Supplementary Materials for

Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom

John K. Brunson, Shaun M. K. McKinnie, Jonathan R. Chekan, John P. McCrow, Zachary D. Miles, Erin M. Bertrand, Vincent A. Bielinski, Hanna Luhavaya, Miroslav Oborník, G. Jason Smith, David A. Hutchins, Andrew E. Allen, Bradley S. Moore

Correspondence to: bsmoore@ucsd.edu (B.S.M.); aallen@jcvi.org (A.E.A.)

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Table of Contents

General Materials and Methods

All chemicals and solvents were used as received from the commercial supplier (Sigma-Aldrich or Fisher). PCR was carried out using standard thermocycling protocols using either Phusion High-Fidelity DNA Polymerase (New England Biolabs), Prime Star MAX (TaKaRa) or OneTaq 2X MM (New England Biolabs) using the templates and buffers as specified for each reaction (detailed in experimental sections). All protein purification from *E. coli* was performed on an ÄKTApurifier instrument (GE Healthcare) with the modules Box-900, UPC-900, R-900 and Frac-900 with all solvents filtered through a nylon membrane 0.2 μm GDWP (Merck) prior to use. FPLC data was analyzed with UNICORN 5.31 (Built 743) software. Ultracentrifugation was performed using a SW 55 Ti rotor (Beckman Coulter) on an Optima L-100 XP (Beckman Coulter) using ultracentrifuge tubes from the manufacturer. All protein quantification was done by method of Bradford using the Protein Assay Dye Reagent Concentrate (Bio-Rad) on protein sample dilutions in MilliQ water.

General LCMS measurements were carried out on an Agilent Technologies 1200 Series system with a diode-array detector coupled to a Bruker Amazon SL ESI-Ion Trap mass spectrometer in negative ionization mode. The LCMS data was processed by Bruker Compass Data Analysis 4.2 SR2. Higher resolution liquid chromatography mass spectrometry (HRMS) measurements were carried out on an Agilent Technologies 1200 Series system with a diode-array detector coupled to an Agilent Technologies 6530 accurate-mass Q-TOF LCMS. In both cases, compounds were separated by reversed-phase chromatography on a Phenomenex Luna C18(2), 5 μm-100 x 4.6 mm column with water + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B) as eluents. We used two specific LC methods for separation of compounds for both general LCMS as well as HRMS, both run at a flow rate of 0.75 mL/min. LC Method A uses the following gradient: hold at 5% B for 1 minute, 5% to 100% B over 20 minutes, hold at 100% B for 1.5 minutes, 100% to 5% B over 2.5 minutes, hold at 5% for 2 minutes. LC Method B uses the following gradient: hold at 5% B for 5 minutes, 5% to 35% B over 15 minutes, 35% to 100% B over 1 minute, hold at 100% B for 1.5 minutes, 100% to 5% B over 2.5 minutes, hold at 5% B for 2 minutes.

For chemical synthesis, all anhydrous reactions were performed under argon atmosphere in flame-dried glassware. Reactions were monitored by LCMS using previously described conditions or TLC (Merck, silica gel 60 F254). Products were visualized by UV at 254 nm or with $KMnO_4$ staining solution (1.5 g $KMnO_4$, 10 g K_2CO_3, 1.25 mL 10% NaOH in 200 mL water). TLC R_f values were rounded to the nearest 0.05. Products were purified by flash column chromatography using silica gel (Alfa Aesar, 60 Å pore size). Concentration under reduced pressure was carried out by rotary evaporation. NMR spectra were recorded on a Bruker Avance III spectrometer (600 MHz) using either a 1.7 mm inverse detection triple resonance (H-C/N/D) cryoprobe or a 5 mm inverse detection triple resonance (H-C/N/D) cryoprobe, or a JEOL spectrometer (500 MHz) using D_2O or $CDCl_3$ as solvents. Chemical shifts (δ) are reported in ppm and are referenced to internal sodium formate for D_2O (δ = 8.44 ppm for ^1H, δ = 171.7 ppm for ^{13}C) or the solvent signal for $CDCl_3$ (δ = 7.26 ppm for ^1H, δ = 77.2 ppm for ^{13}C). Data for NMR spectra was reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, *J* = coupling constant in Hz. Infrared spectra (IR) were recorded on a Nicolet IR 100 FT-IR (Thermo) spectrometer. Absorption maxima are reported as wavenumbers in cm^{-1}.

RNAseq and Differential Expression Analysis Methods

Biomass for *Pseudo-nitzschia multiseries*, cultured across a matrix of pCO_2 conditions and phosphate concentrations, was obtained from experiments conducted by (*17*). Total RNA was extracted using Trizol reagent (Thermo Fisher Scientific), followed by genomic DNA removal using TURBO DNA-free™ Kit (Thermo Fisher Scientific), and Agencourt RNA Clean XP kit (Beckman Coulter) clean up the RNA. Suitable RNA quality was verified using an Agilent 2100 Bioanalyzer.

In order to construct RNAseq libraries from total RNA, ribosomal RNA was removed from 1.5 μg of total using Ribo-Zero Magnetic kits (Illumina). The composition of Removal Solutions constituted a mixture of plant, bacterial, and human/mouse/rat Removal Solution at a ratio of 2:1:1. An Agilent 2100 Bioanalyzer was used to evaluate the quality of rRNA removal. 5 ng of rRNA subtracted RNA was used for library construction using a Scriptseq v2 RNA-seq Library Preparation Kit (Illumina). An Agilent 2100 Bioanalyzer was used to verify the mean size of libraries was within a range of 360-370 bp.

PolyA mRNA transcriptomes were constructed using 300 ng of total RNA as input into the TruSeq RNA Sample Preparation Kit (Illumina), followed by the manufacture's Low-Throughput protocol. The mean size of the final libraries was inspected using an Agilent Bioanalyzer 2100 and confirmed to be within a range of 350-420 bp.

Transcriptomes were sequenced on the Illumina HiSeq platform (Illumina, San Diego, CA, USA). Reads were trimmed of primer sequences and assembled de novo using CLC Genomics workbench. Putative open ready frames on the assembled contigs were called using FragGeneScan (*32*). Putative protein sequences were annotated using hidden markov models and blastp against PhyloDB 1.02 (*33*). *Pseudo-nitzschia multiseries* RNAseq data have been deposited at the NCBI sequence read archive under accession numbers SAMN08773590-SAMN08773622.

Reads were trimmed for quality and filtered to remove primers, adaptors and rRNA sequences using Ribopicker v.0.4.3 (*34*). CLC Assembly Cell (CLCbio) was used to assemble contigs; open reading frames (ORFs) were predicted from the assembled contigs using FragGeneScan (*32*) and additional rRNA sequences were identified and removed. The remaining ORFs were annotated *de-novo* for function and taxonomy via KEGG, KO, KOG, Pfam, TigrFam and the reference dataset PhyloDB 1.02 (*33*).

Reads were aligned to the *Pseudo-nitzschia multiseries* CLN-47 reference assembly Psemu1 (https://genome.jgi.doe.gov/Psemu1) using BWA MEM (*35*). Read counts per gene were aggregated using featureCounts (*36*). Fragments per kilobase of transcript per million mapped reads (FPKM) was calculated as 10^9 * reads / (total_mapped_reads * gene_length) and was used for assessment of relative expression. Differential expression analysis was performed using edgeR (*37*) with the exactTest function, with multiple testing correction by Benjamini-Hochberg to give FDR adjusted p-values.

Next, the model CDS sequences with reads mapped to them were filtered for only those sequences significantly upregulated under limited phosphate ($log2FC > 0$, $FDR < 0.05$), with reference only to the rRNA-depleted libraries. Specifically, we looked for those transcripts which were significantly upregulated under two phosphate and pCO_2 conditions: phosphate limited (0.5 μM) *P. multiseries* grown at 400 ppm pCO_2 and 730 ppm pCO_2, as compared to cultures grown with replete phosphorous (20 μM) at comparable pCO_2 levels. After requiring a cutoff of 50 FPKM for each CDS sequence, with FPKMs summed across all rRNA-depleted libraries (18

libraries, an average of 2.78 FPKM per CDS per library), we were left with a set of 482 CDS which correspond with upregulated transcripts under phosphate limitation.

Next, we looked for genes upregulated under both phosphate limitation and pCO_2 stress. We took the set of 482 CDS upregulated under phosphate limitation and filtered for those CDS which correspond to transcripts significantly upregulated (log2FC > 0, FDR < 0.05) under increasing pCO_2 at low phosphate concentrations. Specifically, we looked for transcripts which were upregulated when *P. multiseries* was grown under 400 ppm pCO_2 or 730 ppm pCO_2, as compared to *P. multiseries* grown under 220 ppm pCO_2, with all cultures being grown under limited phosphate conditions (0.5 μM). This yielded a total of 43 transcripts upregulated under both pCO_2 stress and strictly phosphate limitation.

Molecular Biology/Biochemical Methods

Diatom Strains, DNA extraction, cDNA generation, and *dab* gene amplification

For amplification of *dab* genes, *Pseudo-nitzschia multiseries* isolate 15091C3 was collected from a net tow at Monterey Wharf II in Monterey, CA on April 1st 2015. Once purified to unialgal culture, it was maintained in f/2 enriched sterilized Monterey Bay SW at 15-16 ^{o}C and 75 μmole photon m^{-2} s^{-1} PAR (*38*). The organism was also cultured in Natural Sea Water Media, which is 0.22 μM filtered sea water from Scripps Pier supplemented with 880 μM $NaNO_3$, 55 μM NaH_2PO_4, 100 μM Na_2SiO_3, and AQUIL trace metals and vitamins (*38*). These cultures were also maintained at 15-16 ^{o}C and 75 μmole photon m^{-2} s^{-1} PAR. RNA isolation was performed using the Direct-zol™ RNA MiniPrep kit (ZymoResearch) on *P. multiseries* harvested during late exponential/early stationary phase. cDNA was created using SuperScript™ III Reverse Transcriptase (ThermoFisher). Initial amplification of full-length *dabA* and *PmCPR1* was performed on cDNA samples using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. Primer set DabA_001-2 was used for *dabA* amplification while primer set PmCPR1_001-2 was used for *PmCPR1* amplification. Initial amplification of *dabC* and *dabD* was performed from genomic DNA extracted from stationary phase *P. multiseries* using Plant DNAzol™ Reagent (Invitrogen), again using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. Primer set DabC_001-2 was used to amplify *dabC* while primer set DabD_001-2 was used to amplify *dabD*.

Verification of *dab* cluster in an environmental isolate of *Pseudo-nitzschia multiseries*

PrimeStar MAX polymerase (TaKaRa) and primer sets ampDabC-D_001-002 and ampDabA-C_001-002 were used to amplify the intergenic regions between DabC and DabD and between DabA and DabC, respectively, from gDNA extracted from *P. multiseries* isolate 15091C3. Overhangs for Gibson assembly were added using the primer sets OH_DabA-C_001-002 and OH_DabC-D_001-002. These fragments were then individually assembled into PtGG, a promoter-less vector used here for sequencing purposes (*39*) using Gibson assembly (*40*) and transformed into TOP10 chemically competent *E. coli*. Following colony PCR verification using primer set PtGG_col001-002, we used primer sets DabA-C_seq001-010 and DabC-D_seq001-004 for Sanger sequencing (Eurofins) to achieve full coverage. These sequencing reads, along with reads from sequencing performed on *dab* gene expression vectors, were used to assemble a contig which spans the 3' end of *dabA* to the 3' end of *dabD*. Full cluster sequencing has been deposited to GenBank, accession number MH202990.

Cloning of *dabA* into pET28 backbone

The pET28a vector was linearized using PrimeStar MAX polymerase (TaKaRa) and primer set pET28_BB_001-2. Truncated Ser26 – C terminus *dabA* gene was obtained by PCR with PrimeStar polymerase and primer set DabA_p28_001-002 to incorporate the appropriate Gibson assembly overhangs and remove the putative chloroplast transit peptide predicted by SignalP 3.0 and ASAFind (*41, 42)*. NEBuilder HiFi DNA Assembly mix was used to generate the *dabA* pET28a construct. Assembly of *dabA* into pET28a also incorporated the N-terminal His_6 tag for affinity purification. After confirming the *dabA* sequence by Sanger sequencing, the plasmid was transformed into *E. coli* BL21 cells for subsequent expression.

Expression and purification of DabA

An overnight culture of DabA expressing BL21 *E. coli* cells was used to inoculate 1 L of Terrific Broth in a 2.8 L flask. The flask was shaken at ~200 rpm at 37 ^{o}C until the cells reached an OD_{600} of ~0.6. The incubator was then shifted to 18 ^{o}C and the flasks were allowed to shake for an additional hour. The cells were then induced with 0.5 mM IPTG and were further incubated for ~16 h at 18 ^{o}C. After harvesting by centrifugation, the cells were resuspended in 25 mL of buffer containing 500 mM NaCl, 20 mM Tris-HCl pH 8.0, and 10% glycerol. The cells were either frozen or immediately used for protein purification.

The cells were lysed by sonication using a Qsonica 6 mm tip at 40% amplitude for 12 cycles of 15 sec on and 45 seconds off. The lysate was centrifuged at 15,000 xg for 30 min to remove cellular debris. The cleared lysate was loaded onto a 5 mL HisTrap FF column (GE Healthcare Life Sciences) that was pre-equilibrated with buffer A1 (20 mM Tris-HCl pH 8.0, 1 M NaCl, and 30 mM imidazole). After loading, the column was washed with 40 mL of buffer A1 to remove any weakly bound protein. DabA was eluted using a linear gradient of 100% buffer A1 to 100% buffer B (20 mM Tris-HCl pH 8.0, 1 M NaCl, and 250 mM imidazole) over 40 mL while collecting 5 mL fractions, keeping flow rate at 2 mL/min. Fractions were assessed for purity using SDS-PAGE. The fractions that were at least 90% pure were combined. The protein was concentrated to 2 mL using an Amicon Ultra-15 30 kDa cutoff concentrator and further purified at a flow rate of 1 mL/min with a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol. The protein was concentrated and either immediately used for assays or aliquoted, frozen in dry ice, and stored at -80 ^{o}C.

DabA functional assays

DabA enzyme assays were conducted in 50 mM HEPES (pH 8.0), with 100 mM KCl, 10% glycerol, 10 mM $MgCl_2$, 500 μM geranyl pyrophosphate (or other isoprenoid), 500 μM L-glutamic acid, and 10 μM of purified DabA enzyme. Total reaction volume for small scale assays was brought to 100 μL with MilliQ water. Small scale assays were run at room temperature for ~6 hours. Reactions were quenched with 100 μL of ice cold HPLC-grade MeOH and centrifuged in a table top centrifuge at 15000 xg and 4 ^{o}C for 20 minutes. Supernatant was then removed and subjected to subsequent HPLC and/or LCMS analysis.

Scaled up DabA reactions

Scaled up reactions for NMR characterization were run in 10 mL volume with 10 mg of GPP and 10 mg of L -Glu added as dry weight, also in 50 mM HEPES (pH 8.0) with 100 mM KCl, 10% glycerol and 10 mM $MgCl_2$. Scaled reactions were allowed to run for ~8 hours at room temperature with stirring and were quenched with 10 mL ice cold HPLC-grade MeOH, followed by centrifugation at 10000 xg and 4 ^{o}C for 20 minutes. Supernatant was concentrated *in vacuo* to remove methanol and the remaining aqueous reaction solution was frozen and lyophilized overnight.

HPLC purification of DabA product *N*-geranyl-L-glutamic acid (**2**)

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (5 min), 30 – 45% B (10 min), 45 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **2** (retention time 15.2 minutes) were pooled, concentrated *in vacuo* and lyophilized, affording **2** as a white solid. ^{1}H NMR (600 MHz, D_2O): δ 5.21 (t, J = 7.3 Hz, 1H), 5.12 (t, J = 6.7 Hz, 1H), 3.64 (t, J = 6.9 Hz, 2H), 3.57 (dd, J = 6.6, 5.7 Hz, 1H), 2.49 – 2.39 (m, 2H), 2.12 – 2.05 (m, 5H), 2.02 (ddt, J = 14.5, 6.8, 6.8 Hz, 1H), 1.66 (s, 3H), 1.64 (s, 3H), 1.57 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 176.6, 171.9, 146.4, 132.5, 122.4, 111.5, 58.5, 42.5, 37.6, 29.7, 24.0, 23.9, 23.5, 15.7, 14.3; HRMS (ESI) Calculated for $C_{15}H_{24}NO_4$ 282.1711, found 282.1711 $(M-H)^-$.

Marfey's Analysis of DabA assays

DabA reactions were set up as previously described except using 5 mM GPP and 5 mM L -glutamic acid (Glu), D-Glu, or racemic DL-Glu (2.5 mM each enantiomer) and incubated at room temperature for 24 h. Aliquots (10 μL) of the DabA reaction mixture were diluted with MilliQ-water to a final concentration of 1 mM. Aqueous 1 mM standard solutions of L-Glu, D-Glu, L-NGG, D-NGG, and 2 mM racemic Glu and NGG (1 mM each enantiomer) were also prepared. Saturated sodium bicarbonate (20 μL) was added to aliquots (50 μL) of each DabA reaction mixture or standard, followed by 100 μL of a freshly-prepared 1% w/v solution of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (L-FDAA) in acetone. Reactions were incubated at 37 °C for 90 minutes, then quenched by the addition of 25 μL of 1 N HCl. Reactions were centrifuged (13000 x *g*, 2 min) and the clarified supernatant (10 μL for all assays except 20 μL for DabA reaction with DL-Glu) was analyzed by analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) at a flow rate of 1 mL/min using the following method: 5% B (5 min), 5 – 95% B (20 min), 95 – 5% B (1 min), 5% B (4 min), where A = 0.1% aqueous trifluoroacetic acid, and B = 0.1% trifluoroacetic acid in acetonitrile.

Construction of vector PtpBAD-CTHF

PtpBAD-YFP was first built by amplifying the araC-pBAD region (70048-71265) and the rrnB terminator (4171760-4171962) with primer sets PtpBAD-CTHF_001-2 and PtpBAD-CTHF_003-4 from the K12 genome and assembled into PtpBAD-YFP by ligating the three fragments with an opened portion of the PtPBR1 backbone (*39*). This PtPBR-1 fragment was generated by amplifying the backbone with primer set PtpBAD-CTHF_005-6 in order to eliminate the Amp resistance marker for replacement with the expression cassette. These four fragments were Gibson assembled with a PCR-generated fragment encoding YFP (primers PtpBAD-CTHF_007-8) and containing overhangs to the promoter and terminator fragments.

These four fragments were assembled and after colony PCR and restriction mapping, the vector was validated by Sanger sequencing. In order to generate PtpBAD-CTHF, a synthetic ultramer (uCTHF) was designed based on pNIC-CTHF vector previously described (*43*) but replacing the TEV cleavage site with thrombin. In order to open the backbone and eliminate the YFP gene, primers PtpBAD-CTHF_002 and PtpBAD-CTHF_005 are used to amplify the entire backbone as a single piece. Primer PtpBAD-CTHF_002 amplifies from the 3'end of the pBAD promoter and primer PtpBAD-CTHF_005 amplifies from rrnB terminator and the reaction generates a product of 8,239 bp. The two ultramers encoding the affinity tag region were duplexed according to the manufacturer's protocol (IDT) and the dsDNA fragment was then ligated with the backbone fragment obtained with primers PtpBAD-CTHF_002 and PtpBAD-CTHF_005. This vector was then validated via Sanger sequencing (Eurofins) and designated PtpBAD-CTHF. This vector contains a XhoI site in the uCTHF region for linearization, and proteins assembled into the XhoI site with appropriate overhangs are fused to a carboxy-terminal tag of 21 residues (ALVPRS*GHHHHHHDYKDDDDK) including a thrombin cleavage site (marked with a *), hexahistidine (His_6) and FLAG.

dabC cloning into PtpBAD-CTHF

PtpBAD-CTHF was digested with XhoI (New England Biolabs). Next, *dabC* was amplified with Gibson overhangs with homology to PtPBAD-CTHF using primer set DabC-CTHF_001-2 with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. This amplicon was then assembled into the PtpBAD backbone in order to introduce the C-terminal His_6-FLAG (CTHF) affinity tag and place the CDS under the control of the arabinose-inducible promoter. The assembled vector was then transformed into TOP10 (ThermoFisher) chemically competent *E. coli*, which were then grown on LB plates with tetracycline (10 μg/mL) Following colony PCR using primer set PtpBAD-CTHF_col001-2 and OneTaq 2X MM with standard buffer (New England Biolabs), plasmid isolation using the ZR Plasmid Miniprep Classic Kit (Zymo Research) and Sanger sequencing (Eurofins) of PtpBAD-DabC-CTHF, we transformed the vector into BL21 chemically competent *E. coli* (New England Biolabs), plated on LB with tetracycline (10 μg/mL) and used these transformants for subsequent DabC expression.

DabC expression in BL21 *E. coli*

We inoculated 1 L of Terrific Broth media containing tetracyclin antibiotic (10 μg/mL) with 10 mL of PtpBAD-DabC-CTHF in BL21 grown overnight at 37 ^{o}C, 200 rpm to stationary. Each liter of culture was grown in a shaking refrigerated incubator at 37 ^{o}C, 200 rpm until OD_{600} reached ~0.6. At this time, the temperature on the refrigerated shaker was changed to 18 ^{o}C and flasks were allowed to adjust to temperature for approx. 30 minutes, after which cultures were induced by addition of 5g L-arabinose to a final concentration of 0.5% (w/v). Growth at 18 ^{o}C was continued overnight (~12-18 hr) until cells were harvested by centrifugation at 8000 xg for 5 minutes.

DabC purification

Harvested DabC expression cells were resuspended and lysed using the same procedure as DabA expression cells. Due to weak binding of this DabC construct, the cleared lysate was loaded onto a 5 mL HisTrap FF column (GE Healthcare Life Sciences) that was instead pre-equilibrated with resuspension buffer A2, which lacks imidazole (1 M NaCl and 20 mM Tris-

HCl pH 8.0). After loading, the column was washed with 40 mL of buffer A2. The column was further washed in a stepwise manner with 10 mL of increasing buffer B concentrations (2, 4, 6, 8, 10, 12, and 14% buffer B). DabC was then eluted using a linear gradient to 100% buffer B over 40 mL while collecting 5 mL fractions. Fractions were assessed for purity using SDS-PAGE. The fractions that were at least 90% pure were combined and 2 mM EDTA pH 8.0 was added to remove any co-purified metals. DabC was concentrated, further purified by size exclusion chromatography, and stored in an identical manner to DabA.

DabC assay with *N*-prenylated-L-glutamate substrates

DabC enzyme assays were conducted in 100 mM HEPES (pH 8.0) with 100 mM KCl, 10% glycerol, 1 mM L-ascorbate, and 6.25 mM α-ketoglutarate. *N*-geranyl-L-glutamate (**2**) was added to 1 mM concentration, followed by 25 μM DabC purified as above and 25 μM $FeSO_4$ and allowed to run overnight (~18 hours). Total volume for the reaction is approximately 100 μL. After 18 hours, additional L-ascorbate (approx. 1 mM), DabC (approx. 25 μM) and $FeSO_4$ (approx. 25 μM) was added and allowed to continue for an additional 18-24 hours. The reaction was then quenched with 1 eq. of ice cold MeOH and centrifuged at 15000 x g and 4 ^{o}C for 20 minutes. Reactions were set up in an analogous manner for 7'-carboxy-*N*-geranyl-L-glutamate (**3**), but showed complete substrate consumption after 3 hours and didn't require additional DabC or $FeSO_4$.

DabA and DabC coupled assays to isolate **5a-c**

DabA and DabC coupled assays were carried out at both the 100 μL and 1 mL scale. Reaction contents for both were as follows: 100 mM HEPES pH 8.0, 100 mM KCl, 10% Glycerol, 10 mM $MgCl_2$, 1 mM ascorbate, 6.25 mM μ-KG, 500 μM GPP, 500 μM L-Glu, 10 μM DabA, 25 μM DabC, and 25 μM $FeSO_4$. After 18 hours, additional ascorbate (approx. 1 mM), DabC (approx. 25 μM) and $FeSO_4$ (approx. 25 μM) was added and allowed to continue for an additional 18-24 hours. Both 100 μL and 1mL reactions were quenched with 1 eq. of ice cold MeOH and centrifuged at 15000 x g and 4 ^{o}C for 20 minutes. The clarified supernatant was concentrated *in vacuo* to remove methanol and the remaining aqueous reaction solution was frozen and lyophilized overnight. For scaled up reactions, we set up 10 x 1 mL reactions, as larger volume incubations with light stirring (e.g. 1 x 10 mL as in DabA scale up reactions), were less efficient as these "split up" scaled reactions.

Purification of **5a-c** products from DabA and DabC coupled assays

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 25% B (5 min), 25 – 40% B (10 min), 40 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **5a** (retention time 19.4 minutes), **5b** (retention time 20.6 minutes), and **5c** (retention time 20.4 minutes) were individually pooled, concentrated *in vacuo* and lyophilized. For improved purity, additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) over a gradient of 25 – 40% B (15 min) in the same solvent system. All products **5a-c** were obtained as white solids following lyophilization.

Dainic acid A (**5a**): ^{1}H NMR (600 MHz, D_2O): δ 5.43 (t, J = 7.2 Hz, 1H), 5.16 – 5.13 (m, 1H), 3.93 (d, J = 6.5 Hz, 1H), 3.63 (dd, J = 11.8, 8.1 Hz, 1H), 3.55 (dd, J = 7.8, 7.8 Hz, 1H), 3.42

(dd, *J* = 11.8, 8.1 Hz, 1H), 2.95 (ddd, *J* = 9.0, 6.7, 6.5 Hz, 1H), 2.76 (ddd, *J* = 16.2, 8.0, 7.2 Hz, 1H), 2.62 – 2.58 (m, 1H), 2.40 (dd, *J* = 15.6, 6.7 Hz, 1H), 2.30 (dd, *J* = 15.6, 8.5 Hz, 1H), 1.69 (s, 3H), 1.68 (s, 3H), 1.62 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 179.4, 173.5, 134.0, 130.2, 129.4, 122.2, 65.1, 46.8, 43.7, 40.3, 36.2, 26.3, 24.7, 21.4, 16.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1554 $(M-H)^-$

Dainic acid B (**5b**): ^{1}H NMR (600 MHz, D_2O): δ 5.19 (t, *J* = 7.2 Hz, 1H), 5.15 – 5.13 (m, 1H), 4.00 (d, *J* = 3.2 Hz, 1H), 3.55 (dd, *J* = 11.9, 7.5 Hz, 1H), 3.40 (dd, *J* = 11.4, 11.4 Hz, 1H), 3.01 (ddd, *J* = 8.8, 7.3, 2.5 Hz, 1H), 2.94 (dd, *J* = 7.9, 7.6 Hz, 1H), 2.78 – 2.75 (m, 2H), 2.20 (dd, *J* = 15.4, 6.5 Hz, 1H), 2.11 – 2.07 (m, 1H), 1.69 (s, 3H), 1.68 (s, 3H), 1.64 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 179.8, 173.4, 133.9, 129.8, 126.1, 122.3, 65.7, 47.0, 46.0, 42.2, 36.1, 26.4, 24.7, 16.9, 15.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1555 $(M-H)^-$

Dainic acid C (**5c**): ^{1}H NMR (600 MHz, D_2O): δ 5.19 (t, *J* = 8.6 Hz, 1H), 5.04 (s, 1H), 4.77 (s, 1H), 4.06 (d, *J* = 2.4 Hz, 1H), 3.58 (dd, *J* = 11.9, 7.6 Hz, 1H), 3.42 (dd, *J* = 11.8, 11.8 Hz, 1H), 3.04 – 2.95 (m, 2H), 2.25 (dd, *J* = 15.5, 6.0 Hz, 1H), 2.22 – 2.09 (m, 4H), 2.09 – 2.06 (m, 1H), 1.68 (s, 3H), 1.61 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 179.8, 173.7, 144.1, 134.0, 123.4, 111.6, 65.7, 45.9, 43.9, 41.5, 35.9, 35.4, 25.4, 24.5, 16.6; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1556 $(M-H)^-$

Scaled up DabC assay with 7'-carboxy-*N*-geranyl-L-glutamic acid (**3**) substrate

DabC enzyme assays were conducted as previously described in a 5 mL volume using 1.8 mg of 7'-carboxy-*N*-geranyl-L-glutamic acid (**3**) as a substrate with light stirring. The reaction was quenched as previously described after 5 hours, concentrated *in vacuo* and immediately subjected to RP-HPLC purification.

HPLC purification of DabC product isodomoic acid A (**4**)

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (15 min), 30 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **4** (retention time 12.3 minutes) were pooled, concentrated *in vacuo* and lyophilized, affording **4** as a white solid (1.5 mg, 84%).

Isodomoic acid A (**4**): ^{1}H NMR (600 MHz, D_2O): δ 6.67 (t, *J* = 7.2 Hz, 1H), 5.54 (t, *J* = 7.5 Hz, 1H), 3.97 (d, *J* = 7.3 Hz, 1H), 3.69 (dd, *J* = 11.7, 8.1 Hz, 1H), 3.63 (ddd, *J* = 7.8, 7.6, 7.6 Hz, 1H), 3.46 (dd, *J* = 11.7, 7.7 Hz, 1H), 3.01 (ddd, *J* = 12.7, 10.1, 5.4 Hz, 1H), 2.96 (ddd, 15.6, 7.7, 7.2 Hz, 1H), 2.87 (ddd, 15.8, 7.2, 7.0, 1H), 2.67 (dd, 16.5, 6.5 Hz, 1H), 2.50 (dd, 16.5, 8.6 Hz, 1H), 1.83 (s, 3H), 1.76 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 176.5, 172.9, 172.9, 141.0, 131.0, 128.1, 128.0, 65.1, 46.8, 42.4, 40.2, 34.1, 27.1, 21.3, 11.8; HRMS (ESI) Calculated for $C_{15}H_{20}NO_6$ 310.1296, found 310.1294 $(M-H)^-$

Construction of vector pBEVY-GL-dabD-CTHF-PmCPR1

DabD and PmCPR1 were heterologously expressed in *S. cerevisiae* using the vector pBEVY-GL, an expression vector featuring a bidirectional GAL1/10 promoter (*44*). pBEVY-GL was a gift from Charles Miller (Addgene plasmid #51225). The gene *dabD* was first cloned into PtpBAD-CTHF in order to introduce the C terminal His_6-FLAG affinity tag. Following colony PCR with OneTaq 2X MM polymerase with standard buffer (New England Biolabs) and primer set PtpBAD_col001-2, as well as Sanger sequencing (Eurofins), we then amplified *dabD*-

CTHF from the assembled PtpBAD-dabD-CTHF vector using primer set pBEVY-DabD_001-2 to incorporate Gibson overhangs specific to the pBEVY-GL vector. pBEVY-GL was then linearized with BamHI (New England Biolabs) and the dabD-CTHF amplicon was assembled into the linearized backbone using Gibson assembly in order to insert dabD-CTHF under the GAL1 promoter. Assembled vector was then transformed into Top10 chemically competent *E. coli* cells and plated on ampicillin selection (100 μg/mL). Colony PCR was performed using primer set pBEVY-DabD_col001-2 and OneTaq 2X MM polymerase with standard buffer. Positive clones were sent for Sanger sequencing (Eurofins) to confirm assembly of pBEVY-GL-dabD-CTHF.

Following sequence validation of pBEVY-GL-dabD-CTHF, we moved to clone *PmCPR1* under the GAL10 promoter of pBEVY-GL. *PmCPR1* was PCR amplified with primer set pBEVY-PmCPR1_001-2 to introduce Gibson overhangs specific to the vector pBEVY-GL. Following successful PCR amplification of *PmCPR1*, we used Gibson assembly to clone *PmCPR1* into pBEVY-GL-dabD-CTHF linearized with *SmaI* (New England Biolabs). This assembled vector was then transformed into TOP10 chemically competent *E. coli* cells and plated on ampicillin selection (100 μg/mL). Colony PCR was performed using primers pBEVY-PmCPR1_col001-2 and positive clones were sent for subsequent Sanger sequencing to confirm successful assembly of pBEVY-GL-dabD-CTHF-PmCPR1.

Heterologous expression of DabD and PmCPR1 in *Saccharomyces cerevisiae*

The vector pBEVY-GL-dabD-CTHF-PmCPR1 was then transformed into the protease-deficient *S. cerevisiae* strain BJ5464 (*45*) using standard lithium acetate transformation protocols (*46*). Transformants were then plated onto synthetic drop-out media (Sigma-Aldrich) without amino acid leucine (Leu) to allow selection for transformed yeast colonies. The empty vector pBEVY-GL was also similarly transformed into BJ5464 cells. Screening for DabD expression was performed by growing 50 mL cultures of individual transformants to ~1.0 OD_{600} in -Leu synthetic drop-out media with galactose as the only carbon source (*47*). Cultures were then spun down (8000 xg, 10 minutes, 4 ^{o}C and protein was extracted using a mild alkali treatment followed by resuspension in 200 μL of SDS sample buffer (*48*). Gels for western blots were run using 10 μL of resuspended sample along with the MagicMark XP Western Protein Standard (ThermoFisher) on NuPAGE 4-12% Bis-Tris protein gels (Invitrogen). Transfer to PVDF membrane (Invitrogen) was conducted in an XCell II Blot Module (Invitrogen). Membranes were then blocked and blotted with FLAG mouse monoclonal antibody (Invitrogen, 1:20000 dilution) using standard protocols, diluents and mouse secondary antibodies from the manufacturer (Invitrogen).

For large scale expression, cells were grown according to established protocols for yeast P450 expression (*47*). In short, a single colony was grown in 50 mL of synthetic dropout media (-leu) to stationary phase. A 1:100 dilution of overnight culture was made into 500 mL of YPGE media and cells were allowed to grow for 24-36 hours or until OD_{600} = 1.0. When cells reached the appropriate density, cultures were induced through the addition of 20% galactose (w/v) to a final concentration of 2%. Cells were then induced for at least 24 hours prior to harvesting by centrifugation (8000 xg, 5 minutes, RT).

Isolation of microsomes from *S. cerevisiae*

Following centrifugation of *S. cerevisiae* transformed with either pBEVY-GL-dabD-CTHF-PmCPR1 or empty pBEVY-GL, cells were lysed using the mechanical procedure from

(*47*). In short, cells were incubated in TEK buffer for 5-10 minutes at room temperature and then resuspended in a minimal volume of TES-B. An equal volume of 0.5 mm glass beads were added to the resuspended yeast and were subjected to vortexer-driven bead-beating for a total of 15 minutes of active lysis, with 15 seconds of vortexing followed by 15 seconds of incubation on ice, all performed in 50 mL conical tubes. Following confirmation of cell lysis by light microscopy, the supernatant was decanted and beads were washed with 2-3 volumes of TES-B. The supernatant fraction and bead washes were then combined and spun at 12000 xg for 30 minutes at 4 ^{o}C. The supernatant from the low speed spin was then subjected to ultracentrifugation to collect microsomal fractions (130000 xg, 90 minutes, 4 ^{o}C). Microsomal fractions were then resuspended in the smallest volume of TES-B possible to obtain concentrated microsomal samples.

Microsomal assays

Enzyme assays using ~550 μg of microsomal protein isolated from *S. cerevisiae* strain BJ5464 transformed with either pBEVY-GL-dabD-CTHF-PmCPR1 or empty pBEVY-GL were performed in 100 mM HEPES buffer (pH ~8.0) with 100 mM KCl, 10% glycerol, 1 mM NADPH, 10 μM FAD, 10 μM FMN and 500 μM *N*-geranyl-L-glutamate as substrate. Total reaction volume was brought to 100 μL through addition of MilliQ water. Assays ran for ~8 hours at room temperature, after which the reactions were quenched through addition of 100 μL of ice-cold HPLC grade methanol and centrifuged at max speed (~17000 xg) for 20 minutes at 4 ^{o}C. The supernatant was then removed for subsequent analysis.

DabC relative substrate consumption assay

DabC reactions were set up as previously described utilizing either 1 mM *N*-geranyl-L-glutamic acid (**2**) or 7'-carboxy *N*-geranyl-L-glutamic acid (**3**) as a substrate. Assays were set up in triplicate for each substrate in 500 μL volumes. At regular time intervals (0, 0.17, 0.5, 1, 2, 3, and 6 h), an aliquot (25 μL) of each assay was removed and quenched with a 25 μL solution of 1.0 mM phenol red in methanol. Quenched aliquots were mixed thoroughly, centrifuged at 13000 x *g* for 2 min, and the clarified supernatant (10 μL) was analyzed by analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) at a flow rate of 1 mL/min using the following method: 10% B (5 min), 10 – 25% B (5 min), 25 – 40% B (10 min), 40 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous trifluoroacetic acid, and B = 0.1% trifluoroacetic acid in acetonitrile. The ratio of the peak integrations at 210 nm corresponding to the substrates (**2** or **3**) or major products (**5a** or **4**) to the phenol red internal standard were compared at various time points to the t = 0 h substrate:phenol red ratio to relatively compare substrate consumption and product accumulation over time.

dabB Cloning

The pET28-MBP (Maltose Binding Protein)-TEV vector was linearized using PrimeStar MAX polymerase (TaKaRa) and primer set pET28-MBP_BB_001-2. The *dabB* gene was obtained by PCR with PrimeStar polymerase and primer set DabB_MBP_001-002 to incorporate the appropriate Gibson assembly overhangs. NEBuilder HiFi DNA Assembly mix was used to generate the *dabB* pET28-MBP construct yielding an N-terminal His_6-MBP-His_6-DabB fusion. After confirming the *dabB* sequence by Sanger sequencing, the plasmid was transformed into *E. coli* BL21 cells for subsequent expression.

DabB Expession and Purification

DabB was expressed and purified by a 5 mL HisTrap FF column (GE Healthcare Life Sciences) column using the same procedure as described for DabA. The DabB containing fractions were collected and concentrated to 2 mL and purified by using a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol at a flow rate of 1 mL/min. The DabB containing fractions were collected and the MBP tag was removed by the addition of TEV protease and incubated at 4 ^{o}C overnight. The cut DabB was separated from MBP using a HiLoad 16/60 Superdex 75 prep grade gel filtration column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol at a flow rate of 1 mL/min. The DabB containing fractions were collected, concentrated, and aliquoted before storage at -80 ^{o}C (Fig. S5b).

DabB Activity Assays

DabB activity assays were completed using 10 μM DabB and 500 μM substrate with a buffer containing 100 mM HEPES pH 8.0, 100 mM KCl, and 10% glycerol in 100 μL of total volume. The assays were incubated for 18 h at 23 ^{o}C and quenched with 1 equivalent of methanol. After centifuged to remove precipitated protein, they were analyzed by LCMS method B on the Agilent Technologies 1200 Series system with a diode-array detector coupled to an Agilent Technologies 6530 accurate-mass Q-TOF LCMS. Assays containing DabD were completed using the microsomal assay protocol with the addition of 10 μM DabB.

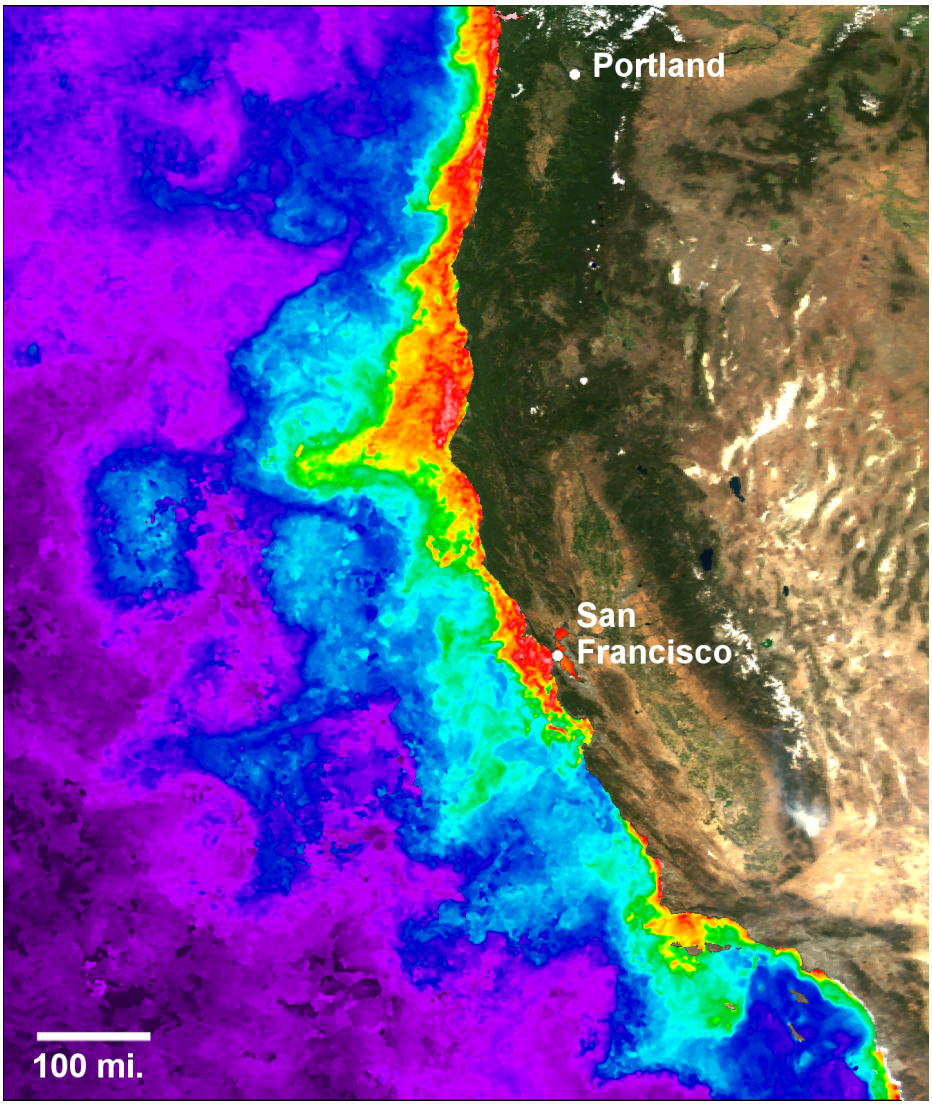


Fig. S1.
Extent of the Pacific Northwest *Pseudo-nitzschia* bloom in late-spring 2015. Gradation of colors purple-blue-green-yellow-red-pink in ocean areas shows increasing chlorophyll concentration towards the coast. Land is filled with quasi-true color data from MODIS-Aqua. Ocean chlorophyll is merged with data from MODIS-Aqua, MODIS-Terra and VIIRS-NPP and then composited over 15 days (May 1st – May 15th 2015). White areas are clouds. Image was processed using methods described previously (*49*). Satellite data courtesy of NASA Ocean Biology Processing Group (OBPG) and LAADS-DAAC. Image courtesy of Dr. Mati Kahru, Scripps Institution of Oceanography.

