Supporting Information for

Enzymatic Formation of Unnatural Novel Chalcone, Stilbene, and Decaketide Benzophenone Scaffolds by Plant Type III Polyketide Synthase

Shepo Shi,[†] Kiyofumi Wanibuchi,[†] Hiroyuki Morita,[†] Kohei Endo,[†] Hiroshi Noguchi,[†] and Ikuro Abe*[,†,¶]

[†] *School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan, and*

[¶] *PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan.*

Table of Contents

Materials and Methods S2

Figure 1. Comparison of the amino acid sequences S8

Figure 2. NMR assignment and key HMBC correlations of Compound **1** – **4** S9

Figure 3 - 8. NMR spectra (^{1}H, ^{13}C, 2D NMR) of Compound **1** S10 - S15

Figure 9 - 12. NMR spectra (^{1}H, ^{13}C, 2D NMR) of Compound **2** S16 - S19

Figure 13 - 17. NMR spectra (^{1}H, ^{13}C, 2D NMR) of Compound **3** S20 - S24

Figure 18 - 24. NMR spectra (^{1}H, ^{13}C, 2D NMR) of Compound **4** S25 - S31

Figure 25. HPLC charts of SEK 15 S32

Materials and Methods

Chemicals. Malonyl-CoA and hexanoyl-CoA were purchased from Moravek Biochemicals (California). 4-Coumaroyl-CoA was chemically synthesized as described previously (Stöckigt, J.; Zenk, M. H. *Z. Naturforsh.* **1975**, *30c*, 352-358). SEK15 (Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 4166-4170) was obtained from Professor Chaitan Khosla (Stanford University).

Enzyme. The enzyme of OKS used in present study was cloned from *Aloe arborescens* as previously described in our report (Abe, I.; Oguro, S.; Utsumi, Y.; Sano, Y.; Noguchi, H. *J. Am. Chem. Soc*. **2005**, *127*, 12709-12716). The recombinant enzyme contains an additional hexahistidine tag at the C-terminal was subcloned to pET-22b (+) (Novagen). The plasmid was transformed into *E. coli* BL21 (DE3) pLysS. The cells harboring the plasmid were cultured to an A_{600} of 0.6 in Luria-Bertani medium containing 100 μg/mL of ampicillin at 23 °C. Then, 1.0 mM isopropyl-β-D-thiogalactopyranoside was added to induce protein expression, and the culture was incubated further at 30 °C for 12 h. The *E. coli* cells were harvested by centrifugation and resuspended in 40 mM potassium phosphate buffer, pH 7.9, containing 0.1 M NaCl. Cell lysis was carried out by sonication and centrifuged at 15 000 *g* for 30 min. The supernatant was passed through a column of Ni Sepharose^{TM} 6 Fast Flow. After washing with 20 mM potassium phosphate buffer, pH 7.9, containing

0.5 M NaCl and 40 mM imidazole, the recombinant OKS was finally eluted with 15 mM potassium phosphate buffer, pH 7.5, containing 10% glycerol and 500 mM imidazole. Protein concentration was determined by the Bradford method (Protein Assay, BioRad) with bovine gamma globulin as standard.

Enzyme reaction. The standard reaction mixture contained 118 nmol of malonyl-CoA, 54 nmol of starter CoA ester and 900 pmol of the purified enzyme in a final volume of 500 μL of 100 mM potassium phosphate buffer, pH 7.5 containing 1 mM EDTA was incubated at 30 °C for 12 h and stopped by adding 50 μL of 20% HCl. The reaction products were then extracted twice with 1000 μL of EtOAc and analyzed by HPLC on a TSK-gel ODS-80Ts column (4.6 ×150 mm, TOSOH) with a flow rate of 0.2 mL/min and an on-line monitoring wavelength at 280 nm. For the analysis of enzyme reaction products from 4-coumaroyl-CoA, gradient elution was performed with H_2O and MeOH, both containing 1.0% HAc: 0-5 min, 30% MeOH; 5-40 min, 30-50% MeOH. For the analysis of enzyme reaction products from hexanoyl-CoA, gradient elution was performed with H_2O and MeOH, both containing 1.0% HAc: 0-5 min, 30%-45% MeOH; 5-40 min, 45-60% MeOH. For the analysis of enzyme reaction products from malonyl-CoA by OKS N222G, gradient elution was performed with H_2O and MeOH, both containing 1.0% HAc: 0-5 min, 30% MeOH; 5-17 min, 30-60% MeOH, 17-25 min, 60% MeOH, 25-27 min, 60-70% MeOH, 27-35 min, 70% MeOH, 35-37 min, 70-100% MeOH, 37-42 min, 100% MeOH. On-line LC-ESIMS spectra were measured with an Agilent Technologies

(Santa Clara, CA, USA) HPLC 1100 series coupled to a Bruker Daltonics (Bremen, Germany) esquire4000 ion trap mass spectrometer fitted with an ESI source as described before (Abe I.; Morita H.; Oguro S.; Noma H.; Wanibuchi K.; Kawahara N.; Goda Y.; Noguchi H.; Kohno T. *J. Am. Chem. Soc.* **2007**, *129*, 5976-5980).

For large-scale enzyme reaction, 12 mg of purified enzyme was incubated with 15 mg of malonyl-CoA and 5 mg starter CoA in 100 mL of buffer (100 mM Tris-HCl buffer, pH 7.5 for 4-coumaroyl-CoA reaction, and 100 mM potassium phosphate buffer, pH 8.0 containing 1 mM EDTA for hexanoyl-CoA reaction, respectively) at 30 °C for 12 h. The reaction was quenched by addition of 20% HCl (10 mL) and extracted with ethyl acetate (200 mL × 3). The enzyme reaction products were purified by reverse-phase HPLC. The gradient elution program was performed with H_2O and MeOH, both containing 1.0% HAc: 0-5 min, 30% MeOH; 5-40 min, 30-46% MeOH; 40-43 min, 46%-90% MeOH; 43-50 min, 90-30% MeOH. For the enzyme reaction of 4-coumaroyl-CoA and OKS wild-type, total 30 times of large-scale reactions were carried out to get 0.6 mg of compound **1** with a yield of 0.96%, and 1.6 mg of compound **2** with a yield of 2.9%. For the enzyme reaction of hexanoyl-CoA, total 6 times of large-scale reactions were carried out to get 1.5 mg of compound **3** with a yield of 13.0%, and 2.0 mg of compound **4** with a yield of 21.2%. NMR spectra were measured by JEOL JNM-A400 (400 MHz for ^{1}H and 100 MHz for ^{13}C) and JEOL JNM-ECA800 (800 MHz for ^{1}H and

200 MHz for ^{13}C) NMR spectrometers with TMS as external standard. The HRFABMS were recorded on a JMS-700 mass spectrometer.

Enzyme kinetics. Steady-state kinetic parameters were determined using [2-^{14}C]malonyl-CoA (1.8 mCi/mmol), malonyl-CoA (164 μM) and 4-coumaroyl-CoA (or hexanoyl-CoA) (11, 22, 33, 44, 55 μM) as substrates. The experiments were carried out in triplicate with the substrates in the assay mixture, containing 20 μg of purified enzyme, 1 mM EDTA, in a final volume of 500 μL of 100 mM K-phosphate buffer, pH 8.0. Incubations were carried out at 30 °C for 60 min. The reaction products were extracted and separated by Si-gel TLC (Merck Art. 1.11798; ethyl acetate/hexane/AcOH = 63:27:5, v/v/v). Radioactivties were quantified by autoradiography using a bioimaging analyzer BAS-2000II (FUJIFILM). Lineweaver-Burk plots of data were employed to derive the apparent K_M and k_{cat} values (average of triplicates) using EnzFitter software (BIOSOFT).

Compound **1** (C_{21} heptaketide chalcone). LC-ESIMS: R_t = 23.5 min, *m/z* 381 $[M + H]^+$. UV: λ_{max} 242 nm, 298 nm, 343 nm. ^1H NMR (400 MHz, CD_3OH): δ 7.40 (2H, d, J = 8.0 Hz), 7.29 (1H, d, J = 15.6 Hz), 6.88 (1H, d, J = 15.6 Hz), 6.76 (2H, d, J = 8.0 Hz), 6.32 (1H, br.s), 6.31 (1H, br.s), 5.79 (1H, br.s), 5.34 (1H, br.s), 3.74 (2H, s). ^{13}C NMR (100 MHz, CD_3OH): δ 198.9, 173.7, 169.2, 166.3, 161.4, 161.2, 159.0, 146.8, 137.2, 131.6, 129.0, 126.5, 119.3, 116.9, 111.0, 102.9, 102.0, 89.9, 38.4. HRMS (FAB, positive): found for $[C_{21}H_{17}O_7]^+$ 381.0986;

calcd. 381.0974.

Compound **2** (C_{19} hexaketide stilbene). LC-ESIMS: R_t = 32.5 min, *m/z* 339 $[M + H]^+$. UV: λ_{max} 244 nm, 297 nm, 327 nm. ^{1}H NMR (400 MHz, CD_3OH): δ 7.24 (2H, d, *J* = 7.6 Hz), 6.90 (1H, d, *J* = 16.0 Hz), 6.79 (1H, d, *J* = 16.0 Hz), 6.72 (2H, d, *J* = 7.6 Hz), 6.66 (1H, br.s), 6.28 (1H, br.s), 6.11 (1H, br.s), 5.38 (1H, br.s). ^{13}C NMR (100 MHz, CD_3OH): δ 173.4, 168.8, 161.5, 161.0, 158.8, 158.4, 141.0, 132.0, 130.3, 129.0, 124.1, 116.6, 112.8, 108.4, 104.5, 102.6, 90.5. HRMS (FAB, positive): found for $[C_{19}H_{15}O_6]^+$ 339.0849; calcd. 339.0868.

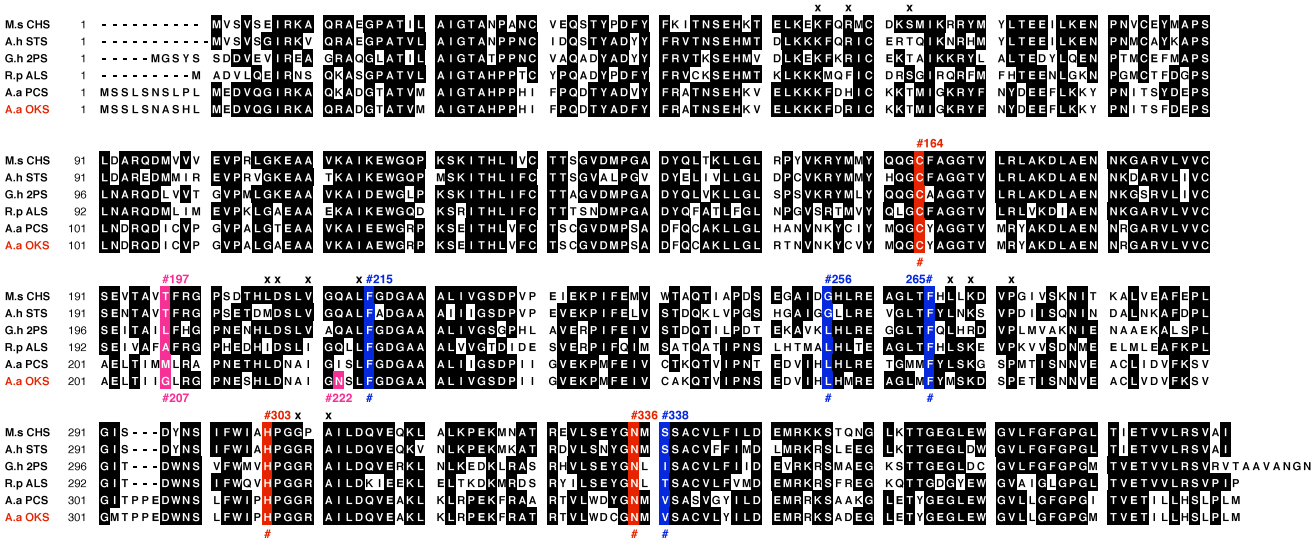
Compound **3** (C_{18} heptaketide phloroglucinol). LC-ESIMS: R_t = 29.6 min, *m/z* 333 $[M + H]^+$. UV: λ_{max} 287 nm. ^{1}H NMR (400 MHz, CD_3OH): δ 6.27 (1H, br.s), 6.25 (1H, br.s), 5.76 (1H, br.s), 5.30 (1H, br.s), 3.75 (2H, s), 2.85 (2H, t, *J* = 7.2 Hz), 1.56 (2H, m), 1.28 (4H, m), 0.88 (3H, t, *J* = 7.2 Hz). ^{13}C NMR (100 MHz, CD_3OH): δ 209.3, 173.5, 168.1, 167.0, 161.4, 159.6, 136.9, 121.6, 111.2, 102.9, 102.5, 89.6, 45.3, 38.2, 32.6, 25.1, 23.4, 14.2. HRMS (FAB, positive): found for $[C_{18}H_{21}O_6]^+$ 333.1359; calcd. 333.1338.

