Supporting Information

Microtermolides A and B from termite-associated *Streptomyces* sp. and structural revision of vinylamycin

Gavin Carr,[†] Michael Poulsen,[‡,§] Jonathan L. Klassen,[‡] Yanpeng Hou,[||] Thomas P. Wyche,[||] Tim S. Bugni,[||] Cameron R. Currie,[‡] and Jon Clardy[*,†]

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706, Biological Institute, Section for Ecology and Evolution, University of Copenhagen, Universitetsparken 15, Building 12, 2100 Copenhagen Ø, Denmark, and Pharmaceutical Sciences Division, University of Wisconsin-Madison, Madison, Wisconsin 53706

Table of Contents:

Experimental

General Experimental Section. NMR spectra were recorded in CD_3OD on a Varian Inova 600 MHz spectrometer equipped with an inverse-detection probe, and chemical shifts were referenced to the solvent peak at δ_H 3.30 and δ_C 49.0. An Agilent 1200 Series HPLC system equipped with a diode array detector and a Phenomenex® C_{18} column (5 μm, 250 x 21.2 mm) was used for preparative HPLC purification. An Agilent 1200 Series HPLC system equipped with a diode array detector and a Discovery® C_{18} column (10 μm, 250 x 10 mm) was used for semi-preparative HPLC purification. LC-MS analyses were performed on an Agilent 1200 Series HPLC system equipped with a diode array detector and a 6130 Series quadrupole mass spectrometer using a Phenomenex® C_{18} column (5 μm, 100 x 4.6 mm). High-resolution mass spectrometry was performed at the University of Illinois Urbana-Champaign Mass Spectrometry Facility.

***Streptomyces* sp. MspM5.** *Streptomyces* sp. strain MspM5 was isolated from a *Microtermes* sp. sampled in South Africa in 2008. A BLAST search of the partial 16S rRNA sequenced revealed that the closest match is *Streptomyces fasiculatus* NBRC 12765 (GenBank accession number AB184128) (99% identical).

Isolation of Microtermolides A and B. The producing organism was grown on solid ISP-2 medium (1.2 L) for 7 days at 30 ^oC. After 7 days, the agar was cut into small squares and soaked overnight in EtOAc. The EtOAc was filtered, dried *in vacuo*, dissolved in 20 mL of 80% MeOH/H_2O and passed through a C_{18} column to remove non-polar components. The eluent from this column, which contained the compounds of interest based on LC-MS analysis, was diluted two-fold with H_2O to give a final MeOH concentration of 40%. This solution was passed through another C_{18} column and washed with additional 40% MeOH/H_2O solution. The

compounds of interest were then eluted from this column with 100% MeOH. This fraction was purified by preparative reversed-phase HPLC using a gradient from 10% CH_3CN to 100% CH_3CN to give two fractions containing crude **1** and **2**, respectively. Crude **1** was dissolved in CH_2Cl_2 and passed through a silica gel column and eluted with a gradient from EtOAc to MeOH. Compound **1** eluted from the silica gel column with ca. 25% MeOH/EtOAc. This fraction was subjected to a final purification by semi-preparative HPLC using a gradient from 10% CH_3CN to 100% CH_3CN to give pure **1** (1.4 mg). Crude **2** was dissolved in CH_2Cl_2 and passed through a silica gel column and eluted with a gradient from EtOAc to MeOH. Compound **2** eluted from the silica gel column with ca. 10% MeOH/EtOAc. This fraction was subjected to a final purification by semi-preparative HPLC using a gradient from 10% CH_3CN to 100% CH_3CN to give pure **2** (3.4 mg).

Marfey's Analysis of 1 and 2. A small sample of 1 (0.1 mg) was hydrolyzed with 6N HCl (1 mL) at 110 ^{o}C for 6 hours. The reaction was allowed to cool to room temperature, dried *in vacuo*, and redissolved in water (100 μL). A solution of Marfey's reagent (180 μL, 0.9 mg/mL) was added, followed by 1 M aqueous $NaHCO_3$ (20 μL). The reaction was heated to 40 ^{oC} for 1 h, cooled to room temperature, and acidified with 1 N HCl (30 μL). The reaction mixture was diluted with MeOH (5 mL) and analyzed by LC-MS using the following gradient: 0-4 min, 10% CH_3CN/H_2O + 0.1% formic acid; 4-24 min, linear gradient from 10-50% CH_3CN/H_2O + 0.1% formic acid. Standards were prepared from the appropriate authentic D- or L- amino acids (0.1 mg) by derivatizing them with Marfey's reagent using the above procedure. The retention times for Marfey's derivatives were as follows: derivative prepared from **1**, 21.21 min, 25.96 min; derivative prepared from authentic L-alanine, 21.25 min; derivative prepared from authentic D-alanine, 22.51 min; derivative prepared from authentic L-valine, 23.91 min;

derivative prepared from D-valine, 25.96 min. In each case, samples prepared from **1** were also co-injected with standards to confirm their assignment. The derivatives prepared from **1** co-eluted with the standards prepared from L-alanine and D-valine as expected. Compound **2** was analyzing using the same procedure and the retention times of Marfey's derivatives were 21.21 min and 25.94 min. The derivatives prepared from **2** co-eluted with L-alanine and D-valine.

UHPLC/HRMS-based Secondary Metabolomics.

Sample Preparation for UHPLC/HRESI-TOF-MS. Each *Streptomyces* sp. was inoculated onto ISP-2 agar from a spore prep. Plates were incubated at 29 ºC for 7-14 days. Two cores (8 mm diameter) of bacteria and agar were obtained from each plate, placed directly into MeOH (2 mL) and extracted for 30 mins. The extract was transferred into a clean vial and evaporated using a SpeedVac concentrator; the residue was dissolved in MeOH (100 µL), followed by the addition of H_2O (1 mL). The solution was then placed on a Gilson GX-271 liquid handling system and subjected to automated SPE (Biotage: EVOLUTE® ABN, 25 mg absorbent mass, 1 mL reservoir volume). The SPE cartridges were washed using H_2O (1 mL) to remove media components and eluted with MeOH (1 mL) directly into an LC/MS-certified vial. Two biological replicates were prepared for each bacterial strain.

UHPLC/HRMS Analysis. LC/MS data were acquired using a Bruker MaXis™ ESI-Q-TOF mass spectrometer coupled with a Waters Acquity™ UPLC system operated by Bruker Hystar software. A gradient of MeOH and H_2O (containing 0.1% formic acid) was employed with a flow rate of 0.3 mL/min on an RP C_{18} column (Phenomenex Kinetex 2.6 µm, 2.1 x 100 mm). The gradient started from MeOH/H_2O (10%/90%), followed by a linear gradient to reach MeOH/H_2O (97%/3%) in 12 mins, and held for 2 mins at MeOH/H_2O (97%/3%). Full scan mass spectra (*m/z* 150 to 1550) were measured in positive ESI mode. The mass spectrometer was operated using the following parameters: capillary: 4.5 kV; nebulizer pressure: 4.0 bar; dry gas flow: 6.0 L/min; dry gas temperature: 200 ^{o}C; scan rate: 2 Hz. Tune mix (Agilent, ESI-L low concentration) was introduced through a divert valve at the end of each chromatographic run for automatic internal calibration.

Data Processing and PCA. Bruker Data Analysis 4.0 was used for analysis of chromatograms. Bucketing LC/MS data and PCA was performed using Bruker ProfileAnalysis 2.0. The line spectra of LC/MS data were used for PCA. Finding molecular features was applied to LC/MS data under these parameters: S/N threshold: 5; correlation coefficient threshold: 0.7; minimum compound length: 10 spectra; smoothing width: 1. The bucket generation was performed under the following parameters. The LC/MS data sets were evaluated in a time range from 120 s to 840 s and in a mass range from *m/z* 200 to 1500. Advanced bucketing was employed using Δ RT = 20 s and Δ *m/z* = 0.02 Da as parameters. Sum of bucket values was applied for normalization in this study, and pareto scaling algorithm was applied.

Analysis of *Streptomyces* sp. MspM5. Extracts from 30 termite-associated strains were subjected to SPE and analyzed using LC/MS-PCA (Figure S1). Principal components were viewed until strain MspM5 was separated from the other analyzed strains. For example in Figure S1, PC1 versus PC8 clearly separated strain MspM5 in the scores plot. Generally, strains with similar chemistry group together in the scores plot and strains with different chemistry separate in the Scores plot. While most of the variance among samples is accounted for in PC1 versus PC2, we have found that differences at higher principal components can be significant. The loadings plot described the variance among the strains. For example, the data points that are geometrically aligned in the direction of strain MspM5 represent the variance or the secondary metabolites responsible for the separation. Using a dynamically linked table, we could use a mouse to highlight the data points that were responsible for the variance and immediately obtain a high-resolution and high-accuracy mass measurement. The *m/z* obtained from the PC was used to search Antibase. While compound **1** was originally identified from the PCA, compound **2** was found in the PCA only after it had been isolated.

