

Supporting Information

Dereplication of Natural Products using Minimal NMR Data Inputs

Russell B. Williams,^a Mark O'Neil-Johnson,^a Antony J. Williams,^b Patrick Wheeler,^c Sergey Golotvin,^c Rostislava Polc^c and Arvin Moser^{c*}

^a*Sequoia Sciences, Inc.*, 1912 Innerbelt Business Center Drive, St. Louis, MO 63114, USA

^bChemConnector Inc., 904 Tamaras Circle, Wake Forest, NC, 27587, USA

^cAdvanced Chemistry Development, Toronto Department, 8King Street E, 107, Toronto, Ontario, M5C 1B5, Canada

*Email: Arvin Moser (arvin@acdlabs.com)

Table of Contents

- Figure S1.** The list of 16 known compounds used to analyze the performance of the dereplication process. The molecular formulae, monoisotopic masses and ChemSpider ID number are included. Structures 2 and 11 are from fungi. All of the others are plant natural products (NPs).
- Figure S2.** Proton NMR spectra (full and inset) of compound **1** in CD₃OD at 600 MHz.
- Figure S3.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **1** in CD₃OD at 600 MHz.
- Figure S4.** Expanded area of a ¹H-¹H COSY spectrum of compound **1** in CD₃OD at 600 MHz.
- Figure S5.** Proton NMR spectra (full and inset) of compound **2** in CD₃OD at 600 MHz.
- Figure S6.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **2** in CD₃OD at 600 MHz.
- Figure S7.** Expanded area of a ¹H-¹H COSY spectrum of compound **2** in CD₃OD at 600 MHz.
- Figure S8.** Proton NMR spectra (full and inset) of compound **3** in CD₃OD at 600 MHz.
- Figure S9.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **3** in CD₃OD at 600 MHz.
- Figure S10.** Expanded area of a ¹H-¹H COSY spectrum of compound **3** in CD₃OD at 600 MHz.
- Figure S11.** Proton NMR spectra (full and inset) of compound **4** in CD₃OD at 600 MHz.
- Figure S12.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **4** in CD₃OD at 600 MHz.
- Figure S13.** Expanded area of a ¹H-¹H COSY spectrum of compound **4** in CD₃OD at 600 MHz.
- Figure S14.** Proton NMR spectra (full and inset) of compound **5** in CD₃OD at 600 MHz.
- Figure S15.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **5** in CD₃OD at 600 MHz.
- Figure S16.** Expanded area of a ¹H-¹H COSY spectrum of compound **5** in CD₃OD at 600 MHz.
- Figure S17.** Proton NMR spectra (full and inset) of compound **6** in CD₃OD at 600 MHz.
- Figure S18.** Expanded area of a ¹H-¹H COSY spectrum of compound **6** in CD₃OD at 600 MHz.
- Figure S19.** Proton NMR spectra (full and inset) of compound **7** in CD₃OD at 600 MHz.
- Figure S20.** Expanded area of a ¹H-¹³C HSQC spectrum of compound **7** in CD₃OD at 600 MHz.
- Figure S21.** Expanded area of a ¹H-¹H COSY spectrum of compound **7** in CD₃OD at 600 MHz.
- Figure S22.** Proton NMR spectra (full and inset) of compound **8** in CD₃OD at 600 MHz.

- Figure S23.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **8** in CD_3OD at 600 MHz.
- Figure S24.** Expanded area of a ^1H - ^1H COSY spectrum of compound **8** in CD_3OD at 600 MHz.
- Figure S25.** Proton NMR spectra (full and inset) of compound **9** in CD_3OD at 600 MHz.
- Figure S26.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **9** in CD_3OD at 600 MHz.
- Figure S27.** Expanded area of a ^1H - ^1H COSY spectrum of compound **9** in CD_3OD at 600 MHz.
- Figure S28.** Proton NMR spectra (full and inset) of compound **10** in CD_3OD at 600 MHz.
- Figure S29.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **10** in CD_3OD at 600 MHz.
- Figure S30.** Expanded area of a ^1H - ^1H COSY spectrum of compound **10** in CD_3OD at 600 MHz.
- Figure S31.** Proton NMR spectra (full and inset) of compound **11** in CD_3OD at 600 MHz.
- Figure S32.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **11** in CD_3OD at 600 MHz.
- Figure S33.** Expanded area of a ^1H - ^1H COSY spectrum of compound **11** in CD_3OD at 600 MHz.
- Figure S34.** Proton NMR spectra (full and inset) of compound **12** in CD_3OD at 600 MHz.
- Figure S35.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **12** in CD_3OD at 600 MHz.
- Figure S36.** Expanded area of a ^1H - ^1H COSY spectrum of compound **12** in CD_3OD at 600 MHz.
- Figure S37.** Proton NMR spectra (full and inset) of compound **13** in CD_3OD at 600 MHz.
- Figure S38.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **13** in CD_3OD at 600 MHz.
- Figure S39.** Expanded area of a ^1H - ^1H COSY spectrum of compound **13** in CD_3OD at 600 MHz.
- Figure S40.** Proton NMR spectra (full and inset) of compound **14** in CD_3OD at 600 MHz.
- Figure S41.** Expanded area of a ^1H - ^{13}C HMQC spectrum of compound **14** in CD_3OD at 600 MHz.
- Figure S42.** Expanded area of a ^1H - ^1H COSY spectrum of compound **14** in CD_3OD at 600 MHz.
- Figure S43.** Proton NMR spectra (full and inset) of compound **15** in CD_3OD at 600 MHz.
- Figure S44.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **15** in CD_3OD at 600 MHz.
- Figure S45.** Expanded area of a ^1H - ^1H COSY spectrum of compound **15** in CD_3OD at 600 MHz.
- Figure S46.** Proton NMR spectra (full and inset) of compound **16** in CD_3OD at 600 MHz.

- Figure S47.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **16** in CD_3OD at 600 MHz.
- Figure S48.** Expanded area of a ^1H - ^1H COSY spectrum of compound **16** in CD_3OD at 600 MHz.
- Figure S49.** Proton NMR spectra (full and inset) of compound **17** in CD_3OD at 600 MHz.
- Figure S50.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **17** in CD_3OD at 600 MHz.
- Figure S51.** Expanded area of a ^1H - ^1H COSY spectrum of compound **17** in CD_3OD at 600 MHz.
- Figure S52.** Proton NMR spectra (full and inset) of compound **18** in $\text{DMSO-}d_6$ at 600 MHz.
- Figure S53.** Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **18** in $\text{DMSO-}d_6$ at 600 MHz.
- Figure S54.** Expanded area of a ^1H - ^1H COSY spectrum of compound **18** in $\text{DMSO-}d_6$ at 600 MHz.
- Figure S55.** Table of ^1H and ^1H - ^{13}C HSQC NMR spectra and m/z and MF search results for NPs 1 to 16.

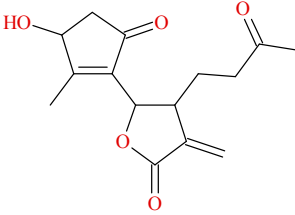
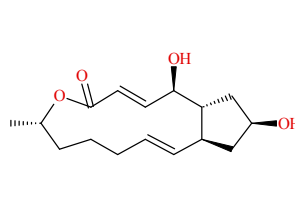
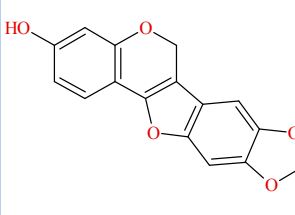
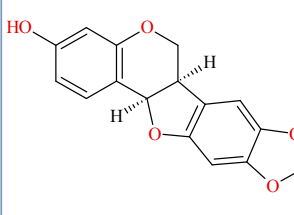
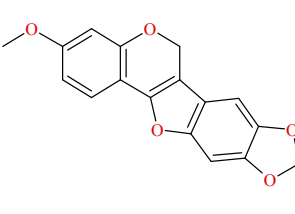
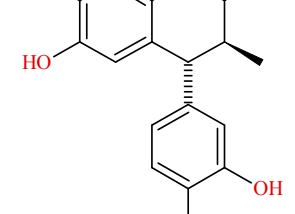
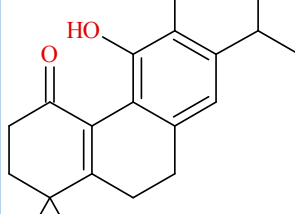
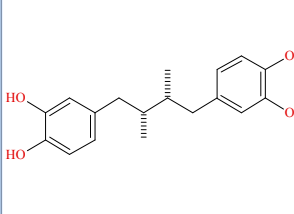
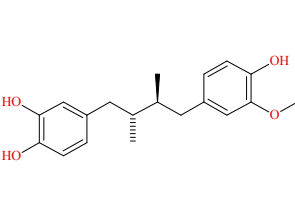
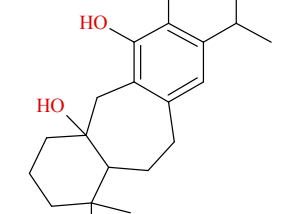
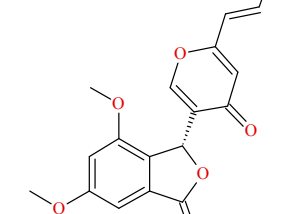
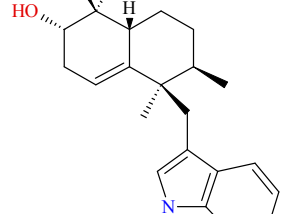
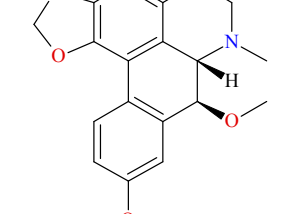
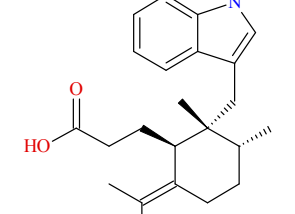
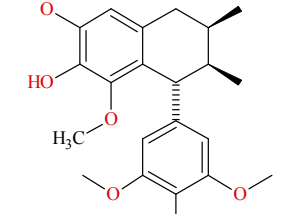
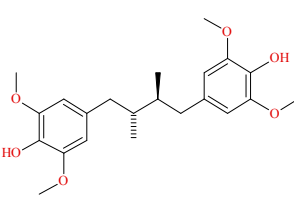
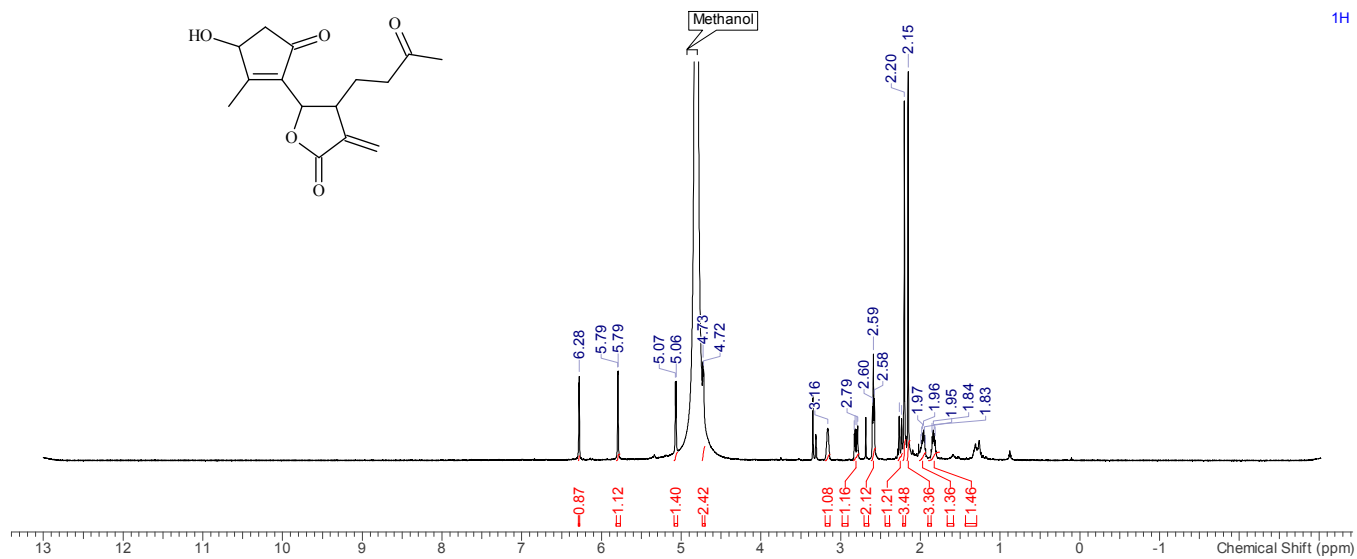
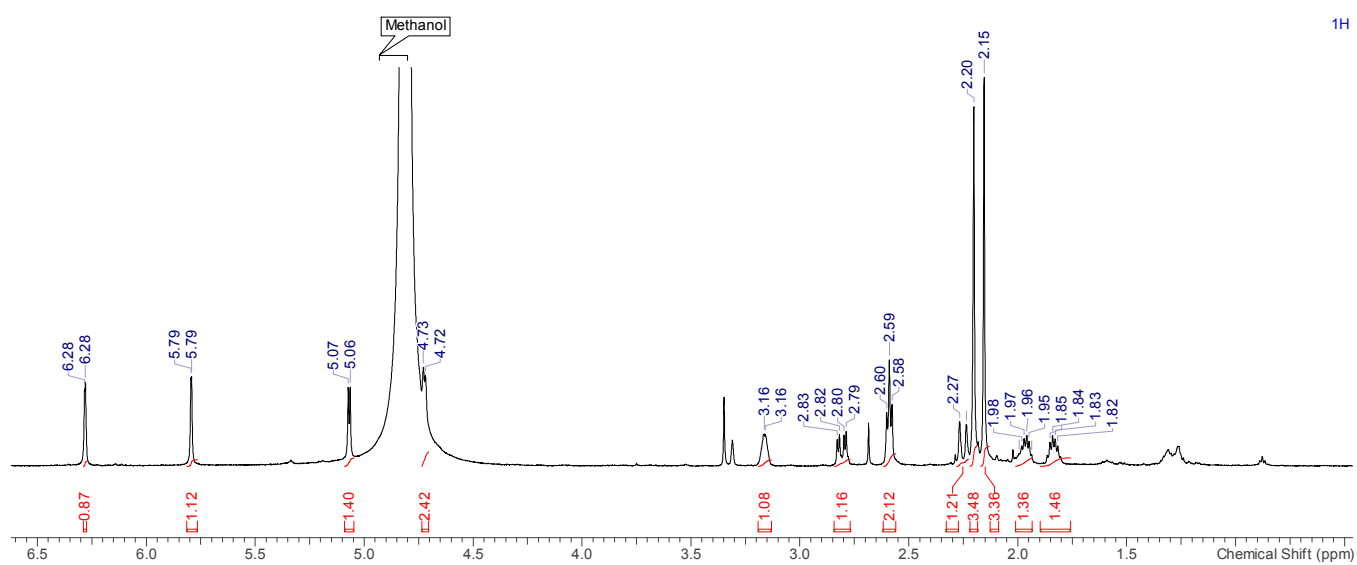
<p>1</p> 	<p>2</p> 	<p>3</p> 	<p>4</p> 
<p>Formula: C₁₅H₁₈O₅ Monoisotopic Mass: 278.115424 ChemSpider ID#: 23339535</p>	<p>Formula: C₁₆H₂₄O₄ Monoisotopic Mass: 280.167459 ChemSpider ID#: 4449949</p>	<p>Formula: C₁₆H₁₀O₅ Monoisotopic Mass: 282.052823 ChemSpider ID#: 24843070</p>	<p>Formula: C₁₆H₁₂O₅ Monoisotopic Mass: 284.068473 ChemSpider ID#: 82631</p>
<p>5</p> 	<p>6</p> 	<p>7</p> 	<p>8</p> 
<p>Formula: C₁₇H₁₂O₅ Monoisotopic Mass: 296.068473 ChemSpider ID#: 9997217</p>	<p>Formula: C₁₈H₂₀O₄ Monoisotopic Mass: 300.136159 ChemSpider ID#: 24702462</p>	<p>Formula: C₁₉H₂₄O₃ Monoisotopic Mass: 300.172545 ChemSpider ID#: 4982318</p>	<p>Formula: C₁₈H₂₂O₄ Monoisotopic Mass: 302.151809 ChemSpider ID#: 599530</p>
<p>9</p> 	<p>10</p> 	<p>11</p> 	<p>12</p> 
<p>Formula: C₁₉H₂₄O₄ Monoisotopic Mass: 318.167459 ChemSpider ID#: 9755278</p>	<p>Formula: C₂₀H₃₀O₃ Monoisotopic Mass: 318.219495 ChemSpider ID#: 9983971</p>	<p>Formula: C₁₈H₁₆O₆ Monoisotopic Mass: 328.094688 ChemSpider ID#: 45785000</p>	<p>Formula: C₂₃H₃₁NO Monoisotopic Mass: 337.240565 ChemSpider ID#: 10234073</p>
<p>13</p> 	<p>14</p> 	<p>15</p> 	<p>16</p> 
<p>Formula: C₂₀H₂₁NO₄ Monoisotopic Mass: 339.147058 ChemSpider ID#: 8241337</p>	<p>Formula: C₂₃H₃₁NO₂ Monoisotopic Mass: 353.235479 ChemSpider ID#: 9374078</p>	<p>Formula: C₂₂H₂₈O₆ Monoisotopic Mass: 388.188589 ChemSpider ID#: 10417064</p>	<p>Formula: C₂₂H₃₀O₆ Monoisotopic Mass: 390.204239 ChemSpider ID#: 4478720</p>

