SUPPORTING INFORMATION

Enzymatic Synthesis of Aromatic Polyketides Using PKS4 from *Gibberella fujikuroi*

Suzanne M. Ma,[1] Jixun Zhan,[1] Kenji Watanabe,[2] Xinkai Xie,[1] Wenjun Zhang,[1] Clay C. Wang,[2] Yi Tang*[1]

I. Experimental Details

Materials. ^{14}C Radiolabeled acyl-CoAs were purchased from American Radiolabeled Chemicals. Nonradiolabeled CoA and acyl-CoAs were purchased from Sigma. All other chemicals were from standard sources. Thin-layer chromatography (TLC) plates (Silica Gel IB2-F) were from J. T. Baker. *G. fujikuroi* was purchased from DSMZ (DSM 893, ATCC 12616).

Primers used in this study. Restriction sites are underlined. The introduced restriction sites created silent mutations that do not change the amino acid sequence of the PKS4 protein.

EXON1-5	5' AACATATGGCGTCCTCAGCAGATGTGTATG 3'
EXON1-3	5' CCTTCGCGCAGTGACAGCAGCGGTCAATGTAGTATCCAATCTCAT 3'
EXON2-5	5' ATGAGATTGGATACTACATTGACCGCTGCTGTCACTGCGCGAAGG 3'
EXON2-3	5' CCTTAAGTGCTTCTGAAACAAGCTCCTCCTG 3'
EXON3-5	5' ACTTAAGGAGTTCTCCAAGGAAAAAAATCTTACATACAGTTCGCGTCCCT 3'
EXON3-3	5' GAGTACTAGGAGCCATAGTCAATCCGTTG 3'
EXON4-5	5' TAGTACTCTCTTCGCCGACATCGCCTTC 3'
EXON4-3	5' TCAGTGGTGGTGGTGGTGGTGGATGAATTTACGCCCATTGTCTCCCGG 3'

Molecular Biology. Genomic DNA from *G. fujikuroi* was used as the template for PCR amplification. *E. coli* XL1-Blue and *E. coli* TOPO10 were used for cloning following standard recombinant DNA techniques. *E. coli* BL21(DE3) was used for protein expression. The four exons of the *pks4* gene were amplified from *G. fujikuroi* genomic DNA. The uninterrupted *pks4* gene were assembled using a combination of splicing by overlap extension PCR and ligation via engineered restriction sites that were introduced using silent mutations. Primers EXON1-5 and EXON1-3 were used to amplify exon1 of the *pks4* gene with an *Nde*I restriction site preceding the gene. Primer pair EXON2-5 and EXON2-3 was used to amplify exon2 of the *pks4* gene containing an engineered *Afl*II restriction site at the 3' of the fragment, and to simultaneously remove an *EcoR*I restriction site. The PCR products from these two reactions were combined to perform splice by overlap extension PCR with primer pair EXON1-5 and EXON2-3 to create a DNA cassette containing exon1 and exon2 as a single open reading frame flaked by restriction sites *Nde*I and *Afl*II. Primer pair EXON3-5 and EXON3-3 was used to amplify exon3 of *pks4* with restriction sites *Afl*II and *Sca*I flanking the gene. Exon4 was amplified with primers EXON4-5 and EXON4-3 to amplify the final exon and to introduce a *Sca*I restriction site preceding, and a hexahistidine tag immediately following the gene cassette. Each gene cassette amplified by PCR was then placed into the pCR blunt vector separately. To construct the entire *pks4* gene in pCR blunt, restriction sites *EcoR*V, from the pCR blunt vector and *Afl*II, from the amplified cassette was utilized to combine exon1-exon2, with exon3. Restriction sites *Sca*I, engineered to precede

exon4, and *EcoR*I, from the pCR blunt vector was used to place exon4 after exons1-exon3 to recreate the uninterrupted open reading frame of PKS4 flanked by *Nde*I and *EcoR*I. The entire *pks4* gene was then inserted into pET24 between *Nde*I and *EcoR*I to create pSMa76, the expression plasmid for *pks4* with a C-terminus hexahistindine tag.

Expression and Purification of PKS4. Following DNA sequence confirmation of *pks4*, pSMa76 was transformed into *E. coli* BL21(DE3) strain for protein expression. For 1 L of liquid culture, the cells were grown at 37°C in LB medium with 35 μg/mL kanamycin to an OD_{600} of 0.4. At which time the cells were incubated on ice for 10 minutes, and then induced with 0.1 mM isopropyl thio-β-D-galactoside (IPTG) for 16 hours at 16°C. The cells were harvested by centrifugation (3500 rpm, 10 minutes, 4°C), resuspended in 30 mL lysis buffer (20mM Tris-HCl pH = 7.9, 0.5 M NaCl, 10 mM imidazole) and lysed using sonication on ice. Cellular debris was removed by centrifugation (15000 g, 1 hour, 4°C). Ni-NTA agarose resin was added to the supernatant (1 mL/L of culture) and the solution was stirred at 4°C for at least 2 hours. The protein resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in Buffer A (50 mM Tris-HCl, pH = 7.9, 2 mM EDTA, 2 mM DTT). Purified proteins were concentrated and buffered exchanged into Buffer A + 10% glycerol with Centriprep filter devices (Amicon Inc.). The final PKS4 enzyme was concentrated to 8 mg/mL, aliquoted and flash frozen. Protein concentrations were determined with the Bradford assay using BSA as a standard.

Labeling of PKS4. Labeling of PKS4 were performed in Buffer L (100 mM NaH_2PO_4, pH = 7.4, 2 mM DTT, 10% glycerol). Enzyme (10 μM) and radiolabeled 2-^{14}C-malonyl-CoA (180 μM, 55 mCi/mmol) were added to Buffer L (final volume 10 μL) and were incubated at room temperature for 10 minutes. The reaction was quenched with 1 volume of SDS-PAGE loading buffer lacking any reducing reagents such as DTT or β-mercaptoethanol. Samples were directly loaded onto a 6% SDS-PAGE gel and electrophoresis was performed at 100 V for 90 minutes. The gel was dried and analyzed using a phosphoimager (Packard InstantImager).

