Supporting Information

Biochemical Analysis of the Biosynthetic Pathway of an Anticancer Tetracycline SF2575

Lauren B. Pickens[1], Woncheol Kim[1], Peng Wang[1], Hui Zhou[1], Kenji Watanabe[2], Shuichi Gomi[3], Yi Tang[1,]*

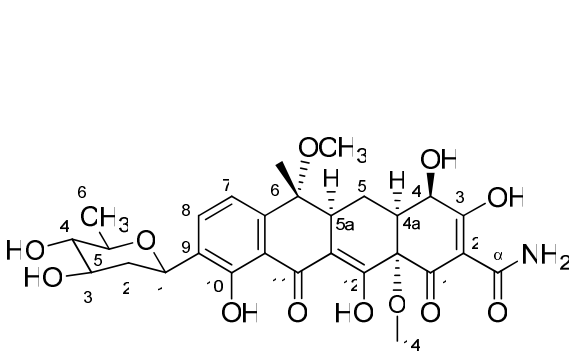
Contents

Table S1: Primers used for amplification of *ssf* genes from cosmids containing genomic DNA encoding the SF2575 gene cluster.

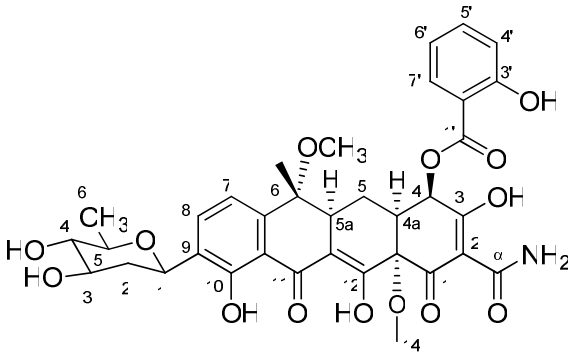
Primers	Sequence[a]
SsfAB_fwd	5'-GG*TTAATTAA*GGAGGAGCCAGC**ATG**CGCGAGGCGCGCCGGGT-3'
SsfAB_rev	5'-GG*TCTAGA***TCA**CGCTGTCTTGCGCACGAT-3'
SsfC_fwd	5'-GG*TCTAGA*GGAGGAGCC*CATATG*TCCGAGTTCGTGATCCA-3'
SsfC_rev	5'-GG*ACTAGT***CTA**GACGCGCTCGGTGATCA-3'
SsfD_fwd	5'-AAA*TCTAGA*GGAGGAGCC*CATATG*TGTGGCATTGCAGGATG-3'
SsfD_rev	5'-TTT*ACTAGT***TCA**CACGTCGAGGGTGACGT-3'
SsfU_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*GAGACCACGAACACGAC-3'
SsfU_rev	5'-AAA*GCTAGC***TCA**GTAGATGCCCAGGCCGCC-3'
SsfO2_fwd	5'-AAGG*ACTAGT*GGAGGAGCACCG**ATG**CAATCACCCGACGACG-3'
SsfO2_rev	5'-AAGG*GCTAGC***TCA**GACCACCAGGTCCTTCT-3'
SsfM1_fwd	5'-AAGG*ACTAGT*GGAGGAGCACCG**ATG**ACCGACGCCGCCGCTATCTC-3'
SsfM1_rev	5'-AAGG*GCTAGC***TCA**GGCCTTGTGGGCGATGG-3'
SsfM3_fwd	5'-AAA*ACTAGT*GGAGGAGCA*CATATG*CCGGACACGGCAGGTGC-3'
SsfM3_rev	5'-AAA*GCTAGC***TCA**TGACGGAAGTCCCTCCTTC-3'
SsfM4_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*ACCAGCACCGACACCGA-3'
SsfM4_rev	5'-AAA*GCTAGC***TCA**GCTCGCCACCGTGCACT-3'
SsfY1_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*TCCACTGGGCAGTCCCGGCA-3'
SsfY1_rev	5'-AAA*GCTAGC***TCA**GCCGGCCTCCACGAACG-3'
SsfY2_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*ACGGTCACGACGGCAGC-3'
SsfY2_rev	5'-AAA*GCTAGC***TCA**GGCGTGGCAGGTGGGCA-3'
SsfY3_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*ACCGGCGCACCGTTCAC-3'
SsfY3_rev	5'-AAA*GCTAGC***TCA**CGCGGCGCCGCCGAGCA-3'
SsfY4_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*CGGTTCCTCGACAGC-3'
SsfY4_rev	5'-AAA*GCTAGC***TCA**CTTGGGCAGGTCGTCCCGGT-3'
SsfL2_fwd	5'-AAA*ACTAGT*GGAGGAGCA*CATATG*GCAACGACAGACTTGACG-3'
SsfL2_rev	5'-AAA*GCTAGC**TCA***GCCCAGCTCGCCGGCCA-3'
SsfL1_fwd	5'-AAA*ACTAGT*GGAGGAGCA*CATATG* GATGAGGGATTCGTGCCC-3'
SsfL1_rev	5'-AAA*GCTAGC**TCA*** TCTGACCAGGTCCCGCA-3'
SsfX3_fwd	5'-AAGG*ACTAGT*GGAGGAAGC*CATATG*ACCACACAGAACACCGC-3'
SsfX3_rev	5'-AAA*GCTAGC***TCA**GCCGCGGACCGGCAGCAC-3'
KS2-LEK-S	5'-*GAATTC*GACGCCATCAAGGCGACGACCGCACG-3'
KS2-LEK-A2	5'-*GGTACC*GCCCCGCACGCGAAGGGCGGCAT-3'
KS2-RPH-S	5'-*CTGCAG*GCGGGCTCGACGCCGTCTCCAAGG-3'
KS2-RPH-A	5'-*AAGCTT*CAGCAGCTGGCCCGGCGTCTTGG-3'

