87.0553	2.3
b1		211.2056	
b2	467.3094	467.3116	4.7
y1	257.1126	257.1132	2.3
y2	160.0963	160.0968	3.1

Figure S24. ESI-QTOF-HRMS analysis (top) and MS^2 fragmentation pattern (bottom) of proposed structure (**9**) with *m/z* 484.

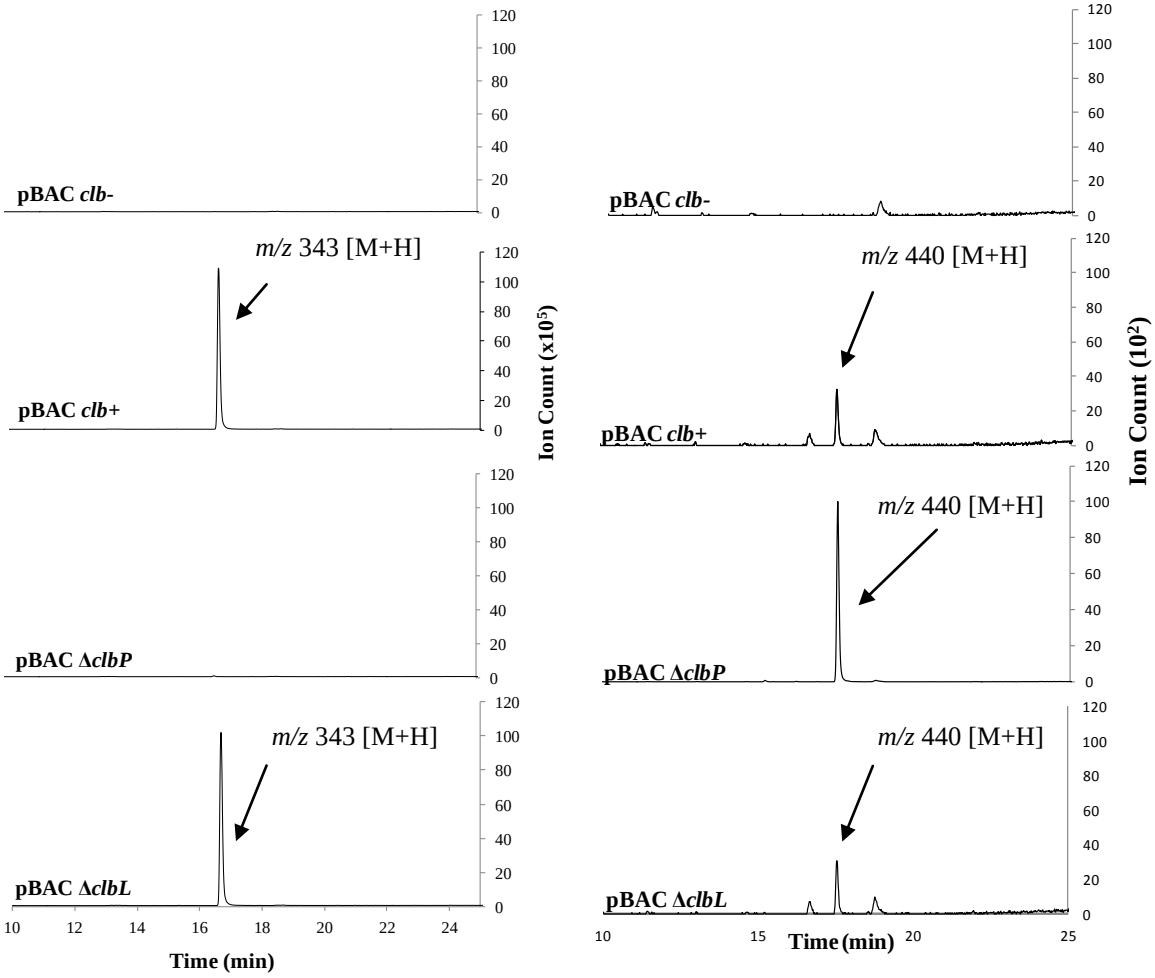


Figure S25. Extracted ion chromatograms of **1** (EIC +343) and **3** (EIC +440) from EtOAc extracts of pBAC *clb*-, pBAC *clb*+, and the two hydrolytic enzyme mutants (pBAC *ΔclbP* and pBAC Δ*clbL*). Deletion of *clbP* resulted in increased presence of **3**, while deletion of *clbL* showed no detectable effect compared to wildtype. Raw abundance of **3** extracted from both pBAC *clb*+ and pBAC *ΔclbL* was 1.0E+06 (data acquired from five biological replicates). **1** is present in only minor amounts when *clbP* is deleted.