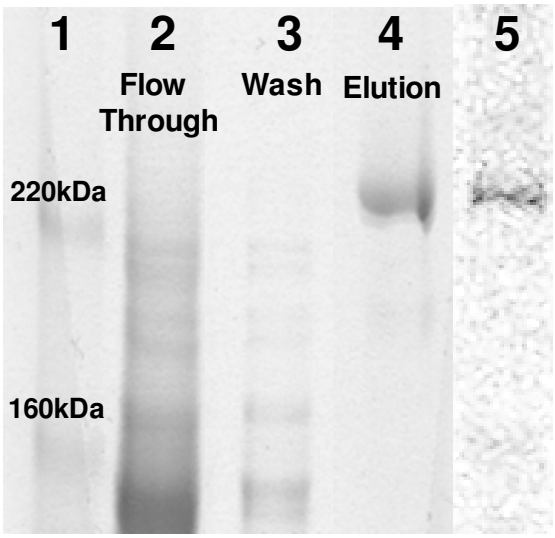


Figure S1. Expression, purification (lanes 2-4) of PKS4. Labelling of PKS4 with [2-^{14}C]malonyl-CoA (lane 5)

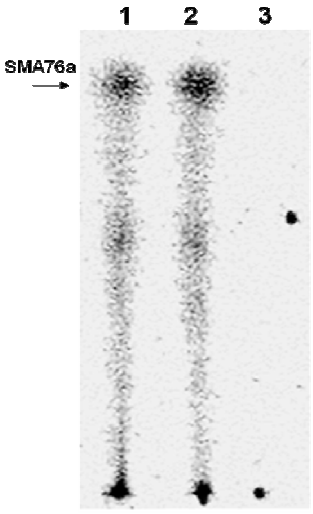


Figure S2. Radio-TLC of products synthesized by PKS4 purified from BL21(DE3)/pSMA76 (lane 1); BAP1/pSMA76 (lane 2); and no PKS4 (lane 3). 2-^{14}C-malonyl-CoA (180 μM, 55 mCi/mmol) was used in all samples. BAP1 contains a chromosomal copy of the broadly specific ppant transferase, Sfp. The dominant product SMA76a (**2**) is shown with an arrow. The production of the expected product by PKS4 purified from BL21(DE3) shows that the ACP domain of PKS4 is correctly phosphopantethienylated by endogenous *E. coli* enzymes, most likely by holo-ACP synthase.

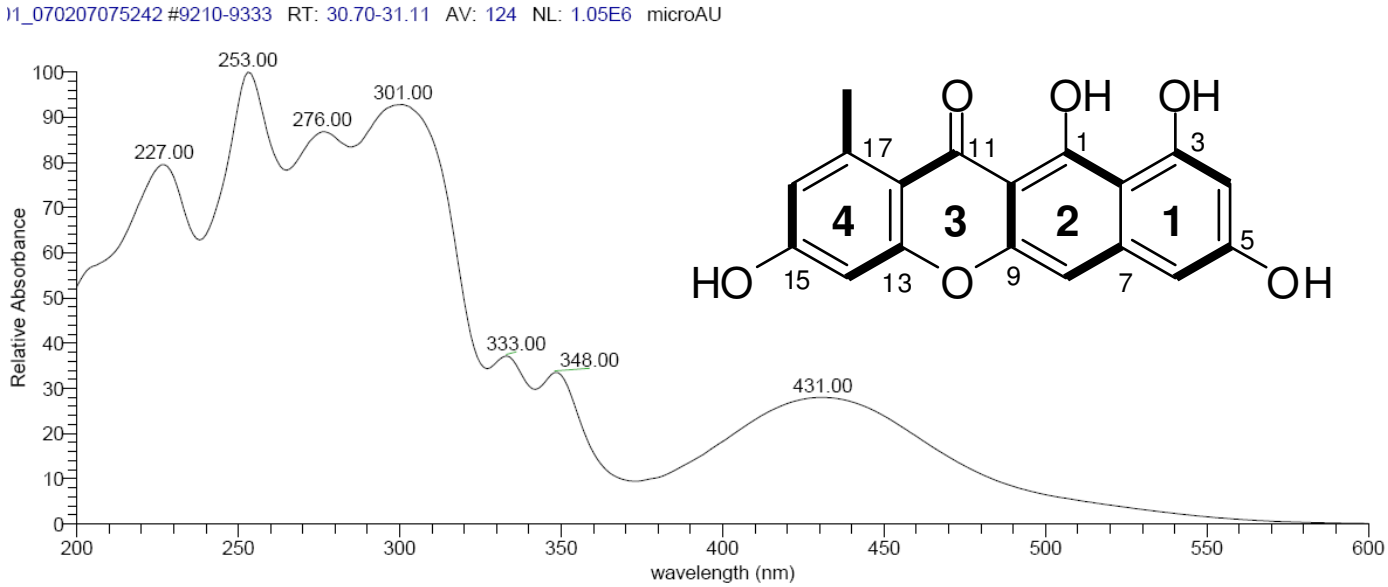
Expression of MatB. The expression plasmid for MatB was transformed into the *E. coli* BL21(DE3) strain for protein expression. For 1 L of liquid culture, the cells were grown at 37°C in LB medium with 100 μg/mL ampicillin to an OD_{600} of 0.4. At which time the cells were incubated on ice for 10 minutes, and then induced with 0.1 mM

isopropyl thio-β-D-galactoside (IPTG) for 16 hours at 16℃. The cells were harvested by centrifugation (3500 rpm, 10 minutes, 4℃), resuspended in 30 mL lysis buffer and lysed using sonication on ice. Cellular debris was removed by centrifugation (15000 g, 1 hour, 4℃). Ni-NTA agarose resin was added to the supernatant and the solution was stirred at 4℃ for at least 2 hours. The protein resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in Buffer A. Purified proteins were concentrated and buffered exchanged into Buffer A + 10% glycerol. The final MatB enzyme was concentrated to 40 mg/ml, aliquoted and flash frozen.