Figure S1. The list of 16 known compounds used to analyze the performance of the dereplication process. The molecular formulae, monoisotopic masses and ChemSpider ID number are included. Structures 2 and 11 are from fungi. All of the others are plant natural products (NPs).



1H



1H

Figure S2. Proton NMR spectra (full and inset) of compound 1 in CD₃OD at 600 MHz.

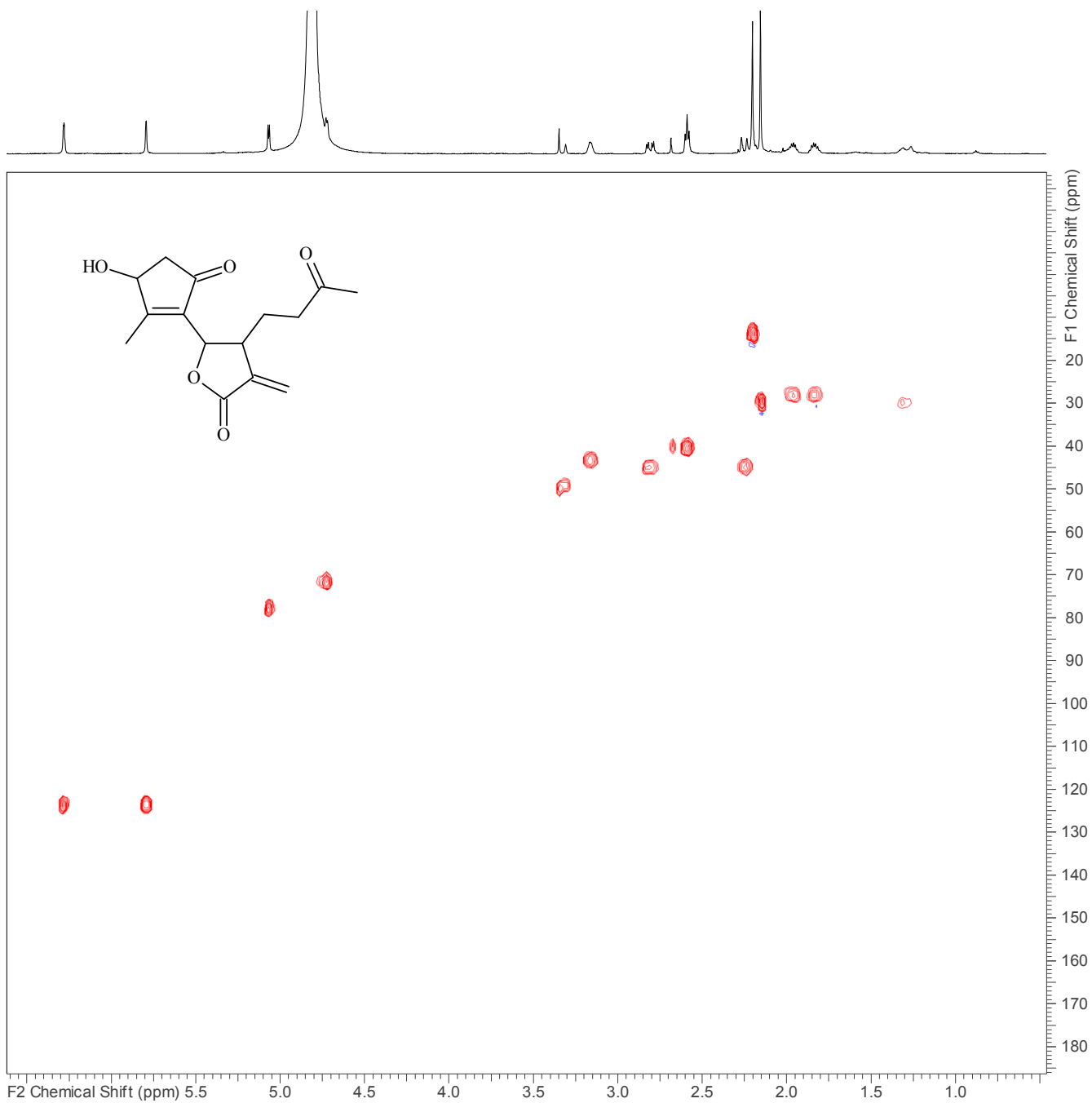


Figure S3. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **1** in CD_3OD at 600 MHz.

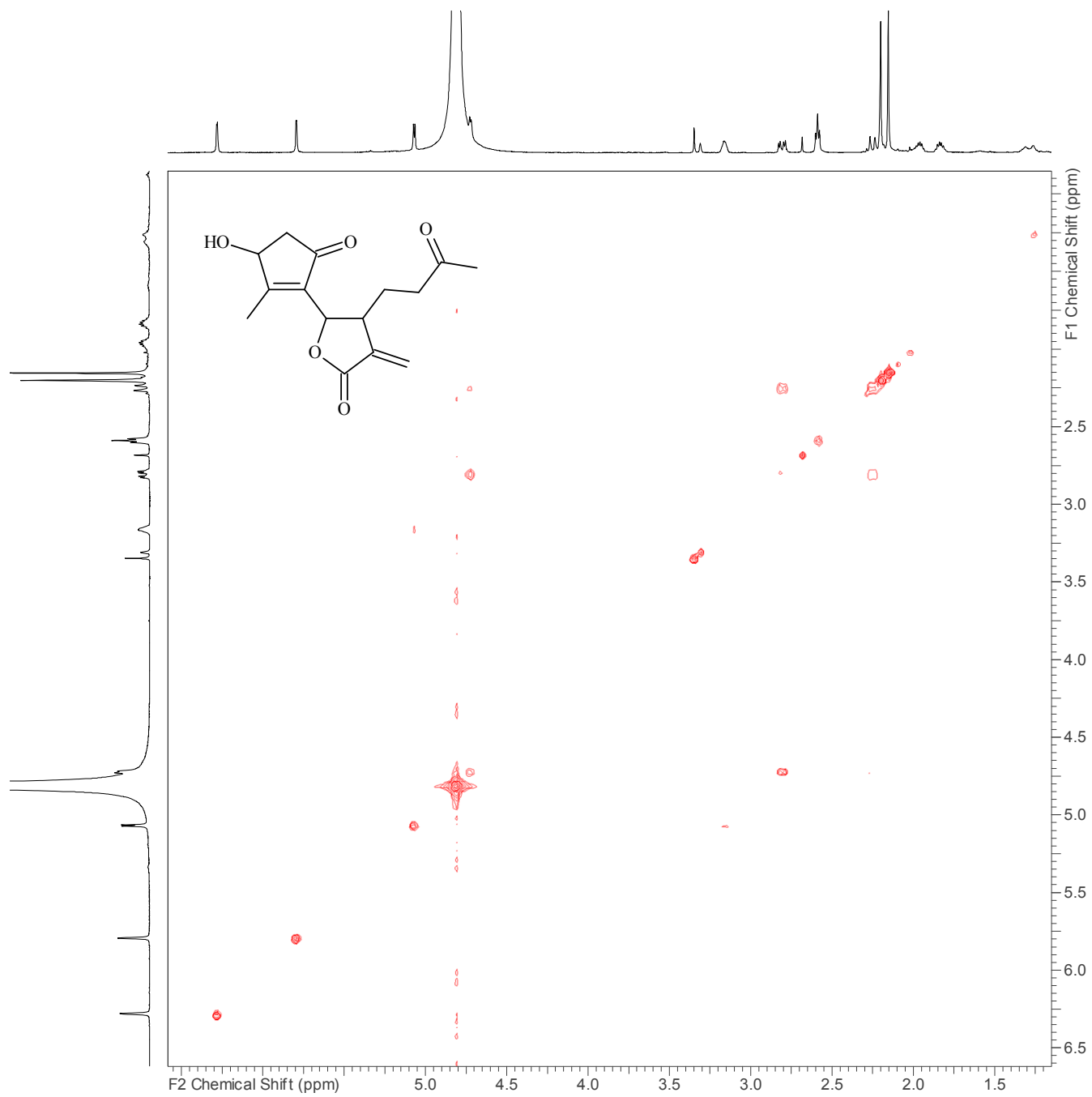


Figure S4. Expanded area of a ^1H - ^1H COSY spectrum of compound **1** in CD_3OD at 600 MHz.

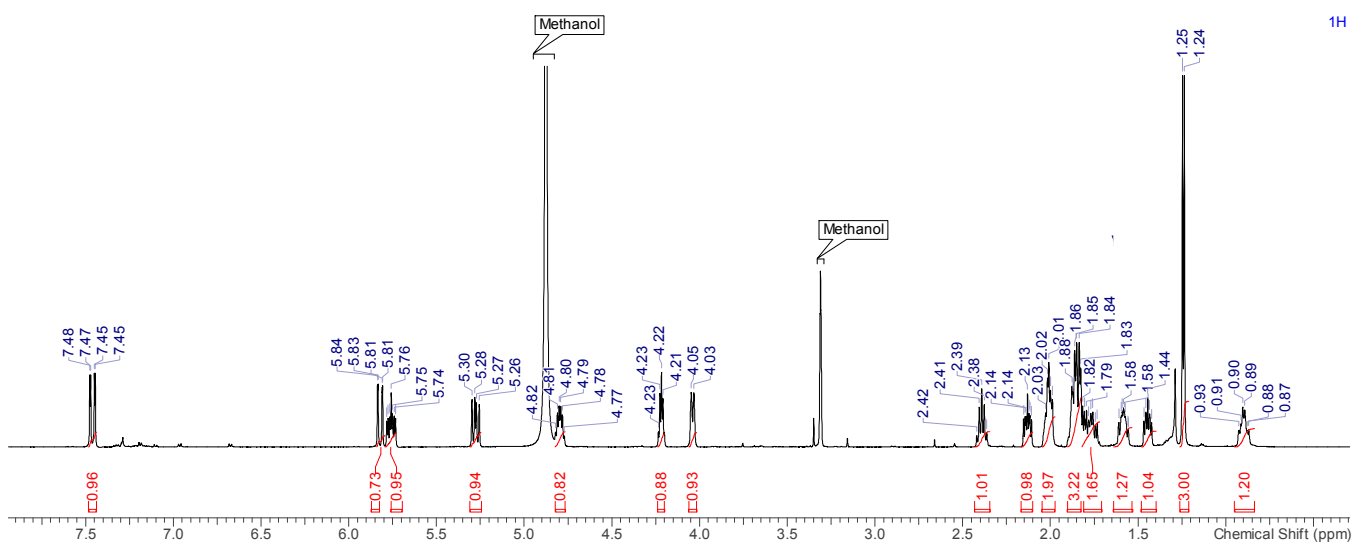
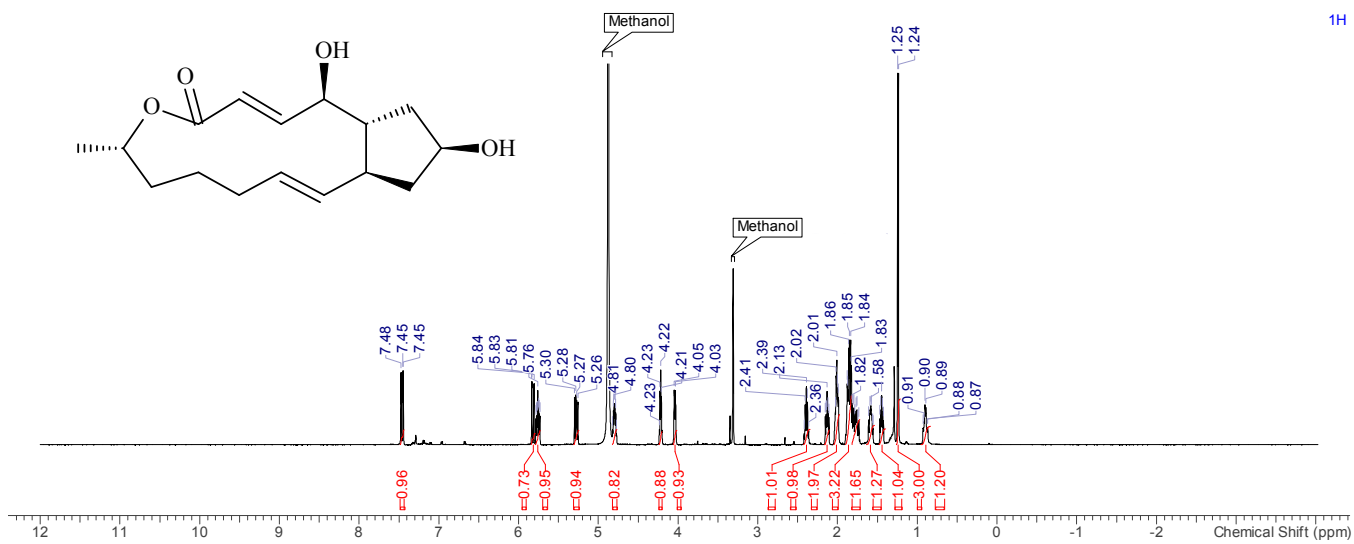


Figure S5. Proton NMR spectra (full and inset) of compound 2 in CD₃OD at 600 MHz.