[a]The introduced restriction sites are shown in italics. The optimal ribosomal binding site GGAGG was introduced at the 5' of each gene and is underlined. The start and stop codons are shown in bold.

Table S2: ^{1}H NMR chemical shifts of **6** and **7**. Measured in DMSO-d_6 at 400MHz



6



7

Proton	^{1}H δ (ppm), multiplicity, (J_{HH}(Hz))	^{1}H δ (ppm), multiplicity, (J_{HH}(Hz))
3-OH	-	-
4-H	4.57, d, (5.1)	6.09, d, (5.1)
4-OH	5.88, br	-
4a-H	2.89, m	3.23, m
5-H	2.11, m; 1.30, m	2.18, m; 1.61, m
5a-H	3.32, m	3.46, m
6-OCH3	3.08, s	3.16, s
7-H	6.95, d. (8.0)	6.95, d, (8.0)
8-H	7.67, d. (8.0)	7.66, d, (8.0)
10-OH	12.11, s	12.07, s
12-OH	14.88, s	14.87, s
12a-OCH3	3.4, s	3.49, s
a-NH2	9.17, s	9.46, s
	9.26, s	9.4, s
14-CH3	0.98, s	1.02, s
1'-H	4.72, d, (10.2)	4.70, d, (10.4)
2'-H	2.13, m; 1.36, m	2.09, m; 1.32, m
3'-H	3.50, m	3.52, m
3'-OH	4.92, br	4.93, br
4'-H	2.89, m	2.84, m
4'-OH	4.92, br	4.93, br
5'-H	3.18, m	3.20, m
6'-CH3	1.23, d, (6.12)	1.21, d, (6.1)
3"-OH	-	10.33, s
4"-H	-	7.03, d, (7.8)
5"-H	-	7.55, m
6"-H	-	7.0, d, (7.2)
7"-H	-	7.84, d, (7.9)