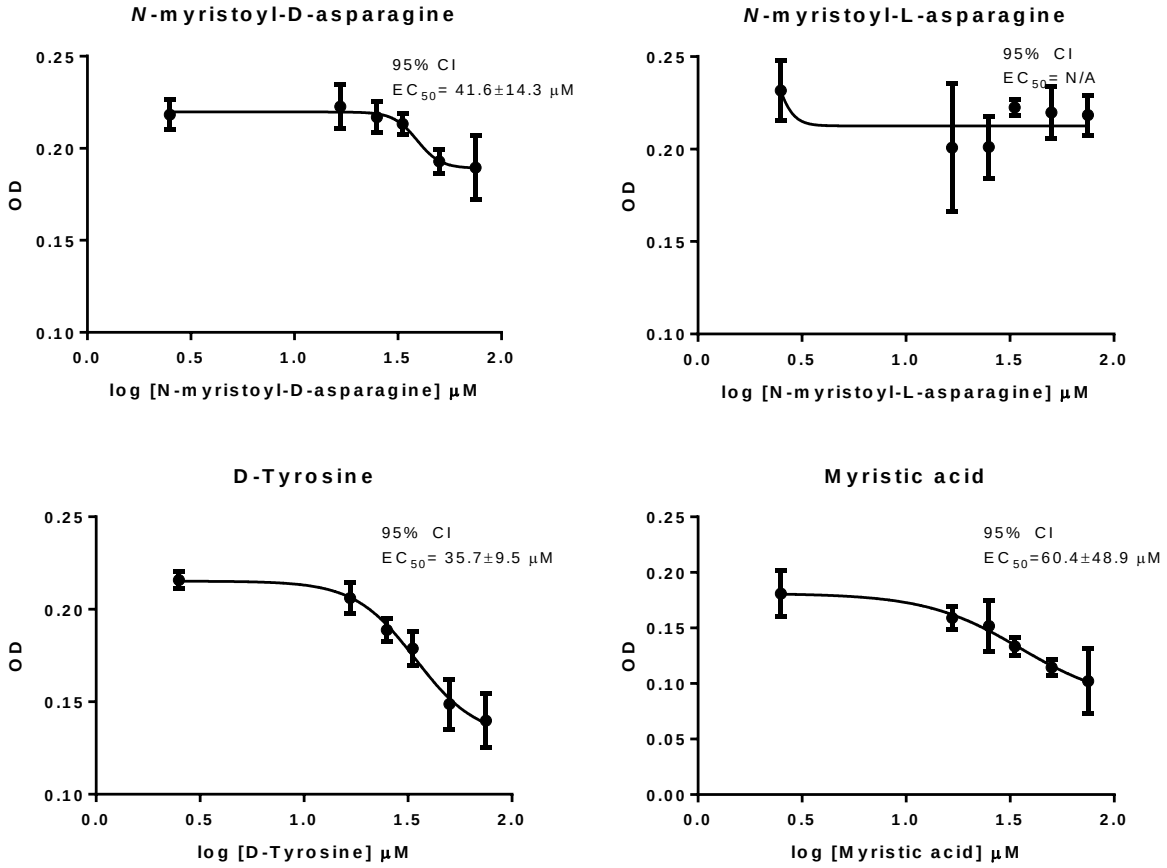
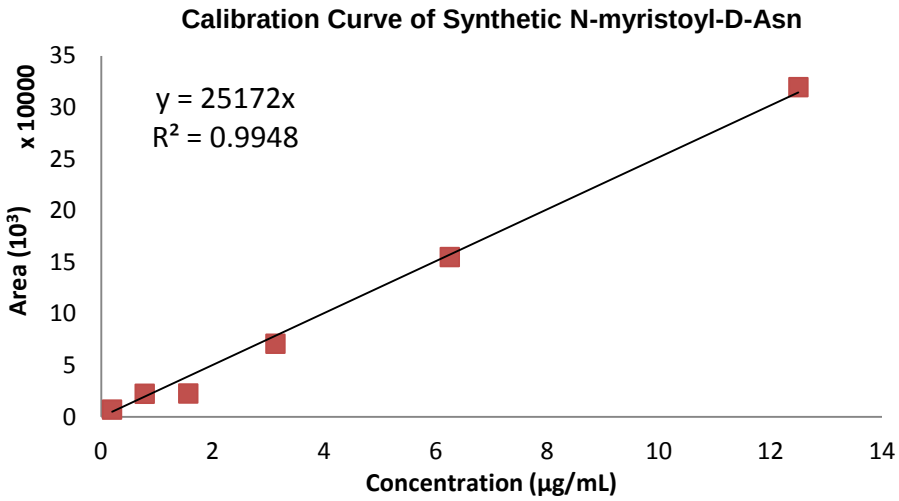


