Supporting Information

Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis

Jaclyn M. Winter,[‡,†] Duilio Cascio,[§] David Dietrich,[∥] Michio Sato,[⊥] Kenji Watanabe,[⊥] Michael R. Sawaya,[§] John C. Vederas,[∥] Yi Tang[*,‡,Δ]

Corresponding Author
Yi Tang
Department of Chemical and Biomolecular Engineering
Department of Chemistry and Biochemistry
University of California, Los Angeles, CA 90095
yitang@ucla.edu

Table of Contents

Experimental Procedures

Strains

Chaetomium globosum CBS 148.51 was obtained from The American Type Culture Collection. *Escherichia coli* BL21 (DE3) was used for expressing the standalone SAT domain from CazM. *Saccharomyces cerevisiae* BJ5464-NpgA[1] (*MATα ura3-52 trp1 leu2-Δ1 his3Δ200 pep4::HIS3 prb1Δ1.6R can1 GAL*), which contains a chromosomal copy of the phosphopantetheinyl transferase NpgA,[2] was used for expressing the NR-PKS CazM, the HR-PKS CazF[3] and for the in vivo production of cazaldehyde A (**10**) and B (**12**).

Chemicals and spectral analysis

All solvents and other chemicals used were of analytical grade. NMR spectra were obtained on Bruker 400-MHz (100 MHz for ^{13}C) and 500-MHz spectrometers and a Varian VNMRS 500-MHz spectrometer (125 MHz for ^{13}C). Mass spectra were recorded on an Agilent 6220 oaTOF and a Shimadzu 2010EV liquid chromatography mass spectrometer.

General Molecular Biology Experiments

PCR reactions were carried out using AccuPrime Pfx DNA polymerase (Invitrogen). Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. pCR®-blunt (Invitrogen) was used to construct recombinant DNA products and DNA sequencing was performed by Retrogen and Laragen, California, USA. DNA manipulation using standard techniques[4] were performed in *E. coli* TOP10 (Invitrogen) and *E. coli* XL1-Blue (Stratagene). RNA extraction was performed using a RiboPure Yeast kit (Ambion) and ImProm-II Reverse Transcription System (Promega) was used to synthesize complementary DNA (cDNA) from total RNA using specific primers.

Construction of SAT and CazM mutants

Site-directed mutagenesis on the active site Cys155 and His277 residues was carried out using a complementary PCR primer method. The wild-type SAT-containing pHis8 vector pJWT-49 was used as template with AccuPrime Pfx (Invitrogen) as the polymerase. Primer pairs SAT_C155A_F and SAT_C155A_R were used for the Cys to Ala mutation, primer pairs SAT_C155S_F and SAT_C155S_R were used for the Cys to Ser mutation, and primer pairs SAT_H277A_F and SAT_H277A_R were used for the His to Ala mutation (Table S1). After a denaturing step of 95 ^{o}C for 30 seconds, 14 rounds of 95 ^{o}C for 30 seconds, 55 ^{o}C for one minute and 68 ^{o}C for six minutes was carried out followed by an incubation at 37 ^{o}C for two minutes. The reaction was digested with DpnI (New England Biolabs) and transformed into *E. coli* TOP10 chemical competent cells (Invitrogen). After sequence verification, the expression vectors pJWT-59 (Cys155Ala), pJWT-71 (Cys155Ser) and pJWT-75 (His277Ala) were transformed into *E. coli* BL21 (DE3) electrocompetent cells and protein expression and purification was performed as described above for wild-type SAT (Protein yields can be found in Table S2 and SDS-PAGE analysis in Figure S4).

The SAT domain makes up the majority of piece one used for the reconstitution of CazM (1,194 out of the 1,313 bp). Therefore, to introduce the C155A, C155S and H277A mutations independently into intact CazM, the mutant SAT expression vectors pJWT-59, pJWT-71 and pJWT-75 were used as template to amplify the slightly shorter piece one carrying the respective

point mutation. Primer pair CazM_P1_F was used with CazM_MT_R for all constructs (Table S1). Pieces 2−5 were amplified as described previously and the five exons were co-transformed with the 2 *μM* URA3 expression plasmid into *S. cerevisiae* BJ5464-NpgA as described previously to yield plasmids pJWT-66 (CazM C155A), pJWT-73 (CazM C155S) and pJWT-80 (CazM H277A). Protein expression and purification was performed as described for wild-type CazM. Protein yields are found in Table S2 and SDS-PAGE analysis in Figure S4.

Cazaldehyde reduction assay

To explain if the conversion of **10** into **12** was enzymatic or spontaneous, a variety of in vitro assays were conducted. The first assay assessed whether the reduction occurred non-enzymatically. In a 100 µL reaction, 2 mM NADPH was incubated with 0.2 µM **10** in 100 mM phosphate buffer pH 7.4 at room temperature for 8 hours. After the incubation, the reaction was extracted two times with 200 µL 99% ethyl acetate:1% acetic acid and the organic layer was dried by speedvac. To identify if the SAT domain mediated the reduction, 25 µM of the wild-type standalone protein was incubated at room temperature in a 100 µL reaction containing 0.2 µM **10**, 2 mM NADPH and 100 mM phosphate buffer pH 7.4. The reaction was incubated for 8 hours and then extracted as described previously. The reducing activity of CazM was evaluated by incubating 25 µM of protein in a 100 µL reaction with 2 mM NADPH and 0.2 µM **10** in 100 mM phosphate buffer pH 7.4. The reaction was incubated at room temperature for 8 hours and then extracted as described above. All extracts were dissolved in 20 µL MeOH and analyzed on a Shimadzu 2010 EV LC-MS with a Phenomenex Luna 5µ 2.0 x 100 mM C18 column using positive and negative mode electrospray ionization with a linear gradient of 5–95% MeCN:H_2O over 30 minutes followed by 95% MeCN for 15 minutes and a flow rate of 0.1 mL/min (Figure S17).

Supplementary Results

Supplementary Table 1. *Primers used in this study*

Primer	Sequence shown 5'–3'
CazM_P1_F	TCAACTATCAACTATTAACTATATCGTAATACCAT/ATGATTTCCG TAGCCGACCT
CazM_P1_R	GCTGTCTTGATGGAC
CazM_P2_F	CGACCTACAACAACGAGGCT
CazM_P2_R	TAGCCGACGAAGGTCCATAG
CazM_P3_F	AGGTGTCGCTGTTCGTG
CazM_P3_R	CTTCATCTCGAGCCAGTG
CazM_P4_F	ACACTGGCTCGAGATGAAG
CazM_P4_R	GTTGTGCGTGGCATGGATG
CazM_P5_F	GGTTCGGGAAGATGTATC
CazM_P5_R	TGTCATTTAAATTAGTGATGGTGATGGTGATGCACCGACCTGAG GAATTCCATCT
CazM_seq1	CGGTAATCGACAAGAG
CazM_seq2	CATCTTAGTGAACCAGTG
CazM_seq3	CGGCTACTTCACCGA
CazM_seq4	TCTCCGACGTGATGC
CazM_seq5	GTCCGTGAGCTGTTG
CazM_seq6	GGATTATCTGCGTGAG
CazM_seq7	TGGAGATGCTCAAGACG
CazM_seq8	ACAAGACACGGTCAAGC
CazM_seq9	CGTAACTCCAAGAAGCG
CazM_seq10	GTGGTGCTGTGCTTTG
CazM_seq11	AGTGCGTGTACGAGTTC
SAT_pHis8_F	GACAGCGAGGAATTCATGATTTCCGTAG
SAT_pHis8_R	GACGGACGGCGGCCGCTCAAGGAGACTGCGG
CazF_Trp_F	TTAGTGGTGGTGGTGGTGGTGCGCCCCTGCAGCTATCTCGAC
CazF_Trp_R	AAACTATAAATCGTGAAGGCATGTTTAAACCTAGGTTAGTGGTG GTGGTGGTGGTG
SAT_C155A_F	ACTTTGGGCTTC**GCT**GCCGGGCTACTA
SAT_C155A_R	TAGTAGCCCGGC**AGC**GAAGCCCAAAGT
SAT_C155S_F	ACTTTGGGCTTC**TCT**GCCGGGCTACTA
SAT_C155S_R	TAGTAGCCCGGC**AGA**GAAGCCCAAAGT
SAT_H277A_F	AAGGGCCGCATC**GCT**AGCCCCGACTCG
SAT_H277A_R	CGAGTCGGGGCT**AGC**GATGCGGCCCTT
CazM_MT_R	AGGAGACTGCGGAGCT

EcoRI and NotI restriction recognition sites are italicized and underlined. Point mutations introduced via primers are indicated in bold.

Supplementary Table 2. *List of enzymes constructed in this work*

Enzyme	Plasmid Name	Expression host	Expression Level (mg/L)
CazM	pJWT-37	*S. cerevisiae* BJ5464-NpgA	2.2 mg/L
CazF 2 μMTrp	pJWT-41	*S. cerevisiae* BJ5464-NpgA	8 mg/L
CazM's SAT	pJWT-49	*E. coli* BL21(DE3)	43 mg/L
SAT C155S	pJWT-71	*E. coli* BL21(DE3)	53 mg/L
SAT C155A	pJWT-59	*E. coli* BL21(DE3)	40 mg/L
SAT H277A	pJWT-75	*E. coli* BL21(DE3)	60 mg/L
CazM C155A	pJWT-66	*S. cerevisiae* BJ5464-NpgA	Not solubly expressed
CazM C155S	pJWT-73	*S. cerevisiae* BJ5464-NpgA	<1.0 mg/L
CazM H277A	pJWT-80	*S. cerevisiae* BJ5464-NpgA	2 mg/L

Supplementary Figure 1. *Correct DNA sequence of cazM (CHGG_07645−07647, new GenBank accession KP764718).* After the gene was reconstituted, sequencing revealed that the original genome contained three extra bases. When these extra bases are removed, the correct DNA sequence of *cazM* does not contain any introns. The three extra bases are capitalized, underlined and highlighted in red

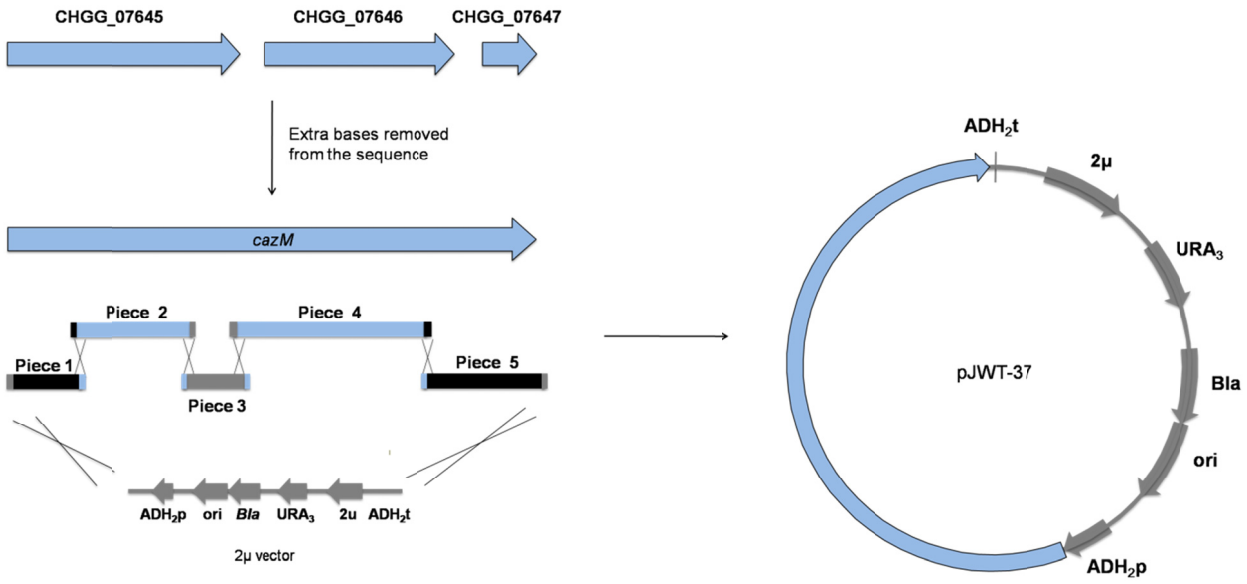
atgatttccgtagccgacctcgattatgcctcccgcaaatccagcattttcctctttgccccccatgtcg
gcactttcaccaagcagtccatggacaagctcgtccggccactcgcggctagcgcccacagggactggat
cctagataccgttgctggtctgccaacatactgggatgccctggcggtcaagatacccaacatagggaac
gcaattccgggccgccgtcaactaactgatctcgacacatggttccggcacggcgcaggcgacgtgacgc
aggacgatgccaccctccccagcatcgtcgtcggcccttggtggtactcatacagctcacccagtactg
gcgctacctcgagctcacacgaccggaccacttagaggactcggcggacttacaagcggacgtggtcacc
cgccagacccagccgggg**A**gccaaggtcgagactttgggcttctgtgccgggctactagccgccgtggct
gtggctagcgccggtaatcgacaagagttccaaaaatatggcgccgtagcggtgcggctagccatgatgg
ccggtgcgctcatcgatggccaagaggcgcgcgacaaggcgacccgcgacggcggctccgtgtcctatgc
catagcatggcgcggccagaagccgggagaggaggctgcgcgcatagtcaaggacctgaacccaaacgcc
tacttcgccgtgctttacgacgaggcccgcgccaccgtgaccacaacccgacgcacggctcccagcctcg
tgaaccggctccgggctgctgatgtcacggtggccgagattgggatcaagggccgcatccacagccccga
ctcggaaaggaagaacaacactgatctactcgtggacctgtgcaagtcattcgaggatctgcaatatgct
gacgccgcctctctcgcgttgccgacctacaacaacgaggctgaaggcaggcctgtgtcgcgggatcgcg
gtaatatgaccgagatggtgatccgcgccatcttagtgaaccagtgcaactggtatgggacgtttaaggg
tgctacggaaggcagggagccctttgtagtcaccttgggcctggagcgtagtgtgccaccgacgctgatg
cgcagtctcggacctcaccaagtgcactacgaggacctggcggacaacggcatccccccagctccgcagt
ctcctcccagggcagtcattgaaacccaaccacagccacaacagcagcaagatacaaccccgccgaaaat
cctcgagaacgacaaagacgccatcgccgtcgtcggcatgtccatcaagacagccggcgccgacgacctg
gacgaattcgtggagatgctcaagacgggccggtcccagcacgagcccgtgacgcgggagcggctcatgc
acgacatgctcttccgggccgccgcggacagcgaccccaaccacaagtactacggctgcttcatccggga
ctcggacgcgttcgaccaccgcttcttcaagcgcagcccgcgcgagtcggccgccatggacccgcagagc
cgcctggccatgcaggccgcctaccaggcggtcgagcagtccggctacttcaccgaaacaaccgccgcgc
ccgagggccgcgacaagaagcacgtgggcgtgtacctgggggtgtgcggtatggactacgaccacaacag
cacgtgccacgagccgagcgccttcacggcgacgggggcgctgcgcagcttcatcacgggccgggtgtcg
cactattttgggtggacgggcccgtccatgaccttcgacacggcctgctcgtcttcggccgtggccatcc