Polyketide Turnover Assay. Example given for a time course analysis: 10 μM of PKS4 was incubated with 2 mM [2-^{14}C]Malonyl-CoA (4.5 mCi/mmol) in Buffer R (100 mM NaH_2PO_4, pH 7.4, 10% glycerol, 2mM DTT) at room temperature. Individual aliquots (20 μL) were removed at different time points and were quenched with 250 μL of 99% ethyl acetate (EA)/1% acetic acid (AcOH). The organic phase was separated, evaporated to dryness, redissolved in 20 μL of EA and the reaction mixture was separated by TLC with 99% EA/1% AcOH as the mobile phase. The resultant TLC plate was imaged and quantified by phosphoimager (InstantImager, Packard). For the starter unit assays, alkylacyl-CoAs were each added to a final concentration of 2 mM. For LC-MS analysis, the same reaction mixture (100 μL) was prepared with cold malonyl-CoA (2 mM). The organic residue were redissolved in methanol and analyzed by LC-MS. LC–MS was conducted with a Finnigan LCQ Deca XP quadrupole ion trap mass spectrometer using negative electrospray ionization and a Waters 2.1 x 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5 to 95% CH_3CN (v/v) over 30 min and 95% CH_3CN (v/v) further 30 min in H_2O supplemented with 0.05% (v/v) formic acid at a flow rate of 0.125 ml/min at room temperature. LC retention times (t_R) were as follows: SMA76a (**2**): 30.9 min; SMA76b (**3**): 30.4 min; SMA76c (**4**): 33 min.

Large Scale Synthesis and NMR. During the scaled-up reaction, malonyl-CoA was made *in situ* with the MatB enzyme. In a typical reaction, the following reagents were added to Buffer R (10~30 mL): 100 mM sodium malonate, 7 mM $MgCl_2$, 20 mM ATP, 5 mM coenzyme A, 20 μM MatB, and 10 μM PKS4. Octanoyl-CoA was added to a final concentration of 1 mM for the synthesis of **3** and **4**. The reaction mixture was shaken gently at room temperature and the reaction progress was followed by HPLC. After the product level had maximized, the reaction mixture was extracted three times with 99% EA/1% AcOH. The resultant organic extracts were combined and evaporated to dryness, redissolved in methanol and purified by reverse-phase HPLC (Alltech Apollo 5u, 250 mm X 4.6 mm) on a linear gradient of 5 to 95% CH_3CN (v/v) over 15 min and 95% CH_3CN (v/v) further for 15 min in H_2O supplemented with 0.1% (v/v) trifluoroacetic acid at a flow rate of 1 mL/min. NMR of **2** was performed on a Varian 400 spectrometer using DMSO-d_6 as solvent. NMR of **3** and **4** were performed on Bruker DRX-500 spectrometer using CD_3OD as the solvent.