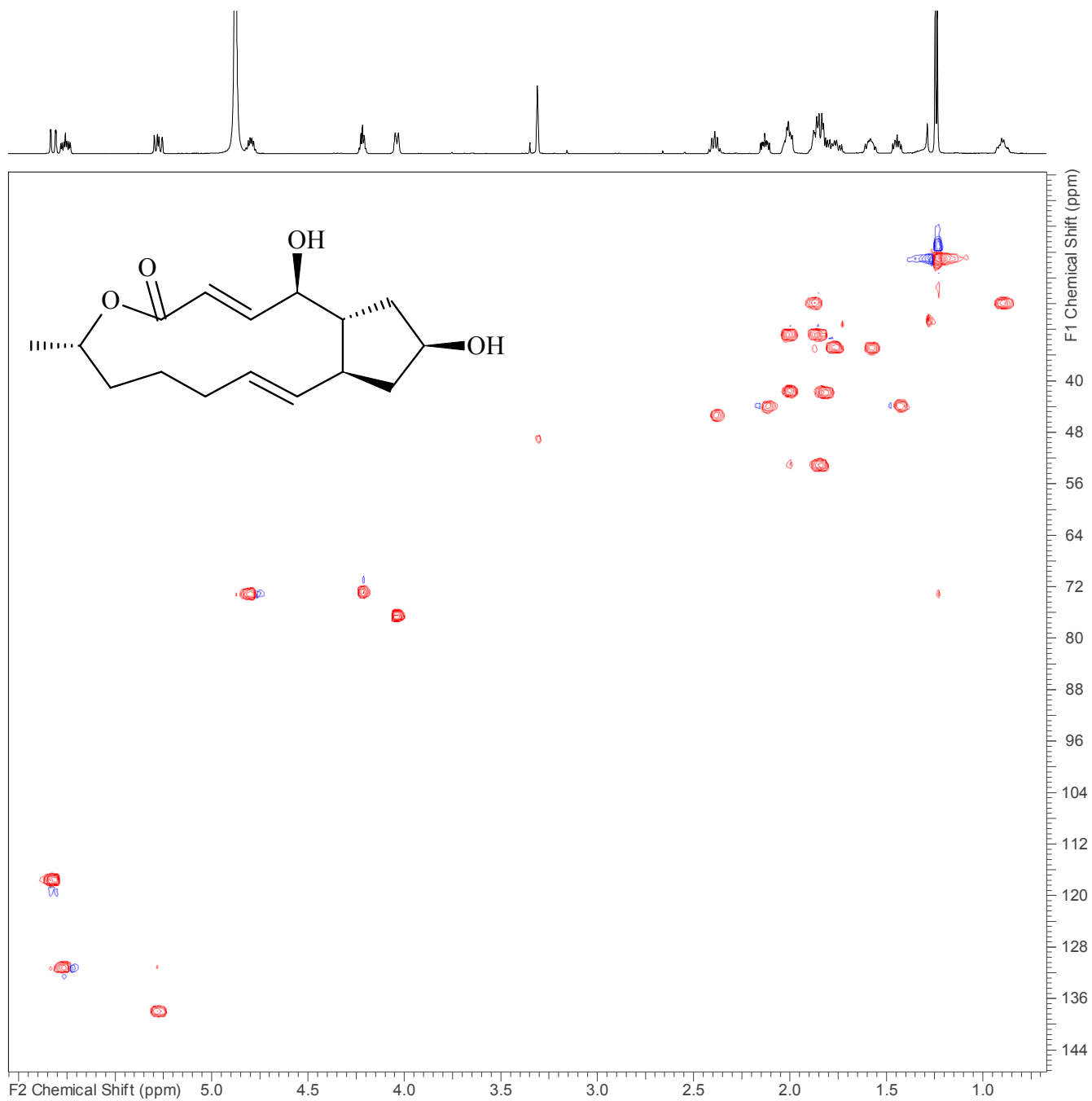


Figure S6. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **2** in CD_3OD at 600 MHz.

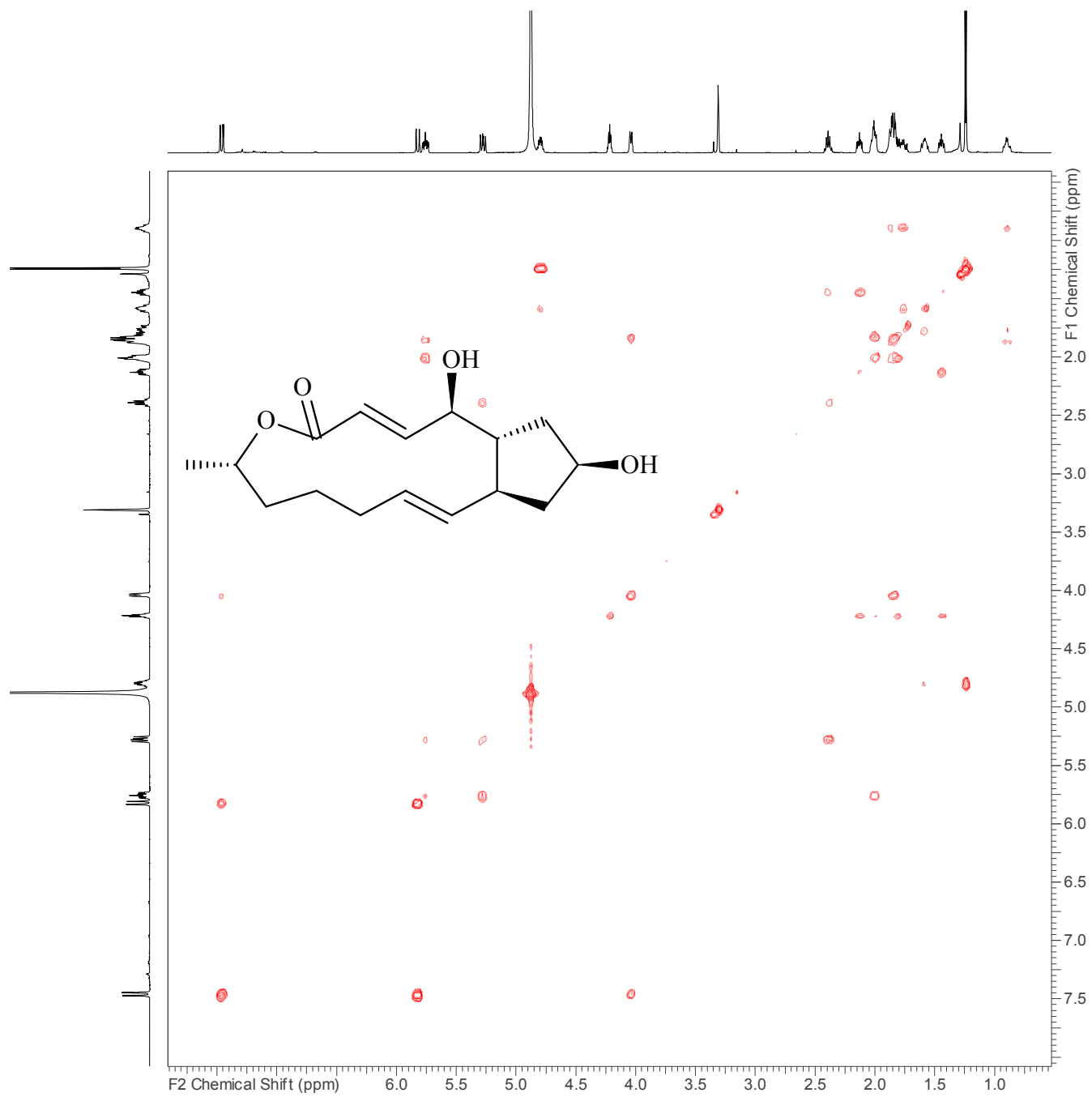
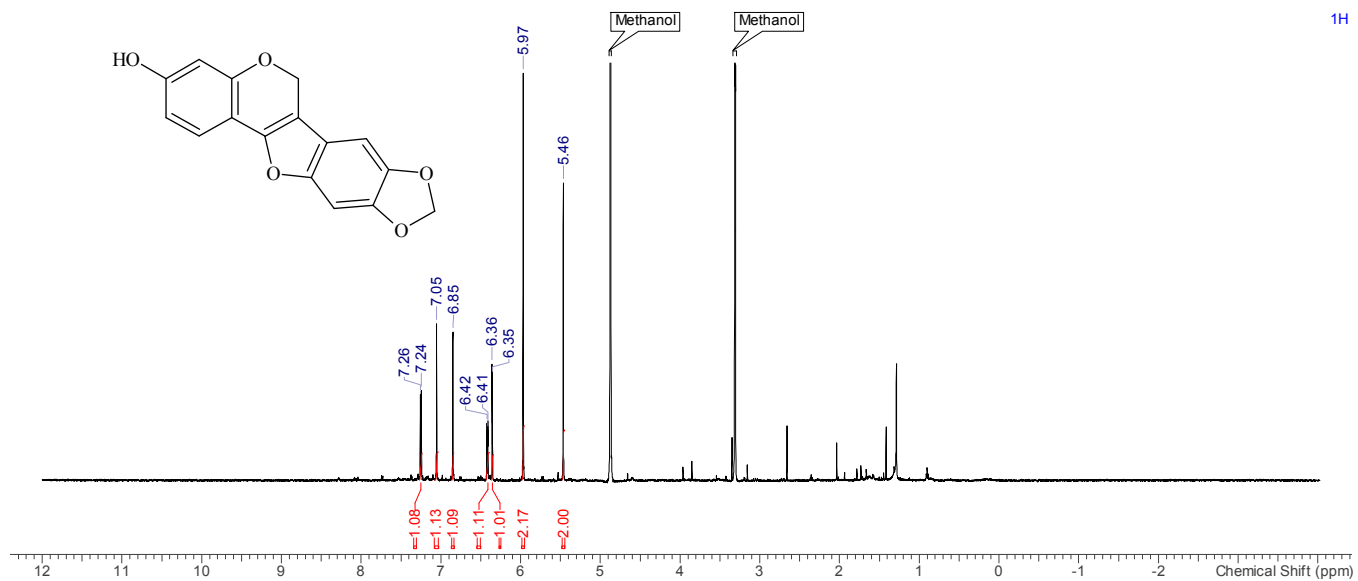
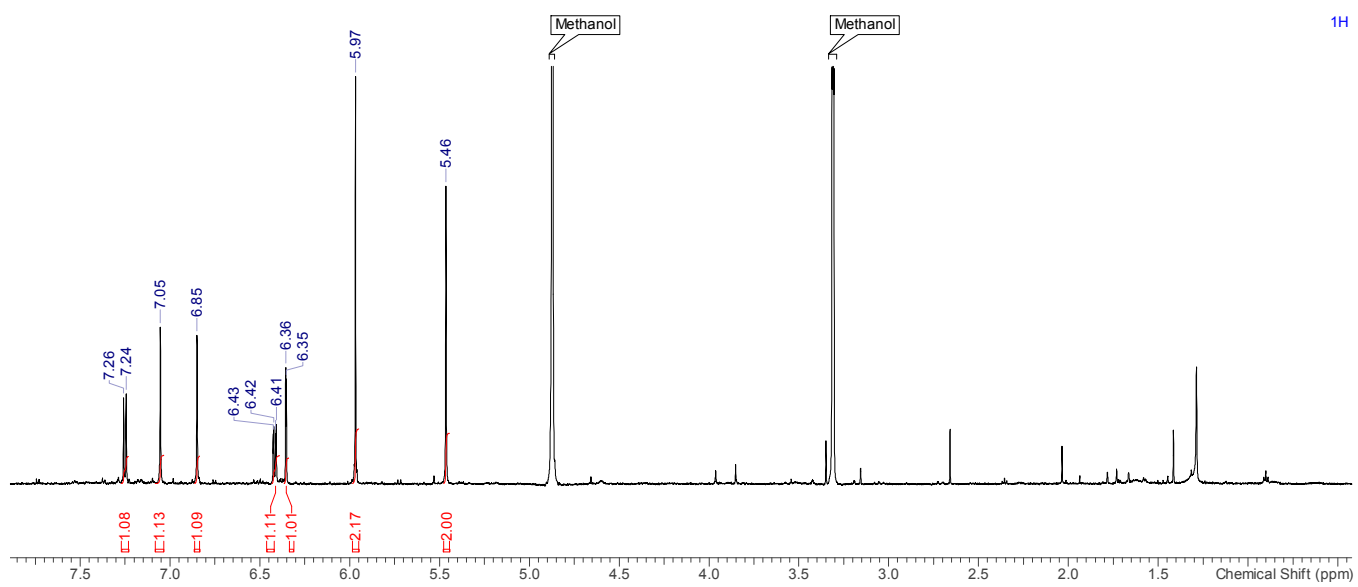


Figure S7. Expanded area of a ^1H - ^1H COSY spectrum of compound **2** in CD_3OD at 600 MHz.