acacggcctgccgcaacctgctctcgggcgagtgcacggcggcgctggccggcggcgccaacaccatcac
cagcatgctttggttccagaacctcagcgcgggcagcttcgtcagcccgacgggccagtgcaagcccttt
gacgacgccgccgacggctactgccgcgccgagggcatggcctttgtgtttttgaagaagctgtccgacg
ccgtgcgcgacggcaaccccgtgctggcaaccatccccagcacggccgtctaccaaaaccagaacagcac
ccccctgttcgtgcccaactccccctccctctcgatgctcttctccgacgtgatgcggcaggccggggtc
gcaccgcgcgacatctcgctggtcgaagcgcacggcacgggcacgcccgtgggcgacccggccgagtacg
agagcatccgcacggcgctgggcgggccgcgcgtgggacgcaccaagccgctgccgatcgggtcggtcaa
gggccacatcgggcacacggagggtgcgtcgggggccatcgcgctgatcaaggtgatcatgatgatgcgc
ggcggcttcatcccgccgcaggccagcttcaccaagatgagccacaacatcgcggtgcggccggacgaca
tgatggaggtggtgacacagctgcgcccctgggccgacccgcacaagatcgccctcctcaacaactacgg
cgcctgcggctccaatgccagcctgatcgtcgcgcagccgcccccggccctgcgcgcaccgacgcgtggc
gcccatcaacaccccgaggggcggcggtatccgttctggatcgcggggctcgacgcccgggcaatttccg
cctacgcggctaagctgatccgctacctcccgtcgtctccccaggacattcccacactcgccgacatctc
ctttagcatgaaccggcagtccaaccgcggcctagcgcaggggttcgtgttcagcagccggtcggtggcc
gagctgcaagagaagctggcgcaggcggccggagcggctgggaaggaggccgcggcgctggtgggcatcg
agccggtcagggccgagcggcccgtggtgctgtgctttggcgggcaggtgtcgctgttcgtggggctgga
cagggcgctgttcgacagcgcggccgtactgcggcggcacctggacgagtgcgatgccgccgtgacggcg
ctcggcctcgacagcatctacccgggcatctttgagcgggagccggtacaagacacggtcaagctgcaga
cgatgctgtttgctatgcagtatgcctctgcgaagagctggatggactgcgggctgcaaggcaaggtggc
cgccgttgtgggccacagcttcggcgagatcacggcgctgtgcgtggccggcgtgcttagcctggaggat
accatcaagcttgtggcggccagggccagactggtgcgtgattcgtggggtcccgatccgggtgccatga
tggctgttgaggccgacgaaggcctcgtccgtgagctgttgcaggaagcgaaccgcagctcggatgggtc
ggccggcattgcttgttacaacggaccgcgcagctttaccctcgccggatcgaccagcgcggtggatgcg
gtgcaggcgaccatgtcggctgacgccaagttcaaagccatcaggagcaagcggctgaacgtgacgaacg
ccttccactccgcccttgtcgagaaactggtggattccctgggacaggtcggaaaggagctgacgttcca
cccacccaccatcccc**C**gtggaacgggcaacggaaacgccctttgacgcggcctcgatggactggacatt
cgtgcctcgccacatgcgggagcccgtattcttcaaccacgccgtgcagcggctggccaagcaacacccg
caagccatctttctcgaagcgggctccaactcgaccatcaccatcatggcaagcagggcgctcgctcaga
gccagacctcccctccggacgcccatcatttccaagccgtctccatcaccaccgcgacgggcttcgacgg
gctgaccgacgccacggtcgccctgtggaagcaaggtctgcgcgtcgccttctgggcccaccacgccctg
cagacgctcgaatacgcgcagctgctcctacccccctaccaattcgacacatccccctcctcgcgacact
ggctcgagatgaagtcaccccagaaggtgatcgacaaagccgtaaaggcactcgtagcggcaggaacacc
gcaccaactgccccagaaccaaaccatcgaccccaaaaccctcccccctatggaccttcgtcggctaccaa
gacgcgccgaccaacaaaacggcgcgcttccgcatcaacaccgagtccgcccagtacaacacctacgtgc
tcagccacgtgatcgcccaaaccgcgcccatctgcccgggcacgctcgaatgcgacatgatgatcgaggc
gctgttcagtctgcacgcggacgcccaggaaaacggtctgcagcccgagatccgcgacctggtcaaccac
acgcccatctgcgtcgacccctcgcgcaccgtctacctggagctgacggcgctgtacaagaagcgcacgc
agtggagcgcgcatttcctcagcgtcgccaacggcggcggcgatgcccataacccgcagacgcacgccga
ggcgcggctgcggctgcgcgcggcgggcgatgcggattatctgcgtgagtttgcgcagtttgagcggctg
gttagtcatgggcggtgtcgcgagttgcttgggttggggctggatgtggaggggggtggaggtgctgcagg
gtcgtggggtgtaccgggcgtttgatccggtggttgagtatggtgaggtgtaccgtggcgtgcgctacct
ggtgggccgcggcaacgagtgtgccggccatgtgcagctgcccgcgcggcaccgtggggatacgtggttg
gatgtcgcgttgagtgacagtttcagccaggtgggtgggttgtgggtgaatttgatgacggatgtaccgc
cgggggatatgtacattgcgacggggtgtgaggtgacgatgcgcgccgcgggccacggcgcggggtga
gggggatgtgtggcatgtgtatggccgccatgcgcgccagagcgacaaggcgtacacgacggatctgttt

gtgtttgatgccgcgaccgggtcgctcgttgaggtcatgttgggcctgcagtatgggcgggtggccaagg
cctcgatgagccggatgctggcgcgcatgactacggatgaaacggtggtgcggaccaaggcatcgttacc
gcagctgccggcgacgaattcttcccctccagctaccgctccggcccccgccatggctgttgtcgaggcg
agccgtaactccaagaagcggagcaagaagaccgaaaagaaggaggccaagcccaaaaagacacaaaagg
attccggatggcgagacattaccgaagaggtgcgcaaccttgtggctcacgtgtccggcattgaagccag
cgagatcgggctcgactcggaaatggccgactttggtatcgattccctgatgggaatggagctgggccgg
gaagtcgagttgacgttcaagtgcaagctggaccaggccgaacagatggaggccacgagtctgcgcaagt
ttgtggcggttgtggccaaggctctctttggtactgatcagcccgctgaggtggaggacgaagccagtga
agctgaagacgaagacgatgatgattcatcctccactggtgaaggagggacatggtcggagccaagccaa
gaagacggaagcaaagagtcctctggtttgcagacgccggattcgagcagctacacccccccgccggaga
aagcccagaaggccgggcccgtgccaggcaacctaaagcccctaaacgcaccaacggcgccgccagccgt
gagcaacctggtactatcgcccgccgacgtgctcgcagccttcggcgaggtgaagatggccaccgacagc
ctcatgcgggaatacgagatcgacaagacggaaaggaccctccttgcgggcagcaaccgtctctgtgcgg
cgctggtggtggaggccctcgacgagctaggctgccccatcggcaccgccgcggcgggccagccactcga
acgggtgccgttcctcccgcagcacggccgcctgatgcagtgcgtgtacgagttcctcgagcgcgacgcg
cggctcatcgacatcgacgtcgcgagcggccagctgacgcgcacgcacgtggcggcgccgcgcaagacga
gccaggccatcctcgaggacctgctgaccagccagcccgagttcgccgtggccaaccggctcgcctacca
cgcgggcaagcagctggcgggggtgctcagcggcaagacggacggcatccgcgtgctcttcggcagcccc
gagggccgcgaggtgacggcggccatgtactgcgagcacccgttcaaccgcatgagctaccggcagatga
gcgacgtggtggagggcctggcggagcgcatccaaagccgcggcggcacgggcgagacgttcaaggtgct
cgagatgggcgccggcacgggaggcaccacgctcgtcatggcgccgctgttggcatcgctcgaagcgcgc
ggcatcatgcgggtcgagtacacttttactgacctgtcgccgtcgctggtcgccaatgctcgtcggcggt
tcgggaagatgtatcccttcatgcggttcgcagtgcatgacatcgagaagtcgccggccgaggagttgag
gggccagcacctggtacttgccagcaatgccatccatgccacgcacaacctggttgtctcggcctcgaat
gtgcgccaggccctgcggccggacggcttcgtaatgattctagagatgaccgaggtggtccccttcatcg
acctggtctttgggctgctcgaaggttggtggctattcgatgacggccggaatcacgcagttgtgccggc
cgagcactgggaacgtgagctgcacactgcaggttttgggcatgttgactggaccgacgggaacttgccc
gagaacgcggtcc**A**aaaaggtcattttcgcgatggcttcagaaccccagggtcctcgcctaccaaaggcc
gccccggagcaaacgatggaactcgaccgaggtgatgtggccgctcgcacggccgaggcggaacggcttg
tgacaaaatactcgaacgcctgggccacgccgagattgcaagccctcgccgcccggcgggaaacaaacca
aggccagccgaagcgcaacaaagcaaatctcggcgccgtggttctcgtgaccggcgcgaccggcagcctc
ggttcgcacctggtgcagaagctggccgagaaccccgaagtcgcgcaggtggtgtgcctcaaccgtcgca
gcagcagcggcctctccgccgtcaagaaacaacaggaggcattcgcatccaggggcatcacgctgtcacc
cggggcacgcgcgaagttgcgggtcctagacgccgacacctcccaggcccagctggccctcccaccgcac
gagtacacctggctcctgcaacacggcacccacatcgtgcacaacgcgtggcccatgagcggaacgcgcc
ccatcagcgccttcgagccccagctgcagggtatgaggaacctcctcgacctcgcgcgggagatggccac
ccgcgatgccgtccacccgttccgcgtcggcttccagtttgtgtcgtccatcggggtggtgggctacgcc
ggggagccgcgcatcctcgaggaccgcgtgcctctttcggccgtcctgcccagcggttacggcgagggca
aatgggtgtgtgagcggctgctggacgacaccctgcataagtaccccagccttttcaggcccatggtcgt
gcggcccggccagatctcgggatcgtcgaacagcggtttctggaatcctgtggagcactttgcgttcctg
gtaaagtcggcgcagaccctgcgggcgtggcccgacttcgacggcgtgctgcagtggattccggtggacc
actgcgcgagcgtcatggcggatttgttgaagatgggggtcgcggacgcgccggatgcctatccggtgta
tcacatcgacaaccccgtaggtcagcagtggaaggccatgtcgccggtgttggcggctgccctcgacatc
ccaccgcacgcgatcatccccttttcgagctggatcaagagggtccgccggtcgccacttcttgccgaga
cggagaatccggcggcgcgtctcgttgacttcctcgacagccacttcgagcgcatgtcttgcggtggcct

gatactagatacgaccaaagcgaaggagcactcccagaccatggccaaggagggtcccgtcgacccagac
ctcgcccgtctatatgtcgcggcttggaagaagatggaattcctcaggtcgtaa

Supplementary Figure2. *Reconstitution of CazM for expression in S. cerevisiae* BJ5464-NpgA. The 8,238bp intron-free gene was assembled and inserted into the 2μ URA3 vector by *in vivo* homologous recombination. Overlapping regions between two flanking segments of CazM ranged from 122-382 bp and overlapping regions between piece 1 and piece 5 with the 2μ vector were 35 bp.