Figure S26. Growth inhibition against *B. subtillis* NCIB 3610. Inhibition was measured by optical density (OD_{600}) of bacteria four hours post infection with synthesized *N*-myristoyl-D-Asn (**1**), the L-enatiomer (**12),** D-tyrosine, or myristic acid. Error bars represent the standard deviation of quadruplicate measurements. 95% confidence intervals for the EC_{50} are displayed on the graph.



	µM	
Time (h)	**pBAC *clb*+ Supernatant**	**pBAC *clb*+ Cell**
6	0.90	0.00
11	5.01	1.30
25	15.86	3.38
48	23.40	4.06

Figure S27. Calibration curve of extracted synthetic **1** (top) and concentration of extracted**1** from pBAC clb+ bacterial cultures at different time points within a 48 h growth period (bottom).

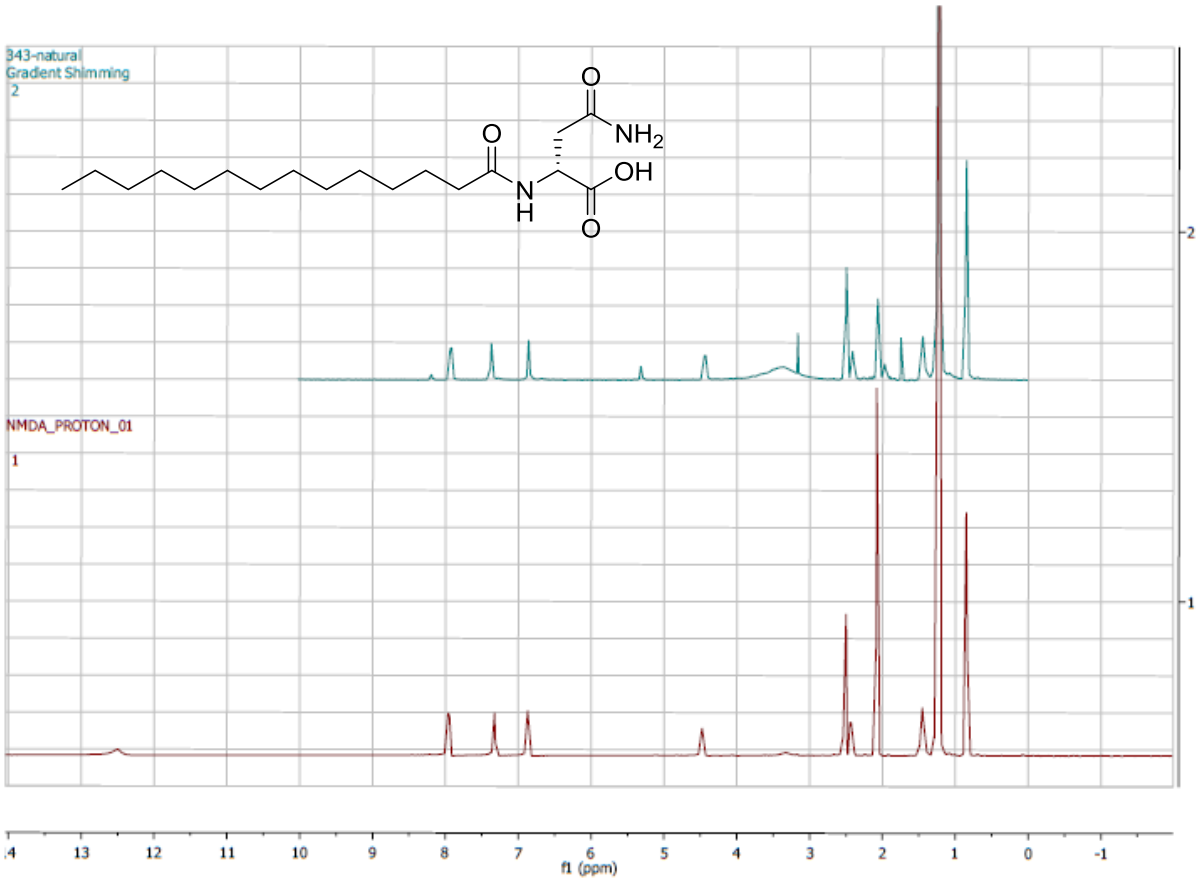


Figure S28. Comparison of ^{1}H NMR spectra from synthetic and isolated metabolite **1**. Synthetic **1** (red spectrum, 400 MHz, DMSO-d_6) has the same NMR spectrum as the isolated natural product (blue spectrum, 600 MHz, DMSO-$_{d6}$).

