

SUPPLEMENTAL INFORMATION:

Identification and Characterization of the Echinocandin B gene cluster from *Emericella rugulosa* NRRL 11440

Ralph A. Cacho¹, Wei Jiang³, Yit-Heng Chooi¹, Christopher T. Walsh^{3*}, and Yi Tang^{1,2,*}

1.) Department of Chemical and Biomolecular Engineering,

University of California, Los Angeles, 420 Westwood Plaza, Los Angeles, CA 90095

2.) Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095

3.) Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 200 Longwood Ave, Boston, MA 02115

Table of Contents:

SUPPLEMENTARY TABLES.....	S3
Table S1: Genome Assembly Statistics for <i>Emericella rugulosa</i>	S3
Table S2: Putative NRPS in <i>Emericella rugulosa</i>	S4
Table S3: A domain selectivity of EcdA A domain as prediction.....	S6
Table S4: NMR data of six amino acid residues in 1 (600 MHz) in CD ₃ OD.....	S7
Table S5: Pyrroline-5-reductase genes in <i>E.rugulosa</i>	S9
Table S6: EcdI homologs that are clustered together with NRPS with T ₀ domains.....	S10
Table S7:List of primers used for this study.....	S11
SUPPLEMENTARY FIGURES.....	S12
Figure S1: LCMS trace of authentic Echinocandin B standard.....	S12
Figure S2: Anti-Candida assay performed on Δ <i>htyA</i> and <i>ecdA</i> mutants.....	S13
Figure S3: PCR screening of Δ <i>ecdA-I-16</i>	S14
Figure S4: SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1.....	S15
Figure S5: SDS-PAGE gel of purified N-terminal-His tagged EcdI.....	S16
Figure S6: SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1 variants.....	S17
Figure S7: Radiogram from EcdA-M1 loading with the SDS-PAGE gel.....	S18
Figure S8: Loading assay of EcdA-M1 with ¹⁴ C-L-ornithine.....	S19
Figure S9: Effect of [CoA] on the reaction rate of EcdI.....	S20
Figure S10: Loading of alternative substrates onto EcdA-M1 by EcdI.....	S21
Figure S11: PCR screening of Δ <i>htyA</i> mutants.....	S22
Figure S12: Relative position of <i>ecd</i> and <i>hty</i> clusters in their respective contigs.....	S23
Figure S13:1D- ¹ H spectrum of 1 (600 MHz)in CD ₃ OD.....	S24
Figure S14: HSQC Spectrum of 1 (600 MHz) in CD ₃ OD.....	S25
Figure S15: gCOSY Spectrum of 1 (600 MHz) in CD ₃ OD.....	S26
Figure S16: HMBC Spectrum of 1 (600 MHz) in CD ₃ OD.....	S27

Table S1: Genome Assembly Statistics for *Emericella rugulosa*.

	Unused Reads	All Contigs	Contigs \geq 100bp	Contigs \geq 1000bp
No. of contigs	25	433	433	373
Min length	88	139	139	1009
Median length		154,740	1547,40	21,740
Mean length	188	74,420	74,420	86,322
N50 length		235,313	235,313	235,313
No. of contigs \geq N50		38	38	38
Length sum	4719	32,224,016	32,224,016	32,198,232

Table S2: Putative NRPS in *Emericella rugulosa*.

Protein annotation	NRPS domains	Closest homolog in NCBI database through BlastP search (Identity, Similarity)	E values, Score
ErNRPS284	C-A-T-C-A-T-C-A-T-E-C-A-T-C-A-T-C _T	EHK21807.1* (47%, 63%), <i>Trichoderma virens</i> Gv29-8	0.0, 1.095 x 10 ⁴
ErNRPS57	A-T-C-A-T-E-C-A-T-C-A-T-C-A-T-C-A-T-R	AN7884.2 (92%,96%), <i>A.nidulans</i> A4	0.0, 1.379 x 10 ⁴
ErNRPS123 (<i>ecdA</i>)	T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C _T	EGR45389* (33%, 52%), <i>Trichoderma reesei</i> QM6a	0.0, 3.404 x 10 ⁴
ErNRPS74	A-T-E-C-A-C-A-T-C-A-T-E-C-T-C-T	AN0016 (92%, 96%) , <i>A.nidulans</i> A4	0.0, 1.127 x 10 ⁴
ErNRPS239	A-T-E-C-A-C-A-T-C-A-T-E-C-T-C-T-C-T	AN1242.2 (91%,94%), <i>A.nidulans</i> A4	0.0, 1.340 x 10 ⁴
ErNRPS99	T-E-C-A-T-E-C-A-T-C-A-T-C-A-T-C-A-T-C _T	AN2545, <i>easA</i> (91%, 94%), <i>A.nidulans</i> A4	0.0, 1.303 x 10 ⁴
ErNRPS32	A-T-C-A-T-C-A-T-C-T-C-T-C _T	AN0607 (91%, 95%) (sidC), <i>A.nidulans</i> A4	0.0, 9032
ErNRPS184	A	AN8433.2 (91%,94%), <i>A.nidulans</i> A4	0.0, 1477
ErNRPS30	A	ATEG09019 (76%, 86%), <i>A.terreus</i>	5 x 10 ⁻⁷⁰ , 265
ErNRPS153a	A-T-Te	AN8513 (92%, 94%) TdiA (terrequinone NRPS) <i>A.nidulans</i> A4	0.0, 1842
ErNRPS153b	A-T-C-A-T-C _T	EED13064.1, GliP-like,(48%, 68%), <i>Talaromyces stipitatus</i> ATCC 10500	0.0, 1896
ErNRPS 92	T-C-A-T-C _T	EGE04746 (38%, 55%)*, <i>Trichophyton equinum</i> CBS 127.97	0.0, 1134
ErNRPS84	A-T-Te	NFIA_045590(69%, 80%), <i>N.fischeri</i> NRRL 181	0.0, 1479
ErNRPS24a	A-T-Te	AN8105.2 (90%, 94%), <i>A.nidulans</i> A4	0.0, 1957
ErNRPS24b	A-T	AN9216 (87%,90%), <i>A.nidulans</i> A4	0.0, 1003

Table S2: Putative NRPS in *Emericella rugulosa* (continued).

Gene annotation	NRPS domains	Closest homolog in NCBI database through BlastP search	E values, Score
ErNRPS105a	A-T-Te	AN5610.2 (99%, 99%), <i>A.nidulans</i> A4	0.0, 2861
ErNRPS105b	A-T	AN5626.2 (98%, 99%), <i>A.nidulans</i> A4	0.0, 1352
ErNRPS146	A-T-Te	AFUA_8G01640 (62%, 75%), <i>A.fumigatus</i>	0.0, 1296
ErNRPS93	A-T-Te	NFIA_005280 (76%, 87%), <i>N.fischeri</i> 181	0.0, 1614
ErNRPS20	A-T	AFLA_066840 (40%, 58%), <i>A.flavus</i> NRRL3357	7 x 10 ⁻¹¹ , 348
ErNRPS100a	C-A	AN2621 (68%, 76%) AcvA synthetase, <i>A.nidulans</i> A4	3 x 10 ⁻¹⁴⁵ , 750
ErNRPS100b	A	AN2621 (63%, 69%)* AcvA synthetase <i>A.nidulans</i> A4	2x10 ⁻¹⁰⁰ , 774
ErNRPS291	A-T	AN5990.2 (98%, 99%), <i>A.nidulans</i> A4	0.0, 1128
ErNRPS247	A-T	AAX09993.1 (52%, 68%)*, <i>Cochliobolus heterostrophus</i>	0.0, 1015
ErNRPS87	A-C	EFY94744.1 (37%, 55%), <i>Metarhizium anisopliae</i> ARSEF 23	6 x 10 ⁻⁶³ , 229
ErNRPS43	A-T-R	AN5318 (98%, 99%), <i>A.nidulans</i> A4	0.0, 2536

Definitions:

A-adenylation domain

T- Thiolation/ Peptidyl Carrier Protein Domain

C – Condensation domain

R – Reductase domain

Te- thioesterase domain

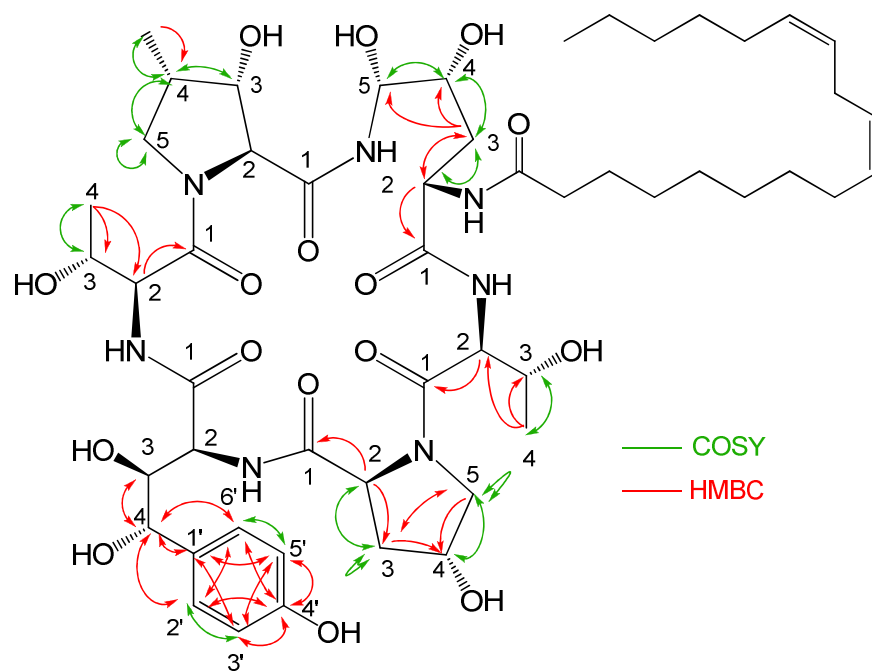
C_T- Terminal Condensation domain

Table S3: A domain selectivity of EcdA A domain as predicted by NRPSpredictor1 and NRPS predictor2 algorithms. Prediction amino acid selectivity based on the bacterial and fungal A domain database are given.

	amino acid in 1	10 aa code	Bacterial		Fungal	
			NRPS1	NRPS2	NRPS1	NRPS2
A1	L-ornithine	DVMELSSITK	hpg	dab	hpg	hpg
A2	L-threonine	DAQTAVAIHK	phe	phe	phe	ala
A3	L-proline	DVSSATTVCK	no prediction	trp	no prediction	pro
A4	L-homotyrosine	DGEAVGCVFK	val,leu	ile	val	leu
A5	L-threonine	DAQTIVAIHK	phe	phe	phe	Ala
A6	4-methyl-L-proline	DNTMITAMSK	no prediction	asn	no prediction	tyr

Hpg = hydroxyphenylglycine, dab= di-aminobutyric acid, phe = L-phenylalanine, ala= L-alanine, pro = L-proline, trp = L-tryptophan, val = L-valine, leu = L-leucine, ile = L-isoleucine, asn = L-asparagine, tyr = L-tyrosine

Table S4: ^1H NMR and 2D-NMR (HSQC, HMBC, gCOSY) data of six amino acid residues in Echinocandin B **1** (600 MHz) in CD_3OD . The parameters in red are from pneumocandin A_0 ¹ and parameters in blue are from ^{13}C -NMR spectrum of **1** from literature as comparison². COSY and HMBC correlation signals that are used to identify **1** are shown in the structure.



4R-OH, 5R-OH-L-ornithine₁

C-1	174.1	174.6	174.0			
C-2	51.3	51.4	52.8	2-H	4.39	4.43
C-3	34.8	34.8	36.73	3-Ha	2.07	2.00
				3-Hb	1.93	2.00
C-4	70.6	70.6	70.87	4-H	3.94	4.00
C-5	74.2	74.0	74.3	5-H	5.25	5.26

L-threonine₂

C-1	172.6	172.7	172.2		
C-2	58.4	58.4	58.6	2-H	4.95
C-3	68.1	68.2	68.3	3-H	4.51
C-4	19.5	19.7	20.1	4-H	1.20

Table S4: continuedL-threonine₅

C-1		172.7	172.2		
C-2	56.7	58.4	56.6	2-H	4.84
C-3	69.5	68.2	69.42	3-H	4.15
C-4	19.5	19.7	19.5	4-H	1.21

4-hydroxy-L-proline₃

C-1	173.3	173.4	173.2		
C-2	62.3	62.5	62.3	2-H	4.58
C-3	38.3	38.5	38.78	3-Ha	2.42
				3-Hb	2.04
C-4	71.1	71.3	70.87	4-H	4.50
C-5	56.8	57.1	56.6	5-Ha	3.96
				5-Hb	3.79

3,4-dihydroxy-L-homotyrosine₄

C1		172.5	169.7			
C-2	56.1	56.4	56.6	2-H	4.32	4.31
C-3	76.7	76.9	76.8	3-H	4.24	4.27
C-4	75.4	75.8	75.51	4-Ha	4.30	4.28
C-1'	132.9	133.0	132.5			
C-2'/C-6'	129.5	129.6	129.5	2'-H/6'-H	7.14	7.13
C-3'/C-5'	116.0	116.2	116.1	3'-H/5'-H	6.76	6.75
C-4'	158.3	158.5	158.2			

3-Hydroxy-4-methyl-L-proline₆

C-1		172.7	172.5			
C-2	69.3	70.2	69.42	2-H	4.34	4.36
C-3	75.4	75.9	75.51	3-H	4.19	4.14
C-4	38.8	39.1	38.78	4-H	2.52	2.48
C-5	52.7	53.0	55.9	5-Ha	3.86	4.09
				5-Hb	3.38	3.34
4-CH ₃	11.1	11.1	11.3	4-CH ₃	1.05	1.04

Table S5: Pyrroline-5-reductase genes in *E.rugulosa*.

Annotation	Closest Homolog	Identity/Similarity
ErP5CR65	AN9279, <i>A.nidulans</i> A4	94%, 96%
ErP5CR54	AN7387, <i>A.nidulans</i> A4	92%, 93%
ErP5CR11	AN4355, <i>A.nidulans</i> A4	93%, 96%
ErP5CR150	AN6025, <i>A.nidulans</i> A4	92%, 95%

Table S6: EcdI homologs that are clustered together with NRPS with initiation T domains. In addition to the NRPS genes, four out of the five EcdI homolog genes are also clustered together with highly-reducing polyketide synthase genes (HR-PKS), suggesting that the HR-PKS product is transferred by the EcdI homolog to the NRPS. This further suggest that the clusters seen above are for the biosynthesis of lipopeptides. Identity and similarity percentages of EcdI homologs in comparison to EcdI sequence are given.

EcdI blast hits	NRPS in the region (domain organization)	PKS in region (domain organization)
PMAA_001090 (45%, 63%) <i>Penicillium marneffei</i>	PMAA_001110 (TCATCATCATCT)	PMAA_001080 (KS-AT-DH-Met-ER-KR-ACP)
AN2549 (46%, 63%) <i>A.nidulans</i> A4	AN2545 (easA) (TECATECATCATCAT)	EasB (KS-AT-DH-Met-ER-KR-ACP)
TSTA_098290 (45%, 63%) <i>Talaromyces stipitatus</i> ATCC 10500	TSTA_098310 (TCATCATCATCT)	TSTA_098280 (KS-AT-DH-Met-ER-KR-ACP)
FOXB_07957 (43%, 53%) <i>Fusarium oxysporum</i> Fo5176	FOXB_07954 (TCATCATCATCATC)	None
EGC42542 (41%, 51%) <i>Ajellomyces capsalatus</i> H88	EGC42544 (TCATCATCATCTCAATC)	EGC42541 (KS-AT-DH-Met-ER-KR-ACP)

Table S7: List of primers used for this study.

Primer Name	Primer Sequence (5' -3')	Notes
ecdA-KO-P1	agggggtcataacaaccgct	For mutant screening and KO cassette cloning
ecdA-KO-P2	gcctgatgccaggcaacctc	For knockout cassette cloning
ecdA-KO-P3	cgccgggtcctgcccgtcaccgagatttaggctgtgggctgatgcgtggt	
ecdA-KO-P4	tgctccttcaatatcatcttctgtcggctggctcgatcatgtcacagta	
ecdA-KO-P5	ggccacaatgttgctacttgccac	
ecdA-KO-P6	gtccagtcaatgtgagtcgca	For mutant screening and KO-cassete cloning
htyA-KO-P1	cgtaatacgcgatgtggcagataaatagtt	For mutant screening and KO-cassete cloning
htyA-KO-P2	ccgtcctccactagctggaat	For knockout cassette cloning
htyA-KO-P3	cgccgggtcctgcccgtcaccgagatttagagctagagtctccgcttccgaa	
htyA-KO-P4	aaagtgctccttcaatatcatcttctgtcggcgtttcagcatcccggtt	
htyA-KO-P5	tctcccaggtgggtggcgga	
htyA-KO-P6	cacaacgcgaggtccccacgaaca	For mutant screening and KO-cassete cloning
bar-f	aagtaaccatgagcccagaacga	For mutant screening
bar-r-stp	ctaaatctcggtagcgggca	For mutant screening and KO cassette cloning
P _{trpC} -F	cgacagaagatgatattgaa	For KO cassette cloning
<i>Nde</i> I-EcdA-T ₀	aaaaacatatgcacagaaccaacgagatggaa	For cloning of EcdA-M1
<i>Eco</i> RI-EcdA-T ₁	ttttgaattcgggtccgctcctgcat	
ecdA-S47A-F	ggaggcgacgcatggctacc	For cloning of EcdA-M1-S47A variant
ecdA-S47A-R	ggtagccatcgctcgctcc	
ecdA-S1127A-F	gggtgatgcaatcgccgcat	For cloning of EcdA-M1-S1127A variant
ecdA-S1127A-R	atcgcgcgattgcatcacc	
ecdI- <i>Nde</i> I-F	aaaaaacatatggtcttacttcccagca	For cloning EcdI
ecdI- <i>Eco</i> RI-R	ttttgaattcttaaatcaactccttcaact	