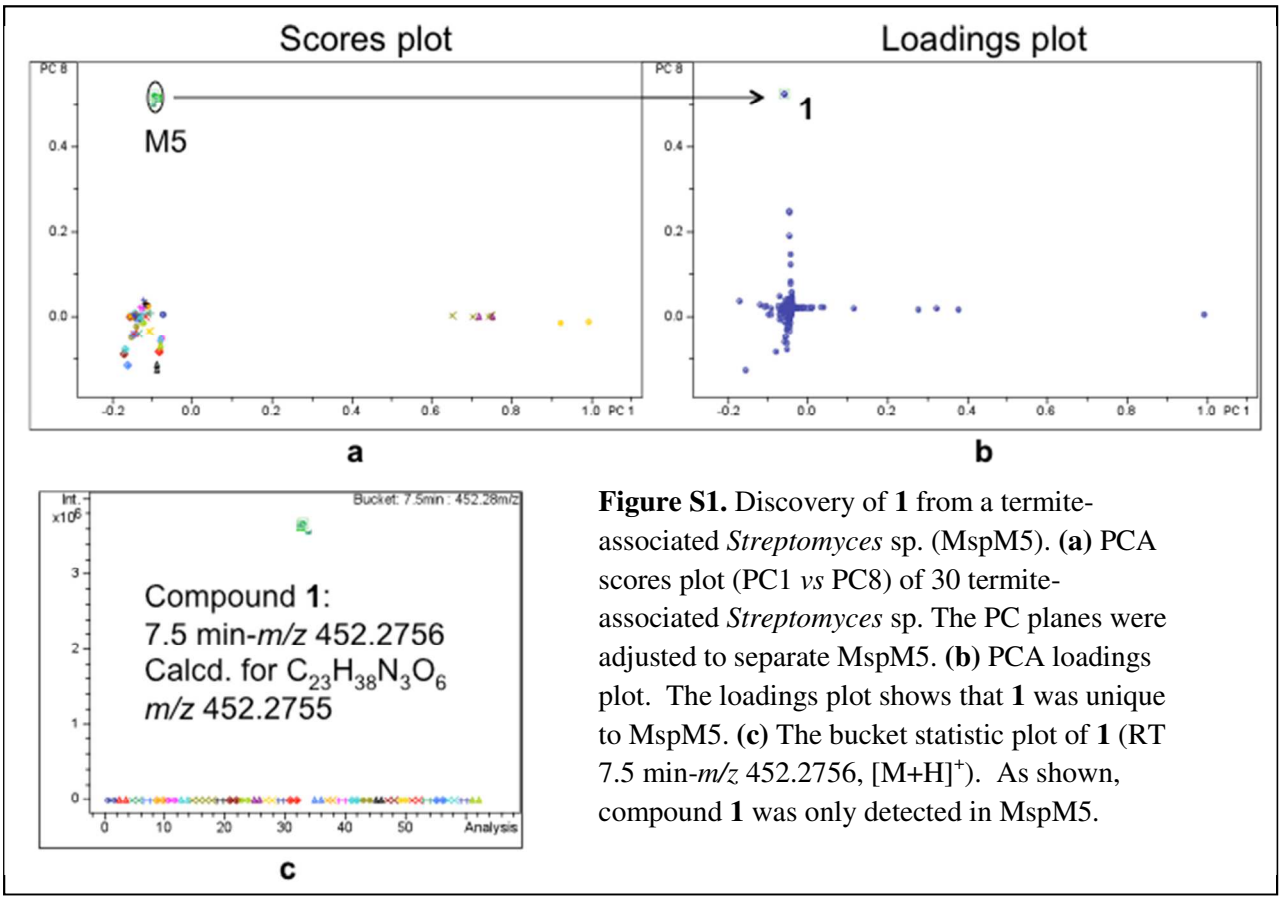


Figure S1. Discovery of **1** from a termite-associated *Streptomyces* sp. (MspM5). **(a)** PCA scores plot (PC1 *vs* PC8) of 30 termite-associated *Streptomyces* sp. The PC planes were adjusted to separate MspM5. **(b)** PCA loadings plot. The loadings plot shows that **1** was unique to MspM5. **(c)** The bucket statistic plot of **1** (RT 7.5 min-*m/z* 452.2756, $[M+H]^+$). As shown, compound **1** was only detected in MspM5.

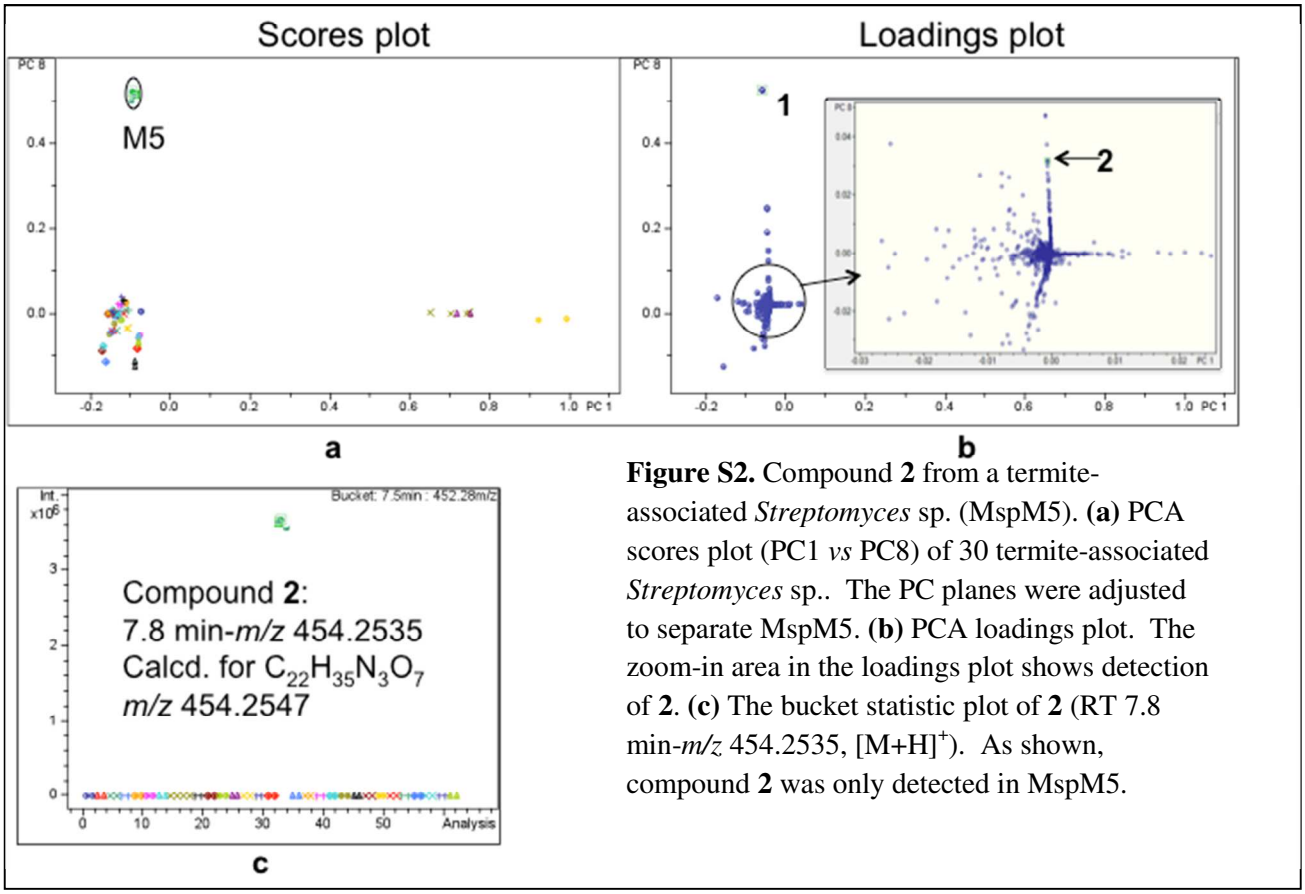


Figure S2. Compound **2** from a termite-associated *Streptomyces* sp. (MspM5). **(a)** PCA scores plot (PC1 *vs* PC8) of 30 termite-associated *Streptomyces* sp.. The PC planes were adjusted to separate MspM5. **(b)** PCA loadings plot. The zoom-in area in the loadings plot shows detection of **2**. **(c)** The bucket statistic plot of **2** (RT 7.8 min-*m/z* 454.2535, $[M+H]^+$). As shown, compound **2** was only detected in MspM5.