SMA76a (**2**) UV Spectrum Measured during LC

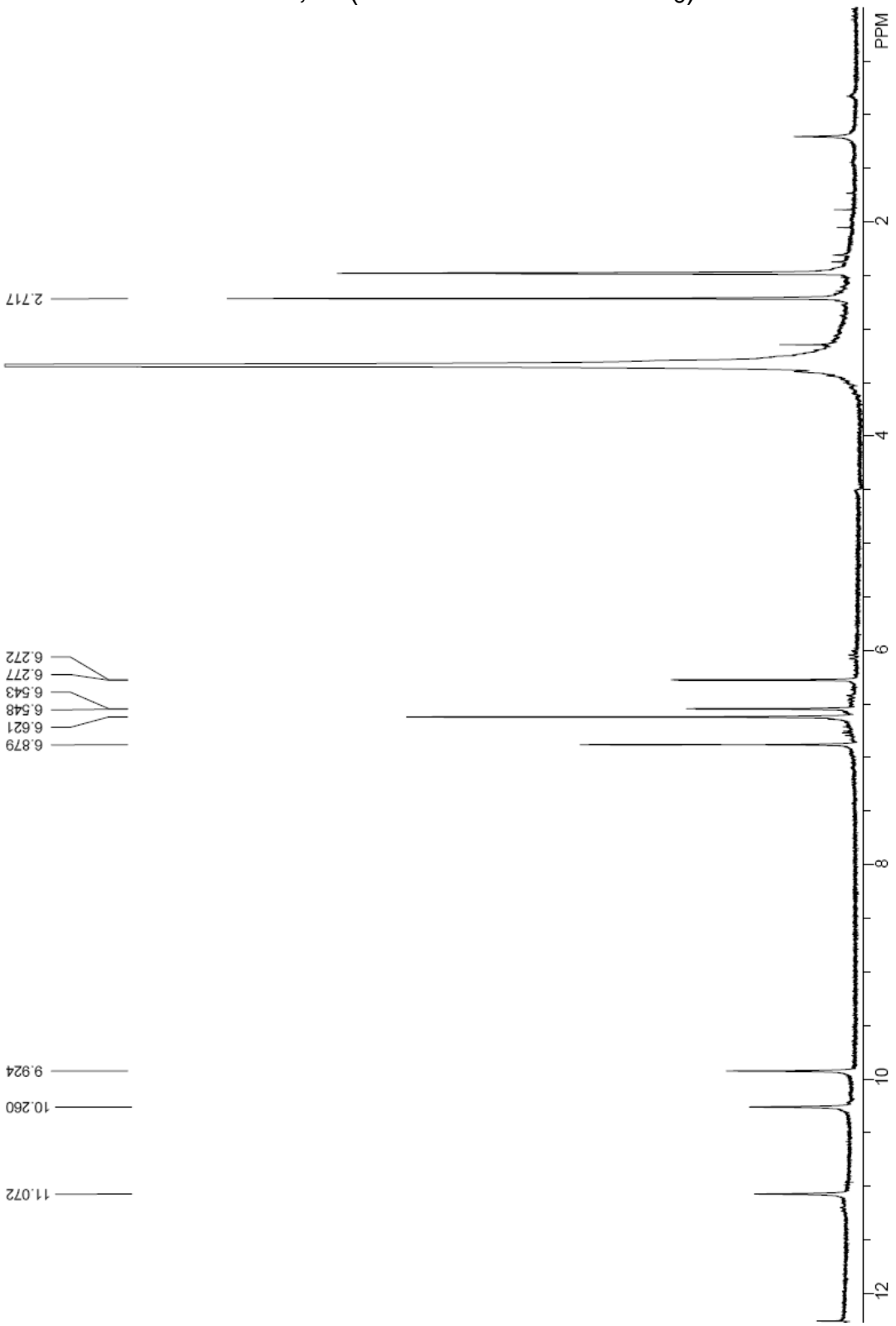


SMA76a (2): HR-ESIMS: *m/z* =323.0550 ($C_{18}H_{11}O_6$ [M -H]$^-$, 323.0556 calculated)

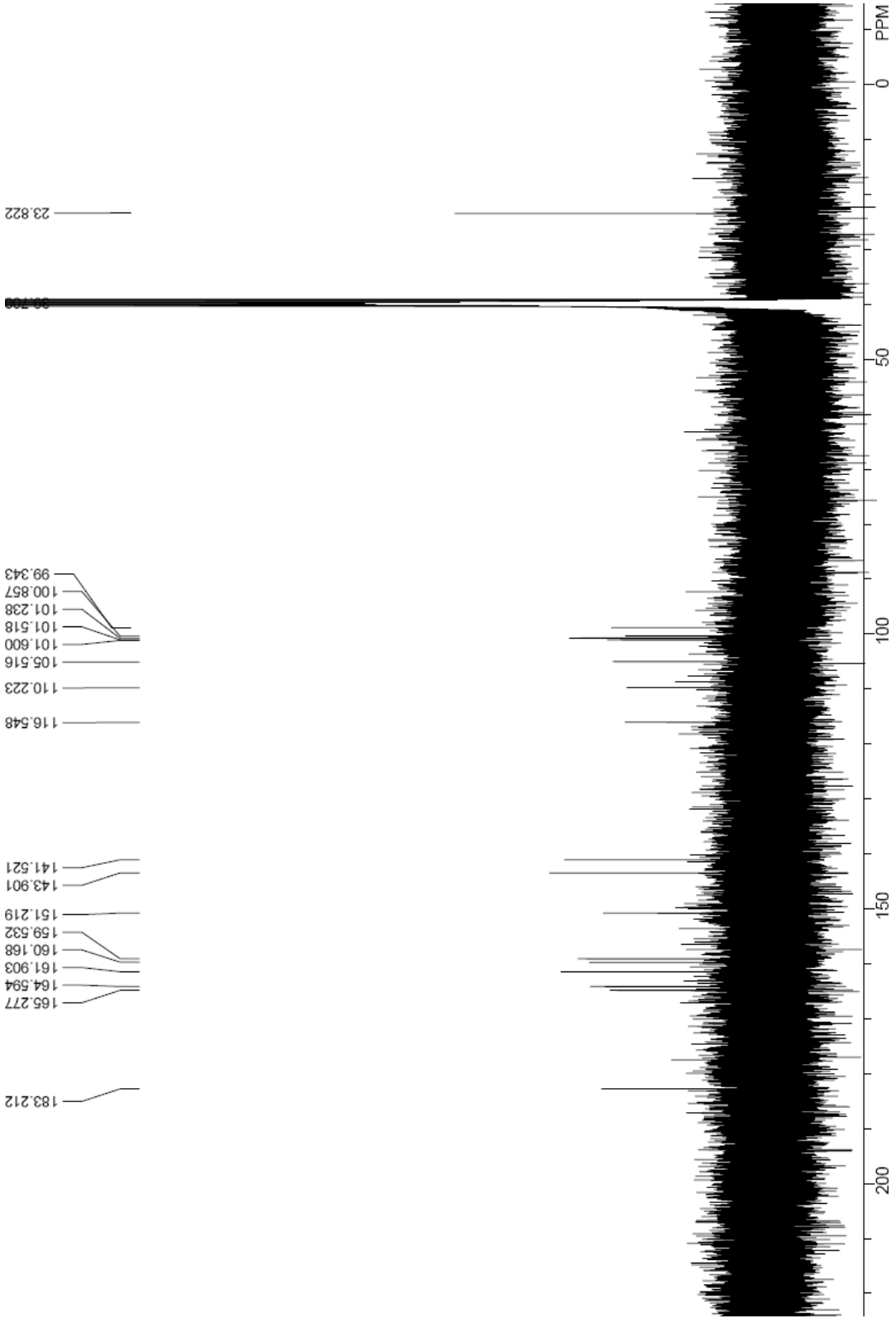
NMR Spectra of **2** were obtained at 400 MHz for proton and 100 MHz for carbon and were recorded in DMSO-d_6