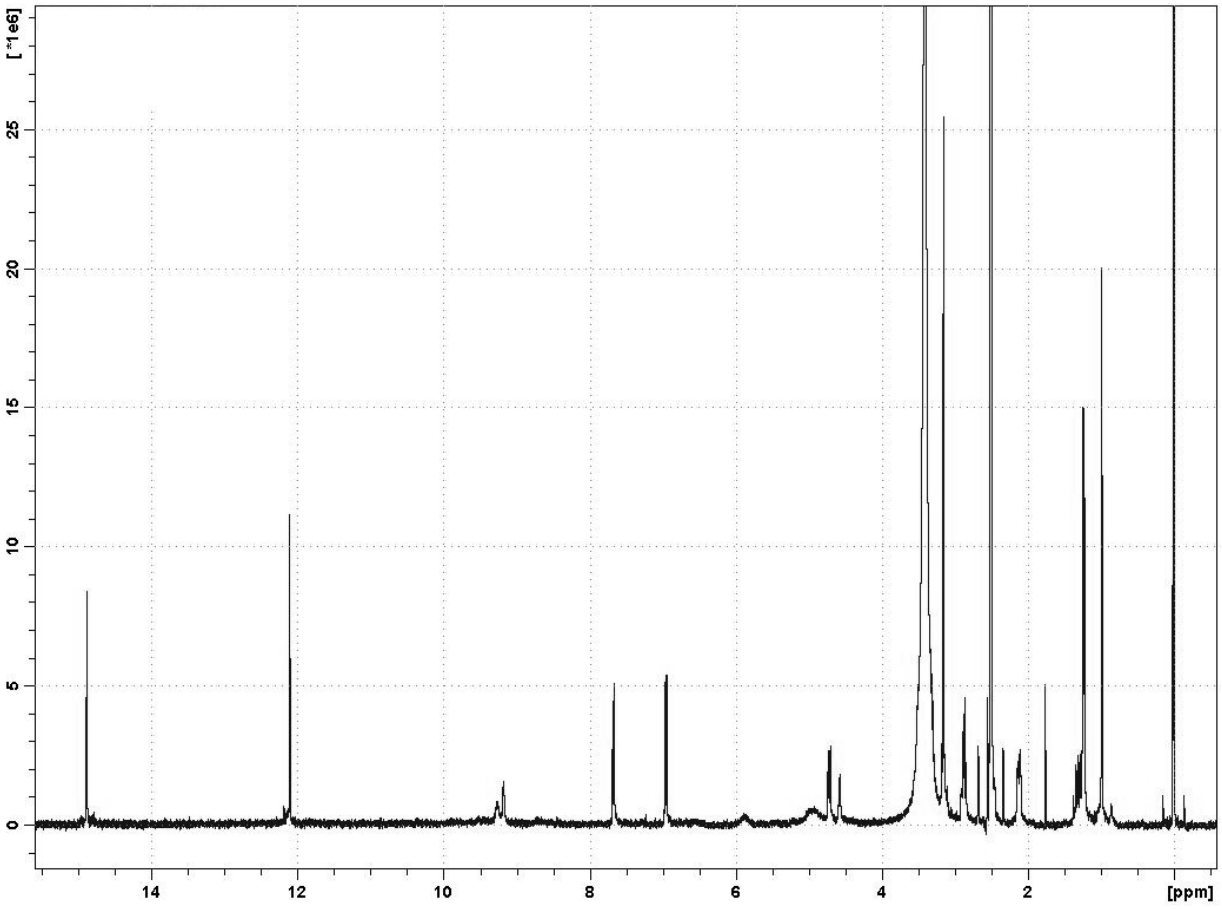


Figure S1: ^{1}H NMR spectrum of **6.** Measured in DMSO-d_6 at 400MHz

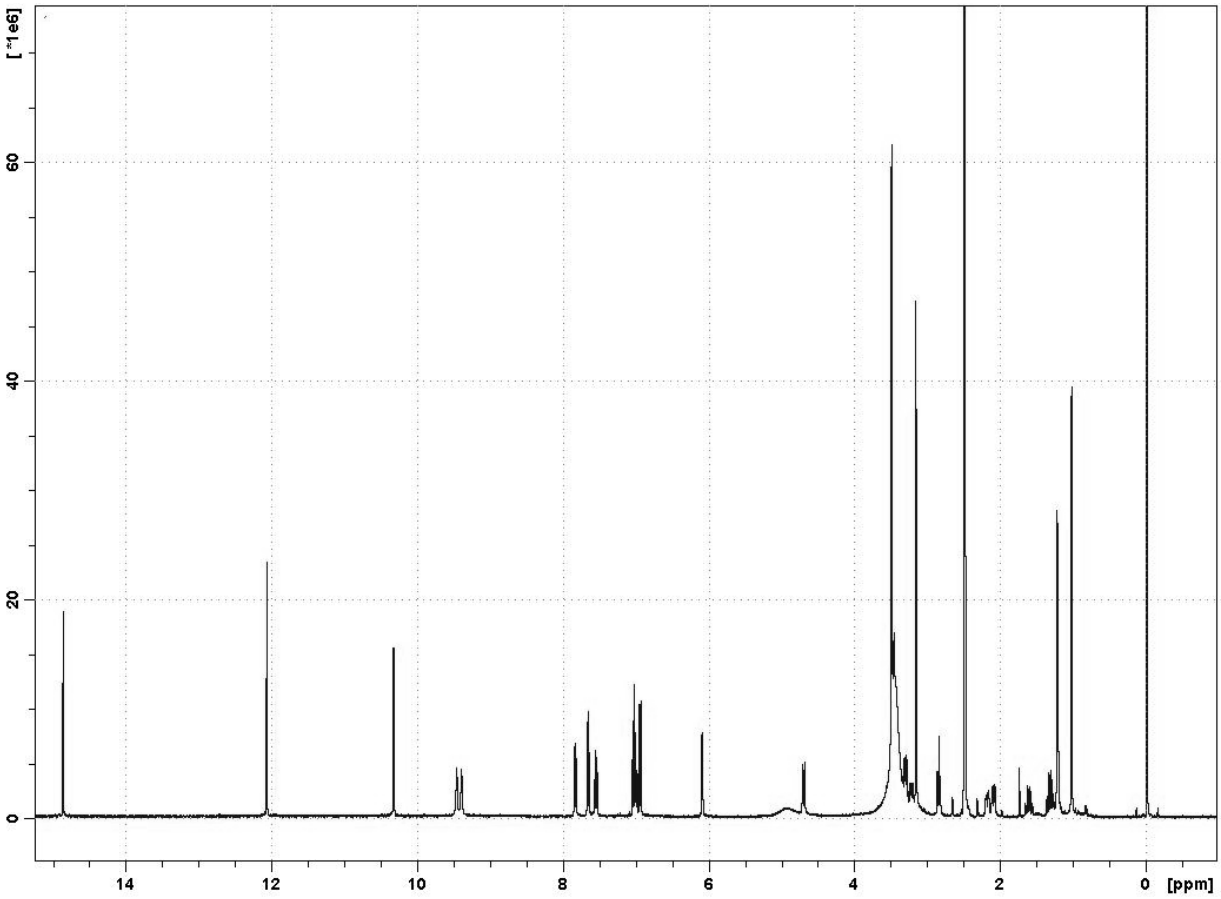
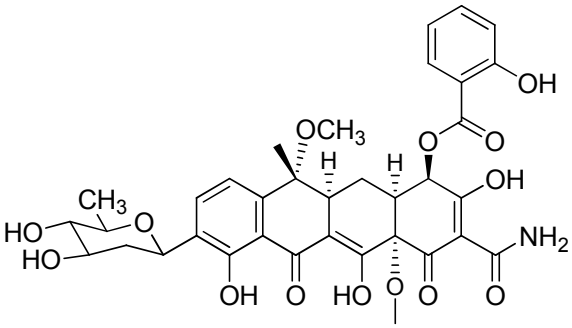
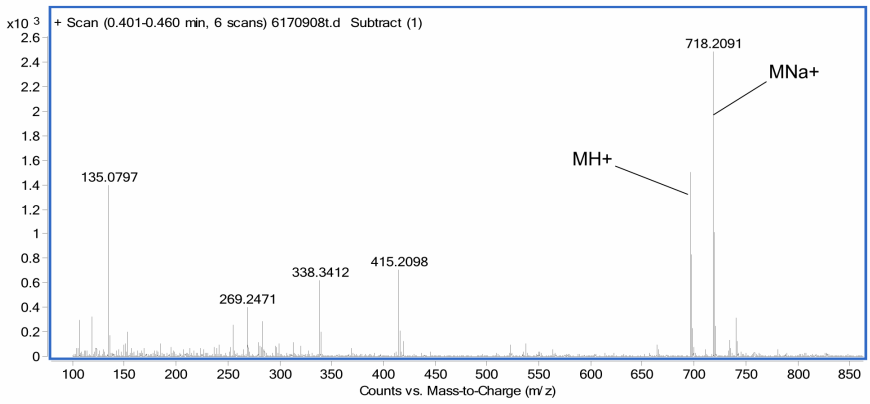
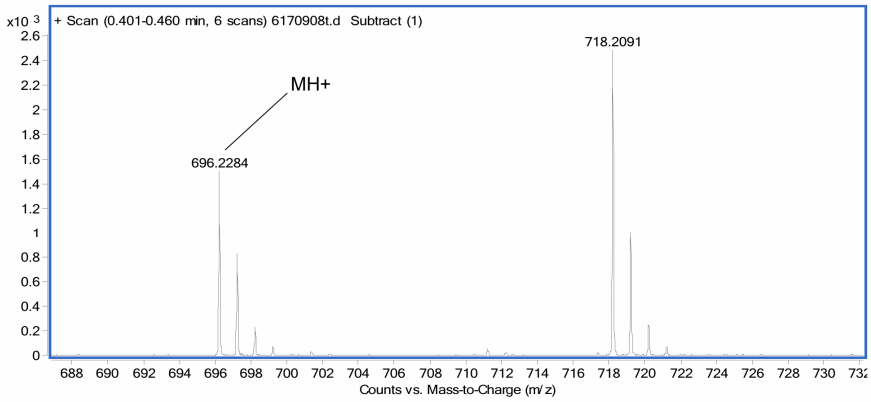