Compound **4** (C_{16} hexaketide resorcinol). LC-ESIMS: R_t = 30.7 min, *m/z* 291 $[M + H]^+$. UV: λ_{max} 296 nm. ^{1}H NMR (800 MHz, CD_3OH): δ 6.12 (1H, d, *J* = 2.4 Hz), 6.10 (1H, d, *J* = 2.4 Hz), 6.00 (1H, d, *J* = 1.8 Hz), 5.28 (1H, d, *J* = 1.8 Hz), 2.38 (2H, t, *J* = 8.0 Hz), 1.40 (2H, m), 1.16 –1.19 (4H, m), 0.77 (3H, t, *J* = 7.2 Hz). ^{13}C NMR (100 MHz, CD_3OD): δ 173.9, 169.4,

162.3, 160.9, 158.1, 145.8, 113.2, 109.0, 106.9, 101.3, 89.9 (this carbon was obtained by checking the sample in CD_3OH), 34.6, 32.7, 31.8, 23.3, 14.3. HRMS (FAB, positive): found for $[C_{16}H_{19}O_5]^+$ 291.1220; calcd. 291.1232.

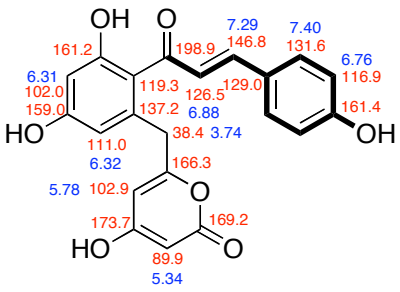
Site-directed mutagenesis. On the basis of the X-ray crystal structure of *A. arborescens* OKS at 2.6 Å resolution (Morita, H.; Kondo, S.; Kato, R.; Wanibuchi, K.; Noguchi, H.; Sugio, S.; Abe, I.; Kohno, T. *Acta Crystal*. **2007**, *F63*, 947-949), OKS N222G mutant was constructed with the QuickChange Site-Directed Mutagenesis Kit (Stratagene), according to manufacturer's protocol, using a pair of primers (mutated codons are underlined): sense 5'-GACAACGCCATCGGA GGT TCTCTTTTCGG-3' and anti-sense as a sense primer, 5'-CCATCTCCGAAAAGAGA ACC TCCGATGGC-3'. The point mutant was functionally expressed in *E. coli* at levels comparable with wild-type enzyme, and purified to homogeneity as in the case of the wild-type OKS.

Figure 1. Comparison of the amino acid sequences of *Aloe arborescens* OKS and other CHS-superfamily type III PKSs.

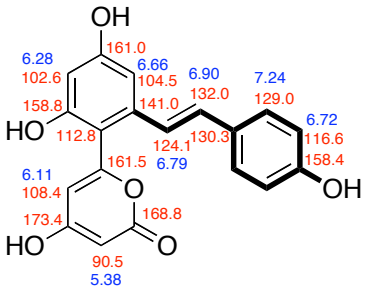


Comparison of the amino acid sequences of *Aloe arborescens* OKS and other CHS-superfamily type III PKSs. M.s CHS, *Medicago sativa* CHS; A.h STS, *Arachis hypogaea* stilbene synthase; G.h 2PS, *Gerbera hybrida* 2PS; R.p ALS, *Rheum palmatum* ALS; A.a PCS, *A. arborescens* PCS; A.a OKS, *A. arborescens* octaketide synthase. The critical active-site residue 197 (in pink), the catalytic triad (Cys164, His303, and Asn336) (in red), and the residues lining the active-site (Phe215, Gly256, F265, Ser338) (in blue) were marked with # (numbering in *M. sativa* CHS), and residues for the CoA binding with +. Asn222 of *A. arborescens* OKS was also marked in pink.

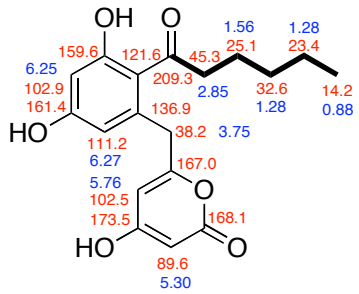
Figure 2. NMR assignment and key HMBC correlations of the enzyme reaction products **1** - **4**.



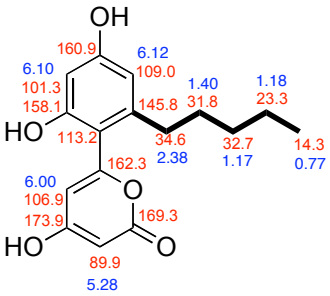
1



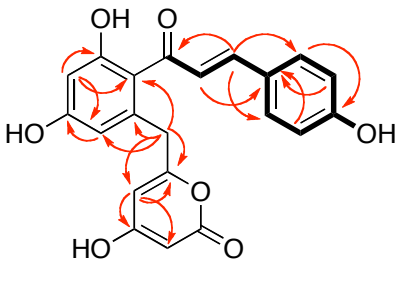
2



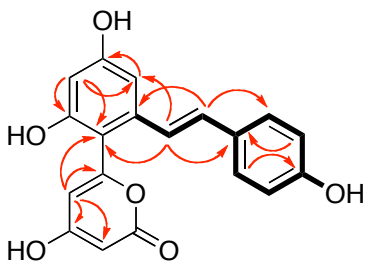
3



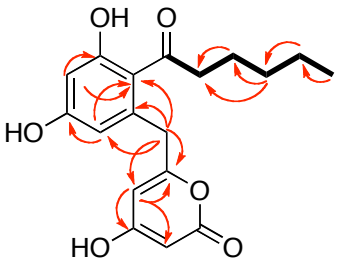
4



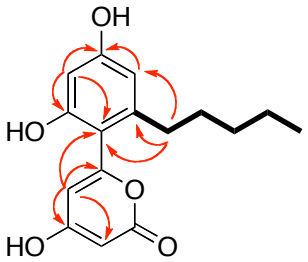
1



2



3



4

Figure 3. ^{1}H NMR spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OH, 400 MHz).

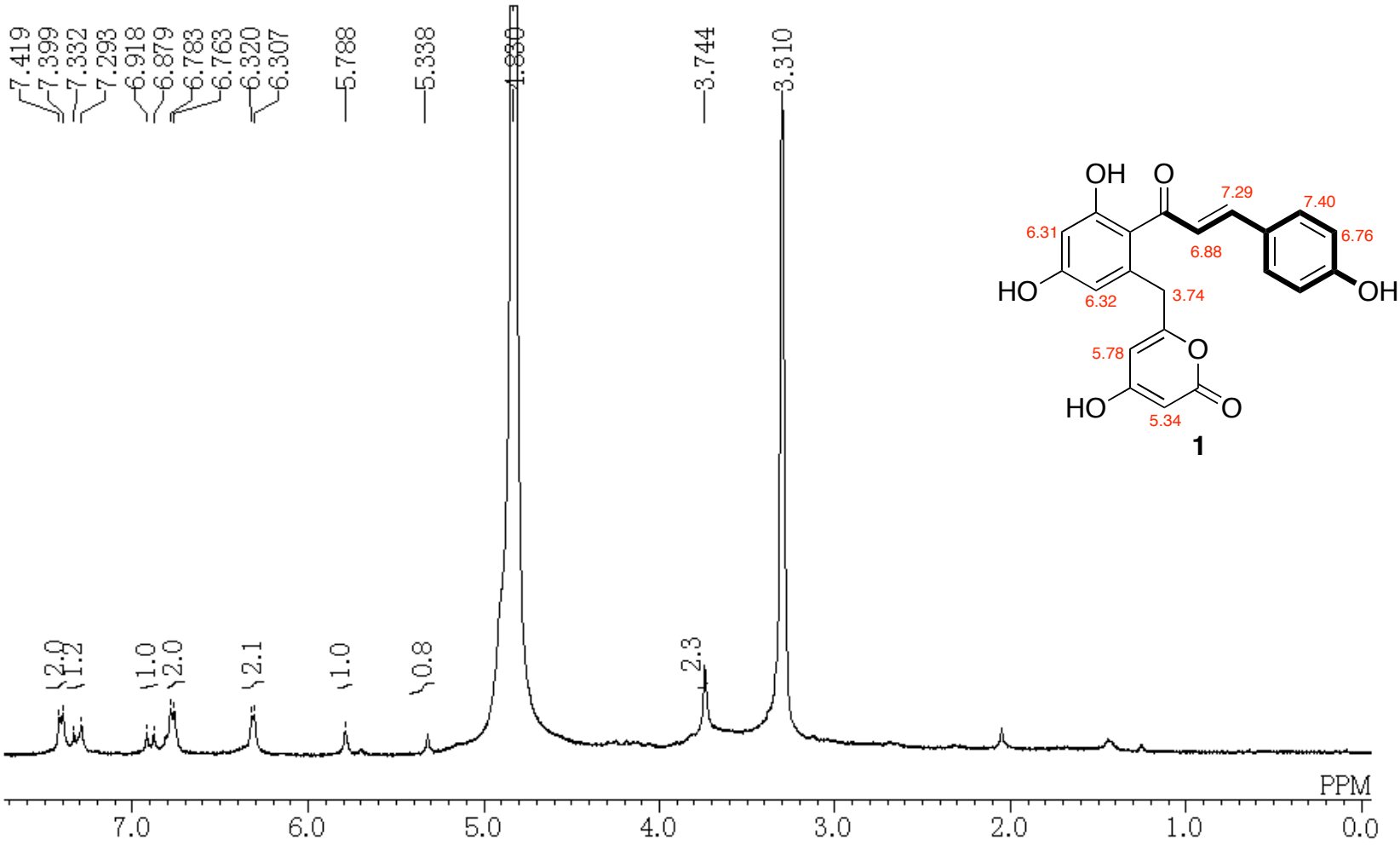


Figure 4. ^{13}C NMR spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OH, 100 MHz).

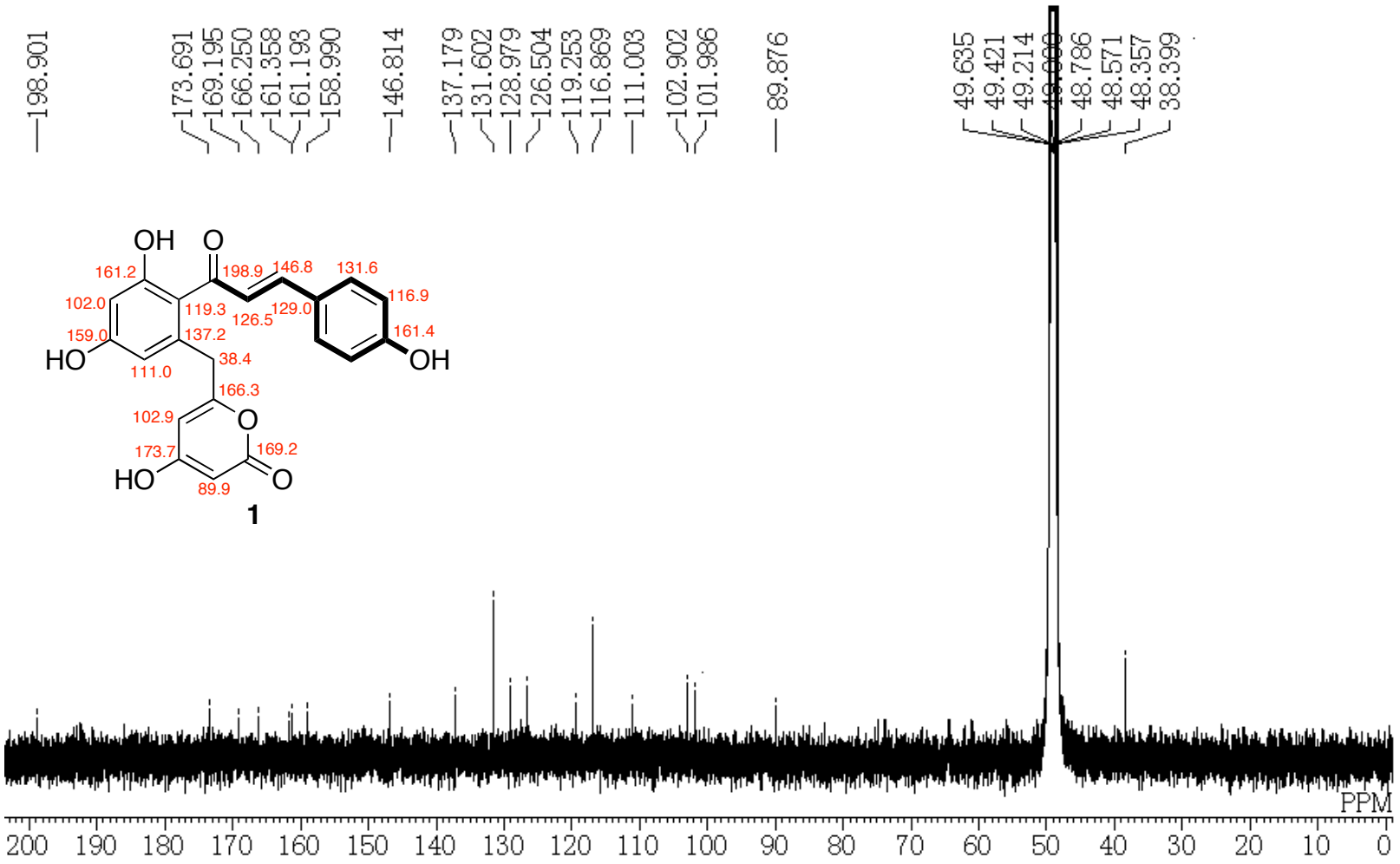


Figure 5. ^{1}H-^{1}H COSY spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OD, 400 MHz).