no	^{13}C δ(ppm)	^{1}H δ(ppm) (m, area, J_{HH} (Hz))	Heteronuclear Multiple Bond Correlation (HMBC)
1	165.3	-	-
2	105.5	-	-
3	159.5	10.26 (s, 1H), OH	C2, C4
4	100.9	6.27 (d, 1H, 2.0)	C2, C3, C5, C6
5	161.9	9.92 (s, 1H), OH	C4, C6
6	101.6	6.54 (d, 1H, 2.0)	C2, C4, C5, C8
7	141.5	-	-
8	99.3	6.88 (s, 1H)	C2, C6, C9
9	151.2	-	-
10	101.5	-	-
11	183.2	-	-
12	110.2	-	-
13	164.6	-	-
14	101.2	6.62 (s, 1H)	C12, C13, C15, C16
15	160.2	11.07 (s, 1H), OH	C14, C16
16	116.5	6.62(s, 1H)	C12, C14, C15, C17
17	143.9	-	-
18	23.8	2.72 (s, 3H)	C12, C16, C17

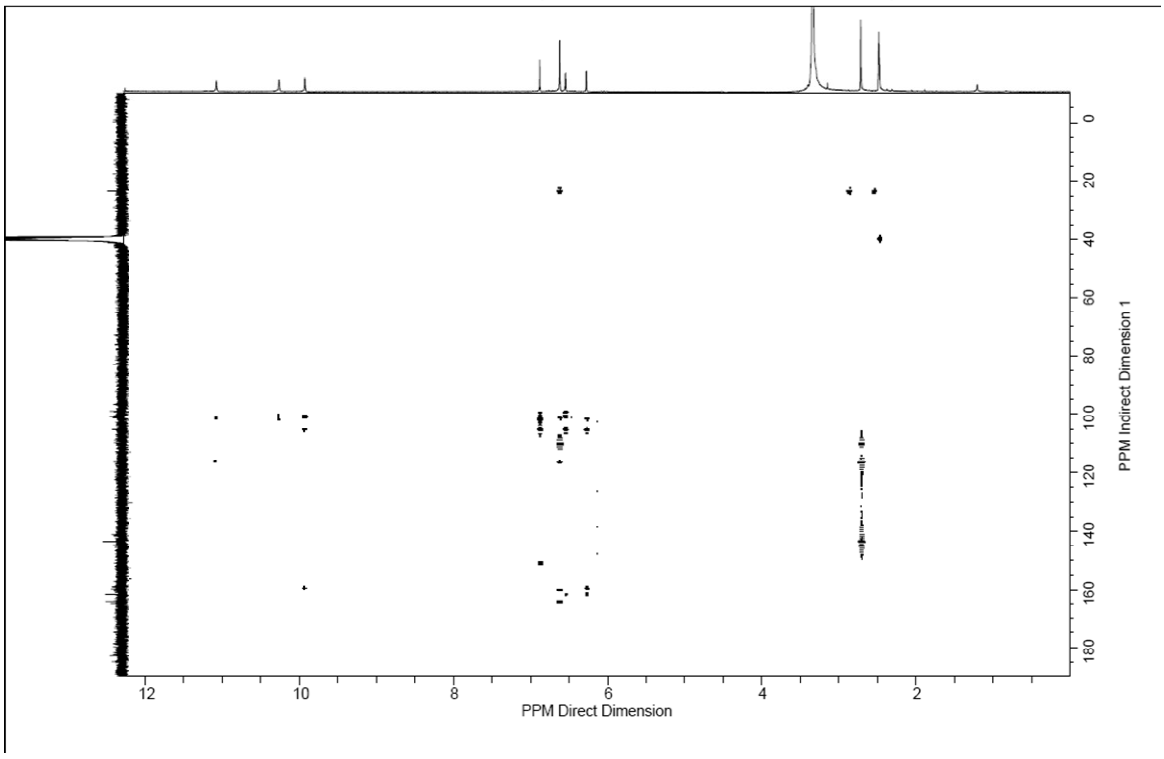
SMA76a, **2** (^{1}H NMR in DMSO-d_6)