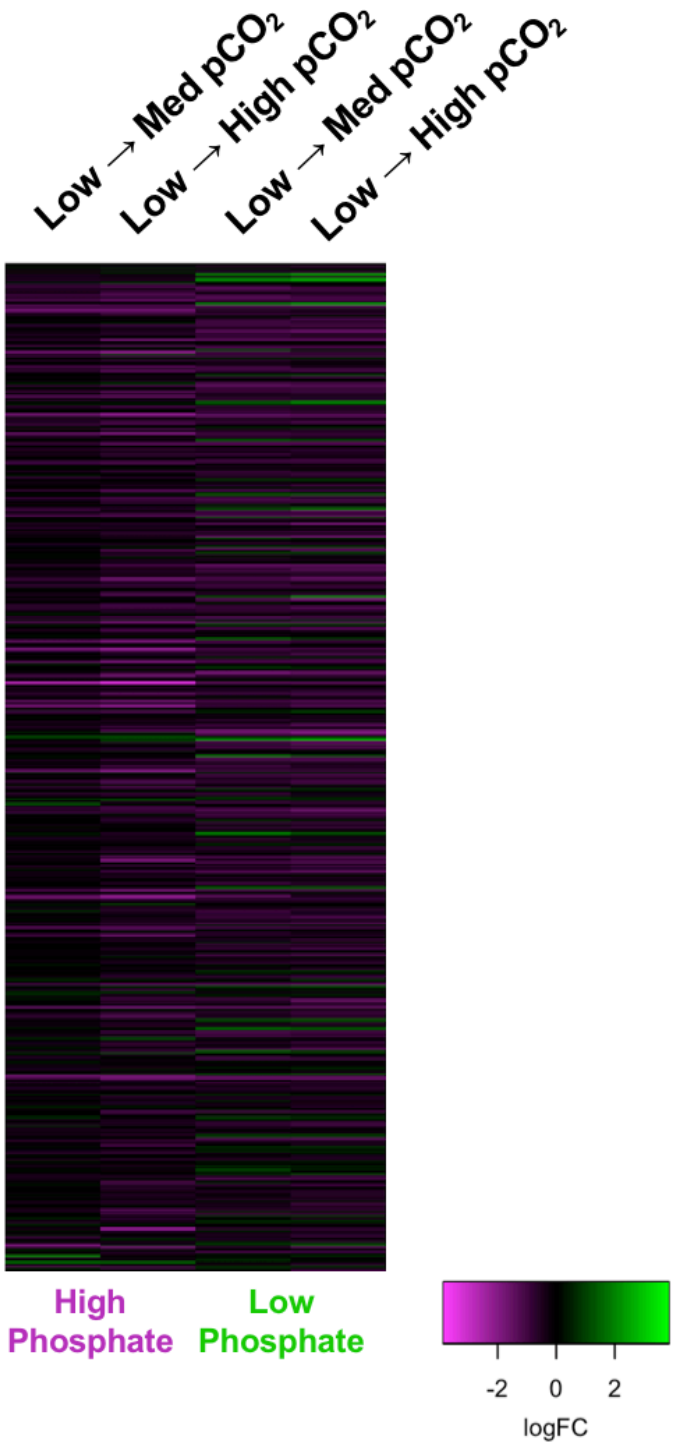


Fig. S2.

Heat map showing differential expression profiles of the 483 transcripts upregulated under phosphate limitation. Log2FC values come from DE analysis of these transcripts with respect to a variety of increasing pCO_2 conditions (low = 220 ppm, med = 400 ppm, high = 730 ppm) with phosphate held constant (high phosphate = 20 mM , low phosphate = 0.5 mM). Ordering of the rows on the heatmap is consistent with **Table S2**.

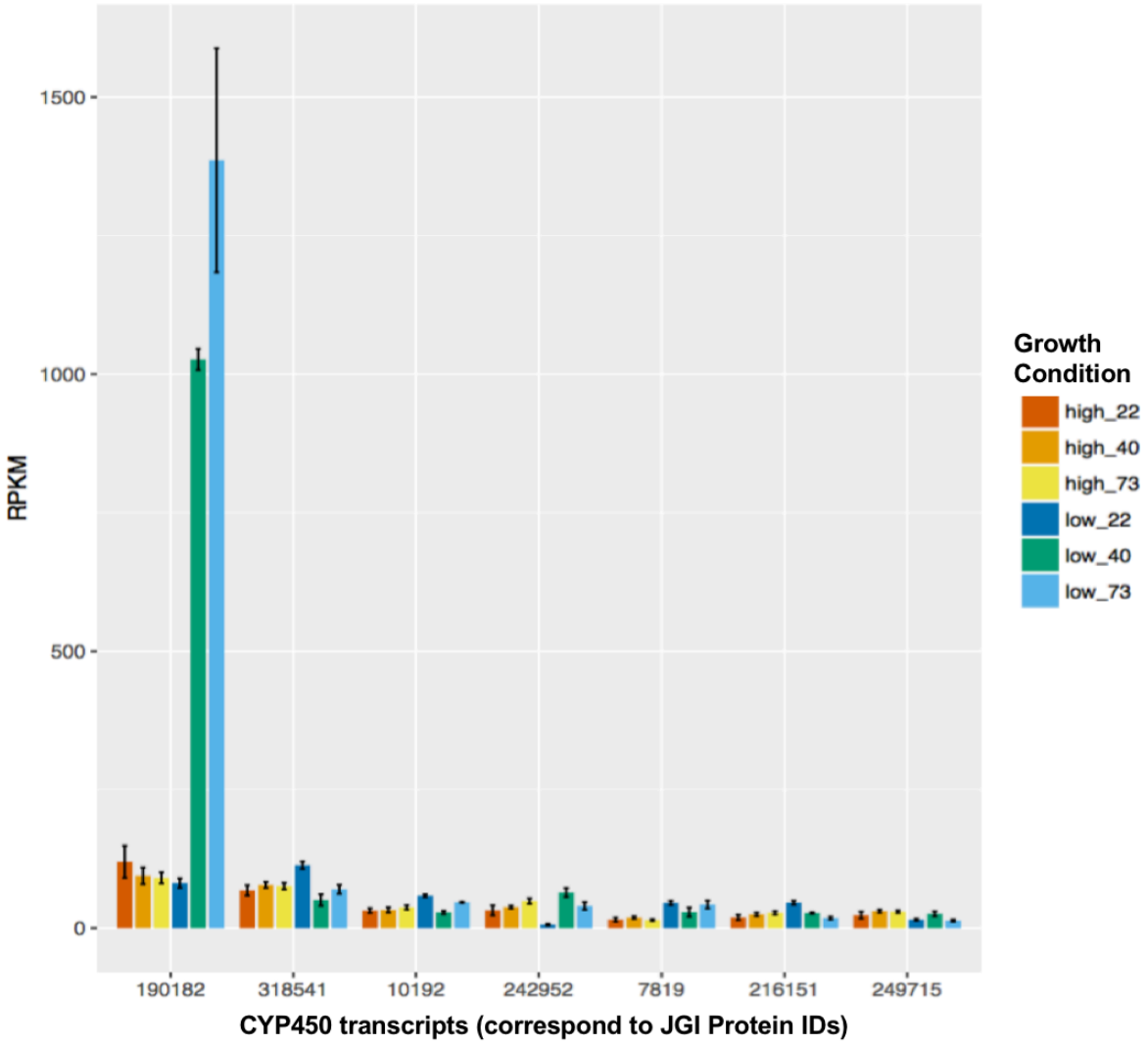


Fig. S3.

Histogram for average expression (RPKM) of the top 7 most highly expressed CYP450 genes in the *P. multiseries* transcriptome under the culturing conditions of high (20 μM) or low (0.5 μM) phosphate as well as 220 ppm, 400 ppm, or 730 ppm pCO_2. RPKM for each culturing condition was averaged across n=3 rRNA-depleted libraries, with error bars representing standard error. Of the other 13 CYP450 transcripts, trends in RPKM expression were either highly similar to those transcripts represented above (with PID: 190182 *dabD* as an exception) or were not expressed at a high enough level to be visualized on this chart. Full RPKM and differential expression analysis is available in **Table S4**.

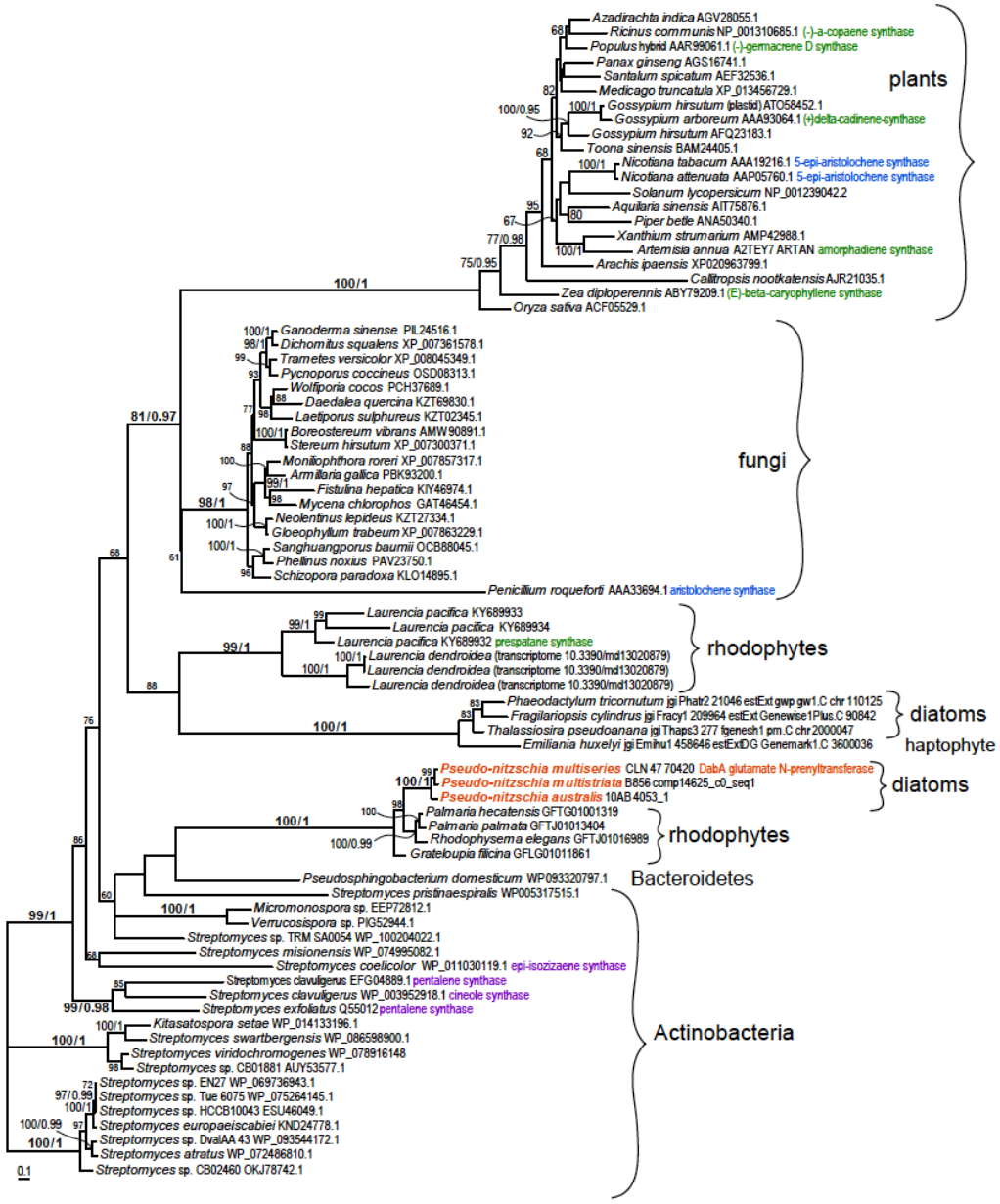
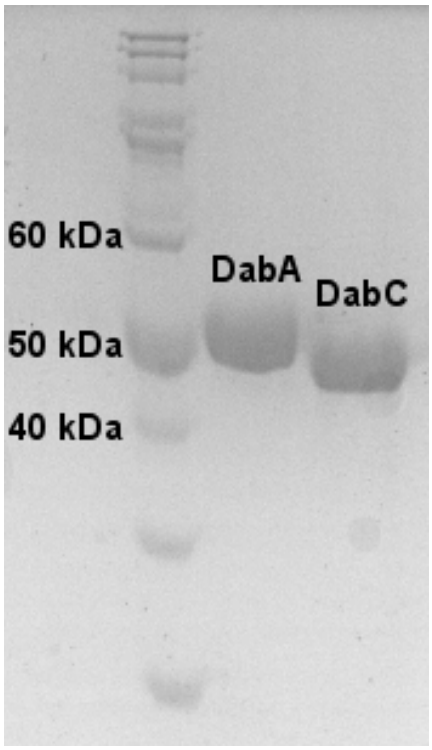


Fig. S4.

Phylogeny for *dabA*, red algal homologues and several putative and characterized sesquiterpene synthases. Due to the divergent amino acid sequence of DabA, candidate proteins were selected to represent a variety of diverse putative sesquiterpene synthases and were not chosen based on sequence similarity. Sequences were aligned using the CLUSTAL algorithm in frame of SeaView program (*50*). Alignment was manually edited by exclusion of gaps and ambiguously aligned regions. Maximum likelihood phylogenetic analysis was performed using IQTree v1.5.3 (*51*) under gamma-corrected LG4M matrix, determined as the best-fit model by the -TEST mode of IQTree (Akaike information criterion). The best topology and branch support values were inferred from 1,000 ultrafast bootstrap replicates (*52*). In parallel, Bayesian inference under the gamma-corrected LG matrix (4 gamma categories) was performed in PhyloBayes v3.3b (*53*).

The best topology was inferred after two independent Markov chain Monte Carlo (MCMC) chains reached 0.2-level convergence and the minimum effective size of the model parameters exceeded 100.

a)



b)

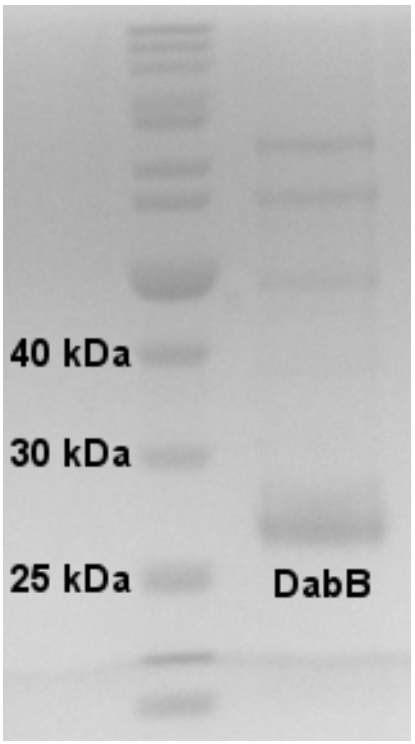
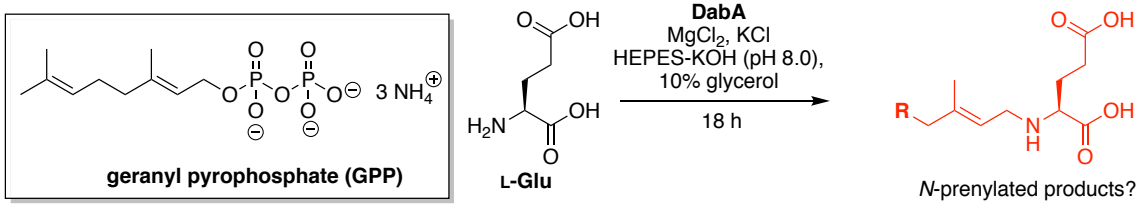


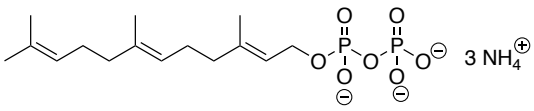
Fig. S5.

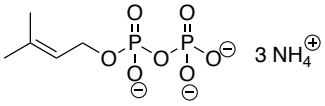
12% SDS-PAGE loaded with **a)** EZ-Run *Rec* Protein Ladder (Fisher Bioreagents), DabA, and DabC and **b)** EZ-Run *Rec* Protein Ladder and DabB.

a) Summary of organic pyrophosphate experiments

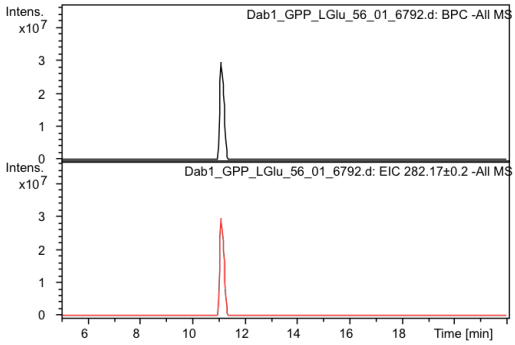


alternative pyrophosphate substrates tested:

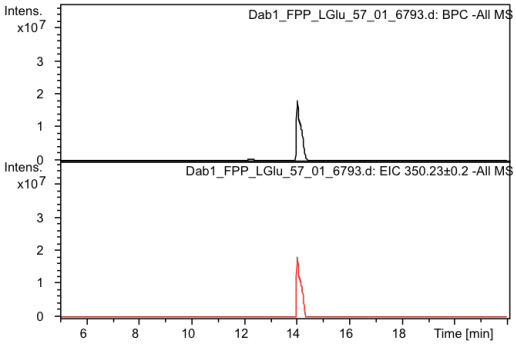




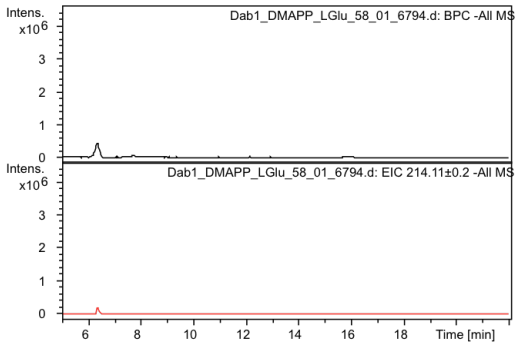
b) geranyl pyrophosphate (GPP)



c) farnesyl pyrophosphate (FPP)

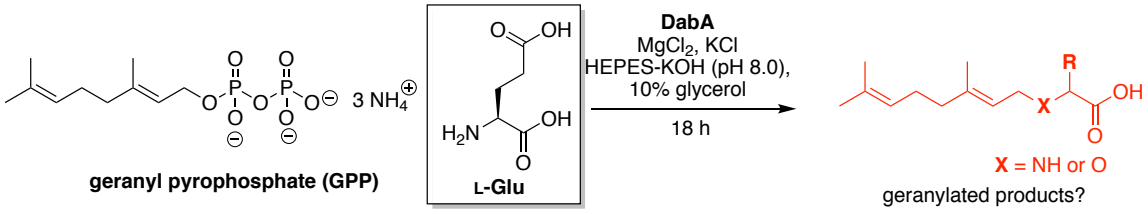


d) dimethylallyl pyrophosphate (DMAPP)*



* traces shown at 10x magnification

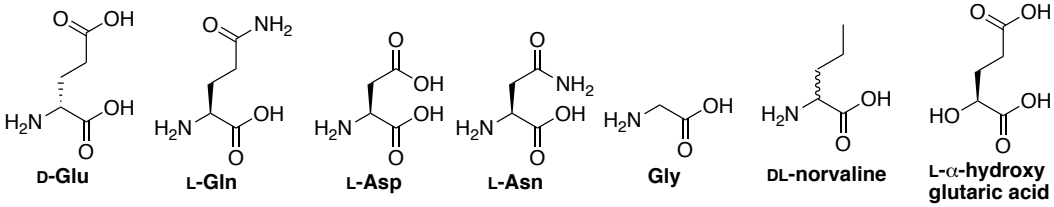
e) Summary of amino acid experiments



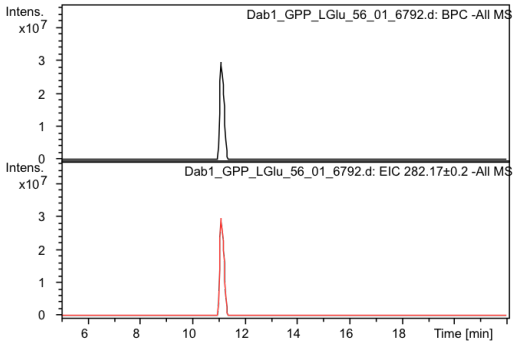
geranyl pyrophosphate (GPP)

geranylated products?

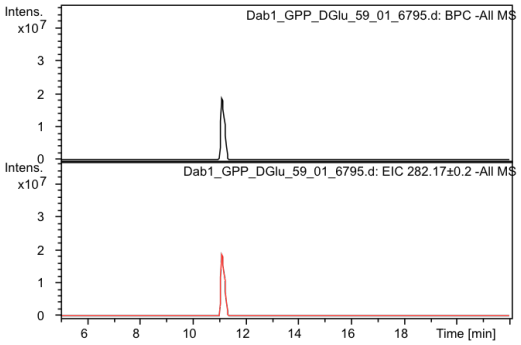
alternative amino acid substrates tested:



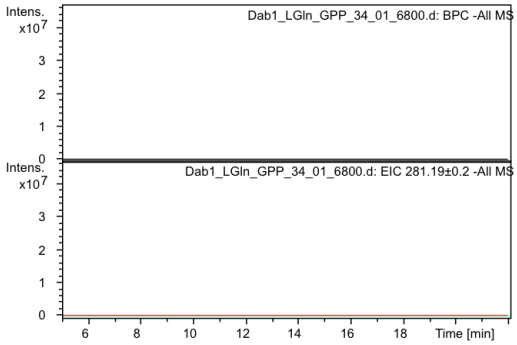
f) L-glutamic acid



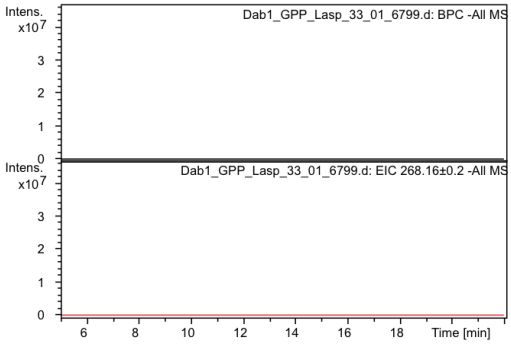
g) D-glutamic acid



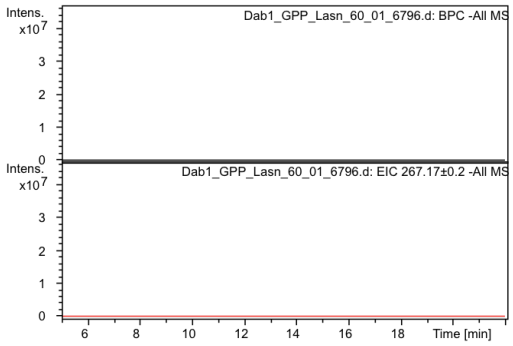
h) L-glutamine



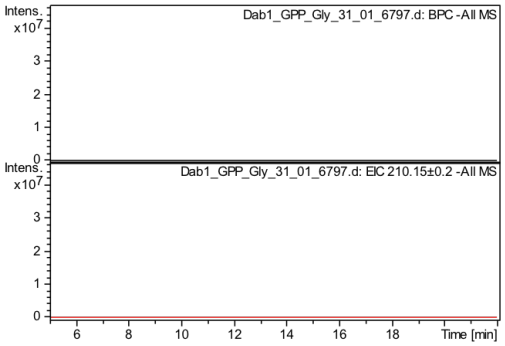
i) L-aspartic acid



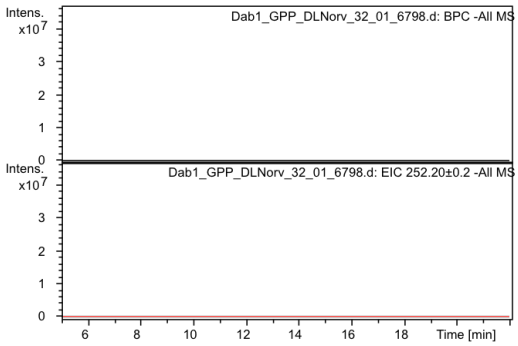
j) L-asparagine



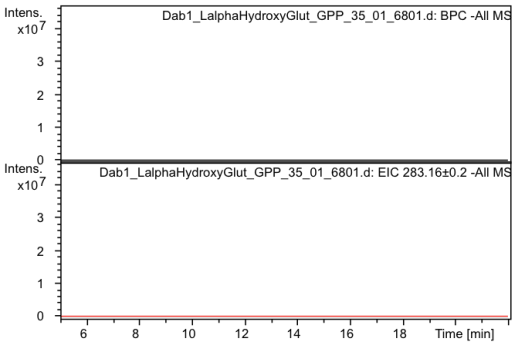
k) glycine



l) DL-norvaline



m) L-α-hydroxyglutaric acid



n) Marfey's analysis of L-Glu and D-Glu substrate specificity

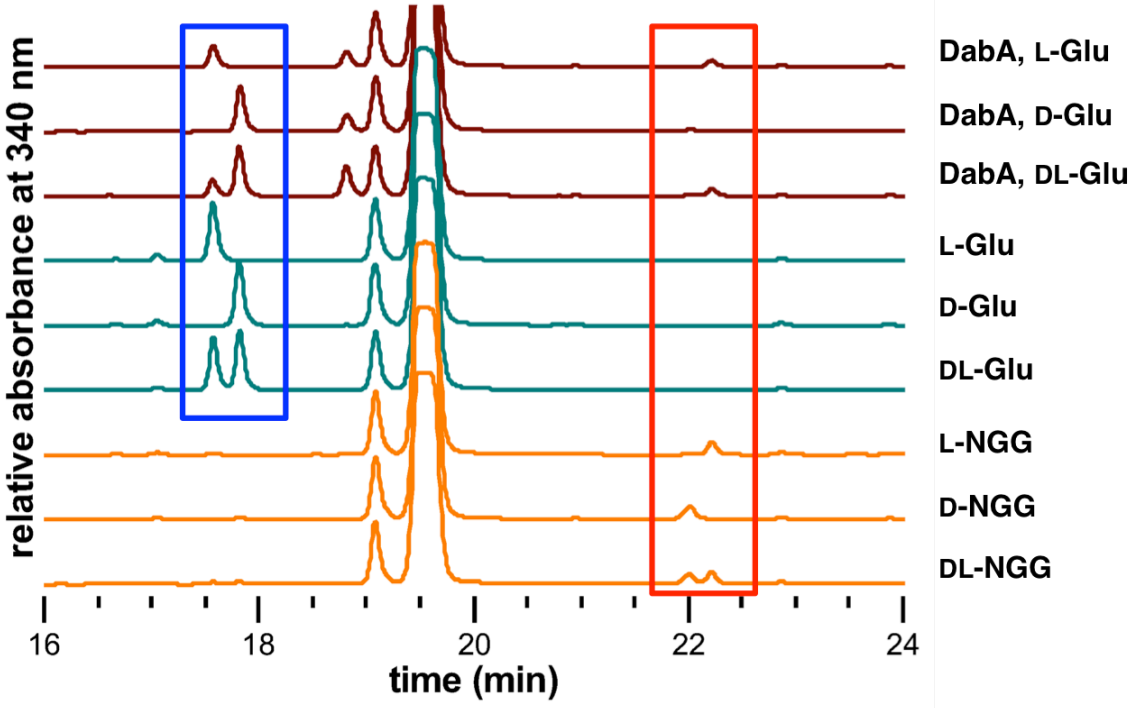
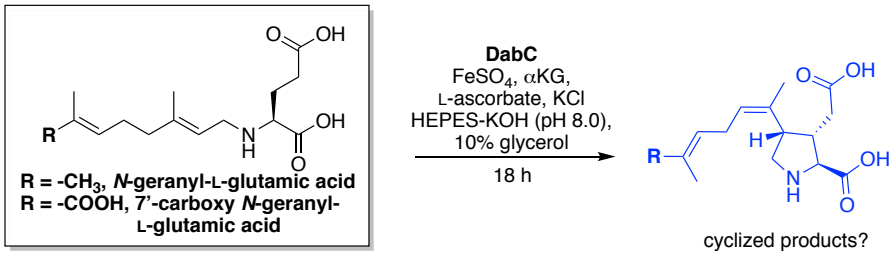
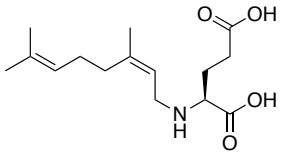


Fig. S6.
a-d) DabA reactions were set up as previously described using various organic pyrophosphates (0.5 mM) and L-glutamic acid (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabA pyrophosphate substrate specificities with L-Glu using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms for anticipated products (EIC ± 0.2 *m/z*, red). **e-m)** DabA reactions were set up as previously described using geranyl pyrophosphate (0.5 mM) and various amino acids/amino acid analogues (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabA amino acid/analogue substrate specificities with GPP using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms for anticipated products (EIC ± 0.2 *m/z*, red). **n)** RP-HPLC (λ = 340 nm) analyses of L-FDAA derivatized DabA reactions with L-, D-, and racemic Glu, and comparison to similarly derivatized Glu (blue box) and NGG (red box) standards. DabA has a ~4:1 preference for L-Glu over D-Glu when incubated with a racemic mixture.

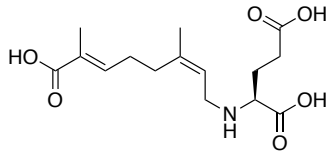
a) DabC stereochemical selectivity experiments



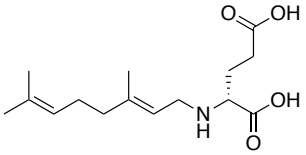
alternative *N*-prenylated glutamic acid substrates tested:



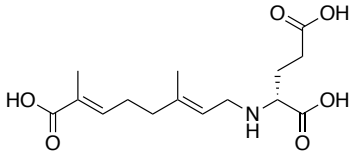
***N*-neryl-L-glutamic acid**



7'-carboxy *N*-neryl-L-glutamic acid

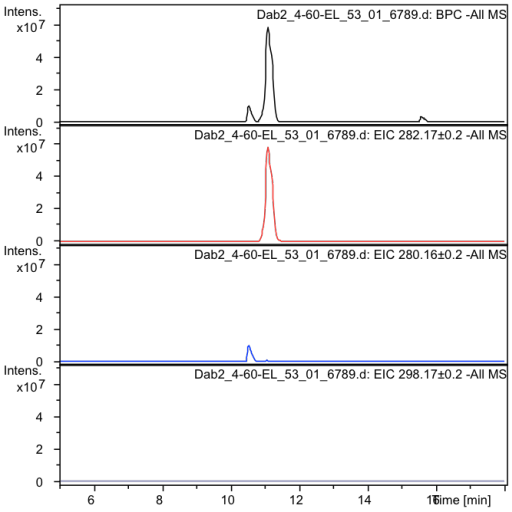


***N*-geranyl-D-glutamic acid**

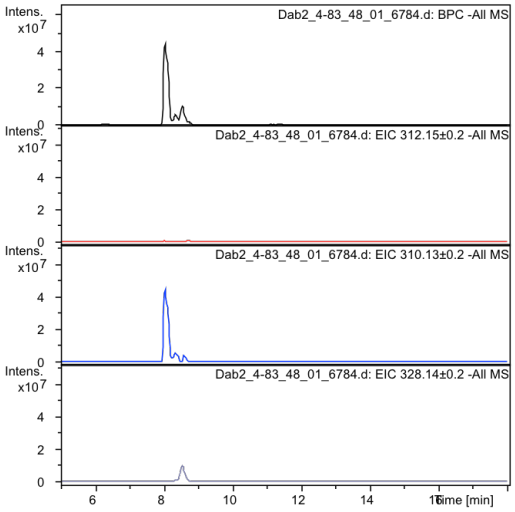


7'-carboxy *N*-geranyl-D-glutamic acid

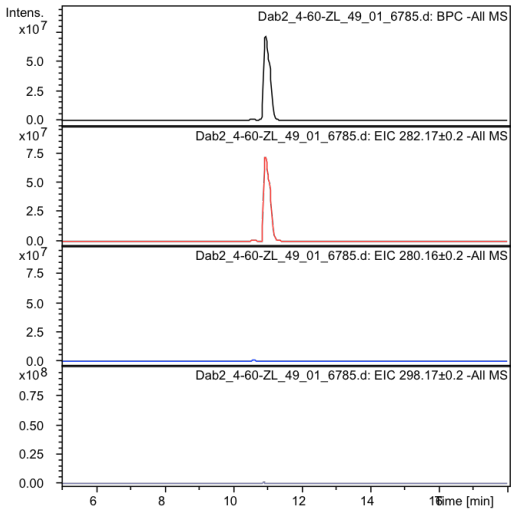
b) *N*-geranyl-L-glutamic acid



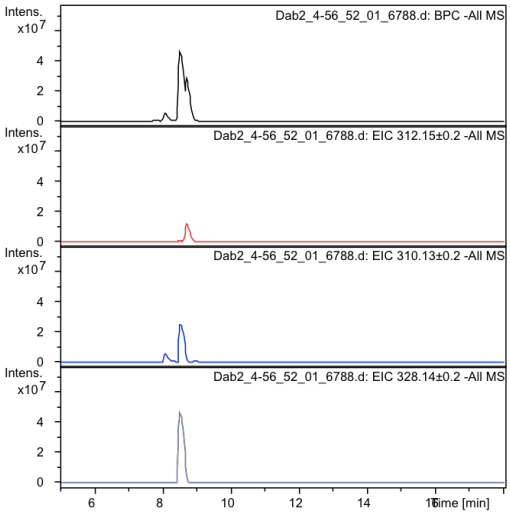
c) 7'-carboxy *N*-geranyl-L-glutamic acid



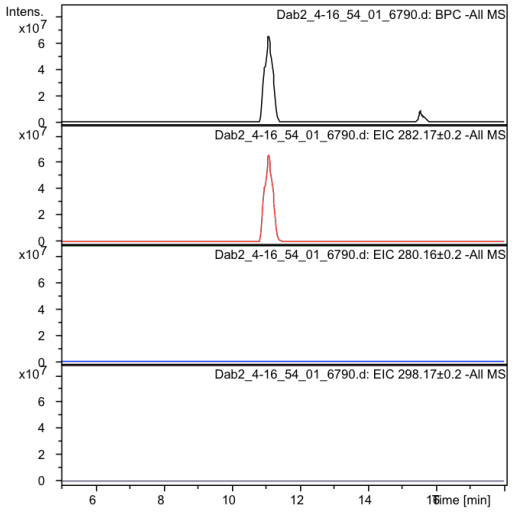
d) *N*-neryl-L-glutamic acid



e) 7'-carboxy *N*-neryl-L-glutamic acid



f) *N*-geranyl-D-glutamic acid



g) 7'-carboxy *N*-geranyl-D-glutamic acid

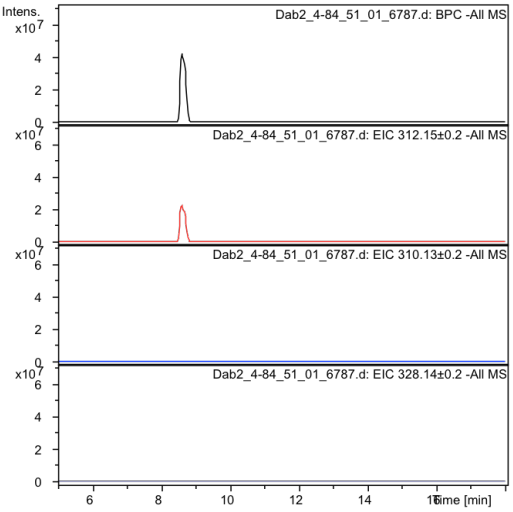


Fig. S7.

DabC reactions were set up as previously described using *N*-prenylated glutamic acid analogues (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabC substrate specificities with various *N*-prenylated glutamic acids using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms (EIC ± 0.2 *m/z*) for starting materials (red), cyclized (blue), or hydroxylated (grey) products.

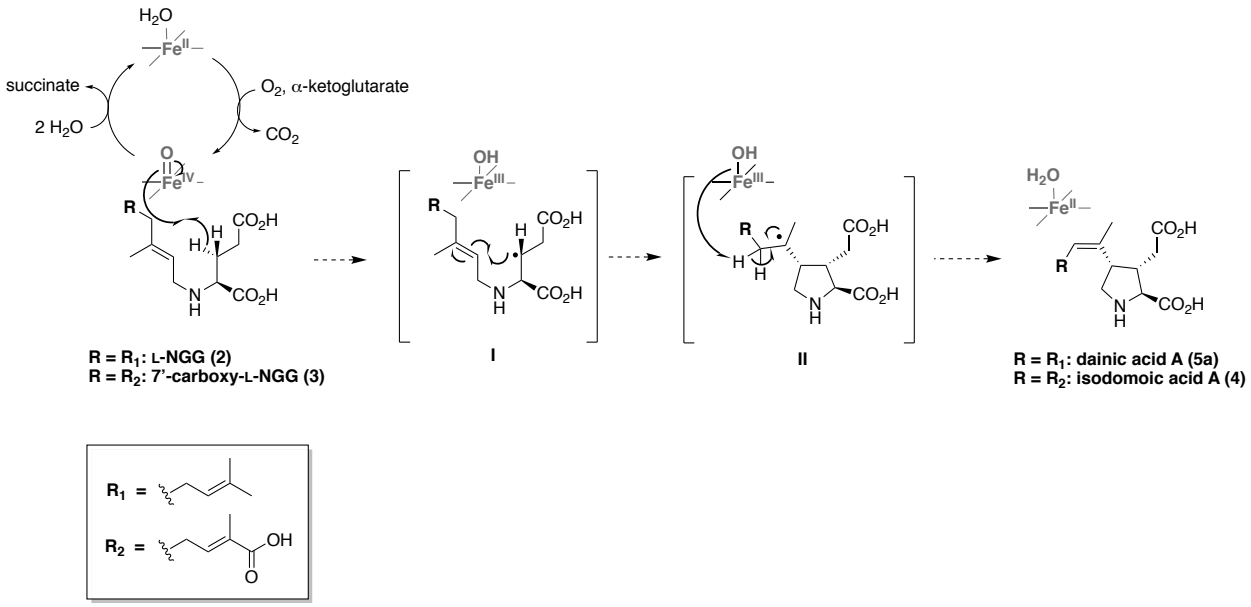
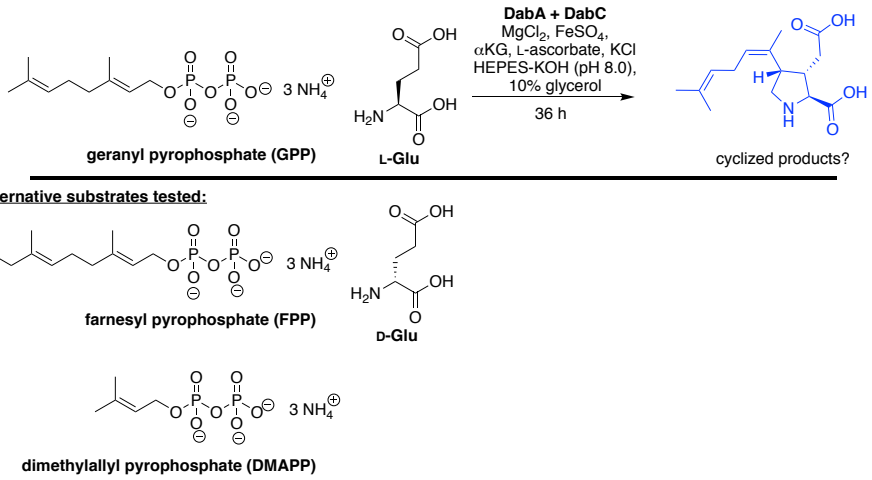


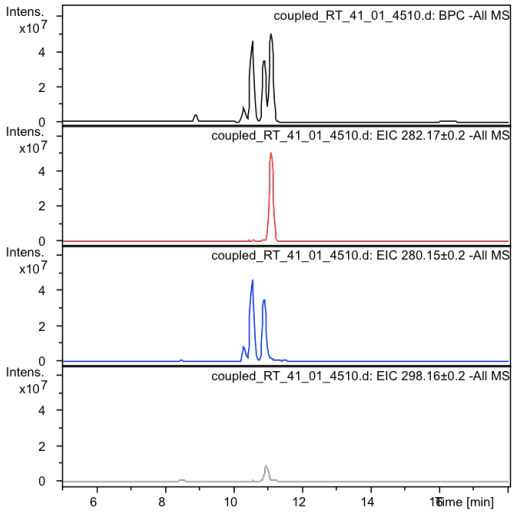
Fig S8.

Proposed mechanism for DabC cyclization of *N*-prenylated glutamic acids **2** and **3** to major products dainic acid A (**5a**) and isodomoic acid A (**4**) respectively. Following radical abstraction of a glutamic acid β-hydrogen (intermediate I), a 5-*exo*-*trig* cyclization generates the pyrrolidine ring and tertiary radical (intermediate II). This can be resolved via a hydrogen atom transfer mechanism from the intermediate Fe(III)-OH species (*54*, *55*) to form the *cis*-double bond of the major products.