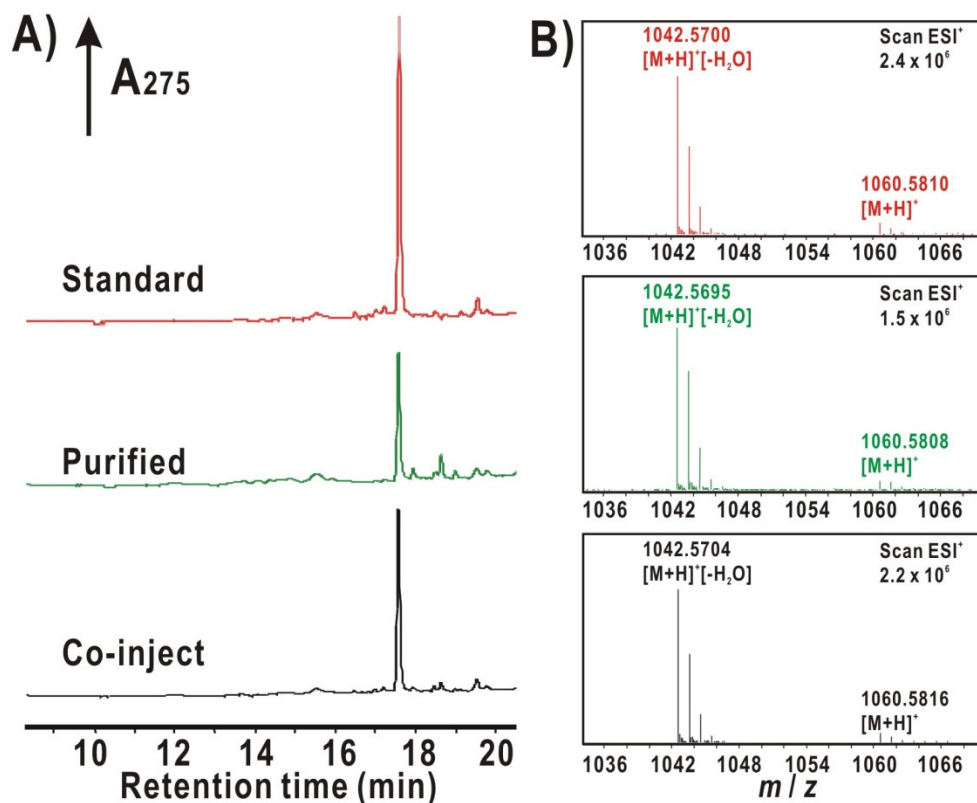


Figure S1: Chromatographic traces at 275 nm (A) and mass spectra (B) of authentic Echinocandin B standard (top, red), Echinocandin B purified from fermentation broth of *E.rugulosa* NRRL 11440 culture (middle, green), and co-injection of the two (bottom, black). The theoretical mass-to charge ratio of Echinocandin B **1** is 1060.5813 for $[C_{52}H_{81}N_7O_{16}+H]^+$ and 1042.5707 for $[C_{52}H_{81}N_7O_{16}+H-H_2O]^+$. Authentic standard for **1** was purchased from Santa Cruz Biotechnology, Inc (Santa Cruz, CA)

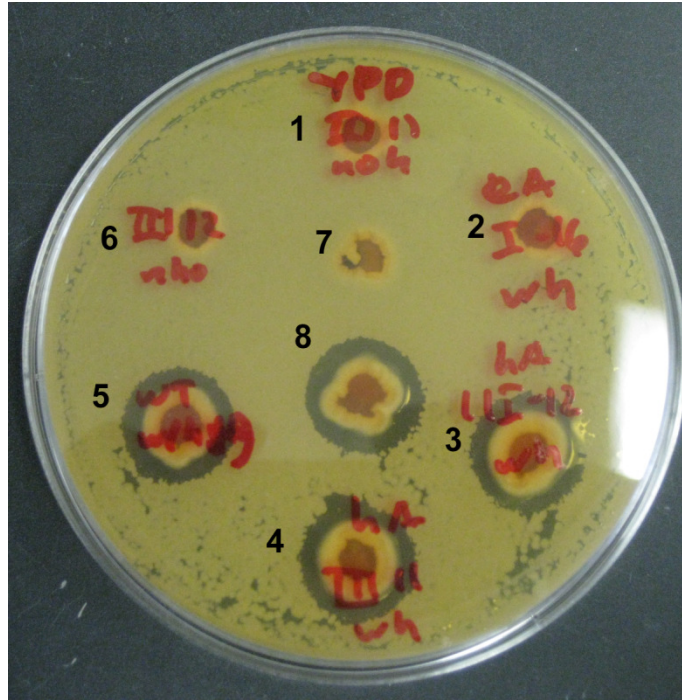


Figure S2: Anti-*Candida* assay performed on $\Delta htyA$ -III-11 without (colony 1) and with (colony 4) L-homotyrosine feeding, $\Delta htyA$ -III-12 without (colony 6) and with (colony 3) L-homotyrosine, on $\Delta ecdA$ -I-16 with (colony 2) and without (colony 7) L-homotyrosine and wild-type *E.rugulosa* with (colony 5) and without (colony 8) L-homotyrosine. Anti-*Candida* bioassay was performed by feeding L-homotyrosine (0.1 mg/mL) to the static liquid cultures of $\Delta htyA$ and $\Delta ecdA$ mutants at day-4 before transferring the mycelial discs from individual clones to the YPD plate pre-inoculated with *C. albicans* at day-7.

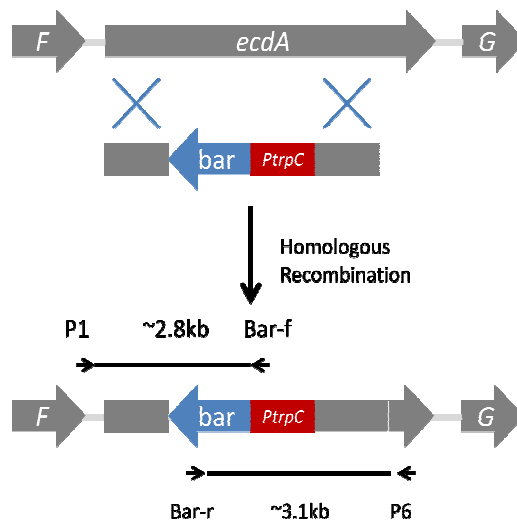
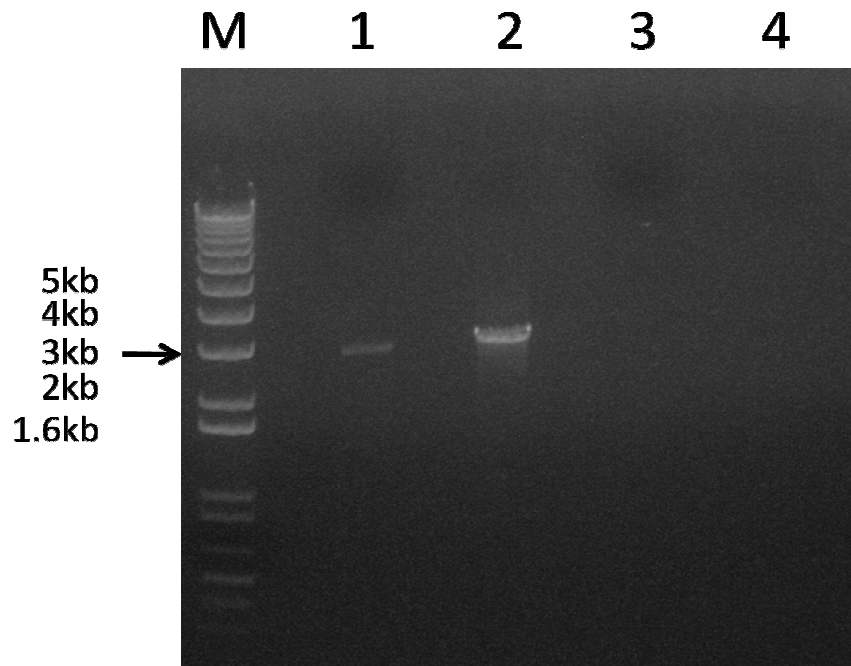


Figure S3: PCR screening of $\Delta ecdA$ -I-16. (M) 1 kb plus ladder (Invitrogen). Amplification of $\Delta ecdA$ -I-16 genomic DNA (gDNA) using primer pair P1 and bar-f showing expected ~ 2.8 kb amplicon (lane 1). Amplification of $\Delta ecdA$ -I-16 gDNA using primer pair ecdA-KO-P6 and bar-r showing expected ~ 3.1 kb amplicon (lane 2). Amplification of wild-type *E.rugulosa* gDNA using primer pair ecdA-KO-P1 and bar-f (lane 3) Amplification of wild-type *E.rugulosa* gDNA using primer pair ecdA-KO-P6 and bar-r (lane 4).