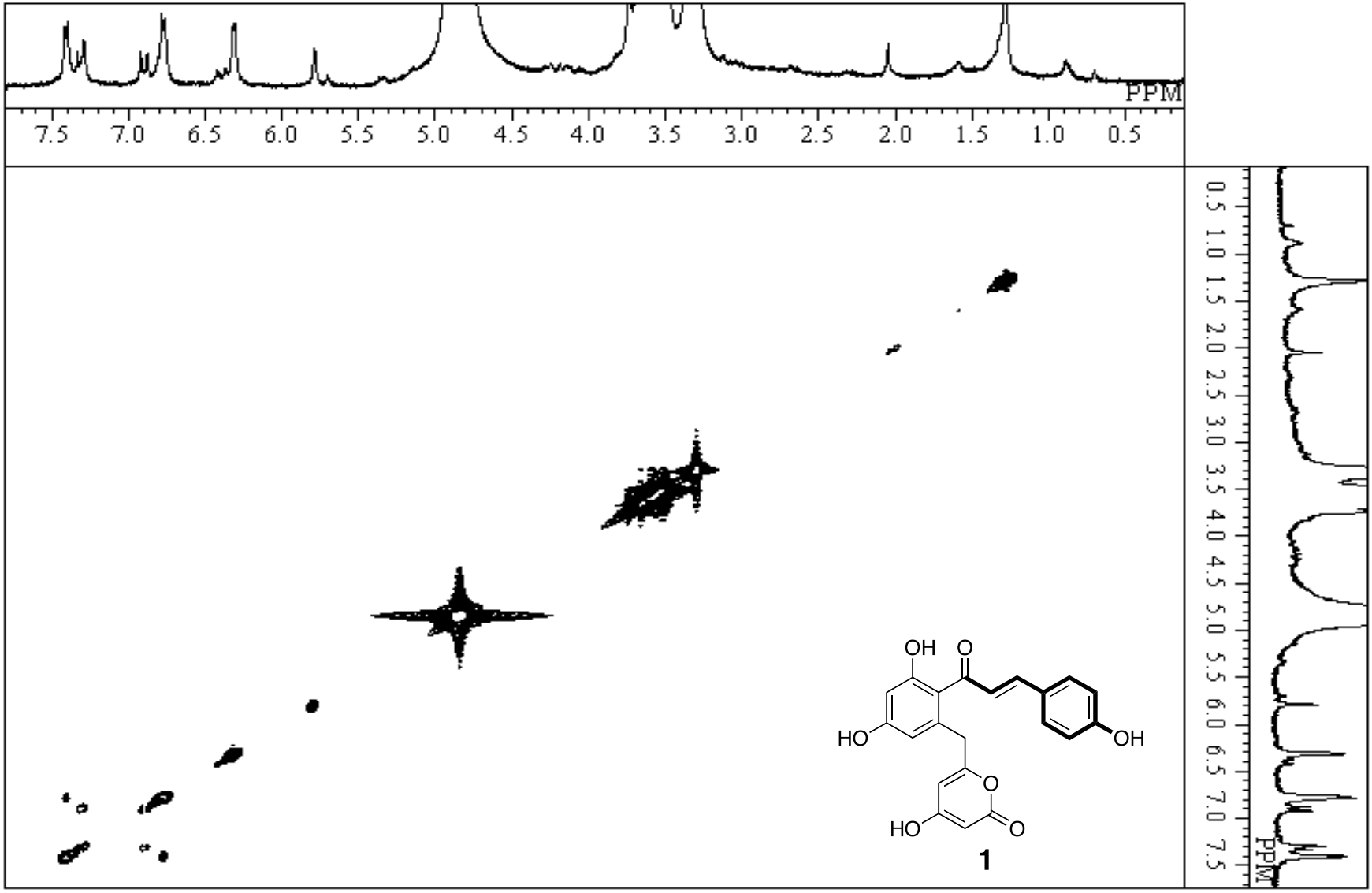


Figure 6. HMQC spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OD, 400 MHz).

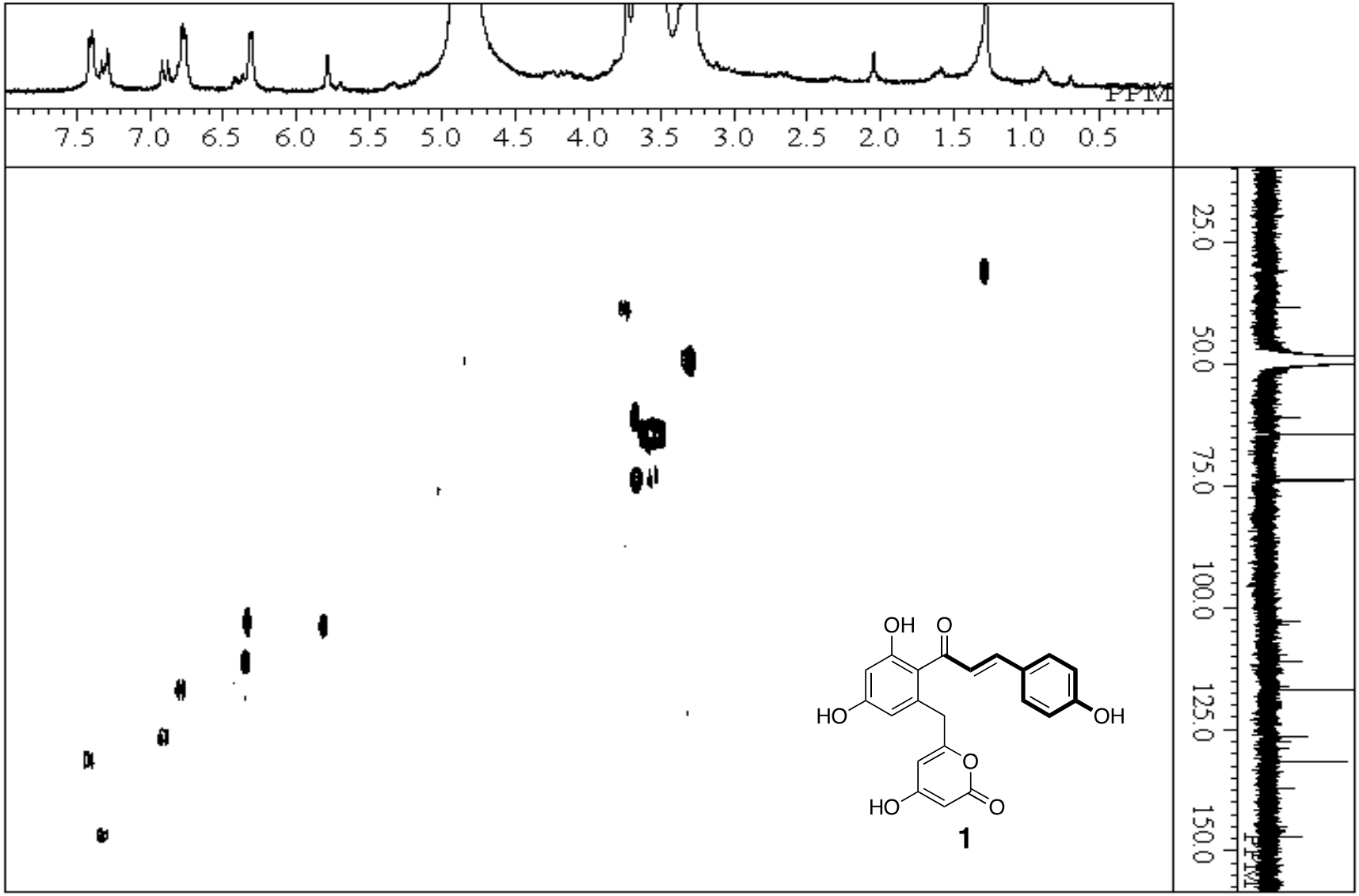


Figure 7. HMBC spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OD, 400 MHz).

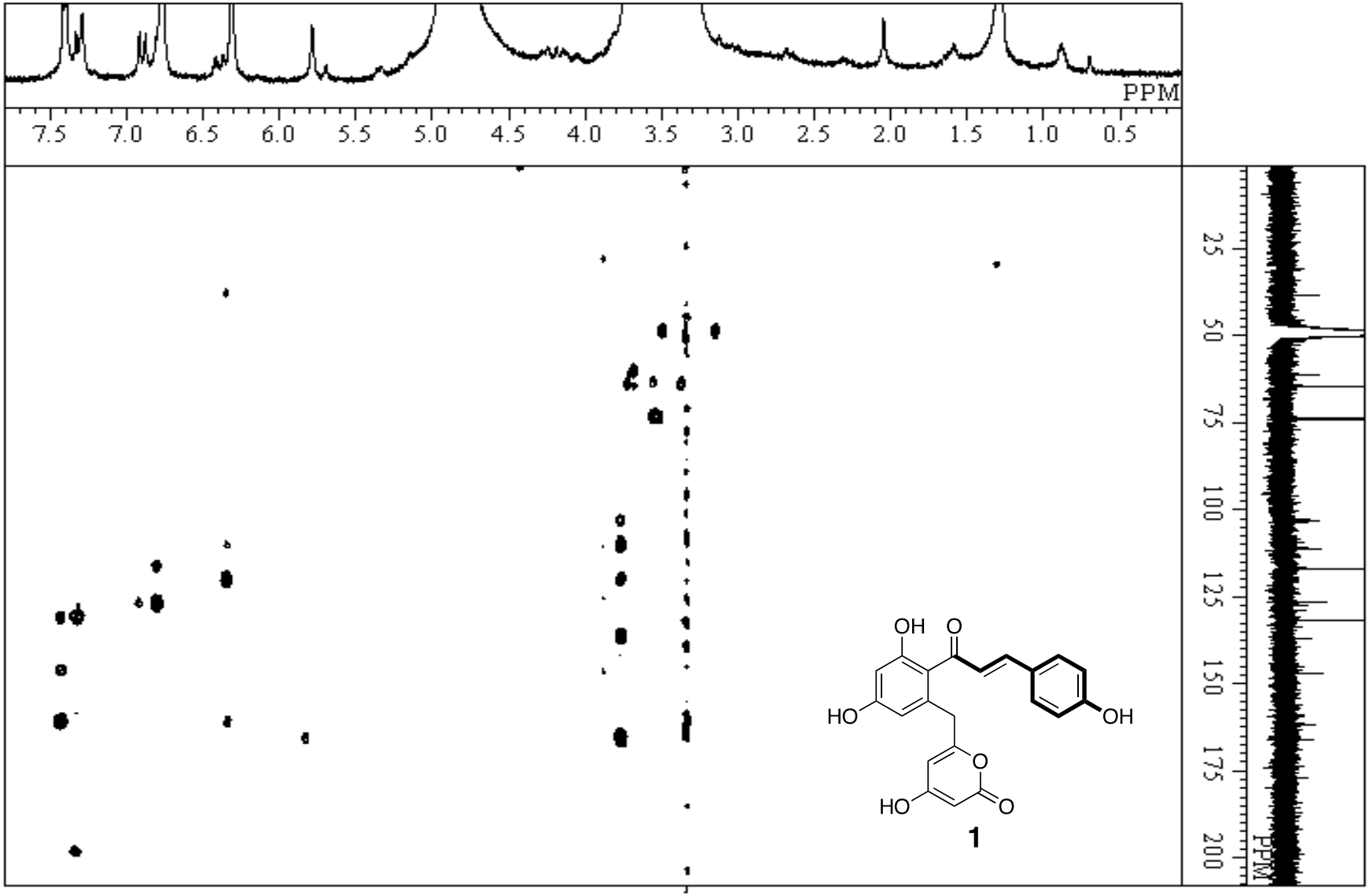


Figure 8. Expended HMBC spectrum of **1** (C_{21} heptaketide chalcone) (in CD_3OD, 400 MHz).