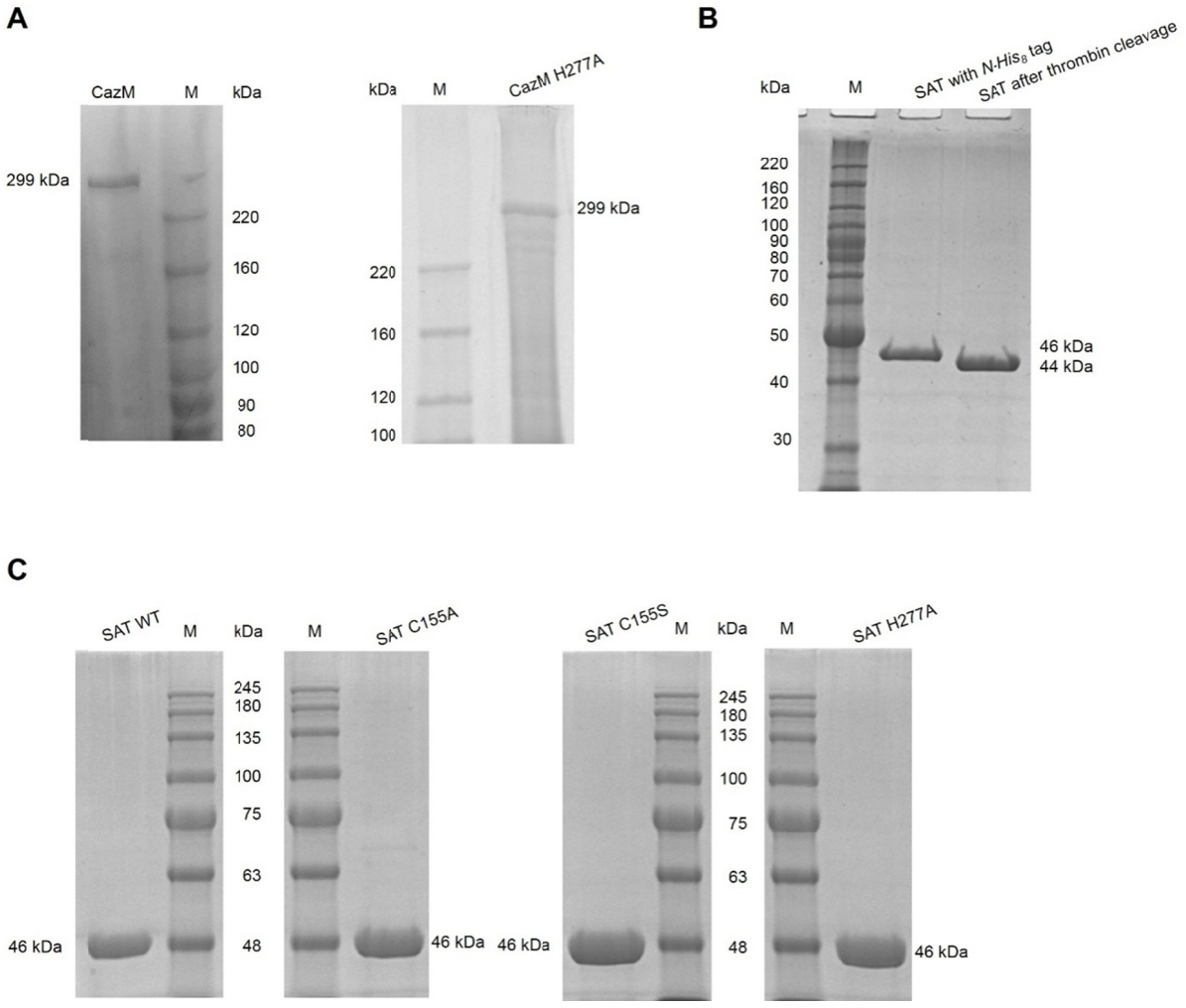


Supplementary Figure 3. *Amino acid sequence of CazM.* The SAT domain with part of the KS-linker is highlighted in gray. The GxCxG binding motif is underlined.

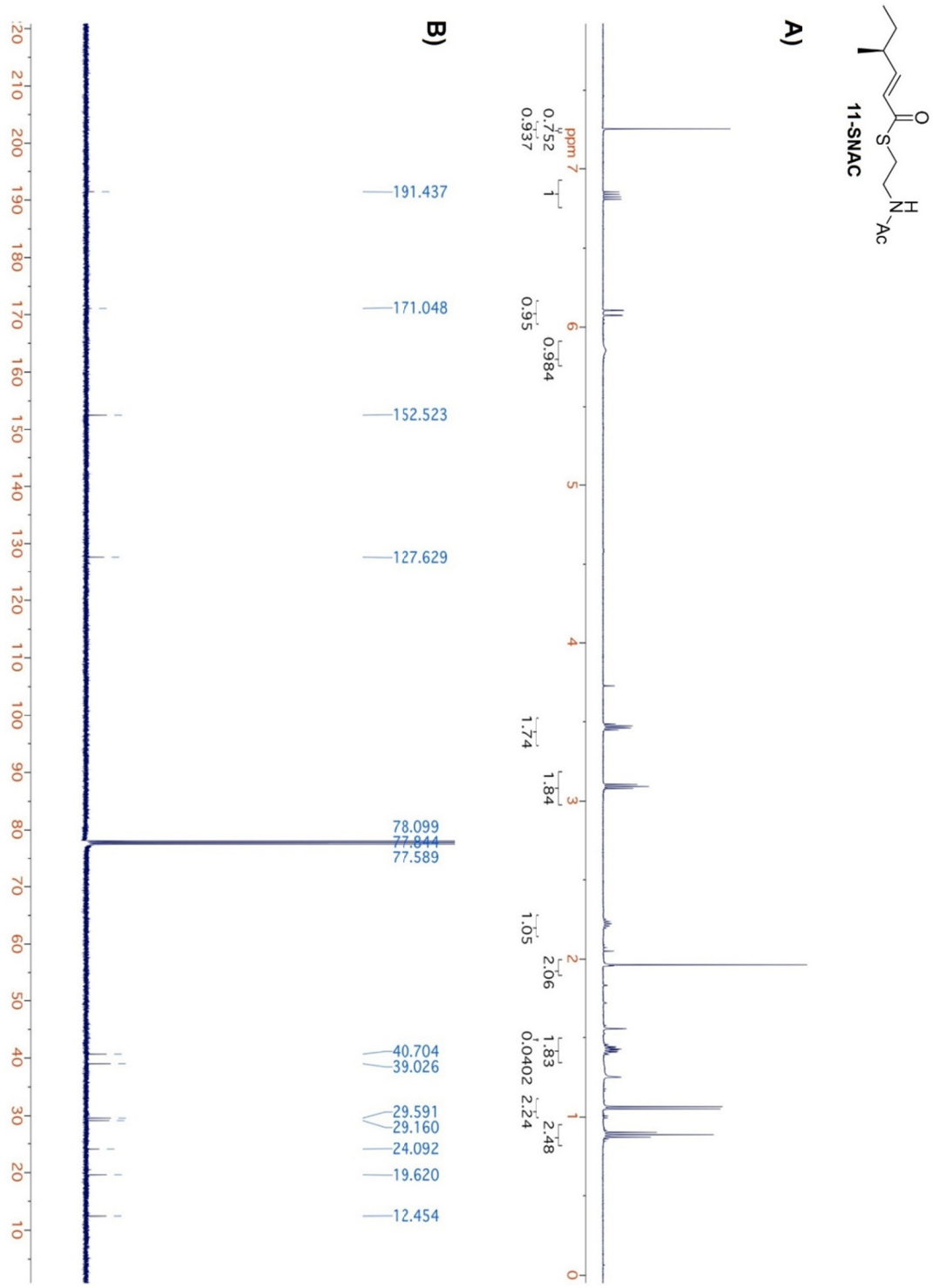
MISVADLDYASRKSSIFLFAPHVGTFTKQSMDKLVRPLAASAHRDWILDTVAGLPTYWDALAVKIPNIGN
AIPGRRQLTDLDTWFRHGAGDVTQDDATLPSIVVGPLVVLIQLTQYWRYLELTRPDHLEDSADLQADVVT
RQTQPGAKVETLGFCAGLLAAVAVASAGNRQEFQKYGAVAVRLAMMAGALIDGQEARDKATRDGGSVSYA
IAWRGQKPGEEAARIVKDLNPNAYFAVLYDEARATVTTTRRTAPSLVNRLRAADVTVAEIGIKGRIHSPD
SERKNNTDLLVDLCKSFEDLQYADAASLALPTYNNEAEGRPVSRDRGNMTEMVIRAILVNQCNWYGTFKG
ATEGREPFVVTLGLERSVPPTLMRSLGPHQVHYEDLADNGIPPAPQSPPRAVIETQPQPQQQQDTTPPKI
LENDKDAIAVVGMSIKTAGADDLDEFVEMLKTGRSQHEPVTRERLMHDMLFRAAADSDPNHKYYGCFIRD
SDAFDHRFFKRSPRESAAMDPQSRLAMQAAYQAVEQSGYFTETTAAPEGRDKKHVGVYLGVCGMDYDHNS
TCHEPSAFTATGALRSFITGRVSHYFGWTGPSMTFDTACSSSAVAIHTACRNLLSGECTAALAGGANTIT
SMLWFQNLSAGSFVSPTGQCKPFDDAADGYCRAEGMAFVFLKKLSDAVRDGNPVLATIPSTAVYQNQNST
PLFVPNSPSLSMLFSDVMRQAGVAPRDISLVEAHGTGTPVGDPAEYESIRTALGGPRVGRTKPLPIGSVK
GHIGHTEGASGAIALIKVIMMMRGGFIPPQASFTKMSHNIAVRPDDMMEVVTQLRPWADPHKIALLNNYG
ACGSNASLIVAQPPPALRAPTRGAHQHPEGRRYPFWIAGLDARAISAYAAKLIRYLPSSPQDIPTLADIS

FSMNRQSNRGLAQGFVFSSRSVAELQEKLAQAAGAAGKEAAALVGIEPVRAERPVVLCFGGQVSLFVGLD
RALFDSAAVLRRHLDECDAAVTALGLDSIYPGIFEREPVQDTVKLQTMLFAMQYASAKSWMDCGLQGKVA
AVVGHSFGEITALCVAGVLSLEDTIKLVAARARLVRDSWGPDPGAMMAVEADEGLVRELLQEANRSSDGS
AGIACYNGPRSFTLAGSTSAVDAVQATMSADAKFKAIRSKRLNVTNAFHSALVEKLVDSLGQVGKELTFH
PPTIPVERATETPFDAASMDWTFVPRHMREPVFFNHAVQRLAKQHPQAIFLEAGSNSTITIMASRALAQS
QTSPPDAHHFQAVSITTATGFDGLTDATVALWKQGLRVAFWAHHALQTLEYAQLLLPPYQFDTSPSSRHW
LEMKSPQKVIDKAVKALVAAGTPHQLPQNQTIDPKTLPLWTFVGYQDAPTNKTARFRINTESAQYNTYVL
SHVIAQTAPICPGTLECDMMIEALFSLHADAQENGLQPEIRDLVNHTPICVDPSRTVYLELTALYKKRTQ
WSAHFLSVANGGGDAHNPQTHAEARLRLRAAGDADYLREFAQFERLVSHGRCRELLGLGLDVEGVEVLQG
RGVYRAFDPVVEYGEVYRGVRYLVGRGNECAGHVQLPARHRGDTWLDVALSDSFSQVGGLWVNLMTDVPP
GDMYIATGCEVTMRAPRATARGEGDVWHVYGRHARQSDKAYTTDLFVFDAATGSLVEVMLGLQYGRVAKA
SMSRMLARMTTDETVVRTKASLPQLPATNSSPPATAPAPAMAVVEASRNSKKRSKKTEKKEAKPKKTQKD
SGWRDITEEVRNLVAHVSGIEASEIGLDSEMADFGIDSLMGMELGREVELTFKCKLDQAEQMEATSLRKF
VAVVAKALFGTDQPAEVEDEASEAEDEDDDDSSSTGEGGTWSEPSQEDGSKESSGLQTPDSSSYTPPPEK
AQKAGPVPGNLKPLNAPTAPPAVSNLVLSPADVLAAFGEVKMATDSLMREYEIDKTERTLLAGSNRLCAA
LVVEALDELGCPIGTAAAGQPLERVPFLPQHGRLMQCVYEFLERDARLIDIDVASGQLTRTHVAAPRKTS
QAILEDLLTSQPEFAVANRLAYHAGKQLAGVLSGKTDGIRVLFGSPEGREVTAAMYCEHPFNRMSYRQMS
DVVEGLAERIQSRGGTGETFKVLEMGAGTGGTTLVMAPLLASLEARGIMRVEYTFTDLSPSLVANARRRF
GKMYPFMRFAVHDIEKSPAEELRGQHLVLASNAIHATHNLVVSASNVRQALRPDGFVMILEMTEVVPFID
LVFGLLEGWWLFDDGRNHAVVPAEHWERELHTAGFGHVDWTDGNLPENAVQKVIFAMASEPQGPRLPKAA
PEQTMELDRGDVAARTAEAERLVTKYSNAWATPRLQALAARRETNQGQPKRNKANLGAVVLVTGATGSLG
SHLVQKLAENPEVAQVVCLNRRSSSGLSAVKKQQEAFASRGITLSPGARAKLRVLDADTSQAQLALPPHE
YTWLLQHGTHIVHNAWPMSGTRPISAFEPQLQGMRNLLDLAREMATRDAVHPFRVGFQFVSSIGVVGYAG
EPRILEDRVPLSAVLPSGYGEGKWVCERLLDDTLHKYPSLFRPMVVRPGQISGSSNSGFWNPVEHFAFLV
KSAQTLRAWPDFDGVLQWIPVDHCASVMADLLKMGVADAPDAYPVYHIDNPVGQQWKAMSPVLAAALDIP
PHAIIPFSSWIKRVRRSPLLAETENPAARLVDFLDSHFERMSCGGLILDTTKAKEHSQTMAKEGPVDPDL
ARLYVAAWKKMEFLRS