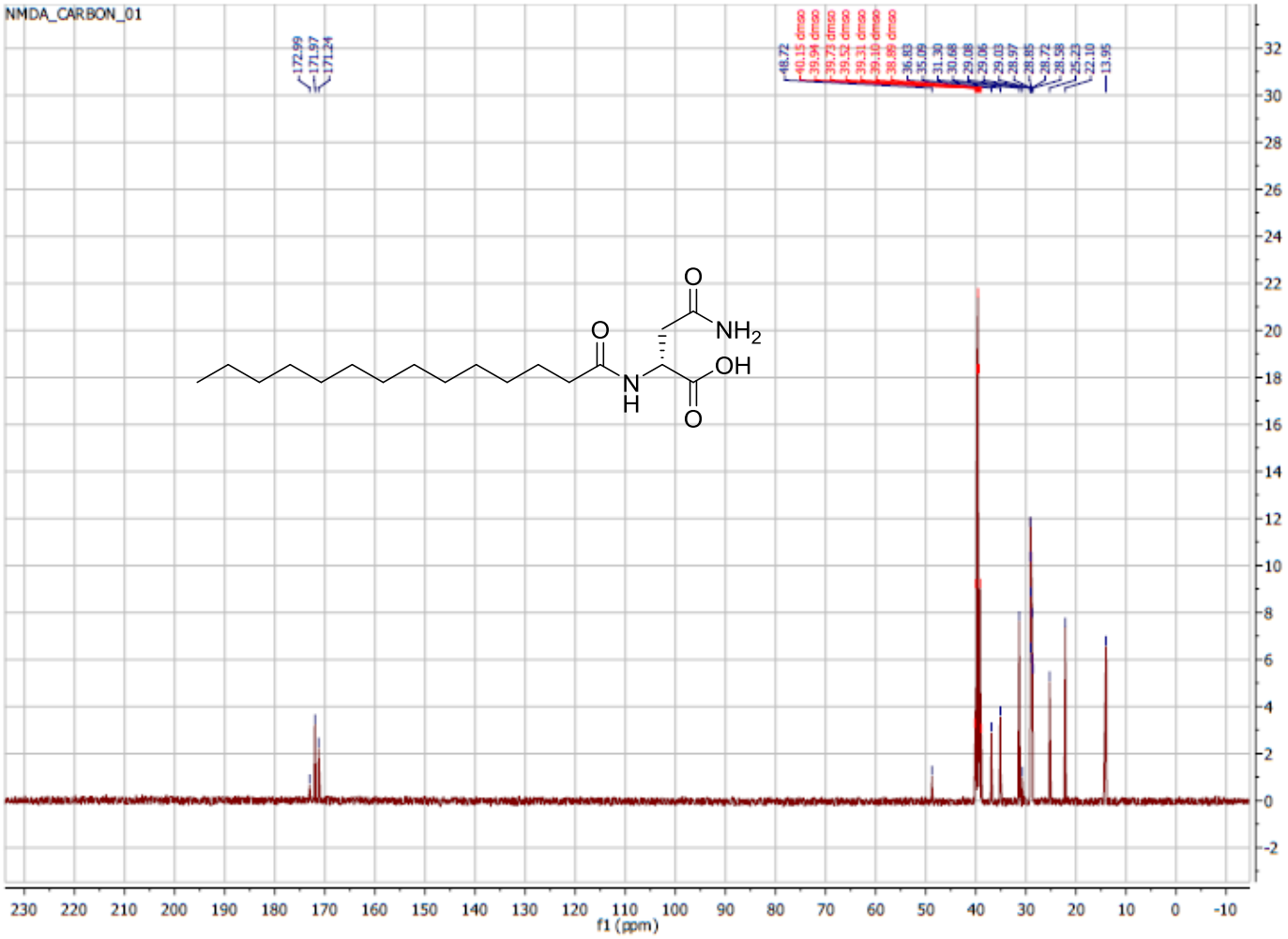


Figure S29. ^{13}C NMR spectrum of synthetic **1** in DMSO-*d6*.