Table S1. NMR Data for Microtermolide A (**1**)

		Microtermolide A (**1**)[a]			
position	δ_C	δ_H (*J* in Hz)	COSY	HMBC	NOESY
1	170.4				
2	59.2	4.42, m	H-3	C-1, C-3, C-4, C-5, C-1'	H-3, H-4, H-5
3	33.4	2.07, dq (13.7, 6.8)	H-2, H-4, H-5	C-1, C-2, C-4, C-5	H-2, H-4, H-5
4	19.7	0.96, d (6.5)	H-3	C-2, C-3, C-5	H-2, H-3
5	18.5	0.95, d (7.0)	H-3	C-2, C-3, C-4	H-2, H-3
1'	175.2				
2'	52.6	4.40, m	H-3'	C-3'	H-3', H-2"
3'	18.5	1.44, d (7.0)	H-2'	C-1', C-2'	H-2'
1"	169.2				
2"	118.2	6.24, d (15.3)	H-3"	C-1", C-3", C-4"	H-2', H-3"
3"	141.6	7.11, d (14.7)	H-2"	C-1", C-4", C-5"	H-2", H-5b"
4"	138.6				
5"	119.3	5.59, s		C-3", C-4"	H-5b"
		5.52, s		C-3", C-4"	H-5a", H-3"
1'"	174.3				
2'"	46.9	3.00, td (10.3, 4.1)	H-3'", H-8'"	C-1'", C-3'", C-8'"	H-3'", H-8'", H-10'"
3'"	78.0	5.48, dd (10.3, 2.1)	H-2'", H-4'"	C-1, C-1'", C-2'", C-4'", C-5'", C-10'"	H-2'", H-4'", H-5a'", H-8'"
4'"	34.9	1.88, m	H-3'", H-5a'", H-5b'", H-10'"	C-5'", C-10'"	H-3'", H-10'"
5'"	37.3	1.34, m	H-4'", H-5b'"	C-3'", C-4'", C-6'", C-10'"	H-3'"
		1.16, m	H-4'", H-5a'", H-6'"	C-3'", C-4'", C-6'", C-10'"	
6'"	21.2	1.42, m	H-5'", H-7'"	C-4'", C-5'", C-7'"	H-7'"
7'"	14.2	0.90, t (7.0)	H-6'"	C-5'", C-6'"	H-6'"
8'"	33.4	1.80, m	H-2'", H-9a'", H-9b'"	C-1'", C-2'", C-3'", C-9'"	H-2'", H-3'", H-9a'", H-9b'"
9'"	59.9	3.66, m	H-8'"	C-2'", C-8'"	H-8'"
		3.56, m	H-8'"	C-2'", C-8'"	H-8'"
10'"	13.4	1.06, d (7.0)	H-4'"	C-3'", C-4'", C-5'"	H-4'"

[a]Spectra collected in CD_3OD at 600 MHz.

Table S2. NMR Data for Microtermolide B (**2**)

			Microtermolide B (**2**)[a]		
position	δ_C	δ_H (J in Hz)	COSY	HMBC	NOESY
1	168.7				
2	134[b]	6.90, d (15.3)[c]	H-3	C-1, C-3, C-4	
3	134.3[b]	6.94, d (15.3)[c]	H-2	C-1, C-2, C-4	
4	166.2				
1'	174.5				
2'	50.5	4.56, q (7.0)	H-3'	C-4, C-1', C-3'	H-3'
3'	18.2	1.39, d (7.0)	H-2'	C-1', C-2'	H-2'
1"	172.2				
2"	59.5	4.35, d (6.5)	H-3"	C-1', C-1", C-3", C-4", C-5"	H-3", H-4", H-5"
3"	31.0	2.21, m	H-2", H-4", H-5"	C-1", C-2", C-4", C-5"	H-2", H-4", H-5"
4"	20.0	0.99, d (7.0)	H-3"	C-2", C-3", C-5"	H-2", H-3"
5"	18.3	0.95, d (7.0)	H-3"	C-2", C-3", C-4"	H-2", H-3"
1'"	178.6				
2'"	41.7	3.03, ddd (10.7, 9.0, 6.7)	H-3'", H-8a'", H-8b'"	C-1'", C-3'", C-8'"	H-3'", H-5a'", H-5b'", H-8a'", H-8b'", H-10'"
3'"	77.1	5.08, dd (7.0, 5.3)	H-2'", H-4'"	C-1", C-1'", C-2'", C-4'", C-8'", C-10'"	H-2'", H-4'", H-5a'", H-5b'", H-8a'", H-8b'", H-10'"
4'"	35.8	1.94, m	H-3'", H-5'", H-10'"	C-5'", C-6'", C-10'"	H-2'", H-3'", H-5a'", H-8a'", H-9b'", H-10'"
5'"	36.6	1.32, m	H-5b'"	C-3'", C-4'", C-6'", C-7", C-10'"	H-2'", H-3'", H-5b'"
		1.13, m	H-4'", H-5a'", H-6'"	C-6'", C-7'", C-10'"	H-2'", H-4'", H-5a'"
6'"	21.0	1.34, m	H-5b'", H-7'"	C-5'", C-7'"	H-7'"
7'"	14.5	0.88, t (7.3)	H-6'"	C-5'", C-6'", C-10'"	H-6'"
8'"	27.0	2.33, m	H-2'", H-8b'", H-9a'", H-9b'"	C-1'", C-2'"	H-2'", H-3'", H-4'", H-8b'", H-9a'", H-9b'"
		2.04, m	H-2'", H-8a'", H-9a'", H-9b'"	C-1'", C-2'", C-3'", C-9'"	H-2'", H-3'", H-8a'", H-9a'", H-9b'"
9'"	67.6	4.31, td (8.7, 2.6)	H-8a'", H-8b'", H-9b'"	C-1'", C-2'", C-8'"	H-8a'", H-8b'", H-9b'"
		4.18, td (9.1, 7.0)	H-8a'", H-8b'", H-9a'"	C-1'", C-8'"	H-4'", H-8a'", H-8b'", H-9a'"
10'"	14.5	0.92, d (7.0)	H-4'"	C-3'", C-4'", C-5'"	H-2'", H-3'", H-4'"

[a]Spectra collected in CD_3OD at 600 MHz.

[b,c]Signals are interchangeable

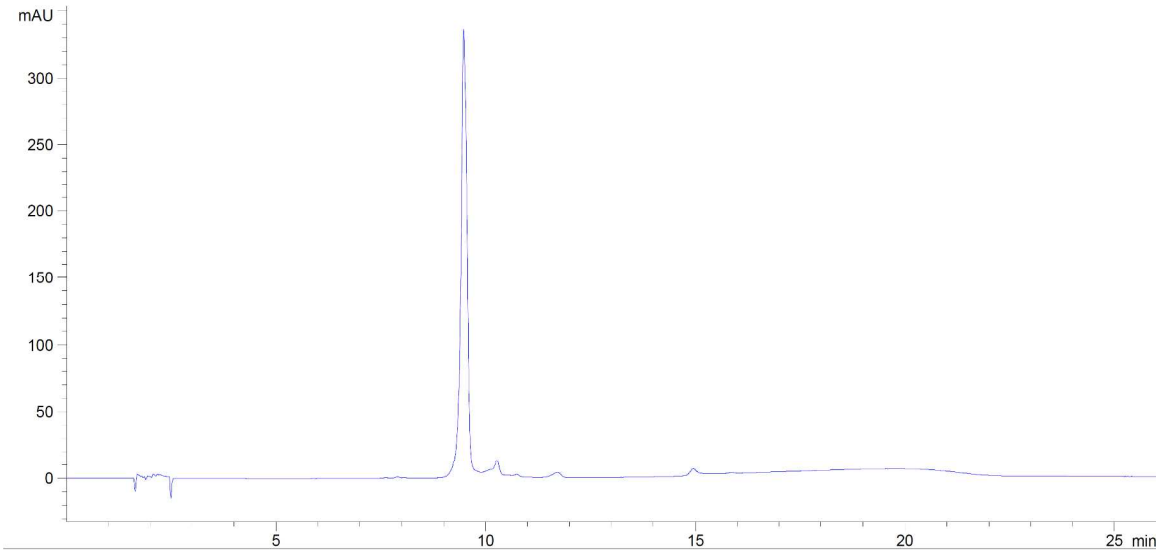


Figure S3. LC-MS trace of **1** at 254 nm. Data acquired on a Phenomenex® C_{18} column (5 µm, 100 x 4.6 mm) using the following gradient: 0-20 minutes, linear gradient from 10% CH_3CN/H_2O + 0.1% formic acid to 100% CH_3CN; 20-25 minutes, 100% CH_3CN + 0.1% formic acid.