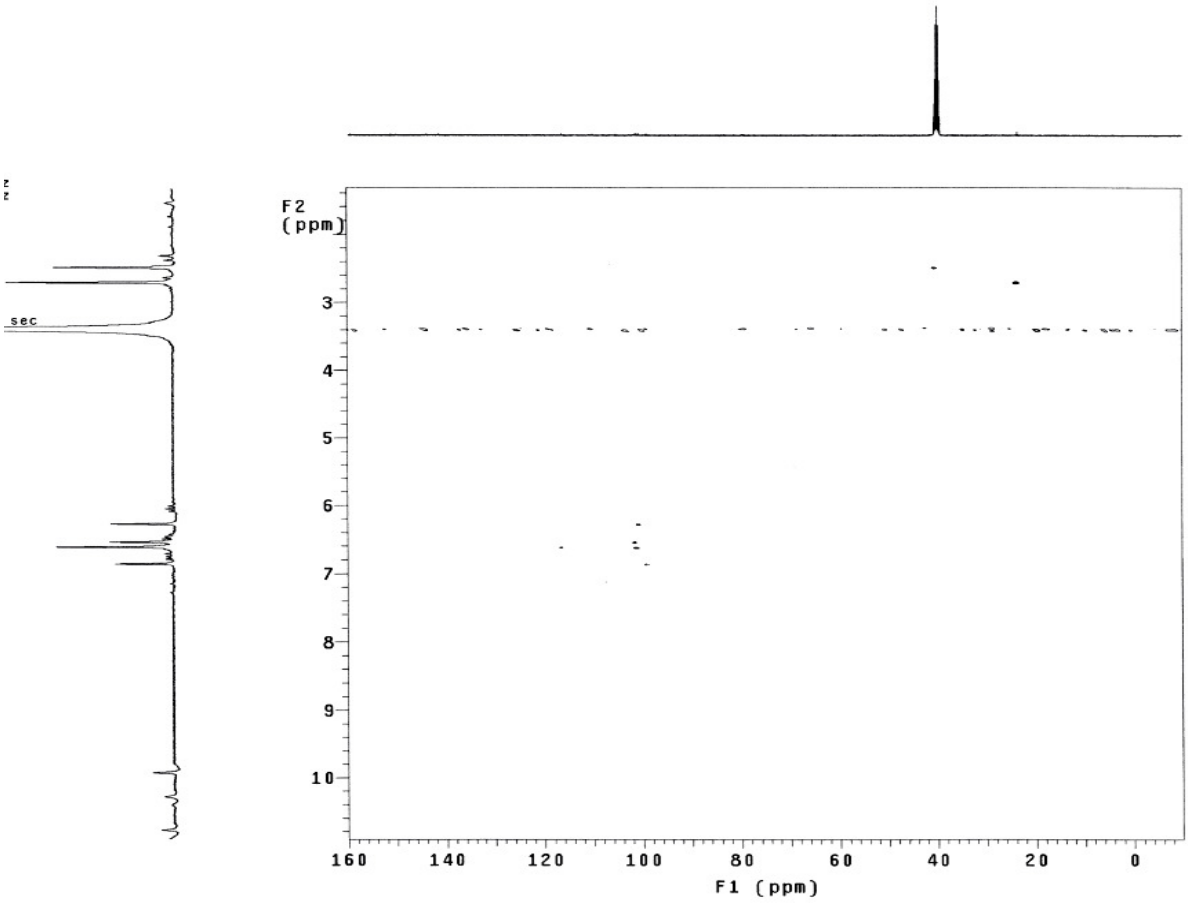


SMA76a, **2** (^{13}C NMR in DMSO-d_6)

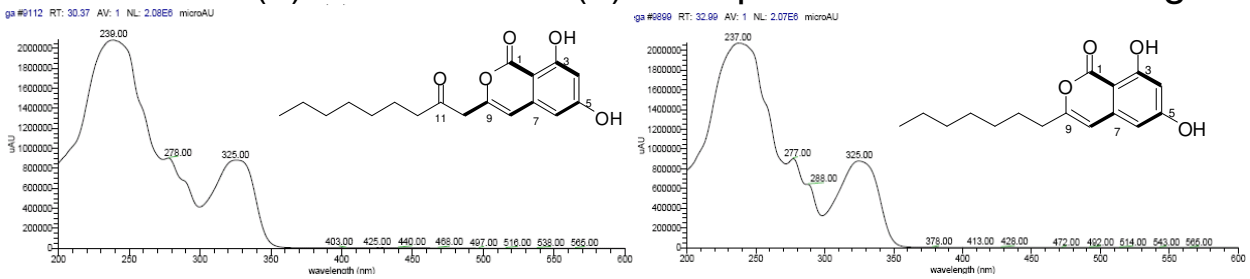


SMA76a, **2** (HMBC and HSQC)

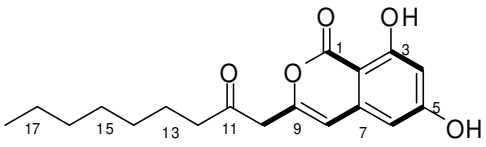
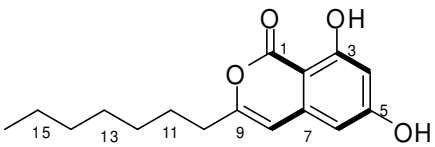