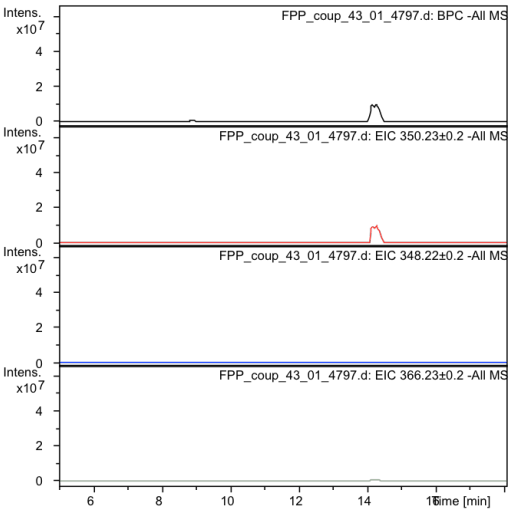
a)



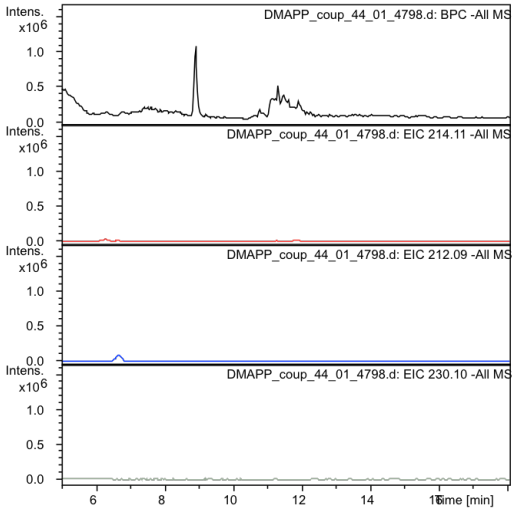
b) L-glutamic acid + GPP



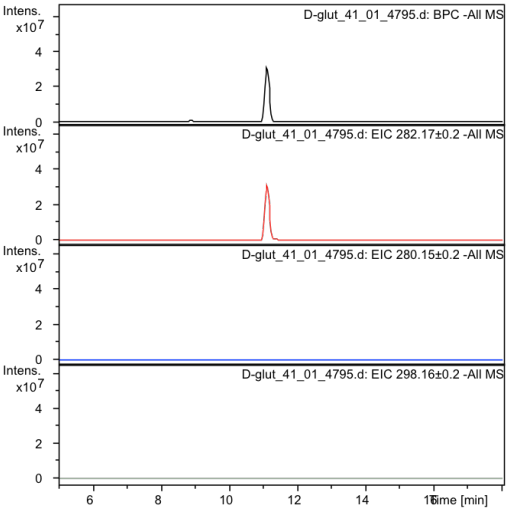
c) L-glutamic acid + FPP



d) L-glutamic acid + DMAPP*



e) D-glutamic acid + GPP



*traces shown at 40x magnification

Fig. S9.
DabA and DabC coupled reactions were set up as previously described using various organic pyrophosphates (0.5 mM) or amino acids (0.5 mM) that showed individual activity in DabA assays. Negative mode LCMS chromatograms of DabA and DabC coupled reactions after 36 h using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms (EIC ± 0.2 *m/z*) for *N*-prenylated glutamic acid (red), cyclized (blue), or hydroxylated (grey) products.

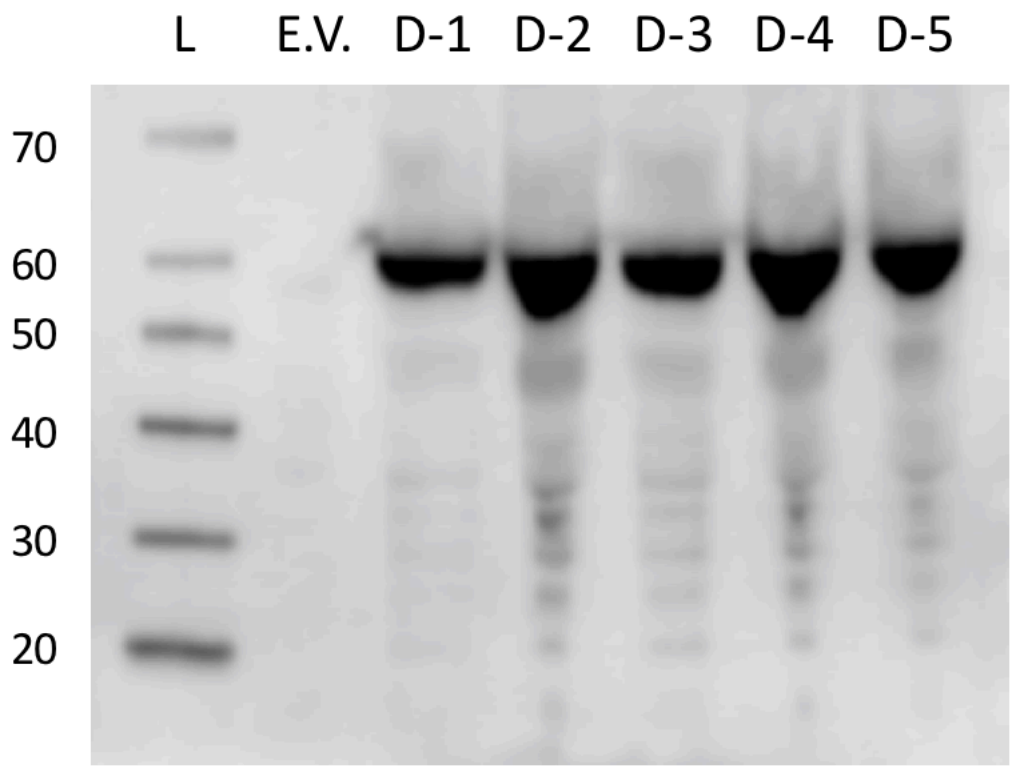


Fig. S10.
Western Blot for DabD using anti-FLAG mouse antibody (Sigma Aldrich) on whole cell lysates of *S. cerevisiae* BJ5464 encoding either pBEVY-GL-dabD-CTHF-PmCPR1 (D-1 through D-5) or an empty pBEVY-GL control (E.V). Ladder (L) is the Magic Mark XP Protein Standard (Thermo Fisher)

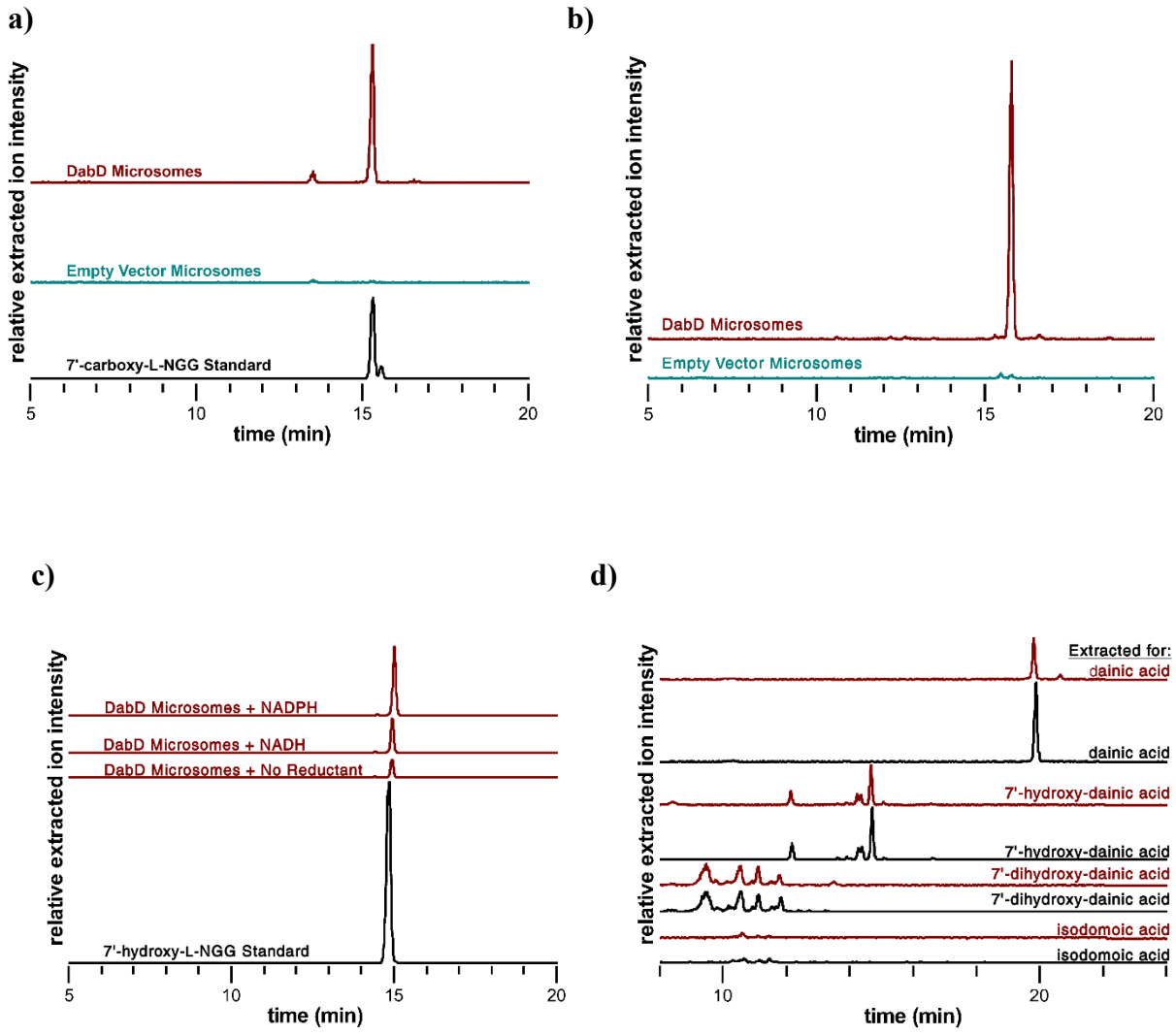


Fig. S11.
Negative mode LCMS chromatograms of DabD microsome assays using LC Method B, showing the extracted ion chromatograms (EIC ± 0.01 *m/z*) for **a)** 7'-carboxy-L-NGG and **b)** 7'-dihydroxy-L-NGG. **c)** Negative mode LCMS chromatograms of DabD microsome assays in the presence of different reducing agents. Traces represent the extracted ion chromatograms (EIC ± 0.01 *m/z*) for 7'-hydroxy-L-NGG. **d)** Negative mode LCMS chromatograms of DabD microsome assays with dainic acid substrate. Traces represent the extracted ion chromatograms (EIC ± 0.01 *m/z*) in the presence or absence of DabD microsomes (red and black traces, respectively). *The intensity of the 7'-carboxy-L-NGG standard trace was divided by 100 to better scale it with the experimental results.

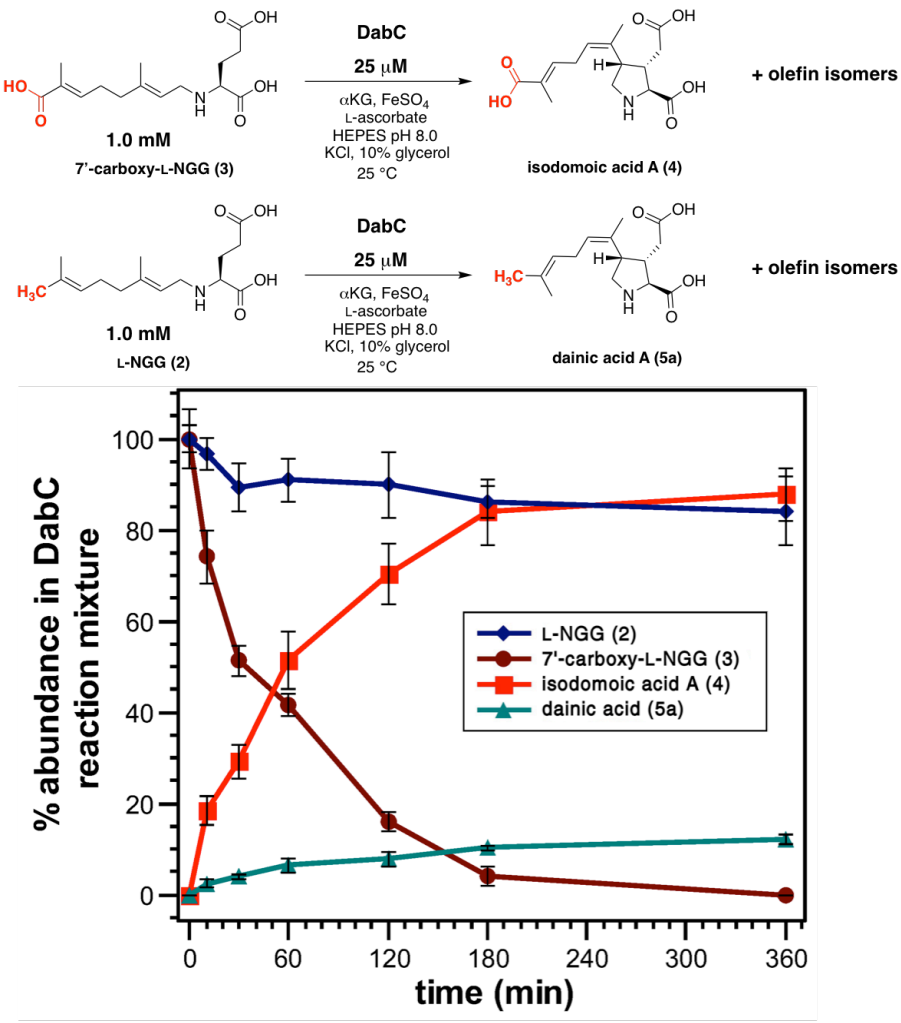


Fig. S12.
Relative DabC substrate consumption of *N*-prenylated substrates **2** or **3** and their respective accumulation of major products **5a** or **4** over time.

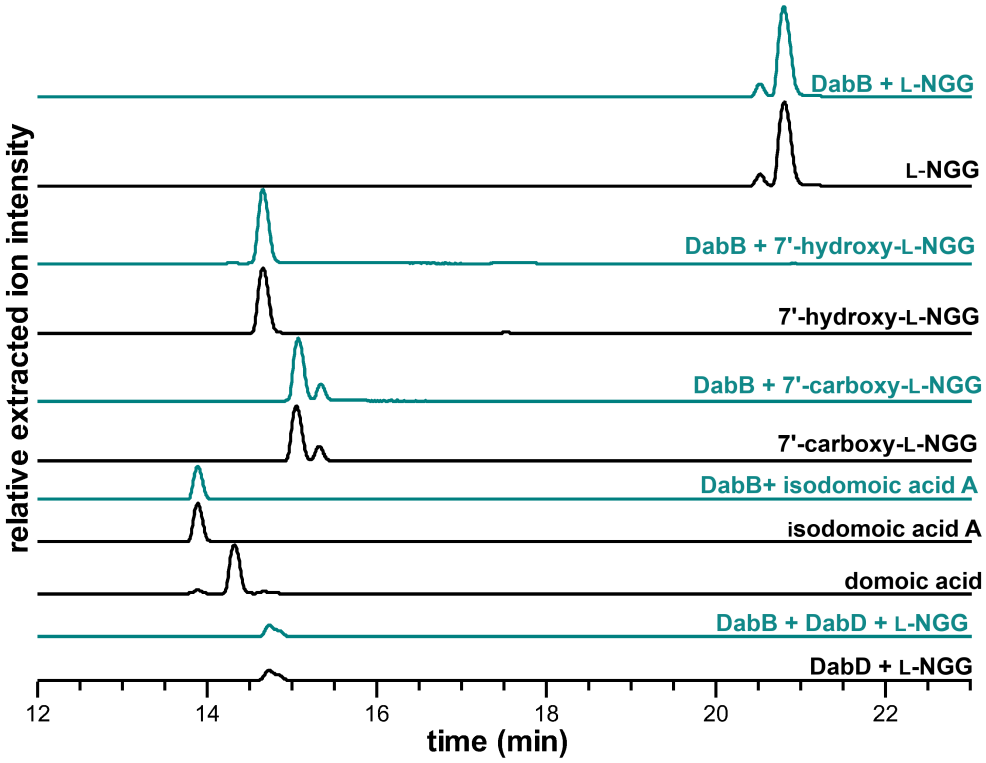


Fig. S13.
Negative mode LCMS chromatograms of DabB assays using LC Method B, showing the extracted ion chromatograms (EIC ± 0.1 *m/z*). The DabD containing assays were extracted for the 7'-hydroxy-L-NGG product. In each case, there did not appear to be change in retention time of the substrate, suggesting that no catalysis had occurred.

JGI Protein ID	InterPro ID	InterPro description	HMMER Pfam description	Phyre2 prediction	Homologs in other non-*Pseudo-nitzschia* diatoms?
190182	IPR001128	Cytochrome P450	cytochrome p450		No
311210			No hit	Modeled on 3TAI: DNA double-strand break repair nuclease NurA, manganese binding (45.1% confidence, 35% coverage)	No
70420	IPR008949	Terpenoid synthase	No hit	Modeled on 1PS1: pentalenene (terpeneoid) synthase (99.3% confidence, 59% coverage)	No
323819			No hit	Terpene Synthase (same ORF as 70420)	No
300256	IPR011876; IPR015797	Isopentenyl-diphosphate delta-isomerase, type 1; NUDIX	NUDIX domain	Modeled on 2PNY: isopentenyl-diphosphate delta-isomerase 2 (100% confidence, 70% coverage)	Yes
317971	IPR000407	Nucleoside phosphatase GDA1/CD39	GDA1/CD39 Nucleoside phosphatase		Yes
65021	IPR005123	2OG-Fe(II) oxygenase	2OG-Fe(II) oxygenase superfamily		No
324484	IPR006025	Peptidase M, neutral zinc metallopeptidases, zinc-binding site	Peptidase M11		Yes
323820			No hit	Modeled on 6BLH: rsv g central conserved region bound to fab cb017.5 (56.7% confidence, 5% coverage)	No
300811	IPR000911	Ribosomal protein L11	Ribosomal L11		Yes
252742	IPR017853	Glycoside hydrolase, catalytic core	No hit	Modeled on 1H4P: exo-1,3-beta glucanase (99.3% confidence, 51% coverage)	Yes
305406			No hit	Modeled on 5B5Y: metal-binding protein with carbonic anhydrase activity, low CO2 inducible protein PtLCIB	Yes
211191	IPR006021	Staphylococcal nuclease (SNase-like)	Staphylococcal nuclease homolog		Yes
25327			No hit	Modeled on 4QS4: Minor Pilin CofB (25.3% confidence, 20% coverage)	No
312725			No hit	Modeled on 4IW3: prolyl-4-hydroxylase (P4H) in complex with elongation factor Tu (EF-Tu) (98.7% confidence, 46% coverage)	Yes
308281	IPR003959	AAA ATPase, core	AAA+-type ATPase		Yes
300319	IPR011046	WD40 repeat-like	G protein beta subunit-like protein		Yes
180541	IPR013116	Acetohydroxy acid	Acetohydroxy acid isomeroreductase,		Yes

		isomeroreductase, catalytic	catalytic domain		
193460	IPR002303	Valyl-tRNA synthetase, class Ia	tRNA synthetase class I		Yes
49425	IPR002013	Synaptojanin, N-terminal	SacI homology domain		Yes
261779	IPR011701; IPR004737	Major facilitator superfamily MFS-1; Nitrate transporter	Major facillitator		Yes
236889	IPR013105	Tetratricopeptide TPR2	No hit	Modeled on 2HR2: tetraticopeptide repeat-like protein (98.2% confidence, 34% coverage)	Yes
327114	IPR001709; IPR001433	Flavoprotein pyridine nucleotide cytochrome reductase; Oxidoreductase FAD/NAD(P)-binding	Flavodoxin; Oxidoreductase NAD-binding domain		Yes
242377	IPR006271	Phosphoserine aminotransferase, Methanosarcina type	No hit	Modeled on 3M5U: phosphoserine aminotransferase (100% confidence, 82% coverage)	Yes
249540	IPR004769	Adenylosuccinate lyase	Adenylosuccinate lyase C-terminus		Yes
301408	IPR001650	DNA/RNA helicase, C-terminal	Helicase C		Yes
294204	IPR005484	Ribosomal protein L18/L5	Ribosomal L18		Yes
66239	IPR001344	Chlorophyll A-B binding protein	Chlorophyll A-B binding protein		Yes
198571	IPR001827	Homeobox protein, antennapedia type	No hit	Modeled on 5UOH: M. tuberculosis serine protease, hydrolase activity (100% confidence, 71% coverage)	Yes
229110	IPR000086	NUDIX hydrolase, core	NUDIX hydrolase	Modeled on 5BON: human protein with 8-oxo-dGTPase activity (100% confidence, 84% coverage)	Yes
327386	IPR001841; IPR014021	Zinc finger, RING-type; Helicase, superfamily 1 and 2, ATP-binding	Ring finger domain; Helicase conserved C-terminal domain		No
247811	IPR006906	Timeless protein	TIMELESS		Yes
251103	IPR001611	Leucine-rich repeat	Leucine-rich repeat		No
5808	IPR010918	AIR synthase related protein, C-terminal	AIR synthase-related protein, C-terminal domain		Yes
242642	IPR001816; IPR002052	Translation elongation factor EFTs/EF1B; N-6 adenine-specific DNA methylase,	Elongation factor TS		Yes

		conserved site			
253335	IPR002035	von Willebrand factor, type A	von Willebrand factor type A domain		Yes
40563	IPR001623	Heat shock protein DnaJ, N-terminal	DnaJ domain		
308790			No hit	Modeled on 1PD7: mad1, transcription related (42% confidence, 12% coverage)	No
189252	IPR001451; IPR003307	Bacterial transferase hexapeptide repeat||eIF4-gamma/eIF5/eIF 2-epsilon	bacterial transferase hexapeptide; eIF4-gamma/eIF5/eIF2-epsilon		Yes
255556	IPR002048	Calcium-binding EF-hand	No hit		Yes
182348			No hit	Modeled on 1CKQ: EcoRI endonuclease (43.1% confidence, 5% coverage)	Yes
163723			No hit	Modeled on 5VO5: Lgd-Shrub complex (transport protein) (65.5% confidence, 16% coverage)	Yes
182666	IPR001164	Arf GTPase activating protein	Putative GTPase activating protein for Arf		Yes

Fig S14.
Further analysis of the 43 transcripts that display significant upregulation under low phosphate and high pCO_2 in an attempt to identify a putative DA isomerase. The analysis platforms InterPro (*56*), pHMMER (*57*), and Phyre2 (*23*) were used to further bioinformatically characterize the translated peptide sequences. In addition, HMMER searches were able to identify putative homologs in other diatom species for select sequences.

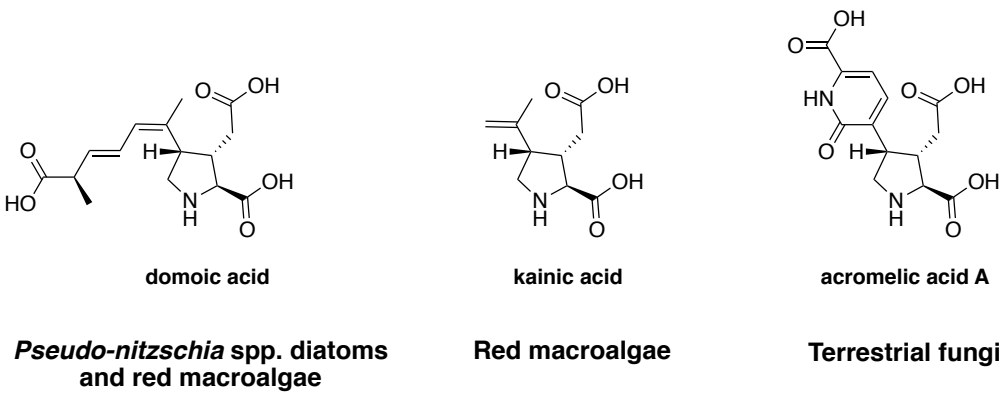


Fig. S15.
Kainoid-ring containing metabolites from diverse biological sources (*12*, *24*, *26*, *27*).

Organism	*dabA* homolog	*dabC* homolog
Grateloupia filicina	GFLG01011861 (incomplete)	GFLG01012884
Palmaria palmata	GFTH01004681	GFTH01007334
Rhodophysema elegans	GFTJ01016989	GFTJ01013404
Palmaria hecatensis	GFTG01026770, GFTG01025549, GFTG01001319 (fragments)	GFTG01010004

Fig. S16.
Table of red algal RNAseq transcript NCBI accession numbers with homology to *dabA* and *dabC*.

>DabA_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:169768..169842,169986..171344 (+ strand) class=CDS length=1434

ATGAAATTTGCAACGTCCATTGTCGCCGCCCTCGCCACTACAGGTGCCGCATTCACCGCGATTCCAAAC
AAGAAGGTTTCTACCGCAGGATTTGGCKTGAAGCACCAGAATCCGCTCTATGCATCGAGTACCTCAGA
CTCCATCAGCAGCCTTACGGCGGAGAGCCCATCTGAAGTCCTTTCTCGAGTCAAAAATACTGGGCTCA
CGTTGACCAACCCCAACGACTTGTACTGGATGGTCGACTTCTGCAAGGAAAAATACTACAACAAAGGC
GACTACTACTACCCCATCAAGACCGTGTGTGATGGTGAGAGTATCGACGTSAAGTTTTATTGCCCATTC
GAGCCCARCCTCAGCCCTCATTATCTCGAGCTCTACGGCTCCCGCGATCAGCGTGCYTCGATGTATGAG
ACAACCATGGAAAAATAYAACCGCATCAACTGCGAAAAAACCTCTGCAATTTGCACMCCTTACTCCA
GTTATGGCGACACCCAGATTGTCGCATACTTTTACTCTATGATGTATTACATCAATGATCAAACCGCCC
ATCTCAAACTACCCGAGAATCAGATTGAAGCGGAKCTCGTCGACATTTTGAACGACGACATTTTGATT
TACTTGAACGAGTTTCTTTCTATCTTCGAACCYAAGAATGACGAAGACTTGGAACGTATTTGGGACTTC
TTGGATTTCTATCAGCCATATTACAACAAGGTTGACGGCAAGATTGTCTTGGACGAGAAATACCAART
CAGAACCCCATCTCAGATGCCCCTAATCAAGACAATCTGYGAGTACATCTCGGAAAAYTTTGCMARCG
AAAAGAATATCACCCAAGTGATTTGGGAGGTTATCCGGTACATCAAGGGAGTRAAGGACGAAATTCA
CATCAGAGGCGACAAGAGTTTCACTCTGTCCCTTCAAGAGTACGAYGACTTCCGTGACAAGGTTACTG
CRAGTCCAATGGCGCAYGCTGTGTCCGATTTGACGCACGAGCGCTTTACCTACAAGGCATACACCAAC
CCACTTTTYATGGAGTTGGAGRACAGGTGTTCCGAGATCATCACATACTTCAACGATGTGTGCACAAG
CGAYAGAGAGCGATTGGATGAAGATCCCTTCAACTCGGTCTTCATTCTCATGGATCTTGACCCCAGTTT
GAATTTCGCAAAATCCTGCGATCTCGTGGTTCARCATGCCYAYGACAAGATGCAGGCATTCTTGAAGC
TCAAGGATGAAATTCTCGAATCGGCCTCGGACGAAGAAGAACGTTTGGCACTTGCTCAGATGATCAAG
ACYAGAGAAGATTCCTTGATTGGATATGTGTTGCATGAAGTCTGCTGTGTCGAAGAGGGSTATGCTCG
TGATCACAAGCCTTTGATGAAAGCGTTTTTGGAGGAGGAATTGGCCAAGTCGCTCTCCGAAAAGGTAT
AA

>DabB_>LN865247.1 Pseudo-nitzschia multistriata strain B856 genome assembly, scaffold: PsnmuV1.4_scaffold_83-size_174963:168137..168859

ATGTTGGTTTGCTTTGTTCCAGGTATTGAAGAAGGAAAATCCTATNCTGCAAAGGATCATTCTGGGCAC
TCAGGGAGCAGCTATCTGAAGAATTTTCTGGCTTTCTTCGCTGGCGCCGCATTCATGGCTGCGGCAATG
TATCTCAATAGTGGCAACTCCATTTCCGTTACCACTTCGGATTTAGTTGGTTCCTTCACCTCATCTGACT
CGCCTGCTATCGATATTGGAATCAATCCAACCCGTCCAAGTGTGGATGCCCGGCAAGACTTTACCAGT
GAAGGTTGGGGTAACCAAACGAGAAATGATGGGTGGCCGAGTTTTCATTTCGAAGTGTTTCACTTTTG
CCCTCACTTGATTAATTTTATGTACGAGAATCAGGTCCATCCACAGTCCCGTGTCGACCAAAGTAACGT
TCTTTATGATGCACAGCATCCTGGTGCTCTAGCGGAAGATGGTTTCGCCGAAGTCTTGCAACAAACNCT
CCCAAAATTCTATGATTACCCCGTAGGATACAGACTCAATGACCGACCCGCAGAAATCTCTGCATATT
GGACGCTCCACATGCGAATTGCACGCGGCAGCGAAGACTCGCATTGGCATGACTTTGCATCGTCTCTA
CCNGAAGGAATTAACAGAATTTATGTGAAAAAGCCCCTACGATGGAGTGGGGATCAGGTCACGCTCTT
CAGCAGCCAAACCGGCTATAATATTCTCGTGTACTAG

>DabC_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:166585..167709 (+ strand) class=CDS length=1125

ATGACAATGAAAGTWGTTAACAACGAAACTGCCRCTTTGACCGCAAACGAGCACGATGTGCAGGTCA
ACAAGGGCAAAACTCTCGAGCAGAGATTTCCACCTCTCGATGGTGAAAAGTTGAAATGGTTTCCGAAG
TCATCTTTGCCTGATGAGATYCCTGCCATTGATATCAGCAAAGTCAACACGAAAGAAGAGCTGGAACA
ATTTTTGGTAGACATCCGCAAGTCAGGACTTTTCTACATCGTCAACCACGGTGTTCCCGAAGACGTCTC
AATTAACGTTTACAACGCCTTCAGAGAATTTCTCTCGTCCACCACCGAAGAAGAGAGAATGAAGTATT
ACACGGACACTCATTTCCAAAATGGTGGATACGTCCCCTTCCAAGGCTCTTCCATTCGCGGAGGAAAT
TTAGGCAAGCCGCAGAAGGACCACGTCGTGAAATACTTTTGGAGAGGGCCTCAAGTTGTTAATAGGAC
CCCMAGCGAGAAATTCACCAAGGCACACGATGYCCATCATACTGAGACCTTCAACGTAGCGGAGAAG
GTGATTAGAACCATTTTCAAGGCCTTAAAGCTTCGCTTCCCTGACTTCGATCCTATGGAATTCGAAGAC
ACTATTAATTCAAAGAAGATGTTCTTCACCAATCGTGTTTACCCTCAGGCTGAAAAGAGTGACGAAGA
AACTATYACTCATCGTCTCGTTCCCCACTTAGATACCAGCTTCATCACGTTAGCGAATCAAGTTCCTGC
CGACAATGGCTTCCAAGGTCTGTTCGTTGAGACTGGAGATGGGAAAAAGGTGAAGGTCCCTGGCATCC
GTAACAGCTATTTGGTCTTCATCGGTCAAAGTCTGTCTTTCCTTACGAAGAACTACCTTCCATCGGCTC
TCCATGGTGTCGACAAGCCTCCCAAGGAAATGTTCGAAGGAAGCGAAAGATCCTCGTTGATCACTTTT
TATGAACCTGCCGAAATCATTATTCCATCCAAGAACATCAATCCCAACCCAGATGAGACCACTGATTC

GTGTCCGTTTTATGATTCTATTGGATTGGATGTAAATGATCCCAAGGGCACTACCTGGGATTTTGTGAA
GAACAAATTCATTACCGGATATTACGCGGATTAA

>DabD_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:164018..165697 (- strand)
class=CDS length=1680
ATGAACTTTGCGACAGCTACTTTATGTGCCGGACTMGCCGTGTCTGGGAAGTTCTTATCGAACTACTAC
CAATTCCTCTCTTCTTCACAGAGCGATGAAGTCGGCTTTCTCGAAACAGTTGACAAGCAGATGGCGAG
CAATGCTGCAGTCGCAGGCGTCGGTCTTGGTCTTATCTTCTACTGTGTCCTCAGCTTTGCAAGAAGCTA
CTACAAGTTTCGCTTCTCTCCCCTTCGTGACGCACCAGGATTTGGTCCGAAATCTTTTGTGTATGGAAT
GTTCTATGAATTCCTCGAGGCGCCATTCATGGAACCTCCAATCGAAGCCCTGAAGAAGCTACGTAAAG
GAGGCAAAGARGTCCCATTTCTTGCCTACACAACTCTTTTTGGAAGCCAAAGGCTTTTGTTATTGGACT
GCGACTTGGTGAAGCACGTCTTTACTTCGCCATCAGGTAGAGATCCAATGAGATACCCCAAGCACTAC
GTTTACTTGAGAGAGGTTGTGGGAGATGGTCTCGTCGTCGTTGAGGGCACTGAATGGAGTCGCCATCG
TCGAATCATCCAGCCKGCGTTTCAATCTGTTTTTCTGAAAGATGCTATRAGTATGGTCGTCCCCGCCTT
GGTAGAGAATCTTGTAGAGGTATGGAAGAAGACTGCTGGGGCAACGATCAATCTGAATGCCCATCTTT
CGCTAATCACTCTTGACGTGATTGGAAAAGTTGCTTTTTCGCACGAGTTCAATGCTTCCAAACTGCTCA
ACGAATGGGCCGAATGTCCRGATAAGGAARTGGGGGAGGTTGACGATCCGTTGATTTCATCCATCGGT
GATGCCTTTTCCTCYAGCCCTCTCAAATTAATGCTTACCGTGTTGAAGCAGCCATGGCTCGAGAAATAC
TTGAGTCCTGCCTTCAGAACTTCACGCAATTTACTGAACAAGGCAGCAGATGACATCGTACAGAATGC
CAAGAACATCGATGATCCTAAACGACGAAGCGTATTGAACCTGATGATGGAAGCTAAGGACGATGAA
TCCAGCAAGGCTCGCAATAAGTTGACAGACACCGAGCTCCGCGATGAAGTRAAAACATTTCTTGTTGC
TGGCCACGAAACAACYTCGACTTGGGTTCATTGGGCTTTATATGTACTTGCGATTCGCCCAGATCTCCA
AGATAAGGTACACGCTGACATCATGAAGCACGCGTCCGACGACAAGACTATTTTACTGGAACAAGCC
GATCAAATGGAATATCTGTGGGCCTTCATGAACGAAACTCTTCGTTTGTACTCACCTCTCGGATTGATC
AGCCGTGTCACTCACAAGGAAGAAAACTTCAAGGGCTACAKGATCCCAGTGGGAACGAATCTGAGAA
TTCCTATTCACTTGATTCATAGACATCCTGATCATTGGAAGGACCCCGAAGAGTTTCTACCCGAACGTT
GGTTCGACAAAGAAGAAACGAGCAAGAGACACAAGTTCGCTTTCGTCCCCTTTGCAGCAGGGGGAAG
AAACTGCATCGGTCAGCGTTTCGCTACGATGGAGGCCAAAATTATACTCGCAAATGTTGCCAAGAATT
TCGAAATCCATTTGGCTGATTCCATGAAGGGCAAGAAAATTACATTCAGTAATTTCATTTCTTTGAAGT
GCAAACCCGAGATTGAAATTCGGGTGACACCTCGAGAATAG

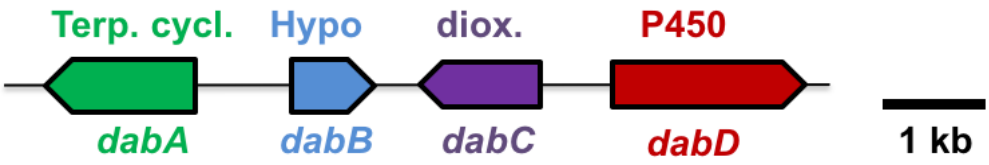


Fig. S17.
Sequences from the publicly available *P. multistriata* genome (*30*).

>DabA_ Pseudo_nitzschia-australis-10249_10_AB-20140214|4053_1
CTACCCTTGCCACCACCGGTGCAGCATTCACTGCAAATCCCCAAAAGGCGTTCACCGCAGGAGCCAGC
GCTAGGCAACAGGTTTCGCTCAATGCATTGAGCACCCTTAGCTCCATCGGCAGCCTTACTGCAGAAAG
CAAATCGGAAATCCTTTCTCGTGTTAAGGAGAACGGACTGACCTTGACGAATCCTAAGGATTTGTACT
GGATGGTTGACCATTTTAAGGAGAATTACTACGACGAAGGTGACTACTACTATCCCATCAAGACGGTG
TGTGATGGTGAAAGCATCGATGTCAAGTTCTACTGTCCTTTCGAGCCCAGCCTCAGCCCGCATTACCTT
CAGTTGTACGGAAGTCGTGACGAGCGTGCTTCGATTTACGAAACGACCATGGCGAAATACAACAAGAT
CAACAGCGAGAAGACTTCCGCGATTTGCACTCCATACTCCAGTTATGGCGACACCCAGATTATCGCGT
ACTTTTACTCCATGATGTATTACATCAACGATCAAACCGCCCATCTCAAGCTTCCCGAAGGAGAGATC
GAAGCTGAACTCATTGACGTCTTGAACGACGATATTTTGATTTACCTCAACGAATTCTTGAGTGTCTTT
GAACCTGAGGACGACGCTGACTTTGAACGCATTTGGGATTTCTTGGAGTTTTACCAGCCATACTTTAAC
AAGGTTGACGGAAAGATTGTGTTGGACGAGAAATACCAGAAGGGAACCCCATCGCAGATGCCCCTGA
TCCAGACCATCTGTTCGTACATCGCCGAACAATTCGCCCCCAACAAGAACATCACTCAAGTAATTTGG
GAGGTCATCAGATACATCAAGGGTGTTAAGAATGAGATCCAGATCAGAGGCGATAAAGGTTTCACTCT
ATCCCTCCAAGAGTACGACGACTTCCGTGACAAGGTTACCGCCAGTCCAATGGCTCATGCGGTTTCTG
ACTTGACCCACAAACGCTTCACCTACGAAGCATACACCGACCCACTTTTCATGGAGTTGGAGAACCGT
TGCTCCGAAATCATCACCTACTTTAACGACGTGTGCACTAGTGATAGAGAGCGATTGGACGACGATCC
GTTCAACTCGGTCTTCATTCTCATGGATCTTGATCCCACTTTGAACTTTGCCGCTTCGTGCGACCTTGTC
GTTGGTCACGCTTACGAAAAGATGGAGAGATTCTTGGAGCTTAAGGAAGAAATTCTCGTATCCGCAAG
AAGCGAGAAAGAAAAGCTTGCCTTTGCCCAGATGATCAAGACCCGAGAAGATTCGTTGATTGGATATG
TTTTGCACGAAGTTTGCTGTGTCGAGGATGGGTATGCTCGTGATCACAAACCTTTGATGAAAGCATTTT
TGGAGGAGGAGTTGACCGAGGCCCTTGAGATGGCTTAA

>DabB_Pseudo_nitzschia-australis-10249_10_AB-20140214|210_1
ATGAGTGCTAGTAAAAATAGTTATCAGTCAGCTCATGTTGAGTCCTATCGCGTTGATGGAGAATACGT
TCCTACTGAAAGGCTTCGTGATGTGTCCTCGA
AAAGCCCTCTGCTCTTGAATTTCGTGACTTTTTTCATTGGCGCTGCATTCATGGCTATAATCAGTTCAAG
AGGGGAATCAGCTGAAGCTGCTATTGATTT
TGGAATCAATCCAACCCGTCCAGGAGTGGATGCCCGGCAAGGATTTAATACCAATGAAGGATGGGGG
AACCAAACTCGAGATGATGGTTGGCCAAGTTTT
CATTTCGAGGTATTTCACTTTTGCCCTCACATGATTAACTTCTTGTACGAGAACCAGGTTCATCCACAA
TCCAAGATCGACGAAAACAATGTTATTTATG
ACGCGCAGCACAATGGTGCTATAGCGGAAGACGGCTTTGCCGAGGTTTTGCAACAAACCCTCCCCAAG
TTTTATGATTATCCCGTAGGATACAGACTTAA
CGACCGCCCCGCAGAAATTTCTGCGTTTTGGACCCTCCACATTCGAATCGCACGTGGTAGTGCGGACTC
CAATTGGCACGACTTTGCATCCTC
TCTTCCC
GAAGGTATTACCAGAATTTACGTCAAAAGGCCCCTGCGATGGGATGGTGATGGTGATGATGTTGCX

>DabC_ Pseudo_nitzschia-australis-10249_10_AB-20140214|55196_1
ATGACTGTCCAAGTAATTAACAACGAGACCAACGTGTTGACCCCCAACGAGGACGACGTCCAAATCA
ACAAGGGCAAAACCTTGGAGACGAGGTTCCCCCCCCTCAAAGCCGACGACCTCAAGTGGTTTCCCCAT
GATTCTCTGCCCGCAGAGATCCCTGCTATTGATATCAGCAAAGTAAACACGAAGGAGGAATTGGAACA
GTTTTTGGTTGACATCCGCAAATCGGGACTTTTCTACATTGTGAACCACGGTGTCCCCGAAGAGGTCTC
CATTAATGTTTACAACGCATTCAGAGAGTTTCTCTCCACCACCACCGAGGAAGAAAGGATGAAGTACT
ACACCGATACGCATTTCCAAAACGGTGGATACGTCCCCTTCCAGGGCTCCTCTATTCGCGGAGGAAAC
TTGGGAAAGCCGCAAAAGGACCACGTCATCAAGTATTTTTGGAGAGGACCACAGGTCGTCAACAGAA
CTCCTAGCGAGAGTTTCACCAAGGCCCATGACGATCACCACACCGAGACCTTCAACGTAGCTGAGAAG
GTTATTAGAACGATCTTCAAGGCTCTCAAACTTCGTTTCCCAGACTTCGACCCGATGGAGTTCGAGAAC
ACGATCAATTCCAAGAAGATGTTCTTCACCAACCGCATCTACCCGCAAGCCGAAAAGTCCGACGAGGA
AGAGCTCACCCATCGTCTTGTTCCCCACCTGGACACCAGTTTCATTACCTTGGCAAACCAGGTTCCCGC
CGATAATCGTTTCCAGGGTCTGTTCGTTGAGACCGGCGATGGGAAAAAGGTCCCCGTCCCGGGCATCC
GCAACAGTTACTTGGTCTTCATTGGTCAGAGTTTGTCGTTCCTCACCAAAAATTACCTTCCATCGGCTC
TTCACGGTGTGGACAAGCCACCCAAGGAAATCTTTGAGGGAAGCGAGAGATCCTCGTTGATCACTTTT
TACGAACCCGCGGAGATCATTATTCCATCTAAGAACATTAATCCCAACCCAGACGAGGTCACCGTGTC

GTGTCCATTCTACGATTCCATCGGCTTGAATGTTAACGACCCAGAGGGTACTACGTGGGATTTTGTCAA
GAACAAATTCATTAAGGGATACTATGCGGATTAA

>DabD_ Pseudo_nitzschia-australis-10249_10_AB-20140214|13962_1
ATGAGCAGCATTAATGTTGTAACGGCTCTTGTATCCATCGGGCTGGCCGTATCCGCTAAATTCGCATCG
AGTTACCACGATTTCCTTTCTTCTCCCAAGAGCGACGAAGTTGGATTTTTGGAAACCCTGGACATCACG
AACACGACGAGCATCGATGCGGTTGCTCACGTCGTTCTCGGTCTTGTCTCGTACTTTGTCCTTAGCTTT
GCAAGGAGCTACTACAAGTTTCGCTTTTCTCCCCTGCGTGATGCTCCGGGCTTTGGTCCGAGATCCTTC
GTGTACGGAATGTTCTATGAATTTCTTGAGGCGCCATTCATGGAACCCCCCATCTTGGCCCTCAAGAAG
CTTCGAGAAGGAGGCAAAGAAATTCCGTTTCTGGCCTACACGACACTGTTTGGAAGCCAACGGCTCTT
GTTGTTGGACTGCGACCTGATCAAGCACGTCTTTACGGCACCGTCCGGGAAGGATCCCATGAGATACC
CCAAGCACTACGTTTACTTGAGGGAGGTAGTTGGAGACGGTCTCGTAGTCGTTGAAGGTCAGGAATGG
AGTCGCCATCGTCGTATCATTCAGCCGGCTTTCCAATCCATGTTTCTCAAAGATGCCATCGGTATGGTC
GTCCCCGCCTTGGTAGAAAATCTCGTGAACGTATGGAAGAAGACCGCTGGGACGACAATCAACATGA
ACGCCCACCTTTCGCTCATCACCCTCGACGTGATTGGAAAGGTAGCGTTTTCGCACGAATTCAACGCTA
GCAAGCTACTCAACCAATGGGCCGAGTCTCCCGACAAAGAATTGGGAGAGGTCGACGATCCCTTGATT
TCGTCCATTGGCGGTTCGTTTTCCTCCAGCCCCCTCAAATTGATGCTCACCGTACTGAAACTTCCGTGG
CTCGAAAAGCACCTCAGTCCCTCGTTCCGAACCACGCGCAACCTACTCAACAAAGCAGCTGACGATAT
TGTGCAAAATGCCAGGAATATCAAAGATCCCTCGAGGCGCAGTGTGTTGAACCTAATGATGGAAGCA
AAGGACGGCGAATCCAGCAAGGCCCGCAATCAGTTGACGGACACGGAGCTTCGAGACGAGGTCAAGA
CATTCCTGGTTGCAGGCCACGAAACAACCTCCACCTGGTCCCACTGGGCGCTCTACGTGCTCGCGATCC
GCCCGGATCTCCAAGAAAAGGTGTACGCCGATGTCATGAAGCACGCACCCCCTAACGATGAGACTATT
GTGCTGGAACAAGCCGACCAAATGGAATACATGTGGGCGTTCATGAACGAAACGCTGCGTTTGTACTC
GCCCCTCGGCCTGATCAGCCGCGTTACCCACCAGGAAGAAAACTTCAAGGGATACACGATTCCGAAAG
GCACCAACCTACGAATTCCGATCCATTTGATTCACAGACATCCCGATCACTGGAAGGACCCCGAGGTT
TTCCGACCCGAGCGTTGGTTCGACAAGGAAGAAACCAGCAAGAGACACAAGTTCGCGTTCATCCCATT
TGCTGCCGGTGGAAGAAACTGCATCGGTCAGCGTTTTGCTACCATGGAGGCCAAGATCATCGTCGCGA
ACGTTGCCAAGAATTTCAAGATTCACTTGGCGGATTCGATGAAGGGCAAGGAAATCACCTTTAGCAAC
TTCATTTCTTTGAAGTGCAACCCCGAGGTTGAAATTCGTGTCGAGGCCCGAAAATAA

Fig S18.
Sequences from the publicly available *P. australis* RNAseq dataset (*31*).