Figure S2: ^{1}H NMR spectrum of **7.** Measured in DMSO-d_6 at 400MHz



Chemical Formula: $C_{35}H_{37}NO_{14}$
Molecular Weight: 695.67

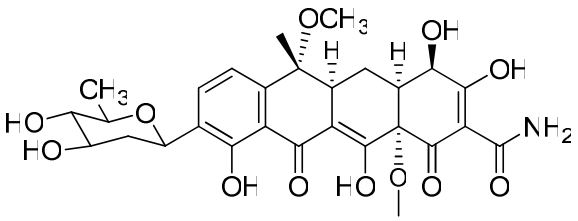




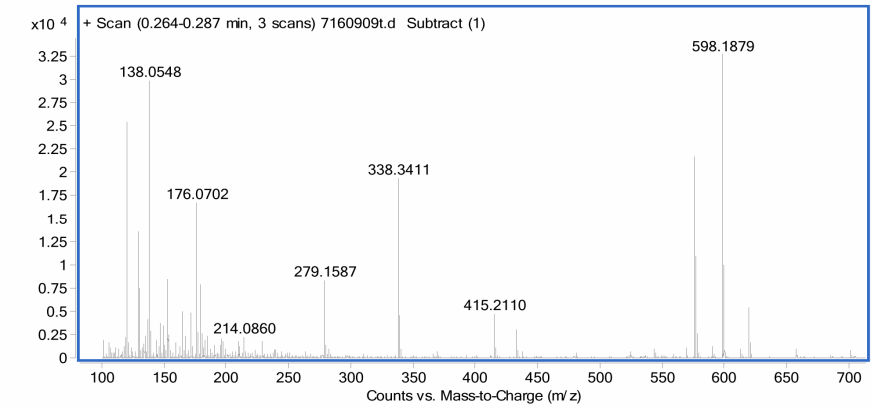
Measured Mass	696.2284	
Element	Low Limit	High Limit
C	30	40
H	30	50
N	0	3
O	12	16