¹H



¹H

Figure S8. Proton NMR spectra (full and inset) of compound **3** in CD₃OD at 600 MHz.

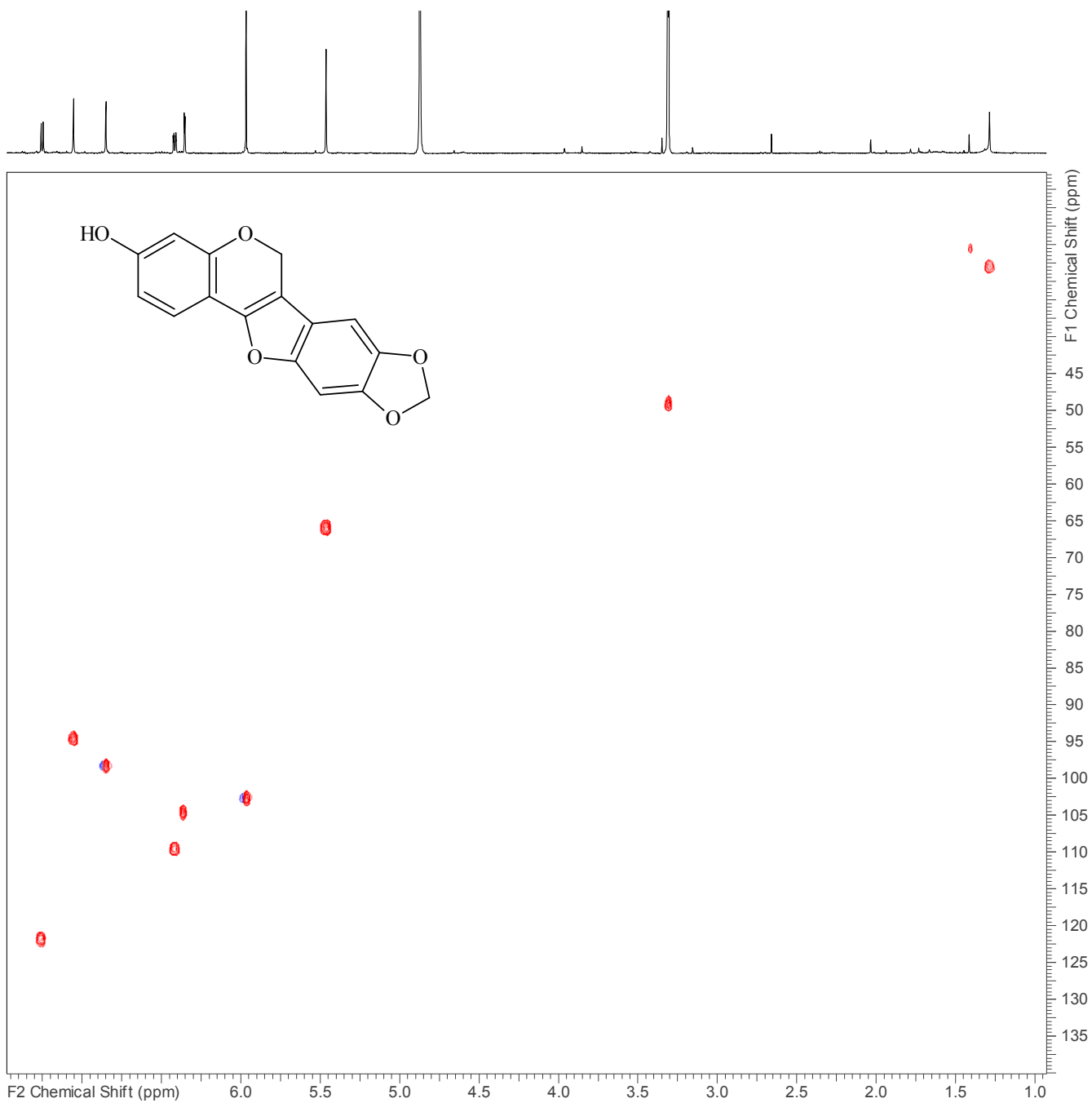


Figure S9. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **3** in CD_3OD at 600 MHz.

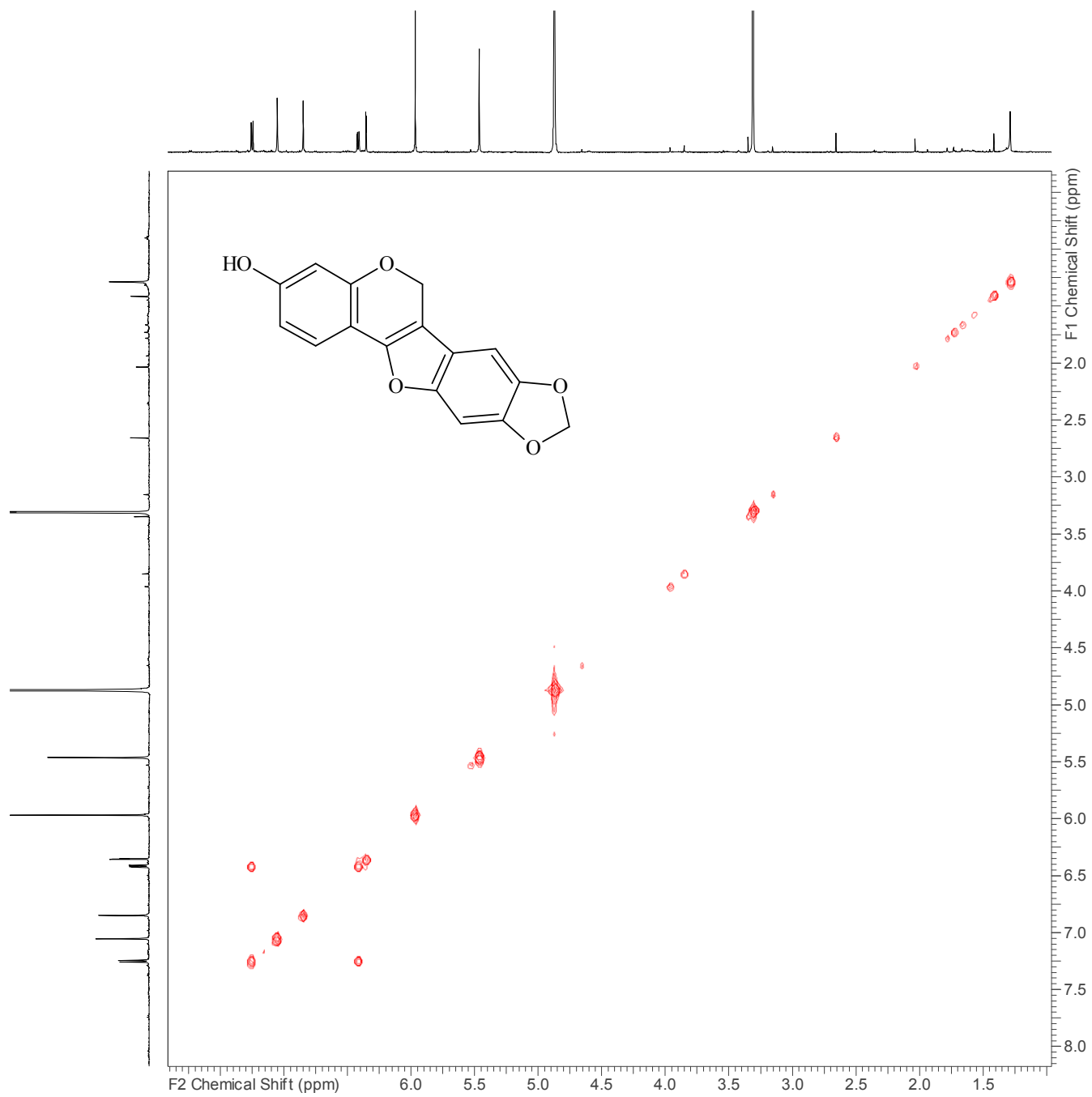


Figure S10. Expanded area of a ^1H - ^1H COSY spectrum of compound **3** in CD_3OD at 600 MHz.

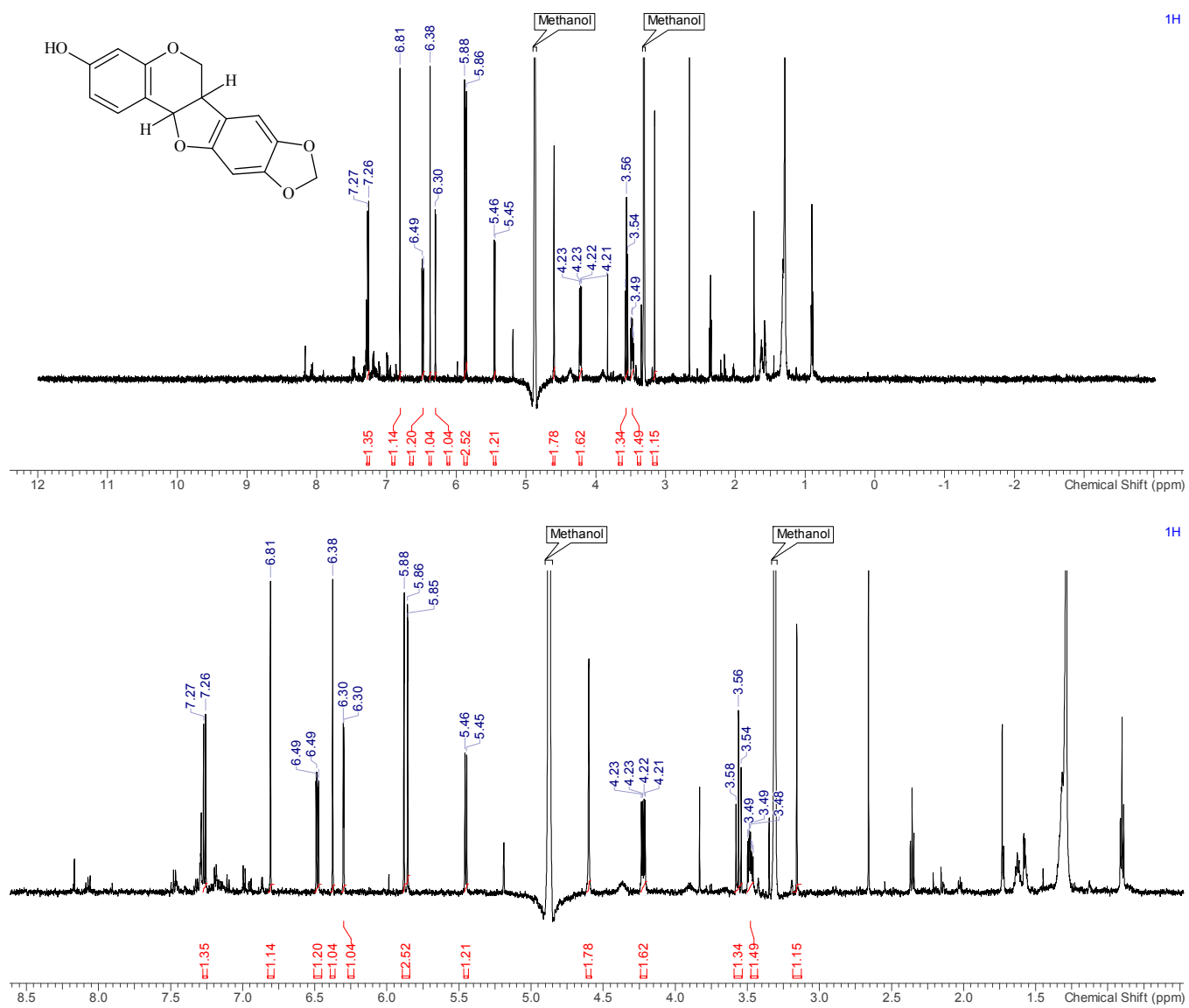


Figure S11. Proton NMR spectra (full and inset) of compound 4 in CD₃OD at 600 MHz.

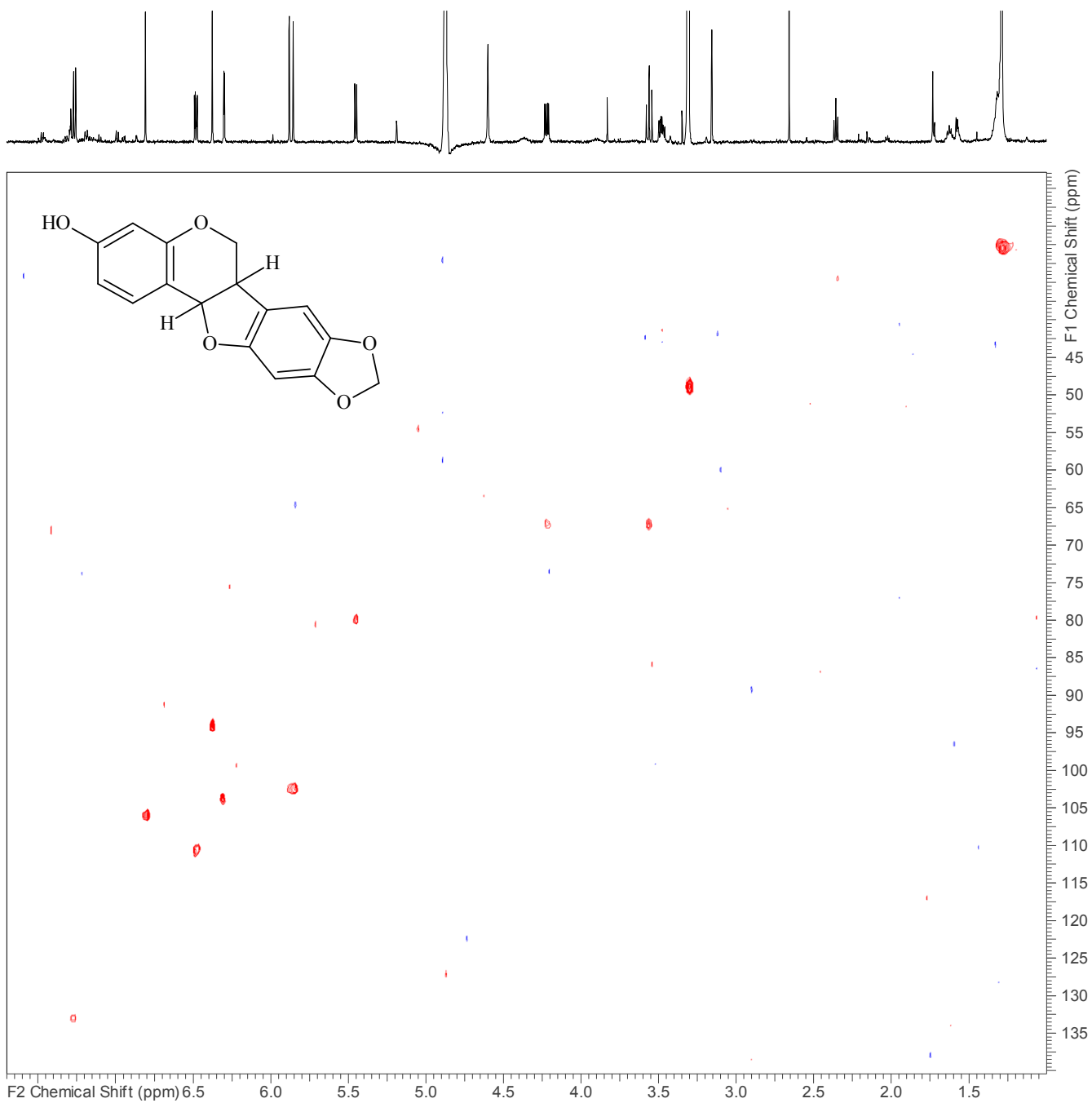


Figure S12. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound 4 in CD_3OD at 600 MHz.