***Pseudo-nitzschia* species/strain/variety**	***dab* gene expression**	**MMETSP Library IDs**
Pseudo-nitzschia arenysensis	No	MMETSP0329
Pseudo-nitzschia australis, Strain 10249 10 AB	Yes, *dabA-D* all expressed	MMETSP0139, MMETSP0140, MMETSP0141, MMETSP0142
Pseudo-nitzschia delicatissima B596	No	MMETSP0327
Pseudo-nitzschia delicatissima, Strain UNC1205	No	MMETSP1432
Pseudo-nitzschia fraudulenta, Strain WWA7	No	MMETSP0850, MMETSP0851, MMETSP0852, MMETSP0853
Pseudo-nitzschia heimii, Strain UNC1101	No	MMETSP1423
Pseudo-nitzschia pungens	No	MMETSP1061
Pseudo-nitzschia pungens cf. cingulata	No	MMETSP1060

Fig. S19.
Table of *Pseudo-nitzschia* transcriptomes analyzed for *dab* gene expression. All transcriptomes were generated as a part of the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) and are available through https://www.imicrobe.us/ using the library IDs in the table above (*31*).

Primer Table

DabA_001	ATG AAG TTT GCA ACA TCC ATC GTC
DabA_002	TCA ATT GAG GCG AAC GGA CT
DabC_001	ATGACTGTGGCAATAAATAACGA
DabC_002	CTA ATC AGC GTA GTA TCC GG
DabD_001	ATG AAC ATT GCA ACA GCT AGT TT
DabD_002	CTA TCC ACG GGG TGT AAC ATA
DabA_p28_001	GGTGCCGCGCGGCAGCCATATG ATG TCG CAC CCA AGC CAG CTC AAT GCC
DabA_p28_002	GCTCGAGTGCGGCCGCAAGCTT TCA ATT GAG GCG AAC GGA CTC AGA CTC AAC CGG
DabB_MBP_001	CTGTACTTCCAATCCGGATCCATGATGGCTGGTAGAAAAGGA TACCAATCAGTTATG
DabB_MBP_002	GGTGGTGGTGGTGCTCGAG TCAGTAGACAAGAATGTTGTAGCCAGTTTGG
DabA_seq_001	CTATATCAATGATCAAACTGCTCATCTCAAA
pET28_BB_001	CATATGGCTGCCGCGCGGCACCAGG
pET28_BB_002	AAGCTTGGATCCGAATTCCCCAGATCTAAAGTTTTGTCGTCT TTCC
pET28-MBP_BB_001	GGATCCGGATTGGAAGTACAGGTTCTCAGATCC
pET28-MBP_BB_002	TGAGATCCGGCTGCTAACAAAGCCCGAAAGG
PtpBAD-CTHF_001	CTACGGGGTCTGACGCTCAGTTATGACAACTTGACGGCTACA TCATTCAC
PtpBAD-CTHF_002	CCAAAAAAACGGGTATGGAGAAACAGTAGAGA
PtpBAD-CTHF_003	GGAACTGCCAGGCATCAAATAAAACGAAAG
PtpBAD-CTHF_004	ATTCTTGAAGACGAAAGGGCCTCGTGATACGCAAAAAGGCC ATCCGTCAGGATG
PtpBAD-CTHF_005	AACAGAATTTGCCTGGCGGCGGAACTGCCAGGCATCAAATA AAACGAAAG
PtpBAD-CTHF_006	TAGCCGTCAAGTTGTCATAACTGAGCGTCAGACCCCGTAGAA AAGAT
PtpBAD-CTHF_007	ACCCGTTTTTTTGGATGGAGTGAAAATGGTGAGCAAGGGCG AGGAGC
PtpBAD-CTHF_008	ATTTGATGCCTGGCAGTTCCTTACTTGTACAGCTCGTCCATGC CGAG
PtpBAD-CTHF_ultramer001	CTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTAACCTC GAGCGCTGGTCCCTCGCGGTAGCCACCACCACCATCACCACG ACTACAAGGATGATGACGATAAGTGAAACAGAATTTGCCTG GCGGC
PtpBAD-CTHF_ultramer002	GCCGCCAGGCAAATTCTGTTTCACTTATCGTCATCATCCTTGT AGTCGTGGTGATGGTGGTGGTGGCTACCGCGAGGGACCAGC GCTCGAGGTTAATTCCTCCTGTTAGCCCAAAAAAACGGGTAT GGAG
PtpBAD-	AAACCAATTGTCCATATTGCATCAGACATTGCC

CTHF_col001	
PtpBAD-CTHF_col002	CTTTCGTTTTATTTGATGCCTGGCAGTTCC
DabC-CTHF_001	GCTAACAGGAGGAATTAACCATGACTGTGGCAATAAATAACGAAACCGTTGT
DabC-CTHF_002	CTACCGCGAGGGACCAGCGCATCAGCGTAGTATCCGGTGATGAACTTGT
DabD-CTHF_001	GCTAACAGGAGGAATTAACCATGAACATTGCAACAGCTAGTTTATGTGCCG
DabD-CTHF_002	CTACCGCGAGGGACCAGCGCTCCACGGGGTGTAACATAAATTTCAATCTCGG
DabCseq_001	AGAGAATGAAGTATTATACGGACACTCATT
DabCseq_002	CTCCGGTCTCAACGAAAAGAC
DabDseq_001	CGTTTCTCAGGGATGCTATCAC
DabDseq_002	AACAAGGAACGTTTTCACTTCGT
ampDabC-D_001	GAAGTATTATACGGACACTCATTTCC
ampDabC-D_002	CATAGAAGCGAAACGCTGAC
ampDabA-C_001	TTCTTCATGGTCGTCTCGTAA
ampDabA-C_002	CAACCCAGAAGAGACCTCTG
OH_DabC-D_001	TTTGTACAAAAAAGCAGGCTGAAGTATTATACGGACACTCATTTCC
OH_DabC-D_002	TTTGTACAAGAAAGCTGGGTCATAGAAGCGAAACGCTGAC
OH_DabA-C_001	TTTGTACAAAAAAGCAGGCTTTCTTCATGGTCGTCTCGTAA
OH_DabA-C_002	TTTGTACAAGAAAGCTGGGTCAACCCAGAAGAGACCTCTG
DabA-C_seq001	CGCGAATTCGCATTAAGAATGACTCAC
DabA-C_seq002	ATGTGAAGGGTCCAATATGCAGAA
DabA-C_seq003	CGTTACAACTCAACTGGTAGGCGAT
DabA-C_seq004	TACAGTAGATATGCACCAGCGAACG
DabA-C_seq005	AATCAATCCAACCCGTCC
DabA-C_seq006	GGAGTCTCGAGTGTTAATCGTC
DabA-C_seq007	CGTTCGCTGGTGCATATCTACT
DabA-C_seq008	GTGAGTCATTCTTAATGCGAATTCGC
DabA-C_seq009	CCATTATTTCCGTAGACACTGCCAC
DabA-C_seq010	CGCTAATACTCCGATGTGCT
DabC-D_seq001	ATTCTCTCTTCCTCGGTGGT
DabC-D_seq002	CGGCACATAAACTAGCTGTTGC
DabC-D_seq003	CGAATCGGACTACGTGTACTAGATAATCTTAC
DabC-D_seq004	CTTAAAGTTCCGACTTCAGCAGTAGGC
PtGG_col001	GCGCAACGTTGTTGCCATTGTCCCAGTCACGACGTTGTAAAACGAC
PtGG_col002	AATGTTGCAGCACTGACCCTTCACTATAGGGGATATCAGCTGGATGG
pBEVY-DabD_001	AAAAAAAAGTAAGAATTTTTGAAAGCCACCATGAACATTGCAACAGCTAGTTTA
pBEVY-DabD_002	GCCTGCAGGTCGACTCTAGATCACTTATCGTCATCATCCTTG

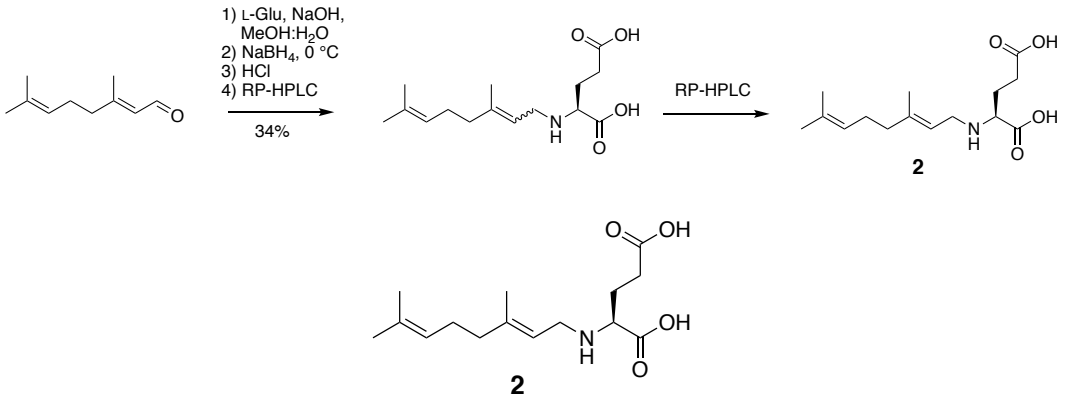
pBEVY-PmCPR1_001	TAACGTCAAGGAGAAAAAACGCCACCATGGTTAACGAAACCACCG
pBEVY-PmCPR1_002	AGTGTCGAATTCGAGCTCGGCTACGACCAAAGCTCTTGAACA
pBEVY-DabD_col001	TTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAAC
pBEVY-DabD_col002	GAGCCAAGGCTGTTTCAATACCGTAAGCATTAATT
pBEVY-PmCPR1_col001	GGGGTAATTAATCAGCGAAGCGATGAT
pBEVY-PmCPR1_col002	AGCGGCTTCTGTAGTTCTCC
pBEVY-seq001	GAAACTTCGAACACTGTCATCTAAAGATGC
pBEVY-seq002	GAAGTGTCAACAACGTATCTACCAACG
PmultCPR1_seq001	GATCAGGAATCACCATCGTTGTCG
PmultCPR1_seq002	TGACGGTGTACTATGCCACC
PmultCPR1_seq003	ACACGGACCCGTTGCTTTCT
PmultCPR_seq004	TTGGTGACGGCTACCGTCA

Chemical synthesis

Synthesis of geranyl pyrophosphate

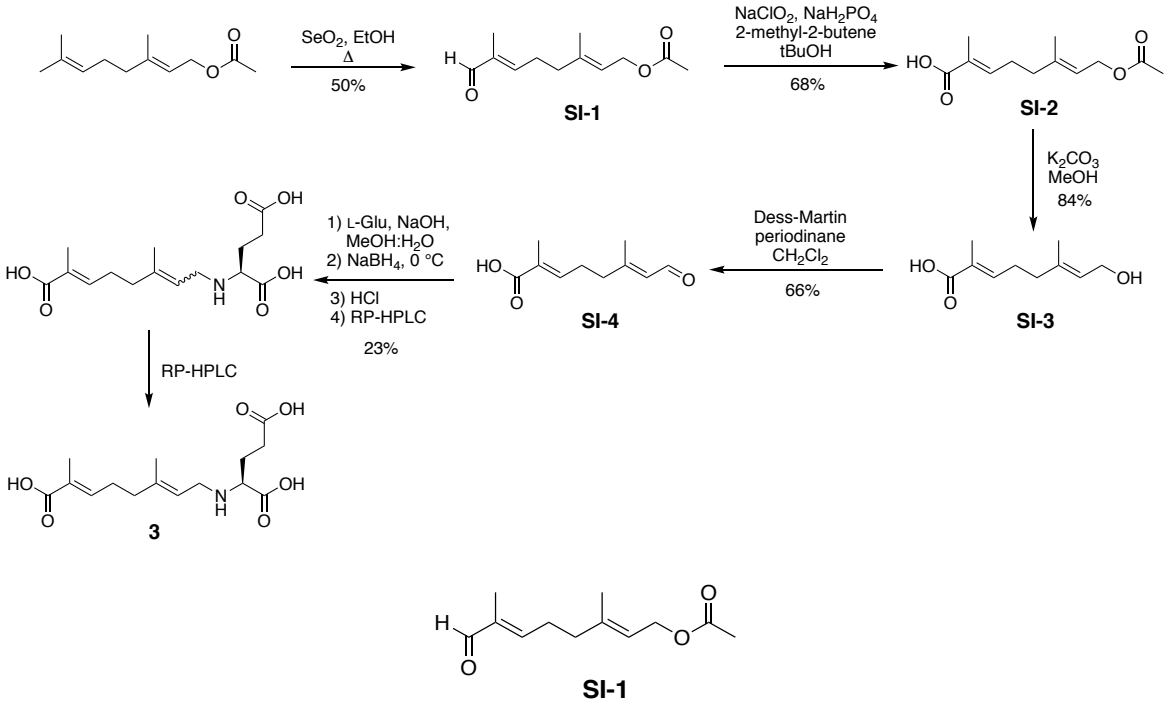
Geranyl pyrophosphate was prepared as a trisammonium salt using established protocols (*58*) and matched literature characterization.

Synthesis of *N*-geranyl-L-glutamic acid (2)



N-geranyl-L-glutamic acid (**2**)
This molecule was synthesized via a modified synthetic procedure (*59*). A solution of citral (0.100 g, 0.66 mmol) in methanol (3 mL) was added to an aqueous solution (3 mL) of L-glutamic acid (0.193 g, 1.31 mmol) and sodium hydroxide (0.079 mg, 1.97 mmol) and stirred for 3 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.032 g, 0.85 mmol) was added at once and stirred for 1 h. 1 N HCl was added until pH 3, and the reaction mixture was concentrated *in vacuo*. Allylic alcohols were removed by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (5 min), 30 – 45% B (10 min), 45 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **2** eluted at 15.2 minutes, whereas minor *cis*-regioisomer *N*-neryl-L-glutamic acid eluted at 14.7 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.0664 g, 0.23 mmol, 36%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). ^{1}H NMR (600 MHz, D_2O): δ 5.21 (t, J = 7.8 Hz, 1H), 5.12 (t, J = 6.8 Hz, 1H), 3.64 (dd, J = 7.8, 5.7 Hz, 2H), 3.57 (dd, J = 6.9, 5.6 Hz, 1H), 2.47 – 2.42 (m, 2H), 2.13 – 2.07 (m, 5H), 2.01 (ddt, J = 14.4, 7.0, 7.0 Hz, 1H), 1.66 (s, 3H), 1.64 (s, 3H), 1.57 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 176.3, 171.8, 146.4, 132.7, 122.5, 111.5, 58.5, 42.5, 37.5, 29.3, 23.9, 23.8, 23.5, 15.7, 14.4; HRMS (ESI) Calculated for $C_{15}H_{24}NO_4$ 282.1711, found 282.1713 $(M-H)^-$
.

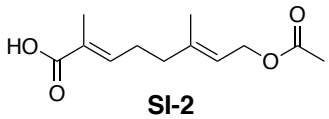
Synthesis of 7'-carboxy-*N*-geranyl-L-glutamic acid (3)



SI-1

(2*E*,6*E*)-3,7-dimethyl-8-oxoocta-2,6-dien-1-yl acetate (**SI-1**)

This molecule was synthesized following a literature procedure (*60*). To a solution of selenium dioxide (3.88 g, 35.0 mmol) refluxing in 95% ethanol (125 mL) at 89 °C was added a solution of geranyl acetate (5.00 mL, 23.3 mmol) in ethanol (25 mL) over 75 minutes. The reaction mixture was stirred under reflux for 16 h, filtered through a pad of silica, and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (4:1 hexanes:EtOAc), isolating aldehyde **SI-1** as a light orange oil (2.48 g, 11.8 mmol, 50%). R_f 0.7 (hexanes/EtOAc 2:1); ^{1}H NMR (500 MHz, $CDCl_3$): δ 9.38 (s, 1H), 6.44 (td, J = 7.3, 1.5 Hz, 1H), 5.37 (tq, J = 7.0, 1.4 Hz, 1H), 4.58 (d, J = 7.1 Hz, 2H), 2.49 (dt, J = 7.4, 7.4 Hz, 2H), 2.23 (t, J = 7.6 Hz, 2H), 2.04 (s, 3H), 1.74 (s, 3H), 1.73 (s, 3H); HRMS (ESI) Calculated for $C_{12}H_{18}NaO_3$ 233.1148, found 233.1147 $(M+Na)^+$.

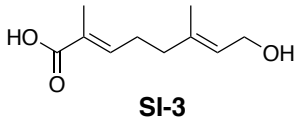


SI-2

(2*E*,6*E*)-8-acetoxy-2,6-dimethylocta-2,6-dienoic acid (**SI-2**)

This molecule was synthesized following a literature procedure (*61*). To a solution of **SI-1** (0.981 g, 4.67 mmol) and 2-methyl-2-butene (5.49 mL, 90%, 46.7 mmol) in *tert*-butanol (20 mL) at 0°C was added an aqueous (20 mL) solution of sodium dihydrogen phosphate (2.24 g, 18.7 mmol) and sodium chlorite (2.11 g, 80%, 18.7 mmol) over 30 minutes. Following addition, the reaction mixture was stirred at 0 °C for 1 h, then warmed to room temperature and stirred for 18 h. Organic compounds were extracted using EtOAc (3 x 25 mL), pooled organic layers were washed with water (25 mL) and brine (25 mL), dried over $MgSO_4$, filtered and concentrated *in*

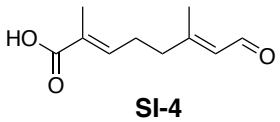
vacuo. The crude reaction mixture was purified by silica flash chromatography (4:1 to 2:1 hexanes:EtOAc + 0.1% AcOH), yielding **SI-2** was a clear colorless oil (0.877 g, 3.88 mmol, 68%). R_f 0.24 (hexanes/EtOAc 2:1); ^{1}H NMR (500 MHz, $CDCl_3$): δ 6.86 (tq, *J* = 7.3, 1.5 Hz, 1H), 5.37 (tq, *J* = 7.2, 1.4 Hz, 1H), 4.59 (d, *J* = 7.1 Hz, 2H), 2.34 (dt, *J* = 7.4, 7.4 Hz, 2H), 2.18 (t, *J* = 7.6 Hz, 2H), 2.06 (s, 3H), 1.84 (s, 3H), 1.72 (s, 3H); HRMS (ESI) Calculated for $C_{12}H_{18}NaO_4$ 249.1097, found 249.1094 $(M+Na)^+$.



SI-3

(2*E*,6*E*)-8-hydroxy-2,6-dimethylocta-2,6-dienoic acid (**SI-3**)

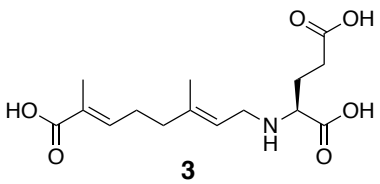
This compound was synthesized following a literature procedure (*62*). To a solution of **SI-2** (0.198 g, 0.88 mmol) in methanol (10 mL) at 0 °C, an aqueous (2 mL) solution of potassium carbonate (0.242 g, 1.75 mmol) was added over 2 minutes. Following addition, the reaction mixture was stirred at 0 °C for 5 minutes, then warmed to room temperature and stirred for 16 h. CH_2Cl_2 (10 mL) was added, followed by 1 N HCl until pH < 3. The layers were separated, and organic compounds were extracted with subsequent CH_2Cl_2 washes (2 x 10 mL). Pooled organic layers were washed with water (10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*, using toluene co-evaporations to remove trace water. Compound **SI-3** was obtained as a light yellow oil (0.136 g, 0.74 mmol, 84%) and was used without further purification. R_f 0.45 (EtOAc + 0.1% AcOH); ^{1}H NMR (500 MHz, $CDCl_3$): δ 6.87 (tq, *J* = 7.3, 1.6 Hz, 1H), 5.49 – 5.39 (m, 1H), 4.17 (d, *J* = 6.8 Hz, 2H), 2.34 (dt, *J* = 7.5, 7.4 Hz, 2H), 2.16 (t, *J* = 7.6 Hz, 2H), 1.84 (d, *J* = 1.2 Hz, 3H), 1.69 (s, 3H). HRMS (ESI) Calculated for $C_{10}H_{16}NaO_3$ 207.0992, found 207.0991 $(M+Na)^+$.



SI-4

(2*E*,6*E*)-2,6-dimethyl-8-oxoocta-2,6-dienoic acid (**SI-4**)

Dess Martin periodinane (0.465 g, 1.10 mmol) was added to a solution of **SI-3** (0.101 g, 0.55 mmol) in dry CH_2Cl_2 (10 mL) at once and stirred at room temperature for 1 h. The reaction mixture was quenched with an aqueous 10% sodium thiosulfate solution (10 mL), and washed with water (10 mL). Organic compounds were extracted from pooled aqueous layers with CH_2Cl_2 (3 x 20 mL). Pooled organic layers were washed with brine (20 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (1:1 hexanes:EtOAc + 0.1% AcOH), yielding **SI-4** as a white waxy solid (0.066 g, 0.36 mmol, 66%). R_f 0.62 (EtOAc + 0.1% AcOH); IR (CH_2Cl_2 cast) 2934 (br), 2656, 1687, 1648, 1422, 1382, 1284, 1196, 1126 cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$): δ 10.00 (d, *J* = 7.9 Hz, 1H), 6.83 (tq, *J* = 7.1, 1.5 Hz, 1H), 5.90 (dq, *J* = 7.9, 1.3 Hz, 1H), 2.45 – 2.41 (m, 2H), 2.40 – 2.34 (m, 2H), 2.19 (d, *J* = 1.3 Hz, 3H), 1.85 (d, *J* = 1.3 Hz, 3H); ^{13}C NMR (151 MHz, $CDCl_3$): δ 191.3, 172.8, 162.2, 142.5, 128.4, 127.7, 39.0, 26.5, 17.8, 12.3; HRMS (ESI) Calculated for $C_{10}H_{13}O_3$ 181.0870, found 181.0870 $(M-H)^-$.

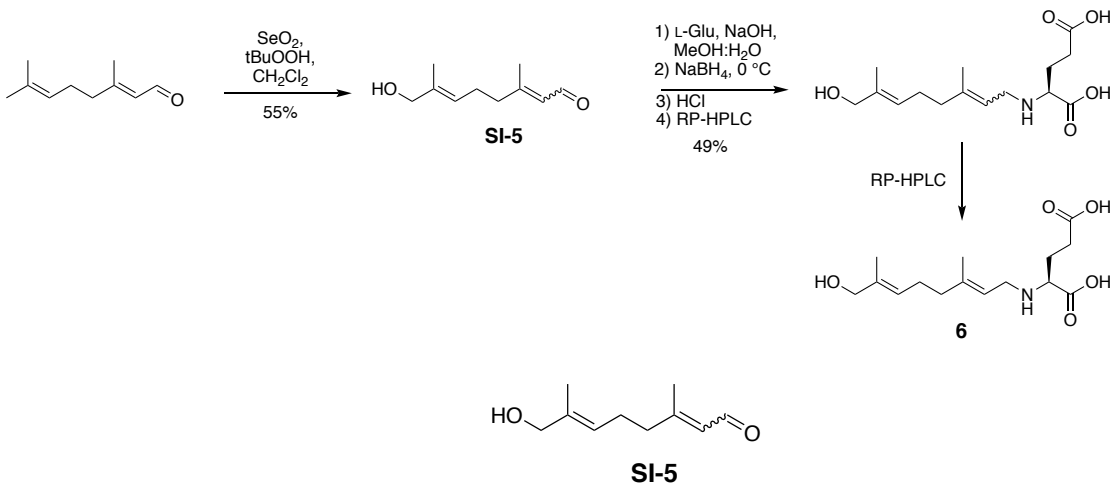


3

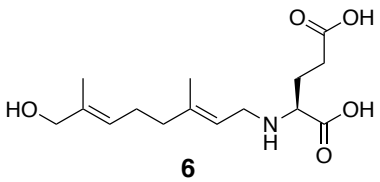
7'-carboxy-*N*-geranyl-L-glutamic acid (**3**)

A solution of **SI-4** (0.0309 g, 0.17 mmol) in methanol (5 mL) was added to an aqueous solution (5 mL) of L-glutamic acid (0.050 g, 0.34 mmol) and sodium hydroxide (0.021 g, 0.52 mmol) and stirred for 2 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.0083 g, 0.22 mmol) was added at once, and the reaction mixture was stirred for 1 h. 1 N HCl was added until pH 3, and the reaction mixture was concentrated *in vacuo*. Alcohol **SI-3** was recovered by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (15 min), 30 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **3** eluted at 13.7 minutes, whereas minor *cis*-regioisomer 7'-carboxy-*N*-neryl-L-glutamic acid eluted at 14.2 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.0123 g, 0.039 mmol, 23%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). ^{1}H NMR (600 MHz, D_2O): δ 6.71 (t, *J* = 6.8 Hz, 1H), 5.27 (t, *J* = 7.6 Hz, 1H), 3.67 (t, *J* = 6.9 Hz, 2H), 3.57 (dd, *J* = 6.2, 6.2 Hz, 1H), 2.41 (dd, *J* = 7.6, 7.4 Hz, 2H), 2.39 (td, *J* = 7.8, 7.1 Hz, 2H), 2.25 (t, *J* = 7.2 Hz, 2H), 2.10 (ddt, *J* = 14.2, 7.0, 7.0 Hz, 1H), 2.03 (ddt, *J* = 14.5, 7.0, 7.0 Hz, 1H), 1.79 (s, 3H), 1.71 (s, 3H); ^{13}C NMR (151 MHz, D_2O): δ 178.6, 173.3, 173.1, 146.8, 143.1, 128.3, 113.6, 60.1, 43.9, 37.4, 31.3, 26.1, 25.4, 15.5, 11.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_6$ 312.1453, found 312.1452 $(M-H)^-$.

Synthesis of 7'-hydroxy-*N*-geranyl-L-glutamic acid (6)



(6*E*)-8-hydroxy-3,7-dimethylocta-2,6-dienal (**SI-5**)
This compound was synthesized following modification of a literature procedure (*63*). To a suspension of selenium (IV) dioxide (0.089 g, 0.80 mmol) in CH_2Cl_2 (5 mL), a solution of citral (0.36 mL, 95%, 2.00 mmol) in CH_2Cl_2 (5 mL), and aqueous *t*-butyl hydrogen peroxide (0.51 mL, 70%, 4.00 mmol) were sequentially added and stirred for 18 h at room temperature. The reaction mixture was diluted with CH_2Cl_2 (20 mL), filtered to remove particulates, and washed with saturated aqueous $NaHCO_3$ (25 mL) and brine (25 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (1:1 hexanes:EtOAc), yielding **SI-5** as a light yellow oil (0.184 g, 1.09 mmol, 55%) as an approximately 1:1 mixture of regioisomers. R_f 0.15 (hexanes/EtOAc 2:1); IR ($CHCl_3$ cast) 3408 (br), 2927, 2863, 1667, 1441, 1387, 1192, 1014 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 9.99 (d, *J* = 8.1 Hz, 1H), 9.89 (d, *J* = 8.1 Hz, 1H), 5.92 – 5.86 (m, 2H), 5.43 – 5.34 (m, 2H), 4.00 (s, 2H), 3.99 (s, 2H), 2.63 (t, *J* = 7.5 Hz, 2H), 2.33 – 2.25 (m, 6H), 2.18 (d, *J* = 1.3 Hz, 3H), 2.00 (d, *J* = 1.4 Hz, 3H), 1.67 (d, *J* = 1.4 Hz, 3H), 1.65 (d, *J* = 1.2 Hz, 3H);^{13}C NMR (125 MHz, $CDCl_3$): δ 191.5, 191.2, 163.9, 163.8, 137.1, 136.3, 128.7, 127.5, 123.7, 123.1, 68.4, 68.3, 40.2, 32.3, 26.6, 25.3, 25.2, 17.7, 13.8; HRMS (ESI) Calculated for $C_{10}H_{16}NaO_2$ 191.1043, found 191.1044 $(M+Na)^+$.

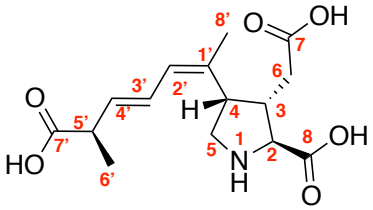


7'-hydroxy-*N*-geranyl-L-glutamic acid (**6**)

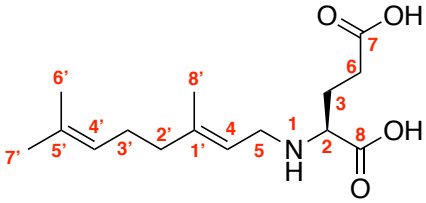
A solution of **SI-5** (0.100 g, 0.59 mmol) in methanol (2.5 mL) was added to an aqueous solution (2.5 mL) of L-glutamic acid (0.175 g, 1.19 mmol) and sodium hydroxide (0.071 mg, 1.78 mmol) and stirred for 3 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.029 g, 0.77 mmol) was added at once and stirred for 1 h. 1 N HCl was added

until pH 3, and the reaction mixture was concentrated *in vacuo*. Allylic alcohols were removed by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 17.5% B (15 min), 17.5 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **6** eluted at 15.7 minutes, whereas minor *cis*-regioisomer 7'-hydroxy-*N*-neryl-L-glutamic acid eluted at 16.5 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.087 g, 0.29 mmol, 49%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). ^{1}H NMR (600 MHz, D_2O): δ 5.35 (tq, J = 5.5, 1.5 Hz, 1H), 5.21 (tq, J = 7.8, 1.6 Hz, 1H), 3.91 (s, 2H), 3.65 (dd, J = 7.8, 3.4 Hz, 2H), 3.59 (dd, J = 7.4, 5.4 Hz, 1H), 2.51 – 2.44 (m, 2H), 2.16 (t, J = 6.9 Hz, 2H), 2.14 – 2.07 (m, 3H), 2.02 (ddt, J = 14.0, 7.0, 7.0 Hz, 1H), 1.66 (d, J = 1.3 Hz, 3H), 1.59 (d, J = 1.3 Hz, 3H); ^{13}C NMR (151 MHz, D_2O): δ 175.3, 171.2, 145.8, 133.2, 124.9, 111.6, 66.2, 58.3, 42.5, 37.1, 28.7, 23.5, 23.5, 14.2, 11.5; HRMS (ESI) Calculated for $C_{15}H_{24}NO_5$ 298.1660, found 298.1661 $(M-H)^-$.

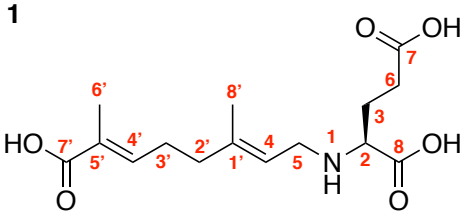
NMR and Compound Characterization



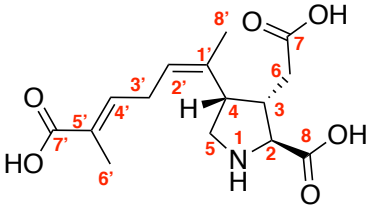
domoic acid
1



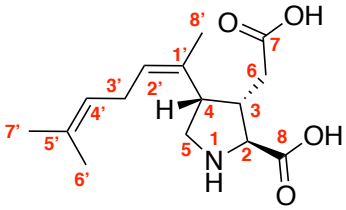
***N*-geranyl-L-glutamic acid**
2



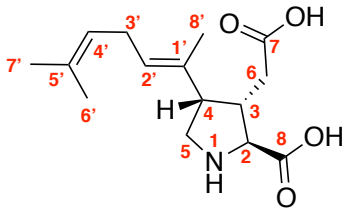
7'-carboxy-*N*-geranyl-L-glutamic acid
3



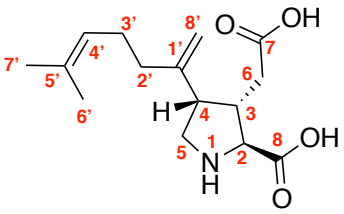
isodomoic acid A
4



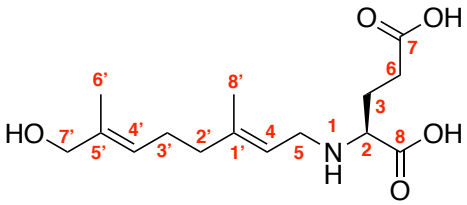
dainic acid A
5a



dainic acid B
5b



dainic acid C
5c



7'-hydroxy-*N*-geranyl-L-glutamic acid
6

Numbering scheme for synthetic and enzyme-isolated compounds **2-6** based on unified numbering scheme for domoic acid (**1**). (*64*)

^{1}H NMR table for compounds 2 – 6

^{1}H	2 (DabA)	2 (synth)	3 (synth)	4 (DabC)	5a (DabC)	5b (DabC)	5c (DabC)	6 (synth)
2	3.57, dd (6.6, 5.7)	3.57, dd (6.9, 5.6)	3.57, dd (6.2, 6.2)	3.97, d (7.3)	3.93, d (6.5)	4.00, d (3.2)	4.06, d (2.4)	3.59, dd (7.4, 5.4)
3	2.08, m 2.02, ddt (14.5, 6.8, 6.8)	2.08, m 2.01, ddt (14.4, 7.0, 7.0)	2.10, ddt (14.2, 7.0, 7.0) 2.03, ddt (14.5, 7.0, 7.0)	3.01, ddd (12.7, 10.1, 5.4)	2.95, ddd (9.0, 6.7, 6.5)	3.01, ddd (8.8, 7.3, 2.5)	3.03, m	2.09, m 2.02, ddt (14.0, 7.0, 7.0)
4	5.21, t (7.3)	5.21, t (7.8)	5.27, t (7.6)	3.63, ddd (7.8, 7.6, 7.6)	3.55, dd (7.8, 7.8)	2.94, dd (7.9, 7.6)	2.97, m	5.21, tq (7.8, 1.6)
5a	3.64, t (6.9)	3.64, dd (7.8, 5.7)	3.67, t (6.9)	3.69, dd (11.7, 8.1)	3.63, dd (11.8, 8.1)	3.55, dd (11.9, 7.5)	3.58, dd (11.9, 7.6)	3.65, dd (7.8, 3.4)
5b	3.64, t (6.9)	3.64, dd (7.8, 5.7)	3.67, t (6.9)	3.46, dd (11.7, 7.7)	3.42, dd (11.8, 8.1)	3.40, dd (11.4, 11.4)	3.42, dd (11.8, 11.8)	3.65, dd (7.8, 3.4)
6a	2.44, m	2.45, m	2.41, dd (7.6, 7.4)	2.67, dd (16.5, 6.5)	2.40, dd (15.6, 6.7)	2.20, dd (15.4, 6.5)	2.25, dd (15.5, 6.0)	2.47, m
6b	2.44, m	2.45, m	2.41, dd (7.6, 7.4)	2.50, dd (16.5, 8.6)	2.30, dd (15.6, 8.5)	2.09, m	2.07, m	2.47, m
2'a	2.10, m	2.09, m	2.25, t (7.2)	5.54, t (7.5)	5.43, t (7.2)	5.14, m	2.14, m	2.12, m
2'b	2.10, m	2.09, m	2.25, t (7.2)	-	-	-	2.12, m	2.12, m
3'a	2.11, m	2.10, m	2.39, td (7.8, 7.1)	2.96, ddd (15.6, 7.7, 7.2)	2.76, ddd (16.2, 8.0, 7.2)	2.76, m	2.16, m	2.16, t (6.9)
3'b	2.11, m	2.10, m	2.39, td (7.8, 7.1)	2.87, ddd (15.8, 7.2, 7.0)	2.60, m	2.76, m	2.16, m	2.16, t (6.9)
4'	5.12, t (6.7)	5.12, t (6.8)	6.71, t (6.8)	6.67, t (7.2)	5.15, m	5.19, t (7.2)	5.19, t (8.6)	5.35, tq (5.5, 1.5)
6'	1.57, s	1.57, s	1.79, s	1.83, s	1.62, s	1.64, s	1.61, s	1.59, d (1.3)
7'	1.64, s	1.64, s	-	-	1.68, s	1.69, s	1.68, s	3.91, s
8'	1.66, s	1.66, s	1.71, s	1.76, s	1.69, s	1.68, s	5.04, s 4.77, s	1.66, d (1.3)

all spectra were collected in D_2O and referenced to sodium formate (δ 8.44).

^{13}C NMR table for compounds 2 – 6

^{13}C	2 (DabA)	2 (synth)	3 (synth)	4 (DabC)	5a (DabC)	5b (DabC)	5c (DabC)	6 (synth)
2	58.5	58.5	60.1	65.1	65.1	65.7	65.7	58.3
3	23.9	23.8	25.4	42.4	43.7	42.2	41.5	23.5
4	111.5	111.5	113.6	40.2	40.3	47.0	43.9	111.6
5	42.5	42.5	43.9	46.8	46.8	46.0	45.9	42.5
6	29.7	29.3	31.3	34.1	36.2	36.1	35.9	28.7
7	176.6	176.3	178.6	176.5	179.4	179.8	179.8	175.3
8	171.9	171.8	173.3	172.9	173.5	173.4	173.7	171.2
1'	146.4	146.4	146.8	131.0	129.4	129.8	144.1	145.8
2'	37.6	37.5	37.4	128.1	130.2	126.1	35.4	37.1
3'	24.0	23.9	26.1	27.1	26.3	26.4	25.4	23.5
4'	122.4	122.5	143.1	141.0	122.2	122.3	123.4	124.9
5'	132.5	132.7	128.3	128.0	134.0	133.9	134.0	133.2
6'	15.7	15.7	11.9	11.8	16.9	16.9	16.6	11.5
7'	23.5	23.5	173.1	172.9	24.7	24.7	24.5	66.2
8'	14.3	14.4	15.5	21.3	21.4	15.9	111.6	14.2

All spectra were collected in D_2O and referenced to sodium formate (δ 171.7). Carbon signals were derived from HSQC and HMBC correlations.