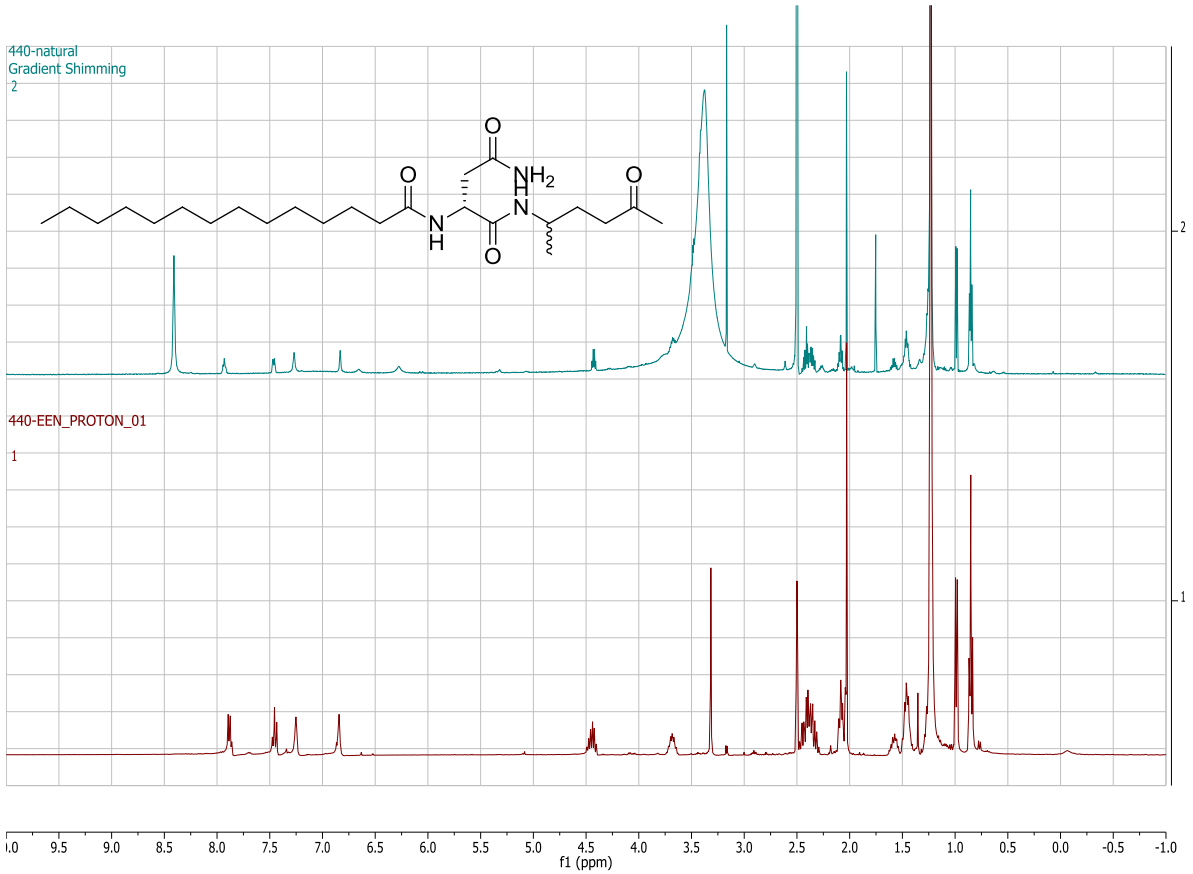


Figure S30. Comparison of ^{1}H NMR spectra from synthetic and isolated metabolite **3** in DMSO-$_{d6}$. Synthetic **3** (red spectrum, 400 MHz, DMSO-$_{d6}$) has same spectrum as the isolated metabolite (blue spectrum, 600 MHz, DMSO-$_{d6}$). Synthetic **3** is a mixture of the (*R*,*S*) and (*R*,*R*) diastereomers in an approximately 2:1 ratio. The peaks corresponding to the major diastereomer align with those of the isolated natural product.