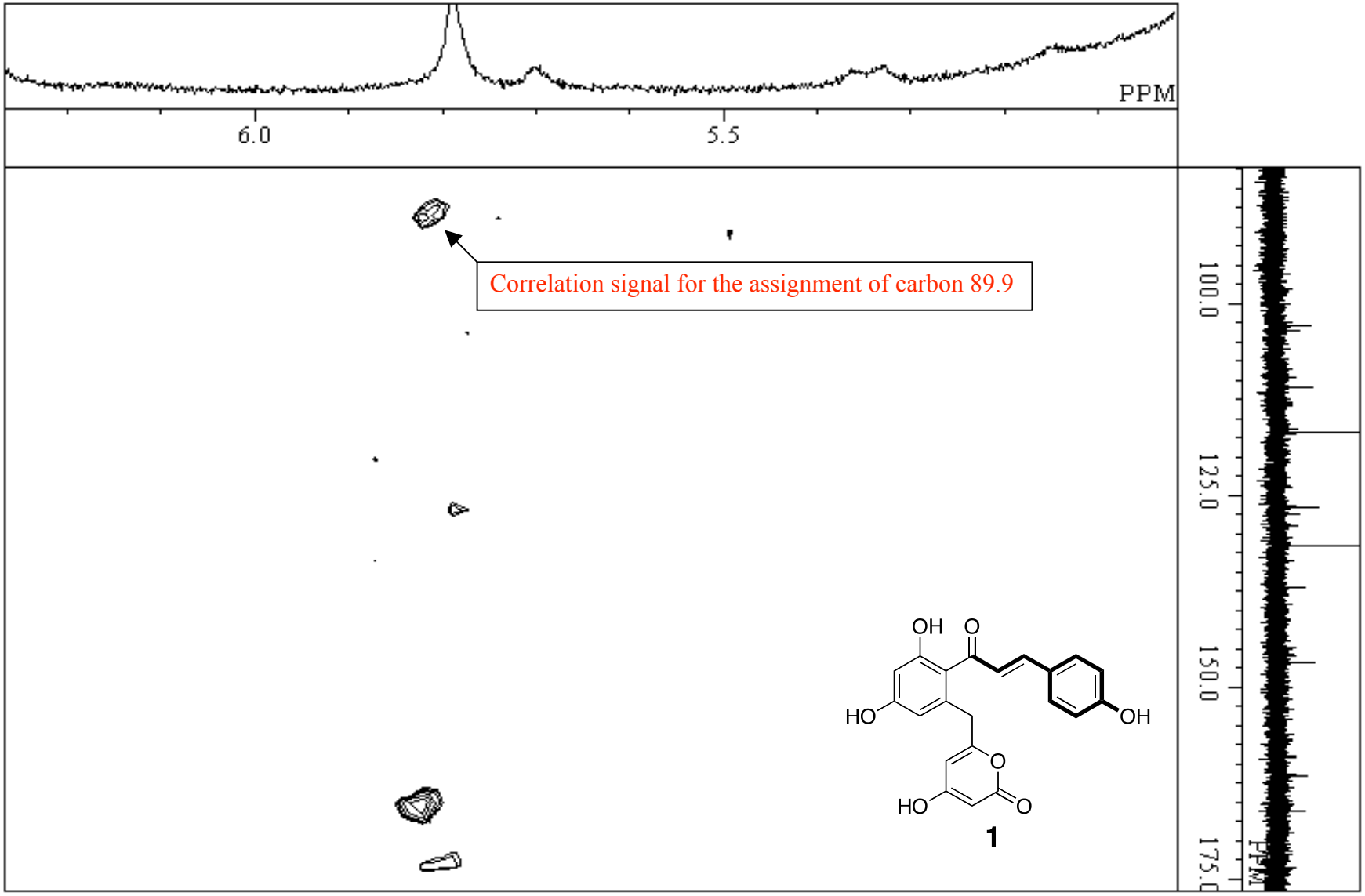


Figure 9. ^{1}H NMR spectrum of **2** (C_{19} hexaketide stilbene) (in CD_3OH, 400 MHz).

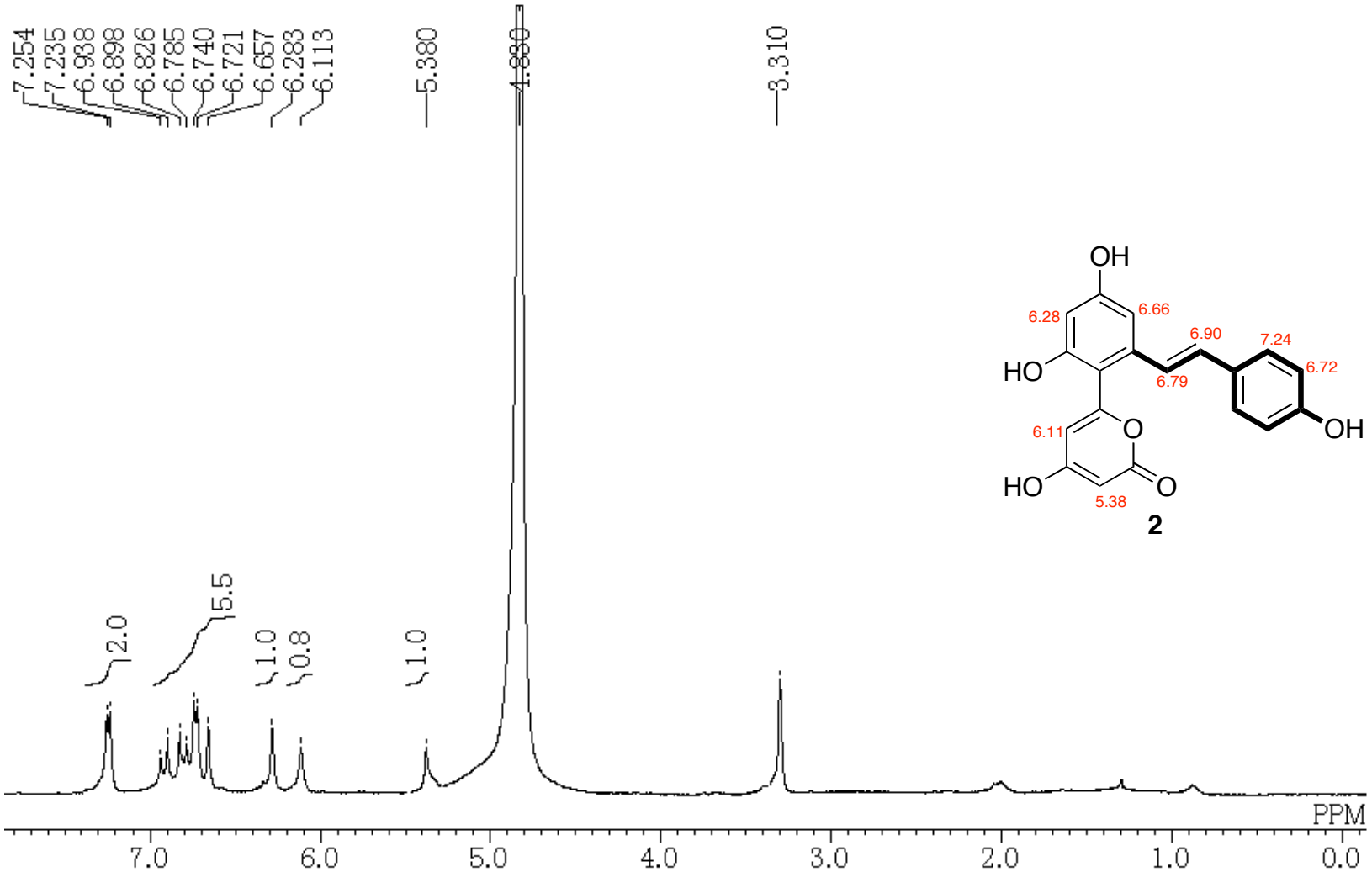


Figure 10. ^{13}C NMR spectrum of **2** (C_{19} hexaketide stilbene) (in CD_3OH, 100 MHz).

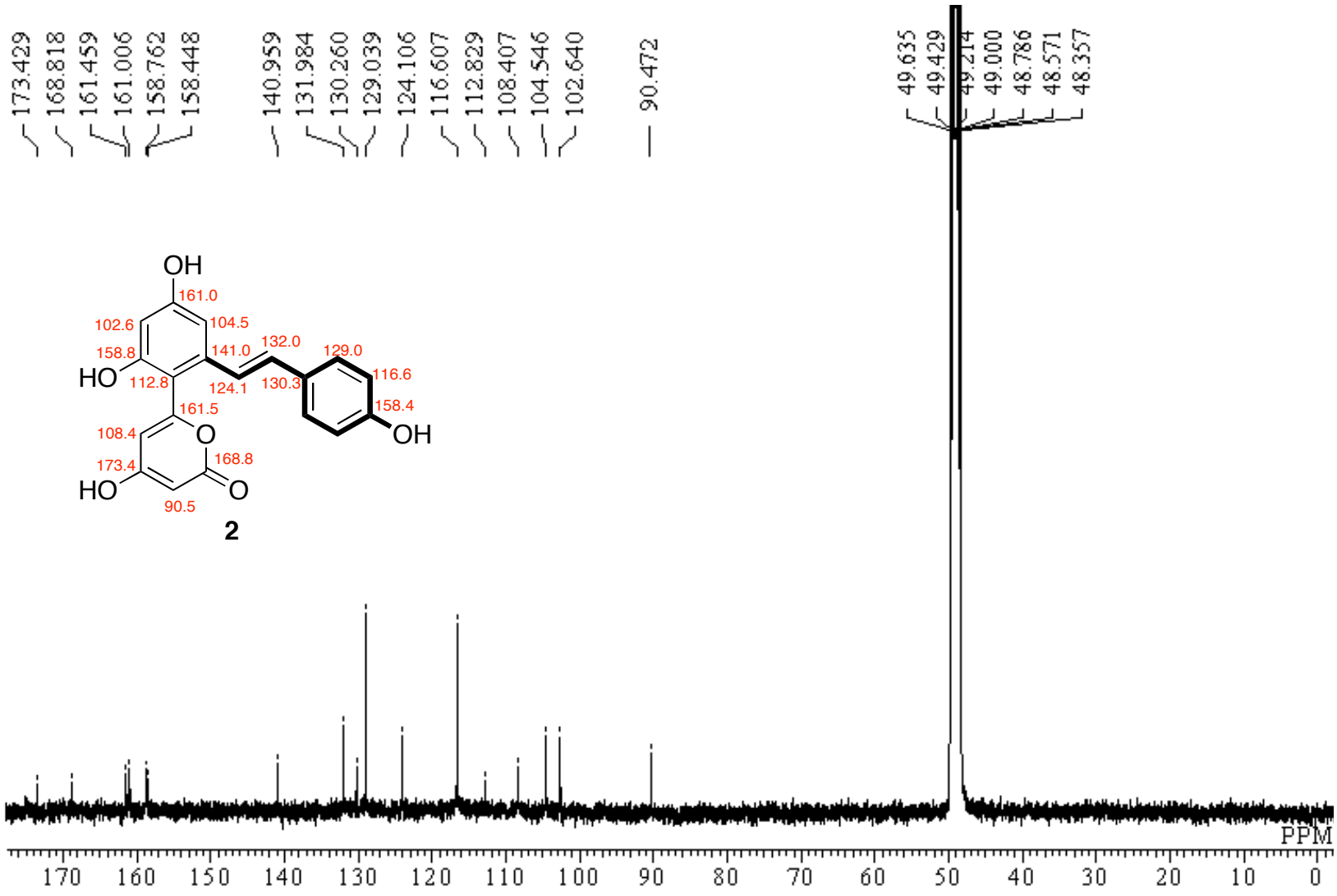


Figure 11. HMQC spectrum of **2** (C_{19} hexaketide stilbene) (in CD_3OD, 400 MHz).

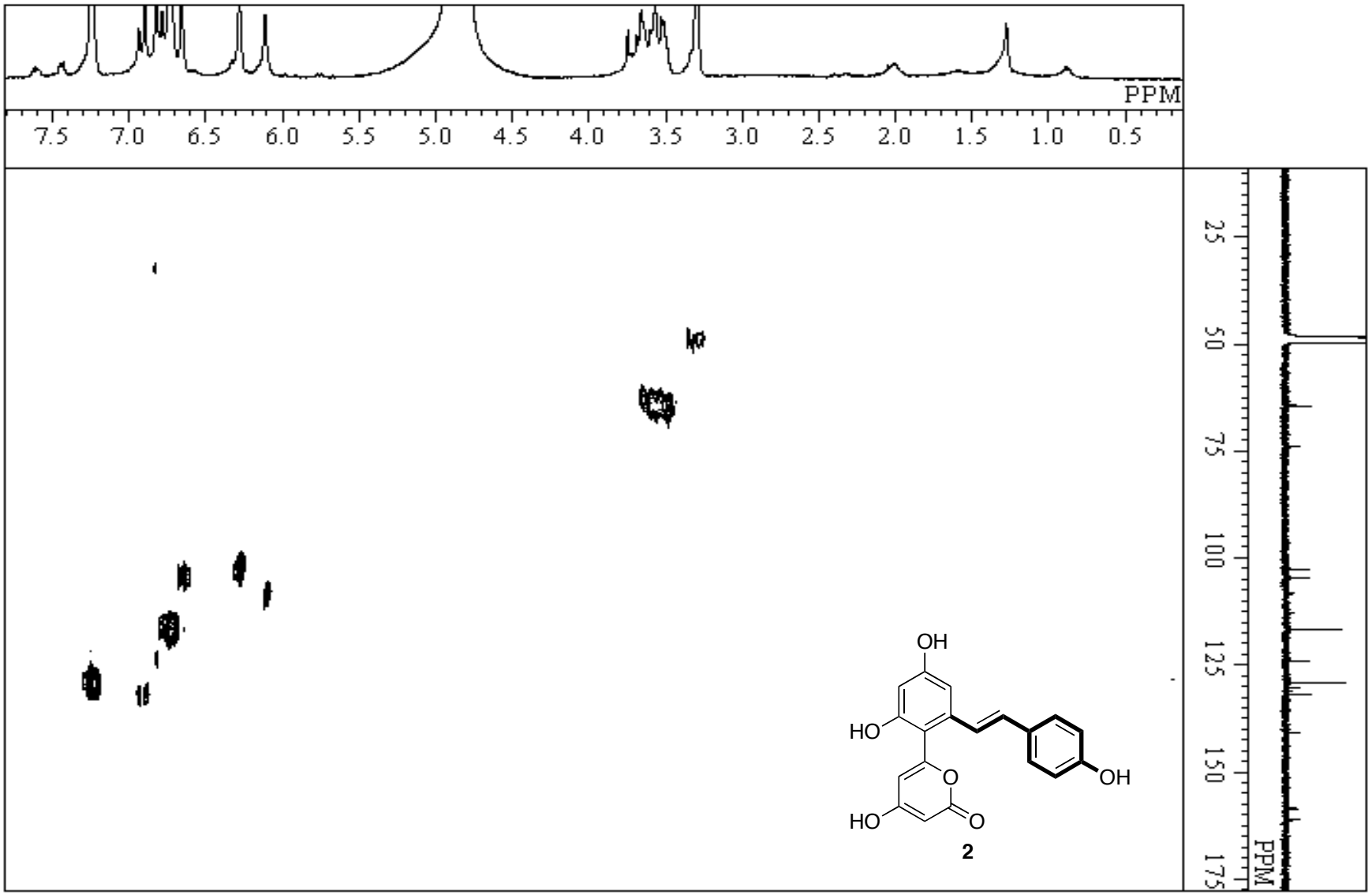


Figure 12. HMBC spectrum of **2** (C_{19} hexaketide stilbene) (in CD_3OD, 400 MHz).