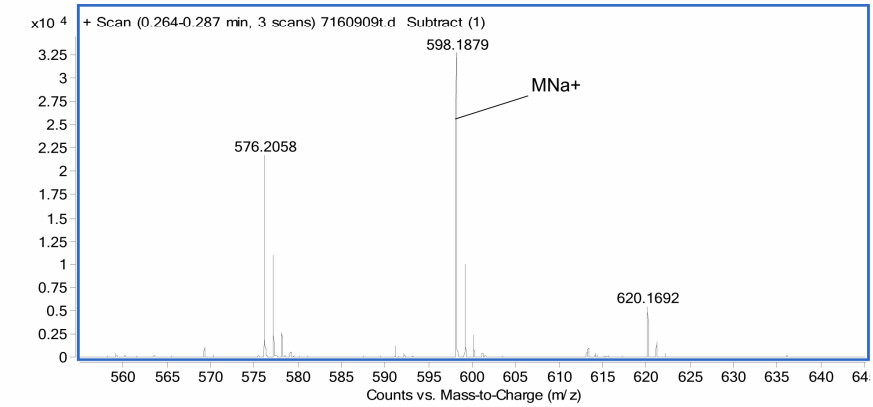
Formula	Calculated Mass	mDaError	ppmError	RDB
C35 H38 N O14	696.2287	-0.3	-0.4	17.5

Figure S3: High accuracy mass spectrometry analysis of semisynthetically prepared **7.**