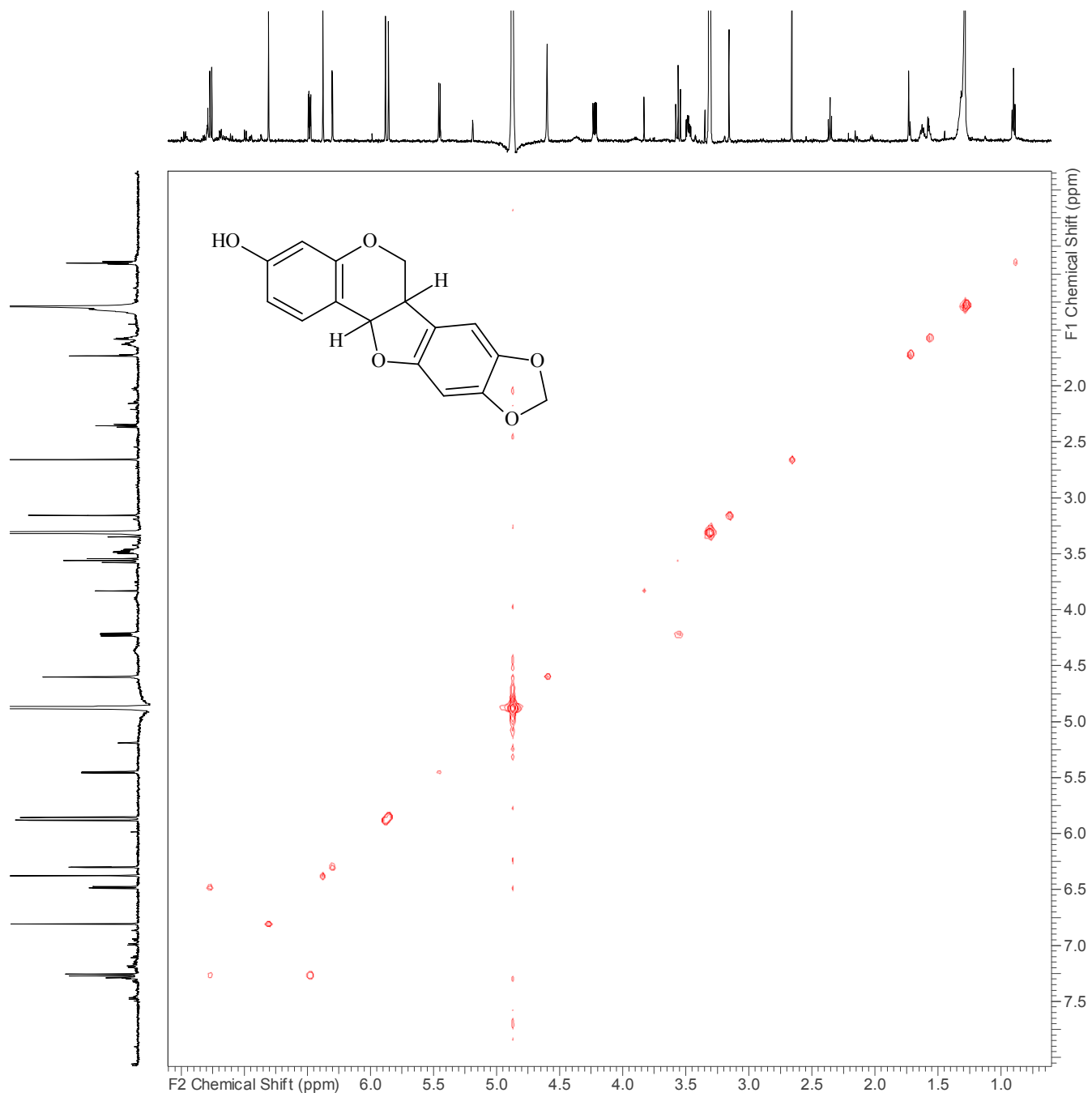
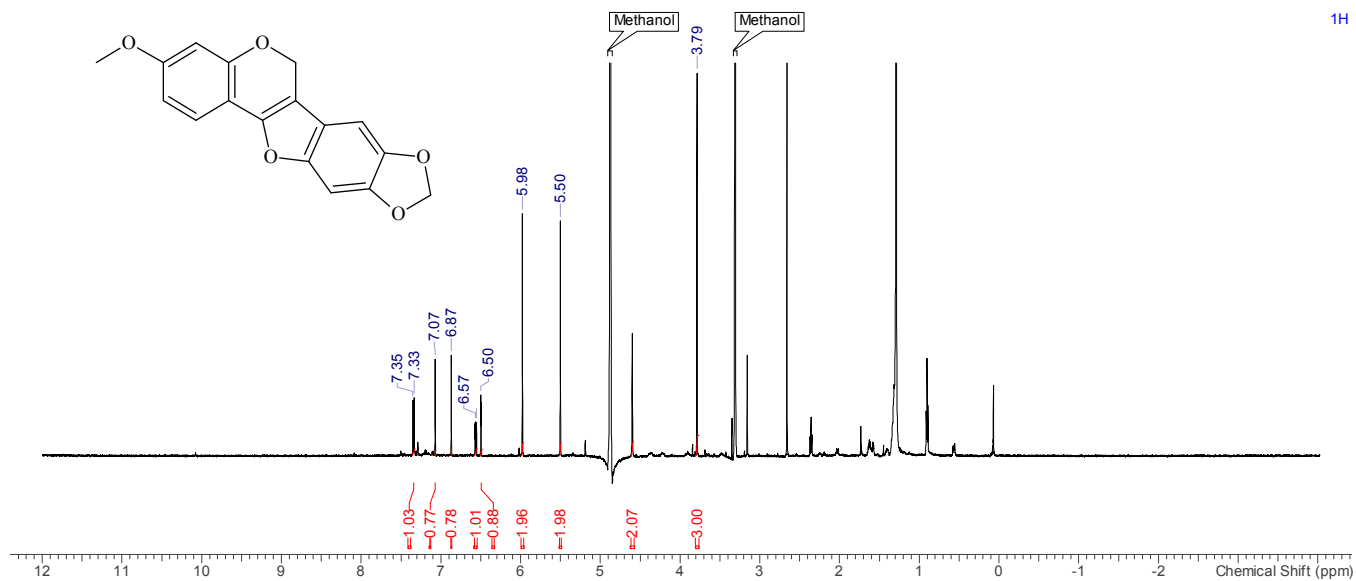


Figure S13. Expanded area of a ^1H - ^1H COSY spectrum of compound **4** in CD_3OD at 600 MHz.

1H



1H

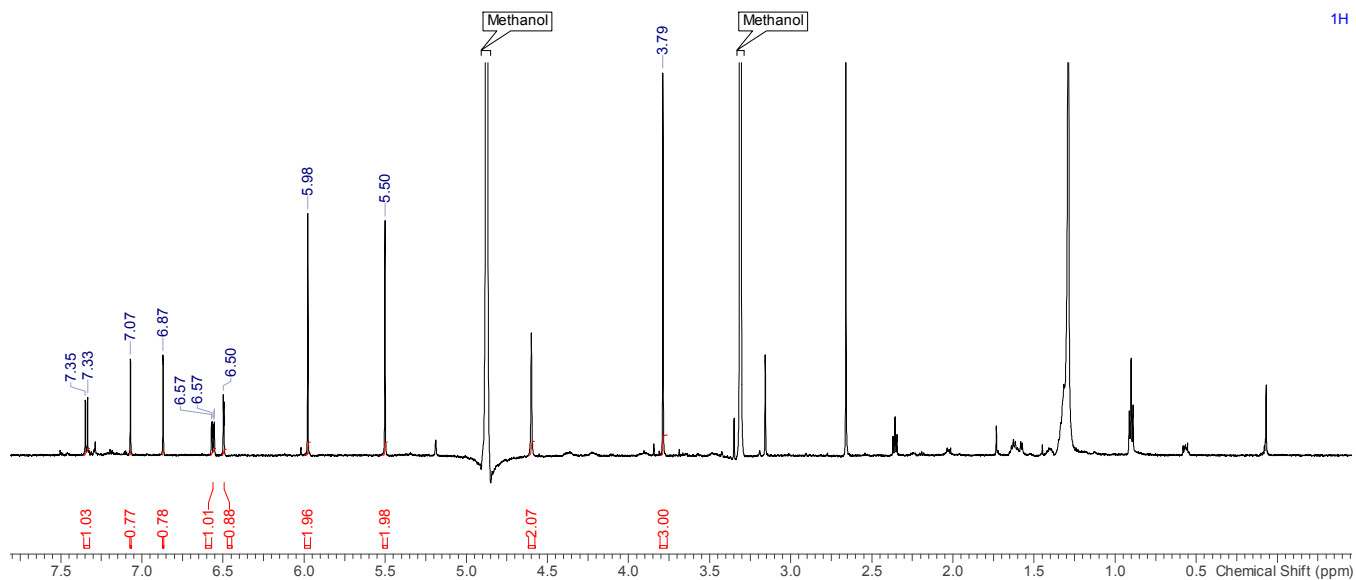


Figure S14. Proton NMR spectra (full and inset) of compound **5** in CD₃OD at 600 MHz.

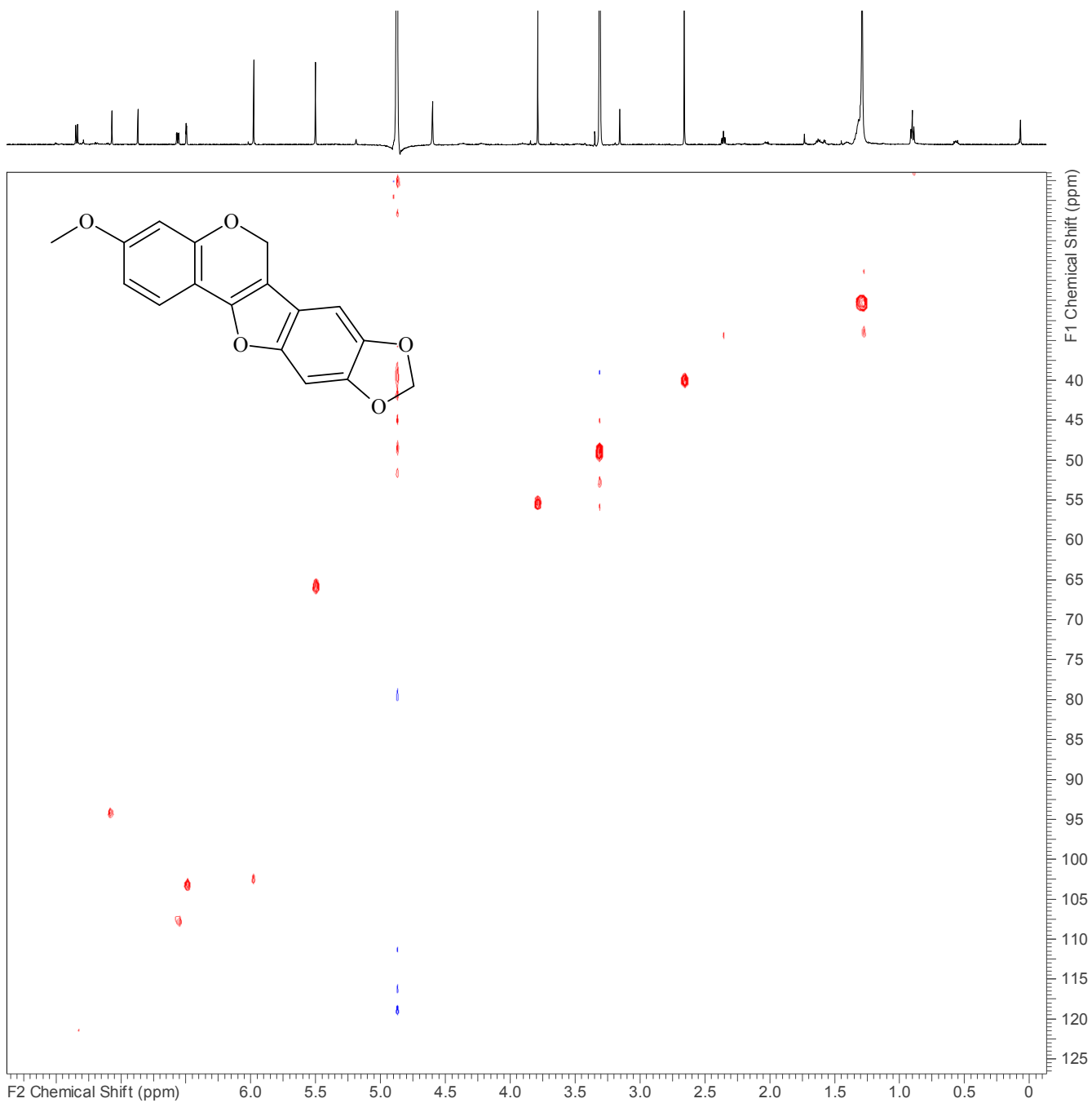


Figure S15. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **5** in CD_3OD at 600 MHz.