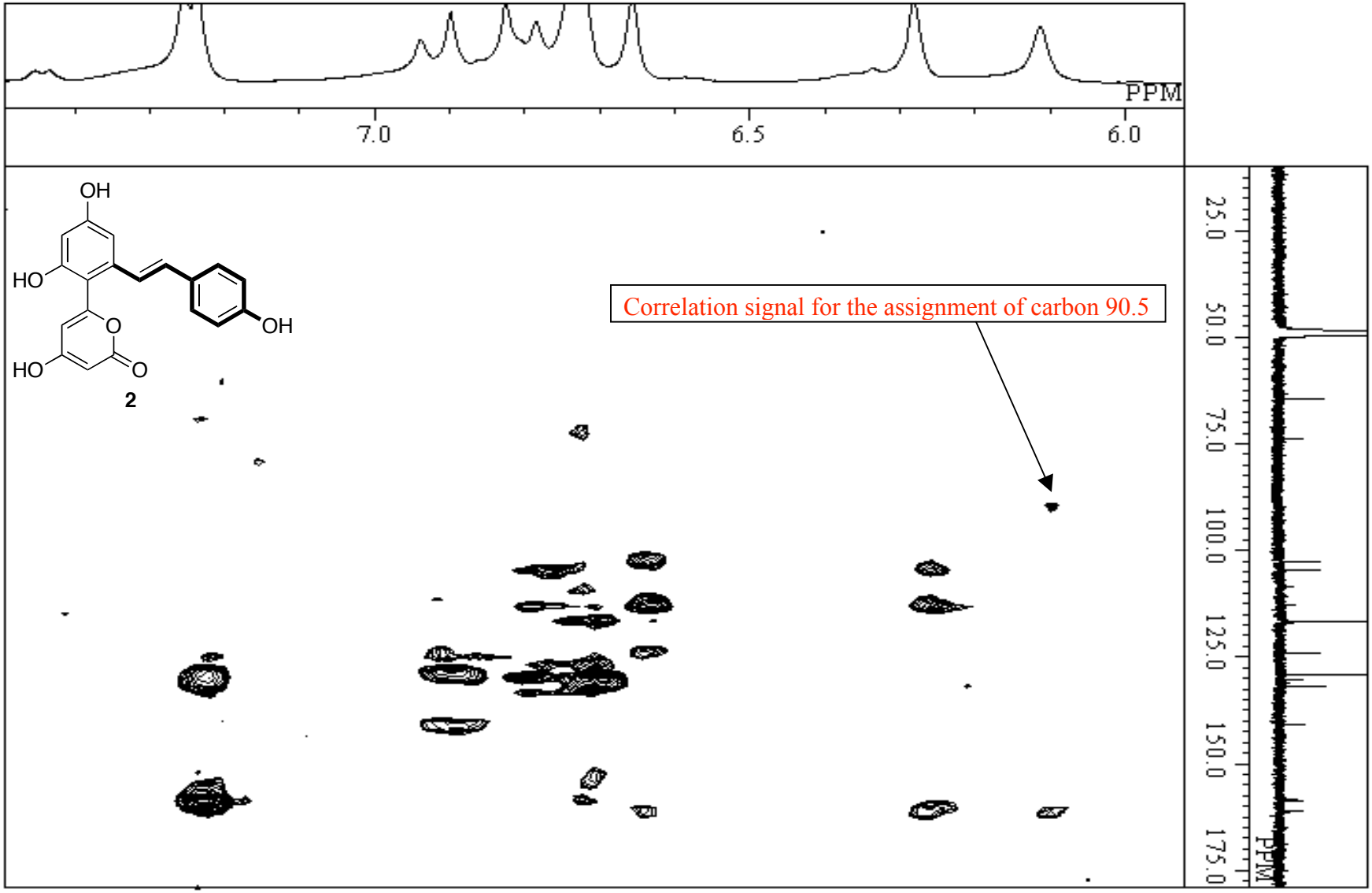


Figure 13. ^{1}H NMR spectrum of **3** (C_{18} heptaketide phloroglucinol) (in CD_3OH, 400 MHz).

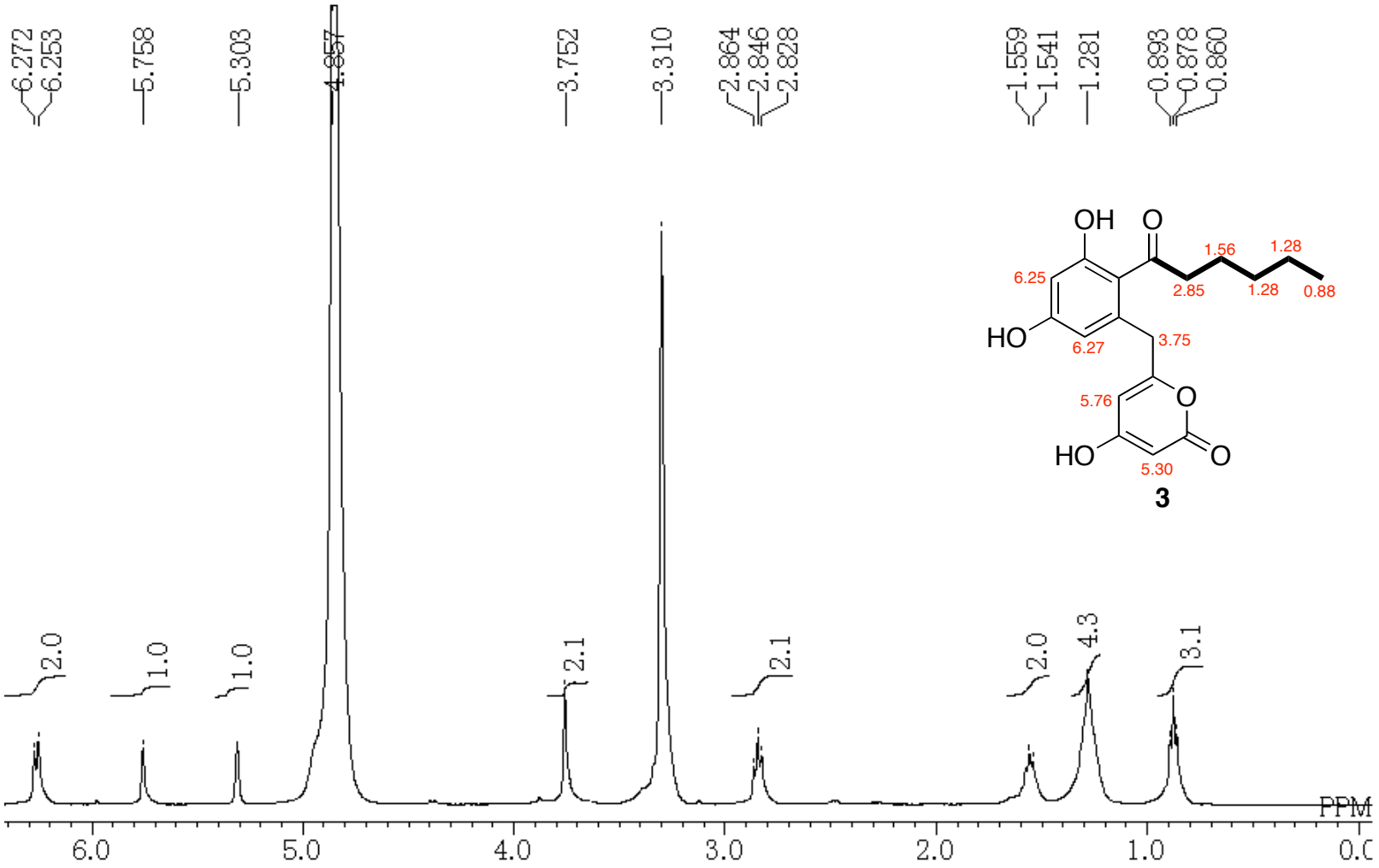


Figure 14. ^{13}C NMR spectrum of **3** (C_{18} heptaketide phloroglucinol) (in CD_3OH, 100 MHz).

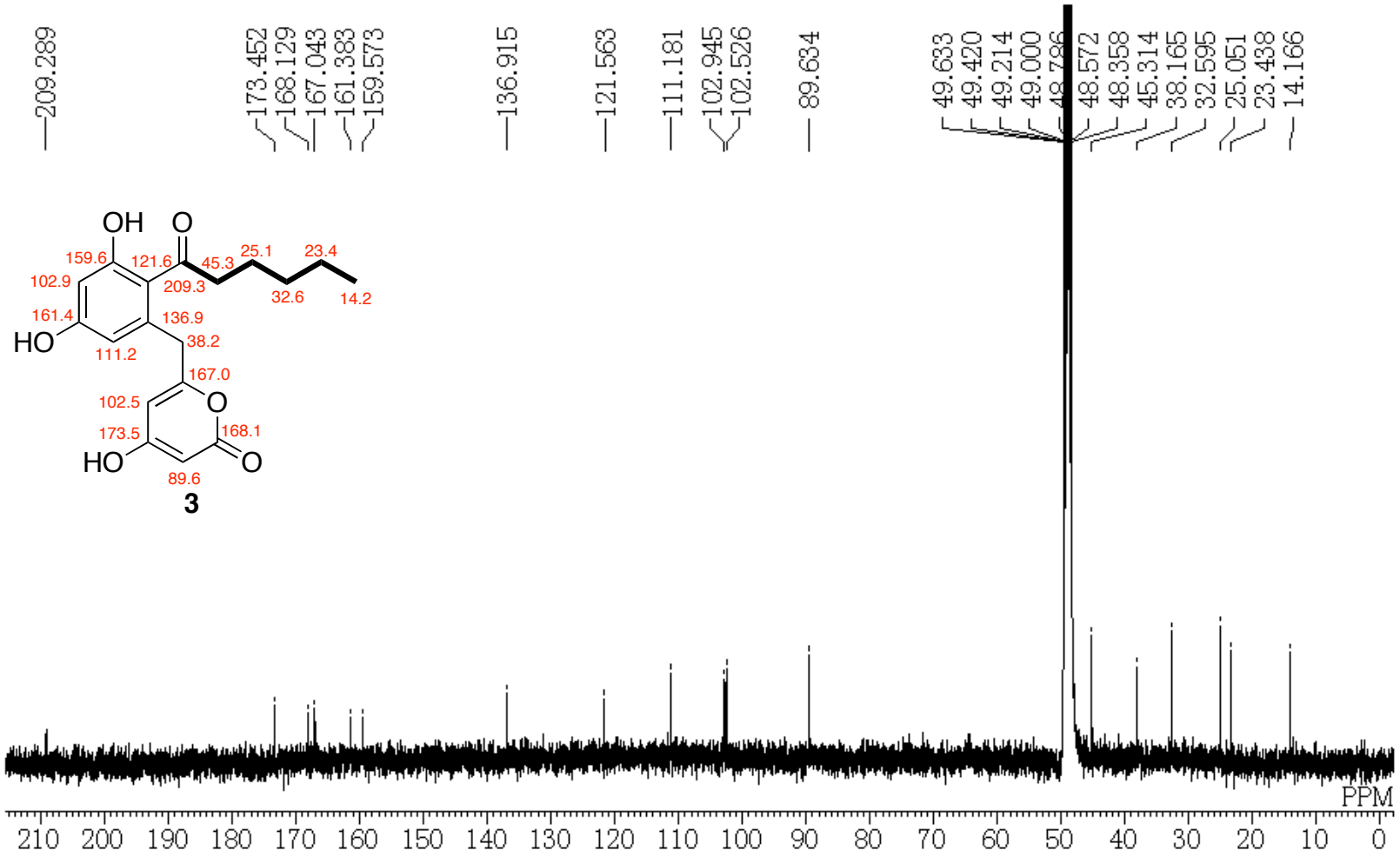


Figure 15. ^{1}H-^{1}H COSY spectrum of **3** (C_{18} heptaketide phloroglucinol) (in CD_3OD, 400 MHz).