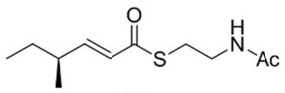
Supplementary Figure 4. *SDS-PAGE of recombinant caz proteins expressed in this study*. A) CazM and CazM H277A (both 299 kDa) were expressed as C-terminal hexahistidyl-tagged proteins in *S. cerevisiae* BJ5464-NpgA, purified by Ni-NTA agarose affinity resin at 2.2 mg/L and 2.0 mg/L, respectively, and analyzed for purity using a 6% SDS-PAGE gel. B) CazM's SAT domain used for crystallization studies was expressed as an N-terminal octahistidyl-tagged protein (46 kDa) in *E. coli* BL21 (DE3) cells. The His-tag was removed via thrombin cleavage to afford the 44 kDa protein, which was analyzed for purity using a 12% SDS-PAGE gel. C) For in vitro assays, the standalone wild-type SAT domain and mutants were expressed as N-terminal octahistidyl-tagged proteins (46 kDa) in *E. coli* BL21(DE3) cells, purified by Ni-NTA agarose affinity resin to yield between 40-60 mg/L and analyzed for purity using a 12% SDS-PAGE gel. In panels A and B, BenchMark ladder (Life Technologies) was used, whereas Bluestain protein ladder 11-245 kDa (GoldBio) was used for all gels in panel C.



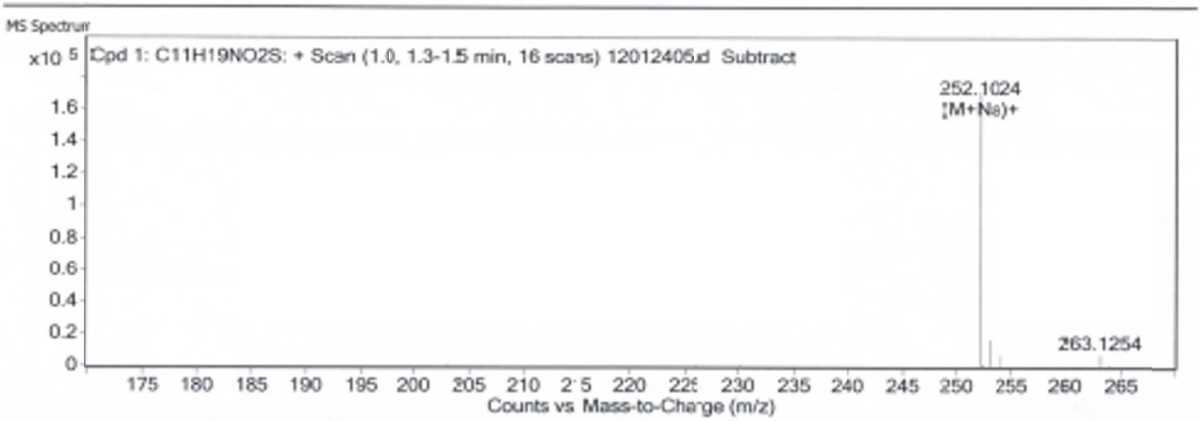
Supplementary Figure 5. *A) ^{1}H NMR spectrum (500 MHz) and B) ^{13}C NMR spectrum (125 MHz) of SNAC ester (**11**) in $CDCl_3$.*



Supplementary Figure 6. *Mass spectrum of SNAC ester (**11**).*



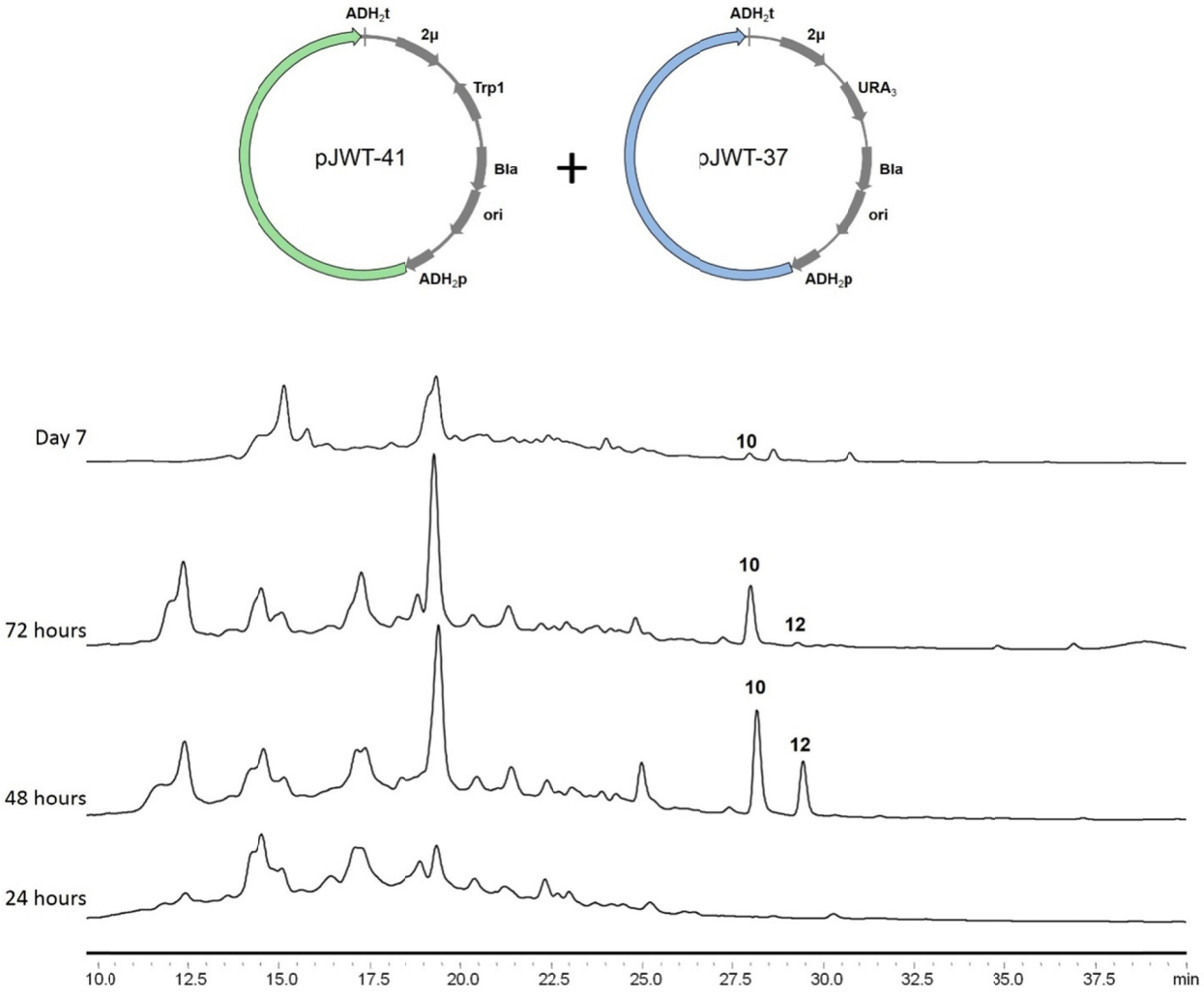
11



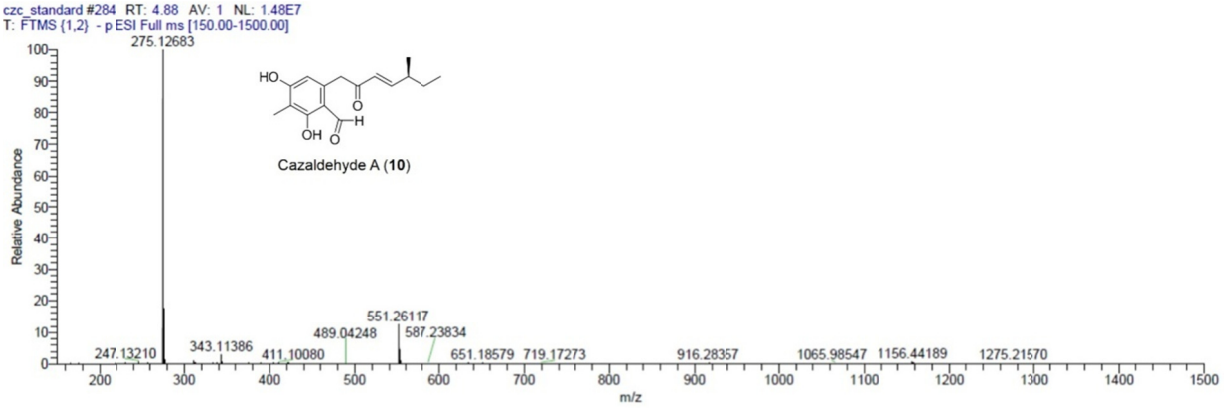
MS Spectrum Peak List

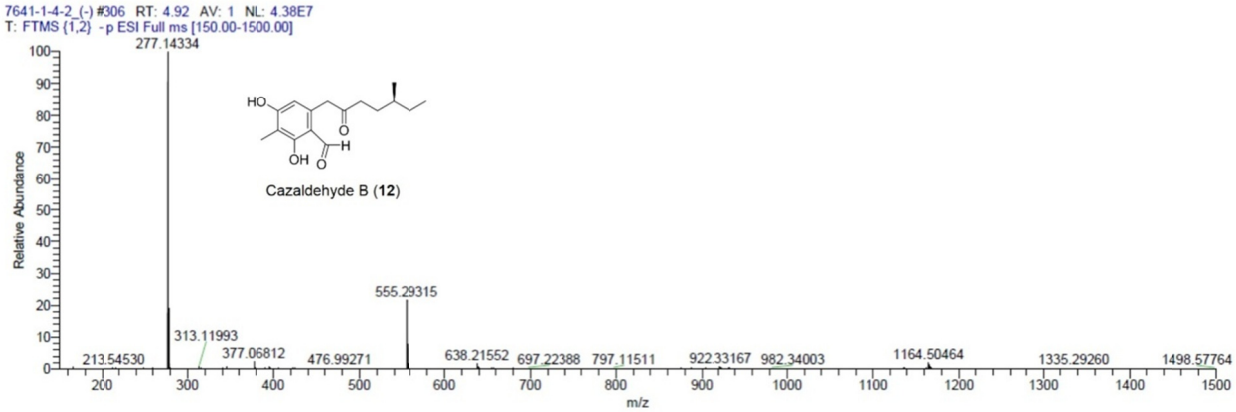
m/z	Calc m/z	Diff(ppm)	z	Abund	Formula	Ion
252.1024	252.1029	-1.76	1	172278	C11 H19 N Na O2 S	(M+Na)+

Supplementary Figure 7. *In vivo synthesis of cazaldehyde A (**10**) and B (**12**) in S. cerevisiae* BJ5464-NpgA. HPLC traces are shown in the same scale and at λ = 290 nm.

