Supplementary Information for ‘Single-bacterial genomics validates the rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts’

Supplementary Materials and Methods

Sponge collection, preparation of ‘Entotheonella’-enriched fraction, and single-cell analysis

Samples of Japanese *T. swinhoei* WA were collected in 2011 by SCUBA off the coast of Hachijo-jima, Japan, as described previously (1). Three *T. swinhoei* samples were collected at Eilat, Red Sea (29°29'57.63"N / 34°54'54.61"E) in 2014 by SCUBA diving at 20 m depth (with a permit from the Nature and Parks Authority), and processed on site at the Interuniversity Institute for Marine Sciences. All further work was performed in a laminar flow hood under sterile conditions.

Japanese *T. swinhoei* WA sponges were subjected to cell separation by differential centrifugation as described previously (1). The resulting 'Entotheonella' fraction was resuspended in 200 ml Ca- and Mg-free artificial sea water (2) and stored at 4 °C. Single ‘Entotheonella’ filaments were sorted by FACS into 96 well plates, and two positive wells G6 and H6 were identified by PCR after multiple displacement amplification (MDA) of the DNA and subsequently sequenced as described previously (1).

Like the Japanese samples, Israeli *T. swinhoei* sponges were subjected to cell separation by differential centrifugation as described previously (1). The resulting 'Entotheonella' fraction was resuspended in 200 ml Ca- and Mg-free artificial sea water (2) and stored at 4 °C. DNA was prepared using the protocol described by Tauch *et al*. (3). For sequencing, a TruSeq PCR-free library (Illumina Inc., Netherlands) was prepared according to the manufacturer's instructions. Sequencing was performed on an Illumina MiSeq platform, using the 2x 300 bp sequencing kit.

For the Japanese TSWA sequences G6 and H6, data was processed as described before (4). In brief, FastQC (5) was applied for quality control of raw sequence reads and quality trimming was

performed using trimmomatic 0.35 (6) and default settings (Remove leading and trailing low quality or N bases (below quality 3); scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15; drop reads below the 36 bases long). Upon sequencing and processing of the obtained data, SPAdes (version 3.9.0.) (7) was used for a *de novo* assembly default settings for single cells and long Illumina reads including kmer 21, 33, 55, 77, 99 and 127. Automatic annotation was performed within the platforms Prokka 1.11 (8) and GenDB 2.0 (9). BUSCO (v3.0.0) with the training set for Bacteria was used to test for completeness of the draft genome assembly (10). BUSCO tests the coverage of Single-Copy Orthologues that are represented as BUSCO gene models used for training of AUGUSTUS (v3.0.3) gene prediction (11).

For the Israeli sample, a total of 194,941,818 reads (3.96 Gbp) were screened against the contigs from the Japanese strains using BlastN, reads (and their mates) with >= 95% identity were retained. This filtering reduced the dataset to a total of 4,085,316 reads (955.2 Mbp) that were assembled using the Newbler (v2.8) *de novo* assembler. The resulting assembly consisted of 937 scaffolds containing 1,528 contigs with a total of 6.57 Mbp. The resulting contigs were automatically annotated with RAST (9).

Sequence analysis

Bioinformatic analysis of natural product genes was carried out as described previously (12). Briefly, manual identification and annotation of natural product biosynthetic genes were conducted with BLAST using validated biosynthetic genes as queries. Automated identification of natural product genes and clusters was performed with Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) 3.0 (13), NaPDoS (14), and manual BLAST analysis of uncertain regions. All manual annotation and routine bioinformatic analysis was performed using Geneious version 8.1.6 created by Biomatters (available from http://www.geneious.com). Scaffold gaps were closed using PCR amplification with Phusion® High-Fidelity or Q5® High-Fidelity DNA polymerase (New England Biolabs) and sequencing (Microsynth). PCR primers used for gap closing were designed from the terminal ends of assembled contigs from 'E. serta' TSWA1 and 'E. serta' TSWB phylotypes from the

Israel collection (Table S8). Average nucleotide identity (ANI) analyses were performed as recently described (15, 16).

Chemical analysis of sponges

3 g of Japanese *T. swinhoei* WA specimen (stored at -80 °C) and approximately 3 g of different specimens of the Israel chemotype WB (stored in ethanol at 4 °C) were individually sliced into small pieces. These were extracted with a 1:1 mixture of dichloromethane and methanol overnight at 4 °C (17). The extracts were filtered and dried under reduced pressure. Extracts were resuspended in acetonitrile and subjected to data-dependent ultra-high performance liquid chromatography-high resolution heated electrospray-tandem mass spectrometry (UPLC HR HESI MS/MS) analysis using a Dionex Ultimate 3000 UPLC system connected to a Thermo QExactive mass spectrometer. A solvent gradient (A = H_2O + 0.1% formic acid and B = acetonitrile + 0.1% formic acid with B at 5% for 0-2 min, 5-95% for 2-14 min and 95% for 11-17 min at a flow rate of 0.5 mL/min) was used on a Phenomenex Kinetex 2.6 µm C18 100A (150 × 4.6 mm) column at 27 °C. The MS was operated in positive ionization mode at a scan range of 600-2000 *m/z* to account for single and double charged theonellamides. The spray voltage was set to 3.7 kV and the capillary temperature to 320 °C. MS^2 data were acquired in a data-dependent fashion with the parent ion scan at a resolution of 70,000 and the MS^2 scan at a resolution of 17,500. The 10 most abundant peaks of each parent ion scan were subjected to CID fragmentation with a normalized collision energy (NCE) of 35 for network analysis, and 20, 25, 30, 40 and 45, respectively, for MS-based structure elucidation and the dynamic exclusion time was set to 10 sec. MS/MS scans were conducted with an AGC target of 3×10^6 or a maximum injection time of 150 ms. Thermo raw files were converted into mzXML file format using MSExport, and uploaded onto the GNPS web server (18). For network analysis, the data were filtered by removing all MS/MS peaks within +/- 17 Da of the precursor *m/z*. MS/MS spectra were window-filtered by choosing only the top 6 peaks in the +/- 50 Da window throughout the spectrum. The data were then clustered with MS-Cluster with a parent mass tolerance of 2.5 Da (to minimize isotopes of halogenated theonellamides appearing as different nodes in the network) and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. Further, consensus spectra that contained less than 2

spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes (18). Data were downloaded and visualized using Cytoscape 3.2. The full LC-MS dataset was uploaded to the MASSIVE database (MSV000081318 PW: 2017). Individual spectra were manually annotated using the software Xcalibur. Spectra of annotated molecules were uploaded to the GNPS library (gnps.ucsd.edu) to make them available for the community.

Phylogenetic analysis of AT domains

To analyze the putative theonellamide loading AT domain, amino acid sequences of 31 AT domains from *cis*-AT PKS modules of the polyketides soraphen, niddamycin, erythromycin, myxothiazole, pellasoren, gulmirecin, cystothiazole, and the theonellamide AT were selected. The AT from the *E. coli.* fatty acid synthase of (FabD) was selected as an outgroup. The sequences were retrieved from the GenBank database and aligned using Geneious 7.1.8 using the MUSCLE algorithm. The alignment allowed for the comparison of the diagnostic motif for AT-substrate specificity as described in (19). The phylogenetic reconstruction was performed with Geneious Tree Builder, employing the NJ algorithm. Bootstrap analysis was performed with 1,000 pseudo-replicate sequences. The substrates of known AT domains were inferred from polyketide structures.

Synthesis of test substrates

***S*-(2-Acetamidoethyl) 2-(4-bromophenyl)ethanethioate); 4-bromophenylacetyl-SNAC (13)**

N-Acetyl cysteamine (NAC) (66.0 µL, 0.59 mmol) was added to a stirred solution of 2-(4-bromophenyl)acetic acid (253.7 mg, 1.18 mmol) and a catalytic amount of 4-DMAP (one crystal) in anhydrous dichloromethane (2.4 mL) at 0 °C. Addition of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (208 µL, 0.59 mmol) was followed at 0 °C. After stirring overnight at room temperature the solution was quenched with saturated aqueous NH_4Cl (4 mL),

extracted with dichloromethane (2 × 4 mL). The organic phase was dried over anhydrous Na_2SO_4, filtered, and the solvent was removed under reduced pressure. Purification of the residue by silica gel chromatography (SiO_2, 1:1 to 2:8 n-hexane/EtOAc, 254 nm, R_f (EtOAc) 0.41) gave thioester **13** (96.2 mg, 52% yield) as a white crystalline solid. Identity was confirmed by ^{1}H NMR (300 MHz, $CDCl_3$) δ (ppm) 7.44 (d, J = 8.4 Hz, 2H), 7.13 (d, J = 8.4 Hz, 2H), 5.99 (br, 1H), 3.76 (s, 2H), 3.37 (q, J = 6.0 Hz, 2H), 2.99 (t, J = 6.6 Hz, 2H), 1.88 (s, 3H) (Fig. S10) and ^{13}C NMR (75 MHz, $CDCl_3$) δ (ppm) 197.2, 170.4, 132.4, 131.9, 131.2, 121.7, 49.8, 39.4, 29.0, 23.2 (Fig. S11). Measured ESI-HRMS *m/z* 316.0001 (calculated for $[M+H]^+$ $C_{12}H_{15}BrNO_2S^+$ 316.0001).

_S_-(2-Acetamidoethyl) 2-phenylethanethioate; phenylacetyl-SNAC (14)

An excess amount of *N*,*N*-diisopropylethylamine (DIPEA) (411 µL, 2.36 mmol) was added slowly to a stirred solution of NAC (66.0 µL, 0.59 mmol) in anhydrous dichloromethane (1 mL) at 0 °C. 2-Phenylacetyl chloride (186.1 mg, 1.18 mmol) was dissolved in anhydrous dichloromethane (1 mL) and added slowly to the solution. After stirring overnight at room temperature the solution was quenched with saturated aqueous NH_4Cl (4 mL) and extracted with dichloromethane (2 × 4 mL). The organic phase was dried over anhydrous Na_2SO_4, filtered, and the solvent was removed under reduced pressure. Purification of the residue by silica gel chromatography (SiO_2, 1:1 to 2:8 *n*-hexane/EtOAc, 254 nm, Rf (EtOAc) 0.45) gave thioester **14** (73.7 mg, 53% yield) as a yellow oil. Identity was confirmed by ^{1}H NMR (300 MHz, $CDCl_3$) δ (ppm) 7.33 (m, 5H), 5.98 (br, 1H), 3.84 (s, 2H), 3.40 (q, J = 6.0 Hz, 2H), 3.01 (t, J = 6.9 Hz, 2H), 1.88 (s, 3H) (Fig. S12) and ^{13}C NMR (75 MHz, $CDCl_3$) δ (ppm) 198.0, 170.4, 133.4, 129.6, 128.8, 127.6, 50.6, 39.5, 28.9, 23.15 (Fig. S13). Measured ESI-HRMS *m/z* 238.0892 (calculated for $[M+H]^+$ $C_{12}H_{16}NO_2S^+$ 238.0896).