SMA76b (**3**) and SMA76c (**4**) UV Spectra Measured during LC

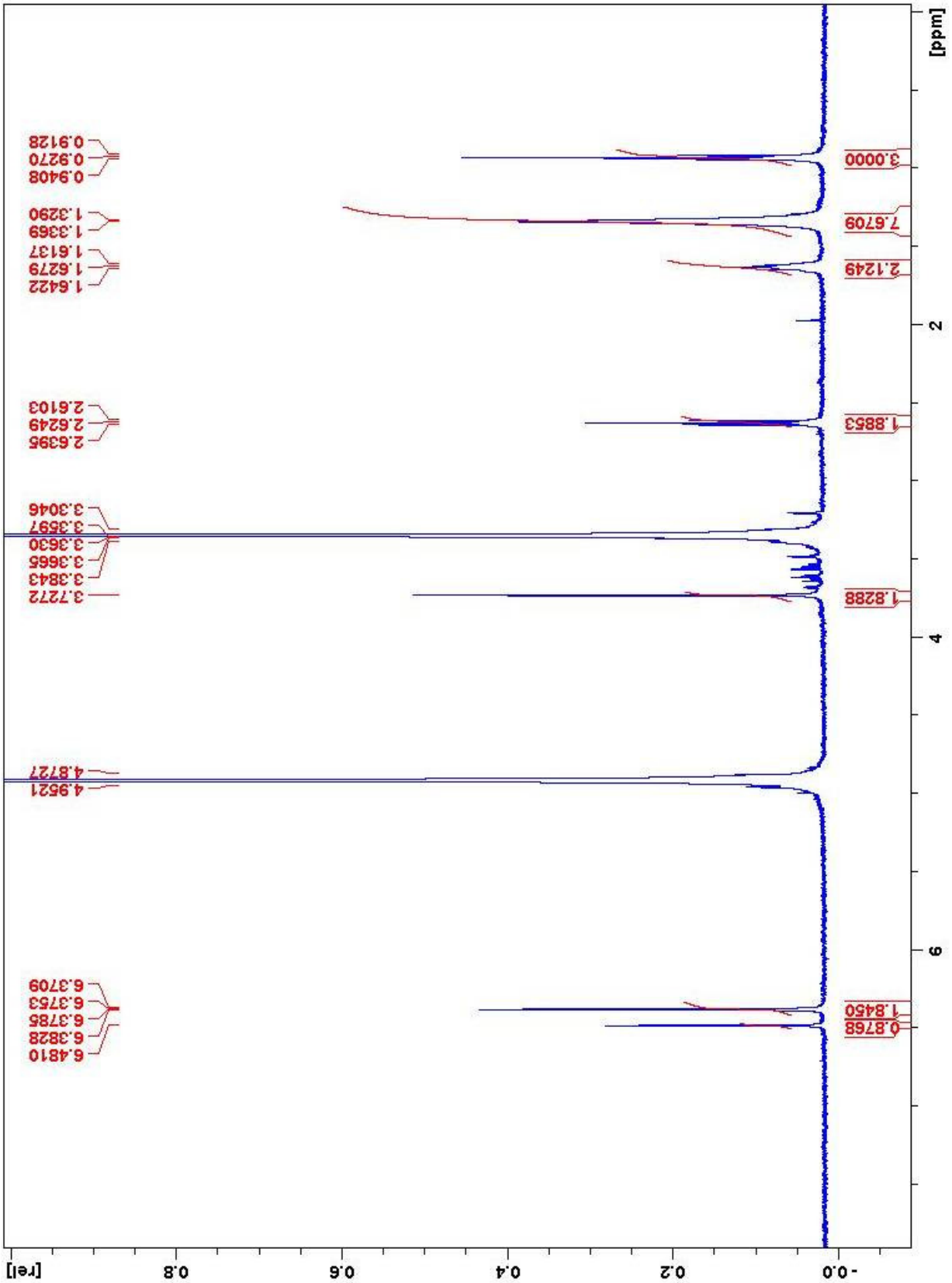


NMR Spectra of **3** and **4** were obtained at 500 MHz for proton and 125 MHz for carbon and were recorded in CD_3OD.

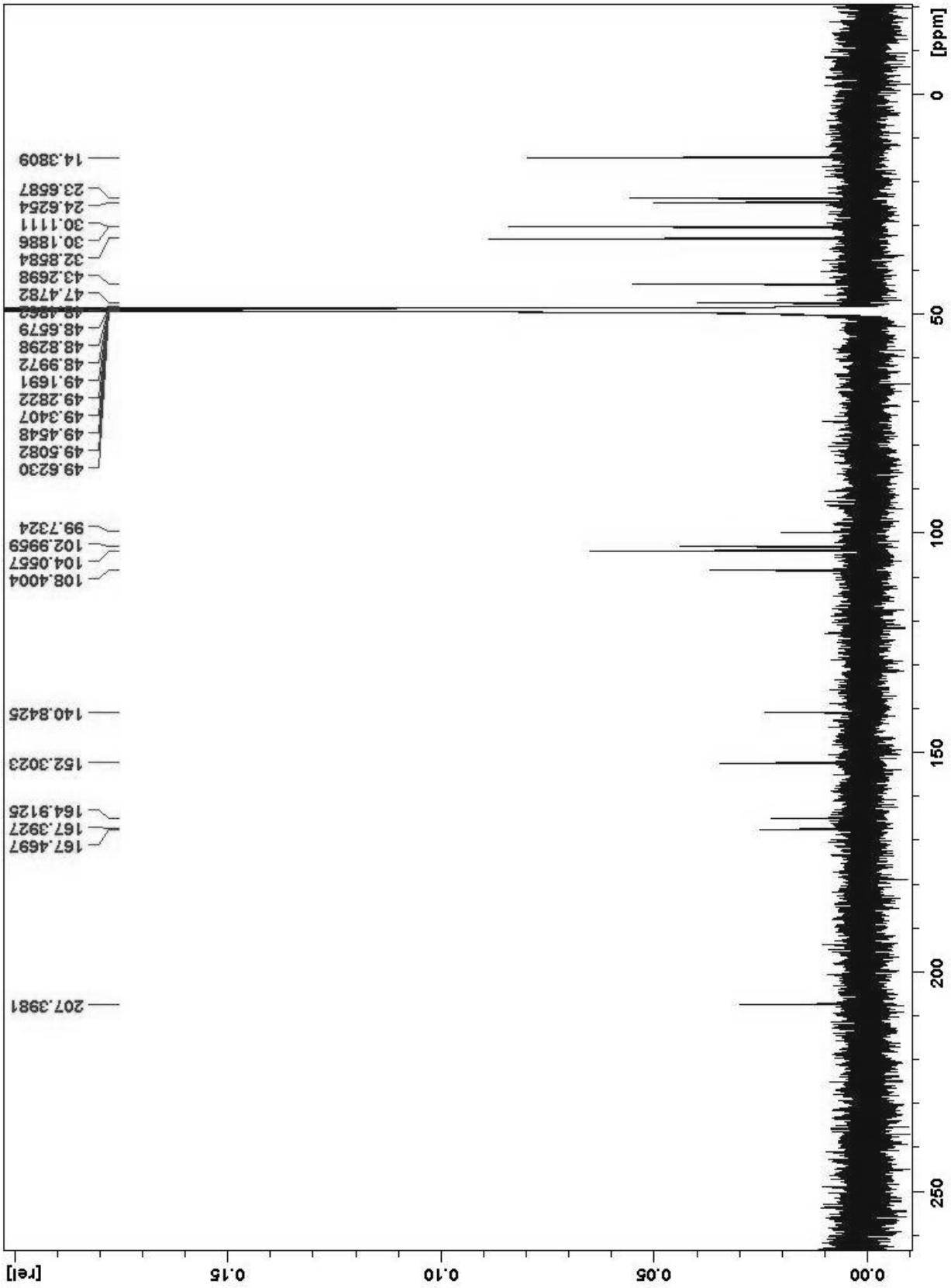
	SMA76b (**3**) 		SMA76c (**4**) 	
no	^{13}C δ (ppm)	^{1}H δ(ppm) (m, area, J_{HH} (Hz))	^{13}C δ (ppm)	^{1}H δ(ppm) (m, area, J_{HH} (Hz))
1	167.5	-	167.8	-
2	99.7	-	99.3*	-
3	164.9	-	165.4	-
4	103.0	6.37 (d, 1H, 2.2)	103.0	6.31 (s, 1H)
5	167.4	-	167.8	-
6	104.1	6.38 (d, 1H, 2.2)	104.1	6.31 (s, 1H)
7	140.8	-	140.6	-
8	108.4	6.48 (s, 1H)	105.6	6.32 (s, 1H)
9	152.3	-	159.5	-
10	47.5	3.73 (s, 2H)	34.5	2.51 (t, 2H, 7.5)
11	207.4	-	28.5	1.66-1.72 (m, 2H)
12	43.3	2.62 (t, 2H, 7.3)	30.5	1.29-1.39 (m, 2H)
13	24.6	1.61-1.64 (m, 2H)	30.6	1.29-1.39 (m, 2H)
14	30.1	1.32-1.35 (m, 2H)	33.4	1.29-1.39 (m, 2H)
15	30.2	1.32-1.35 (m, 2H)	24.2	1.29-1.39 (m, 2H)
16	32.9	1.32-1.35 (m, 2H)	14.9	0.91 (t, 3H, 7.1)
17	23.7	1.32-1.35 (m, 2H)		
18	14.4	0.93 (t, 3H, 7.1)		