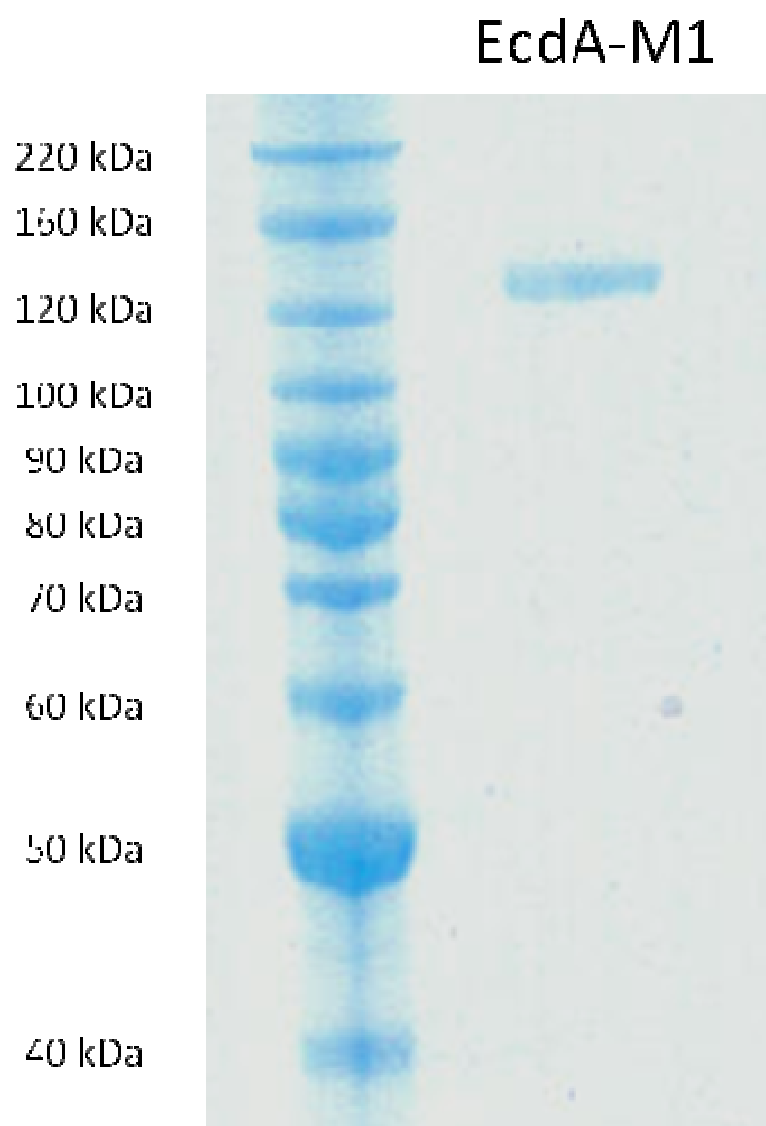


Figure S4: SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1 using a Nickel-NTA resin showing the expected size of 130 kDa.

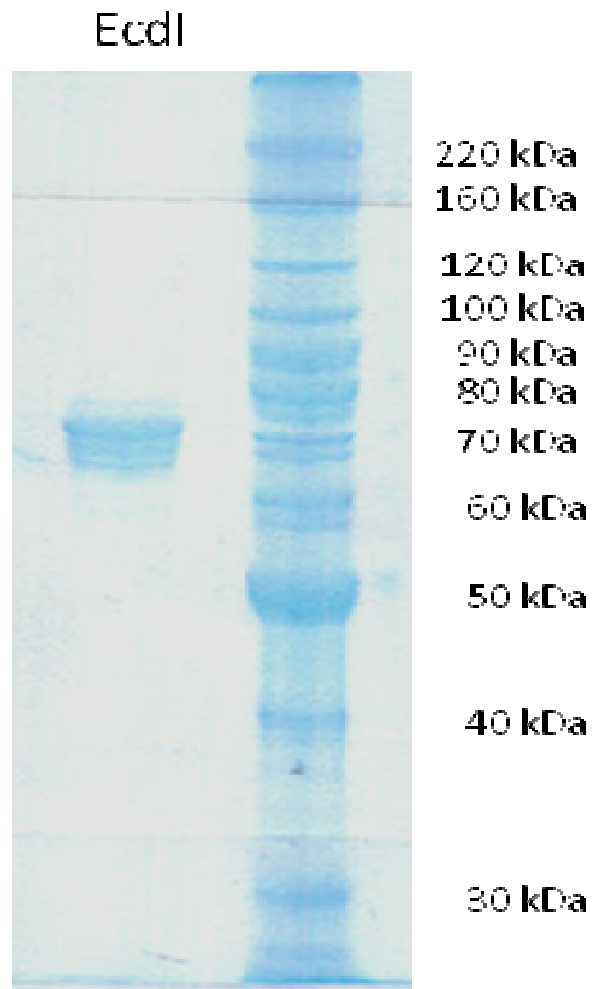


Figure S5: SDS-PAGE gel of purified N-terminal-His tagged EcdI using Nickel-NTA resin showing expected size of 70 kDa.

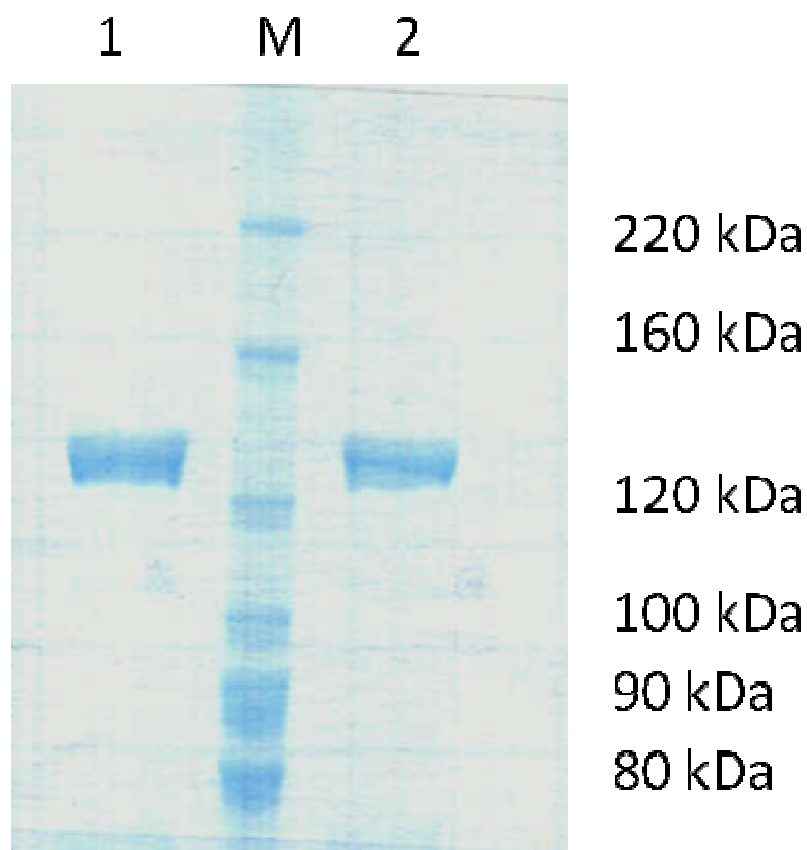


Figure S6: SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1 variants using a Nickel-NTA resin showing the expected size of 130 kDa. **Lane 1:** S47A variant of EcdA-M1 ($T_0^*CAT_1$). **Lane 2:** S1127A variant of EcdA-M1 (T_0CAT_1)

A)