_S_-(2-Acetamidoethyl) (_E_)-3-phenylprop-2-enethioate; cinnamoyl-SNAC (15)

An excess amount of DIPEA (404 µL, 2.32 mmol) was added slowly to a stirred solution of NAC (62.0 µL, 0.58 mmol) in anhydrous dichloromethane (1 mL) at 0 °C. Cinnamoyl chloride (186.2 mg,

1.12 mmol) was dissolved in anhydrous dichloromethane (1 mL) and added slowly to the solution. After stirring overnight at room temperature the solution was quenched with saturated aqueous NH_4Cl (4 mL) and extracted with dichloromethane (2 × 4 mL). The organic phase was dried over anhydrous Na_2SO_4, filtered, and the solvent was removed under reduced pressure. Purification of the residue by silica gel chromatography (SiO_2, 1:1 to 2:8 *n*-hexane/EtOAc) gave a mixture of thioester **15** and cinnamic acid as an orange crystalline solid. Thioester **15** was then recrystallized in acetonitrile (46.8 mg, 32% yield) to yield a crystalline solid. Identity was confirmed by ^{1}H NMR (300 MHz, $CDCl_3$) δ (ppm) 7.62 (d, J = 15.9 Hz, 1H), 7.55 (m, 2H), 7.40 (m, 3H), 6.73 (d, J = 15.9 Hz, 1H), 5.94 (br, 1H), 3.50 (q, J = 6.0 Hz, 2H), 3.16 (t, J = 6.6 Hz, 2H), 1.97 (s, 3H), consistent with previous reports (20). Measured ESI-HRMS *m/z* 250.0884 (calculated for $[M+H]^+$ $C_{13}H_{16}NO_2S^+$ 250.0896);

_S_-(2-Acetamidoethyl) ethanethioate; acetyl-SNAC (16)

Synthesis was performed according to a previously published procedure (21).

S-(2-acetamidoethyl) 3-hydroxybutanethioate; β-Hydroxybutanoyl-SNAC (17)

Synthesis was performed according to a previously published procedure (22).

***S*-(2-Acetamidoethyl) 4-oxopentanethioate (18), *S*-(2-acetamidoethyl) 2-methyloxazole-4-carbothioate (19), and S-(2-acetamidoethyl) 2-methylthiazole-4-carbothioate (20)**

Syntheses were performed according to a previously published procedure (23).

Construction of the *tna* AT-ACP expression vector

The DNA sequence corresponding to the *tnaA* AT-ACP loading didomain was amplified from the metagenomic DNA from the filamentous bacterial fraction of *T. swinhoei* WA using the primer pair

AT_fwd_*Bam*HI (5'-GTC GGA TCC CTT GCA GCA TTA TGA CGA TGT TC-3') and ACP_rev_*Hin*dIII (5'-CGT AAG CTT CTA GGT CTC TTG CCA TGG AGT C-3'). The gel purified gene fragment was digested with *Bam*HI and *Hin*dIII and cloned into pCDFDuet-1 (EMD Biosciences, Darmstadt, Germany), yielding the plasmid *tnaA*AT-ACP/pCDFDuet-1. The plasmid was isolated and introduced into the expression strains *E. coli* BL21(DE3) and *E. coli* BAP1. These strains were used for the expression of N-terminally His_6-tagged TnaA AT-ACP in the apo (*E. coli* BL21) or holo (*E. coli* BAP1) form.

Heterologous gene expression and protein purification of the *holo*-AT-ACP didomain

The *E. coli* expression strains were grown in TB medium supplemented with 50 µg/ml spectinomycin until an optical density (OD_{600}) of 0.8 was reached, after which the culture was cooled on ice for 30 min. Gene overexpression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a concentration of 0.1 mM. Induced cultures were grown for additional 24 h at 16 °C and afterwards harvested by centrifugation. Cell pellets were either processed directly or frozen in liquid N_2 and stored at -80 °C.

All purification steps were carried out at 4 °C. Cells were resuspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% glycerol) and disrupted by sonication using a Sonicator Q700 (QSonica, Newton, USA). The lysate was centrifuged for 45 min at 18,000 g to remove cell debris. The supernatant was incubated with Ni-NTA agarose (Macherey-Nagel, Oensingen, Switzerland) for 60 min and transferred to a fretted column. The resin was washed once with 3 ml lysis buffer (LB) and subsequently with 2 ml wash buffer 1 (same as LB, yet 20 mM imidazole) and finally eluted trice with 0.5 ml elution buffer (250 mM imidazole). Elution fractions were checked for the eluted protein by SDS-PAGE gel electrophoresis, after which a buffer exchange was conducted using a PD-minitrap column (GE Healthcare, Frankfurt a. M., Germany).

Competitive substrate depletion assays

To investigate the substrate specificity of the theonellamide loading module, we conducted substrate depletion assays, comparable to those of Zheng and co-workers (24). The purified AT-ACP didomain was simultaneously incubated with eight precursors (**13 - 20**) activated as *N*-acetylcysteamine thioesters (SNACs) to mimic naturally occurring CoA activated substrates. These included the predicted substrates bromophenylacetyl-SNAC (**13**) and phenylacetyl-SNAC (**14**), as well as the rather common PKS precursors cinnamoyl-SNAC (**15**) and acetyl-SNAC (**16**) and five further unusual precursors (**17**-**20**) as negative controls. All substrates were weighed in and individually dissolved in water containing 5 % DMSO to a stock concentration of 1 mM. Using the stock concentration, a master mix containing 100 µM of each respective substrate was prepared in water. All respective in vitro assays were set up from the substrate master-mix and the protein stock solution. The competitive depletion assay was set up in triplicate in a volume of 100 µl, containing 20.0 µM of the AT-ACP didomain, 20 µM of each respective substrate, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 4% [v/v] glycerol and 100 mM phosphate buffer (pH 8.0). Two negative controls were included in which the enzyme was either boiled at 98 °C for 10 min prior to addition of the substrates or substituted with buffer. Reactions were incubated at 30 °C for 20 min, quenched by addition of 20 µL concentrated formic acid and subsequently analyzed by HRMS. To prepare HPLC-MS samples, the precipitated protein was removed by centrifugation (4 °C, 15 min, 20000 × g) and the supernatant analyzed by HPLC-MS. Measurements were conducted on a QExactive Orbitrap MS (Thermo Scientific, Reinach, Switzerland) coupled to a UltiMate 3000 UHPLC system (Dionex, Reinach, Switzerland) and equipped with a Kinetex ® XB-C18 column (150 × 4.6 mm; Phenomenex, Torrance, CA, USA). The mobile phase consisted of water as solvent A and acetonitrile as solvent B, both supplemented with 0.1% formic acid. The gradient was isocratic at 5% solvent B for 3 min and increased to 95% B in 10 min, stayed isocratic at 95% B for 3 min, linearly decreased to 5% B in 0.1 min, followed by isocratic conditions for 2.9 min. MS measurement was conducted in positive ionization mode in a mass range of 100 - 1000 *m/z*. Collected Data of all MS experiments was analyzed using the Thermo Xcalibur 2.2. software.

Supplementary Tables

Table S1. 'E. serta' TSWA1 draft genome statistics

		G6	H6	Assembled
Contig Number	all contigs	5,239	4,805	3,146
	large contigs (≥ 1000bp)	1,759	1,170	1,570
Bases in contigs		7,686,988	5,112,621	8,973,789
G+C content [%]		54.8	54.6	54.9
Contig size (bp)	average	1,467	1,054	2,852
	smallest	50	50	201
	largest	23,891	37,334	55,291
N50 / *L50*		3885 / 513	2585 / 452	6936 / 358

Table S2. 'E. serta' TSWA1 draft single copy phylogenetic markers

present:

ffh	Signal recognition particle protein
infB	Translation initiation factor IF-2
lepA	Elongation factor 4
pheS	Phenylalanine-tRNA ligase alpha subunit
pheT	Phenylalanine-tRNA ligase beta subunit
pyrG	CTP synthase
rnhB	Ribonuclease HII
tgt	Queuine tRNA-ribosyltransferase
tpiA	Triosephosphate isomerase
tsaD	tRNA N6-adenosine threonylcarbamoyltransferase
rplA	50S ribosomal protein L1
rplB	50S ribosomal protein L2
rplC	50S ribosomal protein L3
rplE	50S ribosomal protein L5
rplF	50S ribosomal protein L6
rplK	50S ribosomal protein L11
rplN	50S ribosomal protein L14
rplO	50S ribosomal protein L15
rplP	50S ribosomal protein L16
rplR	50S ribosomal proteinGenes L18
rplV	50S ribosomal protein L22
rplX	50S ribosomal protein L24
rpsB	30S ribosomal protein S2
rpsC	30S ribosomal protein S3
rpsD	30S ribosomal protein S4
rpsH	30S ribosomal protein S8
rpsI	30S ribosomal protein S9
rpsK	30S ribosomal protein S11
rpsL	30S ribosomal protein S12
rpsM	30S ribosomal protein S13
rpsO	30S ribosomal protein S15
rpsQ	30S ribosomal protein S17
rpsS	30S ribosomal protein S19

Genes absent:

rplD	50S ribosomal protein L4
rplJ	50S ribosomal protein L10

Table S3. Comparison of completeness between 'Entotheonella' genomes.

	BSCG+RP		BUSCO v3		
	Present	Duplicated	Complete	Duplicated	Fragmented
'E. serta' TSWA1	33/35 (94.3%)	2	106/148 (71.6%)	3	9
'E. factor' TSY1	35/35 (100%)	1	126/148 (85.1%)	1	9
'E. gemina' TSY2	35/35 (100%)	2	90/148 (60.8%)	2	17

Table S4. Natural product biosynthetic domains and enzymes of 'E. serta' TSWA1 compared to the two previously sequenced 'Entotheonella' variants.

	'E. serta' TSWA1	'E. factor' TSY1	'E. gemina' TSY2
NRPS	*151*	*147*	*41*
Adenylation	46	44	11
Condensation	43	40	11
Epimerization	3	4	3
N-methyltransferase	1	1	0
Peptidyl-carrier protein	45	50	11
Thioesterase	11	5	3
MtbH-like protein	2	3	2
PKS	*153*	*115*	*25*
Ketosynthase	34	25	5
Adenyltransferase	17	14	6
***trans*-AT docking**	19	12	0
Ketoreductase	21	19	4
Dehydratase	11	8	1
Enoylreductase	2	2	1
O-methyltransferase	2	5	0
C-methyltransferase	5	0	0
Aminotransferase	3	4	2
Acyl-carrier protein	33	21	3
Thioesterase	4	4	2
Type III PKS system	2	1	1
Other	*4*	*13*	*13*
Ectoine synthase	0	2	0
Cyanobactin synthase	0	0	2
YcaO cyclodehydratase	0	1	0
Lassopeptide cyclase	0	1	1
RiPP amino acid epimerase	1	3	2
Lanthionine synthase	0	1	2
Nitrile hydratase leader peptide	0	3	4
Nif11-like leader peptide	0	1	1
Staurosporin dimerization enzyme	1	0	0
Terpene cyclase	2	1	1
Total	**308**	**275**	**79**

Table S5. NRPSpredictor2 results for A domains from 'E. serta TSWA1'. TNA refers to A domains from theonellamide gene cluster contigs obtained by gap closure PCR, and are separated into TNA1 and TNA2 corresponding to regions 1 and 2 of the theonellamide BGC. A domains from theonellamide found on initial sequencing contigs are marked with [tna1] or [tna2]. When multiple A domains were detected on a contig, they are listed in sequence order. Not all A domains detected by antiSMASH and BLAST-based annotation yielded good predictions from NRPSpredictor.