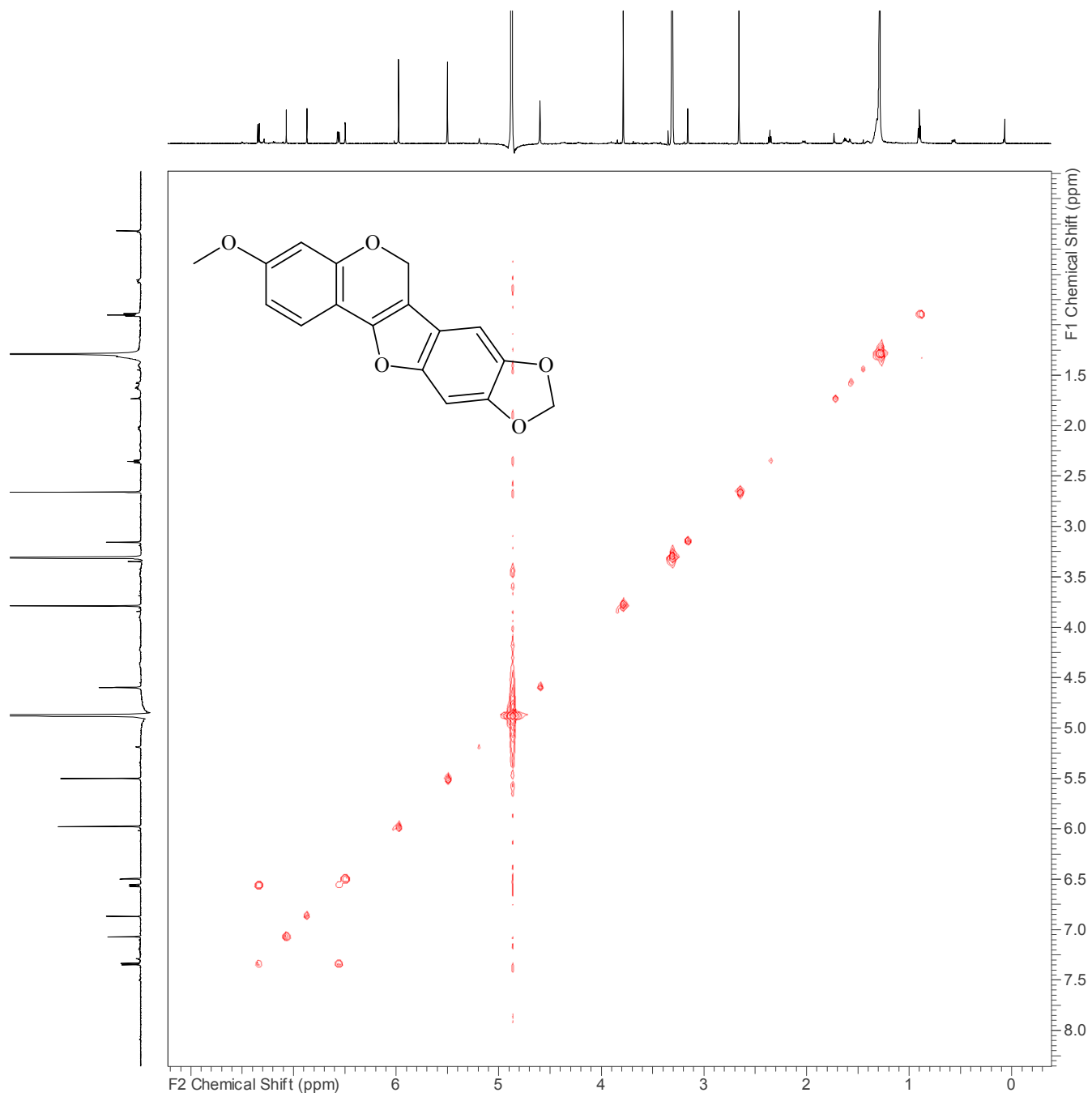


Figure S16. Expanded area of a ^1H - ^1H COSY spectrum of compound **5** in CD_3OD at 600 MHz.

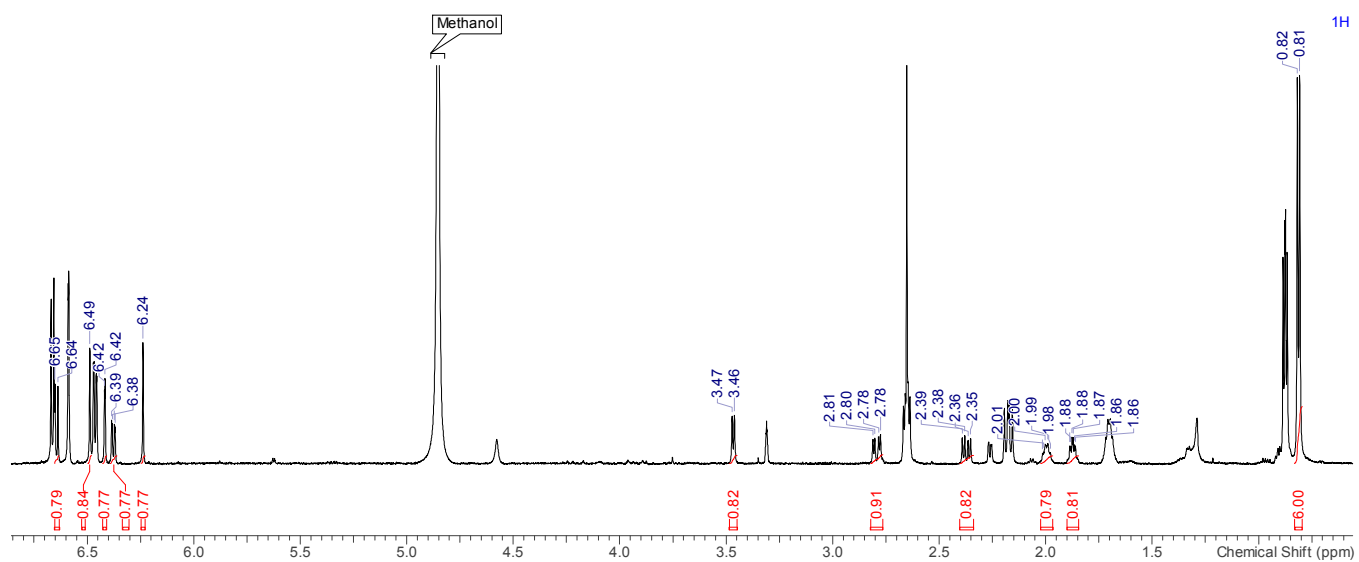
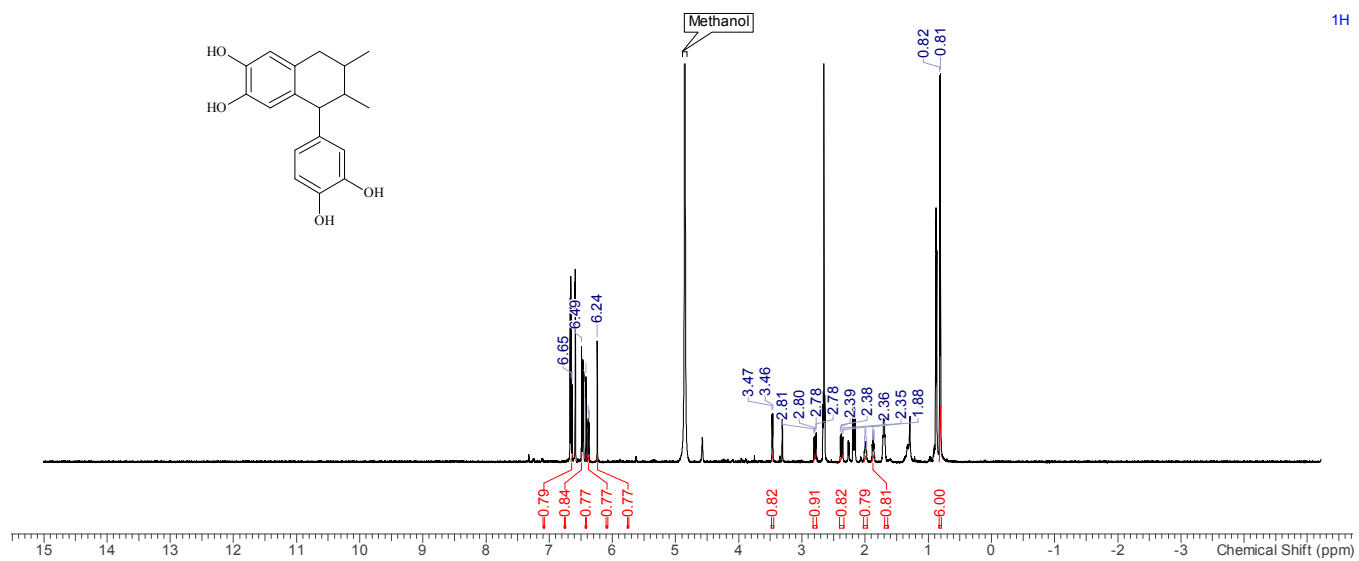


Figure S17. Proton NMR spectra (full and inset) of compound 6 in CD₃OD at 600 MHz.

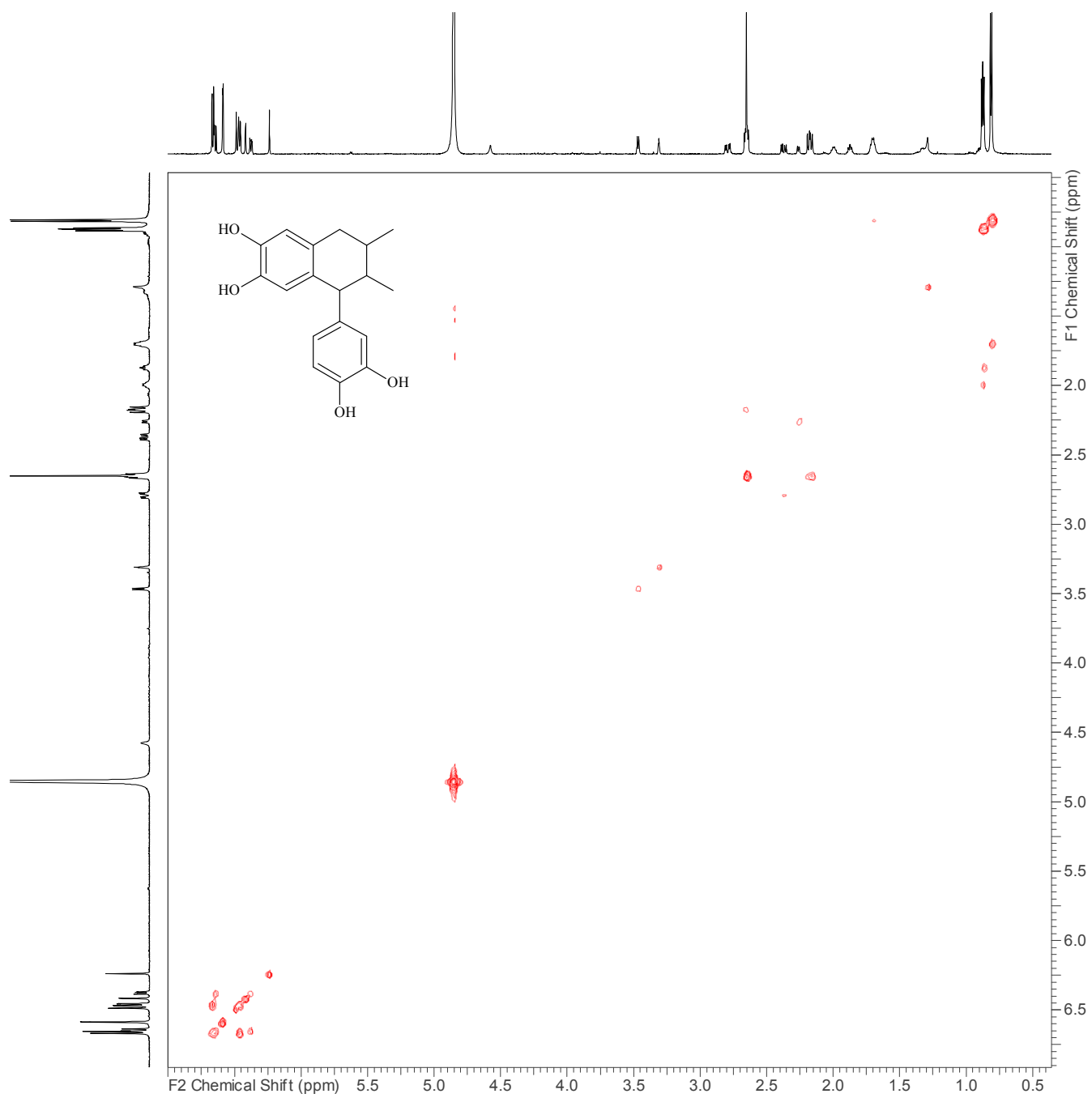


Figure S18. Expanded area of a ^1H - ^1H COSY spectrum of compound **6** in CD_3OD at 600 MHz.