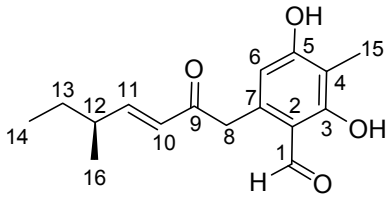


Supplementary Figure 8. *Mass spectra data of* ***10*** *and* ***12***





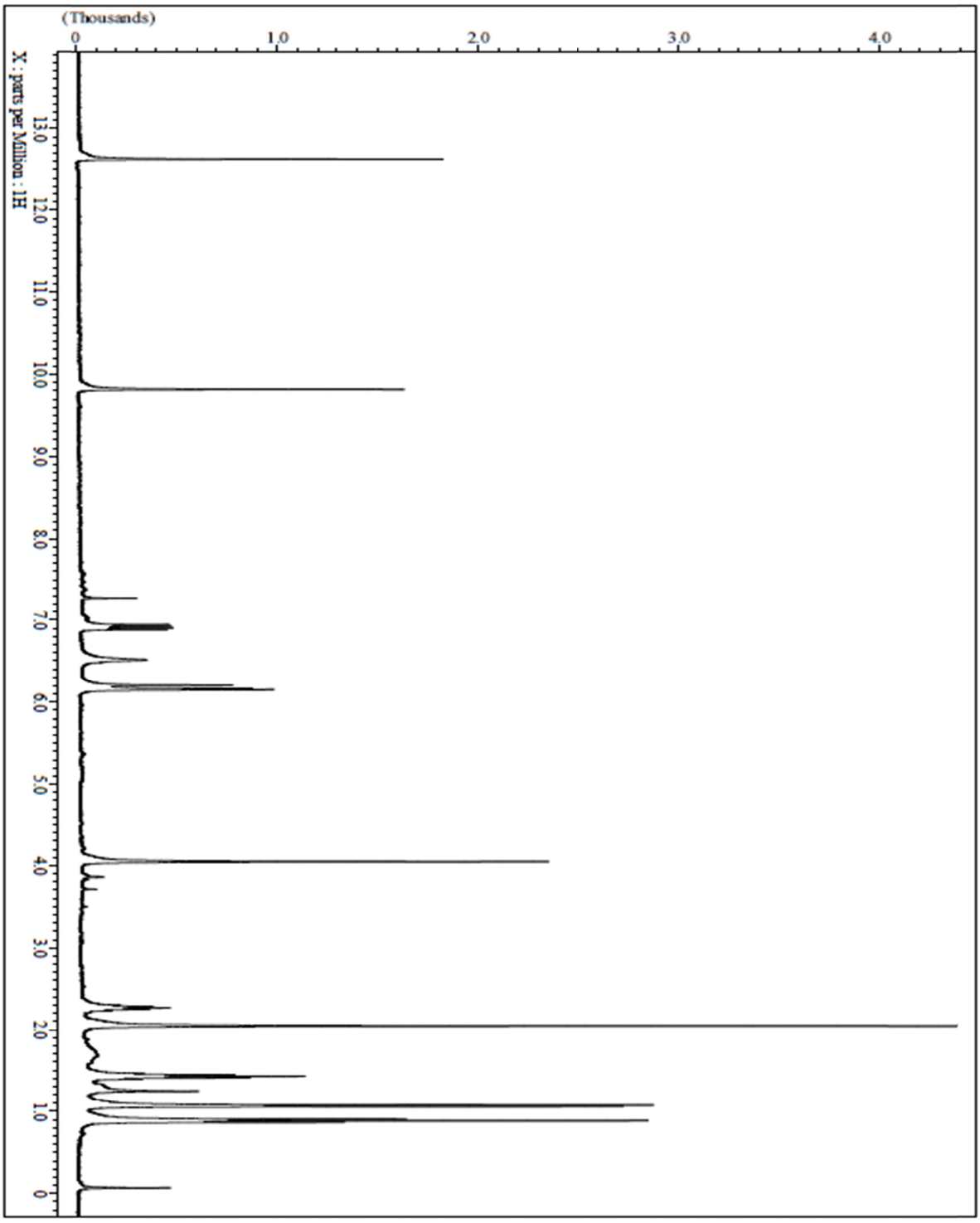
Supplementary Table 3. *Assignments of ^{13}C and ^{1}H NMR data for cazaldehyde A (**10**)*



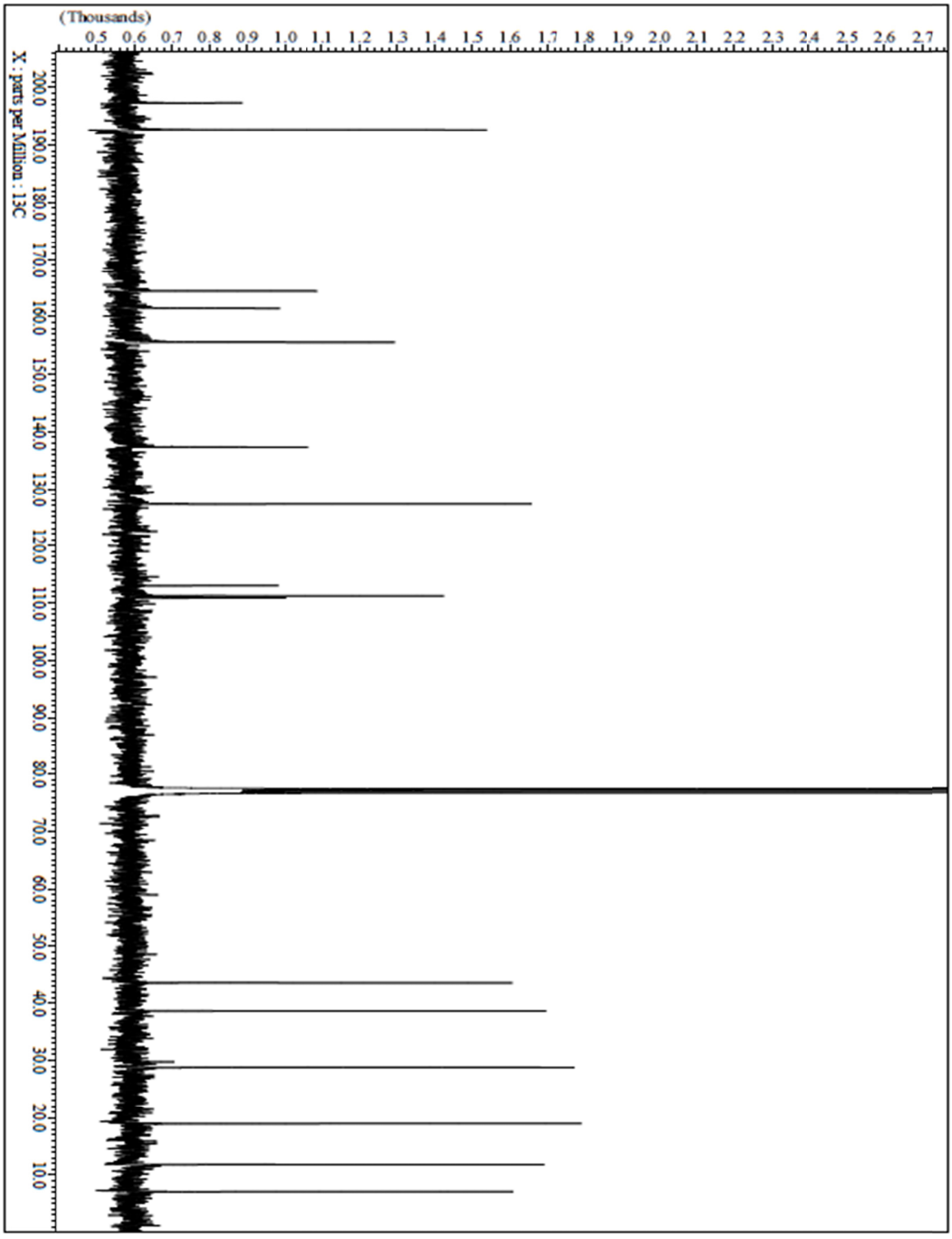
Position[a]	δC ppm[b],	δH (ppm)[c] (J, Hz)	^{1}H-^{1}H COSY	^{1}H-^{13}C HMBC
1	192.7	9.82 (s)		2, 3
2	113			
3	161.5			
4	111.2			
5	164.5			
6	110.8	6.15 (s)		4, 8
7	137.2			
8	43.6	4.05 (s)		6, 7, 9
9	197.4			
10	127.3	6.2 (d, 16.0)	11	9, 12
11	155.6	6.94 (d.d. 16.4, 8.4)	10, 12	9, 12, 13, 16
12	38.7	2.28 (m)	11, 13, 16	11
13	28.9	1.45 (m)	12, 14	15
14	11.8	0.9 (t, 7.2)	13	12, 13
15	7.1	2.04 (s)		3, 4
16	19	1.09 (d, 7.2)	12	11, 12, 13
3-OH		12.7 (s)		3,4
5-OH		6.5 (br, s)		

[a] The numbering scheme for **10** was adopted from reference[3]; [b] ^{13}C NMR spectral data was taken at 100 MHz in $CDCl_3$; [c] ^{1}H NMR spectral data was taken at 400 MHz in $CDCl_3$.

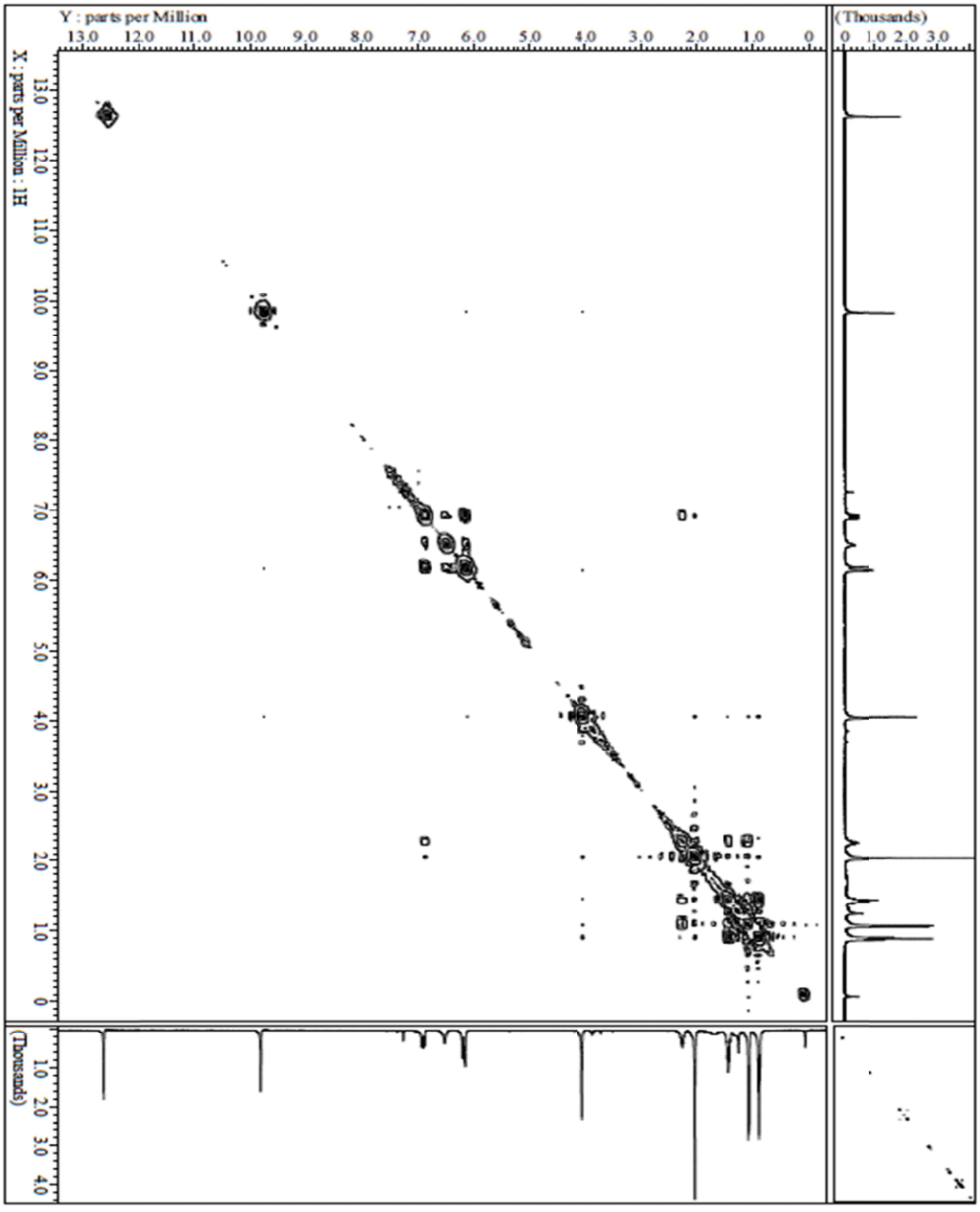
Supplementary Figure 9. *^{1}H NMR spectrum (400 MHz) of cazaldehyde A (**10**) in $CDCl_3$*