Sequence id	Diagnostic residues	Prediction
NODE_19_length_23254_cov_84.159	DLYNMSLIW-	Cys
NODE_19_length_23254_cov_84.159	DLFNNALTY-	Ala
NODE_21_length_22347_cov_80.998	GVFWLAASA-	Oiv
NODE_77_length_15626_cov_82.1233	DATKVGHVGK	Asn
NODE_77_length_15626_cov_82.1233	DIVQLGLIW-	Gly
NODE_102_length_13681_cov_28.1532	DVSFMGAIMK	Phe
NODE_128_length_12577_cov_15.9641	DVWHISLIDK	Ser [tna1]
NODE_128_length_12577_cov_15.9641	DATKVGEVGK	Asn [tna1]
NODE_131_length_12428_cov_62.3546	--WLYNDDVK	Leu
NODE_178_length_10938_cov_674.184	DIFFIGIVL-	Cys/Ile
NODE_178_length_10938_cov_674.184	DILQ------	Gly
NODE_188_length_10616_cov_57.6286	DVWHISLVDK	Ser [tna1]
NODE_188_length_10616_cov_57.6286	DAWTVAAVCK	Phe [tna1]
NODE_217_length_9792_cov_222.438	DVWHLSLIDK	Ser
NODE_217_length_9792_cov_222.438	DVWHLSLIDK	Tyr
NODE_358_length_7044_cov_7.49125	IFYYLAGAS-	Iva
NODE_370_length_6864_cov_58.92	DATKVGEVGK	Asn [tna2]
NODE_370_length_6864_cov_58.92	DAAIIAAVC-	Phe [tna2]
NODE_419_length_6363_cov_170.727	AYGYVSADIK	Ser
NODE_464_length_5864_cov_191.233	---FYAHVVK	Pro
NODE_464_length_5864_cov_191.233	VDWVIS-LG-	Ala-b
NODE_493_length_5509_cov_291.31	VDWATSLAD-	Ala-b
NODE_700_length_4040_cov_3.50192	IYMYMGGPV-	Pip
NODE_710_length_3990_cov_152.12	DASCVAGLL-	Bht

NODE_859_length_3100_cov_25.3946	DAFWLG----	Val
NODE_871_length_2800_cov_15.4219	DAFFLGVTFK	Ile
NODE_933_length_2766_cov_2.88139	SAAHYAAIFK	Phe
NODE_948_length_2720_cov_91.9819	DLFYLALVC-	Iva
NODE_1050_length_2084_cov_6.40208	DVQFNAHAIK	Pro
NODE_1172_length_1732_cov_5.76923	DATKVGEVGK	Asn [tna1]
TNA1	DFWNIGMVHK	Thr
TNA1	DVWHISLVDK	Ser
TNA2	VDWVVSLGDK	Ala-b
TNA2	DPRMFVLRAK	Aad/Gln
TNA2	DSVLIAEVWK	His

Abbreviations: Oiv, 2-oxo-isovaleric acid; Iva, isovaleric acid; Ala-b, β-alanine; Pip, pipecolic acid; Bht, β-hydroxytyrosine; Aad, aminoadipic acid

Table S6. ORFs detected on the 'E. serta' TSWA1 loci containing the *tna* genes and their putative functions.

ORF	Protein size [aa]	Proposed protein function	Closest protein homologue [source organism]	Identity [%]	GenBank accession number
tnaA	2729	PKS : AT-ACP-KS-AT-DH-KR-cMT-ACP	Curacin polyketide synthase CurJ [*Moorea producens*]	40.7	WP_008191795
tnaB	1852	PKS: KS-AT-DH-KR-ACP	6-deoxyerythronolide-B synthase [*Methylobacter tundripaludum*]	44.8	WP_006892897
tnaC	4817	PKS/NRPS hybrid: KS-AT-ACP-AmT-C-PCP-C-A(Ser)-PCP-C-A(Asn)-PCP-E	Uncharacterized protein [*Scytonema hofmannii* PCC 7110]	44.6	ANNX02000017
tnaD	354	Dioxygenase	Taurine catabolism dioxygenase TauD/TfdA [*Microcoleus vaginatus*]	55.8	WP_006635045
tnaE	1802	NRPS: C-A(Asn)-PCP-C-PCP	Non-ribosomal peptide synthase [*Scytonema* sp. HK-05]	41.2	WP_073634571
tnaF	4027	C-C-A(Thr)-PCP-C-A(Ser)-PCP-C-A(Phe)-PCP-TE	Malonyl CoA-acyl carrier protein transacylase [*Archangium gephyra*]	44.4	WP_047857045
tnaG	339	Monooxygenase	LLM class flavin-dependent oxidoreductase [*Bradyrhizobium* sp. YR681]	47.4	WP_008142486
tnaH	835	Cyclase	Guanylate cyclase [*Candidatus* Rokubacteria bacterium CSP1-6]	40.6	KRT67164

tnaI	441	Glycosyltransferase	Glycosyltransferase [*Methylobacter tundripaludum*]	36.9	WP_027150830
ORF1	207	DUF2063-containing hypothetical protein	Hypothetical protein CBB70_10885 [*Planctomycetaceae* bacterium TMED10]	34.5	OUT65940
(ORF2)	45	Hypothetical protein	[none]		
ORF3	106	Acyl-coA dehydrogenase	Hypothetical protein AUI83_03445 [*Armatimonadetes* bacterium 13_1_40CM_3_65_7]	67.3	OLD59118
ORF4	187	Transporter	Major Facilitator Superfamily protein [bacterium YEK0313]	57.4	CEJ15702
ORF5	151	Transporter	Cyanate permease [*Rhizobiales* bacterium GAS113]	56.8	SDR24766
tnaP	5379	NRPS: C-A(Asn)-PCP-C-A(Phe)-PCP-C-A(Ser)-PCP-C-A(Aad)-PCP-C-A(His)-PCP	Uncharacterized protein [*Archangium* sp. Cb G35]	41.6	WP_073560596
(ORF6)	120	NRPS (fragment)	Non-ribosomal peptide synthetase [*Calothrix* sp. HK-06]	35.2	WP_073622402
tnaR	534	Methyltransferase	PhpK family radical SAM P-methyltransferase [*Pseudomonas fluorescens*]	50.2	WP_003177935
ORF7	415	Enamine deaminase-like protein	Hypothetical protein [*Bosea* sp. RAC05]	47.0	WP_083247627

ORF8	190	Hypothetical protein	Hypothetical protein [*Rhodospirillales* bacterium URHD0088]	41.2	WP_051474581
(ORF9)	101	Metallo-hydrolase (fragment)	MBL fold metallo-hydrolase [*Paenibacillus ferrarius*]	70.8	WP_079413376
ORF10	287	Metallo-hydrolase	MBL fold metallo-hydrolase [*Bradyrhizobium* sp. NAS80.1]	46.4	WP_084805835
ORF11	288	Universal stress protein	Universal stress protein [*Planctomyces* sp. SH-PL14]	35.5	WP_075091812
ORF12	132	Transcriptional regulator	PadR family transcriptional regulator [*Anabaena cylindrica*]	39.2	WP_015364289
ORF13	316	Transcriptional regulator	Transcriptional repressor [*Pyrinomonas methylaliphatogenes*]	42.5	WP_041976756
ORF14	358	Oxidoreductase	gfo/Idh/MocA family oxidoreductase [*Microtetraspora malaysiensis*]	37.1	WP_067138380
ORF15	227	Hypothetical protein	Hypothetical protein [*Cesiribacter andamanensis*]	30.2	WP_009194155
ORF16	128	Glutathione transferase	Glutathione transferase [*Roseivivax lentus*]	30.8	WP_076448207
ORF17	413	Amidase	Amidase [*Rhodoplanes* sp. Z2-YC6860]	55.4	WP_068014341
ORT18	185	Osmotically inducible protein C	Osmotically inducible protein C [*Bradyrhizobium* sp. Ec3.3]	70.7	WP_027526663

(ORF19)	56	Oxidoreductase (fragment)	Pyridine nucleotide-disulfide oxidoreductase [*Mycobacterium kansasii*]	66.7	WP_083103196
ORF20	372	Oxidoreductase	Pyridine nucleotide-disulfide oxidoreductase [*Bradyrhizobium* sp. Ec3.3]	60.3	WP_027526662

Table S7. LC-HRMS data of identified theonellamide analogs

Compound name	Sum formula	Calculated	Detected	Mass error (ppm)	Isotope ratio			
					simulated		found	
					M0/M1	M0/M2	M0/M1	M0/M2
Theonellamide B (**8**)	$C_{70}H_{89}BrN_{16}O_{23}$	801.2808 1601.5543	1601.5507	3.6	0.75	0.97	0.68	1.0
Theonellamide A (**4**)	$C_{76}H_{99}BrN_{16}O_{28}$	881.8033 1763.6071	1763.6020	5.1	0.86	0.55	0.89	0.56
Theonegramide (**11**)	$C_{75}H_{97}BrN_{16}O_{26}$	859.3044 1717.6016	1717.6029	0.973	0.81	0.97	0.84	1.0
Theonellamide H* (**12**)	$C_{69}H_{87}BrN_{16}O_{22}$	786.2755 1571.5437	1571.5429	0.84	0.74	0.97	0.84	1.1
Deoxy-**8***	$C_{70}H_{89}BrN_{16}O_{22}$	793.2833 1585.5593	1585.5613	1.93	0.75	0.97	0.83	1.1
Dideoxy-**8***	$C_{70}H_{89}BrN_{16}O_{21}$	785.2859 1569.5644	1596.5657	1.24	0.75	0.97	0.86	1.12
Deoxy**-12***	$C_{69}H_{87}BrN_{16}O_{21}$	778.2780 1555.5488	1555.544	2.27	0.74	0.97	0.74	1.14