Chemical Formula: $C_{28}H_{33}NO_{12}$
Molecular Weight: 575.56



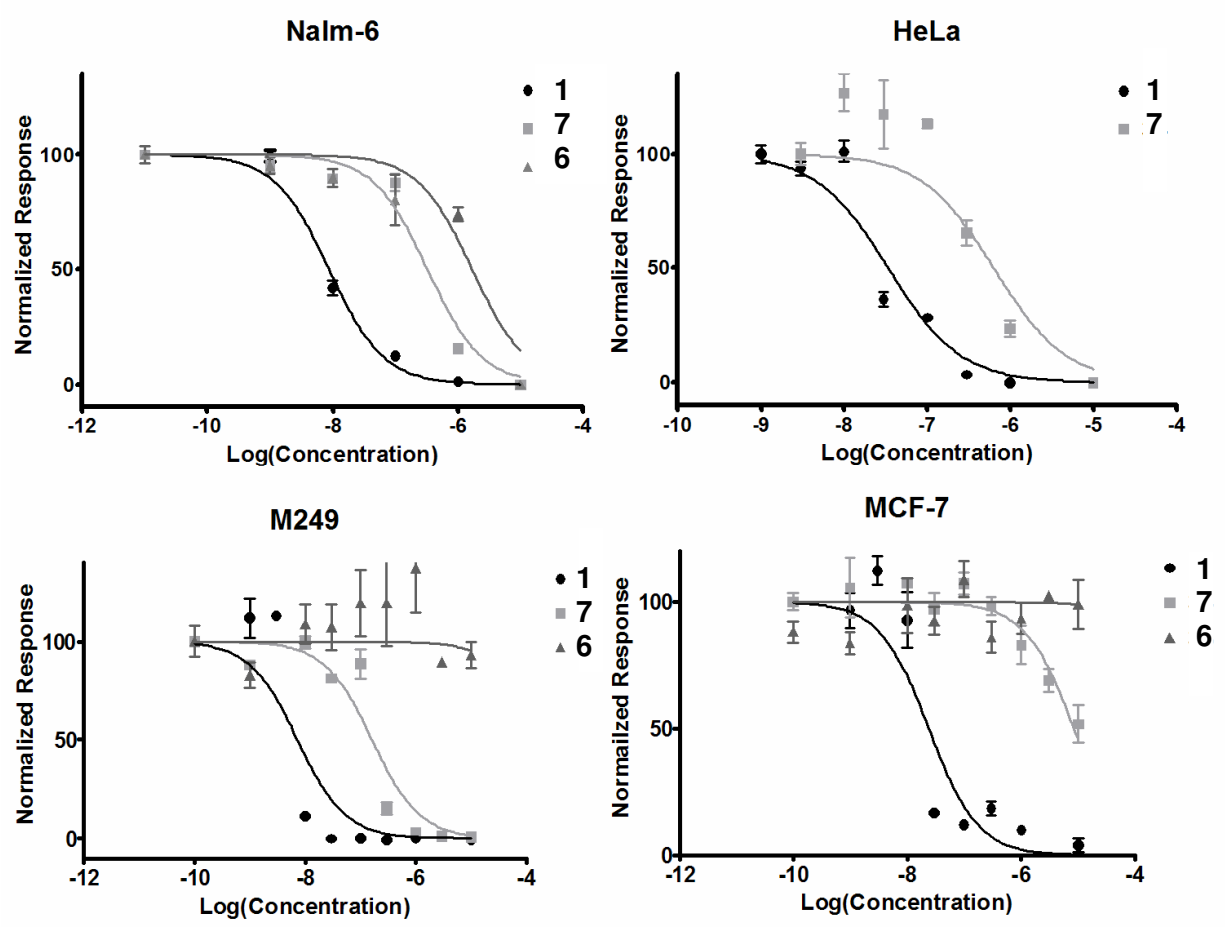


Measured Mass 598.1879

Element	Low Limit	High Limit
C	23	33
H	25	45
N	0	2
O	10	14
Na	0	1

Formula	Calculated Mass	mDaError	ppmError	RDB
C28 H33 N O12 Na	598.1895	-1.6	-2.7	12.5

Figure S4: High accuracy mass spectrometry analysis of semisynthetically prepared **6.**



		Log(IC_{50}) ± Std. Error		
Cell line	Type	**1**	**7**	**6**
Nalm-6	human B cell precursor leukemia	-8.056 ± 0.0518	-6.485 ± 0.0753	-5.285 ± 0.0958
HeLa	human cervix carcinoma	-7.466 ± 0.0769	-6.196 ± 0.153	> -5.0
MCF-7	human breast adenocarcinoma	-7.619 ± 0.112	-5.068 ± 0.0813	> -5.0
M249	human melanoma	-8.134 ± 0.144	-6.788 ± 0.0921	> -5.0

Figure S5: Dose response curve was obtained from cell proliferation experiments. Cells were analyzed 72 hours after treatment with **1**, **7**, or **6**. Data is normalized to the response of untreated cells during the same time period. Adherent cells (HeLa, MCF-7, and M249) were analyzed by MTS assay. Nalm-6 suspension cells were analyzed by direct cell counting. Data was analyzed using Prism software (GraphPad).