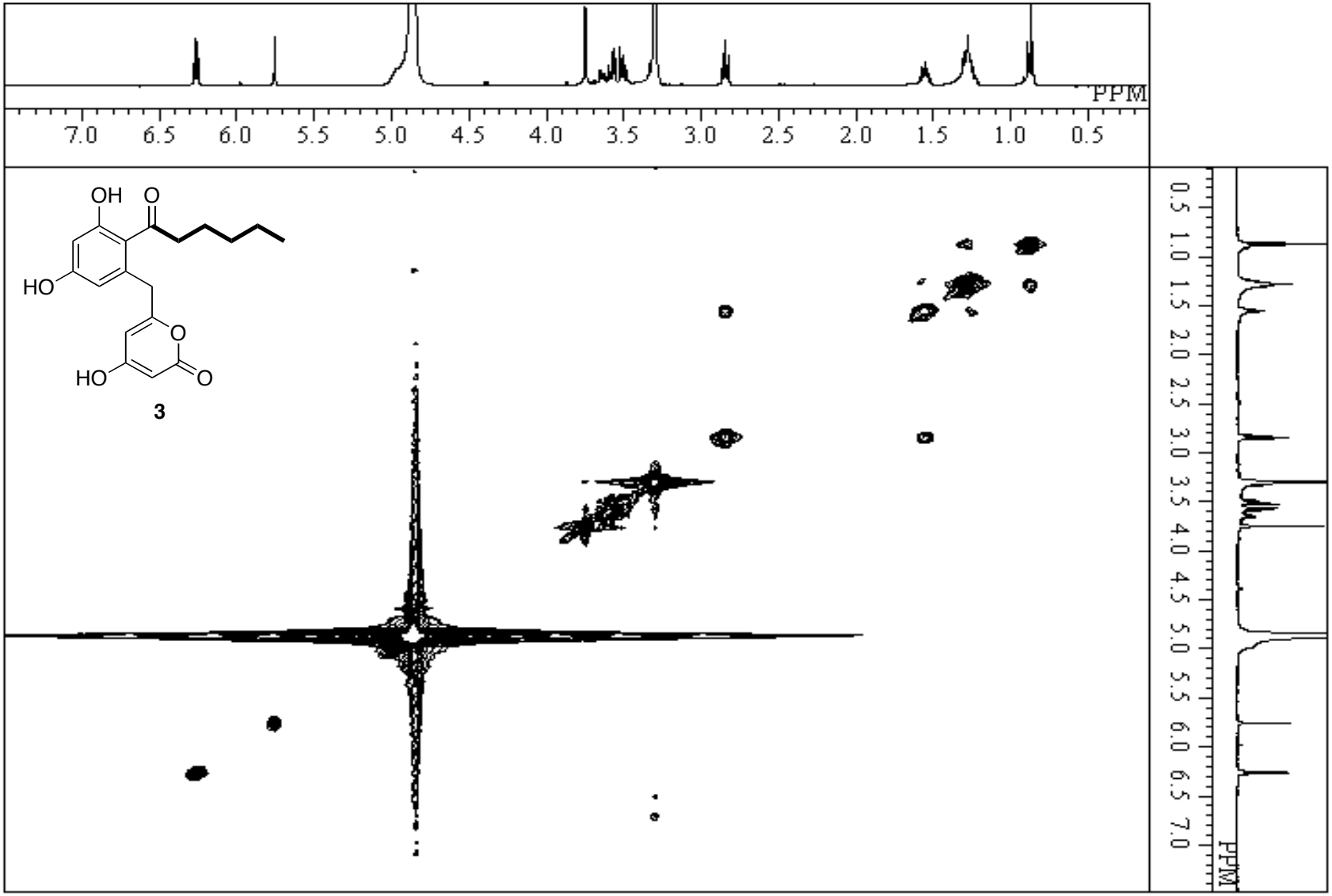


Figure 16. HMQC spectrum of **3** (C_{18} heptaketide phloroglucinol) (in CD_3OD, 400 MHz).

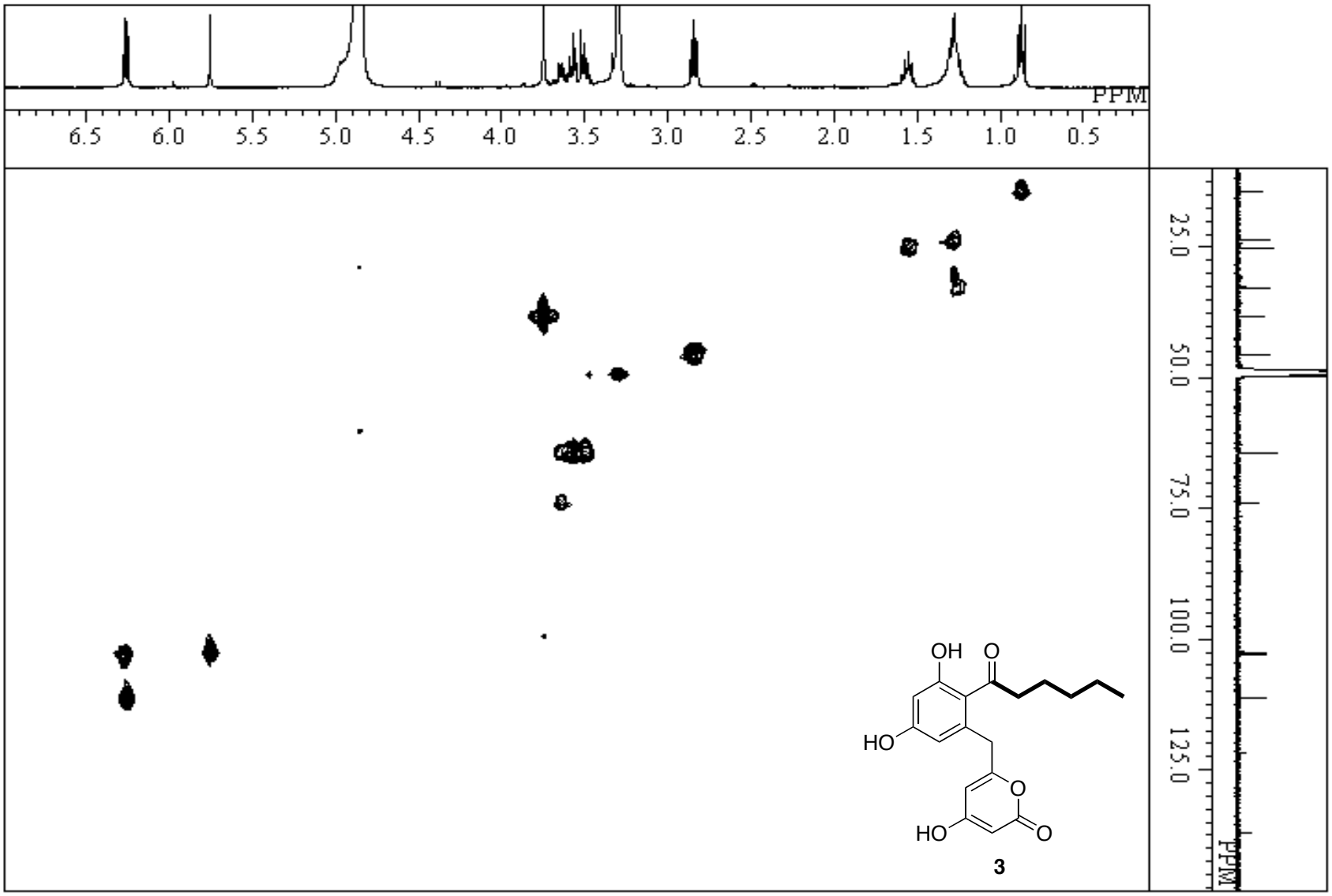


Figure 17. HMBC spectrum of **3** (C_{18} heptaketide phloroglucinol) (in CD_3OD, 400 MHz)**.**

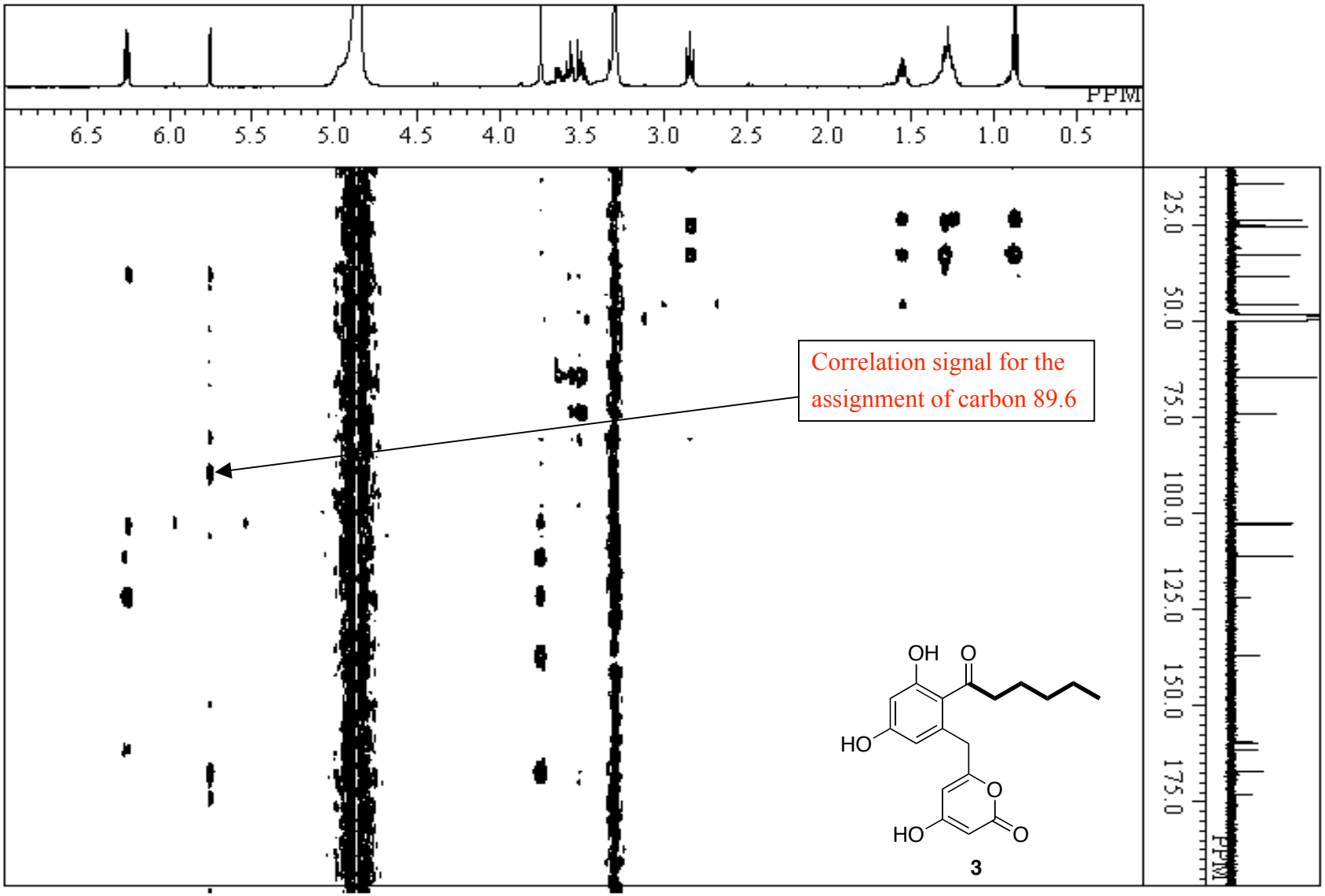


Figure 18. ^{1}H NMR spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OD, 400 MHz).