* new theonellamides

Table S8. Primers used for gap closing in *tna* pathway from 'E. serta' TSWA1

Primer Name	Primer Sequence	Primer Name	Primer Sequence
1_5381 F	GACCGTCAGATGCTGTTGC	2_03_F	CAACCTCGCCTATATCATCT
1_6125 R	GTCATGCCGAAGATCAAATCG	2_03_R	TGAACGCGCAACATATCT
2_8232 F	AATCCCTCTCTCCTATCAAACG	2_05a_F	TTTGGCTGGCGGACTTGA
2_8904 R	TGGCTTTTGGAAAAGTTGATGC	2_05a_R	GTATAGGGAATTGCGGCTGA
3_9848 F	ATCGCGTTTCTTTTCAGTGG	14.2_F	GCCGACATCTGGAACGAA
3_10531 R	TTAGGCGAACATCAAGGATCG	14.2_R	TGCCGCTACAATCGAACT
5_13758 F	CTTTCTGAACGAATATCTCCCG	2_04b_F	CAGTACGCTGATTTCGCC
5_14432 R	GATAGAGCACCTGCAAAAGC	2_04b_R	GTTGTTGTCGAGACGGATG
6_25456 F	CATCTCCAGAATCCGAGACG	2_04d_F	AACCTCTATGGCTCTACT
6_26124 R	GTATGGACAATTCAGCAAGTGG	2_04d_R	TGTTGAAGACGGAGGTAT
7_30392 F	CTTGATCAACGAATATGGTCCC	2_05b_F	CCTTTATCGAGCTACTCC
7_31049 R	CAAAAGCCTATTCCAAAGGTCG	2_05b_R	GTAAAGCCAATAACCGCA
4_12270 F PH	ACTCCAAGATCATATCACCACC	2_05c_F	CAGGCCAAAATCAAACTC
4_12809 R PH	TCGTAATAGGCATCCTTGTCC	2_05c_R	ATTTCGGGGATTTCGATG
7_30392 F2 PH	CTTGATCAACGAATATGGTCCC	2_05d_F	CCTTCTTGACCAATCCCT
7_31049 R2 PH	CAAAAGCCTATTCCAAAGGTCG	2_05d_R	TCTCGACTGCTGAATCCT
11_F	GCGAGTTGTTGGGTGAGA	2_05f_F	CGCTTATCGGCTTCTTTG
11_R	TGGATGATGGGTTGCGTT	2_05f_R	CTGCGGTGTGAAATCTGT
12_F	GAGCGGTTGTTGATCAGT	2_07_F	TTTAGGCGGTCACTCTTT
12_R	TGTTGGGCCATAGGCATT	2_07_R	CCGTACTCTTCCCTTTTTT
17_F	ATACACCTTCTCCCACAC	2_08_F	CAAGCTGCTGATGTCCTC
17_R	GTCCTAAATGGCTCGTGA	2_08_R	TTCTTCGTCGTGCTTATCC
18_F	TGGCCAAAATCCCGATCT	9.2c_F	GTCTTCGCCGTAATGCTC
18_R	TTGACTCCATCACCAGGTT	9.2d_R	TTCAGGGTGGATTCGATT
15_F	ATCTCTATGCTGGCCTTT	8.3b_F	TACAGAGACCAGAGGCCA
15_R	CTTTAGGATGAGGGTCGTT	8.3b_R	GGAGAGGTTATAAGAGGCA
16_F	TTGGCGGGTTCTAGAGAT	2_05e_F	GCTTACAACATTCCCGCT
16_R	TGCTGTTGTTCTTGGGTT	2_05e_R	AAACATGGCCTGGAACAC
20_F	TATTGTTCACTCGGCAGC	9.2c_F	GTCTTCGCCGTAATGCTC
20_R	TGGTCGAGTCTCATGGTT	9.2d_R	TTCAGGGTGGATTCGATT
8.2a_F	CCCCACCCTTTTTGCAAC	8.2b_F	GTTTGAGTTGGTGGCGTG
8.2a_R	ACGAATCCCTCCAGCAAT	8.2b_R	GCTGGGCAAATGATAGAGG
9.2a_F	CCGAACTGCCGATACAAT	2_05a_F	TTTGGCTGGCGGACTTGA
9.2a_R	GATGGTAGGAAAGGGGAGA	2_05b_R	GTAAAGCCAATAACCGCA
9.2b_F	TGTATCGGCTCTCACGTT	9.2c_R	TCCAAAAAGATGCCGTGT
9.2b_R	TCAGGTCATCTCGCAGCA	9.2e_F	GGTCTTCGTGGGTGGGAAA
10_F	GACGTTGTTGGCTGCTTG	2_04b_F	CAGTACGCTGATTTCGCC
10_R	CGGTATTTGGGTGTCGCT	2_04d_F	AACCTCTATGGCTCTACT
13_F	CGTTGTAGCAGGTGAGAG	2_04c_F	GATTGACATGGTGGTGGG
13_R	GAAGAGGAACGAGGGGTGA	2_04d_R	TGTTGAAGACGGAGGTAT
19_F	TCCGGACTCAACTCAACCT	2_05f_F	CGCTTATCGGCTTCTTTG
19_R	CGCCCACACTCAAAACCA	2_06_R	CTCAGCCAGTTGTTCCTC
9.2e_F	GGTCTTCGTGGGTGGGAAA	8.2b_F_800	GTTTGAGTTGGTGGCGTG
9.2e_R	AAGGACAGAGGTTGGGGA	8.2b_R_800	GCTGGGCAAATGATAGAGG
9.2c_F	GTCTTCGCCGTAATGCTC		
9.2d_R	TTCAGGGTGGATTCGATT		
9.2c_F	GTCTTCGCCGTAATGCTC		
9.2d_R	TTCAGGGTGGATTCGATT		
2_02_F	CGAGAATCAAAGCCATCA		
2_02_R	AGAGAATTCAGAACGGATAG		

Supplementary Figures

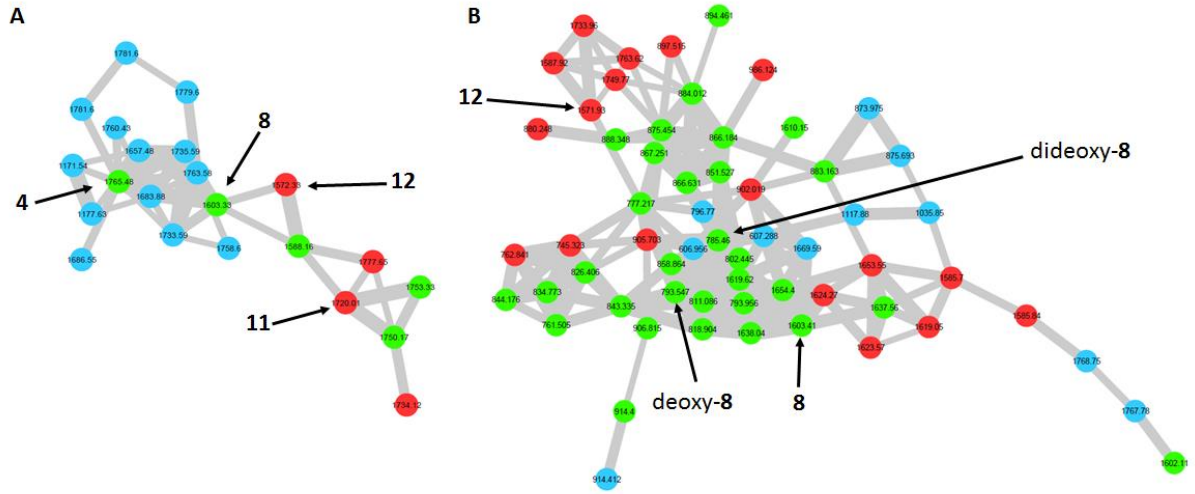

Fig. S1. Molecular network of singly (A) and doubly charged (B) theonellamides from sponge samples from Japan and Israel. Nodes are color coded according to the sponge sample they were isolated from: blue, TSW Japan; red, TSW Israel; green, metabolites present in both sponges. Numbers within each node indicate the averaged exact masses over 2.5 Da and may include isotopes. As a result, the masses indicated are not to be interpreted as exact masses. The edge line width indicates the relatedness between two metabolites (cosine 0.7).

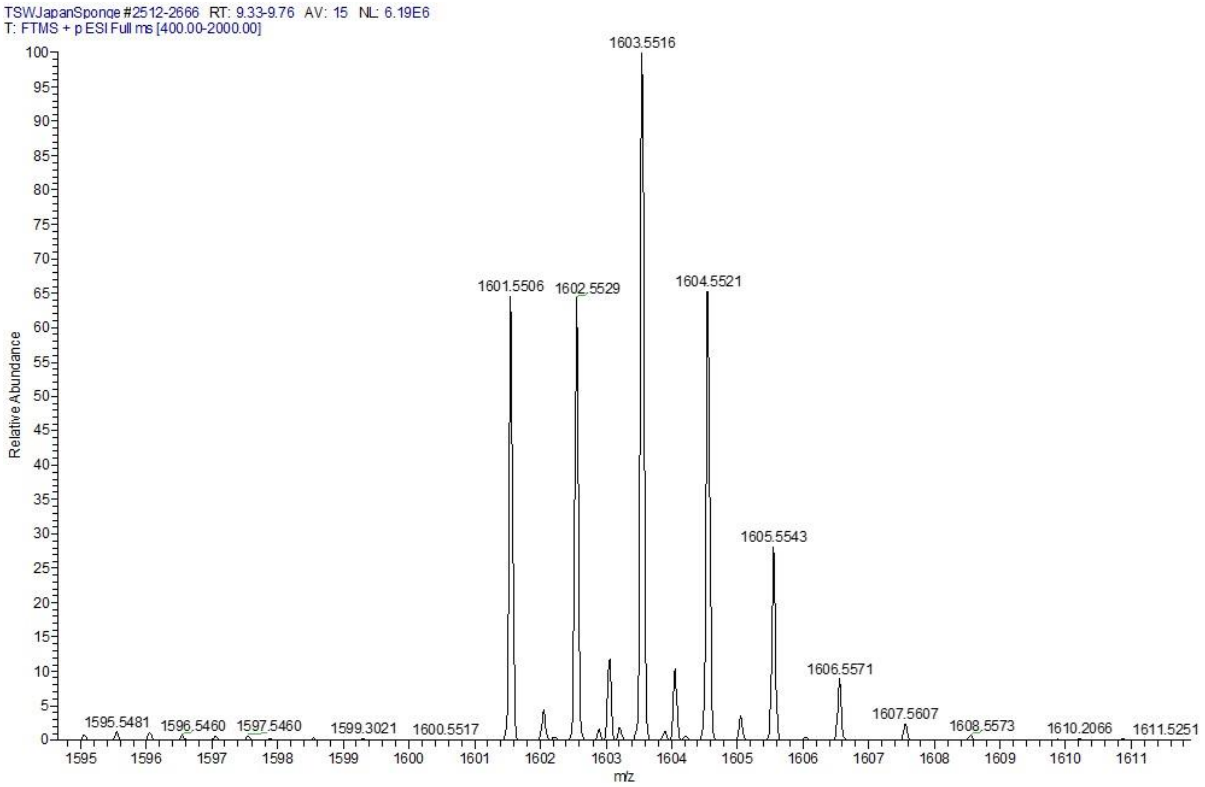

Fig. S2. Mass spectrum of theonellamide B (**8**), showing characteristic isotopic pattern of theonellamides. Detailed structural analysis of this compound is laid out in Figs. S7-9.