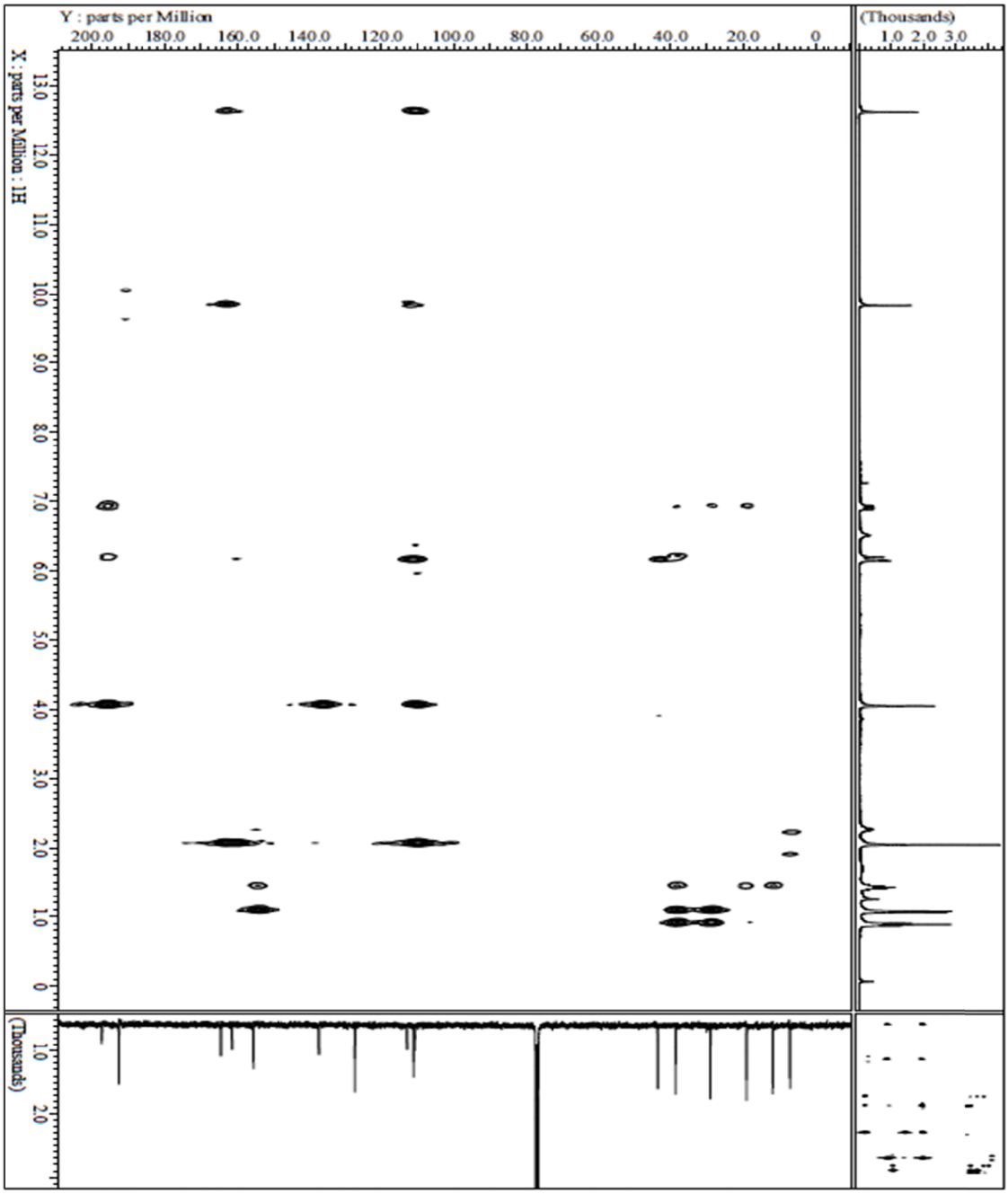
Supplementary Figure 10. *^{13}C NMR spectrum (100 MHz) of cazaldehyde A (**10**) in $CDCl_3$*



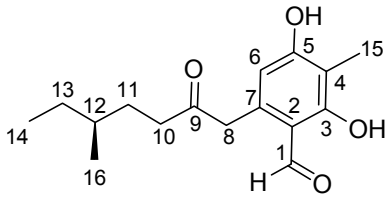
Supplementary Figure 11. *^{1}H-^{1}H COSY NMR spectrum (400 MHz) of cazaldehyde A (**10**) in $CDCl_3$*



Supplementary Figure 12. *^{1}H-^{13}C HMBC NMR spectrum (400 MHz) of cazaldehyde A (**10**) in $CDCl_3$*



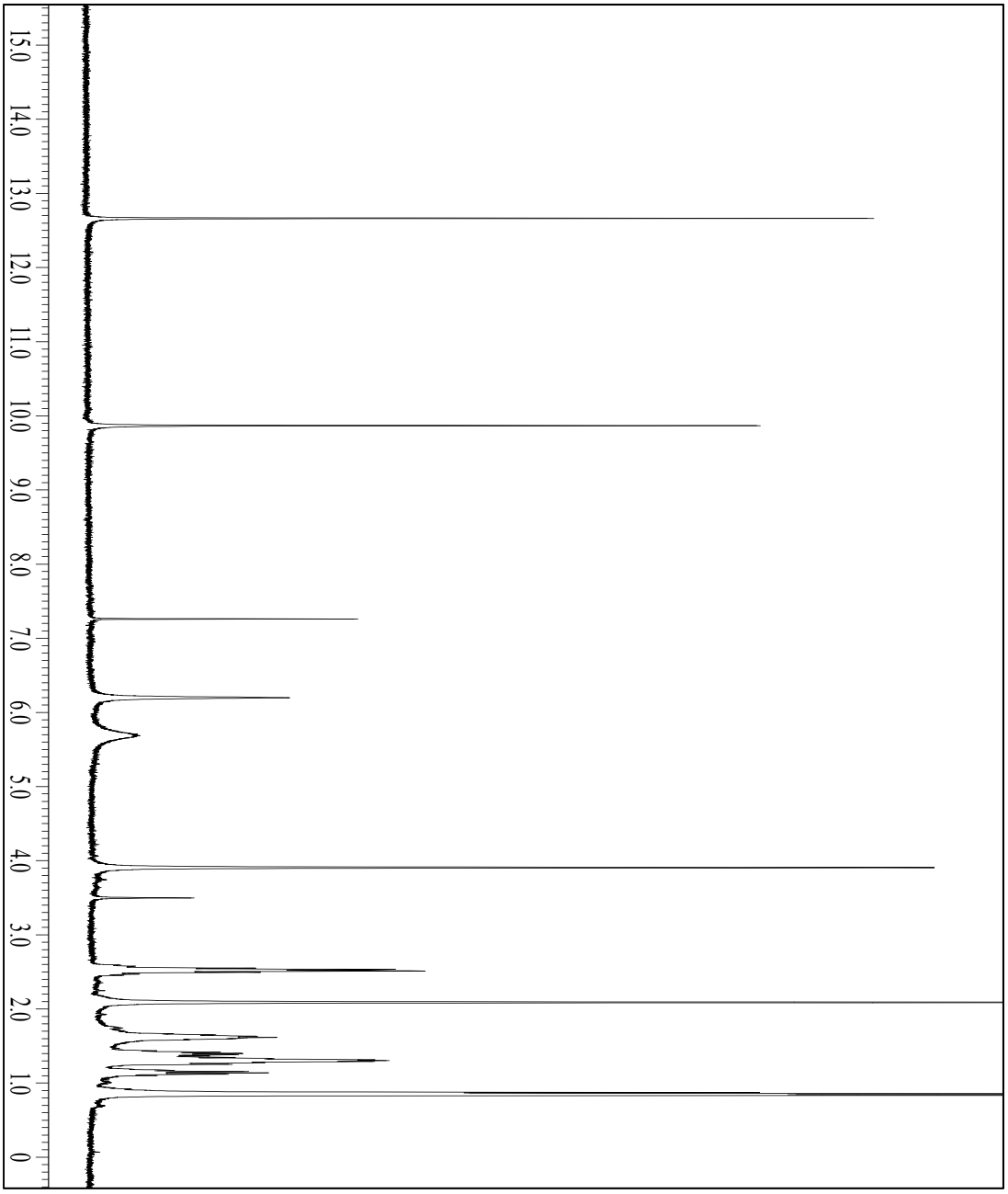
Supplementary Table 4. *Assignments of ^{13}C and ^{1}H NMR data for cazaldehydeB (**12**)*



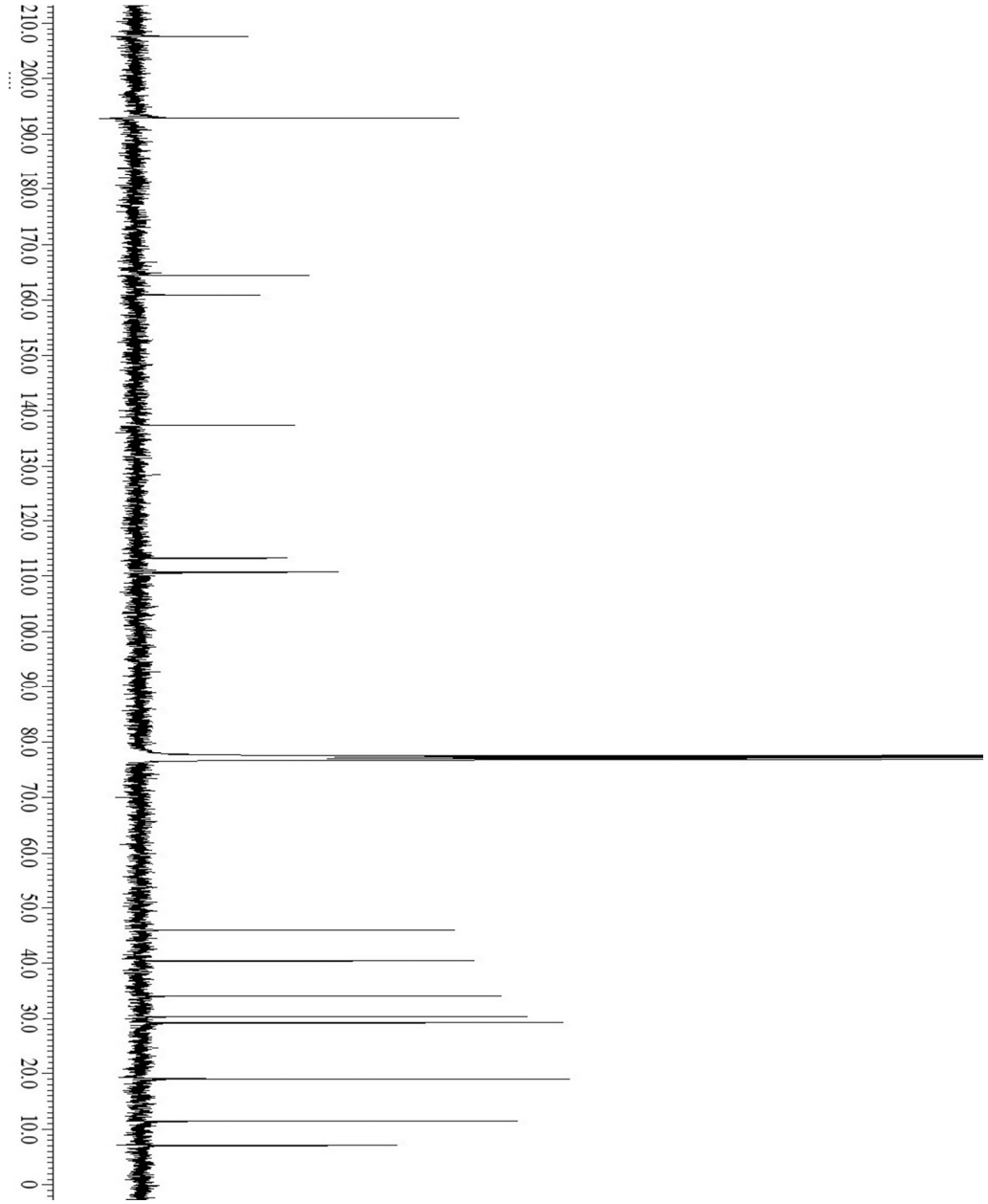
Position[a]	δCppm[b],	δH(ppm)[c] (J, Hz)	^{1}H-^{1}H COSY	^{1}H-^{13}C HMBC
1	192.9	9.87 (s)		3, 7
2	137.3			
3	164.5			
4	110.6			
5	161.0			
6	110.7	6.20 (s)		7, 8
7	113.2			
8	46.0	3.91 (s)		2, 6, 9
9	207.6			
10	40.5	2.52 (d,d, 15.4, 7.3)	11	9, 11, 12
11	30.4	1.62 (m)	10, 12	10, 12, 13, 16
12	34.1	1.30 (m) 1.33 (m)	11, 13, 16	10, 11, 13, 16
13	29.4	1.40 (m)	12, 14	11, 12, 14, 16
14	11.4	1.14 (m) 0.86 (t, 7.6)	13	12
15	7.10	2.09 (s)		3, 4, 5
16	19.1	0.84 (d, 6.1)	12	11, 12
3-OH		12.7 (s)		3,4
5-OH		5.7 (br s)		

[a] The numbering scheme for **12** was adopted from reference[3]; [b] ^{13}C NMR spectral data was taken at 100 MHz in $CDCl_3$; [c] ^{1}H NMR spectral data was taken at 400 MHz in $CDCl_3$.