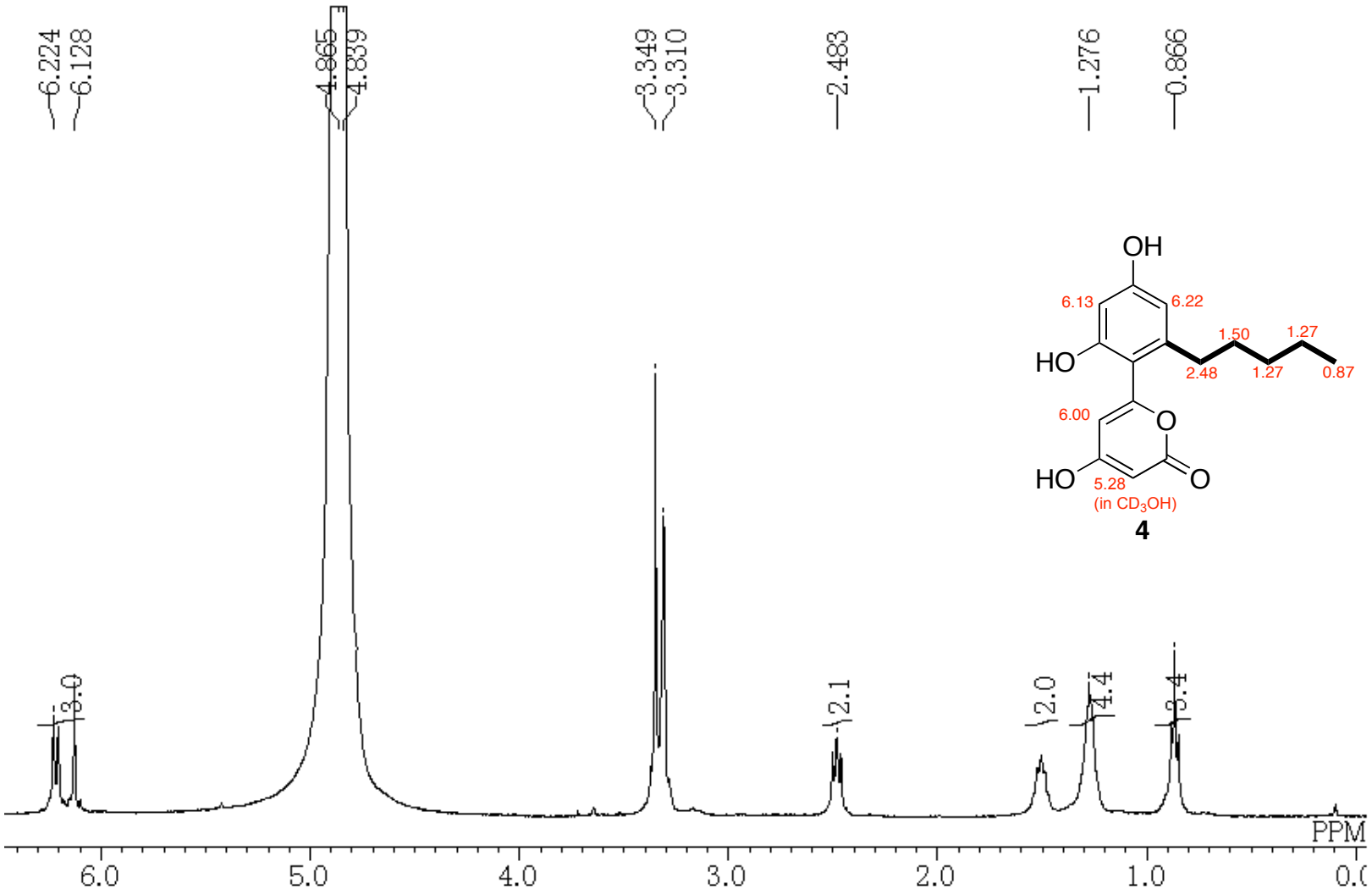


Figure 19. ^{1}H NMR spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OH, 800 MHz).

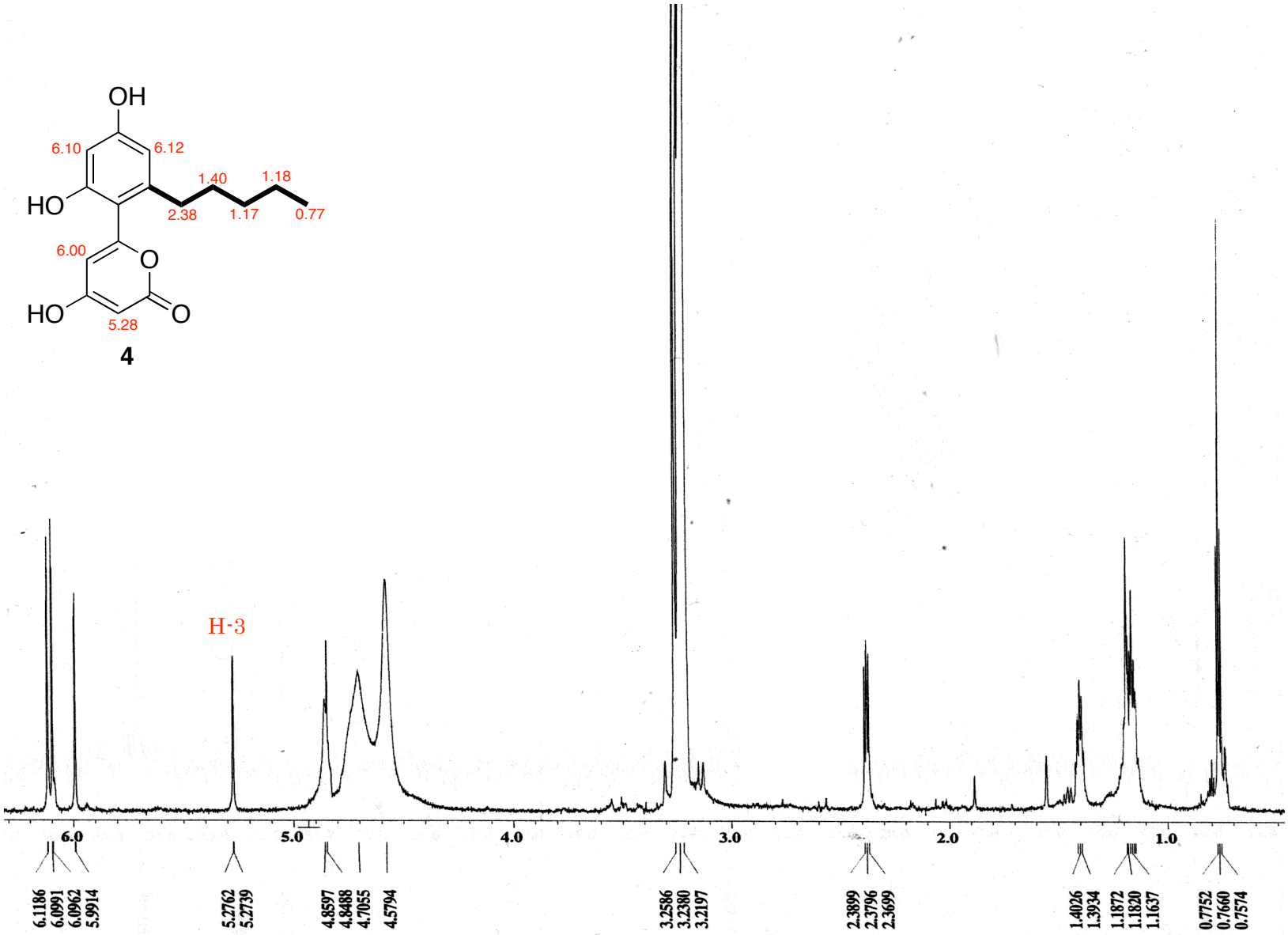


Figure 20. ^{13}C NMR spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OD, 100 MHz).

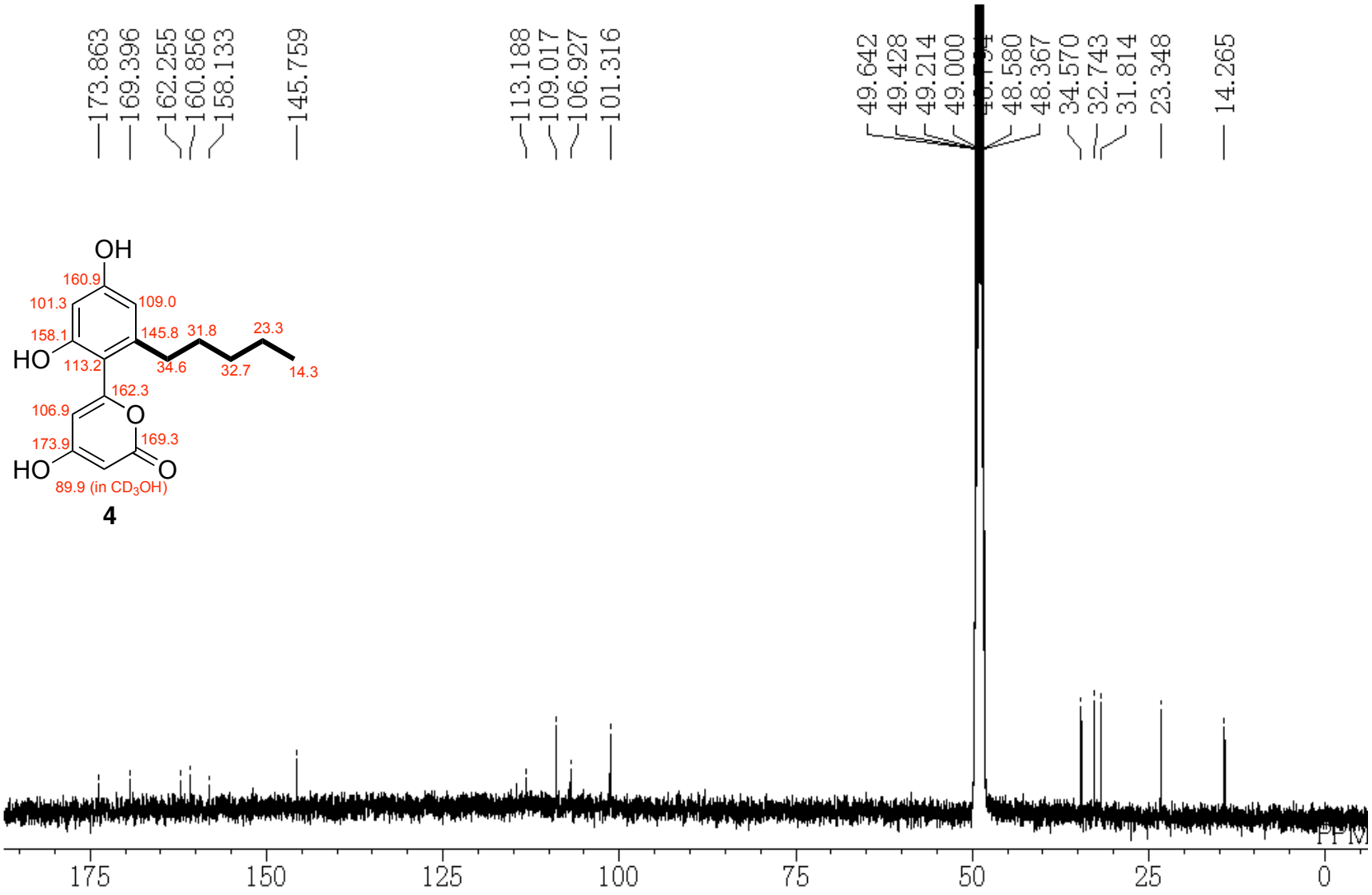


Figure 21. ^{13}C NMR spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OH, 200 MHz).

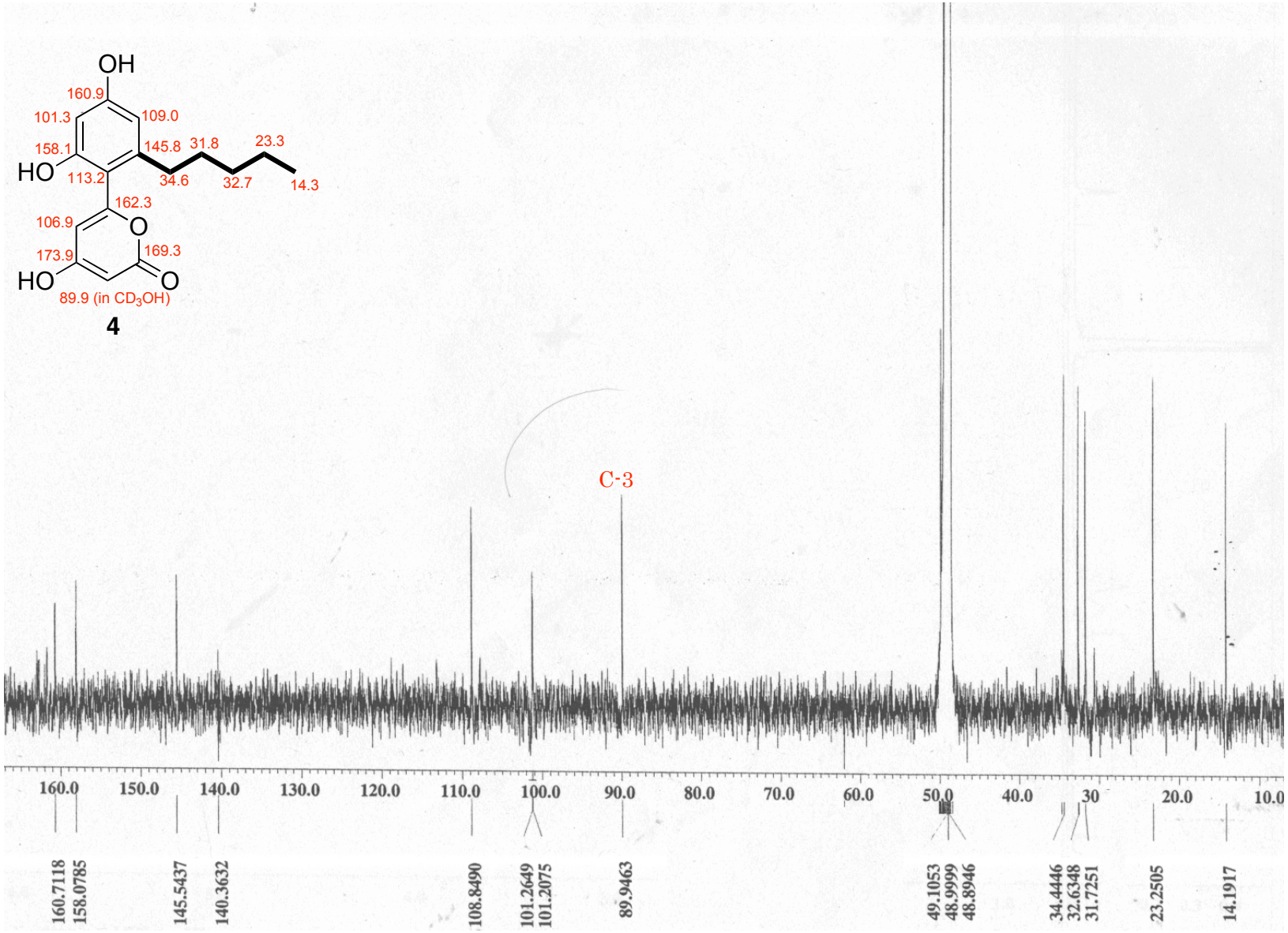


Figure 22. ^{1}H-^{1}H COSY spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OD, 400 MHz).