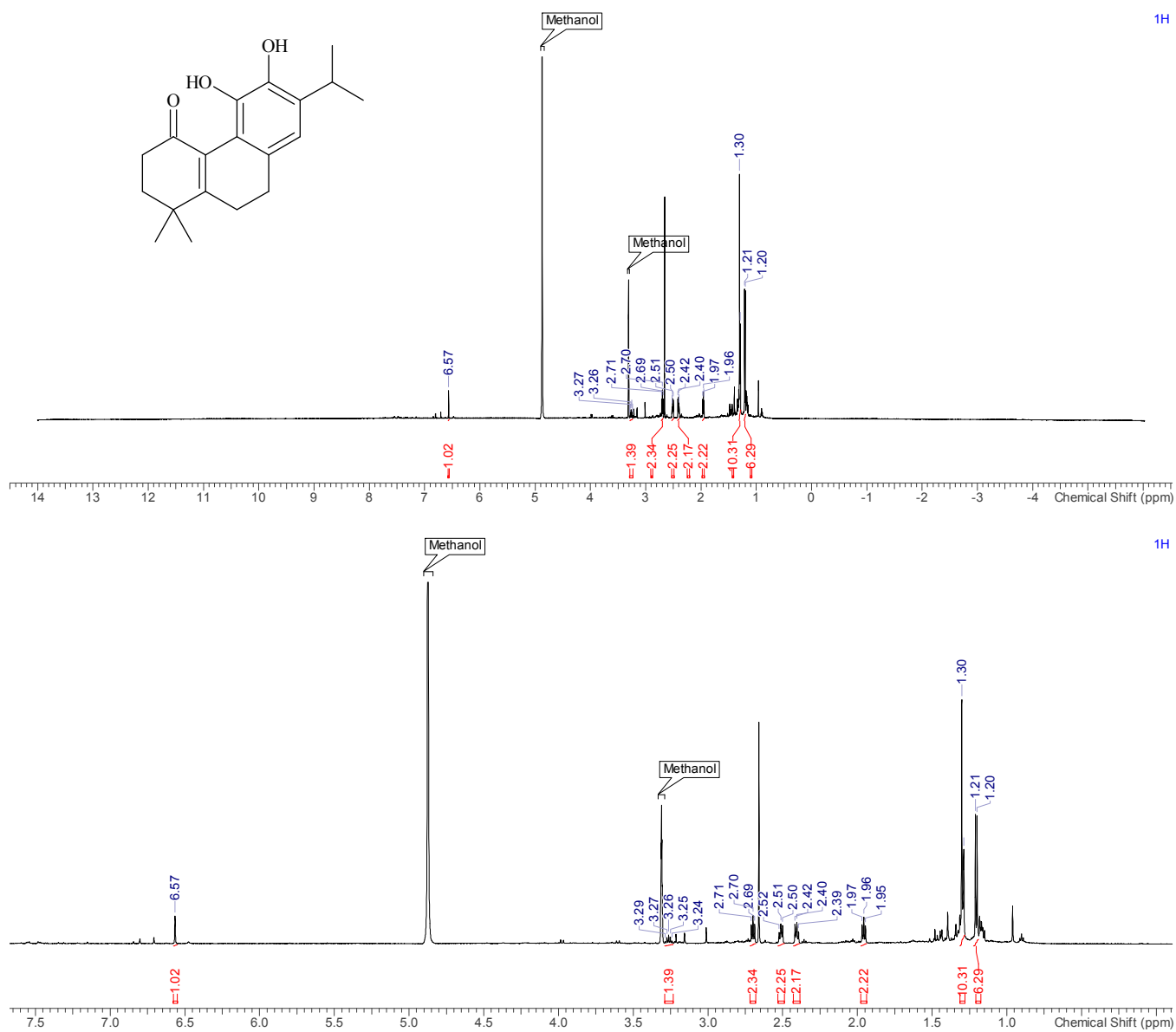


Figure S19. Proton NMR spectra (full and inset) of compound 7 in CD₃OD at 600 MHz.

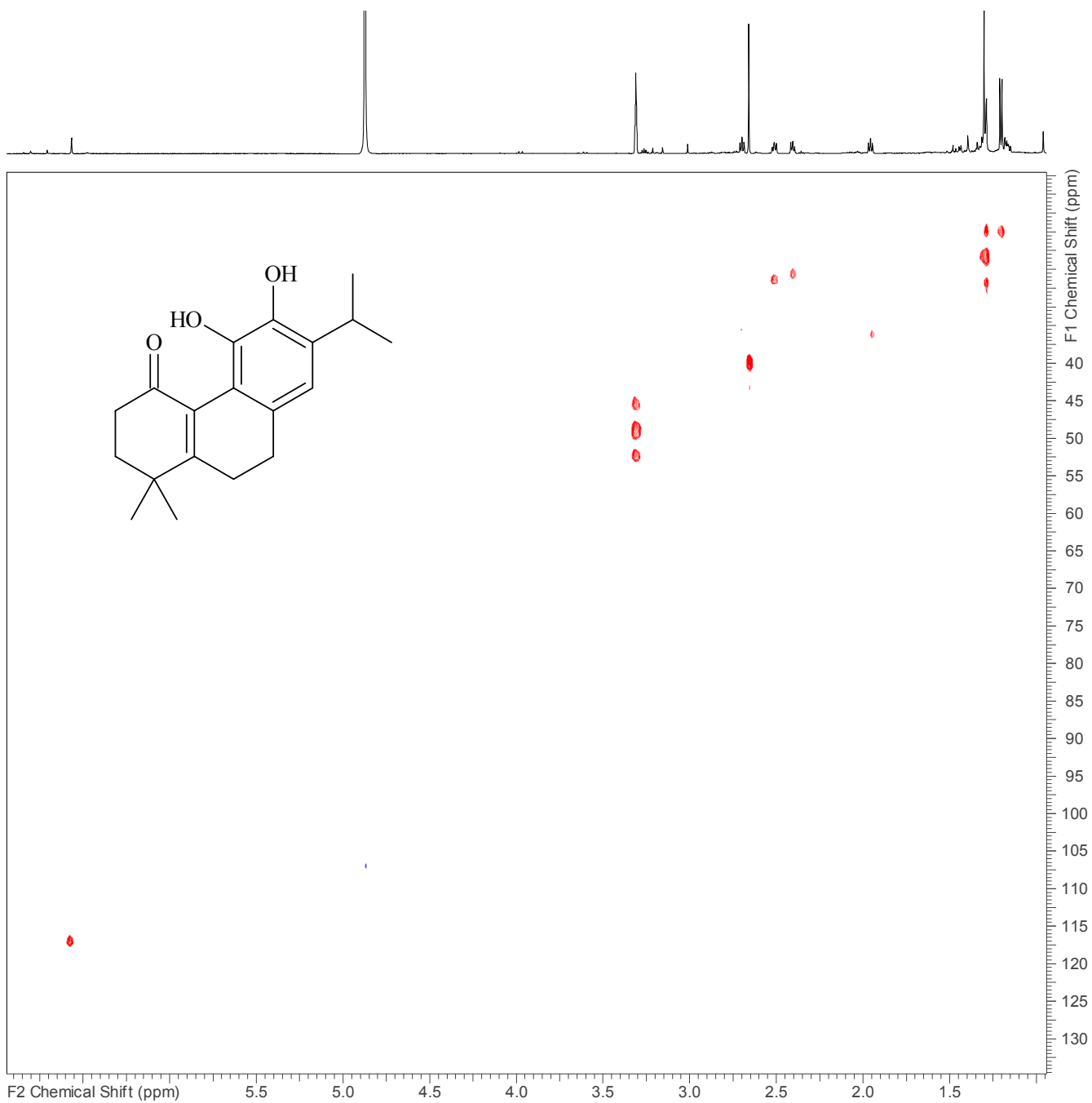


Figure S20. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **7** in CD_3OD at 600 MHz.

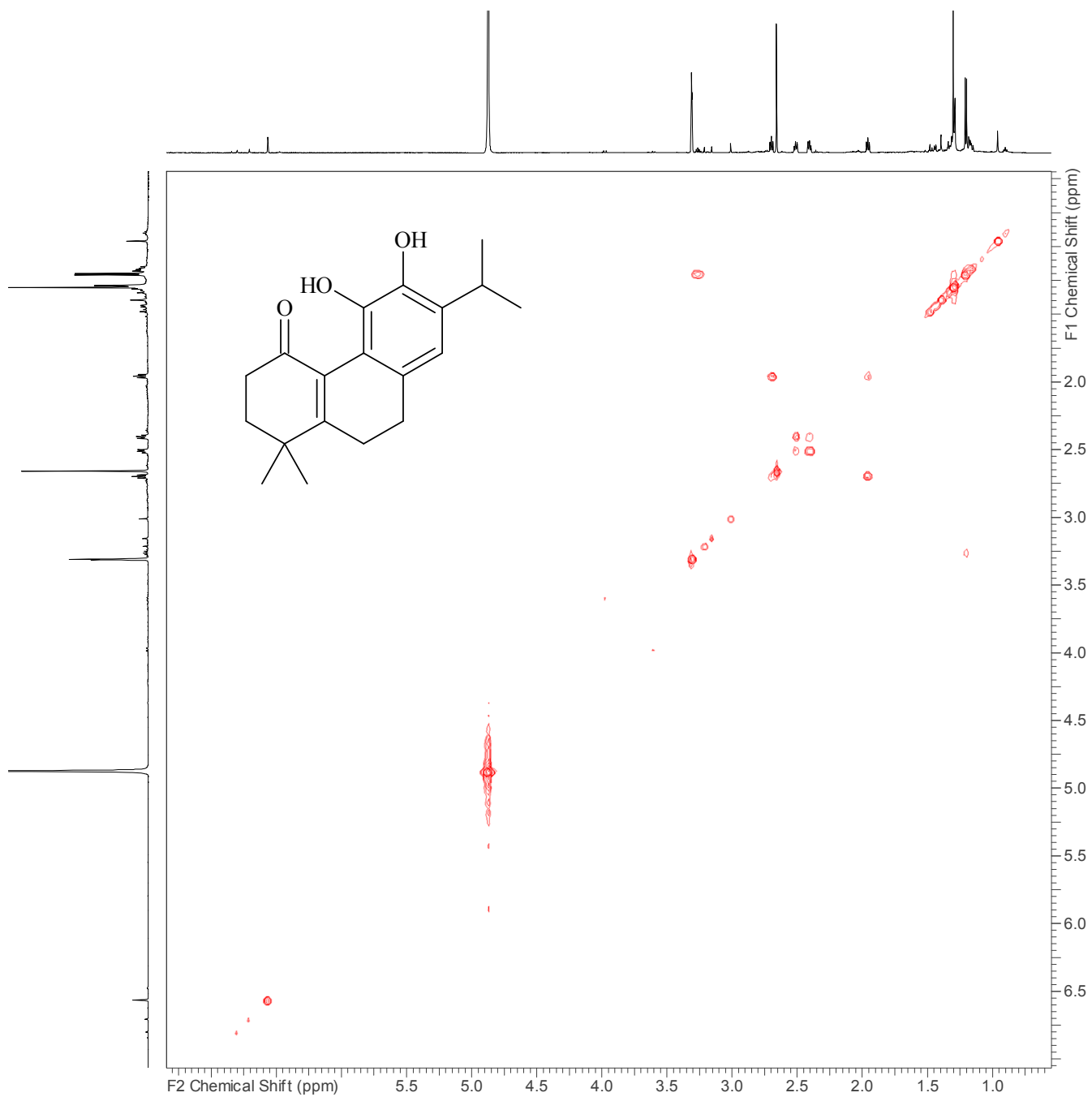


Figure S21. Expanded area of a ¹H-¹H COSY spectrum of compound 7 in CD₃OD at 600 MHz.

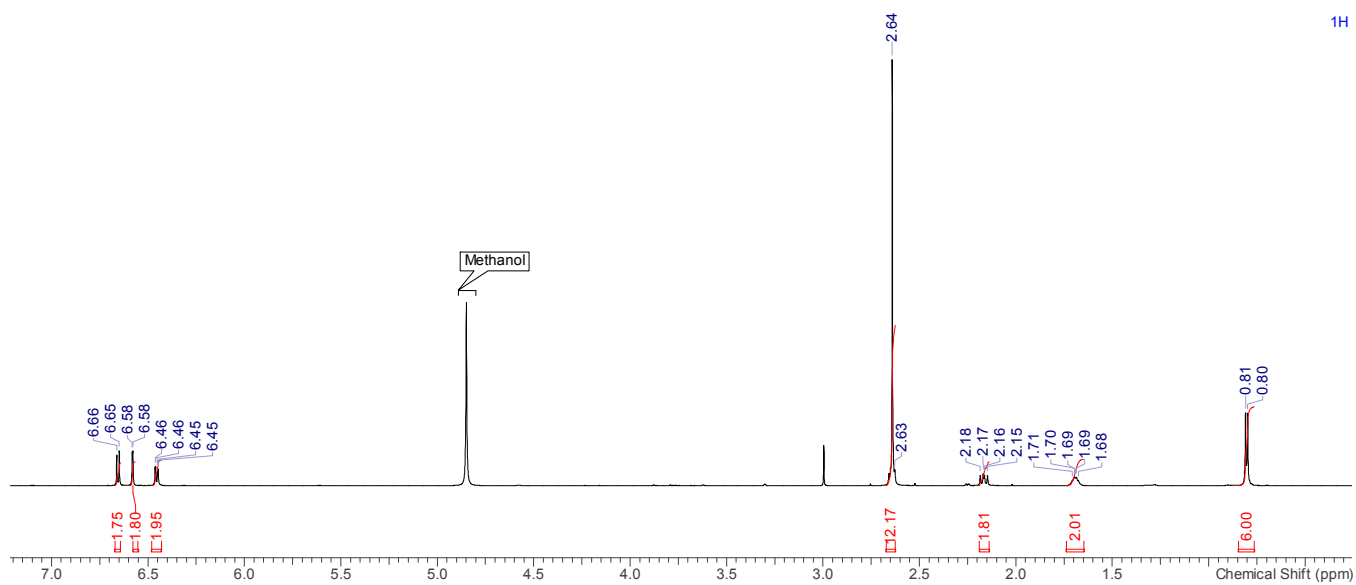
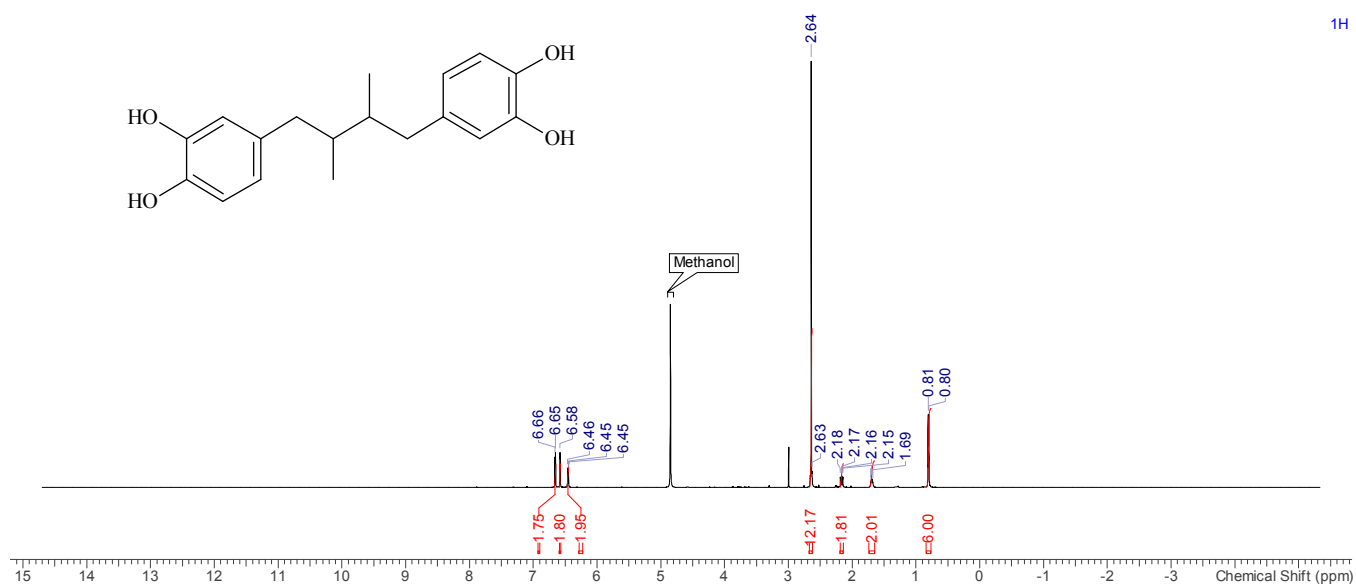


Figure S22. Proton NMR spectra (full and inset) of compound **8** in CD₃OD at 600 MHz.