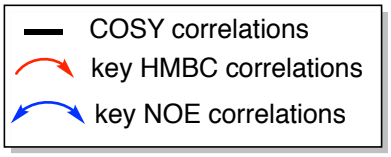
HMBC correlations table for Dab enzyme-isolated compounds 2, 4, 5a-c

^{13}C	2 (DabA)	4 (DabC)	5a (DabC)	5b (DabC)	5c (DabC)
2	3, 5, 6	6b	3, 4, 5a, 5b, 6a, 6b	5a, 6a, 6b	6a, 6b
3	2, 6	2, 4, 6b	2, 4, 5a, 5b, 6a, 6b, 8'	4, 5a, 6a, 6b	5a, 6a, 6b
4	5, 2', 8'	5a, 6a, 6b, 2', 8'	2, 3, 5a, 5b, 6a, 6b, 2', 8'	2, 5b, 6a, 6b, 2', 8'	5b, 8'a, 8'b
5	2, 4	4	2, 3, 4	4	-
6	2, 3	2, 4	2, 3, 4	2, 4	-
7	3, 6	6a, 6b	3, 6a, 6b	6a, 6b	6a, 6b
8	2, 3	2	2, 3	2	2
1'	5, 2', 3', 8'	4, 5a, 5b, 2', 8'	3, 5a, 5b	4, 5b, 3', 8'	4, 5b, 2'b
2'	4, 3', 8'	4, 3'a, 3'b, 8'	4, 3'a, 3'b, 8'	4, 3', 8'	8'a, 8'b
3'	2', 4'	2', 4'	2', 4'	4'	-
4'	2', 3', 6', 7'	3'a, 3'b, 6'	2', 3'a, 3'b, 6', 7'	3', 6', 7'	6', 7'
5'	3', 6', 7'	6'	3'a, 3'b, 6', 7'	6', 7'	6', 7'
6'	7'	4'	4', 7'	4', 7'	7'
7'	4', 6'	4', 6'	4', 6'	4', 6'	6'
8'	4, 2'	4, 2'	4, 2'	4, 2'	4, 2'b

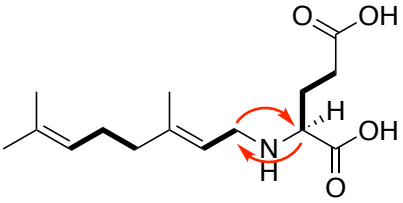
HMBC correlations table for Dab synthetic compounds 2, 3, 6

^{13}C	**2 (synthetic)**	**3 (synthetic)**	**6 (synthetic)**
2	3, 5, 6	3, 5, 6	3, 5, 6
3	2, 6	2, 6	2, 6
4	5, 2', 8'	5, 2', 8'	5, 2', 8'
5	2, 4	2	2
6	2, 3	2, 3	2, 3
7	3, 6	3, 6	3, 6
8	2, 3	2, 3	2, 3
1'	5, 2', 3', 8'	5, 2', 3', 8'	5, 2', 3', 8'
2'	4, 3', 8'	4, 3', 8'	4, 3', 8'
3'	2', 4'	2'	2'
4'	2', 3', 6', 7'	2', 3', 6'	2', 3', 6', 7'
5'	3', 6', 7'	3', 6'	6', 7'
6'	4', 7'	4'	4', 7'
7'	4', 6'	4', 6'	4', 6'
8'	4, 2'	4, 2'	4, 2'

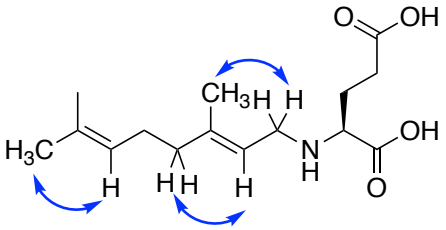
NMR correlations for compounds **2 – 6**



N-geranyl-L-glutamic acid – DabA product (**2**):

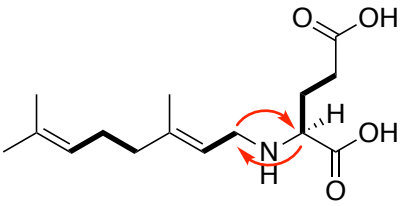


***N*-geranyl-L-glutamic acid**
2
(enzymatic)

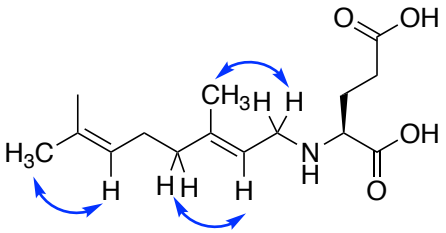


***N*-geranyl-L-glutamic acid**
2
(enzymatic)

N-geranyl-L-glutamic acid – synthetic (**2**):

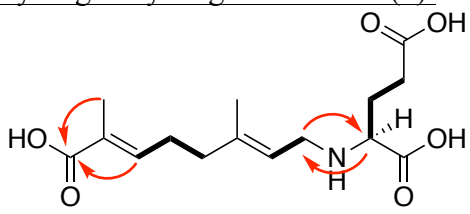


***N*-geranyl-L-glutamic acid**
2
(synthetic)

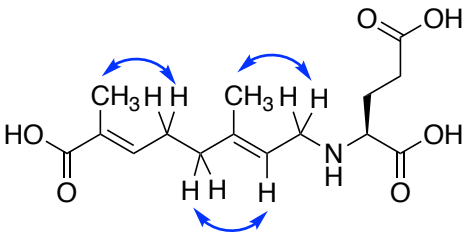


***N*-geranyl-L-glutamic acid**
2
(synthetic)

7'-carboxy-*N*-geranyl-L-glutamic acid (**3**):

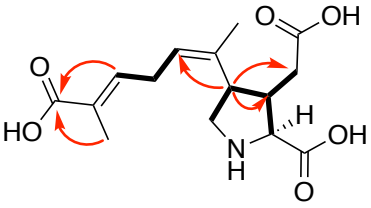


7'-carboxy-*N*-geranyl-L-glutamic acid
3

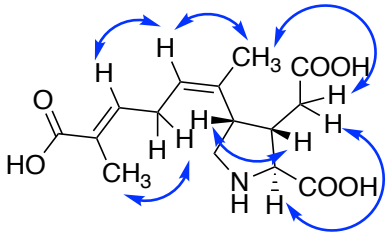


7'-carboxy-*N*-geranyl-L-glutamic acid
3

isodomoic acid A (**4**):

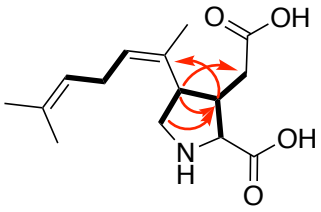


isodomoic acid A
4

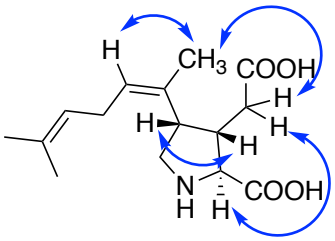


isodomoic acid A
4

dainic acid A (**5a**):

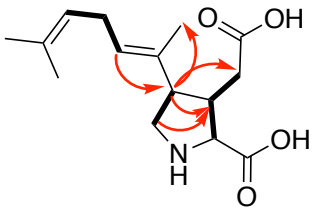


dainic acid A
5a

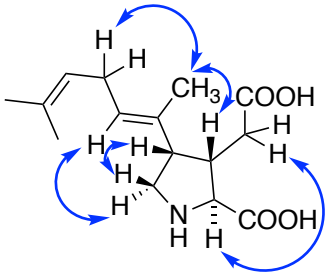


dainic acid A
5a

dainic acid B (**5b**):

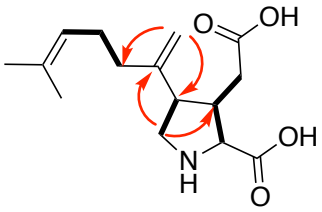


dainic acid B
5b

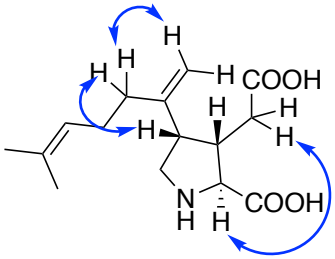


dainic acid B
5b

dainic acid C (**5c**):

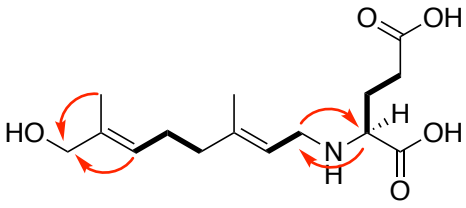


dainic acid C
5c

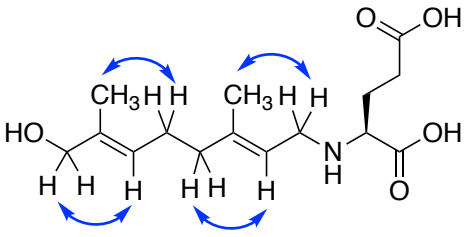


dainic acid C
5c

7'-hydroxy-*N*-geranyl-L-glutamic acid (**6**):

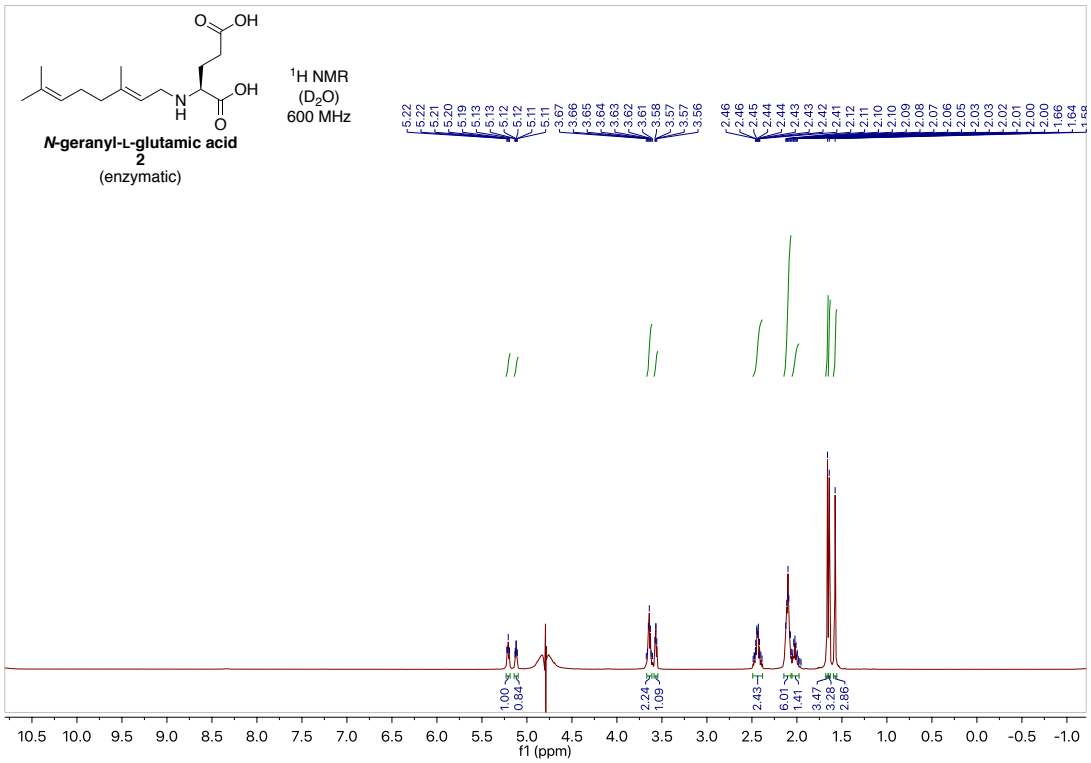


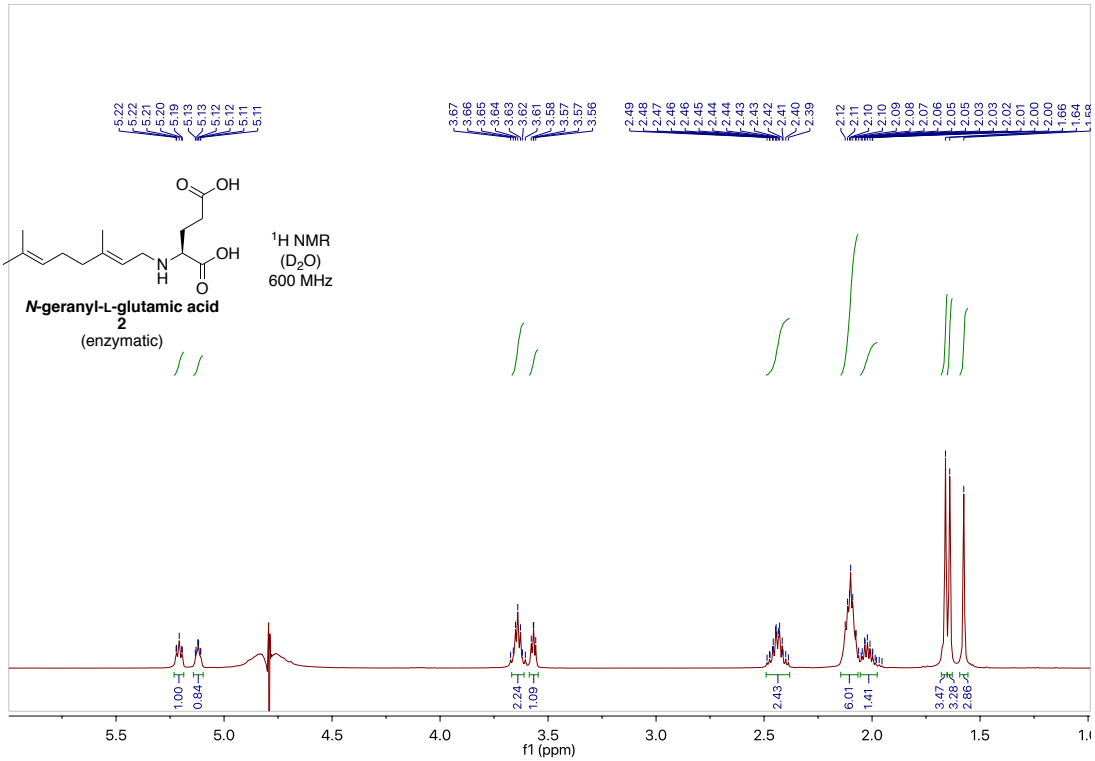
7'-hydroxy-*N*-geranyl-L-glutamic acid
6

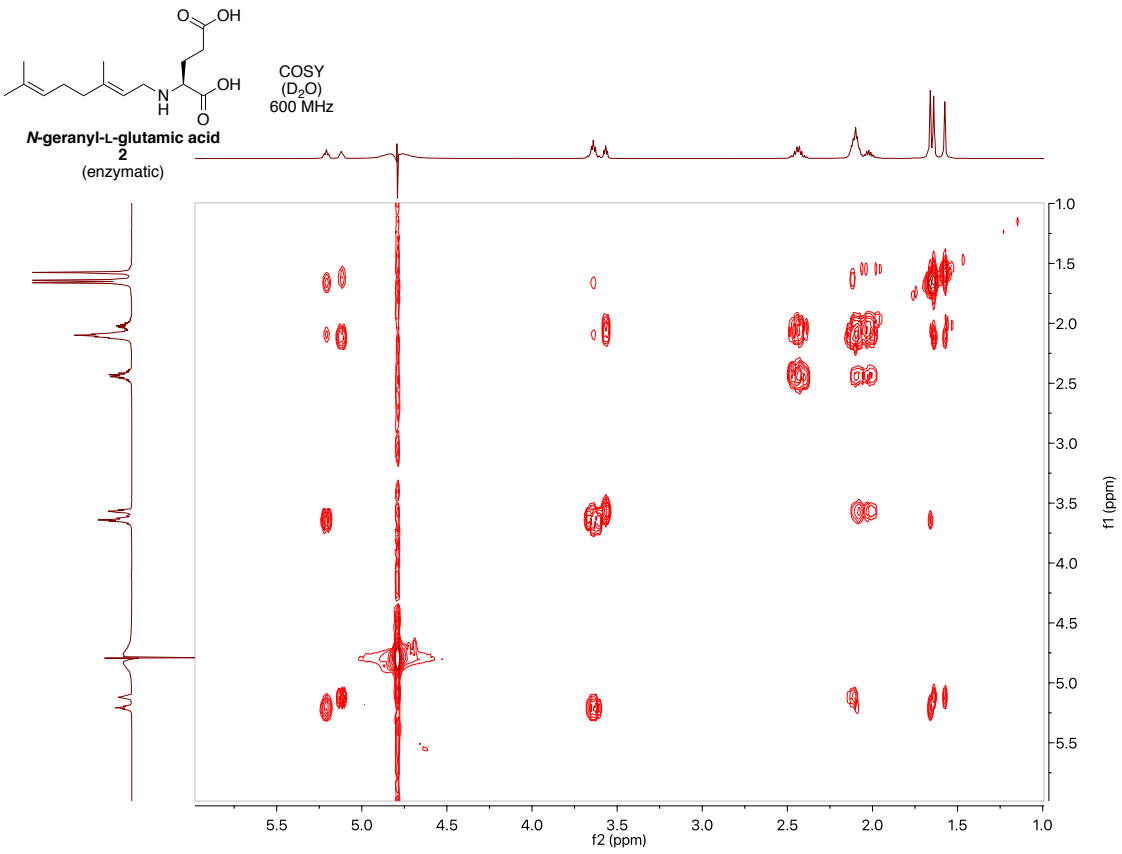


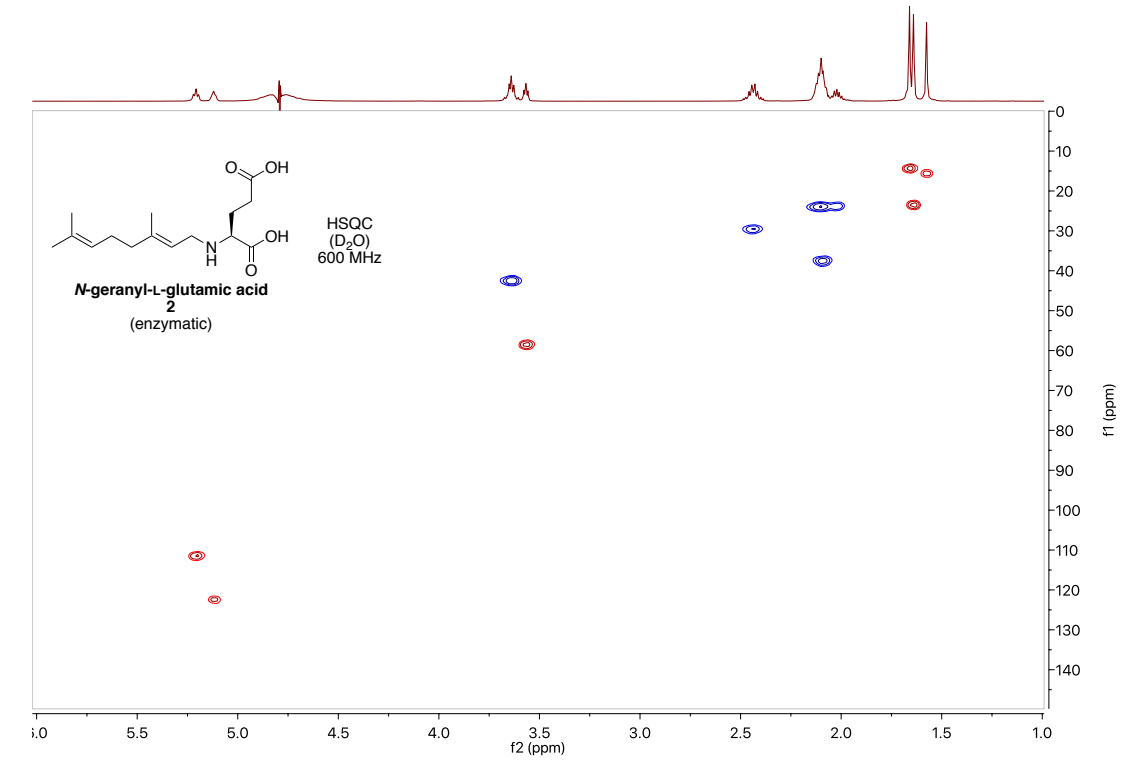
7'-hydroxy-*N*-geranyl-L-glutamic acid
6

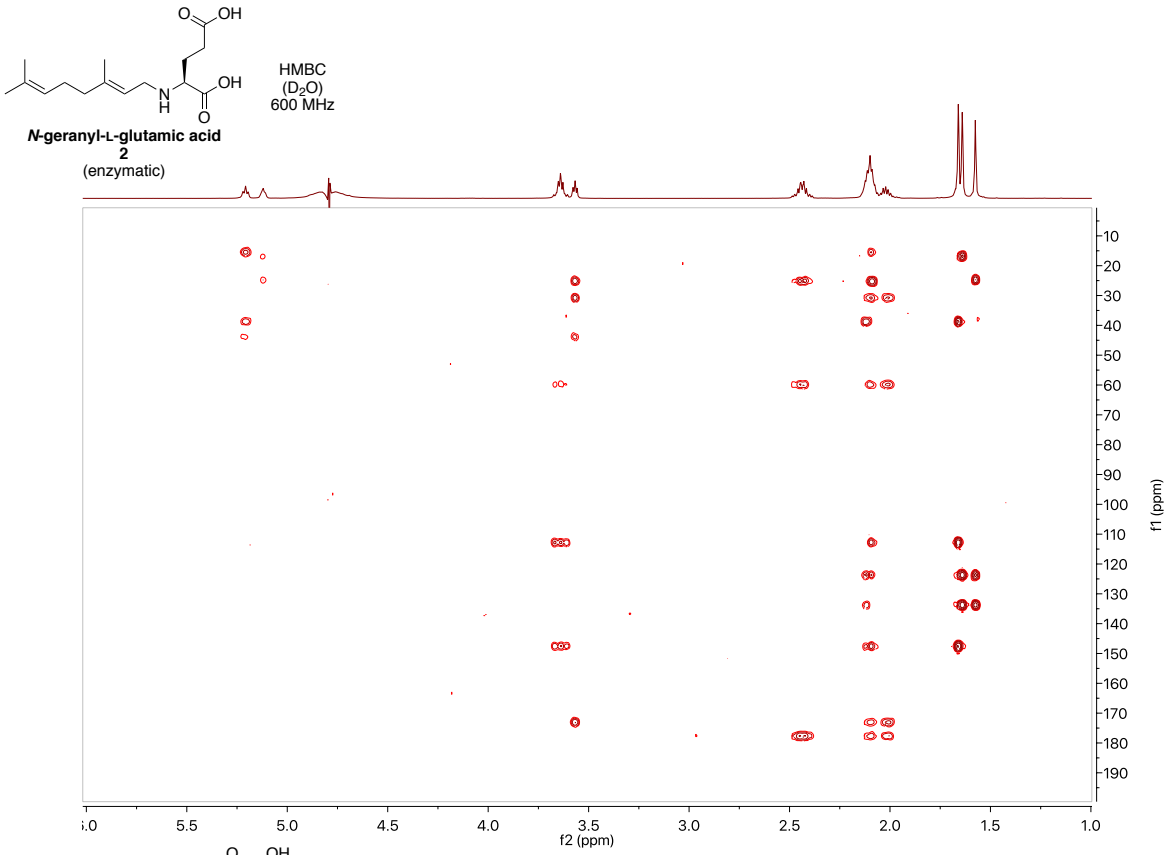
N-geranyl-L-glutamic acid – DabA product (**2**):

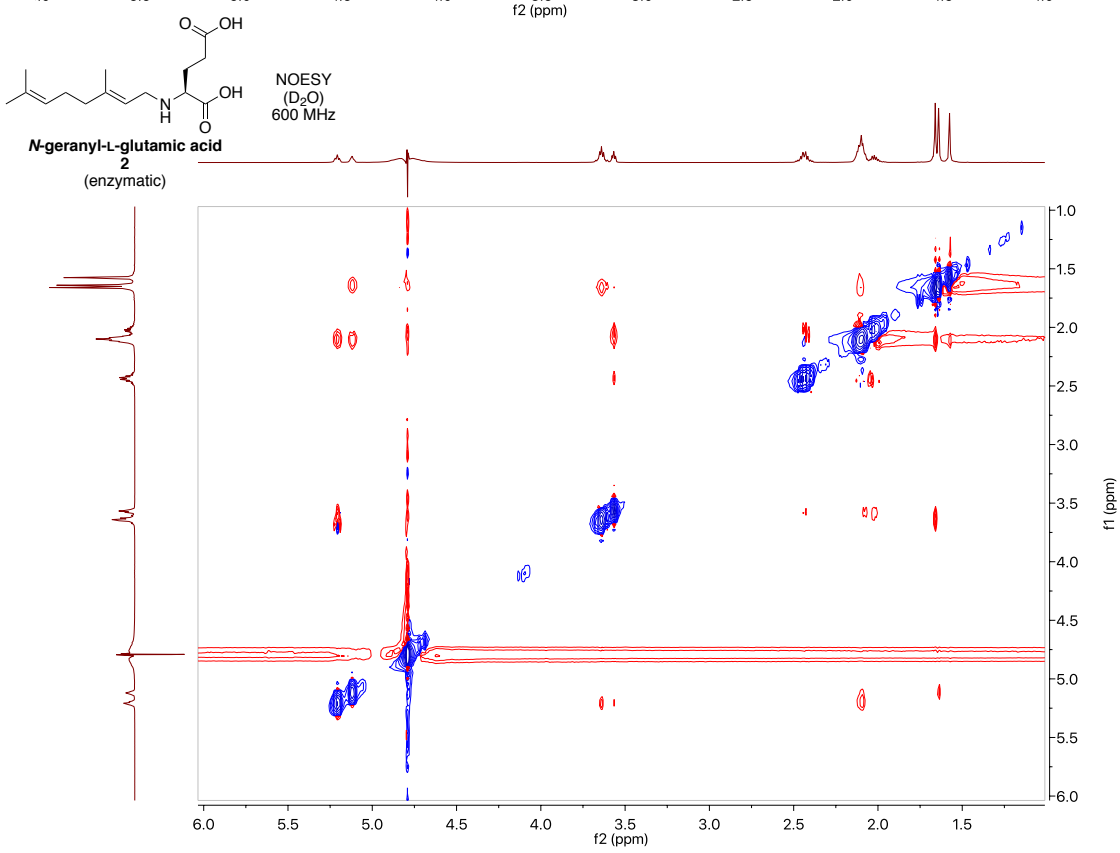




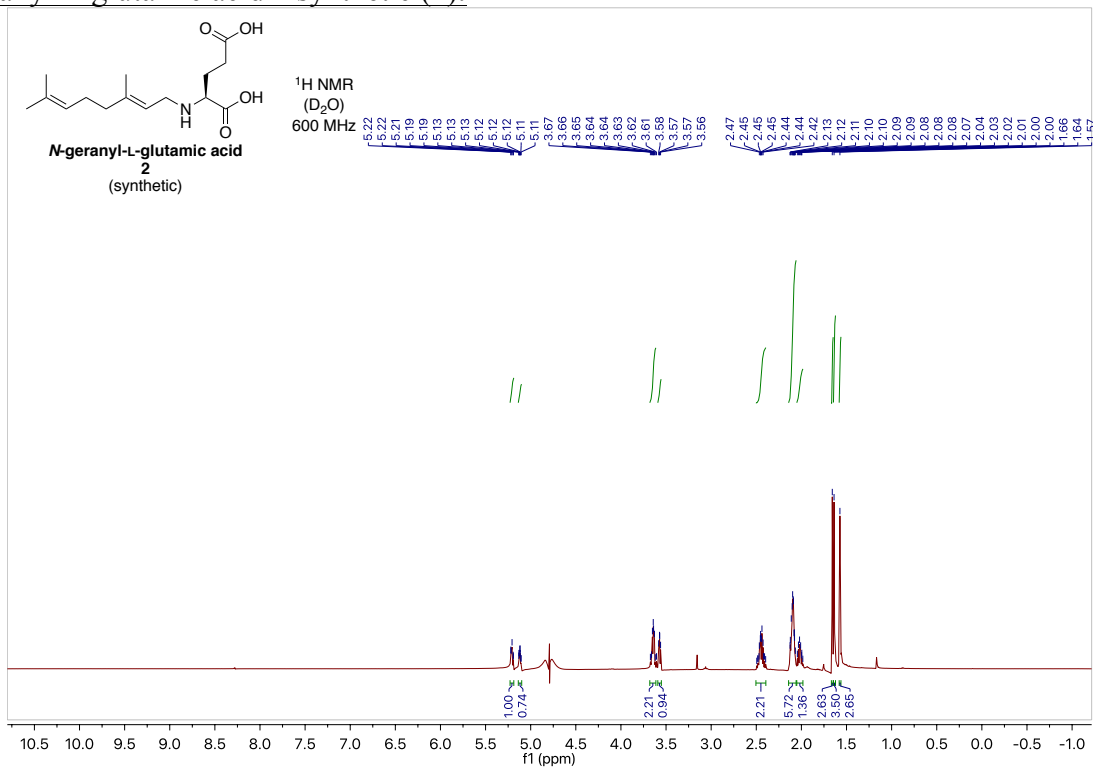

COSY
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

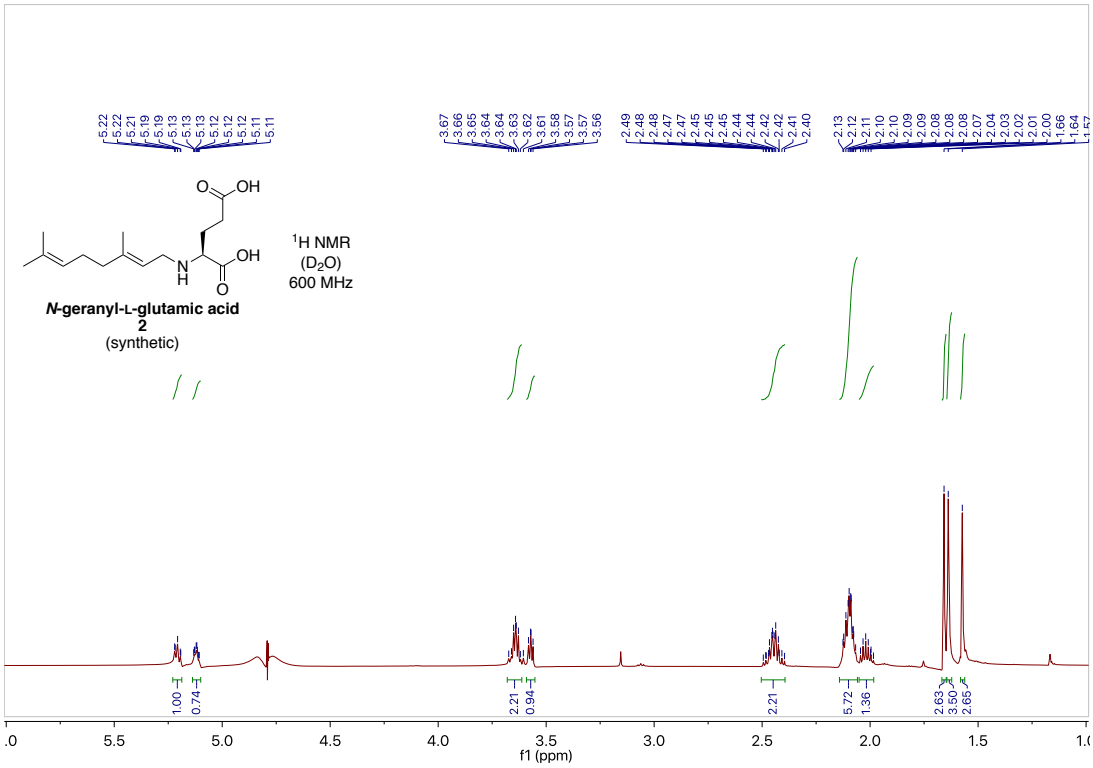

HSQC
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

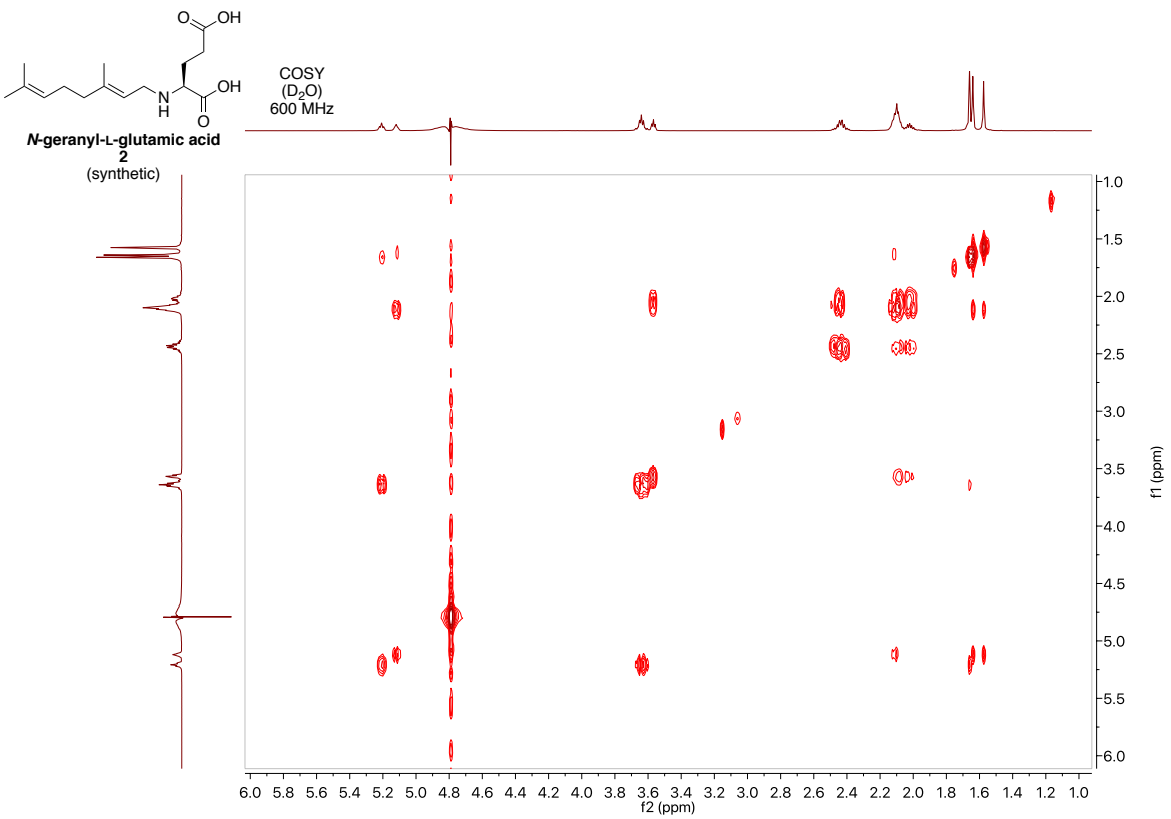

HMBC
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

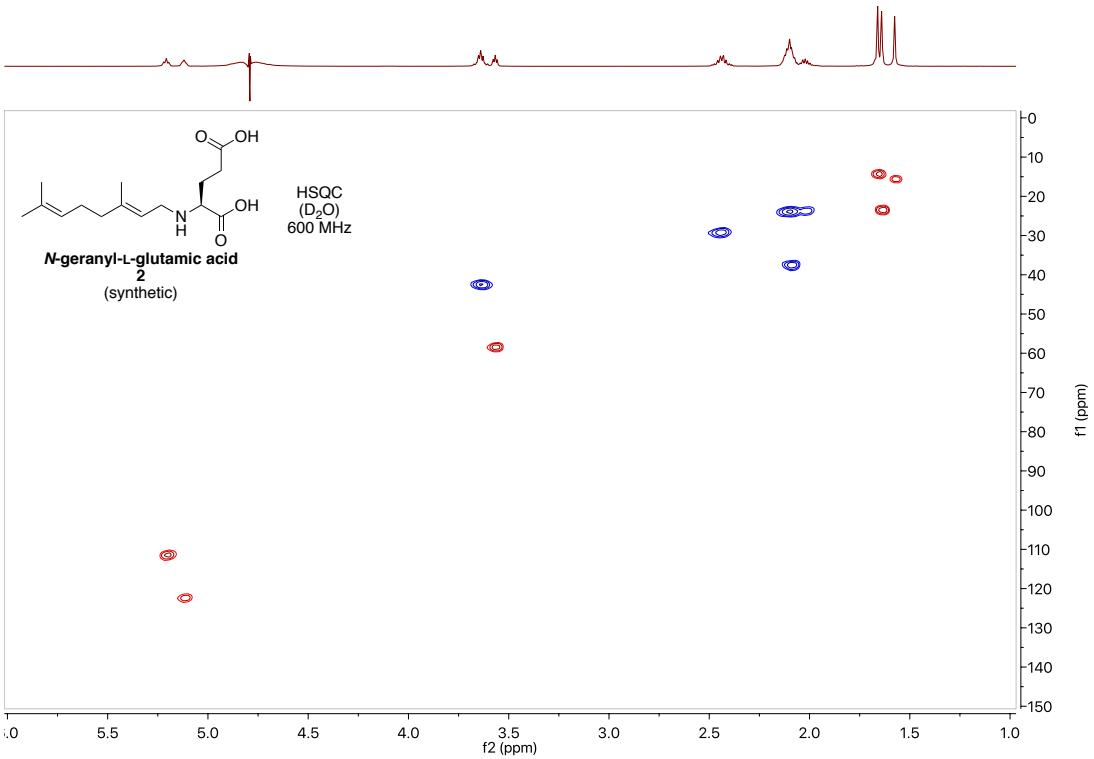

NOESY
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

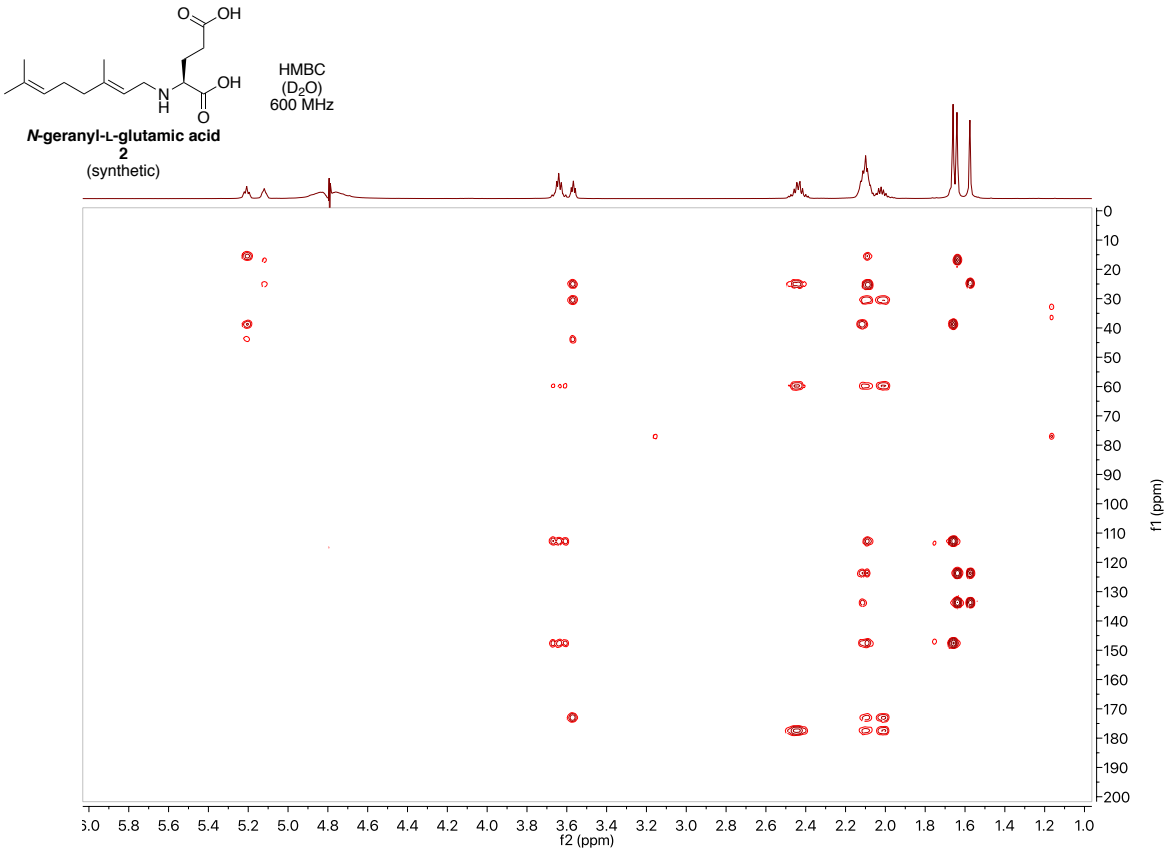
N-geranyl-L-glutamic acid – synthetic (**2**):