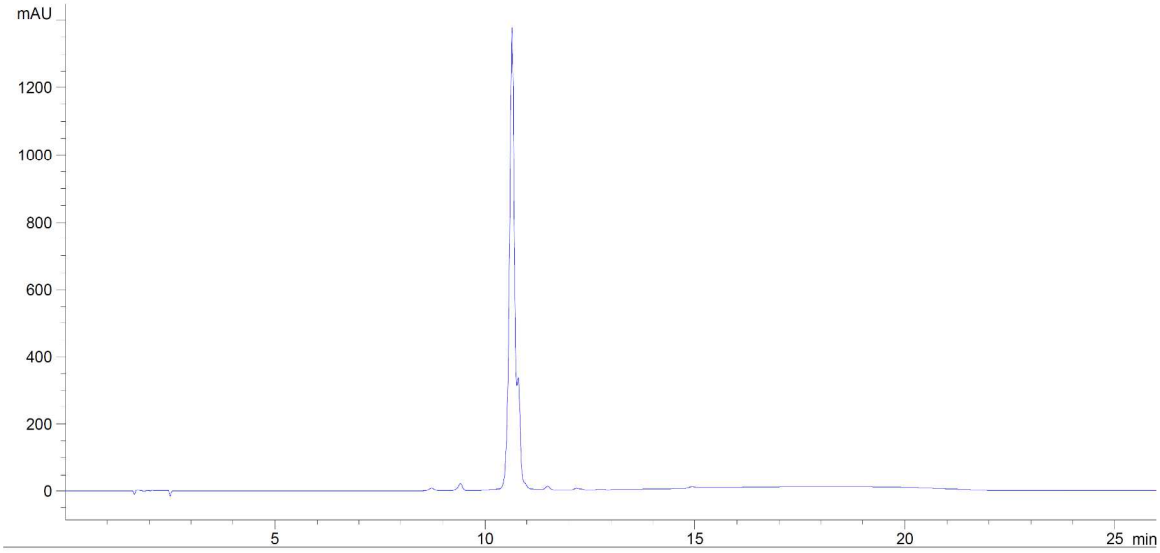
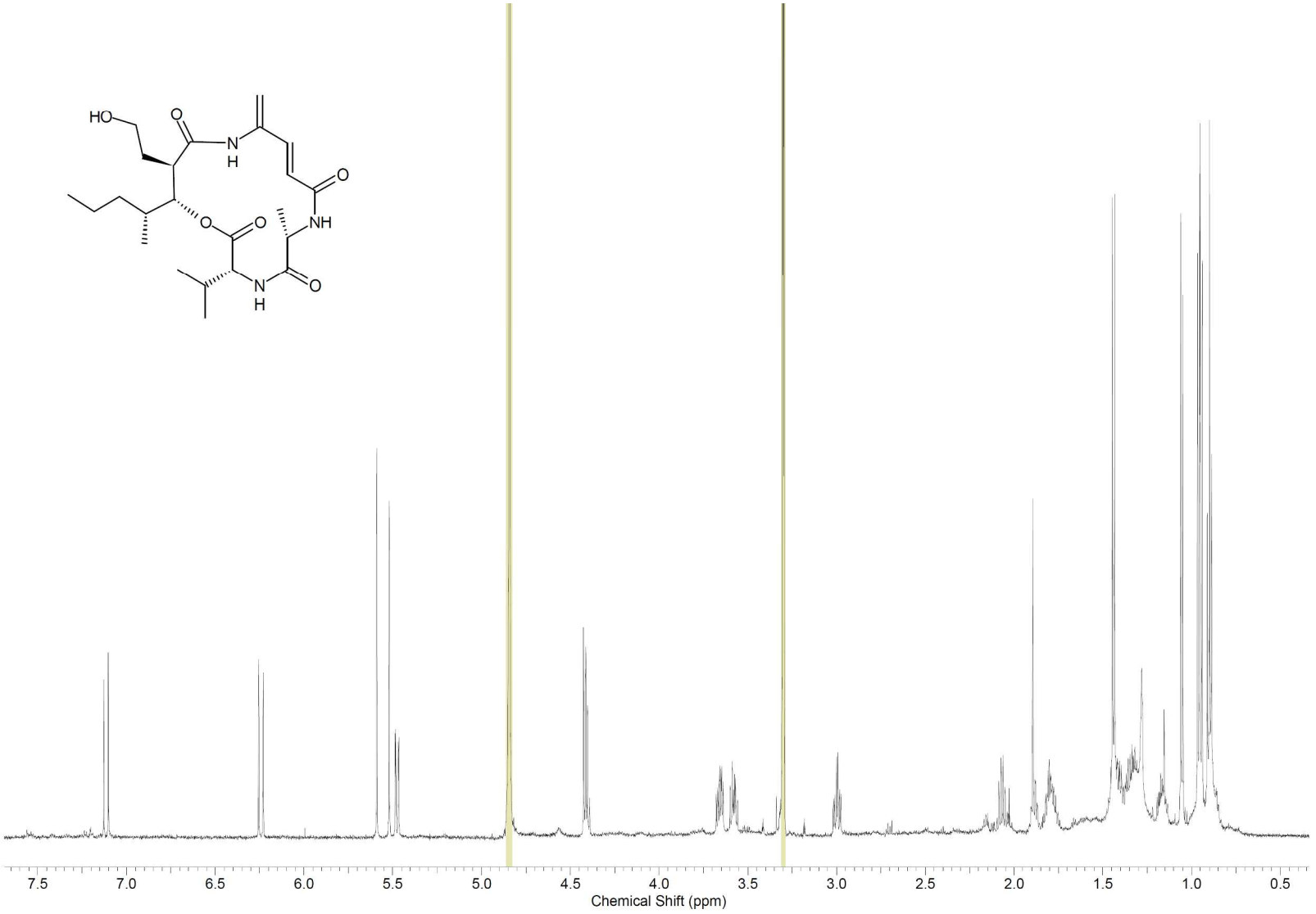
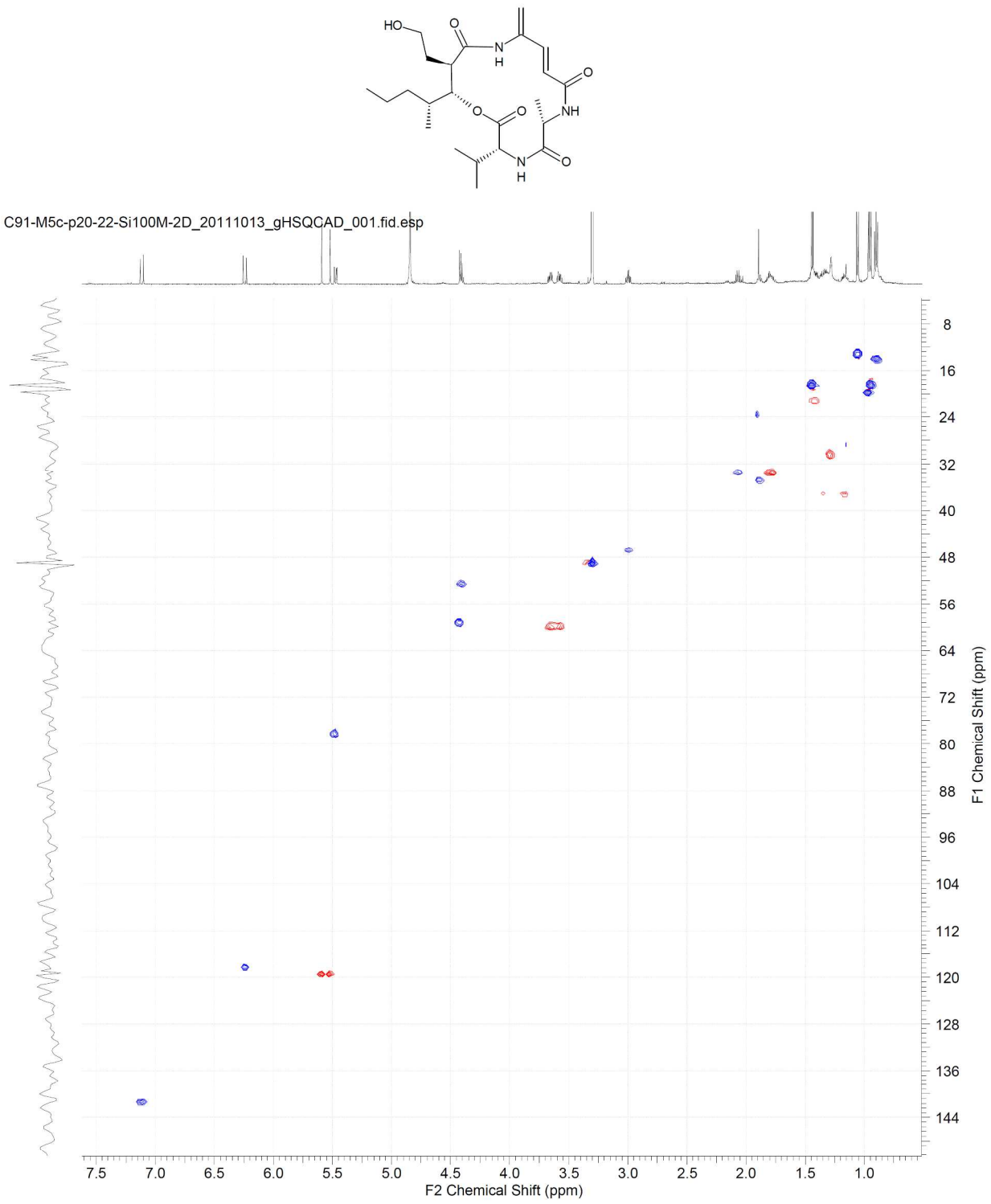