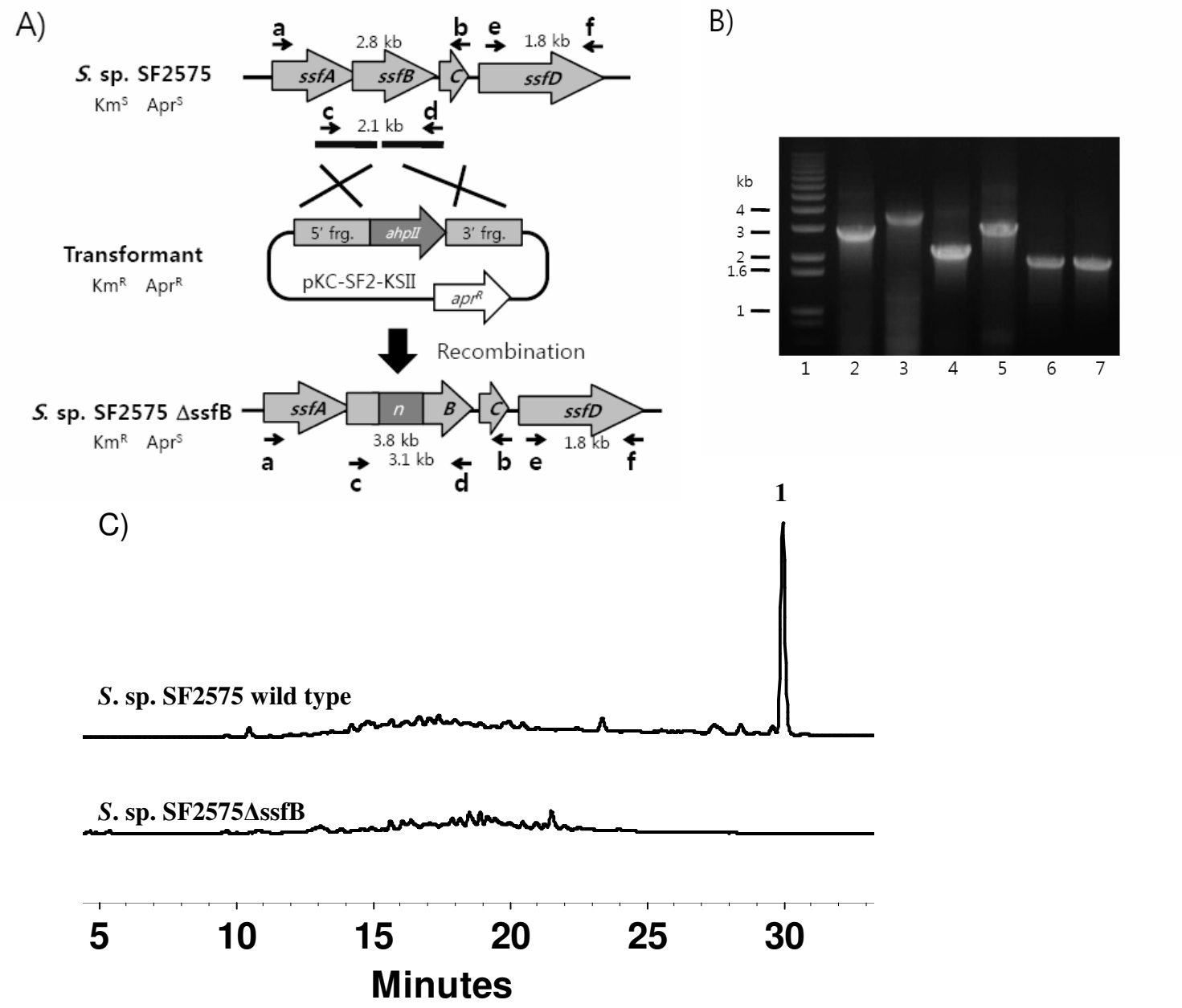


Figure S6: Construction and verification of *ssfB* disruption mutant. A) Wild type *S*. sp. SF2575 were transformed with disruption vector and double-crossover recombinants were screened based on the antibiotic resistance and verified the insertion of neomycin resistance gene by PCR. B) Chromosomal DNA was used as template for both wild type and mutant. 1, 1kb plus ladder; 2, *S*. sp. SF2575 wild type with primer a and b; 3, *S*. sp. SF2575 ΔssfB with primer a and b; 4, *S*. sp. SF2575 wild type with primer c and d; 5, *S*. sp. SF2575 ΔssfB with primer c and d; 6, *S*. sp. SF2575 wild type with primer e and f; 7, *S*. sp. SF2575 ΔssfB with primer e and f. C.) HPLC trace of extracts of *S*. sp. SF2575ΔssfB culture following 7 days of growth on solid Benette's media, shown with wild type *S*. sp. SF2575 for comparison. Disruption of *ssfB* resulted in complete loss of SF2575 production.

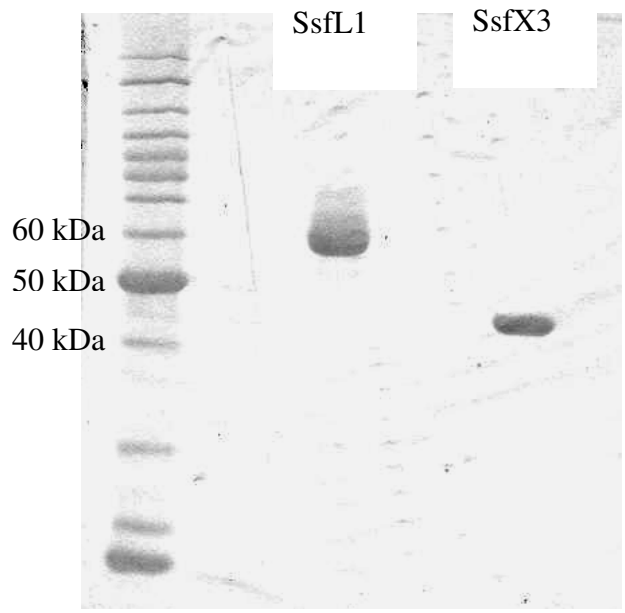
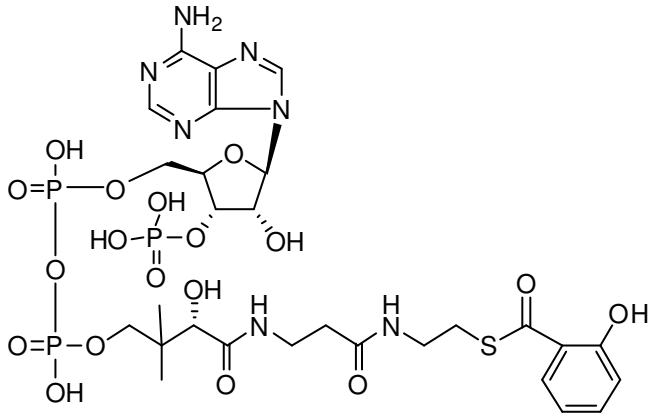


Figure S7: Protein gel showing purified protein SsfL1 (58 kDa) and SsfX3 (40 kDa). Lane 1: Benchmark protein ladder (Invitrogen); lane 2: SsfL1; lane 3: SsfX3.