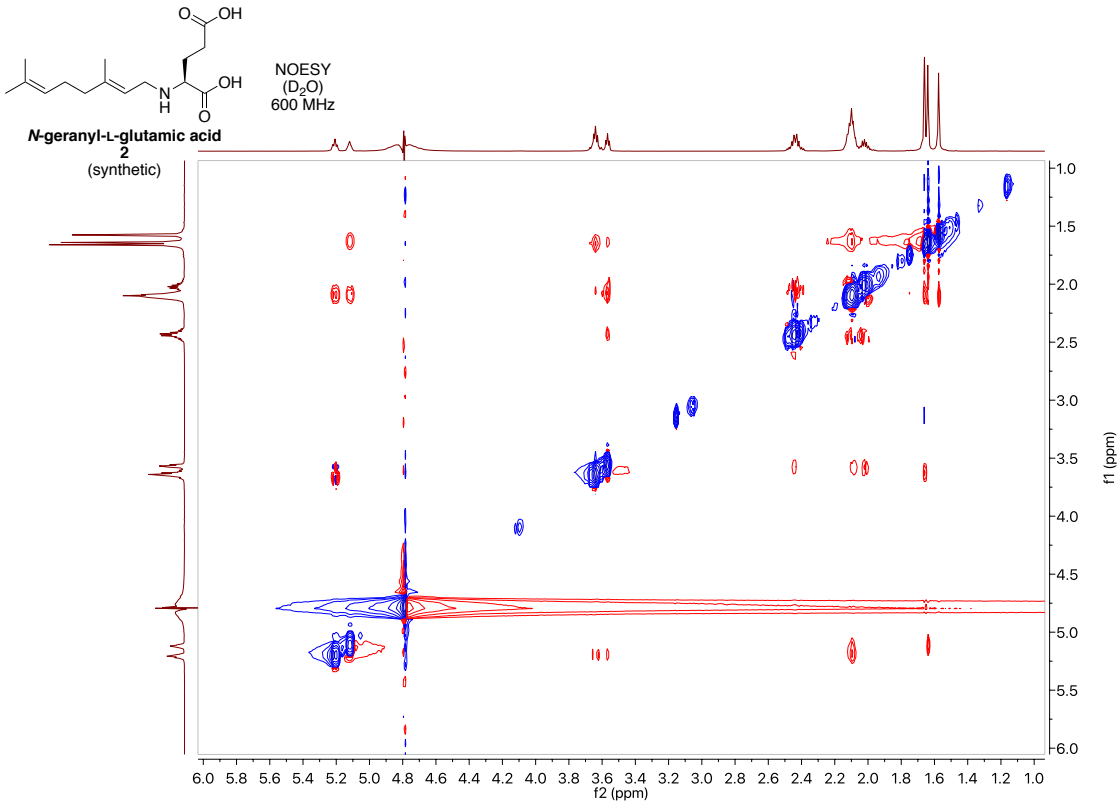




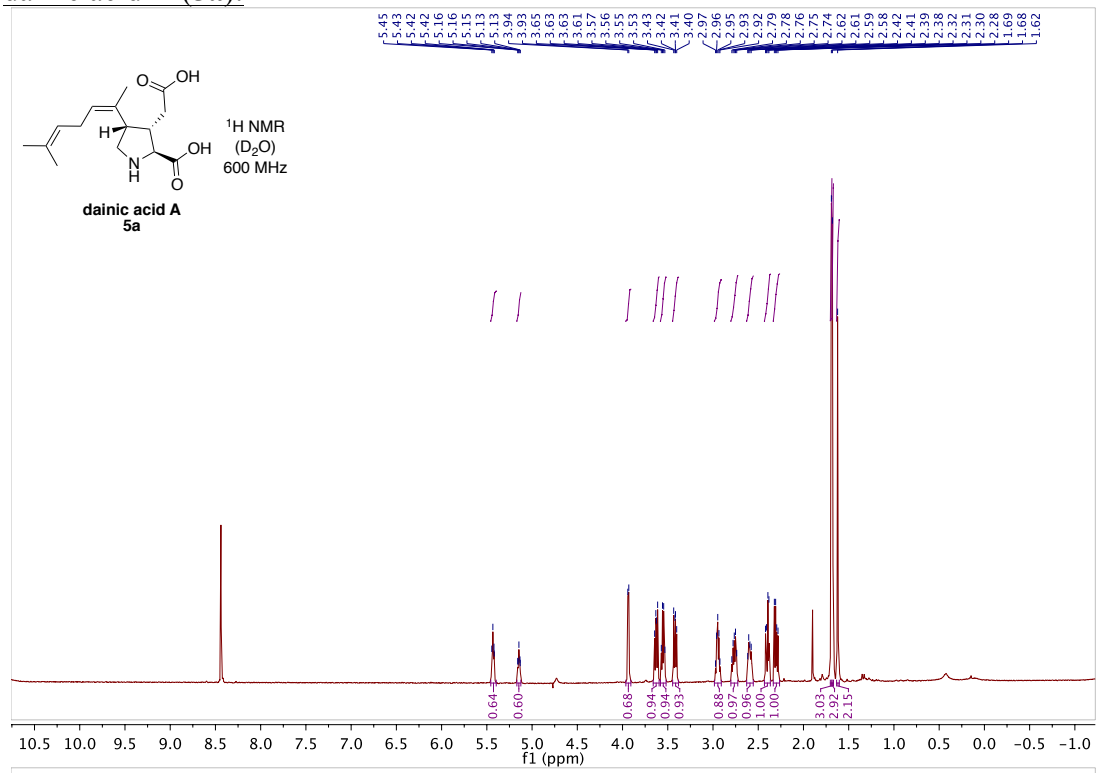

N-geranyl-L-glutamic acid
2
(synthetic)
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

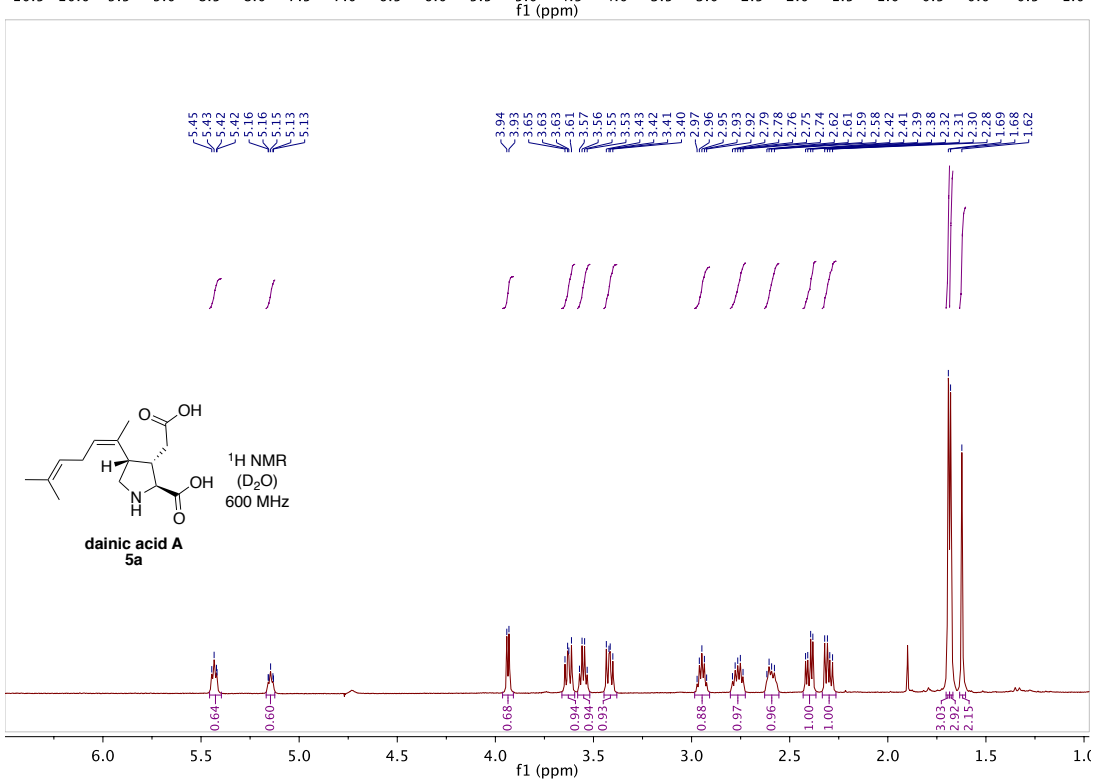

N-geranyl-L-glutamic acid
2
(synthetic)
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

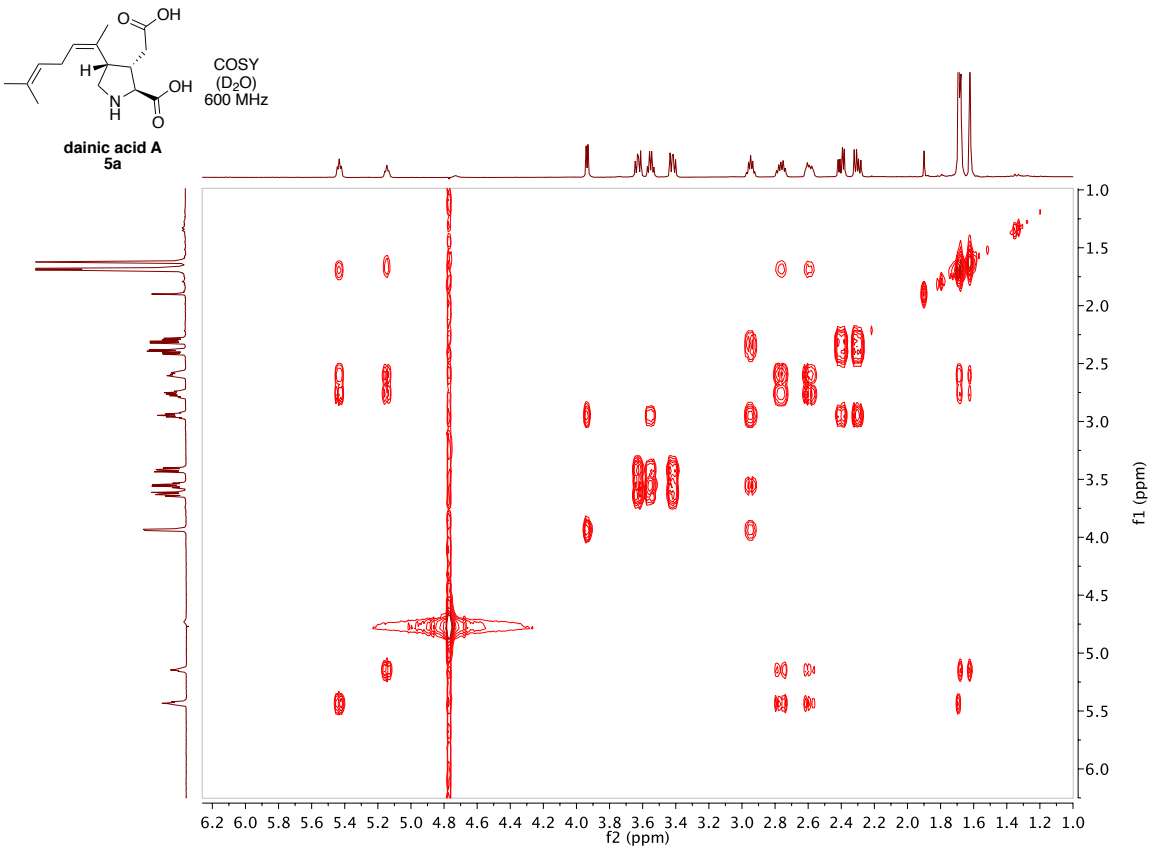

HMBC
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(synthetic)
f2 (ppm)
f1 (ppm)

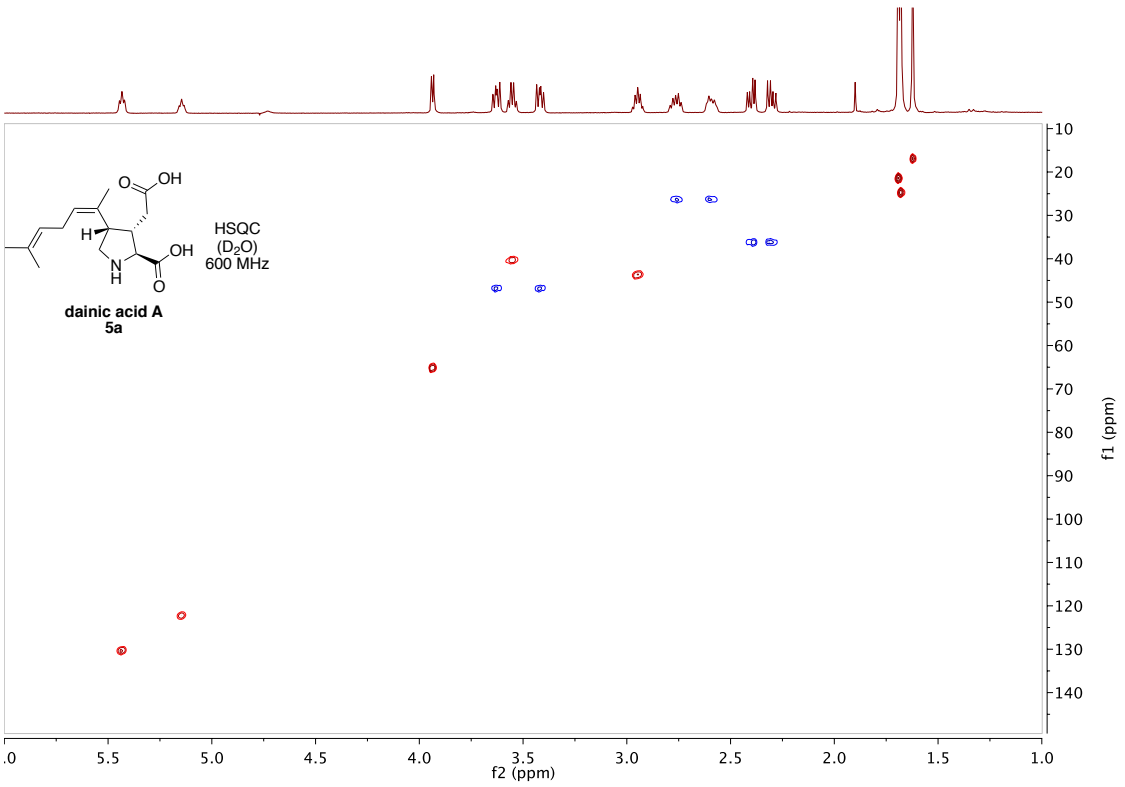

NOESY
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(synthetic)
f2 (ppm)
f1 (ppm)

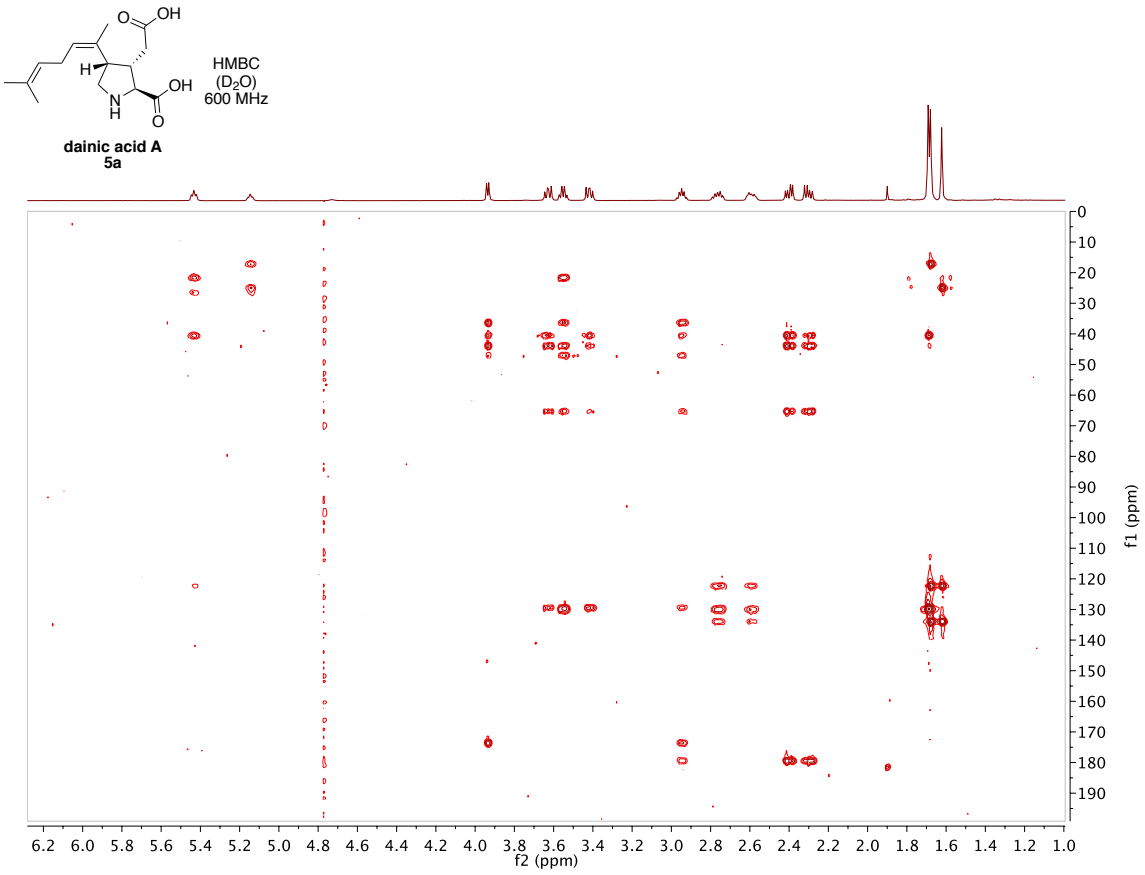
dainic acid A (**5a**):

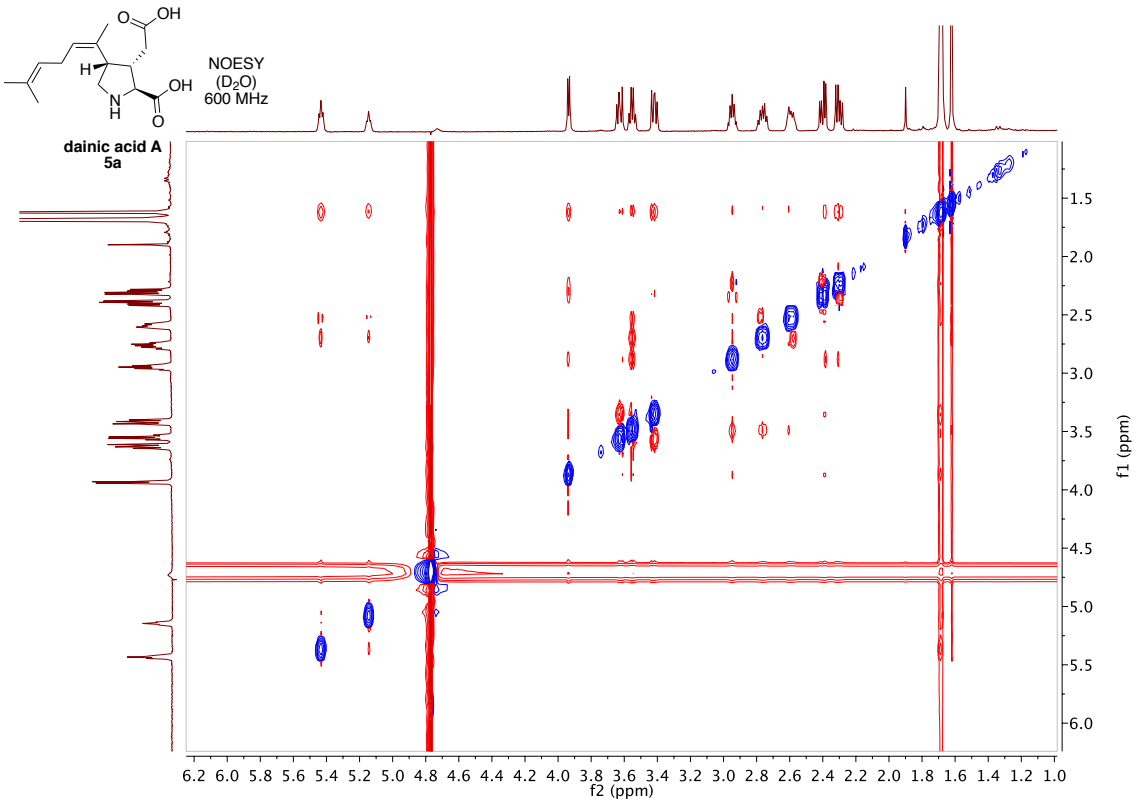




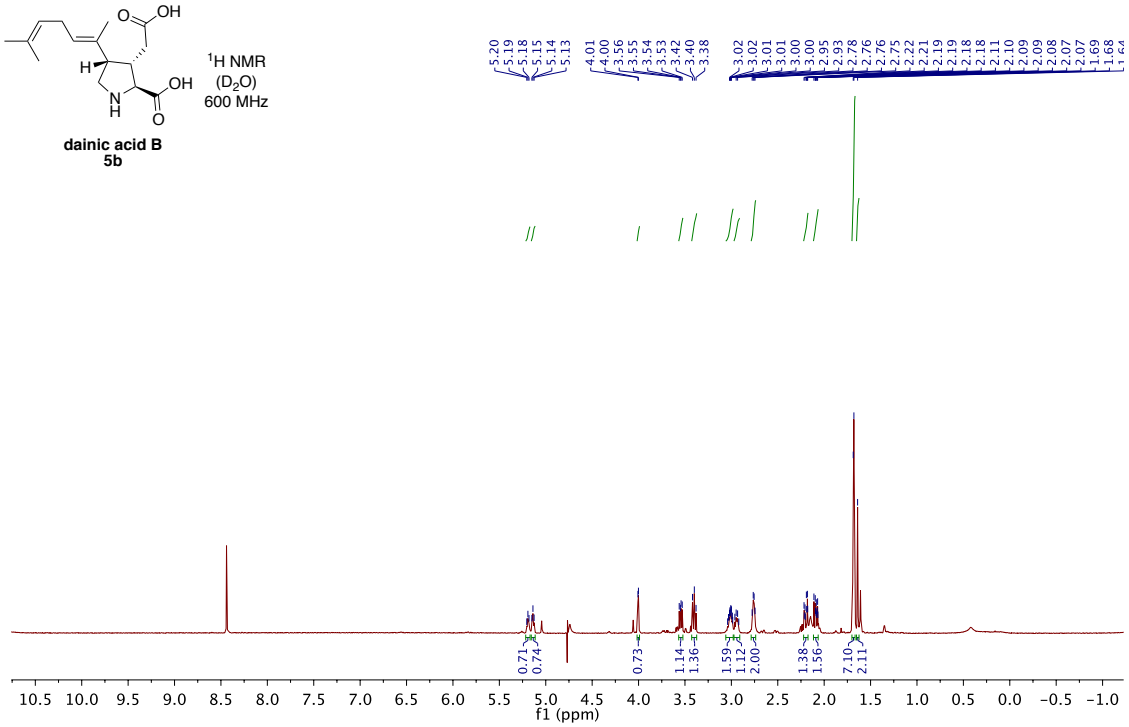

COSY
(D2O)
600 MHz
dainic acid A
5a
f1 (ppm)
f2 (ppm)

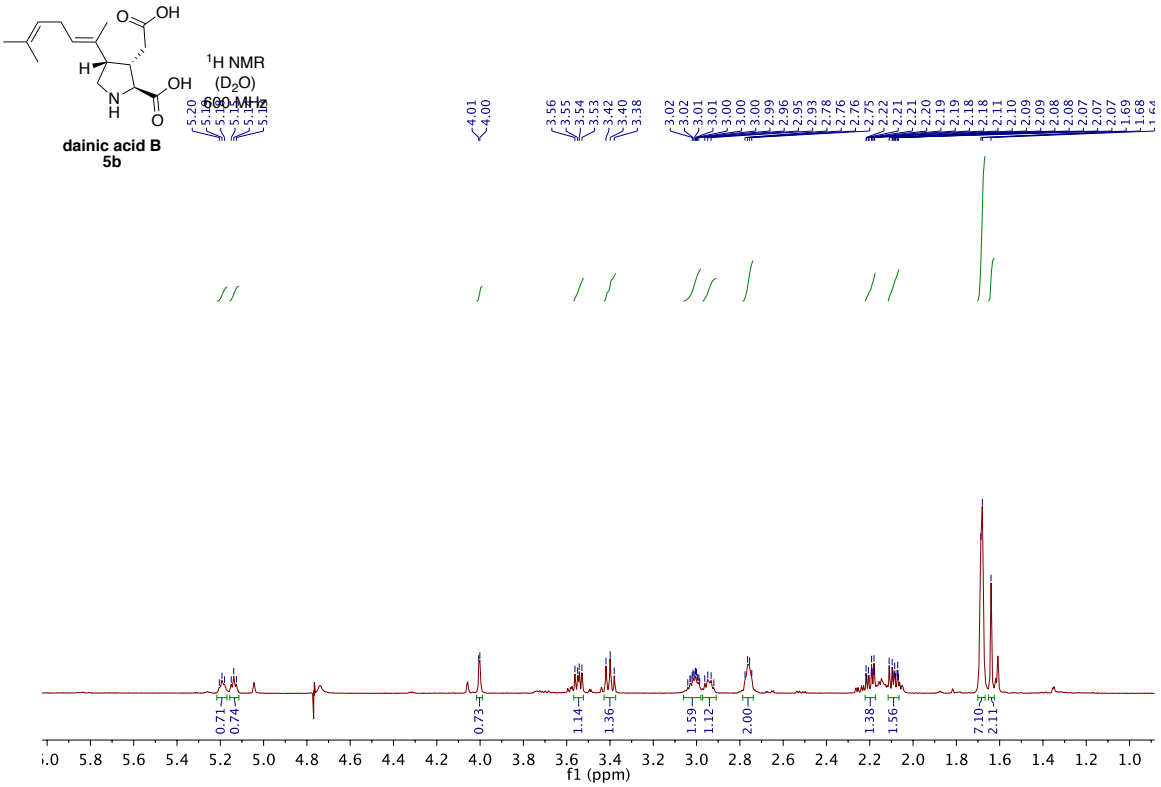

HSQC
(D2O)
600 MHz
dainic acid A
5a
f1 (ppm)
f2 (ppm)

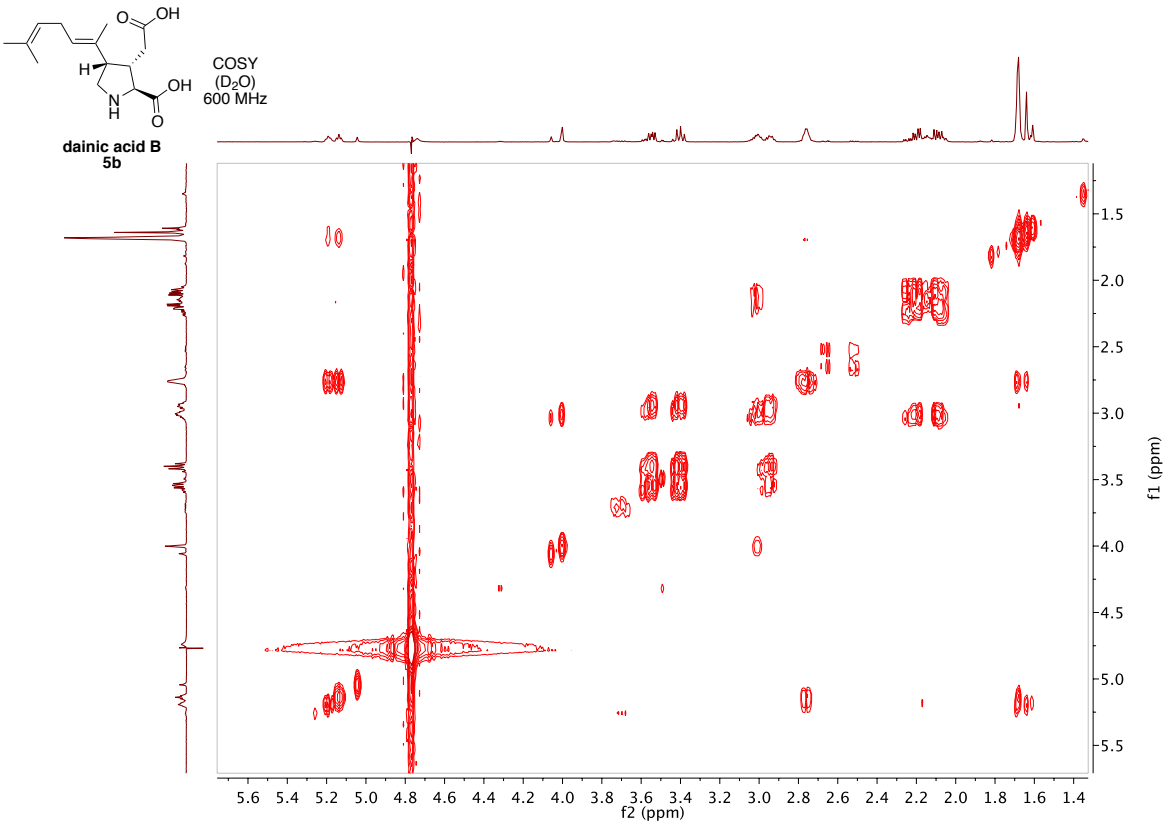

HMBC
(D2O)
600 MHz
dainic acid A
5a
f2 (ppm)
f1 (ppm)

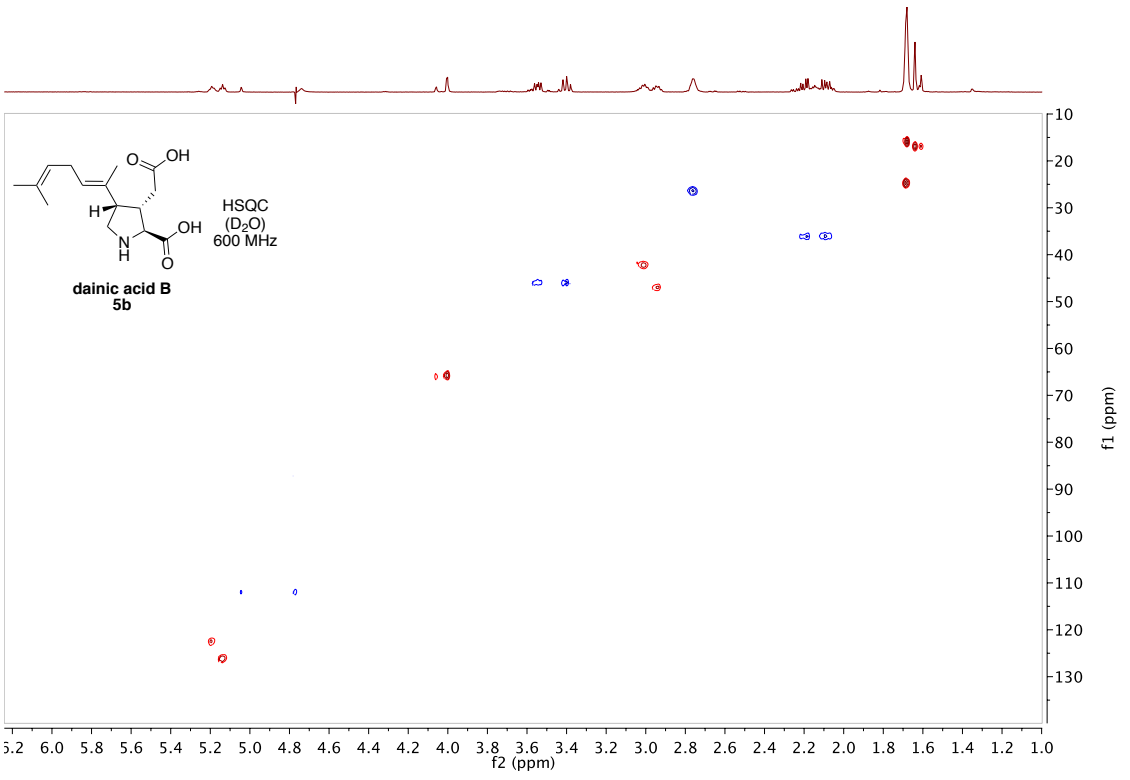

NOESY
(D2O)
600 MHz
dainic acid A
5a
f2 (ppm)
f1 (ppm)

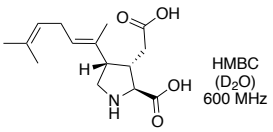
dainic acid B (**5b**):





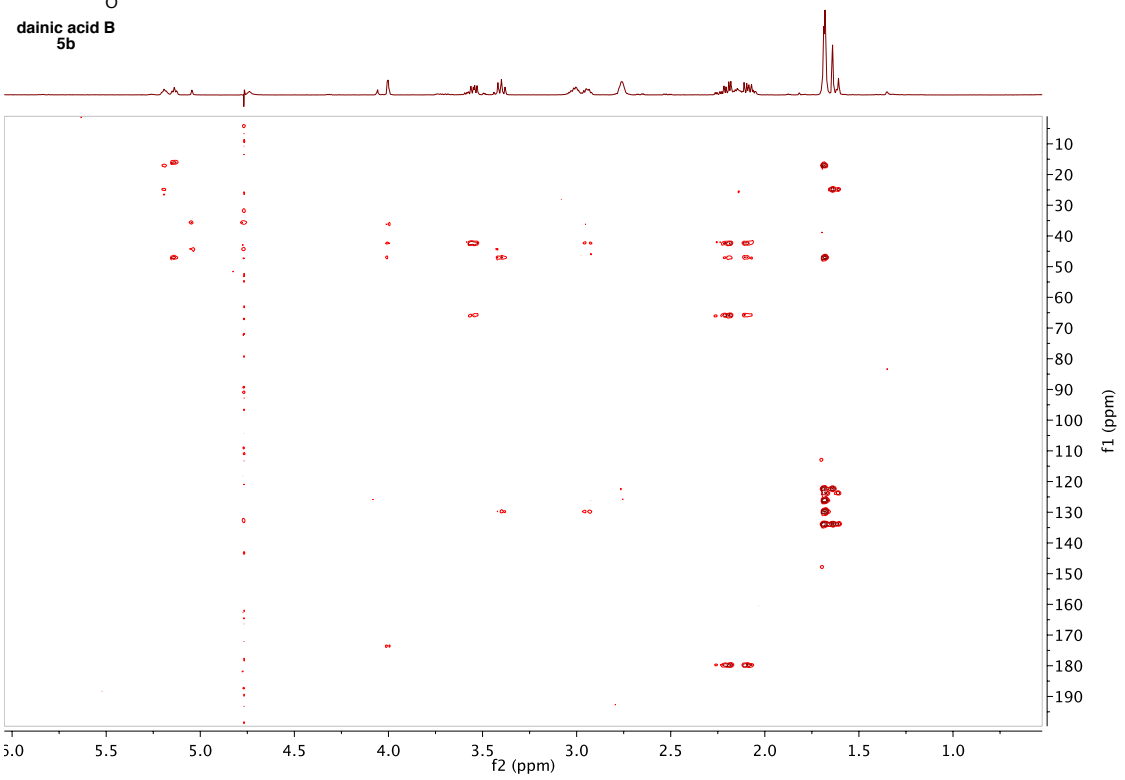

dainic acid B
5b
COSY
(D2O)
600 MHz
f1 (ppm)
f2 (ppm)

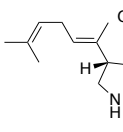

dainic acid B
5b
HSQC
(D2O)
600 MHz
f1 (ppm)
f2 (ppm)