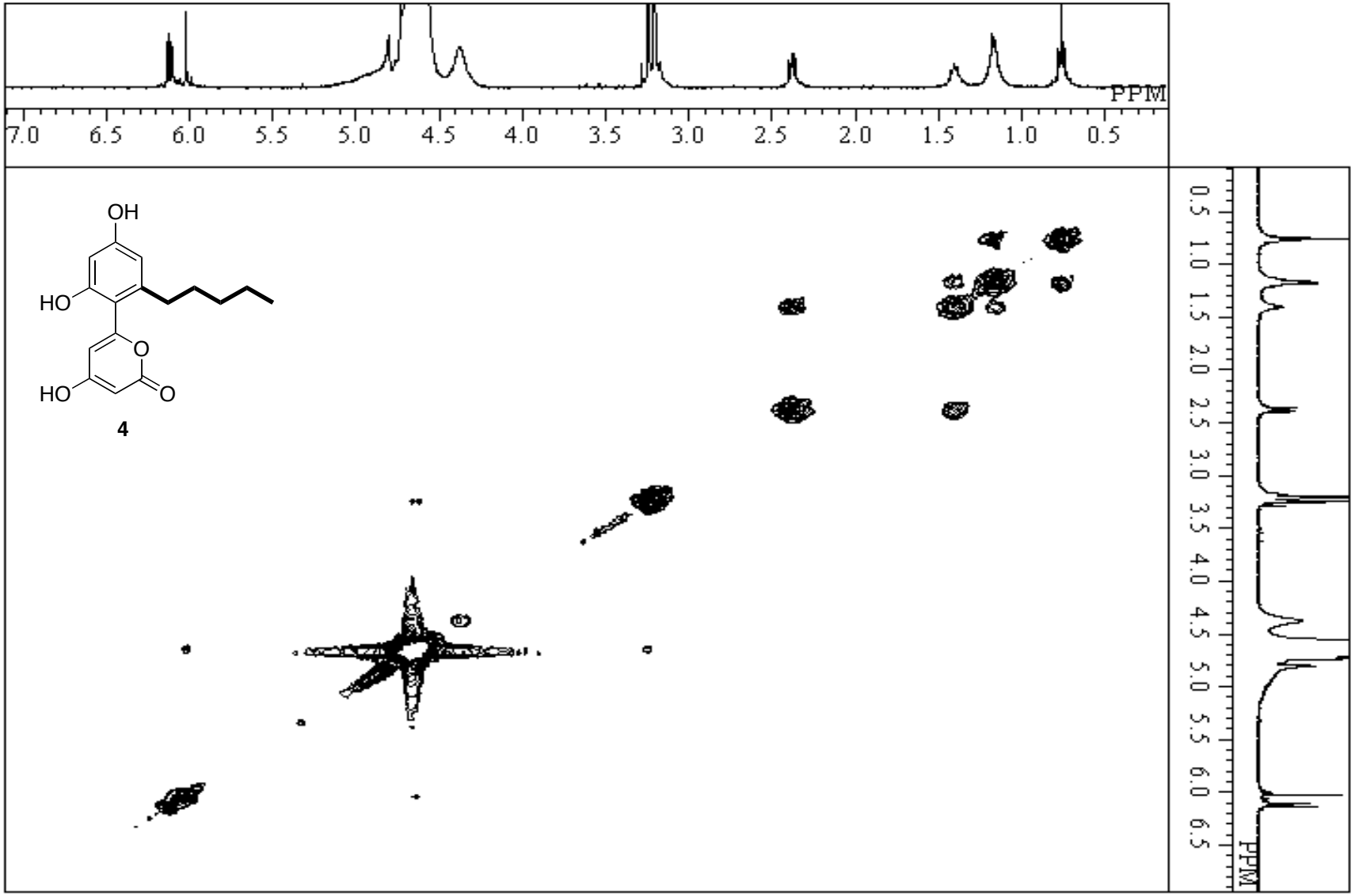


Figure 23. HMQC spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OD, 400 MHz).

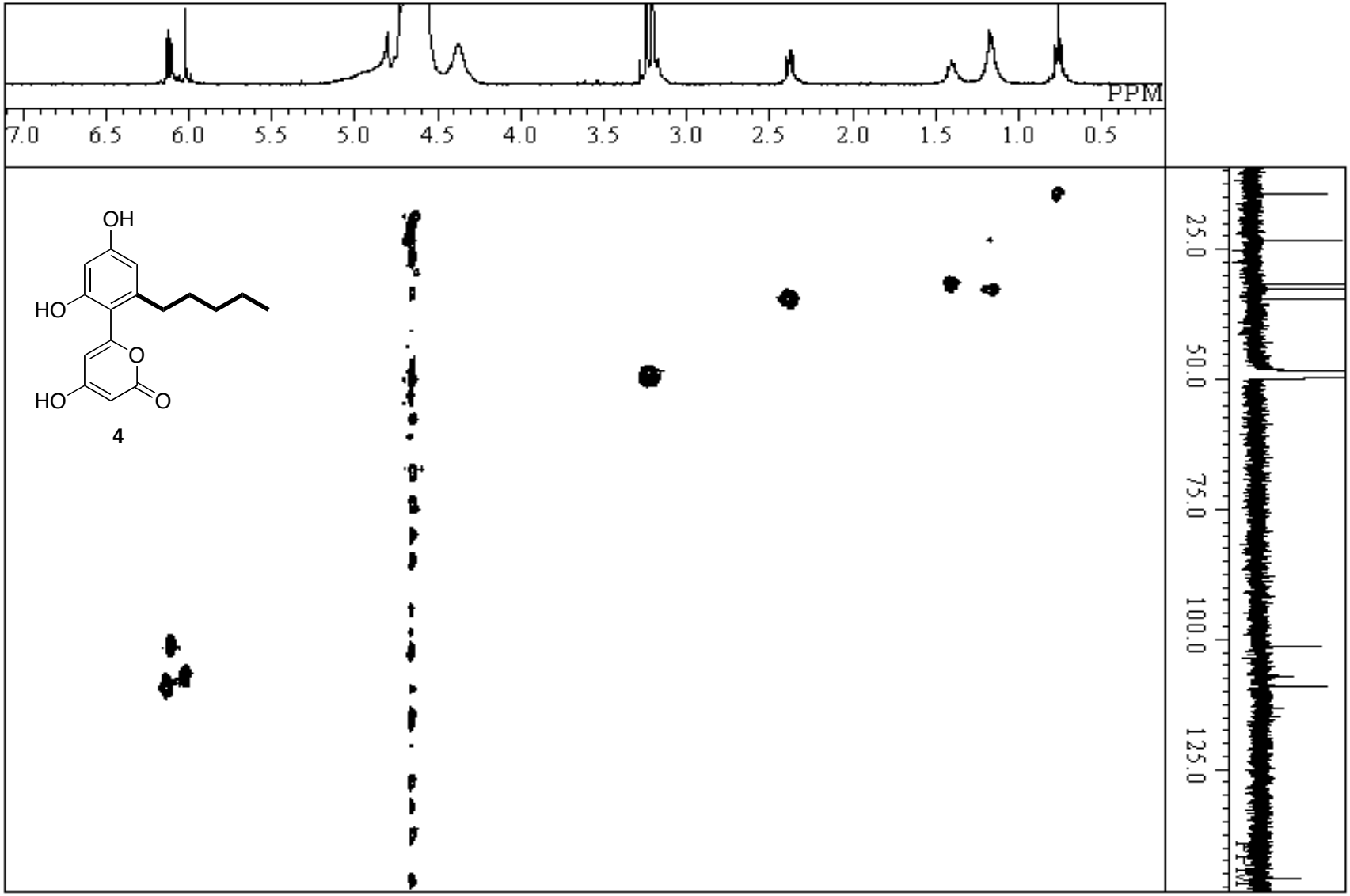


Figure 24. HMBC spectrum of **4** (C_{16} hexaketide resorcinol) (in CD_3OD, 400 MHz).

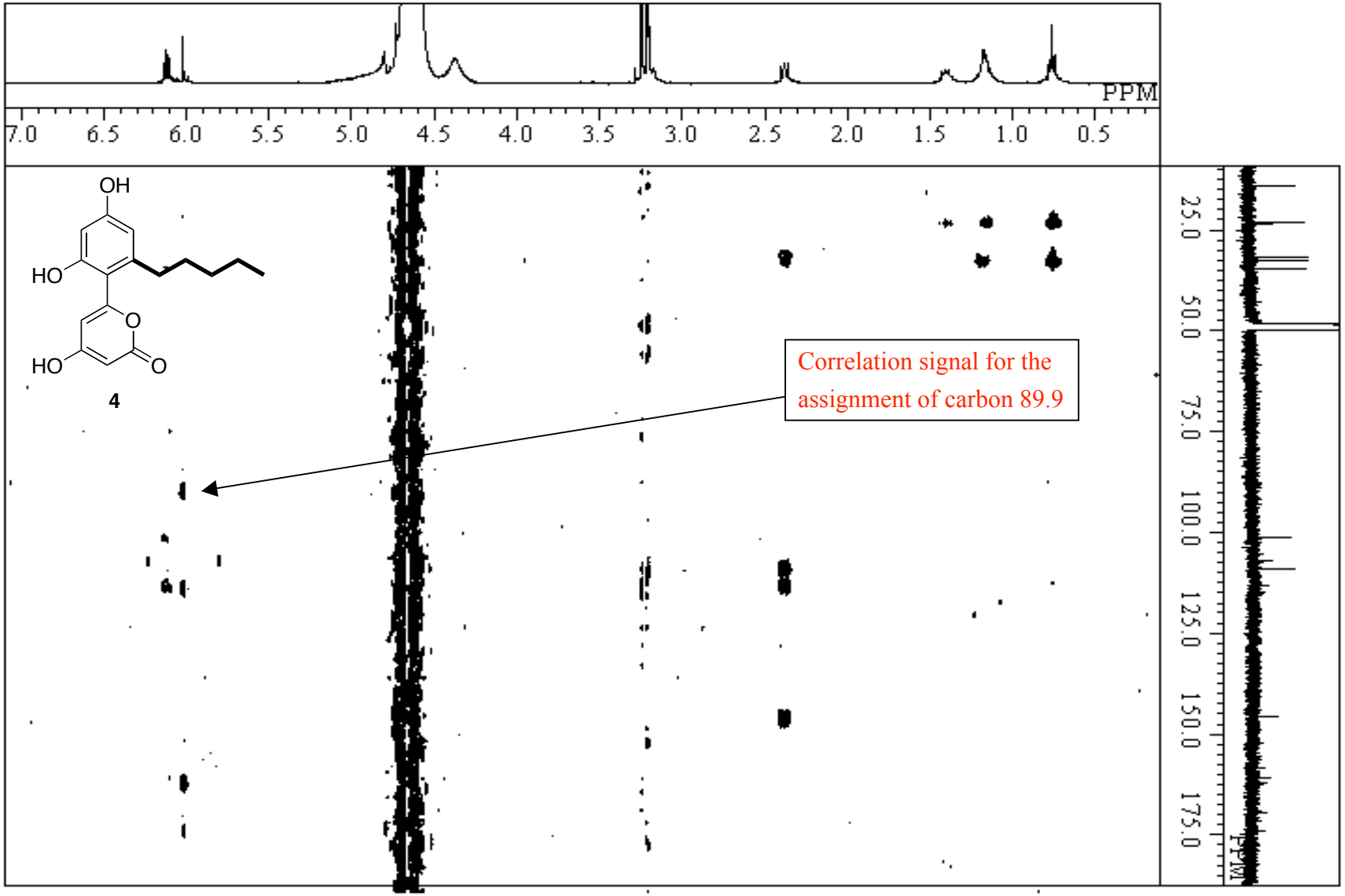


Figure 25. HPLC chromatograms: (A) enzyme reaction product of OKS N222G, (B) authentic SEK 15, and (C) co-injection of the enzyme reaction product and the authentic SEK 15.