Chemical Formula: $C_{28}H_{40}N_7O_{10}P_3S$

Molecular Weight = 887.64

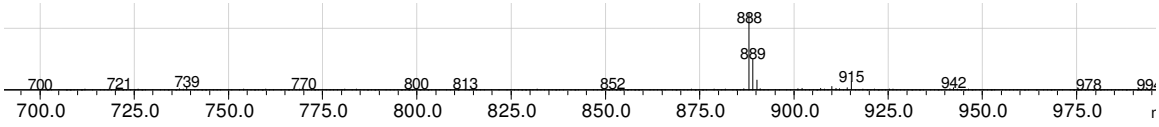


Figure S8: MS spectrum obtained from LCMS analysis of enzymatic formation of salicylyl-CoA by SsfL1 using positive electrospray ionization. [M+H] *m/z* = 888

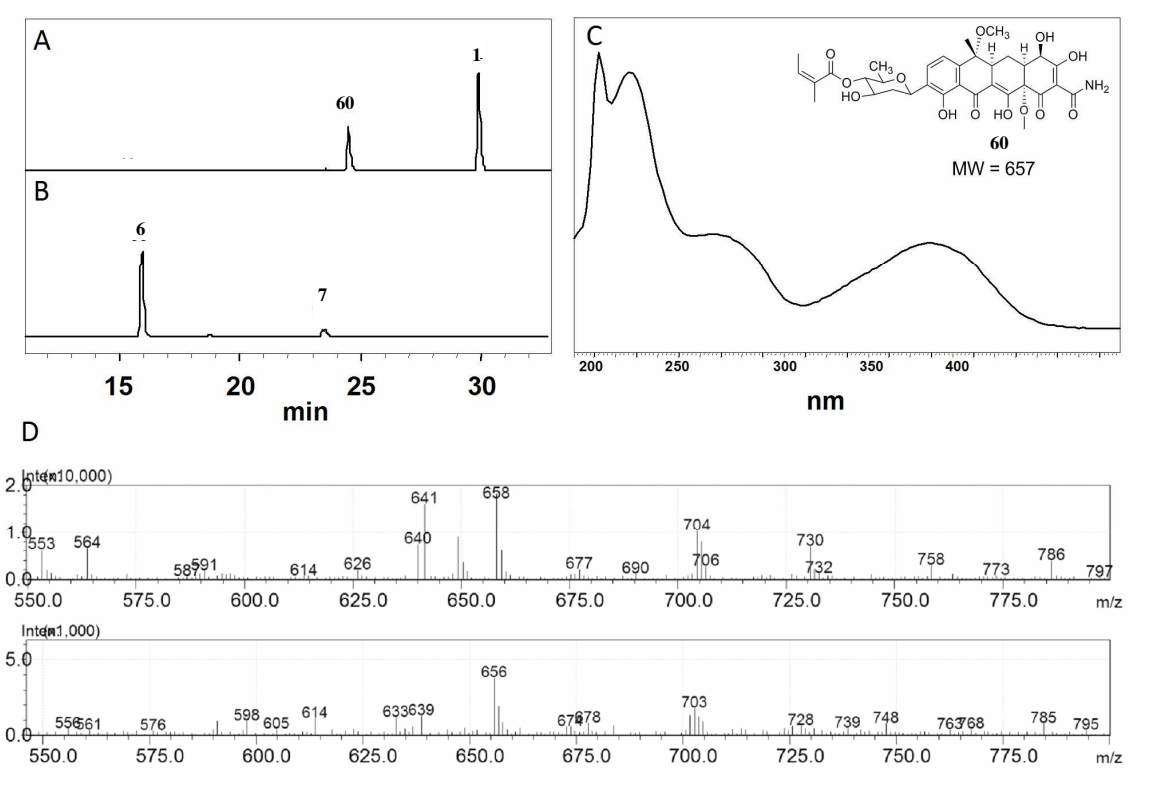


Figure S9: Enzymatic hydrolysis of **1** and **7** by SsfX3. (**A**) Treatment of **1** with SsfX3 resulted in incomplete hydrolysis of **1** and a new peak **60** with MW=657 confirmed by mass spectrometry which corresponds to the loss of salicylic acid. **(B.)** Treatment of **7** with SsfX3 resulted in nearly complete conversion of **7** to **6** as confirmed by mass spectrometry and comparison to semisynthetic standards. Reaction mix includes 50 mM HEPES, pH 7.9, 10 mM $MgCl_2$ 5 μM SsfX3, and 20 μM of either **1** or **7**. All reactions were incubated at 25°C overnight, extracted with organic solvent and analyzed by HPLC (358 nm). (**C**) UV spectra of **60** formed from enzymatic hydrolysis of **1**. (**D**) LCMS mass spectrometry data **60.** Mass spectrometry data was obtained by electrospray ionization ion scan in both positive and negative mode. [M+H] *m/z* = 658; [M-H] *m/z* = 656.