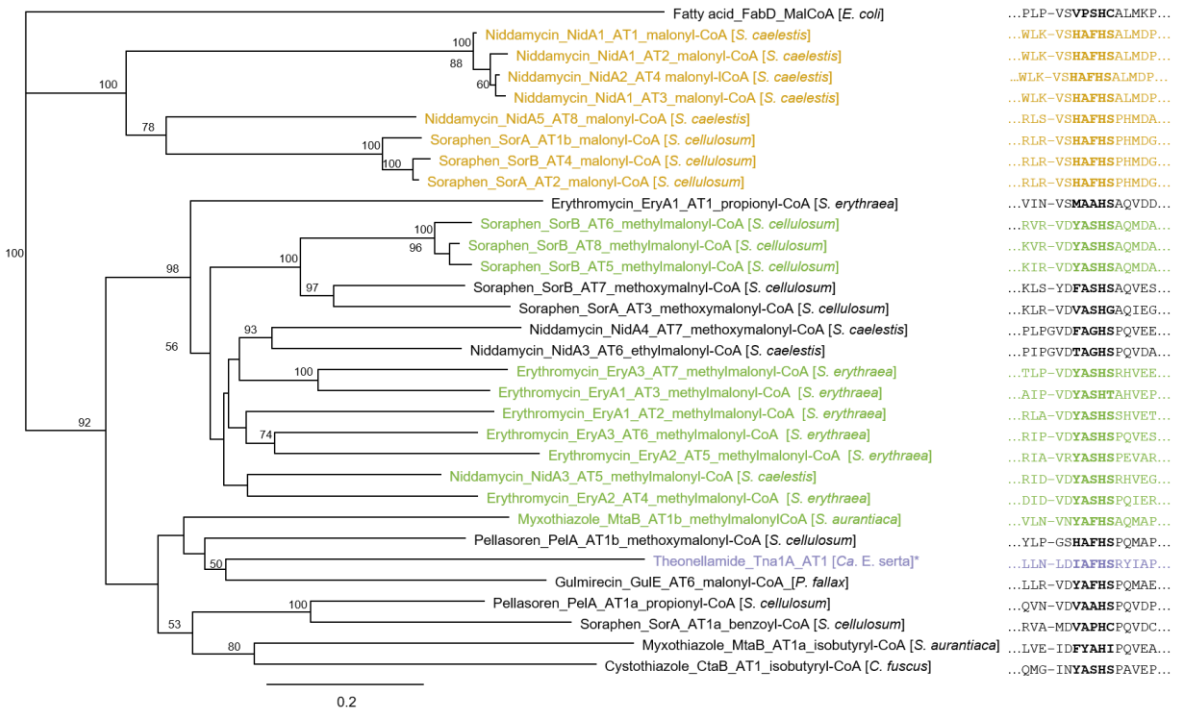

Fig. S3. Phylogram of AT domains from various pathways. Tip labels consist of the product of the relevant biosynthetic pathways, protein name, module number, putative substrate and organism. On the right is a sequence alignment of the corresponding AT region that contains residues previously identified (19, 25-27) as diagnostic for substrate specificity. The outgroup is the *E. coli* AT from fatty acid biosynthesis (FabD). In the alignment, the sequence of the AT substrate specificity motif is marked in bold. Yellow denotes ATs selecting malonyl-CoA units and green refers to methylmalonyl-CoA specificity. The putative loading AT from the 'E. serta' theonellamide biosynthetic pathway is marked in purple.

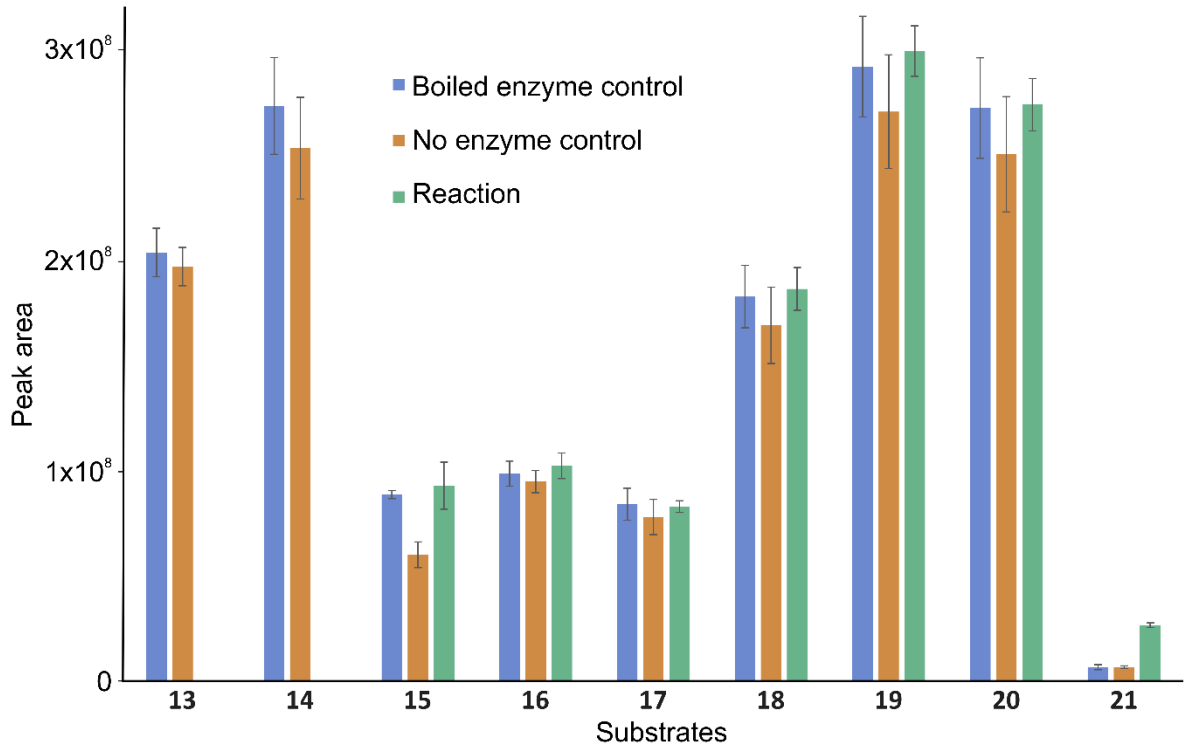

Fig. S4. Competitive TnaA AT-ACP depletion assay performed in triplicates. Depletion assay was carried out with substrates **13-21**, as shown in Fig. 4 and Figs. S5 and S6. Negative controls contain either boiled or no enzyme. Bar heights represent the average peak area of each compound. Error bars represent the standard deviation between the triplicates. Peak areas were assigned using the Avalon algorithm implemented in the Xcalibur software package and manual correction when necessary.

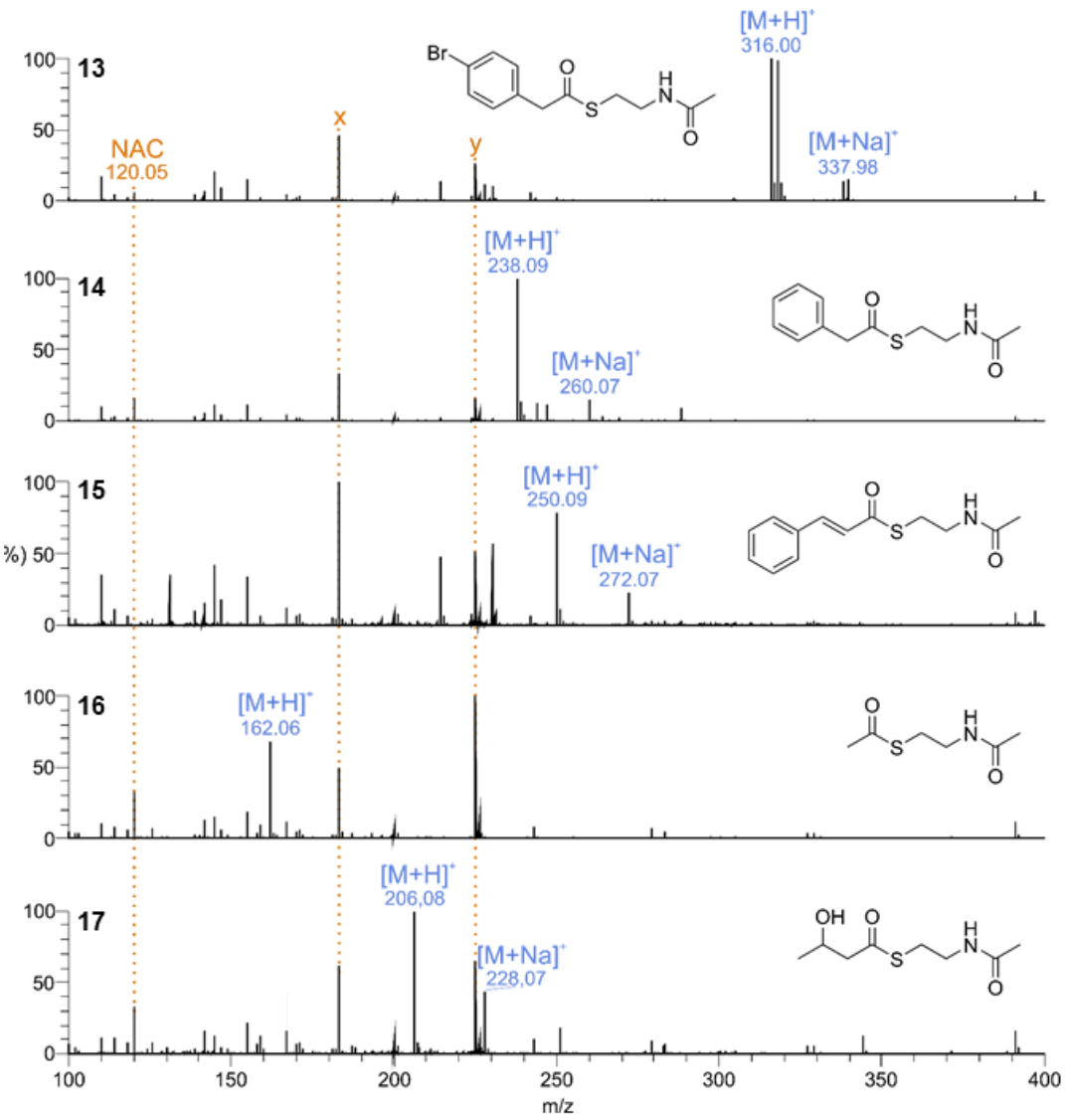

Fig. S5. Mass spectra of compounds **13-17** from the extracted ion chromatogram (EIC) of the boiled enzyme control (labeled a in Fig. 4). Monoisotopic peaks corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of the respective substrates labeled in blue. Orange labels mark background peaks from the MS instrument (x and y) and free NAC stemming from fragmentation during ionization.