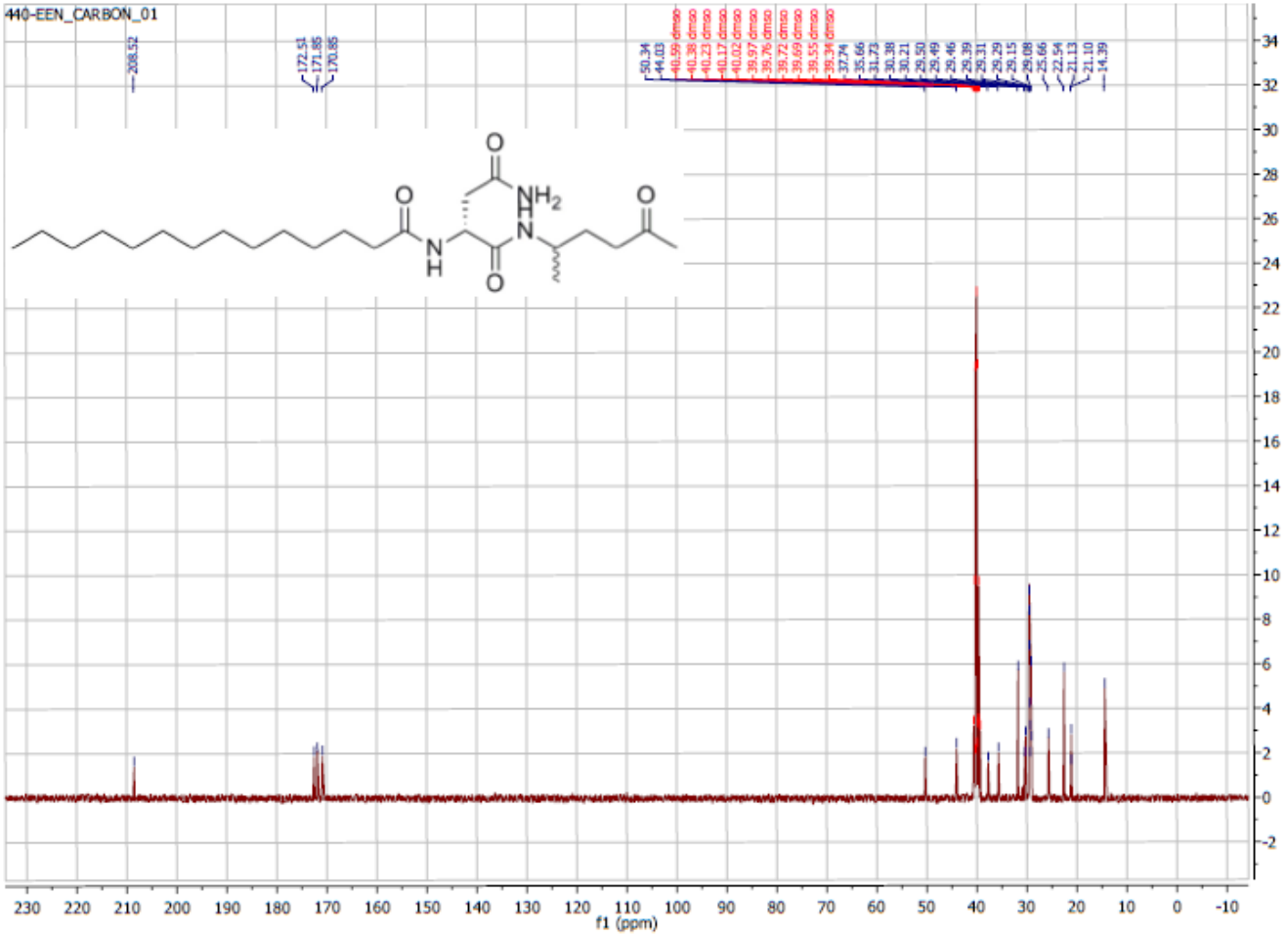