Sample	1	2	3	4	5	6	7
EcdA	-	+	+	+	+	T ₀ *CAT ₁ T ₀ CAT ₁ *	
EcdI	+	-	+	+	+	+	+
Sfp	+	+	-	+	+	<i>holo-</i>	<i>holo-</i>
ATP	+	+	+	-	+	+	+
*LA	+	+	+	+	+	+	+

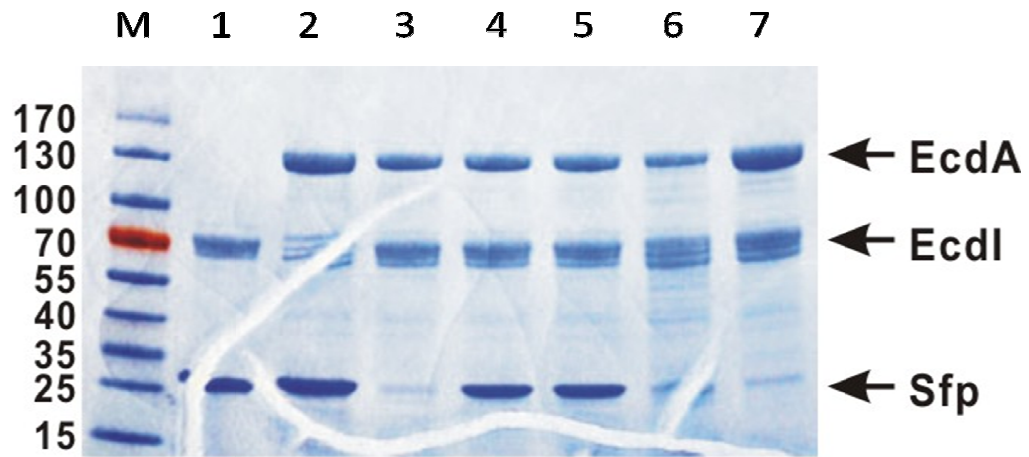
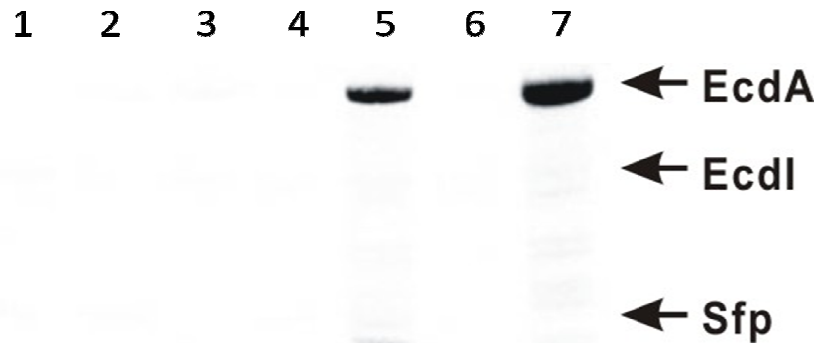
B)**C)**

Figure S7: A) The components of the assay mixture used in the radiography B) SDS-PAGE gel from EcdA-M1 ¹⁴C-linoleic acid loading assay showing the protein present in the assay mixture and C) its matching radiogram used for Figure 3A of main text. EcdA variants S47A and S1127A are in *holo-* form due to expression in *E.coli* BAP1.

Sample	1	2	3	4
EcdA	+	+	T ₀ *CAT ₁	T ₀ CAT ₁ *
Sfp	+	+	<i>holo-</i>	<i>holo-</i>
ATP	-	+	+	+
*L-Orn	+	+	+	+



Figure S8: Loading assay of EcdA-M1 with ¹⁴C-L-ornithine showing incorporation of [14C]-L-ornithine to wild-type EcdA-M1 and T₀*CAT₁ (S47A) but not to T₀CAT₁* (S1127A) as expected. EcdA variants S47A and S1127A are in *holo-* form due to expression in *E.coli* BAP1.

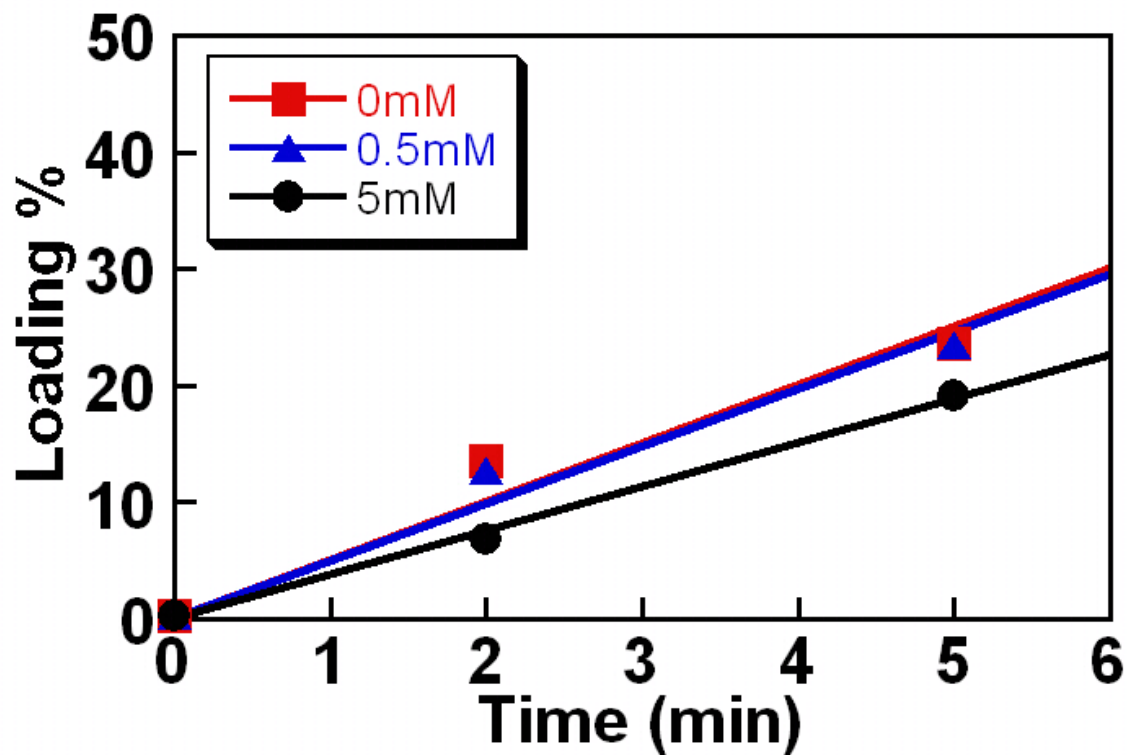


Figure S9: No apparent effect of [CoA] on the reaction rate of loading of linoleic acid onto EcdA-M1 by EcdI. After the conversion of *apo* to *holo* form, the residual CoA is removed by buffer exchange through a ultracentrifugal filter (Molecular weight cutoff of 10 kDa). 0, 0.5, or 5 mM CoA was added afterwards and the loading percentages were measured at 2 and 5 min.

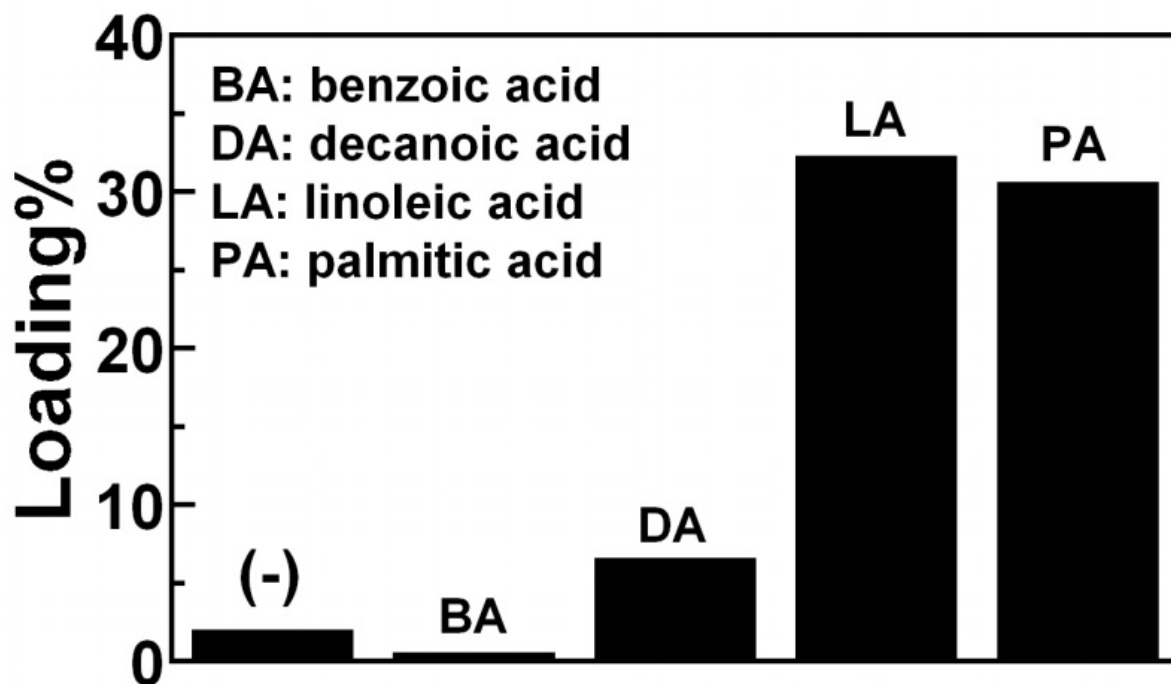


Figure S10: Loading of alternative substrates onto EcdA-M1 by EcdI. The assay procedure was described in Materials and Methods. The reactions were run at ambient temperature for 30 min and subject to quantification. The negative control sample contains linoleic acid but no addition of ATP.