HMBC
(D_2O)
600 MHz

dainic acid B
5b

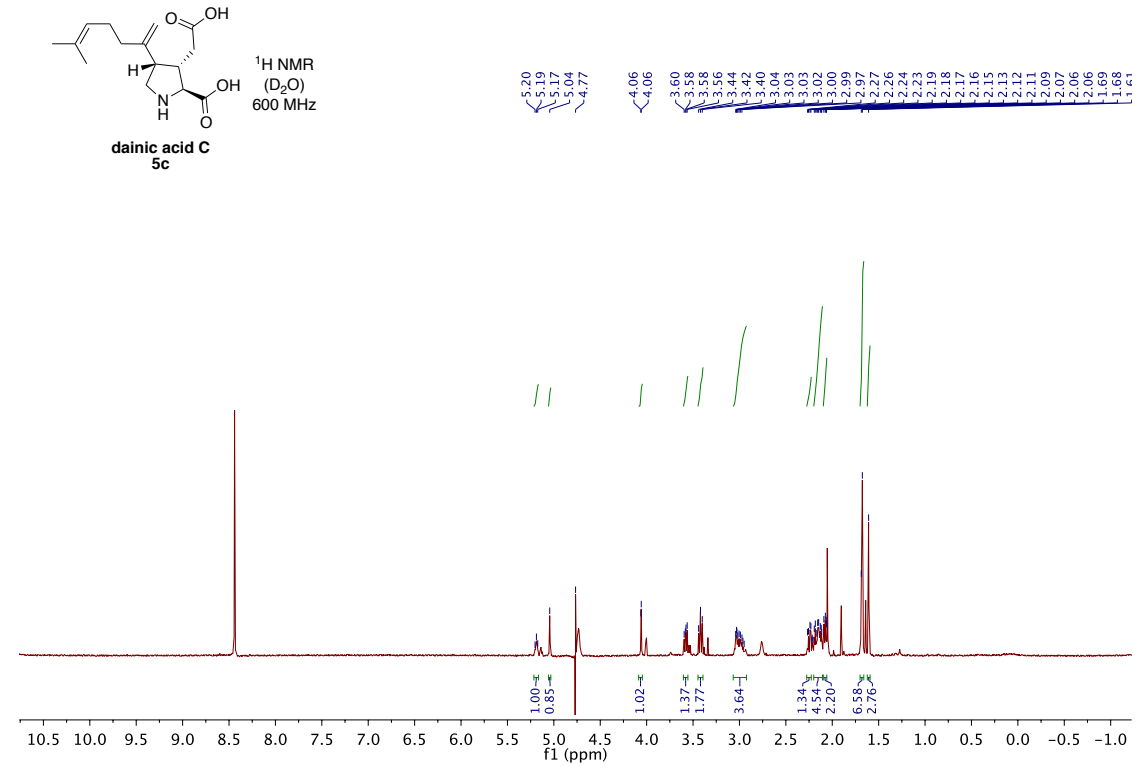


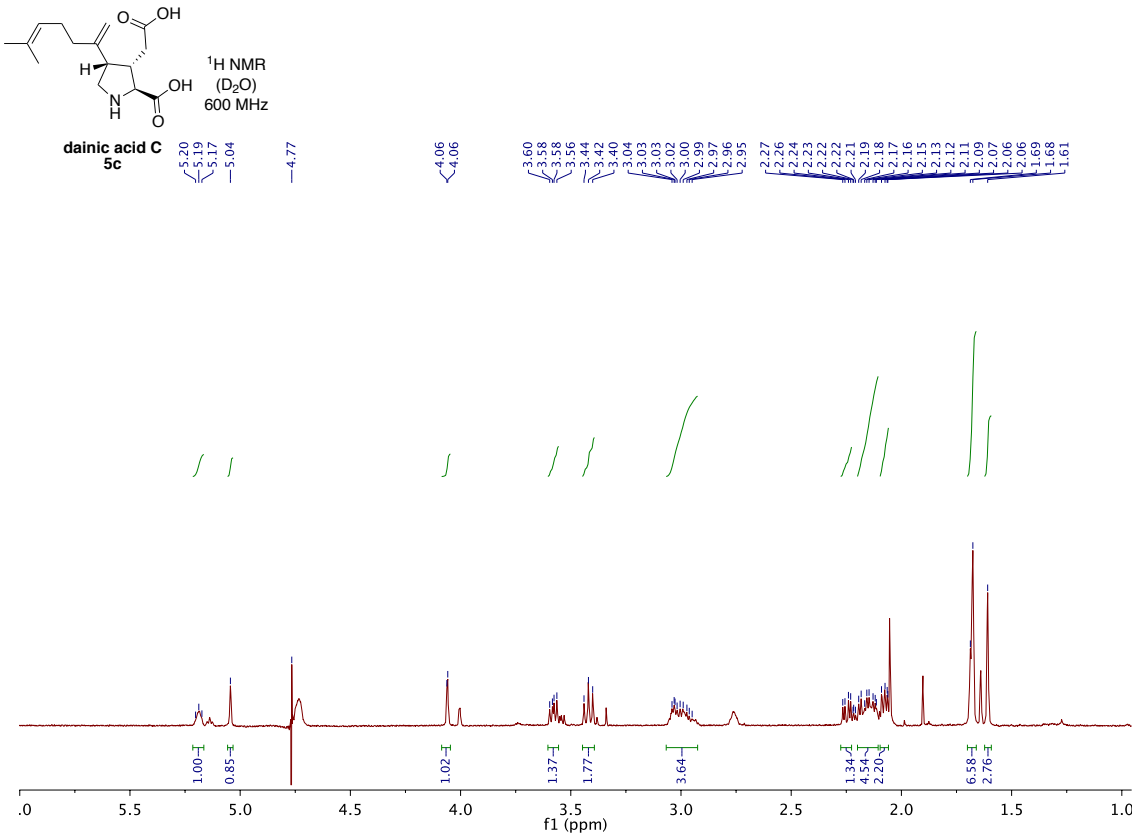


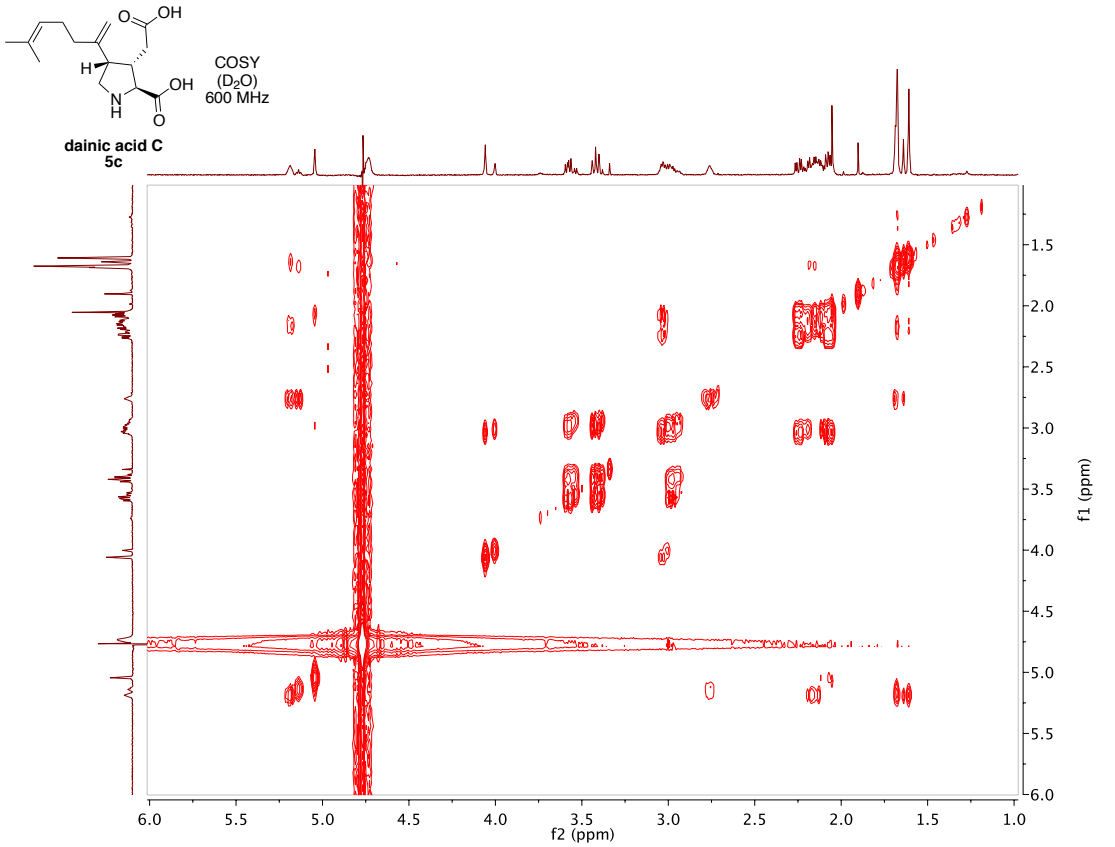
NOESY
(D_2O)
600 MHz

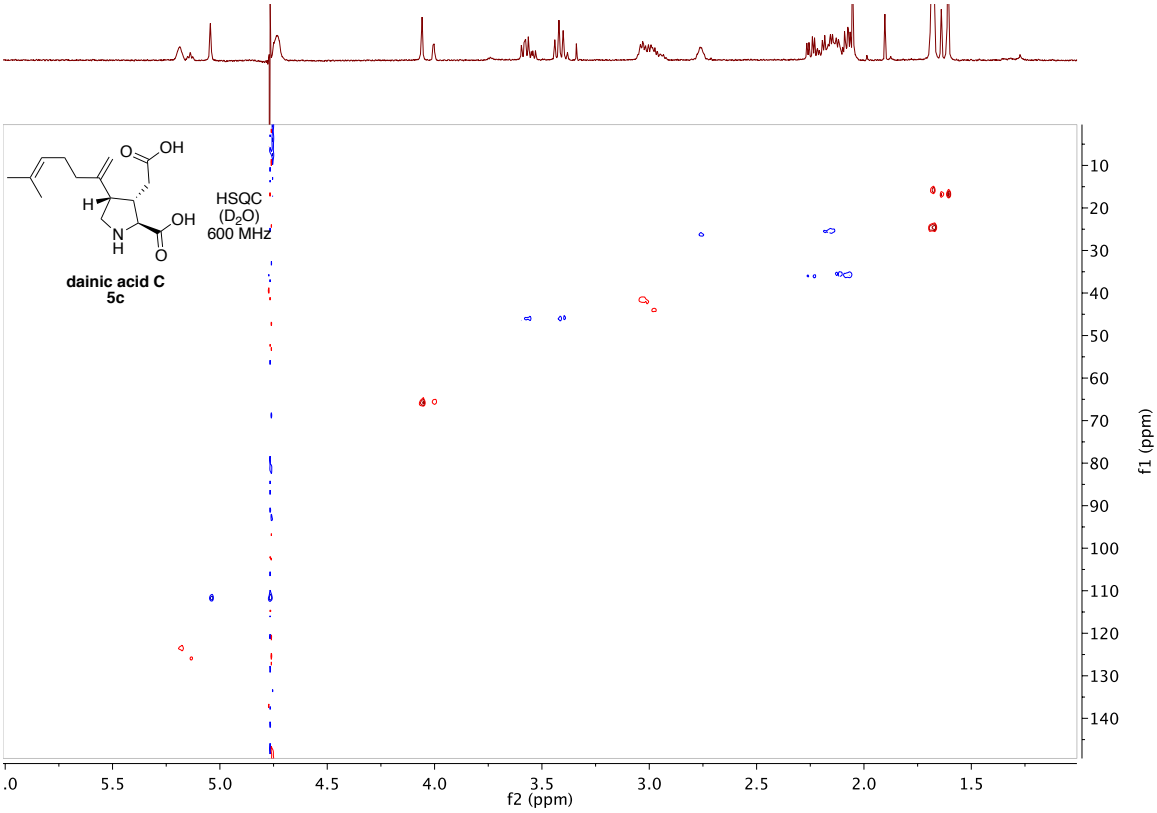
dainic acid B
5b

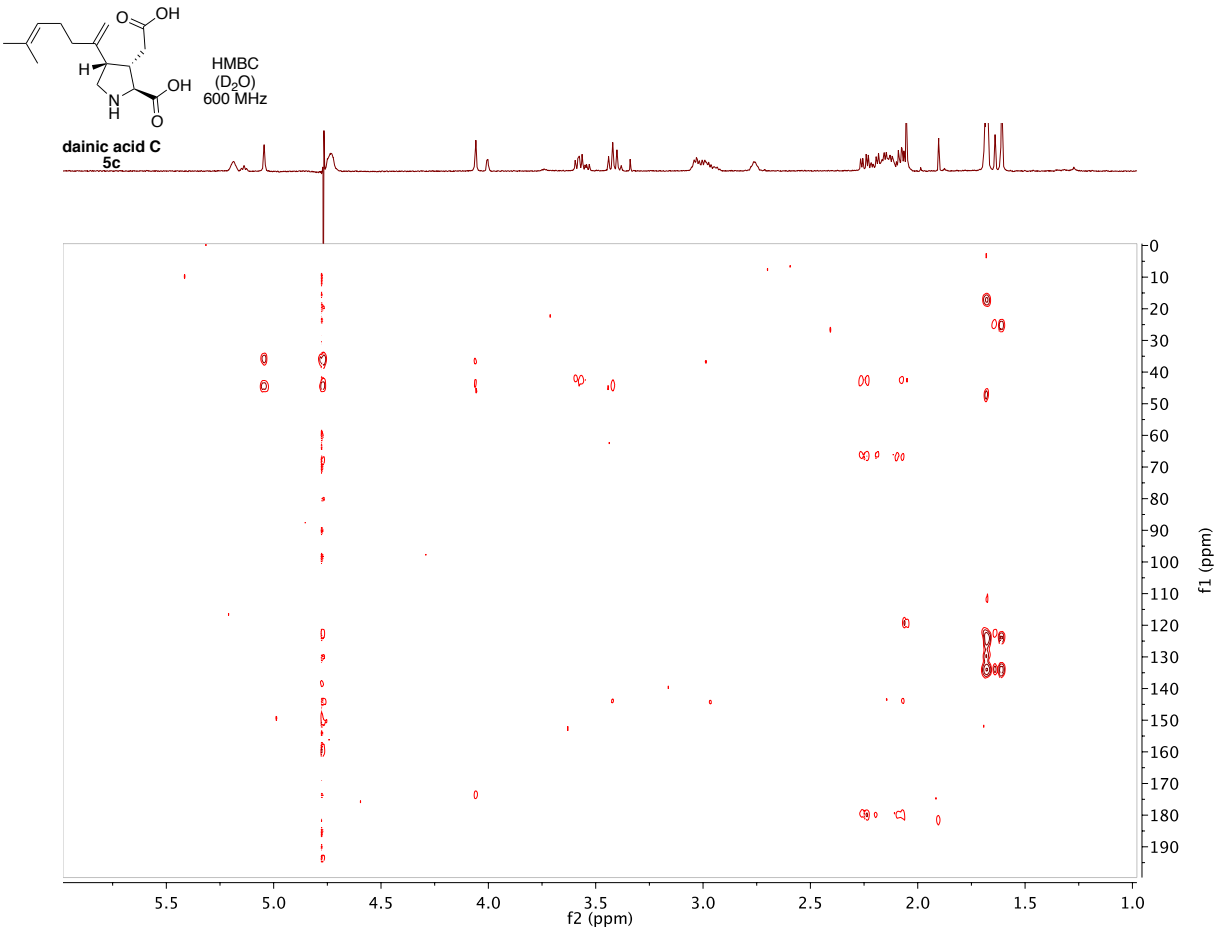
dainic acid C (**5c**):

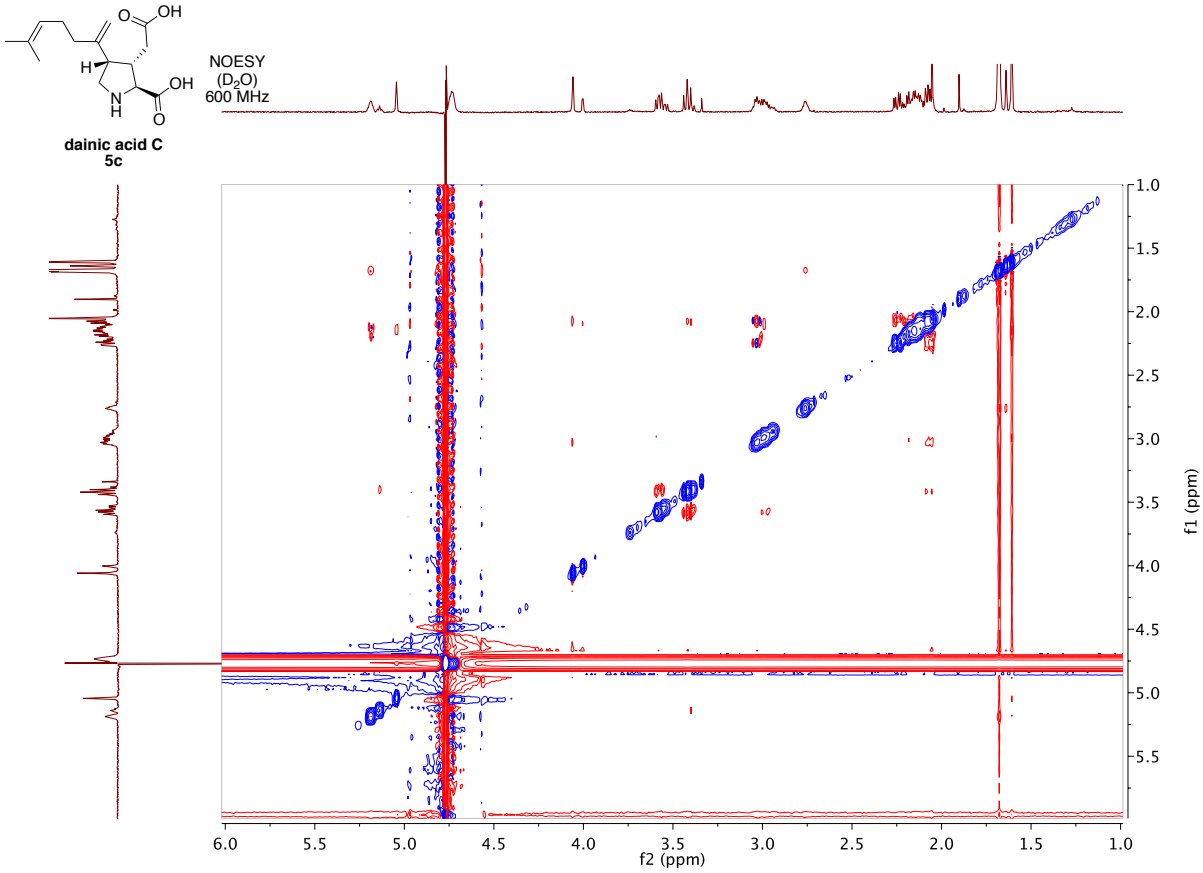




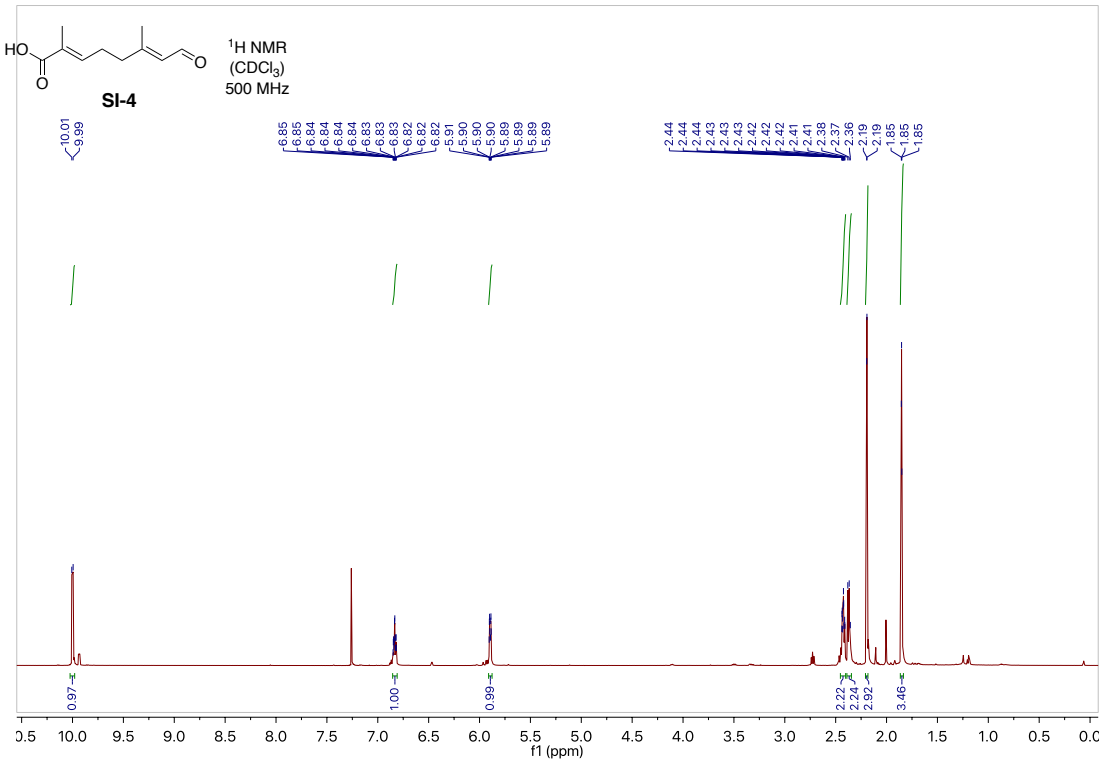

dainic acid C
5c
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

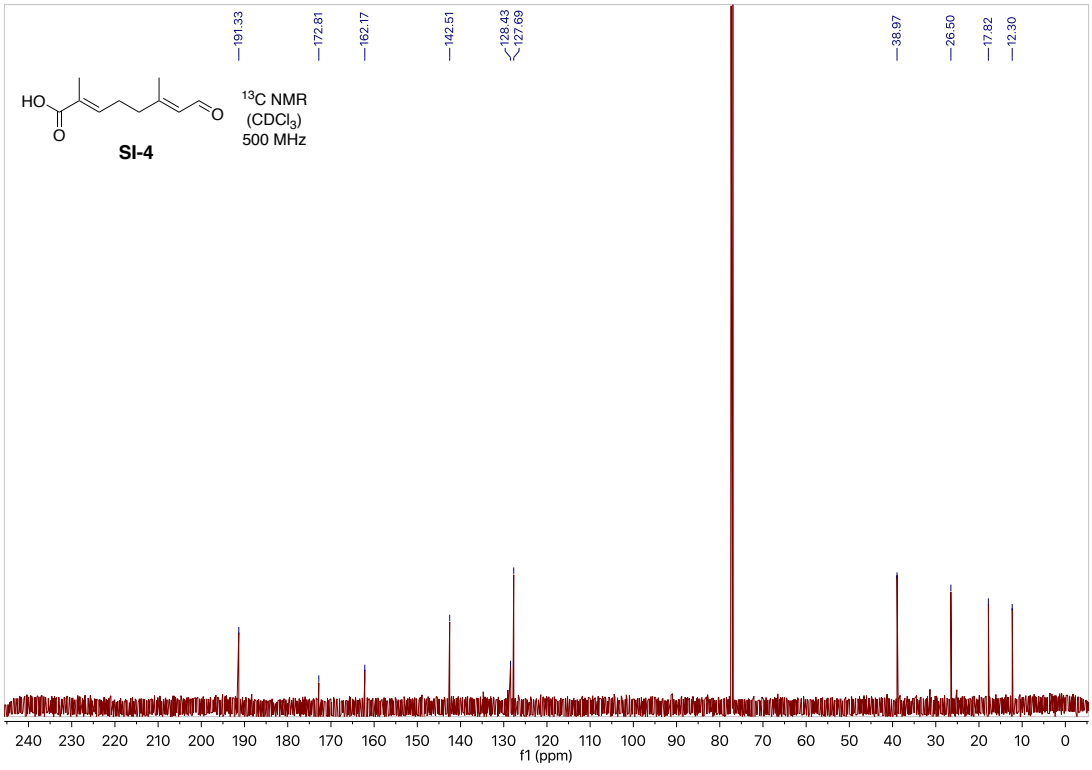

dainic acid C
5c
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)

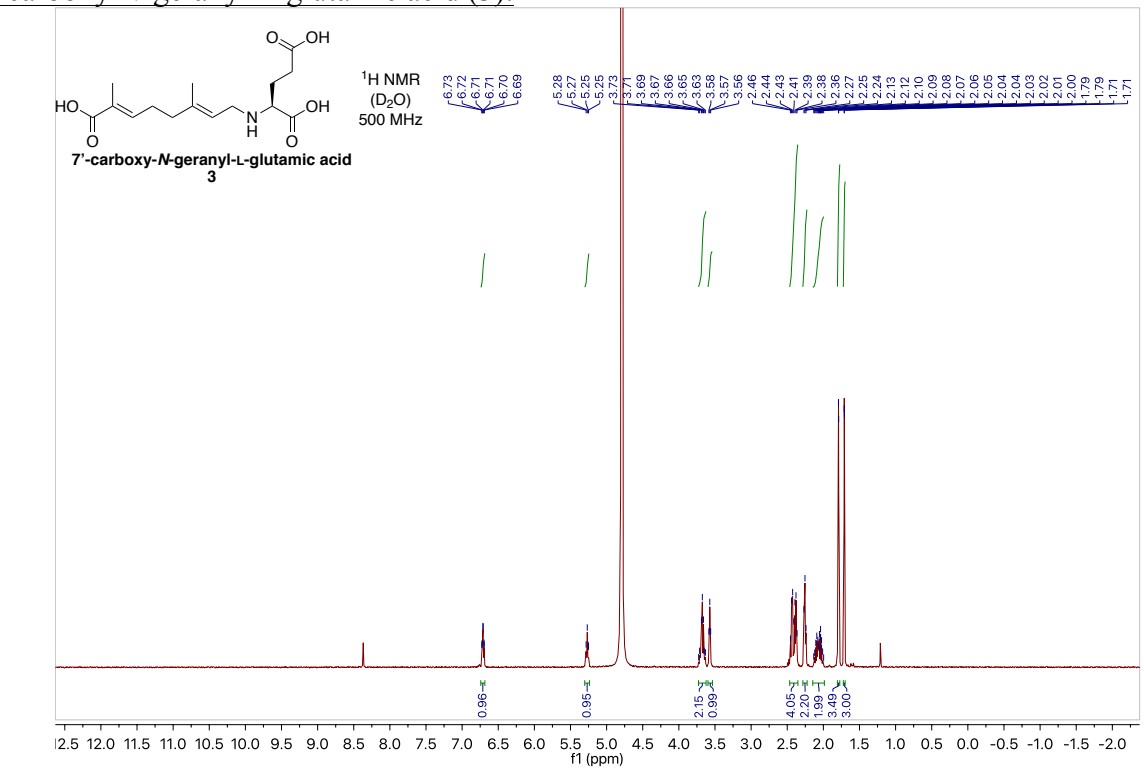

NOESY
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)

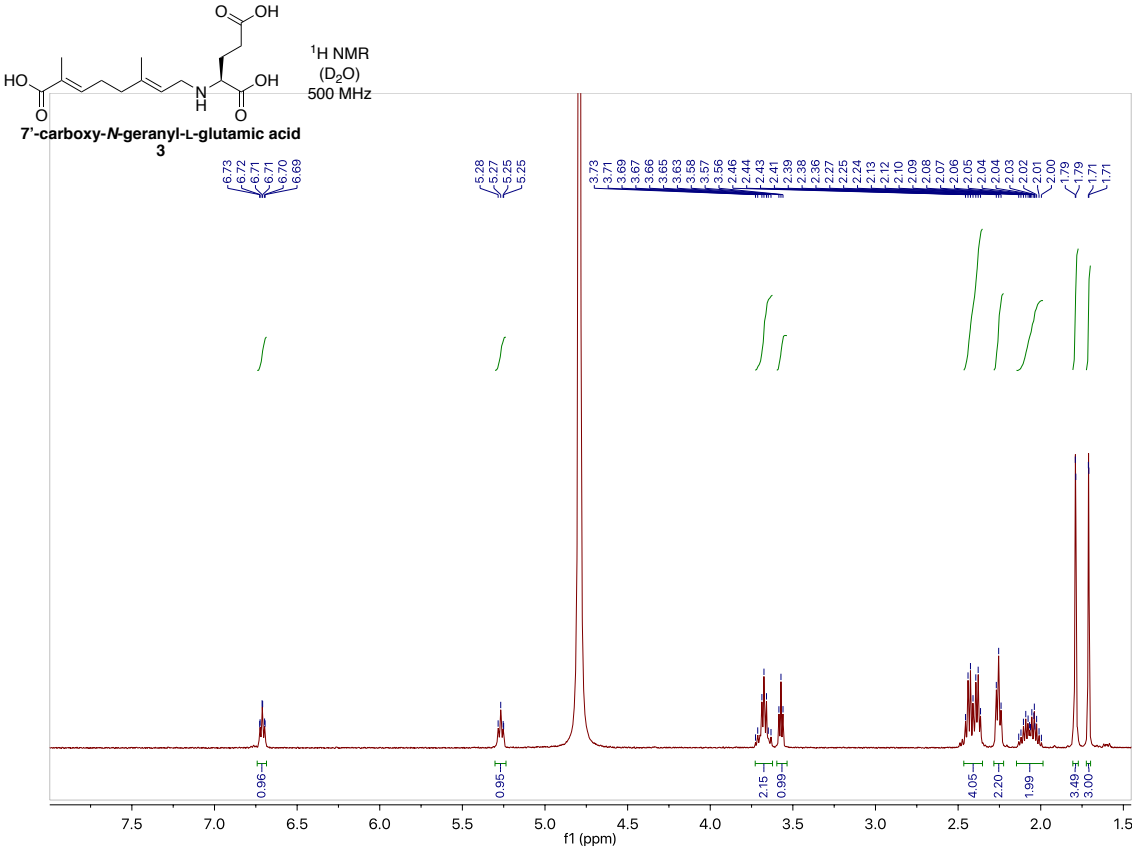
(2*E*,6*E*)-2,6-dimethyl-8-oxoocta-2,6-dienoic acid (**SI-4**):

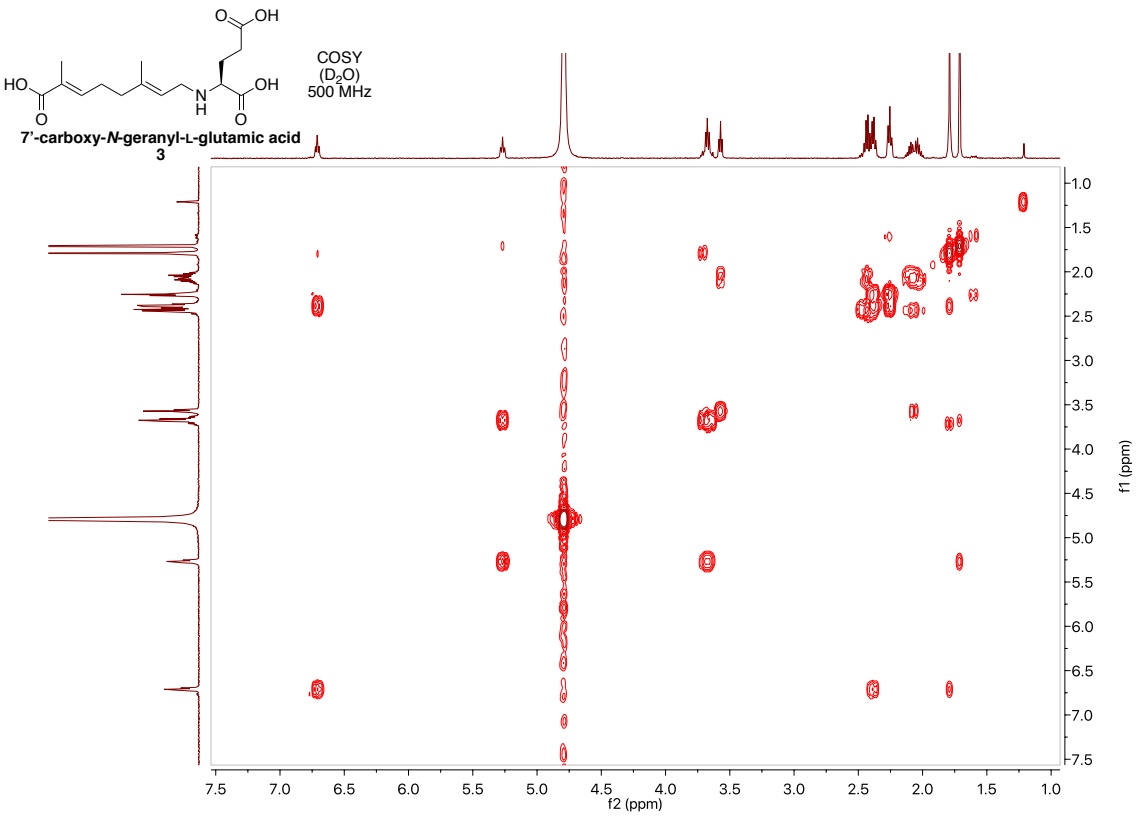


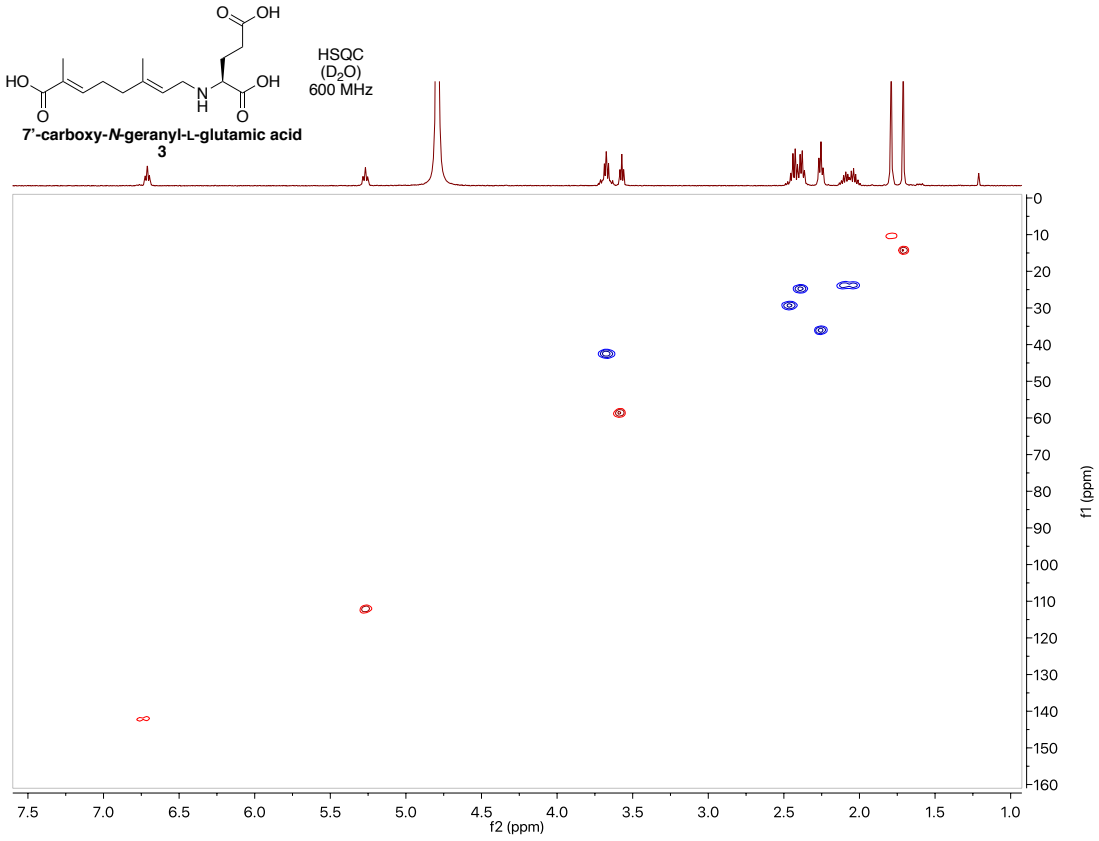


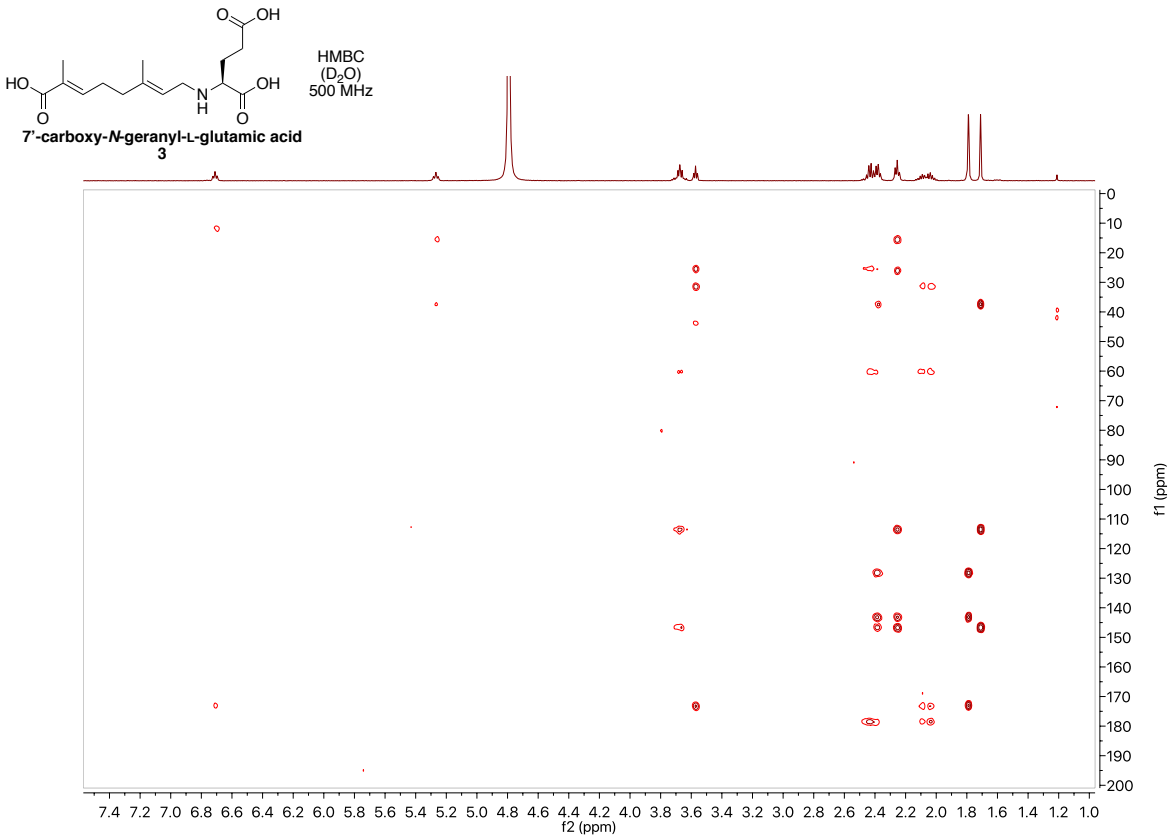
7’-carboxy-*N*-geranyl-L-glutamic acid (**3**):

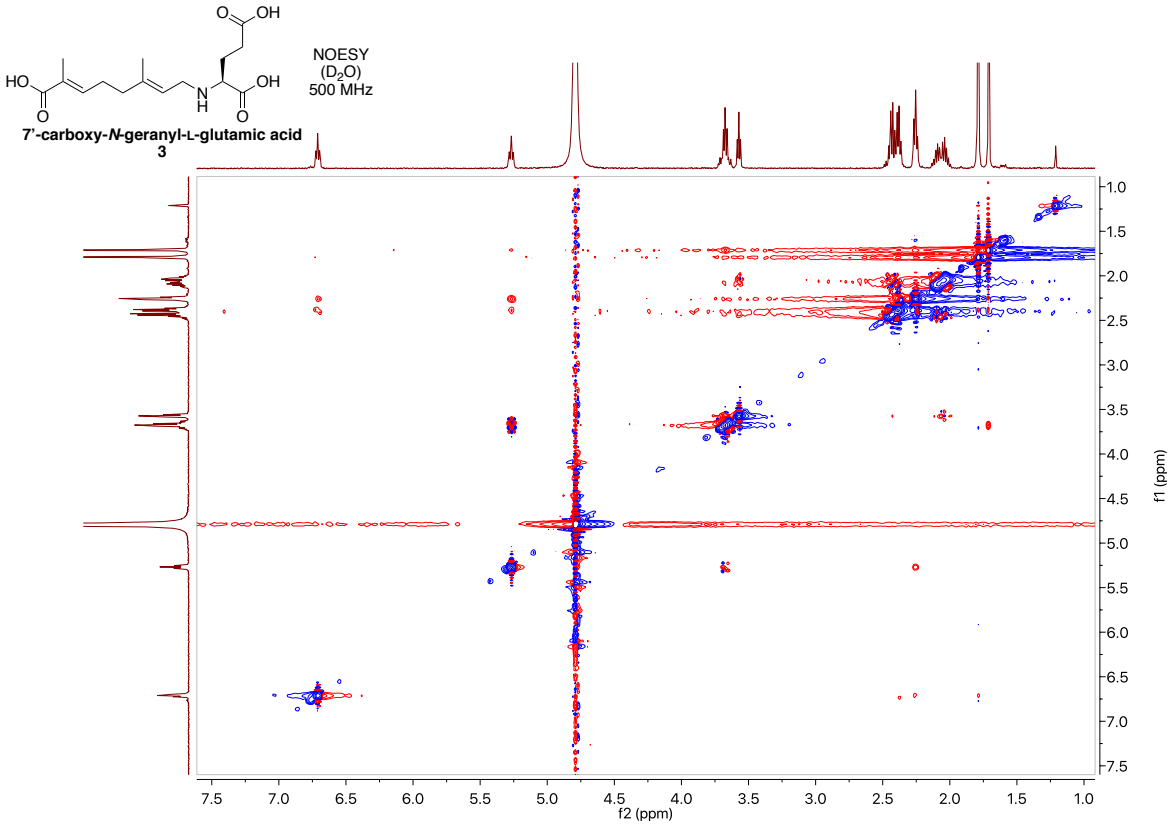


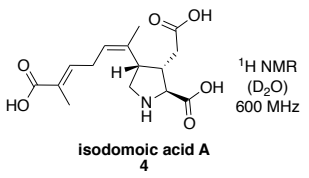



7'-carboxy-N-geranyl-L-glutamic acid
3
COSY
(D2O)
500 MHz
f2 (ppm)
f1 (ppm)


7'-carboxy-N-geranyl-L-glutamic acid
3
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

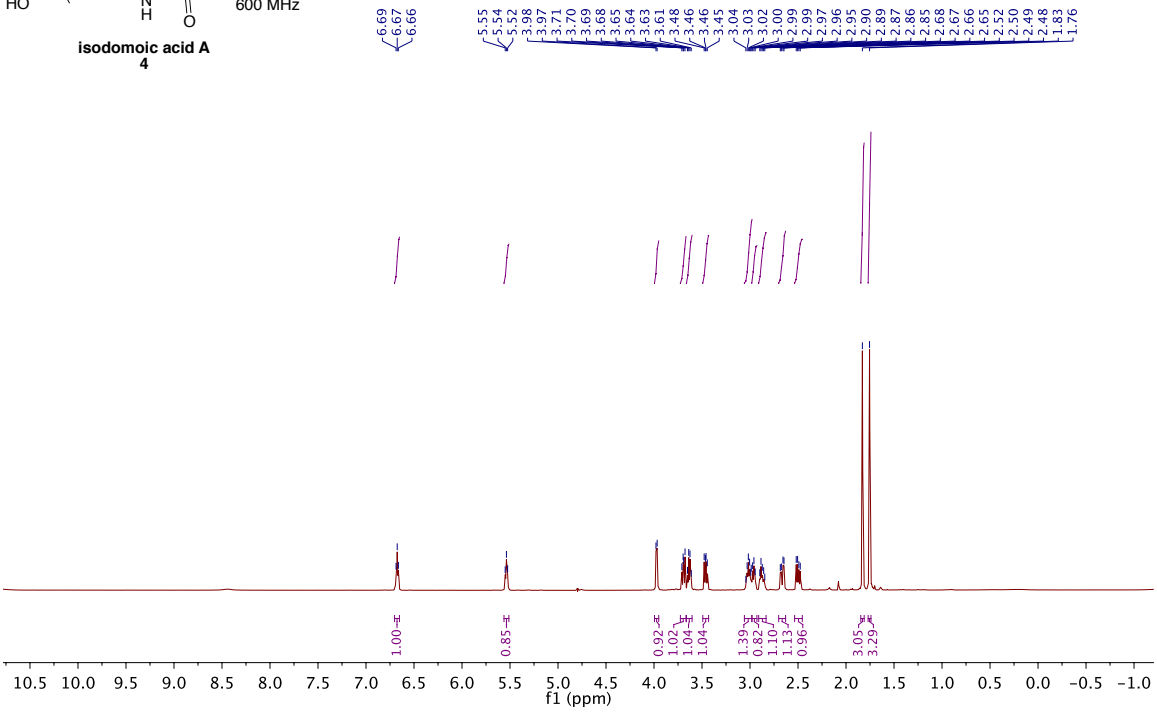

HMBC
(D2O)
500 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)

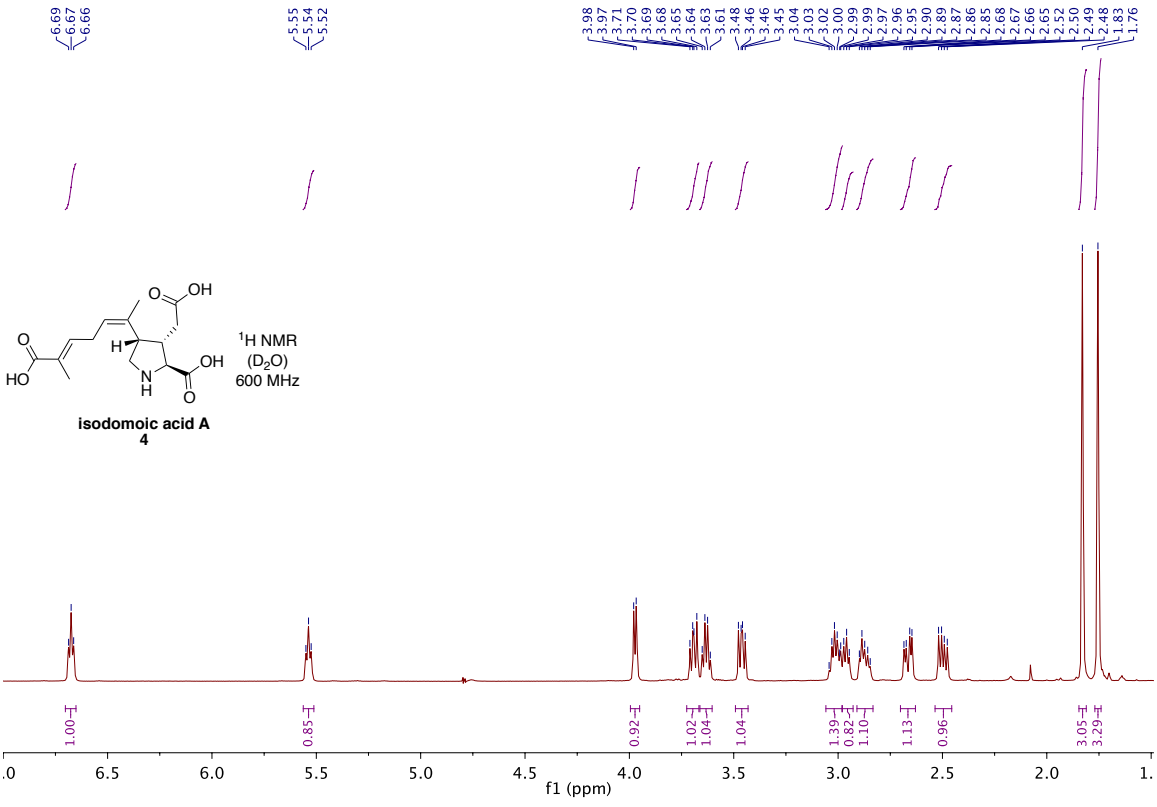

NOESY
(D2O)
500 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)