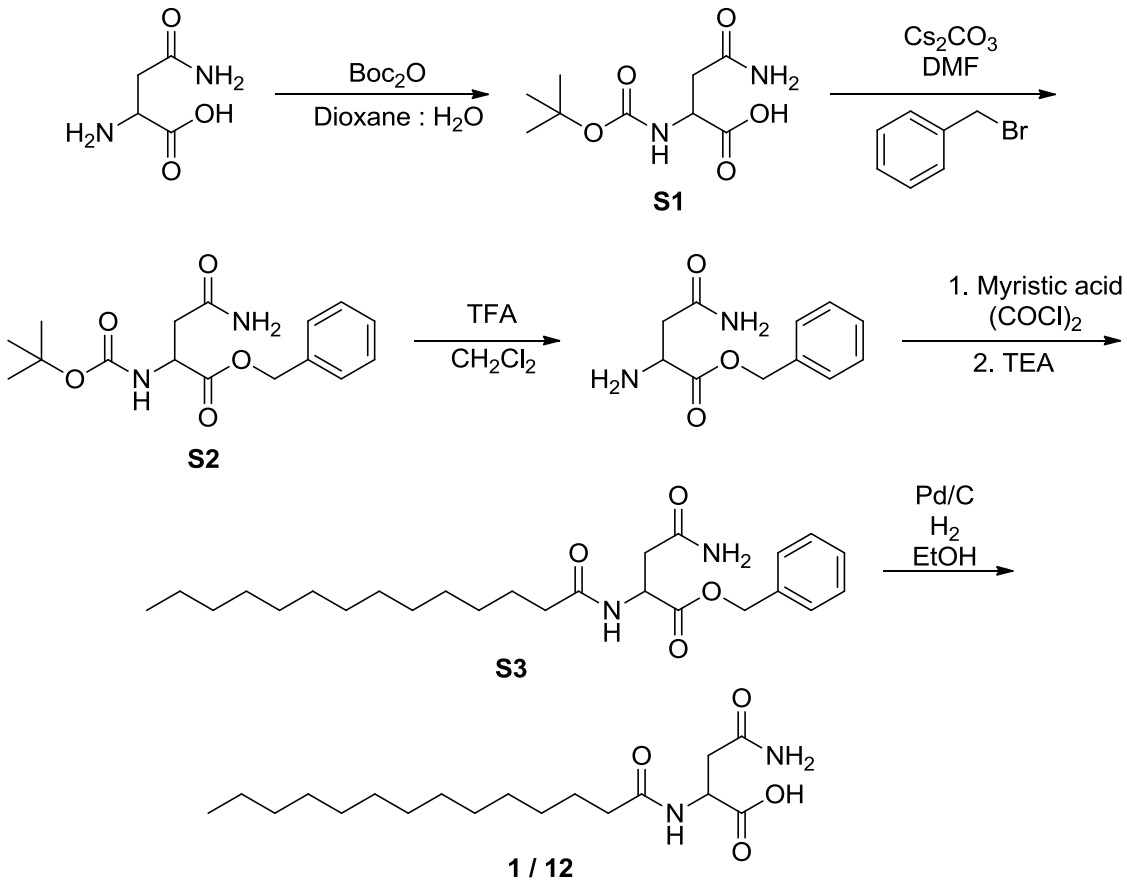