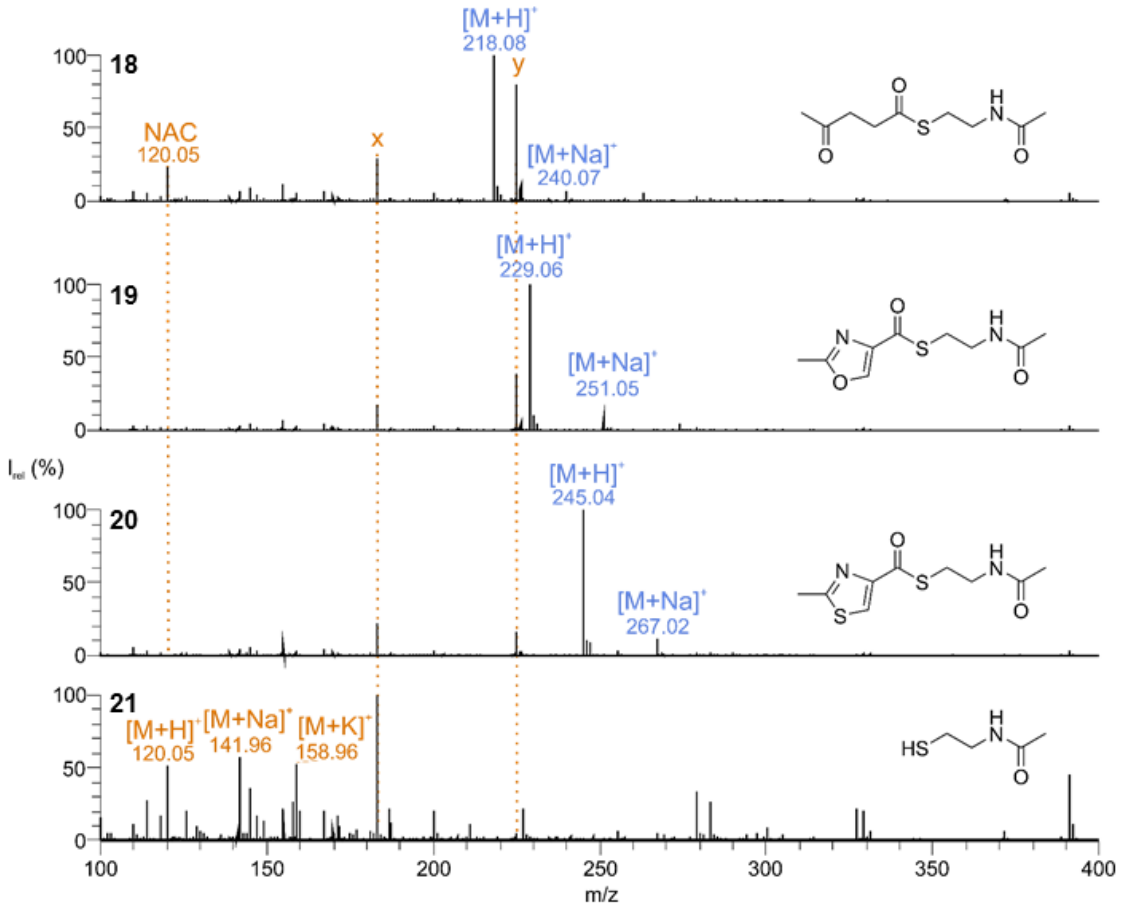

Fig. S6. Mass spectra of compounds **18**-**20** from the EIC of the boiled enzyme control (labeled a in Fig. 4), and free NAC (**21**) from the reaction (labeled c) in Fig. 4). Monoisotopic peaks corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of the respective substrates labeled in blue and of free NAC in orange. Orange labels mark background peaks from the MS instrument (x and y) respectively, and free NAC stemming from fragmentation during ionization.

Detailed Structural Analysis of Theonellamide B

In this section, we describe in detail the MS-based structural assignment of theonellamide B, as an example of the analysis performed for each labelled node in Fig. 3. MS^2 spectra for structurally assigned theonellamides were uploaded to the GNPS library (gnps.uscd.edu). Theonellamides were identified based on their characteristic isotope pattern resulting from bromination. As a result of the chemical complexity and halogenation, the isotopic pattern of theonellamide B (**8**; $C_{70}H_{89}BrN_{16}O_{23}$) is composed of six signals (Fig. S2), $[\mathbf{8}+H]^+$ (65% of relative abundance), $[\mathbf{8}+H+1]^+$ (65% of relative abundance), $[\mathbf{8}+H+2]^+$ (100% of relative abundance), $[\mathbf{8}+H+3]^+$ (65% of relative abundance), $[\mathbf{8}+H+4]^+$ (30% of relative abundance) and $[\mathbf{8}+H+5]^+$ (10% of relative abundance). We observed theonellamides as either single or double charged ions. The fragmentation pattern, however, differed significantly (collision energy was set to 25 eV). Fig. S7 shows the MS^2 spectra of theonellamide B (*m/z* 1601.5506 Da $[\mathbf{8}+H]^+$ and *m/z* 801.26 Da $[\mathbf{8}+2H]^{2+}$), either singly or doubly charged, as an example. MS-based structure elucidation was achieved by combining the information obtained from different collision energies.

In the MS^2 spectrum at *m/z* 1601.5506 $[\mathbf{8}+H]^+$, the most intense fragment signal at *m/z* 1583.5417 corresponds to a H_2O loss from the molecular ion, and two additional H_2O losses were observed with low intensities. Other fragments correspond to the (5*E*,7*E*)-3-amino-4-hydroxy-6-methyl-8-phenyl- 5,7-octadienoic acid (Apoa) moiety (Fig. S8). Interestingly, in the MS^2 spectrum of $[\mathbf{8}+2H]^{2+}$, the most intense signal is also related to Apoa, (Fig. S9), however Apoa is observed here as a signal at *m/z* 226.1224 Da and the most intense signal at *m/z* 209.0959 corresponds to $[Apoa\text{-}NH_3]^+$.

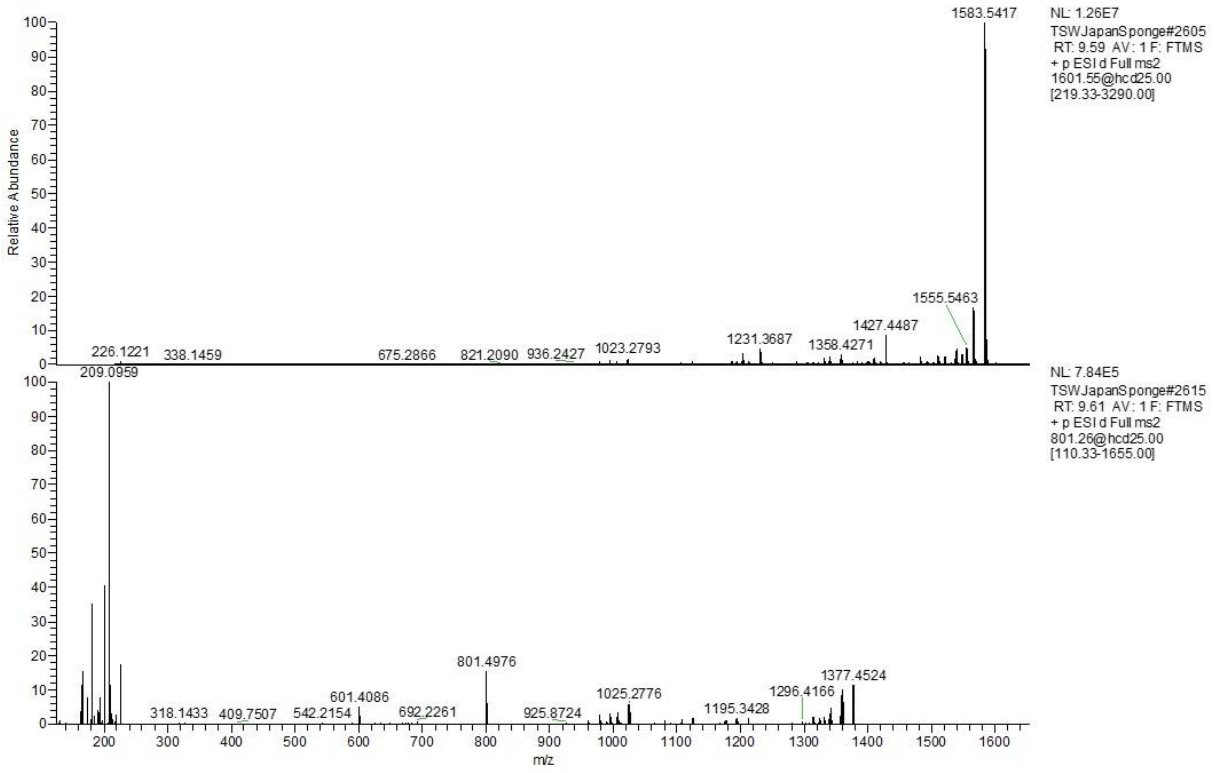

Fig. S7. Comparison of MS fragmentation patterns between [**8**+H]$^{+}$, (top) and [**8**+2H]$^{2+}$ (bottom).