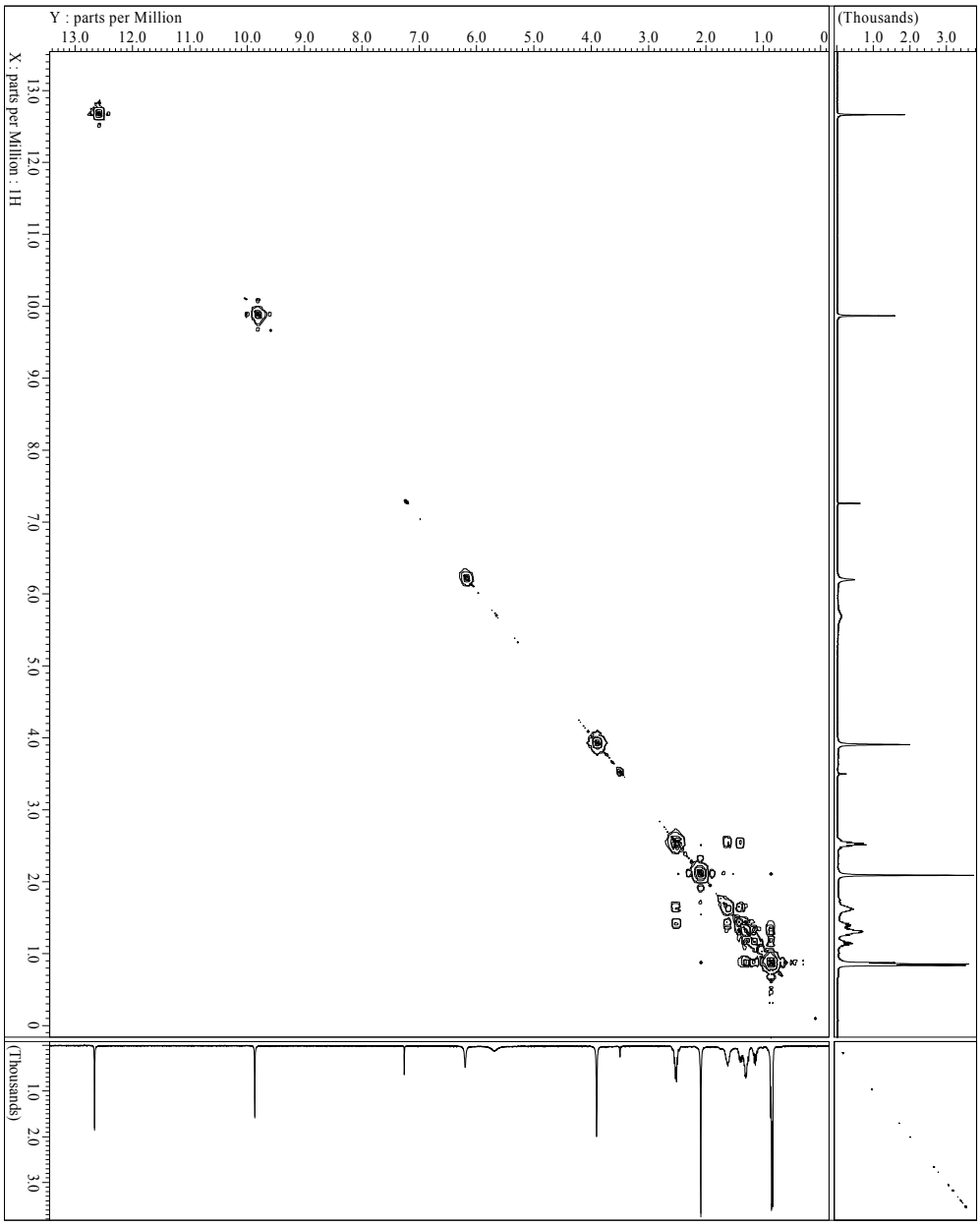
Supplementary Figure 13. *^{1}H NMR spectrum (400 MHz) of cazaldehyde B (**12**) in $CDCl_3$*



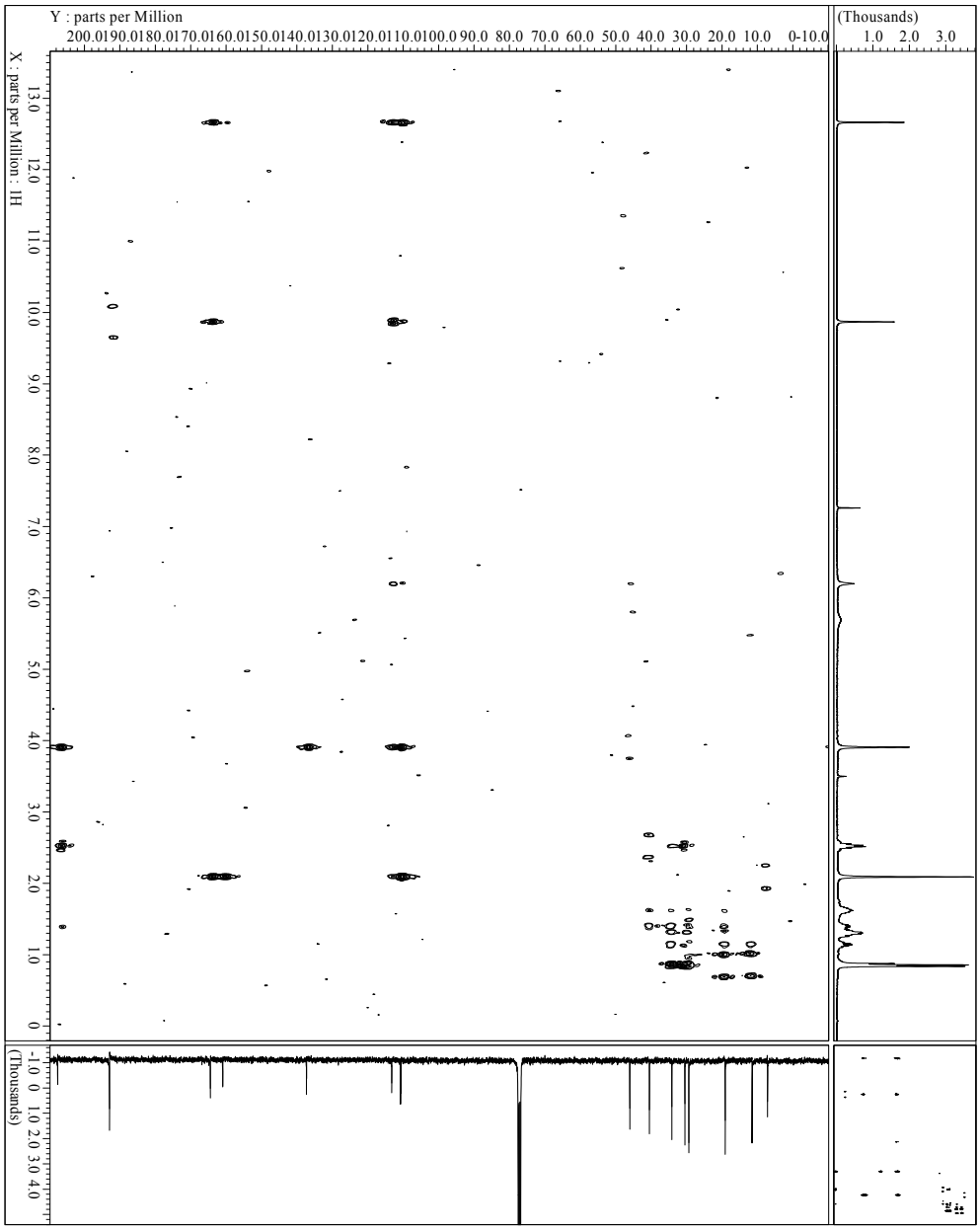
Supplementary Figure 14. *^{13}C NMR spectrum (100 MHz) of cazaldehydeB (**12**) in $CDCl_3$*



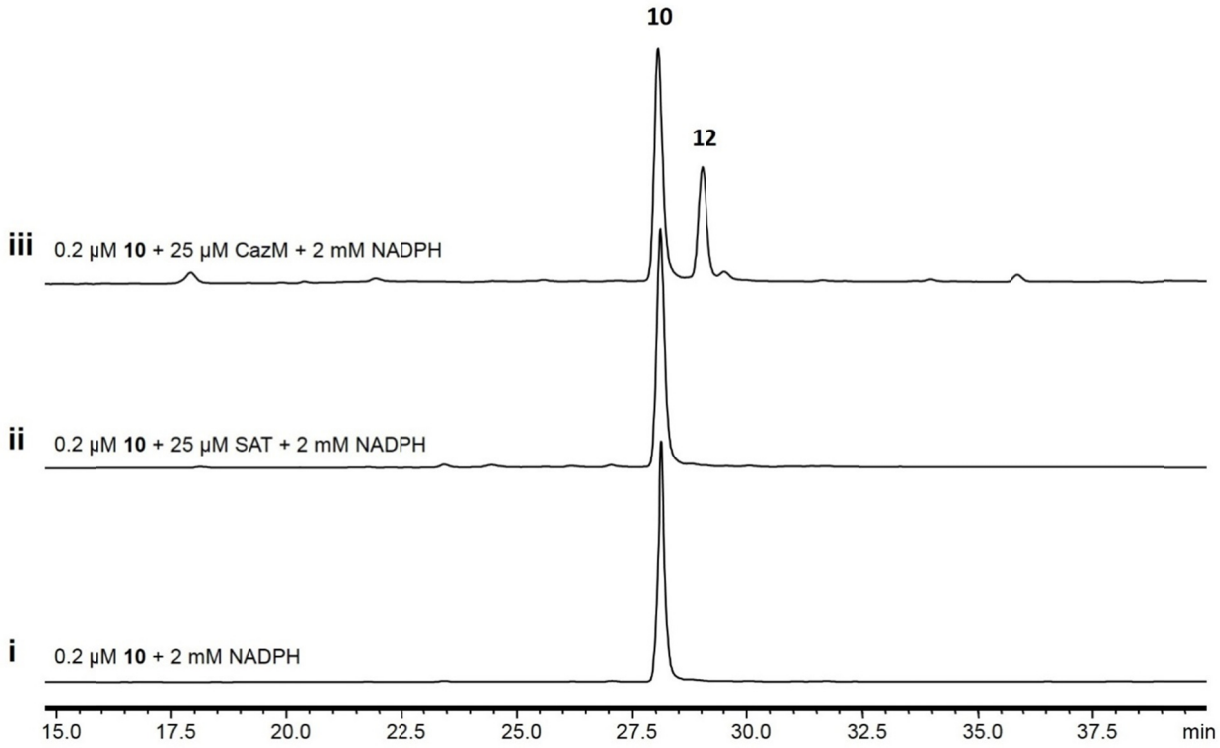
Supplementary Figure 15. *^{1}H-^{1}H COSY NMR spectrum (400 MHz) of cazaldehyde B (**12**) in $CDCl_3$*