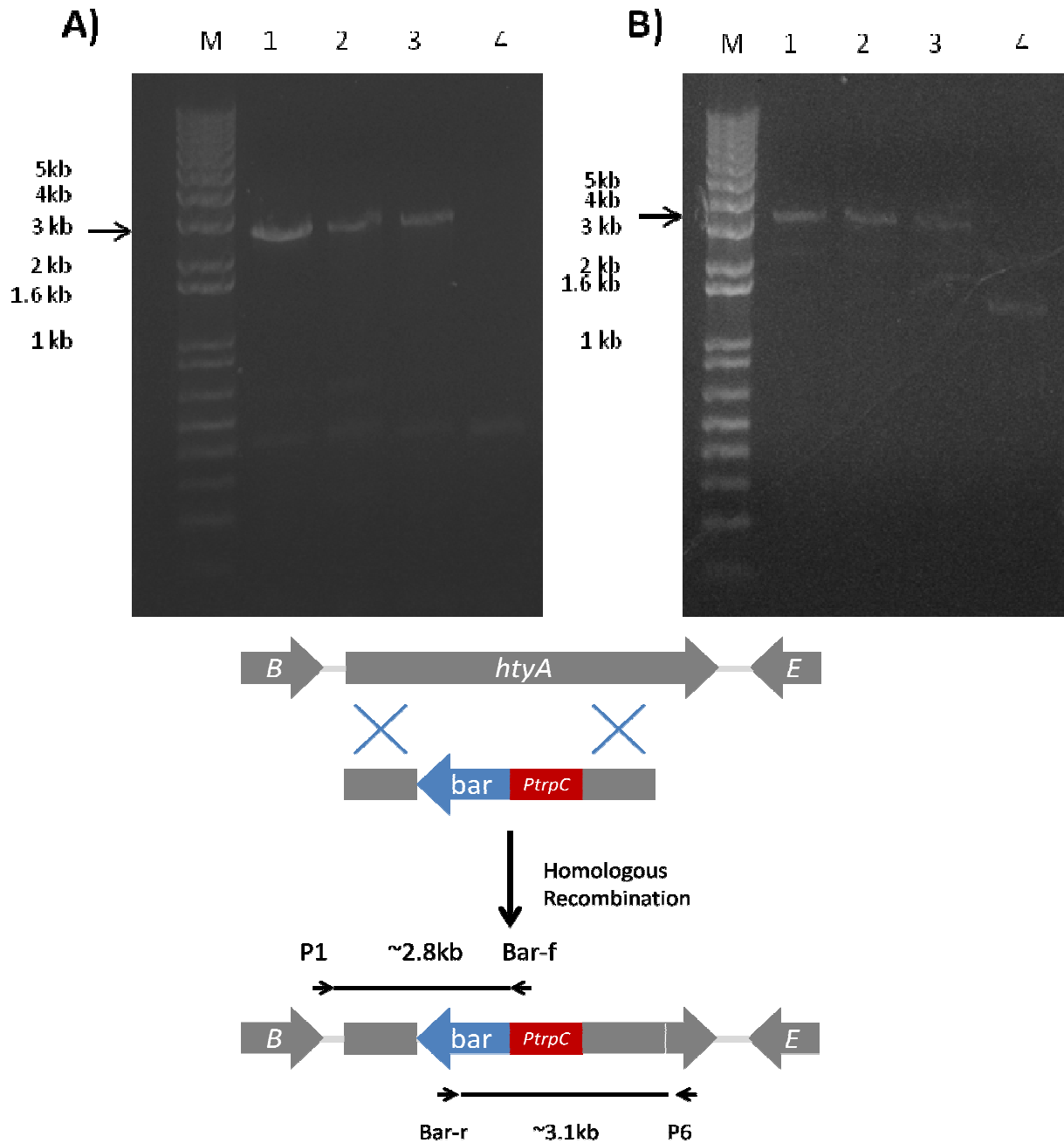
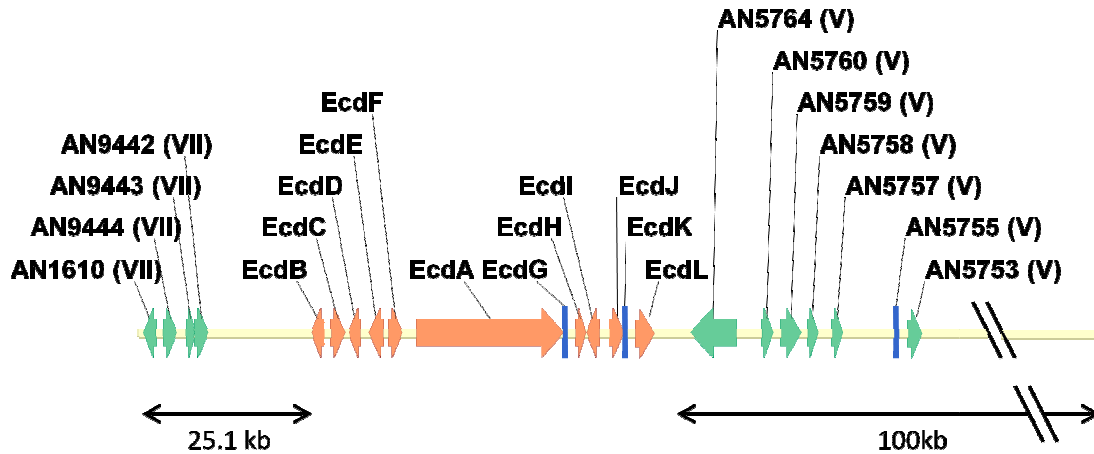
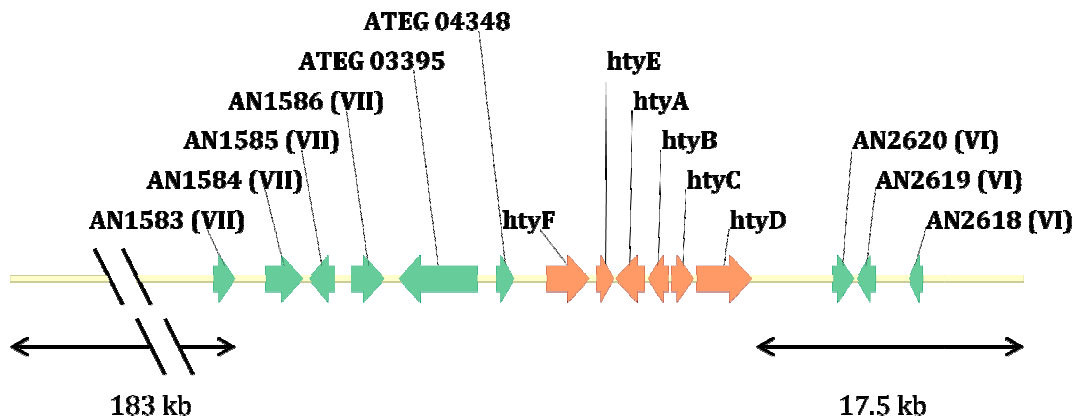


Figure S11: PCR screening of $\Delta htyA$ mutants **A.)** Amplification of gDNA from $\Delta htyA$ III-7 (lane 1), III-11(lane 2), III-12(lane 3),and wild-type (lane 4) using primer pairs htyA-KO-P1 bar-f. $\Delta htyA$ mutants show the expected ~ 2.8 kb amplicon **B.)** Amplification of gDNA from of (lane 1) $\Delta htyA$ III-7, (lane 2) III-11, (lane 3) III-12, and (lane 4)wild-type using primer pairs htyA-KO-P6 bar-r. $\Delta htyA$ mutants show the expected ~ 3.1 kb amplicon



Contig 123 (Echinocandin B cluster)
 (total length of Contig 123 179124 bp)



Contig 66 (L-homotyrosine gene cluster)
 (total length of Contig 66 216241 bp)

Figure S12: The *ecd* and *hty* cluster are found on two separate contigs (123 and 66). Based on the available genome sequence information, we cannot determine if the two gene clusters are located on the same or separate chromosomes. The shorter flanking regions of *ecd* and *hty* clusters are ~25 kb and ~17.5 kb respectively. Thus, if the two gene clusters are co-located on a single chromosome on *E. rugulosa* the minimal distance between them will be ~42.5 kb. The presence of microsyntenic blocks flanking the *ecd* and *hty* cluster corresponds to different chromosomes of *A. nidulans* A4 suggests that there are chromosomal translocation events during the divergence of the two species, thus we cannot map the chromosomal location of the two cluster using *A. nidulans* as reference genome.