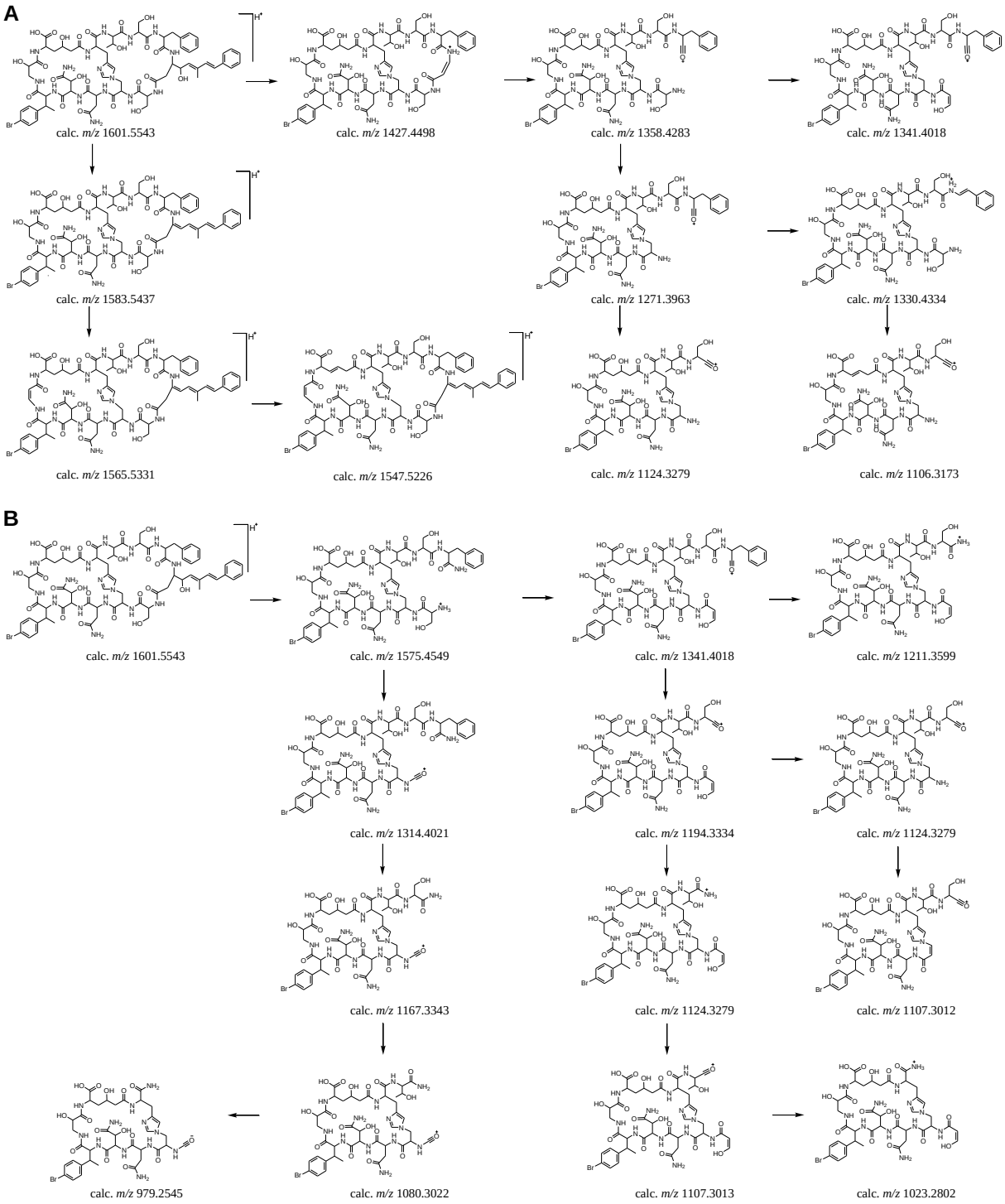

Fig. S8. Observed fragments for [**8**+H]$^+$.

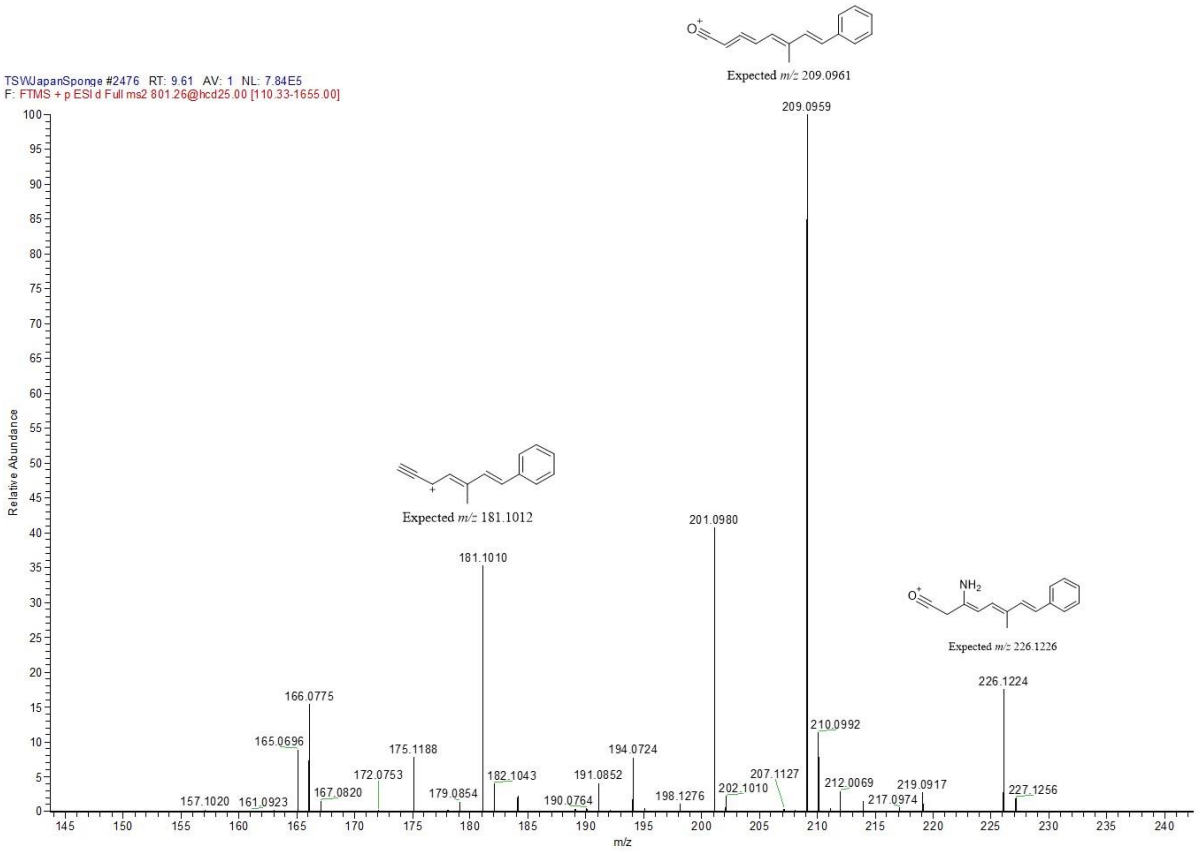

Fig. S9. Characteristic MS^2-fragments of $[\mathbf{8}+2H]^{2+}$.

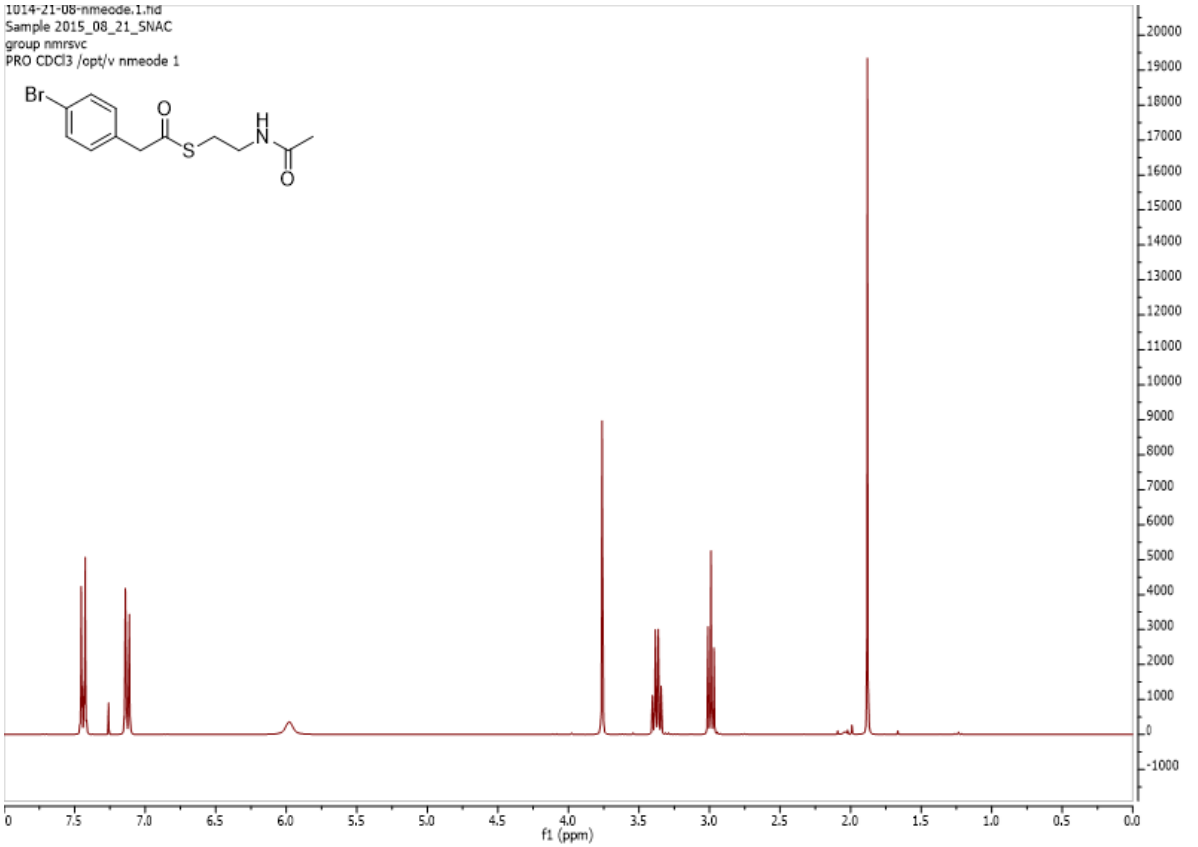

Fig. S10. ^{1}H NMR (300 MHz, $CDCl_3$) spectrum of **13**.

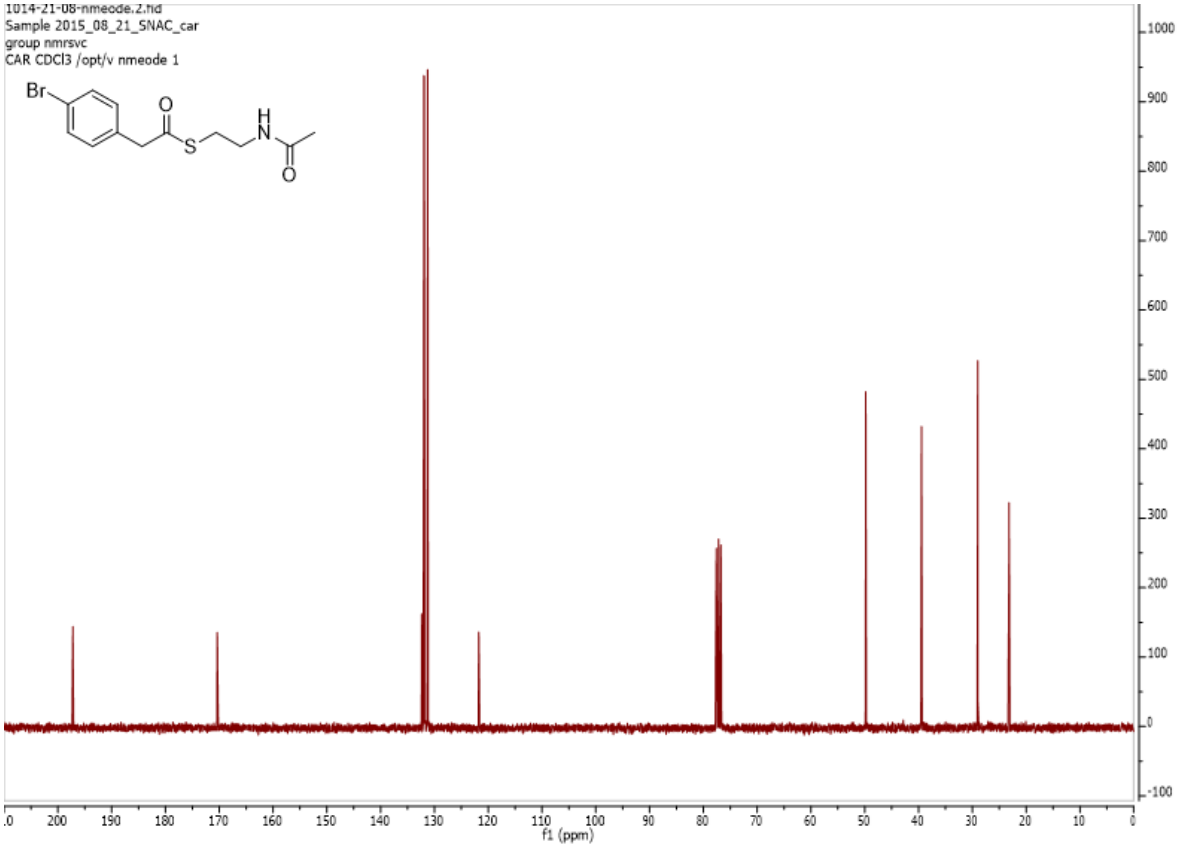

Fig. S11. ^{13}C NMR (75 MHz, $CDCl_3$) spectrum of **13**.