Figure S4. LC-MS trace of **2** at 254 nm. Data acquired on a Phenomenex® C_{18} column (5 µm, 100 x 4.6 mm) using the following gradient: 0-20 minutes, linear gradient from 10% CH_3CN/H_2O + 0.1% formic acid to 100% CH_3CN; 20-25 minutes, 100% CH_3CN + 0.1% formic acid.

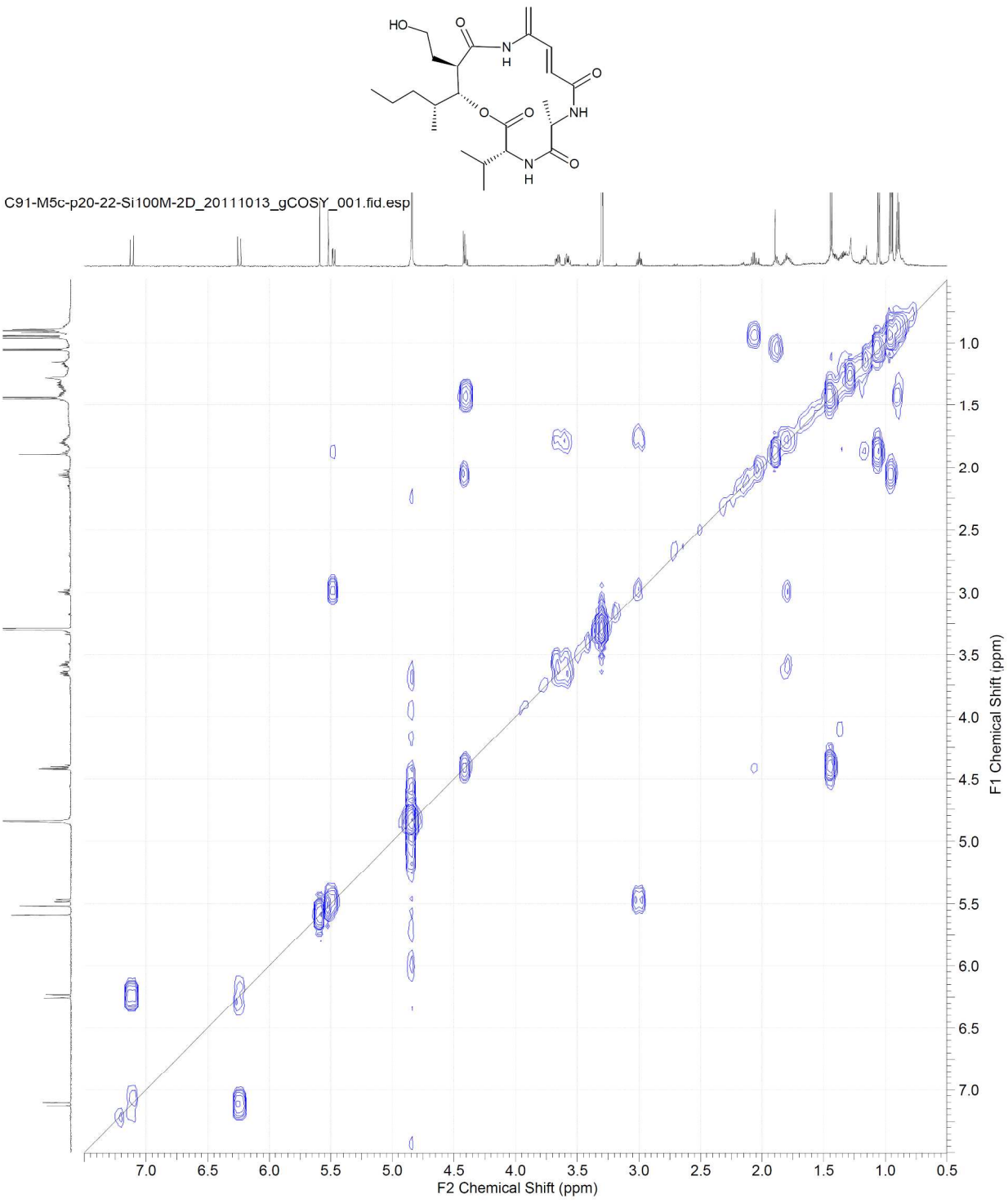
^{1}H NMR Spectrum of **1** at 600 MHz in CD_3OD



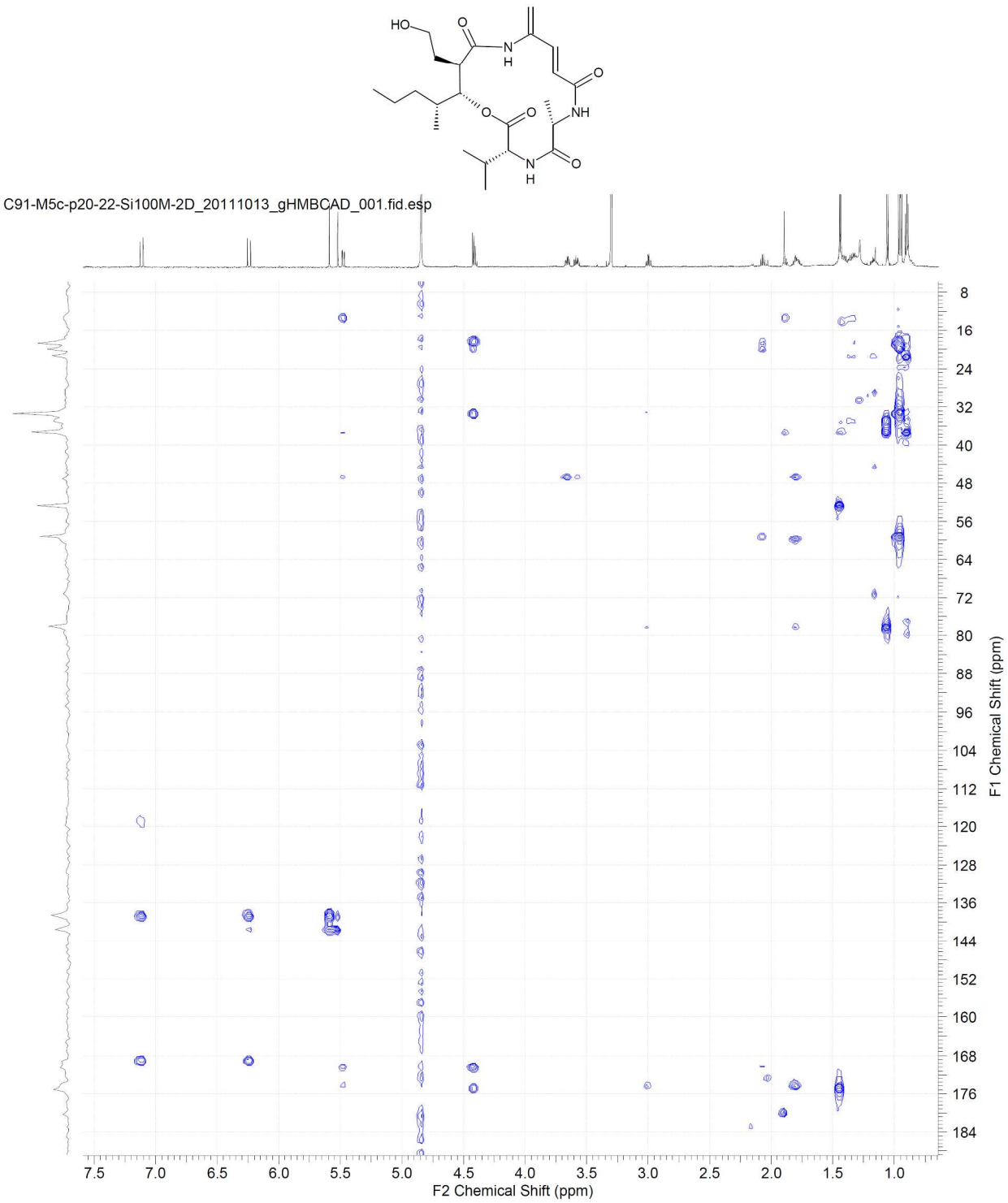
HSQC Spectrum of **1** at 600 MHz in CD_3OD