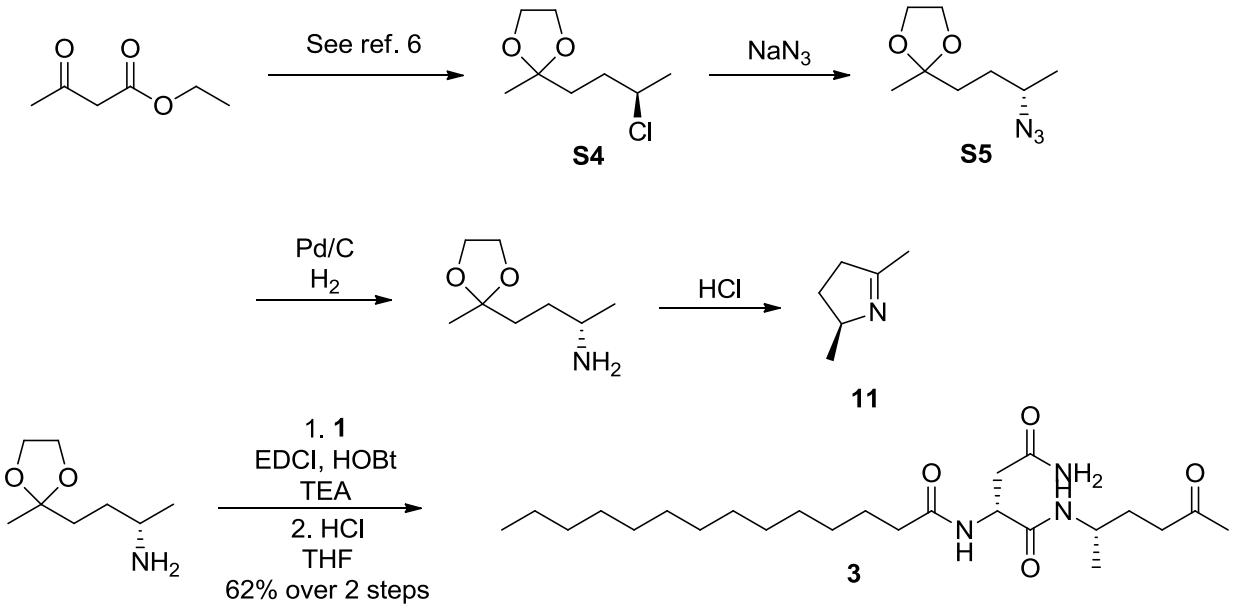
Figure S31. ^{13}C NMR spectrum of synthetic **3** in DMSO-$_{d6}$. The solution is a mixture of the (*R*,*S*) and (*R*,*R*) diastereomers in approximately a 2:1 ratio.

4. Supplemental Schemes

Scheme S1. Synthesis of the D-enantiomer (**1**) and L-enantiomer (**12**) of *N*-myristoyl-asparagine



Scheme S2. Synthesis of **3** and 2,5-dimethylpyrroline (**11**)



5. Supplemental References

1. Datsenko, K. A.; Wanner, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 6640-5.
2. Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 1541.
3. Nougayrede, J. P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Science* **2006,** *313*, 848.
4. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, E1743.
5. Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Anal. Chem.* **1997,** *69*, 5146.
6. Cornish, C. A.; Warren, S. *J. Chem. Soc., Perkin Trans.1* **1985,** 0, 2585.
7. Evans, G. G. *J. Am. Chem. Soc.* **1951,** *73*, 5230.
8. Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. *J. Bacteriol.* **1995,** *177*, 4121.
9. Piel, J. *Nat. Prod. Rep.* **2010,** *27*, 996.