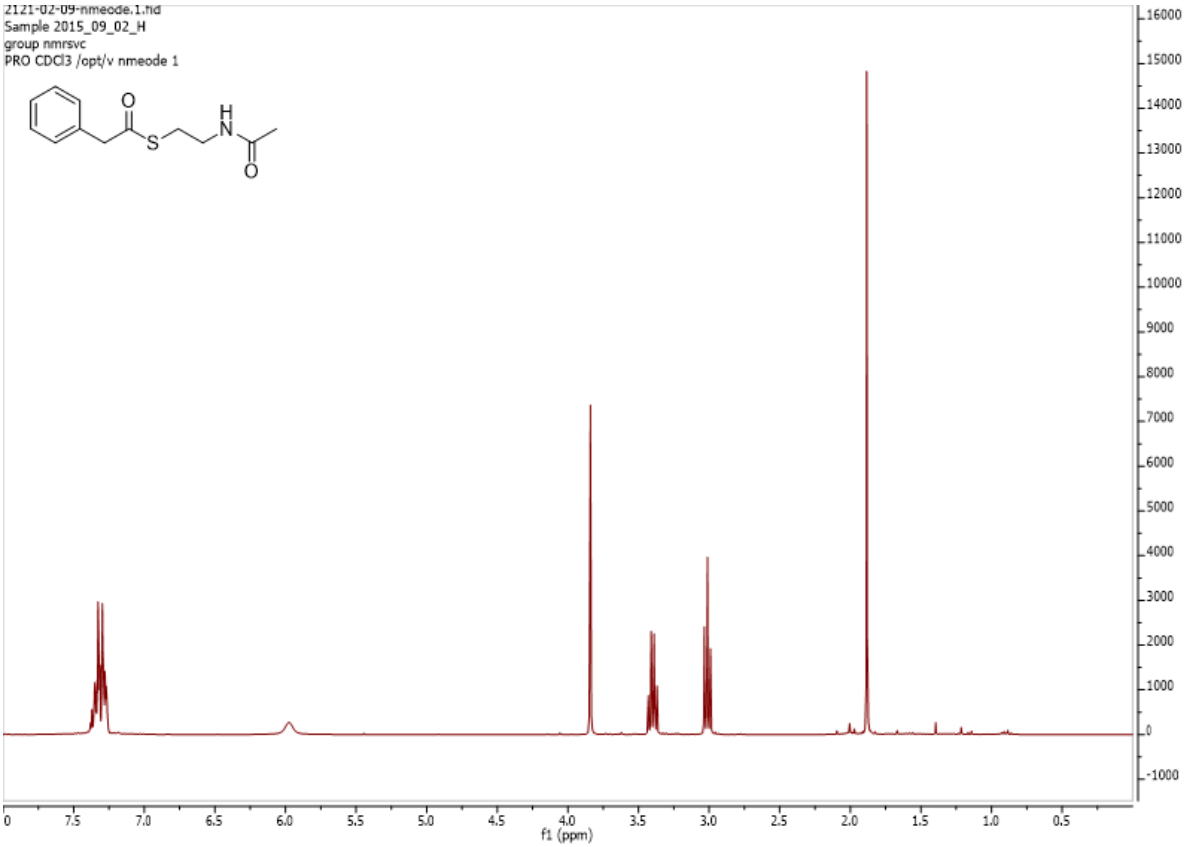

Fig. S12. ^{1}H NMR (300 MHz, $CDCl_3$) spectrum of **14**.

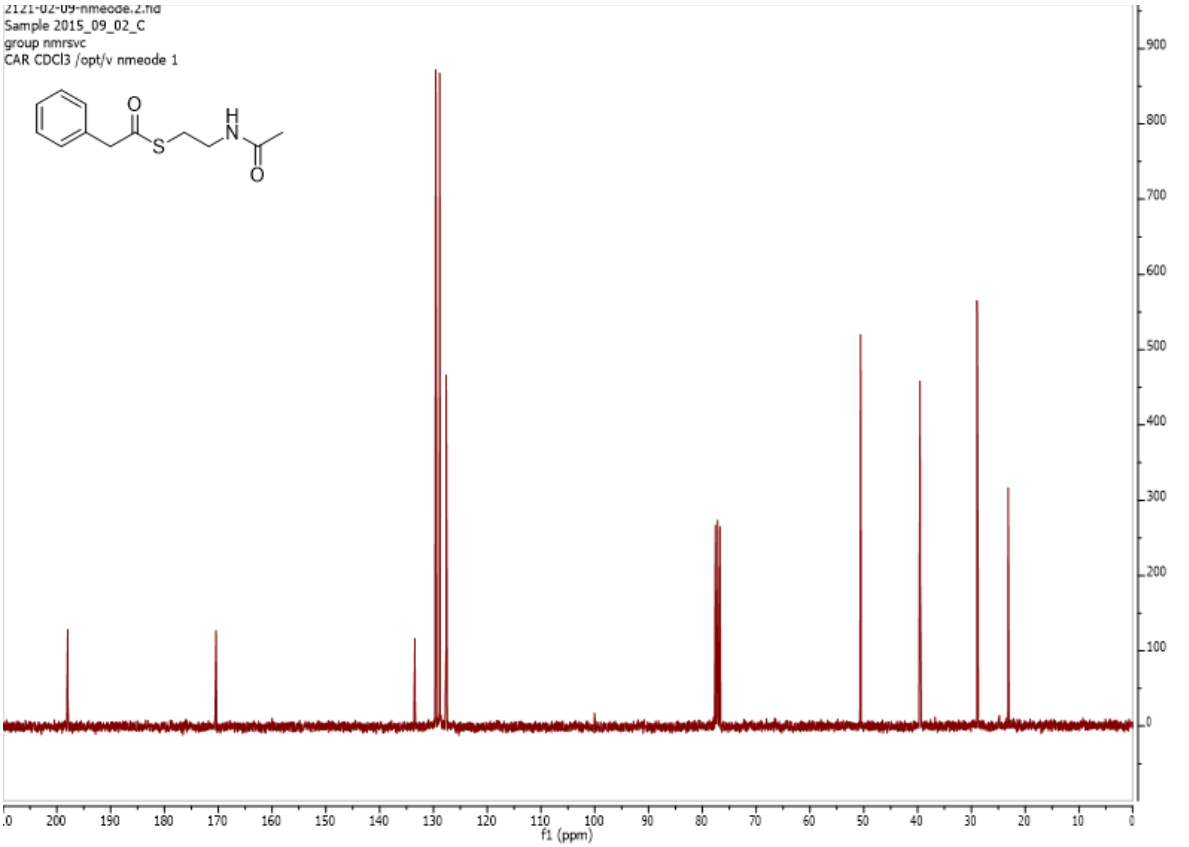

Fig. S10. ^{13}C NMR (75 MHz, $CDCl_3$) spectrum of **14**.

Supplementary References

1. Ueoka R*, et al.* (2015) Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nature Chemical Biology* 11(9):705-712.
2. Spiegel M & Rubinstein N (1972) Synthesis of RNA by dissociated cells of the sea urchin embryo. *Experimental Cell Research* 70(2):423-430.
3. Tauch A, Kassing F, Kalinowski J, & Pühler A (1995) The *Cornyebacterium xerosis* Composite Transposon Tn*5432* Consists of Two Indentical Insertion Sequences, Designated IS*1249*, flanking the Erythromycin Resistance Gene *ermCX*. *Plasmid* 34: 119-131.
4. Maus I*, et al.* (2015) Complete genome sequence of the strain *Defluviitoga tunisiensis* L3, isolated from a thermophilic, production-scale biogas plant. *Journal of Biotechnology* 203:17-18.
5. Andrews S (2015) FastQC Blog: A Quality Control Tool for High ThroughputSequence Data. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
6. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
7. Bankevich A*, et al.* (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19(5):455-477.
8. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069.
9. Aziz RK, *et al*. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.
10. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210-3212.
11. Stanke M, Steinkamp R, Waack S, & Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* 32(Web Server issue):W309-312.
12. Wilson MC*, et al.* (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506(7486):58-62.
13. Weber T*, et al.* (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* 43(W1):W237-W243.
14. Ziemert N*, et al.* (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS ONE* 7(3):e34064.
15. Wibberg D*, et al.* (2015) Development of a *Rhizoctonia solani* AG1-IB specific gene model enables comparative genome analyses between phytopathogenic *R. solani* AG1-IA, AG1-IB, AG3 and AG8 isolates. *PloS ONE* 10(12):e0144769.
16. Wibberg D*, et al.* (2017) Draft genome sequence of the potato pathogen *Rhizoctonia solani* AG3-PT isolate Ben3. *Archives of Microbiology*:1-4.
17. Youssef DT*, et al.* (2014) Theonellamide G, a potent antifungal and cytotoxic bicyclic glycopeptide from the Red Sea marine sponge *Theonella swinhoei*. *Marine drugs* 12(4):1911-1923.
18. Wang M*, et al.* (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 34(8):828-837.
19. Haydock SF*, et al.* (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Letters* 374(2):246-248.
20. Munde T*, et al.* (2013) Biosynthesis of tetraoxygenated phenylphenalenones in Wachendorfia thyrsiflora. *Phytochemistry* 91:165-176.
21. Jenner M*, et al.* (2013) Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angewandte Chemie International Edition* 52(4):1143-1147.
22. Grüschow W, Buchholz TJ, Weufert W, Dordick JS, & Sherman DH (2012) Substrate profile analysis and ACP-mediated acyl transfer in *Streptomyces coelicolor* Type III polyketide synthases. *ChemBioChem* 8(8):863-868.

23. Kohlhaas C*, et al.* (2013) Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. *Chemical Science* 4(8):3212-3217.
24. Wang F, *et al*. (2015) Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase. *ACS Chemical Biology* 10(4):1017-1025.
25. Sundermann U*, et al.* (2013) Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chemical Biology* 8(2):443-450.
26. Del Vecchio F*, et al.* (2003) Active-site residue, domain and module swaps in modular polyketide synthases. *Journal of Industrial Microbiology & Biotechnology* 30(8):489-494.
27. Reeves CD*, et al.* (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* 40(51):15464-15470.