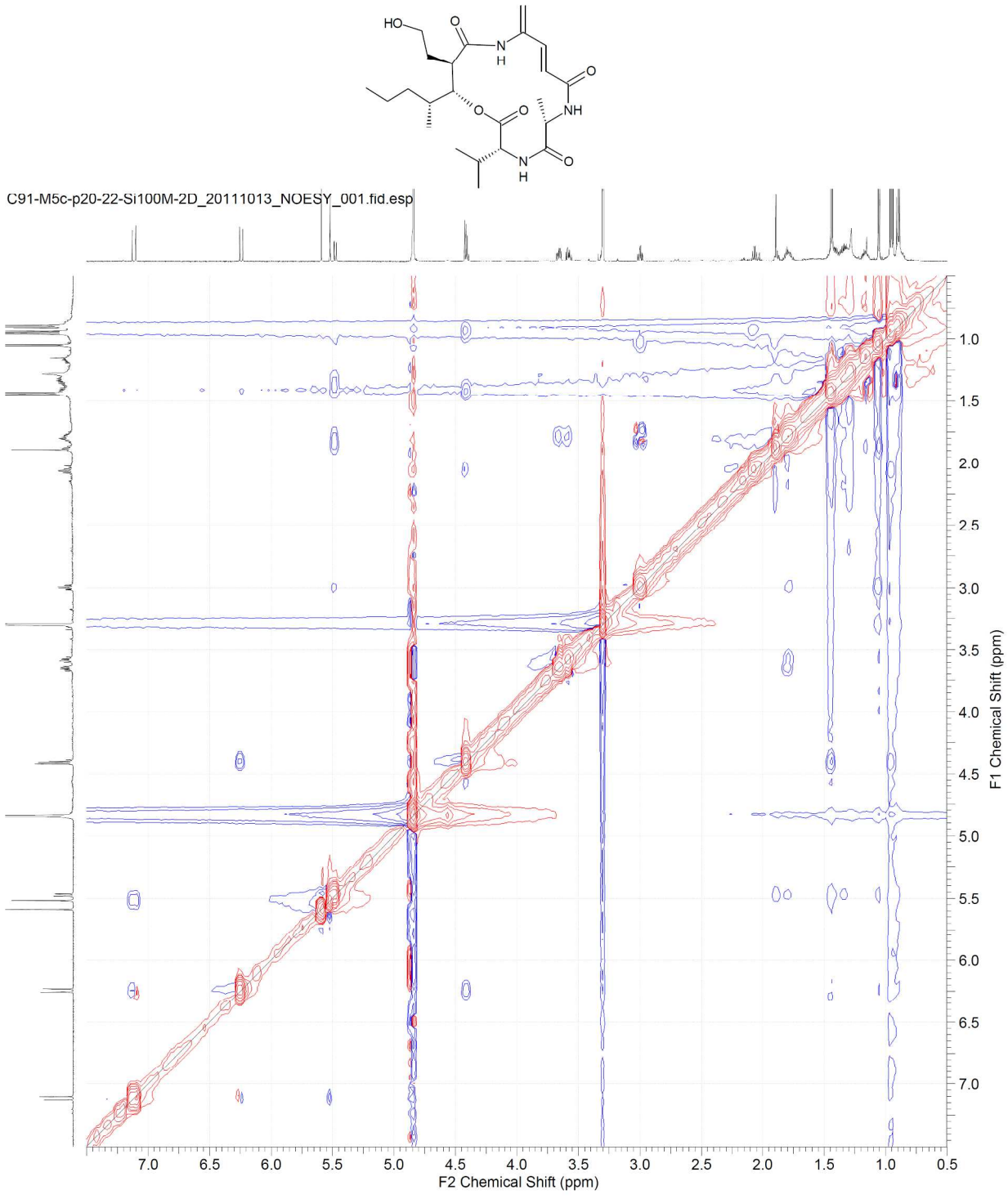
COSY Spectrum of **1** at 600 MHz in CD_3OD



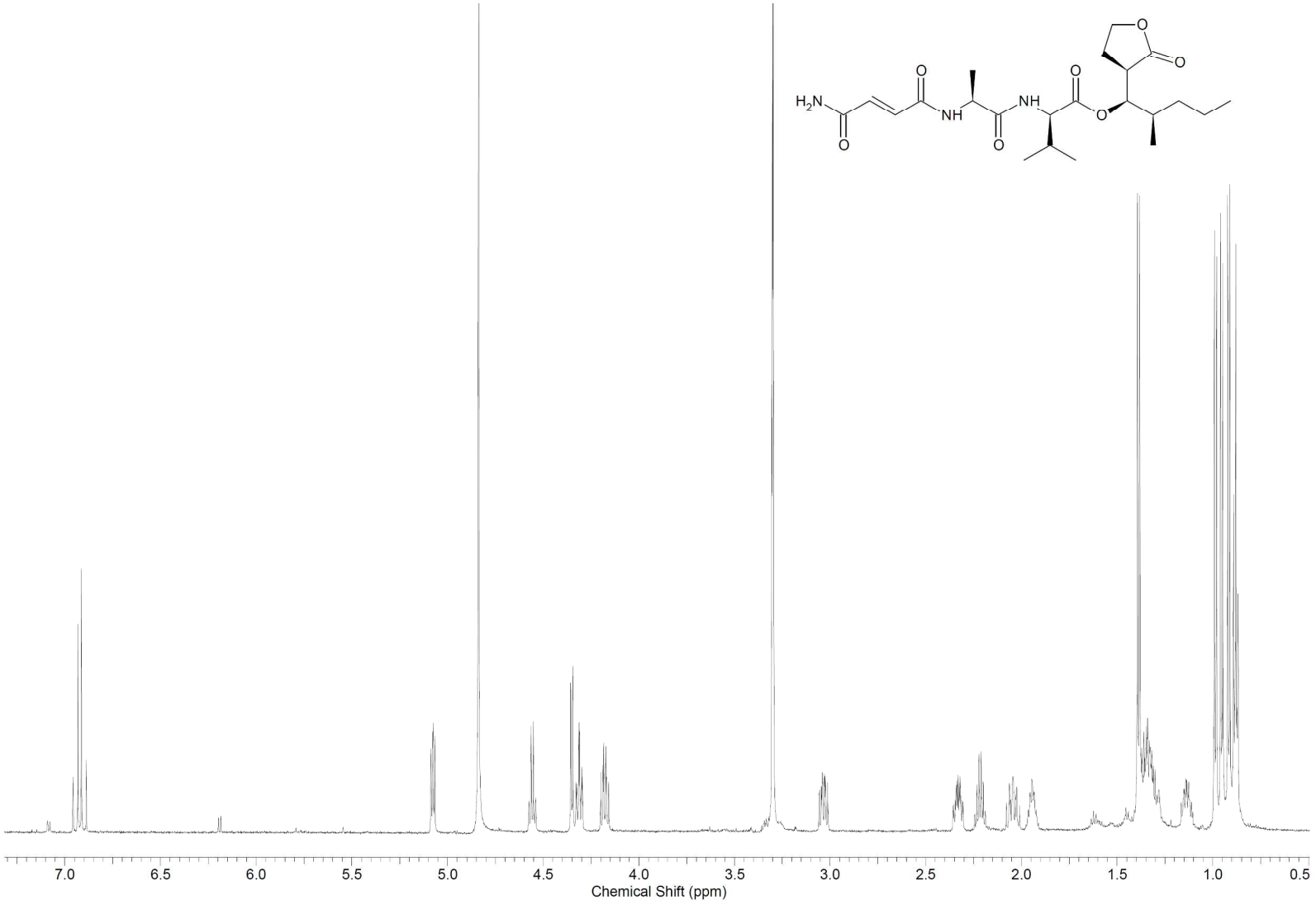
HMBC Spectrum of **1** at 600 MHz in CD_3OD