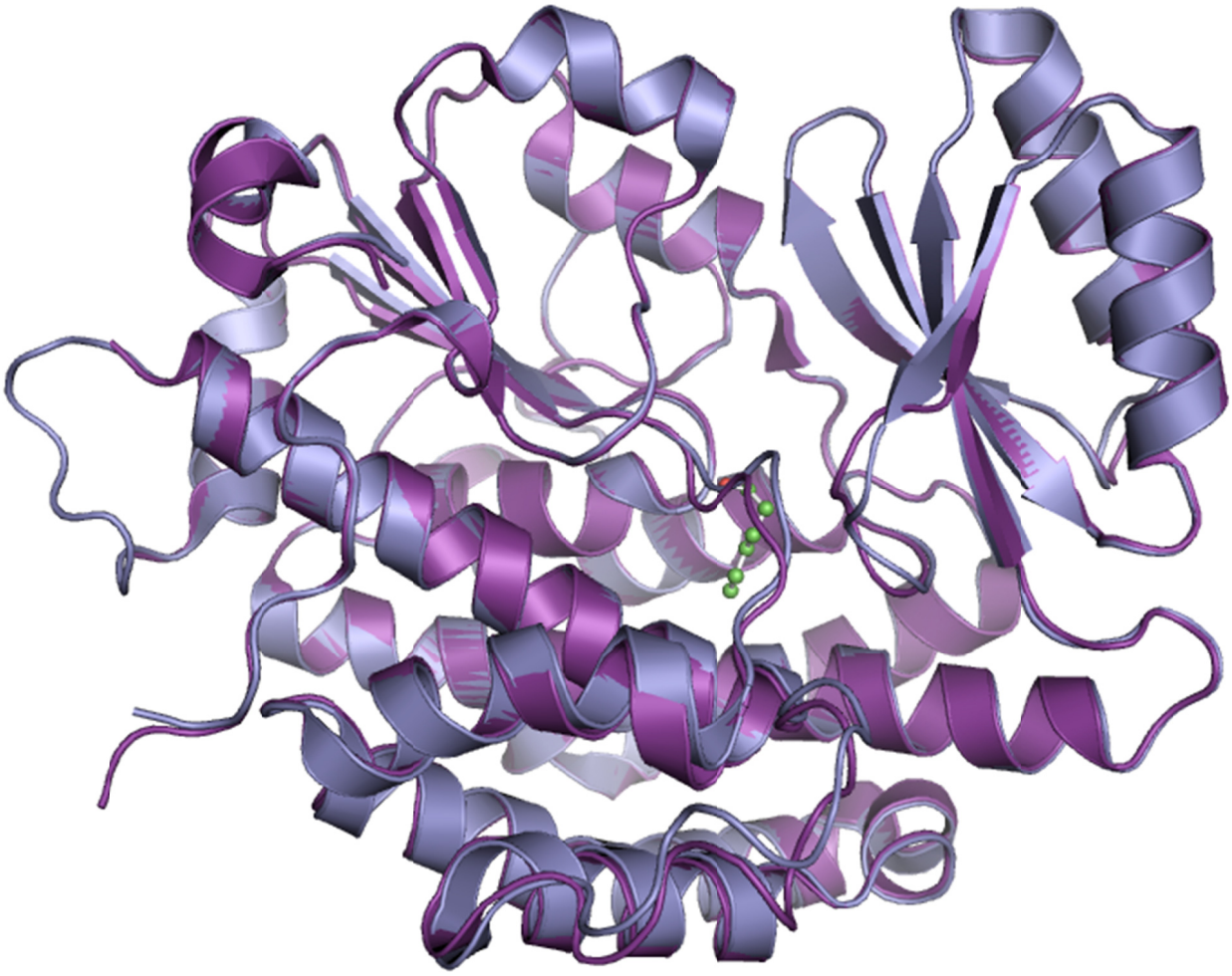
Supplementary Figure 16. *^{1}H-^{13}C HMBC NMR spectrum (400 MHz) of cazaldehyde B (**12**) in $CDCl_3$*



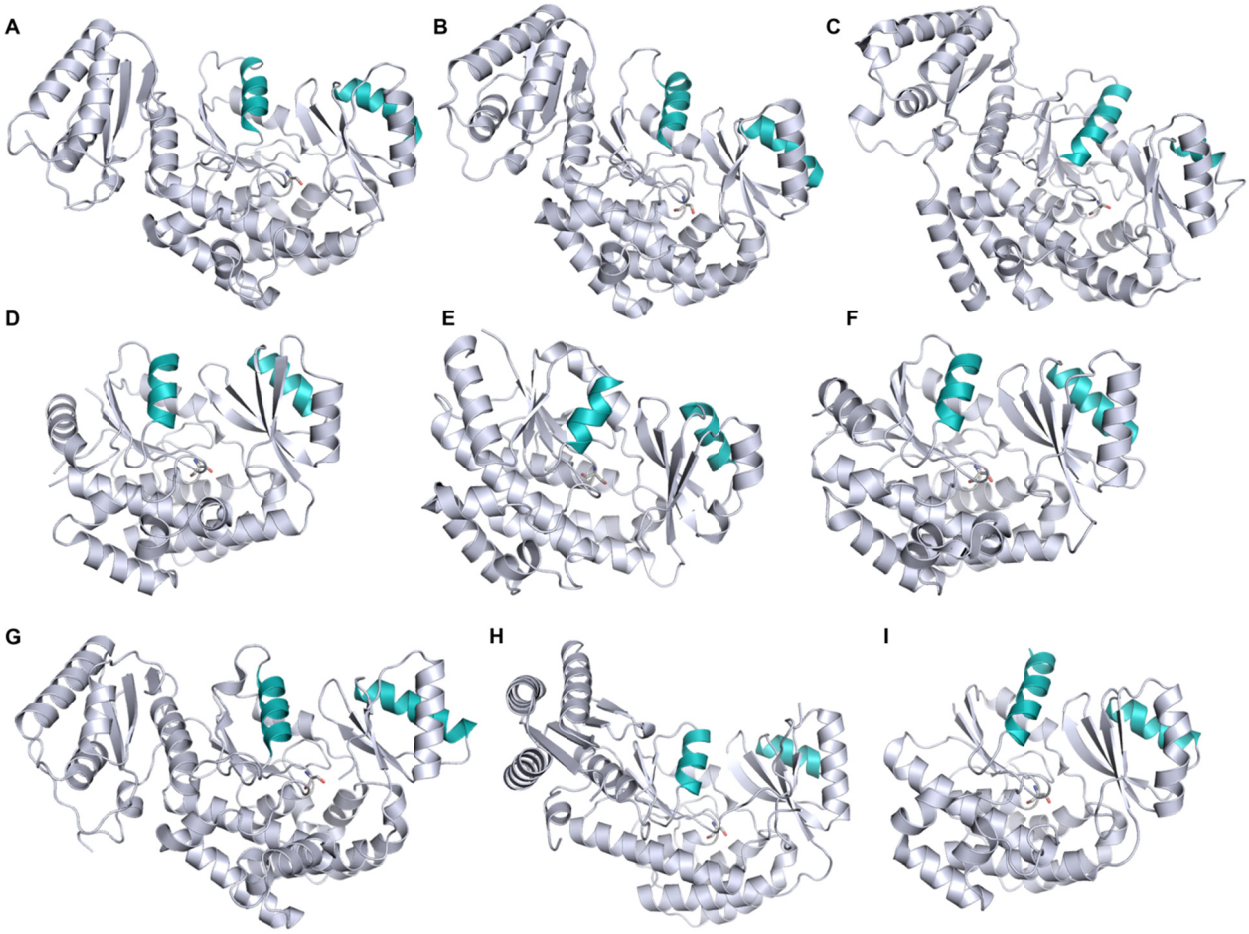
Supplementary Figure 17. *CazaldehydeA (**10**) reduction assay.* i) No reduction of **10** was observed when the compound was incubated with 2 mM NADPH indicating the reduction is enzymatic; ii) No reduction of **10** was observed when the compound was incubated with the standalone wild-type SAT domain in the presence of NADPH indicating that the SAT domain did not catalyze the observed reduction; iii) Reduction of **10** to **12** was observed after **10** was incubated with CazM in the presence of NADPH indicating that the observed reduction most likely takes place via the R domain. HPLC traces are shown in the same scale and at λ = 290 nm.



Supplementary Figure 18. Alignment of the apo and hexanoyl-bound SAT structures. The structures are shown as cartoons where the apo (4R05) is colored magenta and the hexanoyl-bound (4RPM) SAT is colored in light purple. Hexanoyl is shown as ball and stick. The structures were analyzed using PyMOL.[5]



Supplementary Figure 19. Crystal structures of malonyl-CoA:ACP transacylases, acyltransferase domains and trans-acyltransferases. A) Acyltransferase from module 3 of 6-deoxyerthronolide B synthase (2QO3).[6] B) Acyltransferase from module 5 of 6-deoxyerythronolide B synthase (2HG4).[7] C) Acyltransferase from PKS13 (3TZW).[8] D) Malonyl-CoA:ACP transacylase from *Streptomyces coelicolor* A3(2) (1NM2).[9] E) Malonyl-CoA:ACP transacylase from *Helicobacter pylori* (2H1Y).[10] F) Malonyl-CoA:ACP transacylase from *Xanthomonas oryzae* pv. Oryzae KACC10331 (3R97).[11] G) Acyltransferase from zwittermicin polyketide synthase ZmaA (4QBU).[12] H) Acyltransferase from the enediyne polyketide synthase Dyn_{10} (4AMM).[13] I) Trans-acyltransferase involved in disorazole synthesis (3RGI).[14] Each structure is shown as a cartoon and helices corresponding to αJ and αP in the SAT structure are highlighted in teal. The active site Ser residues for each structure is shown as a stick. The structures were analyzed using PyMOL.[5]



Supplementary Table 5. *X-ray data collection and refinement statistics for the SAT domain*

	SeMethionine *apo* SAT	**SeMethionine Hexanoyl-SAT**
Data Collection		
PDB accession code	4RO5	4RPM
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions		
a=*b*= *c* (Å)	43.13, 52.16, 163.09	45.67,52.36,168.03
α=β= γ (°)	90.0	90.0
Reflections observed	156783	283472
Unique reflections	87226	79353
Wavelength (Å)	0.9792	0.9789
Resolution (Å)	81.5-1.9	84.0-1.4
Highest Resolution Shell (Å) [a].	1.64-1.60	1.44-1.40
R_{sym} (%) [b]	4.7(26.4)	4.9(53.9)
CC(1/2)	98.8(85.5)	99.8(80.0)
I/σ	11.53(2.98)	12.79(2.01)
Completeness (%)	92.7(88.1)	98.5(96.6)
Wilson B value (Å**2)	12.8	16.2
Refinement		
Resolution (Å)	81.5-1.60	84.0-1.4
Resolution (Å) (last shell)	1.63-1.60	1.42-1.40
Reflections Used	48603	79350
R_{work}(%)/R_{free}(%)[c]	18.7(20.0) 21.8(25.0)	18.7(25.86) 20.4(27.33)
Protein Molecules in Asymmetric Unit	1	1
Number of non-H atoms		
Protein	2989	3040
Non-protein	189	185
RMS deviations		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	1.067	1.018
Average *B*-factor (Å^2)		
Protein atoms	15.37	22.29
Non-protein atoms	19.70	26.9
Ramachandran plot regions[d]		
Favored	99.22	98.74
Allowed	0.78	1.26
Outliers	0.0	0.0

[a] Highest resolution shell shown in parenthesis. [b] $Rsym = 100\ x \sum (I - I(mean))^2 / \sum I^2$ where *I* is the observed intensity of the reflection HKL and the sum is taken over all reflections

HKL.[c] $Rfactor = 100\ x \sum \left| |F_{obs}| - |F_{calc}| \right| / \sum |F_{obs}|$ F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement. [d]Percentage of residues in Ramachandran plot regions were determined using MOLPROBITY.[15]

Supplementary References

(1) Lee, K. K.; Da Silva, N. A.; Kealey, J. T. *Anal. Biochem.* **2009**, *394*, 75.

(2) Keszenman-Pereyra, D.; Lawrence, S.; Twfieg, M. E.; Price, J.; Turner, G. *Curr. Genet.* **2003**, *43*, 186.

(3) Winter, J. M.; Sato, M.; Sugimoto, S.; Chiou, G.; Garg, N. K.; Tang, Y. *J. Am. Chem. Soc.* **2012**, *134*, 17900.

(4) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning, a Laboratory Manual* Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

(5) DeLano, W. L. The PyMOL User's Manual **2002**, DeLano Scientific, San Carlos, CA, USA.

(6) Tang, Y.; Chen, A. Y.; Kim, C-Y.; Cane, D. E.; Khosla, C. *Chem. Biol.* **2007**, *14*, 931.

(7) Tang, Y.; Kim, C-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. USA.* **2006**, *103*, 11124.

(8) Bergeret, F.; Gavalda, S.; Chalut, C.; Malaga, W.; Quémard, A.; Pedelacq, J-D.; Daffé, M.; Guilhot, C.; Mourey, L.; Bon, C. *J. Biol. Chem.* **2012**, *287*, 33675.

(9) Keatinge-Clay, A. T.; Shelat, A. A.; Savage, D. F.; Tsai, S-C.; Miercke, L. J. W.; O' Connell III, J. D.; Khosla, C.; Stroud, R. M. *Structure.* **2003**, *11*, 147.

(10) Zhang, L.; Liu, W.; Xiao, J.; Hu, T.; Chen, J.; Chen, K.; Jiang, H.; Shen, X. *Protein Sci.* **2007**, *16*, 1184.

(11) Natarajan, S.; Kim, J-K.; Jung, T-K.; Doan, T. T. N.; Ngo, H-P-T.; Hong, M-K.; Kim, S.; Tan, V. P.; Ahn, S. J.; Lee, S. H.; Han, Y.; Ahn, Y-J.; Kang, L-W. *Mol. Cells.* **2012**, *33*, 19.

(12) Park, H.; Keavany, B. M.; Dyer, D. H.; Thomas, M. G.; Forest, K. T. *PLoS One.* **2014**, *9*, e110965.

(13) Liew, C. W.; Nilsson, M.; Chen, M. W.; Sun, H.; Cornvik, T.; Liang, Z-X.; Lescar, J. *J. Biol. Chem.* **2012**, *287*, 23203.

(14) Wong, F. T.; Jin, X.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Biochemistry.* **2011**, *50*, 6539.

(15) Chen, V. B. W; Arendall, B.; Headd, J. F.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Cryst.***2010**, *D66*,12.