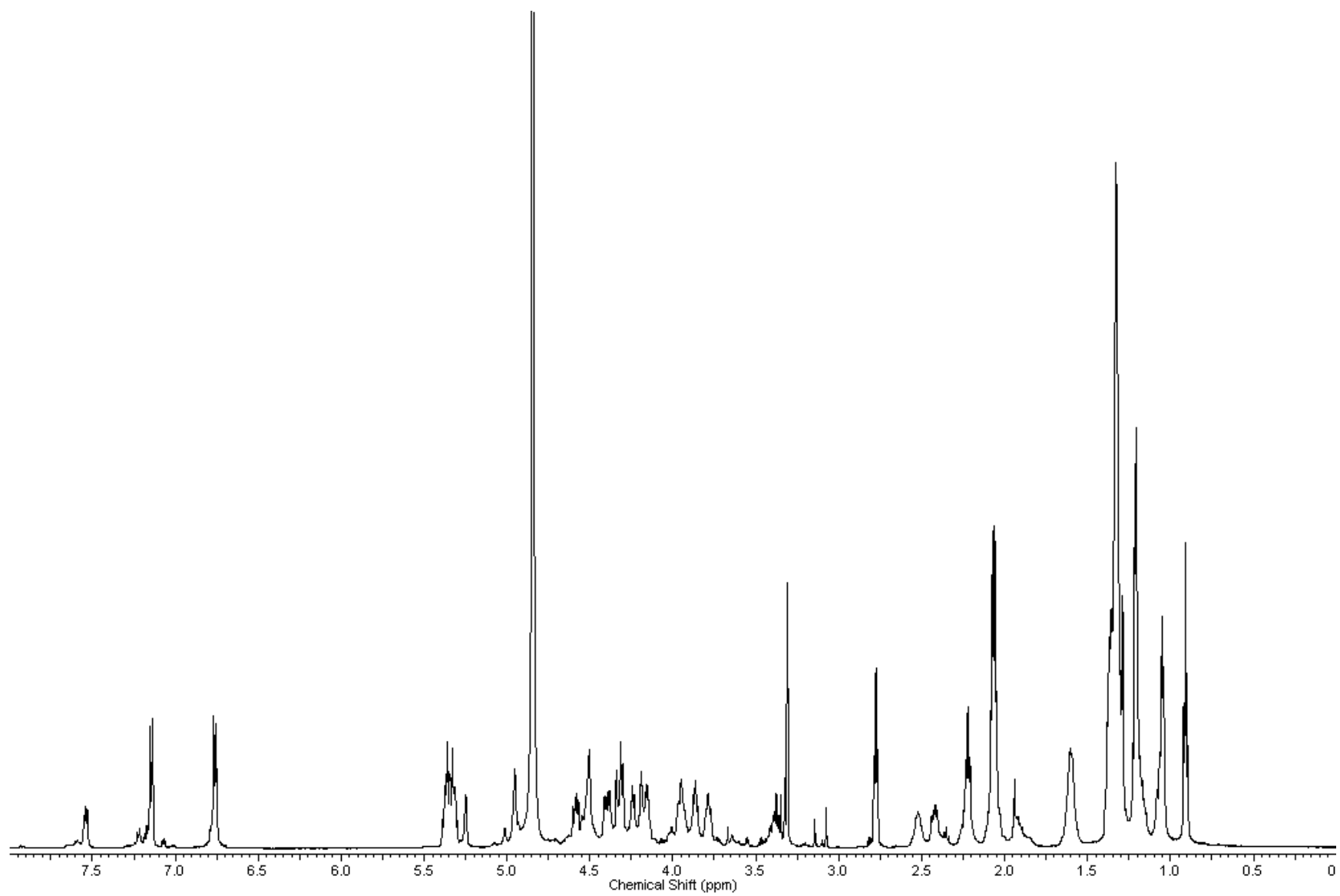


Figure S13: ^1H spectrum of **1** (600 MHz) in CD_3OD

Figure S14: HSQC Spectrum of **1** (600 MHz) in CD₃OD

WJ-III-41_20120322_HSQC.fid.esp

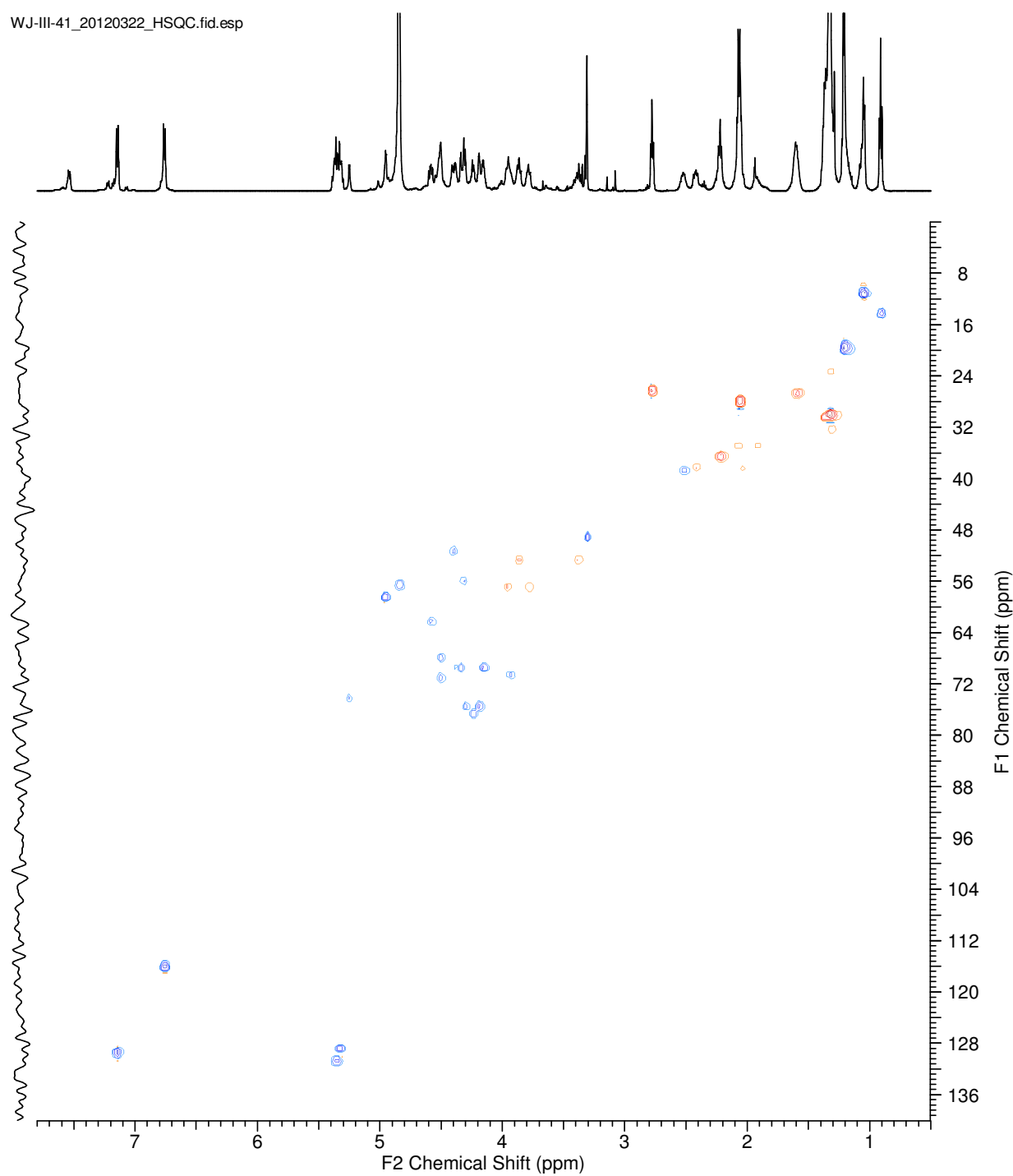


Figure S15: gCOSY Spectrum of **1** (600 MHz) in CD₃OD

WJ-III-41_20120322_gCOSY.fid.esp

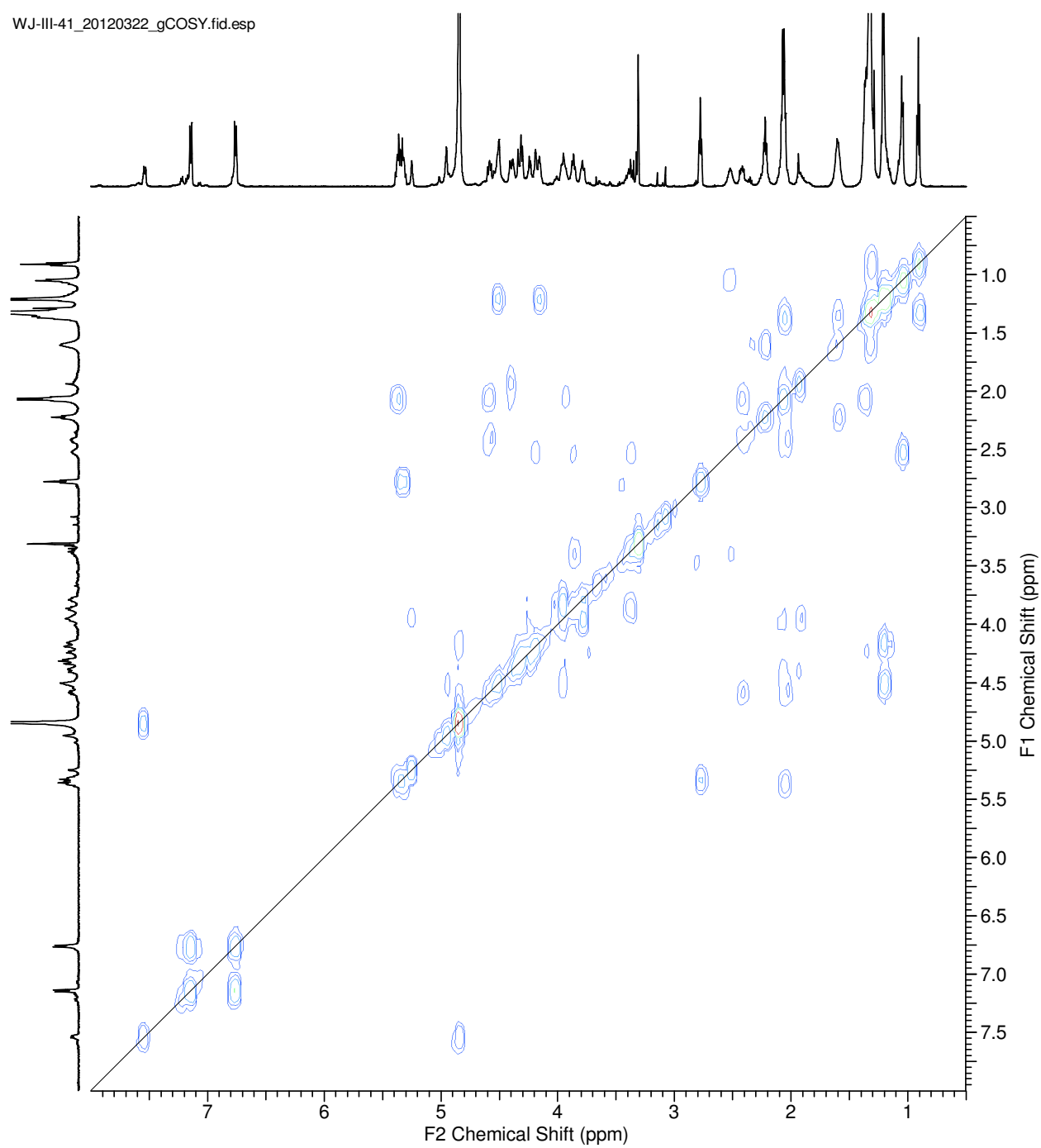
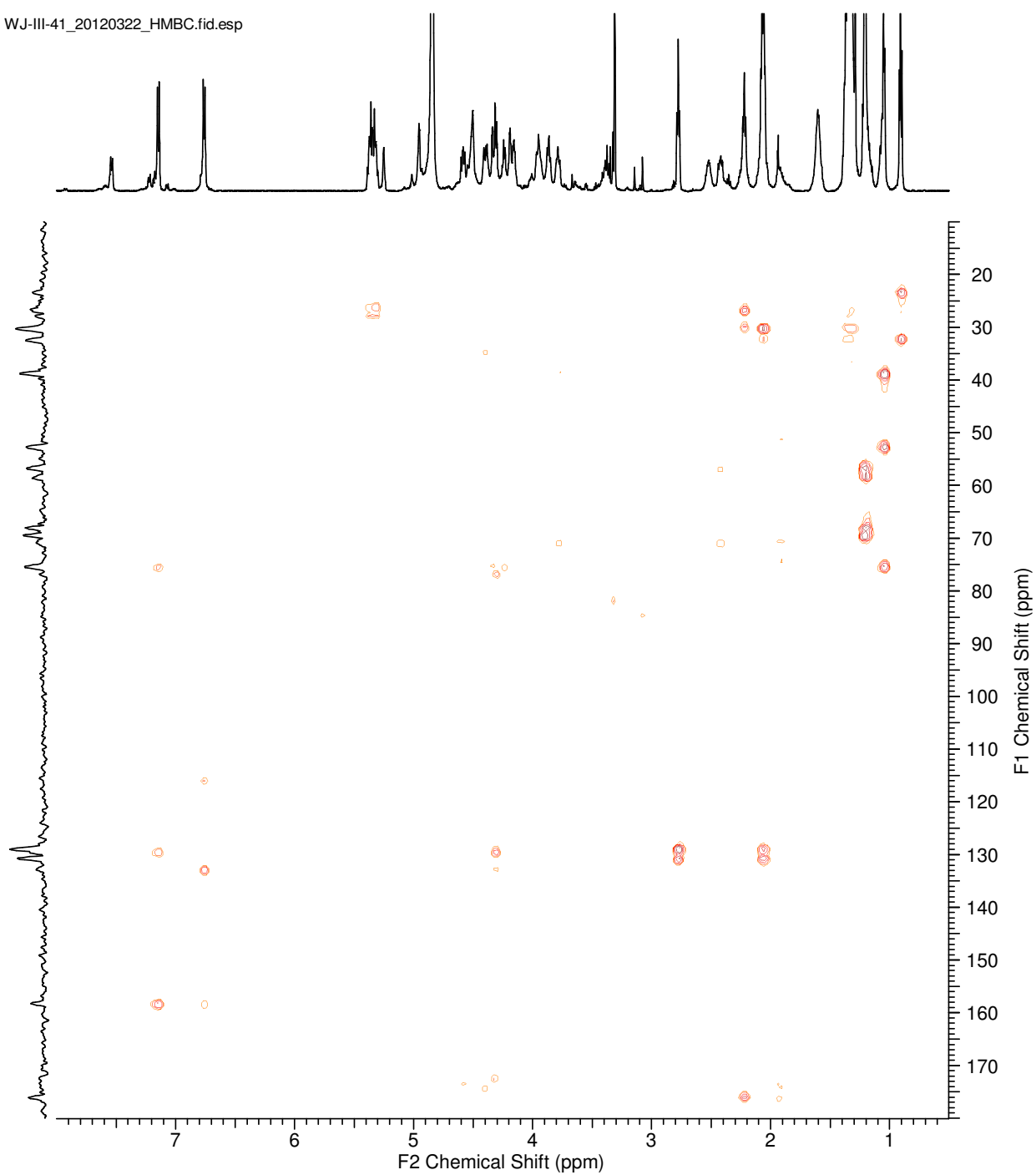


Figure S16: HMBC Spectrum of **1** (600 MHz) in CD₃OD

WJ-III-41_20120322_HMBC.fid.esp



Supplemental Information References:

(1) Hensens, O. D.; Liesch, J. M.; Zink, D. L.; Smith, J. L.; Wichmann, C. F.; Schwartz, R. E. *J Antibiot (Tokyo)* **1992**, *45*, 1875.

(2) Benz, F.; Knüsel, F.; Nüesch, J.; Treichler, H.; Voser, W.; Nyfeler, R.; Keller-Schierlein, W. *Helv. Chim. Acta* **1974**, *57*, 2459.