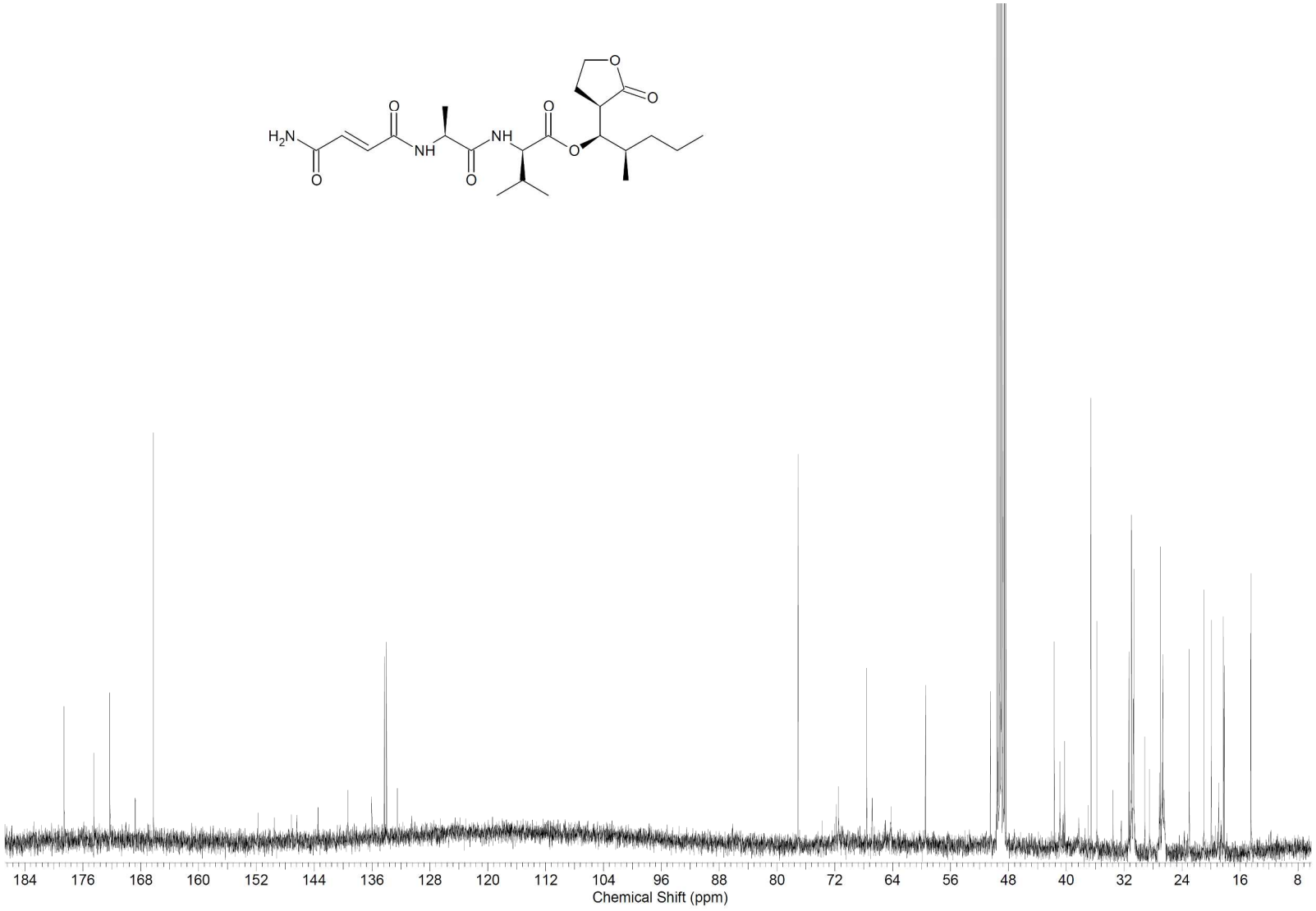
NOESY Spectrum of **1** at 600 MHz in CD_3OD



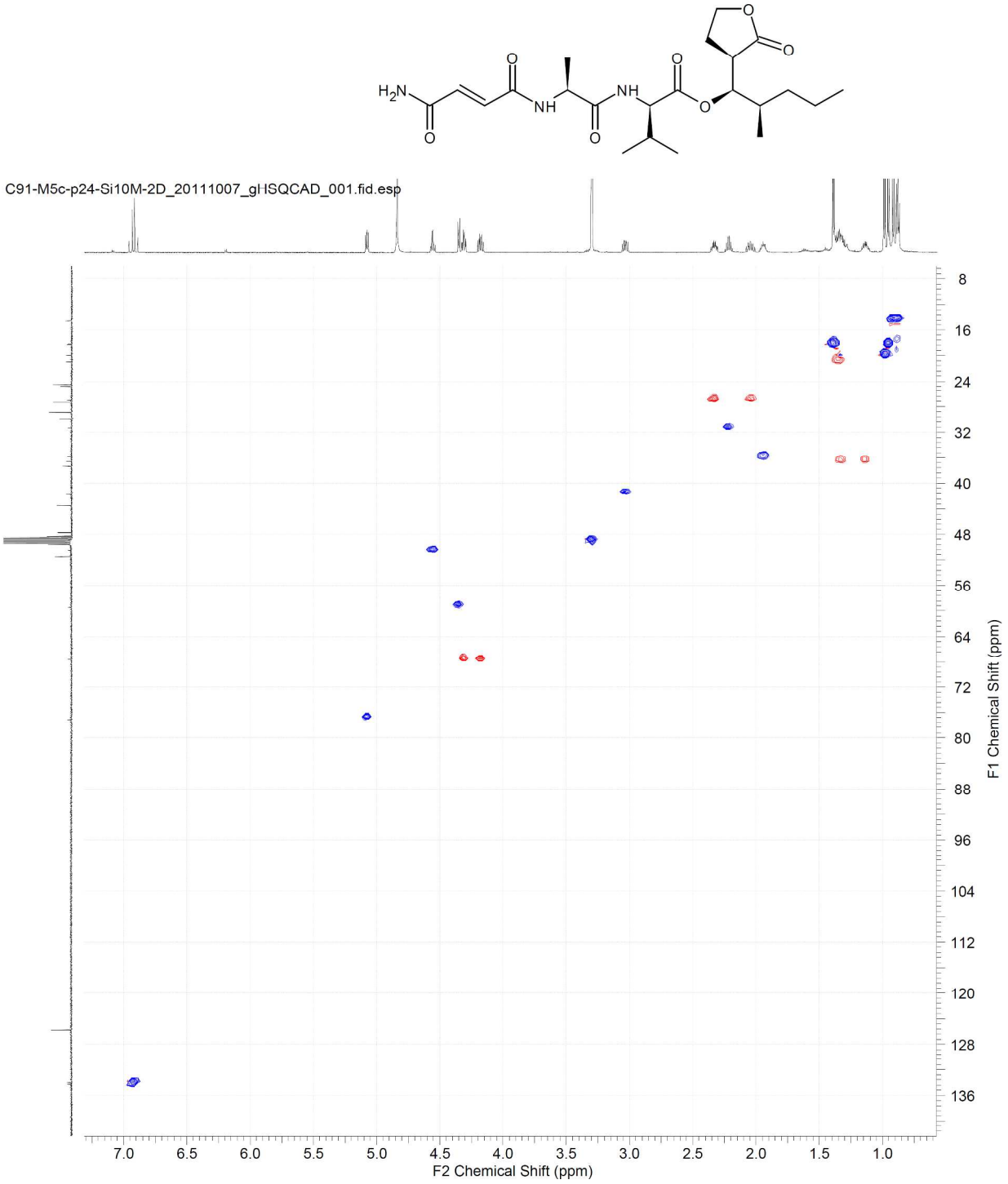
^{1}H NMR Spectrum of **2** at 600 MHz in CD_3OD