^{1}H NMR
(D_2O)
600 MHz

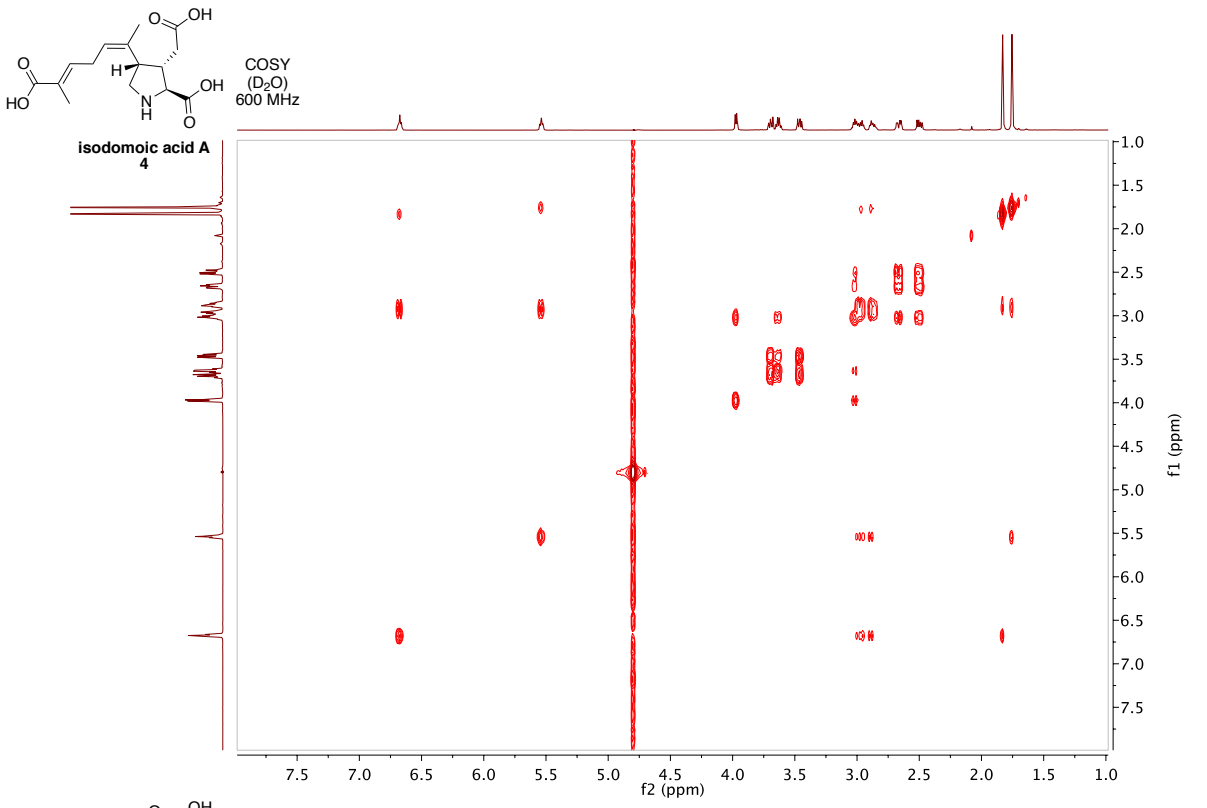
isodomoic acid A
4

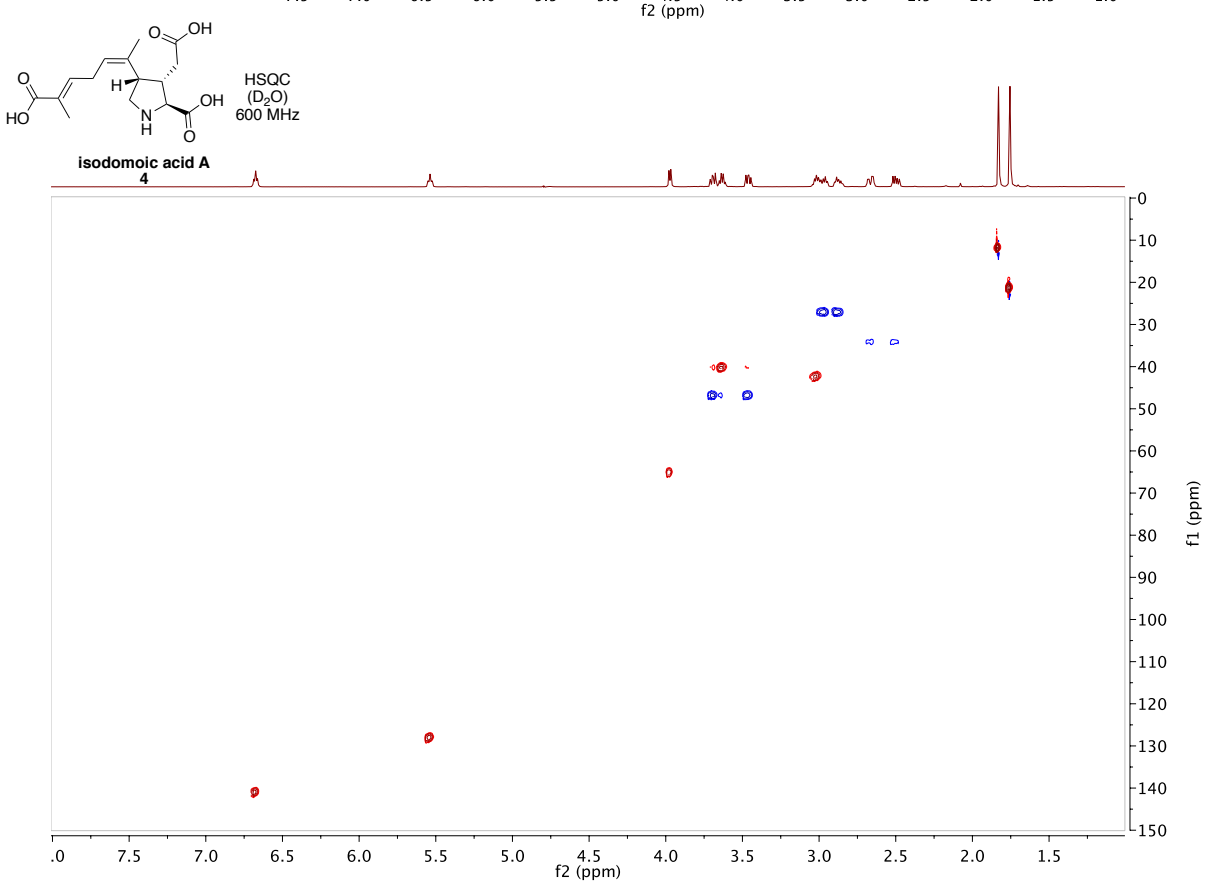


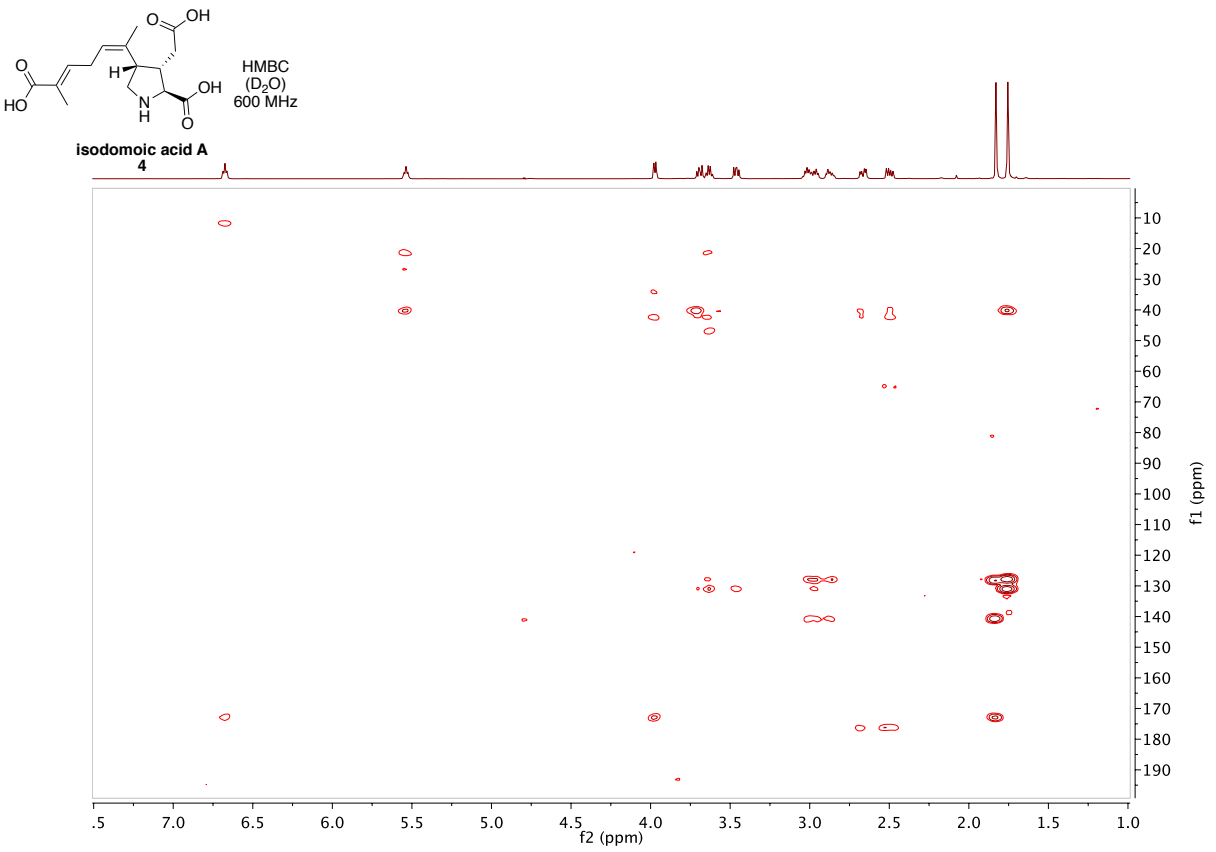


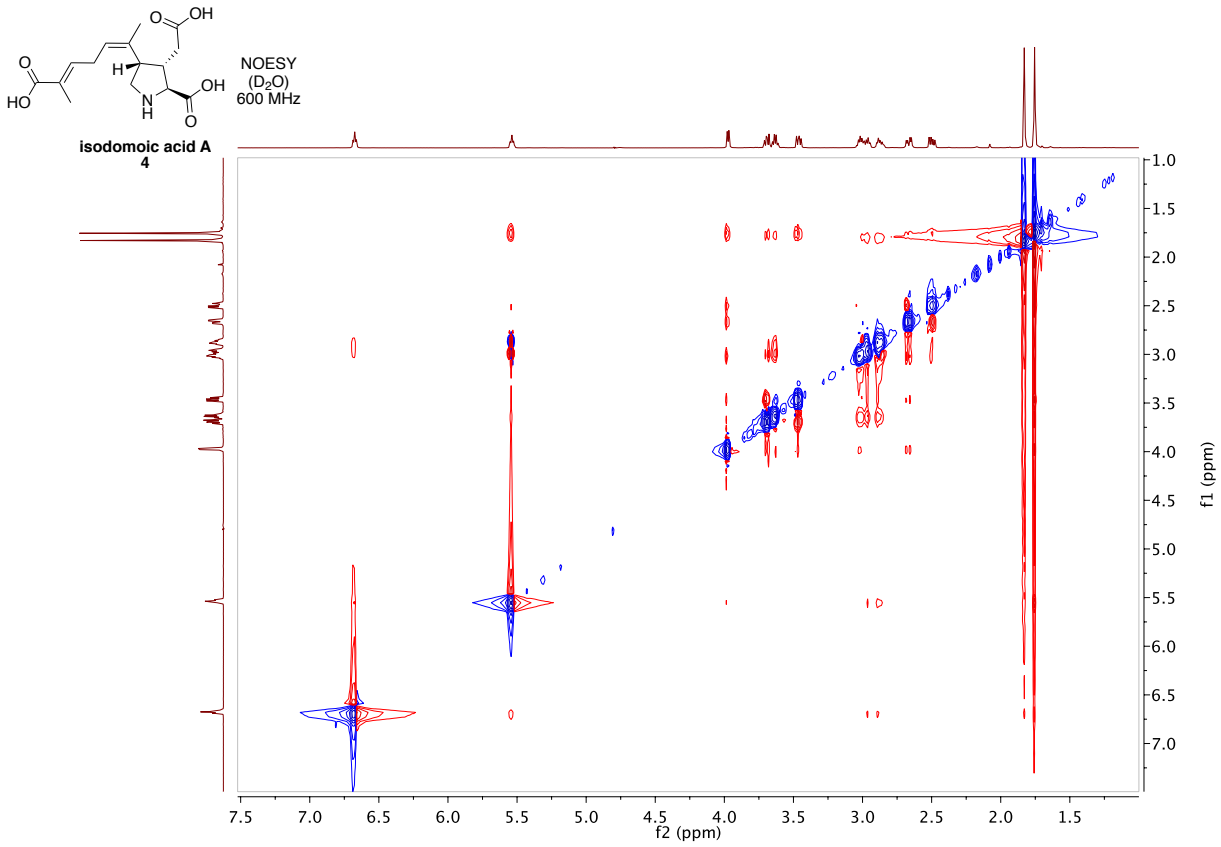
^{1}H NMR
(D_2O)
600 MHz

isodomoic acid A
4

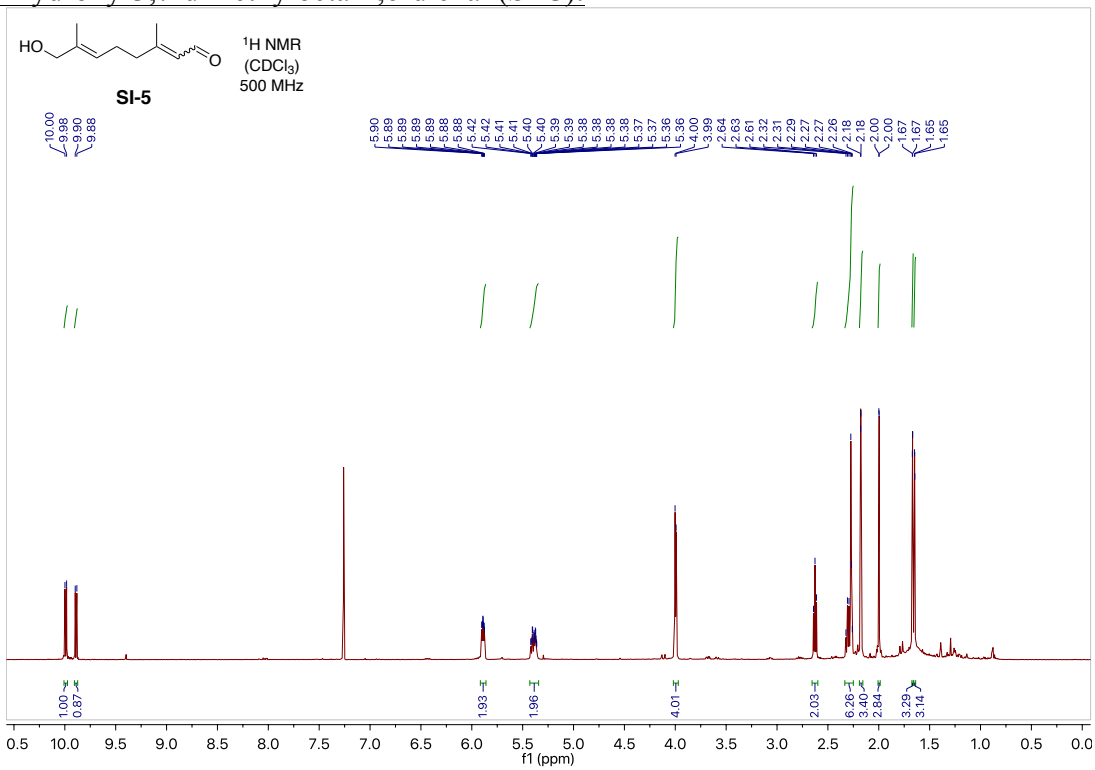

COSY
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

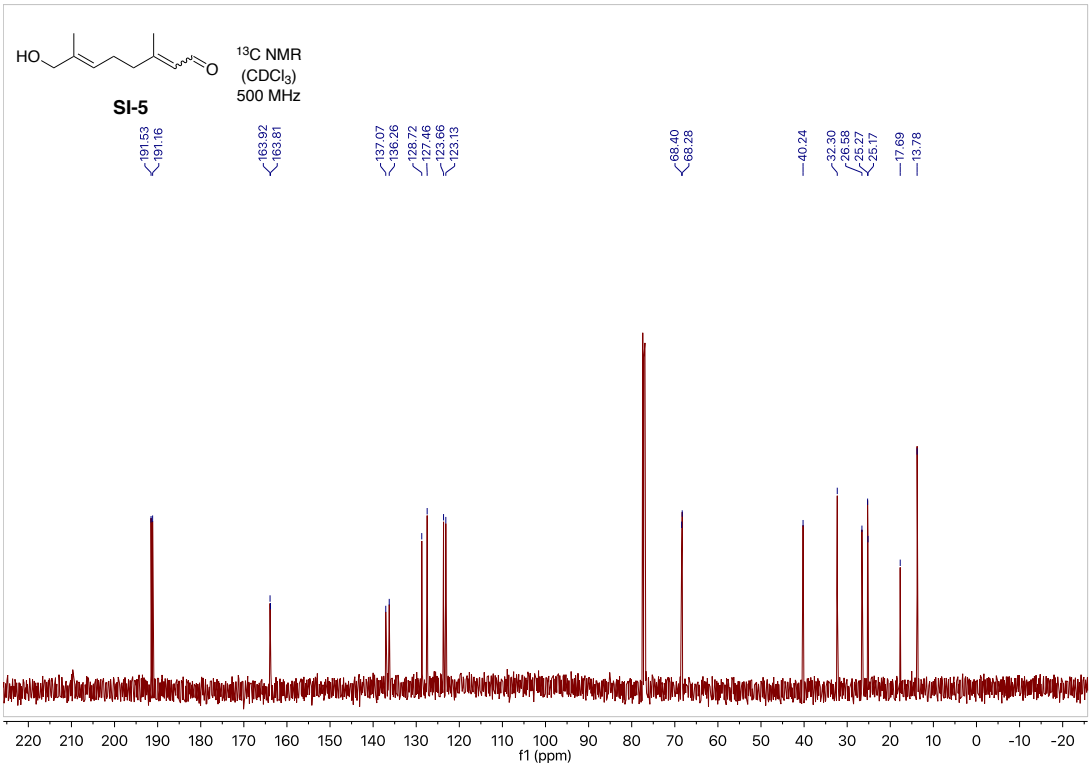

HSQC
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

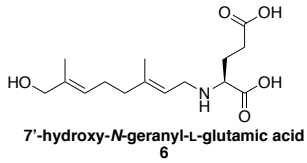

NOESY
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

(6*E*)-8-hydroxy-3,7-dimethylocta-2,6-dienal (**SI-5**):



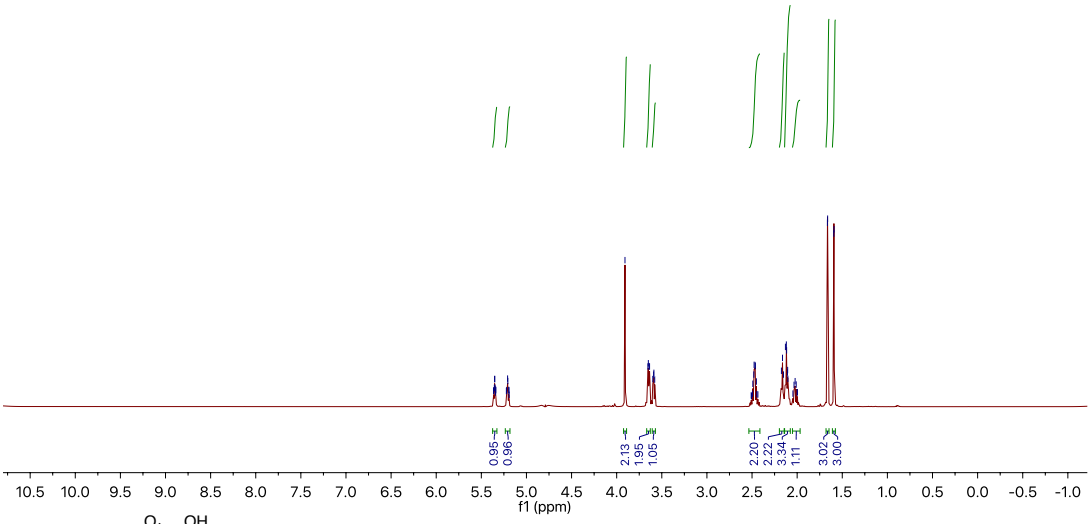


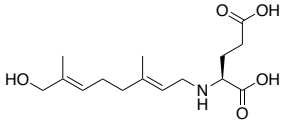
7’-hydroxy-*N*-geranyl-L-glutamic acid (**6**):



7’-hydroxy-*N*-geranyl-L-glutamic acid
6

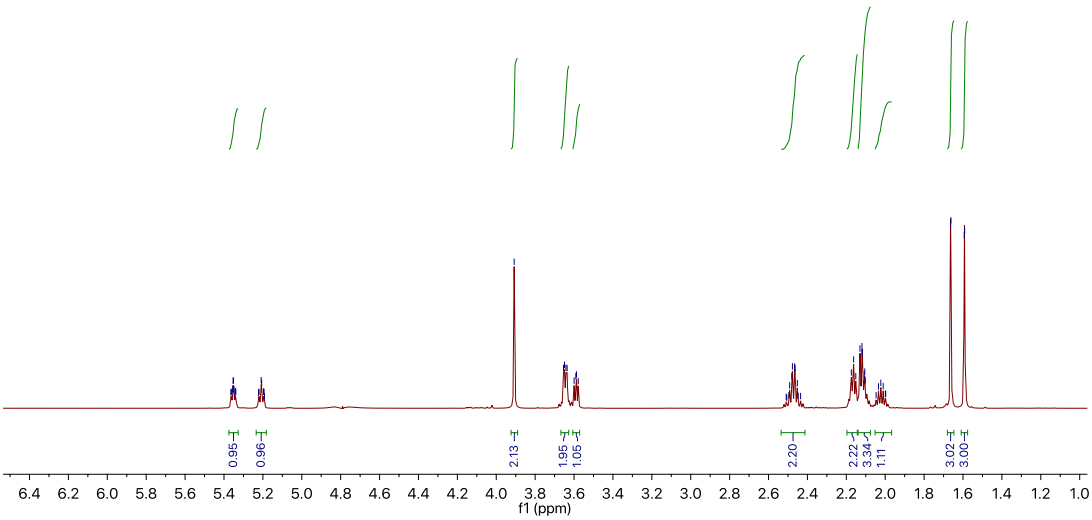
^{1}H NMR
(D_2O)
600 MHz

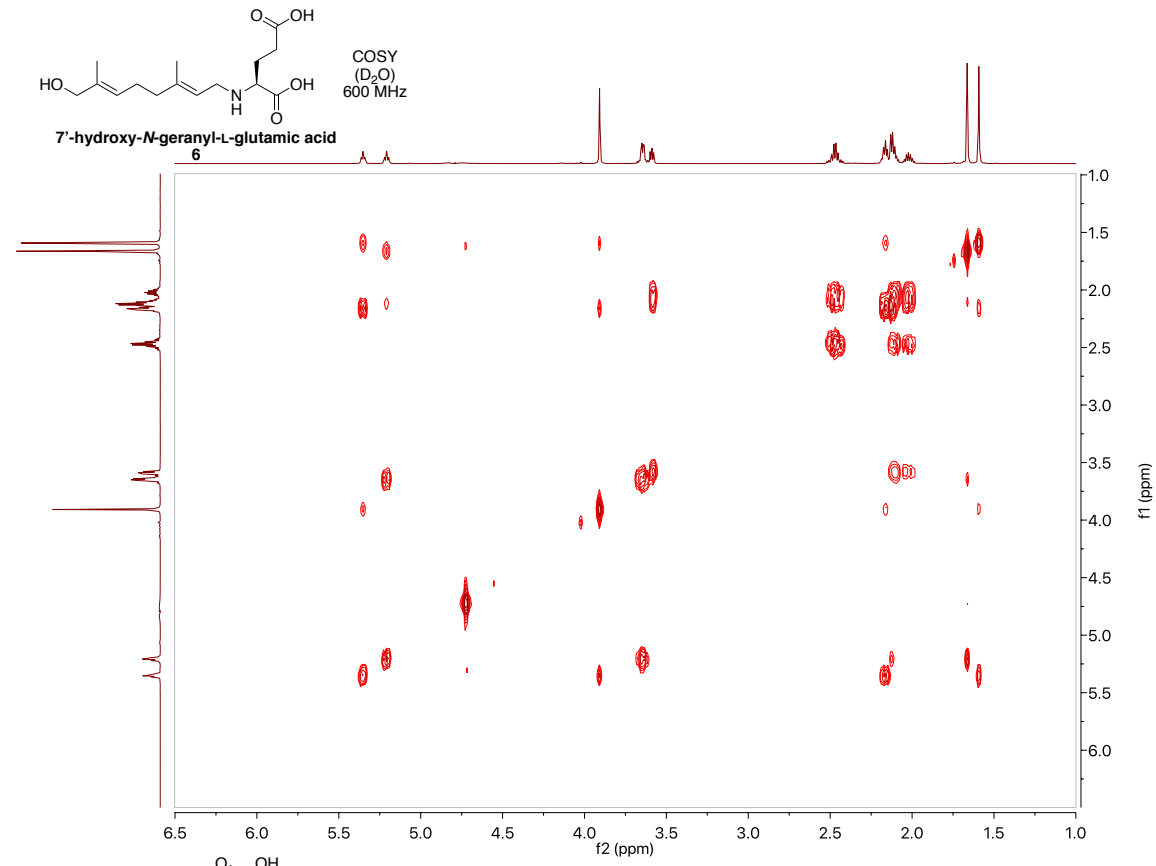


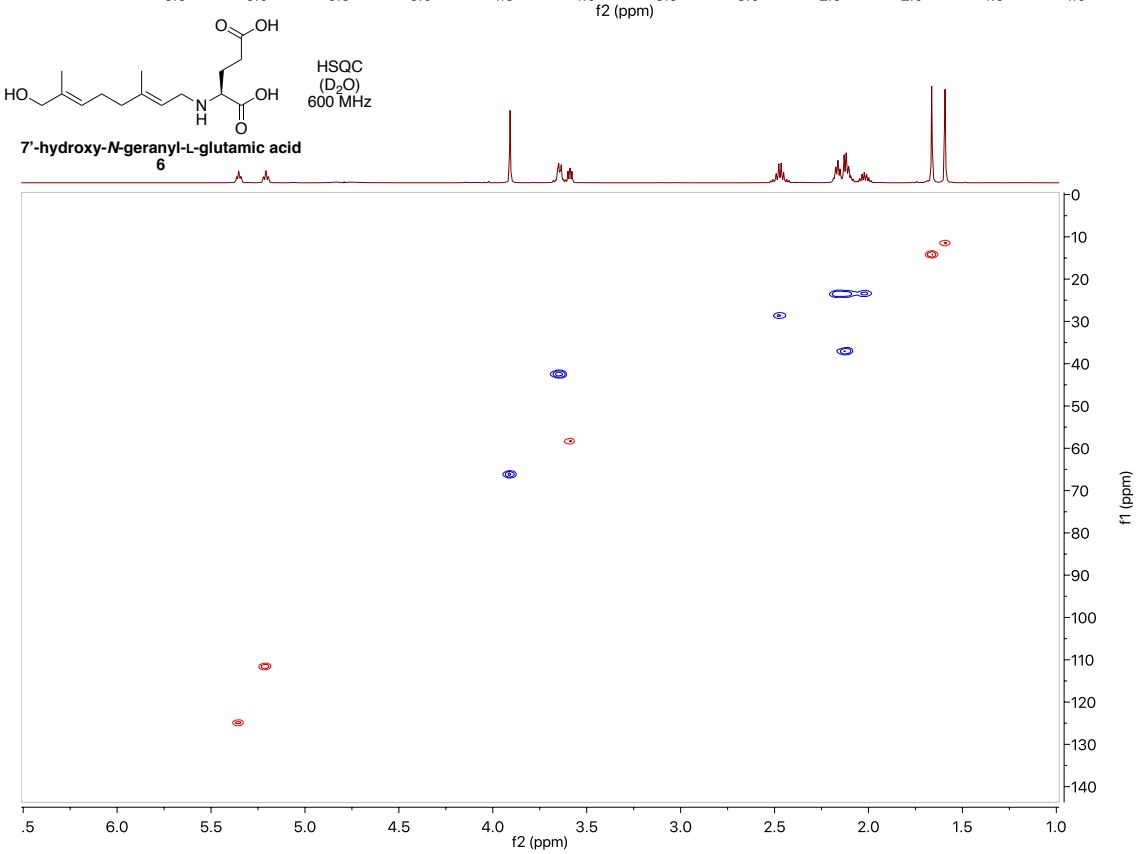


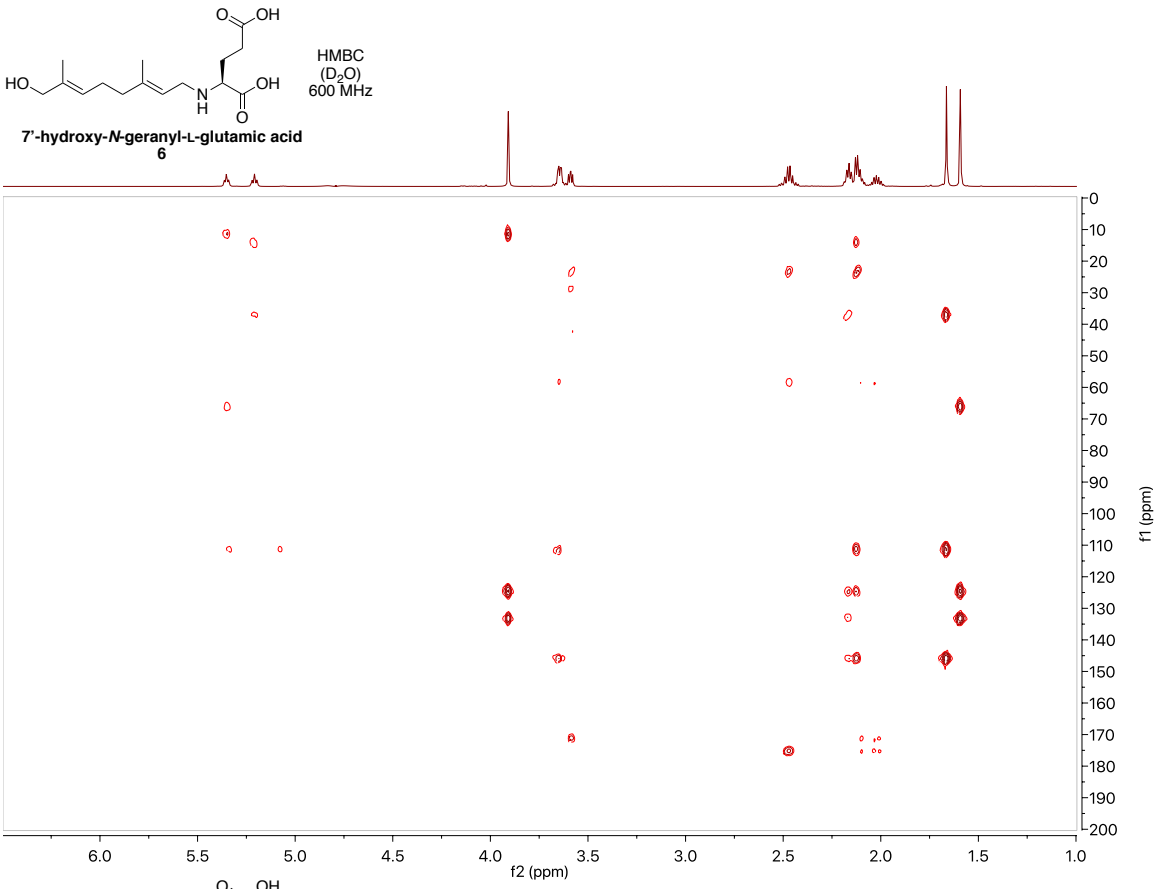
7’-hydroxy-*N*-geranyl-L-glutamic acid
6

^{1}H NMR
(D_2O)
600 MHz




COSY
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)


HSQC
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)

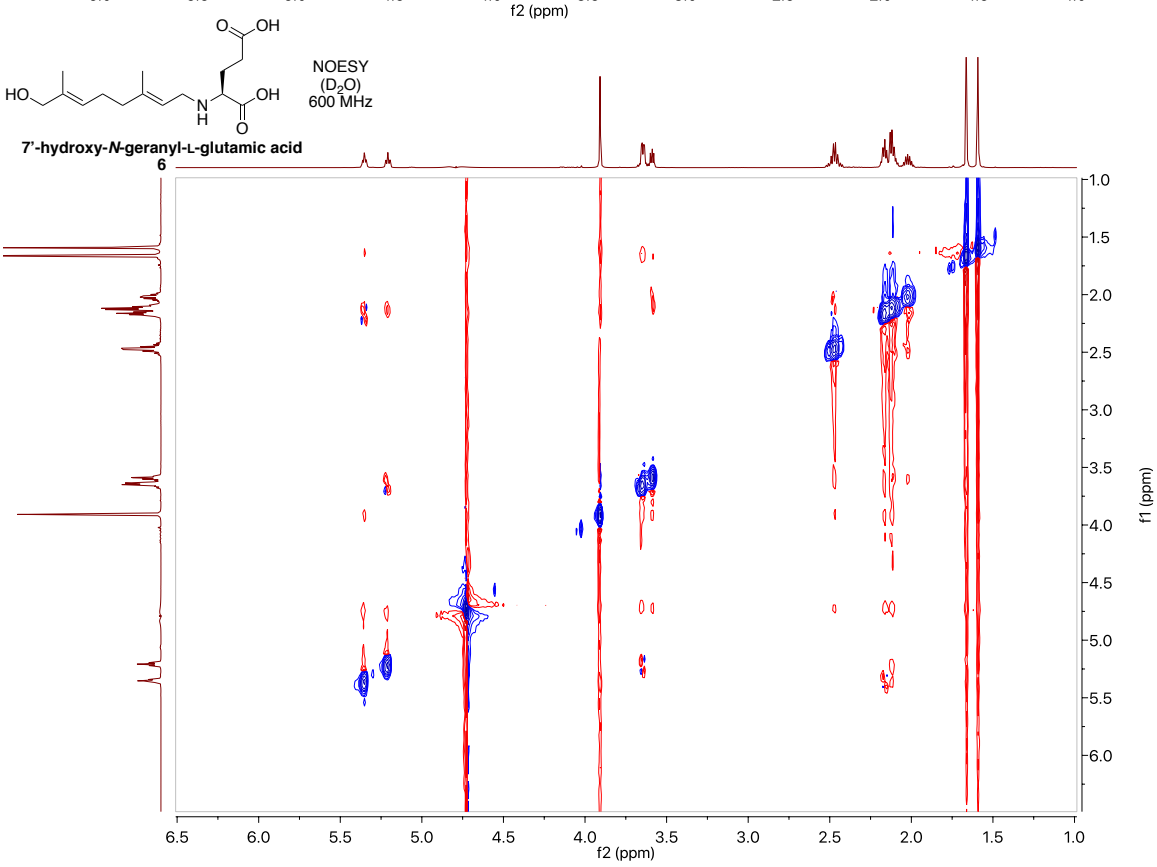

NOESY
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)

Table S1. (separate file)

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for all 19704 *P. multiseries* transcripts under the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

Table S2. (separate file)

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for *P. multiserieis* transcripts upregulated under phosphate starvation. Expression data (RPKM, TMM) is also shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

Table S3. (separate file)

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for *P. multiserieis* transcripts upregulated under phosphate starvation and under increasing pCO_2 with constant low phosphate. Expression data (RPKM, TMM) is also shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

Table S4. (separate file)

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for all CYP450 annotated transcripts. Expression data (RPKM, TMM) is shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

References

1. M. L. Wells *et al.*, Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae*. **49**, 68–93 (2015).

2. S. M. McKibben *et al.*, Climatic regulation of the neurotoxin domoic acid. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 239–244 (2017).

3. C. J. Gobler *et al.*, Ocean warming since 1982 has expanded the niche of toxic algal blooms in the North Atlantic and North Pacific oceans. *Proc. Natl. Acad. Sci.* **114**, 4975–4980 (2017).

4. R. M. McCabe *et al.*, An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10366-10376 (2016).

5. G. R. Stewart, C. F. Zorumski, M. T. Price, J. W. Olney, Domoic acid: A dementia-inducing excitotoxic food poison with kainic acid receptor specificity. *Exp. Neurol.* **110**, 127–138 (1990).

6. J. A. Larm, P. M. Beart, N. S. Cheung, Neurotoxin domoic acid produces cytotoxicity via kainate- and AMPA-sensitive receptors in cultured cortical neurones. *Neurochem. Int.* **31**, 677–682 (1997).

7. J. A. Funk *et al.*, Characterization of renal toxicity in mice administered the marine biotoxin domoic acid. *J. Am. Soc. Nephrol.* **25**, 1187–1197 (2014).

8. K. A. Lefebvre *et al.*, Chronic low-level exposure to the common seafood toxin domoic acid causes cognitive deficits in mice. *Harmful Algae*. **64**, 20–29 (2017).

9. J. S. Ramsdell, T. S. Zabka, *In utero* domoic acid toxicity: a fetal basis to adult disease in the California sea lion (*Zalophus californianus*). *Mar. Drugs*. **6**, 262–90 (2008).

10. L. M. Grattan *et al.*, Repeated dietary exposure to low levels of domoic acid and problems with everyday memory: research to public health outreach. *Toxins (Basel).* **10**, 103 (2018).

11. P. F. Cook *et al.*, Algal toxin impairs sea lion memory and hippocampal connectivity, with implications for strandings. *Science*. **350**, 1545–7 (2015).

12. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia*. **51**, 168–216 (2012).

13. M. P. Sison-Mangus, S. Jiang, K. N. Tran, R. M. Kudela, Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J.* **8**, 63–76 (2014).

14. U. P. Ramsey, D. J. Douglas, J. A. Walter, J. L. Wright, Biosynthesis of domoic acid by the diatom *Pseudo-nitzschia multiseries*. *Nat. Toxins*. **6**, 137–46 (1998).

15. T. J. Savage, G. J. Smith, A. T. Clark, P. N. Saucedo, Condensation of the isoprenoid and amino precursors in the biosynthesis of domoic acid. *Toxicon*. **59**, 25–33 (2012).

16. X. Zhang, S. Li, Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **34**, 1061–1089 (2017).

17. J. Sun *et al.*, Effects of changing pCO_2 and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseries*. *Limnol. Oceanogr.* **56**, 829–840 (2011).

18. M. H. Medema *et al.*, Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–31 (2015).

19. E. V. Armbrust *et al.*, The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. **306**, 79–86 (2004).

20. C. Bowler *et al.*, The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*. **456**, 239–244 (2008).

21. A. E. Allen *et al.*, Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci.* **105**, 10438–10443 (2008).

22. G. Sapriel *et al.*, Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *PLoS One* **4**, e7458 (2009).

23. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–58 (2015).

24. Y. Maeno *et al.*, Six domoic acid related compounds from the red alga, *Chondria armata*, and domoic acid biosynthesis by the diatom, *Pseudo-nitzschia multiseries*. *Sci. Rep.* **8**, 356 (2018).

25. J. Clayden, B. Read, K. R. Hebditch, Chemistry of domoic acid, isodomoic acids, and their analogues. *Tetrahedron*. **61**, 5713–5724 (2005).

26. I. Nitta, H. Watase, Y. Tomiie, Structure of kainic acid and its isomer, allokainic acid. *Nature*. **181**, 761–762 (1958).

27. K. Konno, H. Shirahama, T. Matsumoto, Isolation and structure of acromelic acid A and B. New kainoids of *Clitocybe acromelalga*. *Tetrahedron Lett.* **24**, 939–942 (1983).

28. G. W. Saunders, C. Jackson, E. D. Salomaki, Phylogenetic analyses of transcriptome data

resolve familial assignments for genera of the red-algal Acrochaetiales-Palmariales Complex (Nemaliophycidae). *Mol. Phylogenet. Evol.* **119**, 151–159 (2018).

29. N. Matasci *et al.*, Data access for the 1,000 Plants (1KP) project. *Gigascience*. **3**, 17 (2014).

30. S. Basu *et al.*, Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* **215**, 140–156 (2017).

31. P. J. Keeling *et al.*, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

32. M. Rho, H. Tang, Y. Ye, FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191–e191 (2010).

33. E. M. Bertrand *et al.*, Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.* **112**, 9938–9943 (2015).

34. R. Schmieder, Y. W. Lim, R. Edwards, Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*. **28**, 433–435 (2012).

35. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013) (available at http://arxiv.org/abs/1303.3997).

36. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).

37. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).

38. R. A. Anderson, Ed., *Algal Culturing Techniques* (Elsevier Inc., 2005).

39. R. E. Diner, V. A. Bielinski, C. Dupont, A. E. Allen, P. D. Weyman, Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front. Bioeng. Biotechnol.* **4**, 65 (2016).

40. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).

41. J. D. Bendtsen, H. Nielsen, G. Von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).

42. A. Gruber, G. Rocap, P. G. Kroth, E. V. Armbrust, T. Mock, Plastid proteome prediction

for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81**, 519–528 (2015).

43. P. Savitsky *et al.*, High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* **172**, 3–13 (2010).

44. C. A. Miller, M. A. Martinat, L. E. Hyman, Assessment of aryl hydrocarbon receptor complex interactions using pBEVY plasmids: expression vectors with bi-directional promoters for use in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **26**, 3577–83 (1998).

45. E. W. Jones, Tackling the protease problem in *Saccharomyces cerevisiae*. *Methods Enzymol.* **194**, 428–53 (1991).

46. D. C. Amberg, D. J. Burke, J. N. Strathern, *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

47. D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol.* **272**, 51–64 (1996).

48. V. V Kushnirov, Rapid and reliable protein extraction from yeast. *Yeast*. **16**, 857–860 (2000).

49. M. Kahru, R. M. Kudela, M. Manzano-Sarabia, B. Greg Mitchell, Trends in the surface chlorophyll of the California Current: merging data from multiple ocean color satellites. *Deep Sea Res., Part II.* **77**–**80**, 89–98 (2012).

50. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).

51. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

52. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

53. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. **25**, 2286–2288 (2009).

54. N. P. Dunham *et al.*, Two distinct mechanisms for C–C desaturation by iron(II)- and 2-(oxo)glutarate-dependent oxygenases: importance of α-heteroatom assistance. *J. Am. Chem. Soc.* **140**, 7116–7126 (2018).

55. M.-C. Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang, Oxidative cyclization in natural product biosynthesis. *Chem. Rev.* **117**, 5226–5333 (2017).

56. R. D. Finn *et al.*, InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).

57. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).

58. A. B. Woodside, Z. Huang, C. D. Poulter, Trisammonium geranyl diphosphate. *Org. Synth.* **66**, 211 (1988).

59. G. Verardo, P. Geatti, E. Pol, A. G. Giumanini, Sodium borohydride: A versatile reagent in the reductive N -monoalkylation of α-amino acids and α-amino methyl esters. *Can. J. Chem.* **80**, 779–788 (2002).

60. G. R. Labadie, R. Viswanathan, C. D. Poulter, Farnesyl diphosphate analogues with ω-bioorthogonal azide and alkyne functional groups for protein farnesyl transferase-catalyzed ligation reactions. *J. Org. Chem.* **72**, 9291–9297 (2007).

61. Y.-J. Zhao, T.-P. Loh, Practical synthesis of 1,5-dimethyl substituted conjugated polyenes from geranyl acetate. *Tetrahedron*. **64**, 4972–4978 (2008).

62. H. L. Jackson, G. T. Nadolski, C. Braun, S. F. Lockwood, Efficient total synthesis of lycophyll (ψ,ψ-Carotene-16,16‘-diol). *Org. Process Res. Dev.* **9**, 830–836 (2005).

63. S. Xie, S. Uesato, T. Fujita, H. Inouye, Biosynthesis of iridoid glucosides in *Patrinia gibbosa*. *J. Nat. Prod.* **52**, 701–705 (1989).

64. J. A. Walter, M. Falk, J. L. C. Wright, Chemistry of the shellfish toxin domoic acid: characterization of related compounds. *Can. J. Chem.* **72**, 430–436 (1994).