*: weak signal

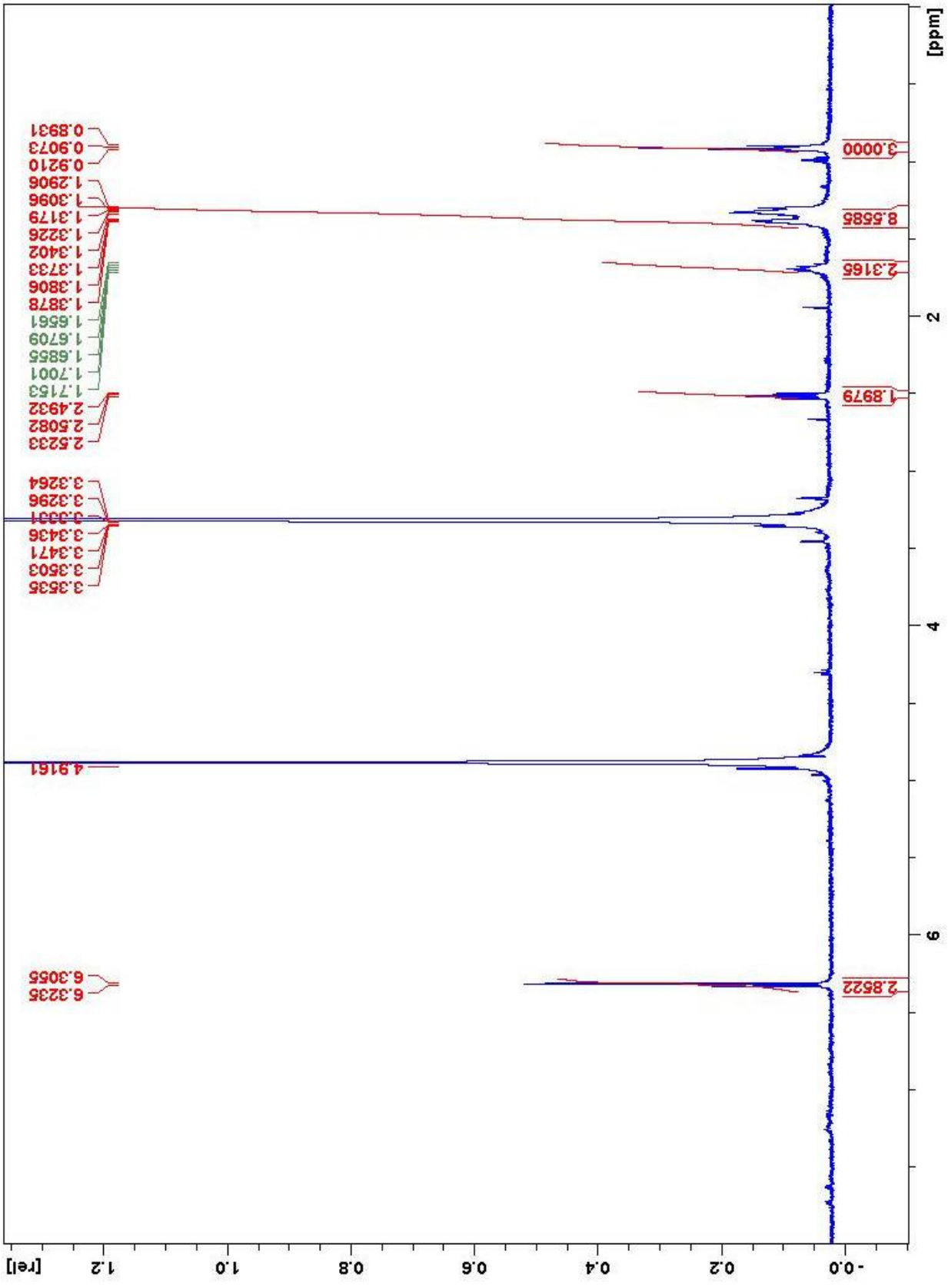
SMA76b, **3** (^{1}H NMR in CD_3OD)



SMA76b, **3** (^{13}C NMR in CD_3OD)



SMA76c, **4** (^{1}H NMR in CD_3OD)



SMA76c, **4** (^{13}C NMR in CD_3OD)