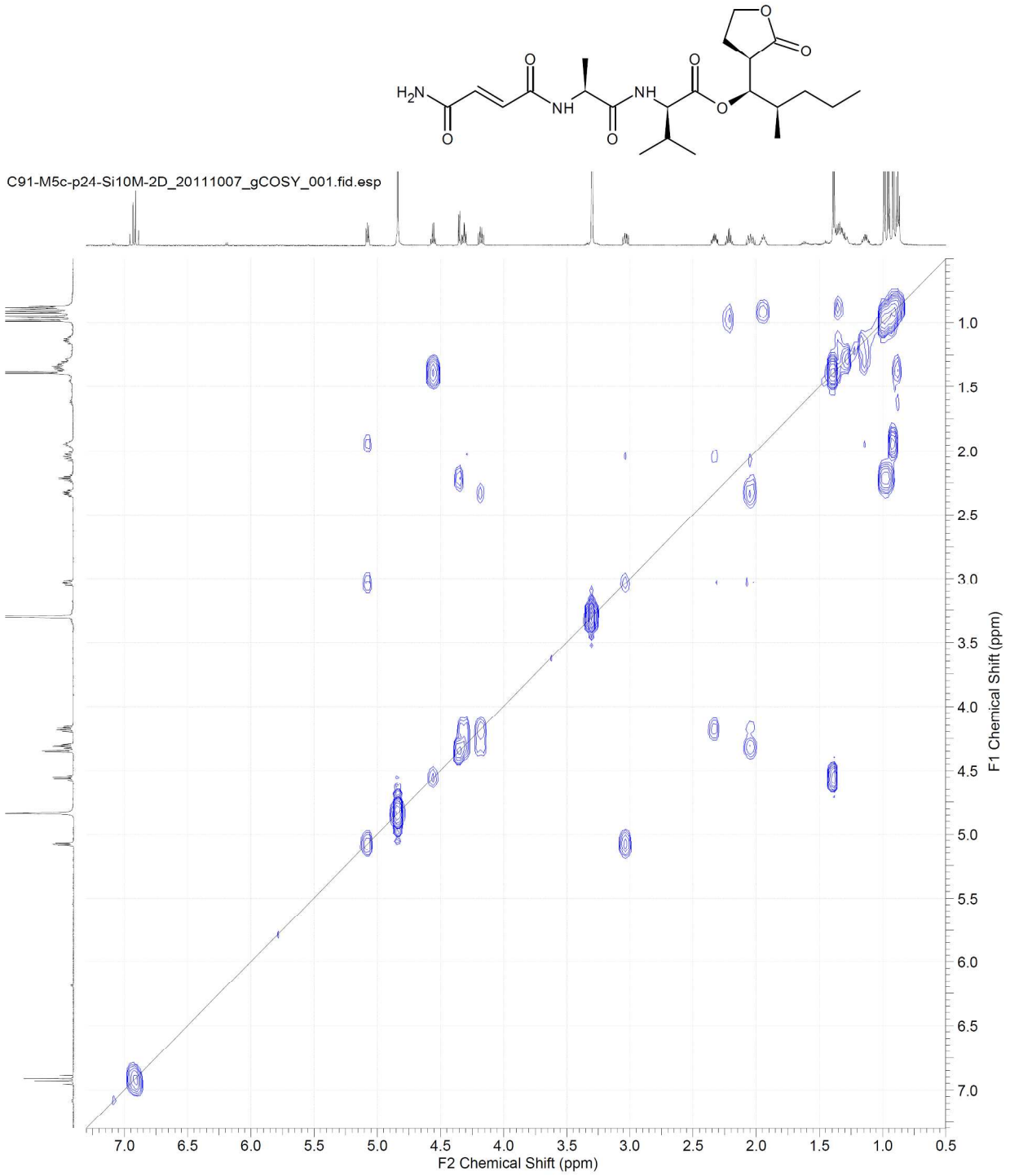
^{13}C NMR Spectrum of **2** at 600 MHz in CD_3OD



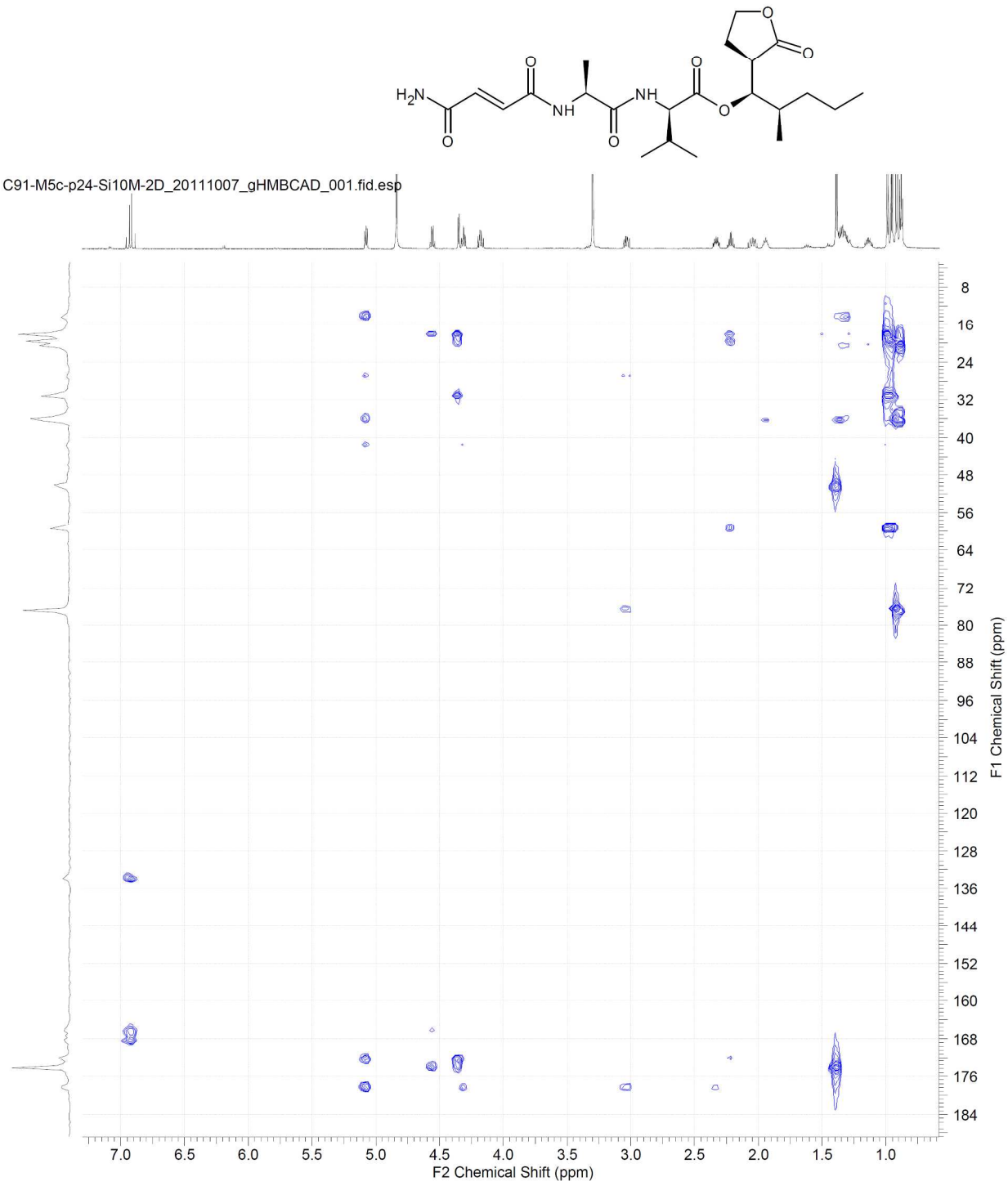
HSQC Spectrum of **2** at 600 MHz in CD_3OD



COSY Spectrum of **2** at 600 MHz in CD_3OD



HMBC Spectrum of **2** at 600 MHz in CD_3OD



NOESY Spectrum of **2** at 600 MHz in CD_3OD