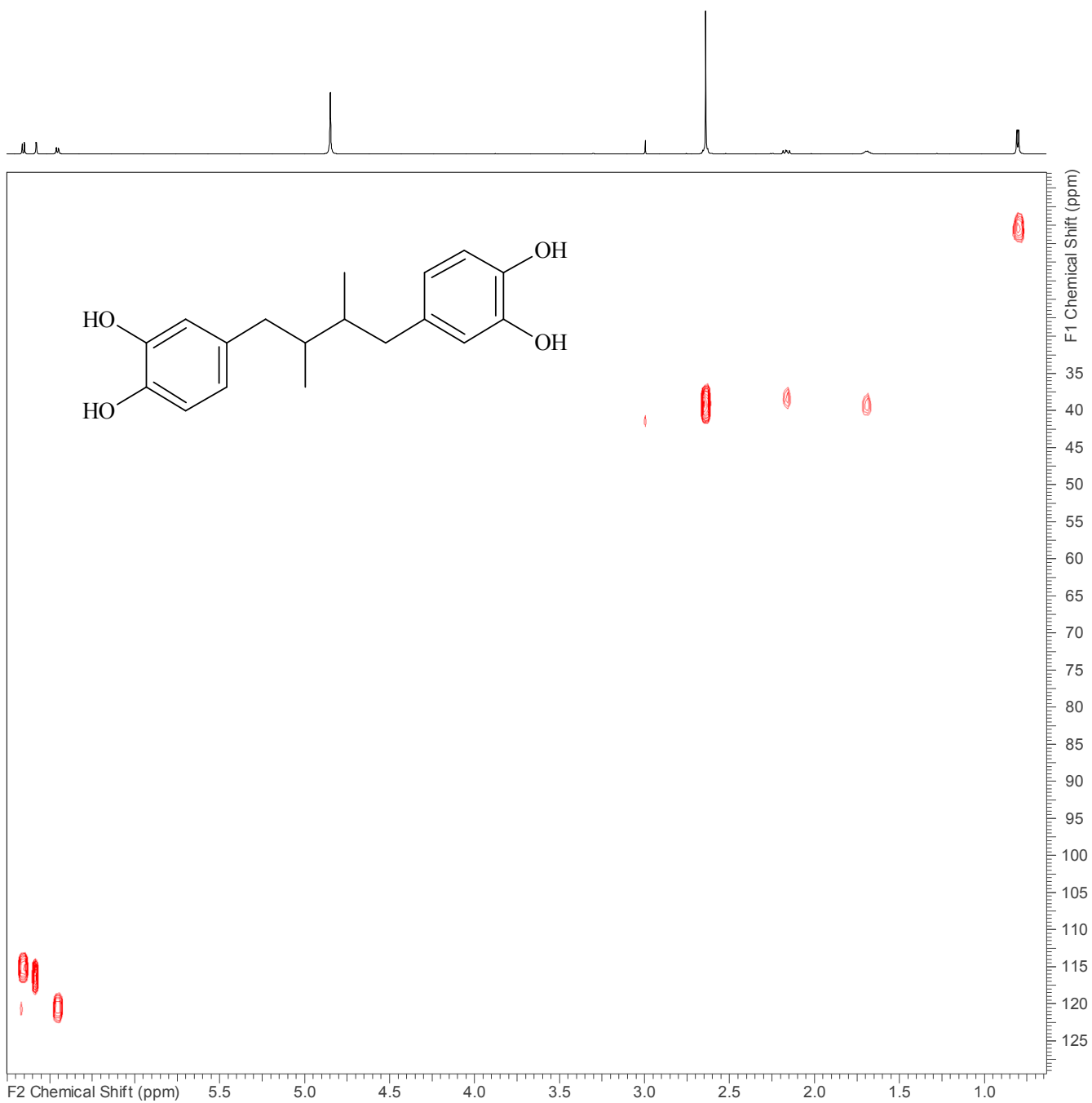


Figure S23. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **8** in CD_3OD at 600 MHz.

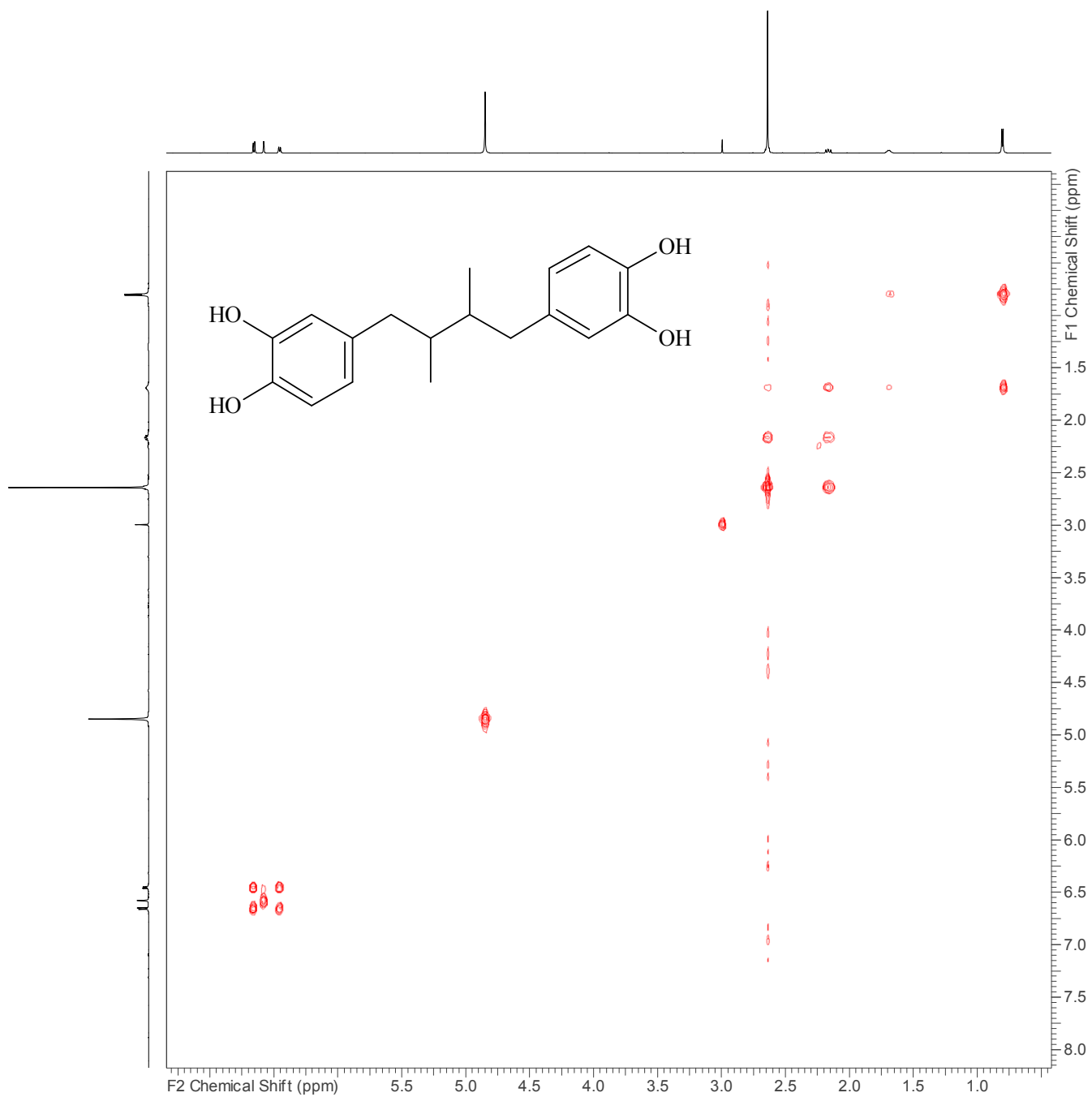


Figure S24. Expanded area of a ^1H - ^1H COSY spectrum of compound **8** in CD_3OD at 600 MHz.

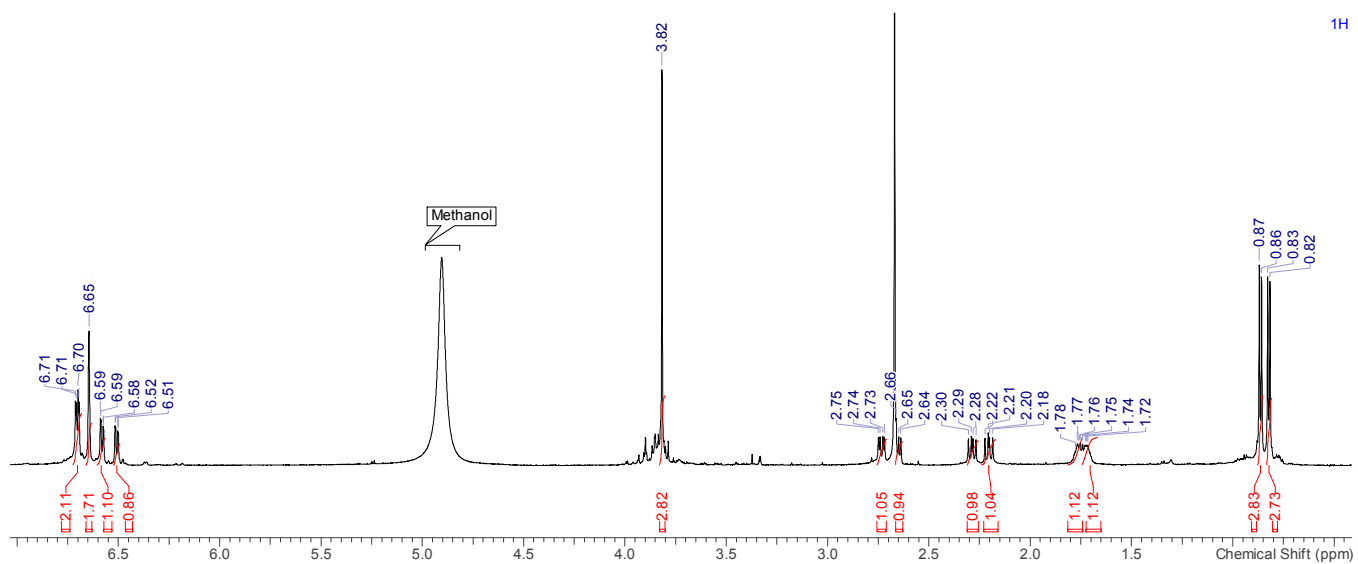
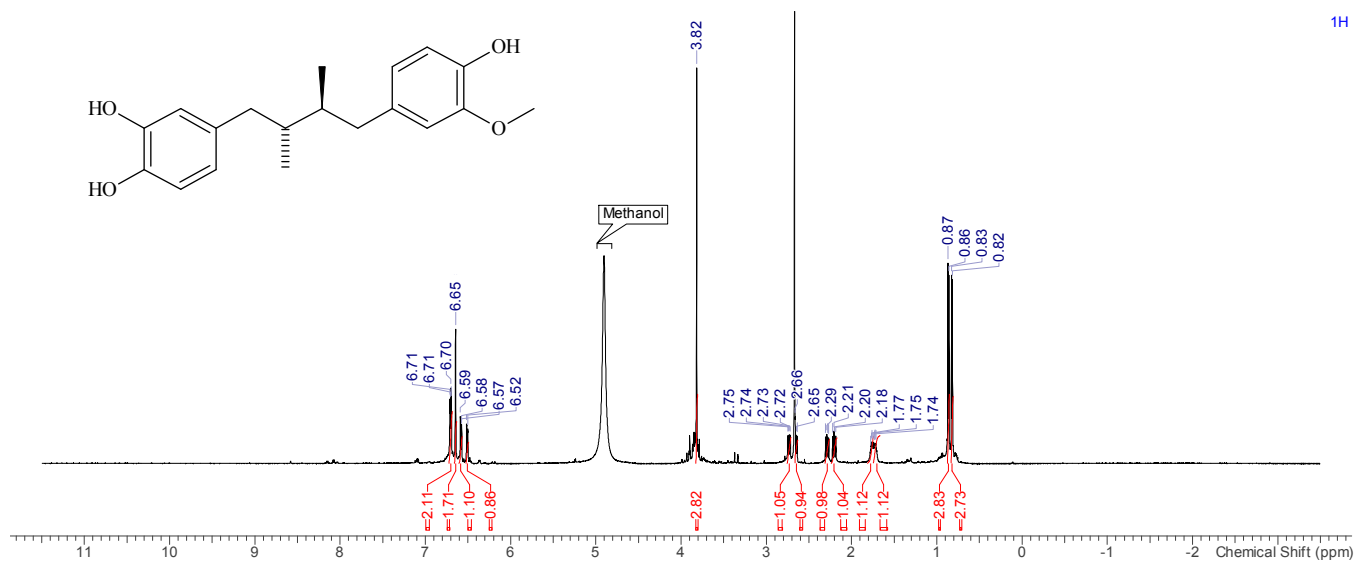


Figure S25. Proton NMR spectra (full and inset) of compound **9** in CD₃OD at 600 MHz.

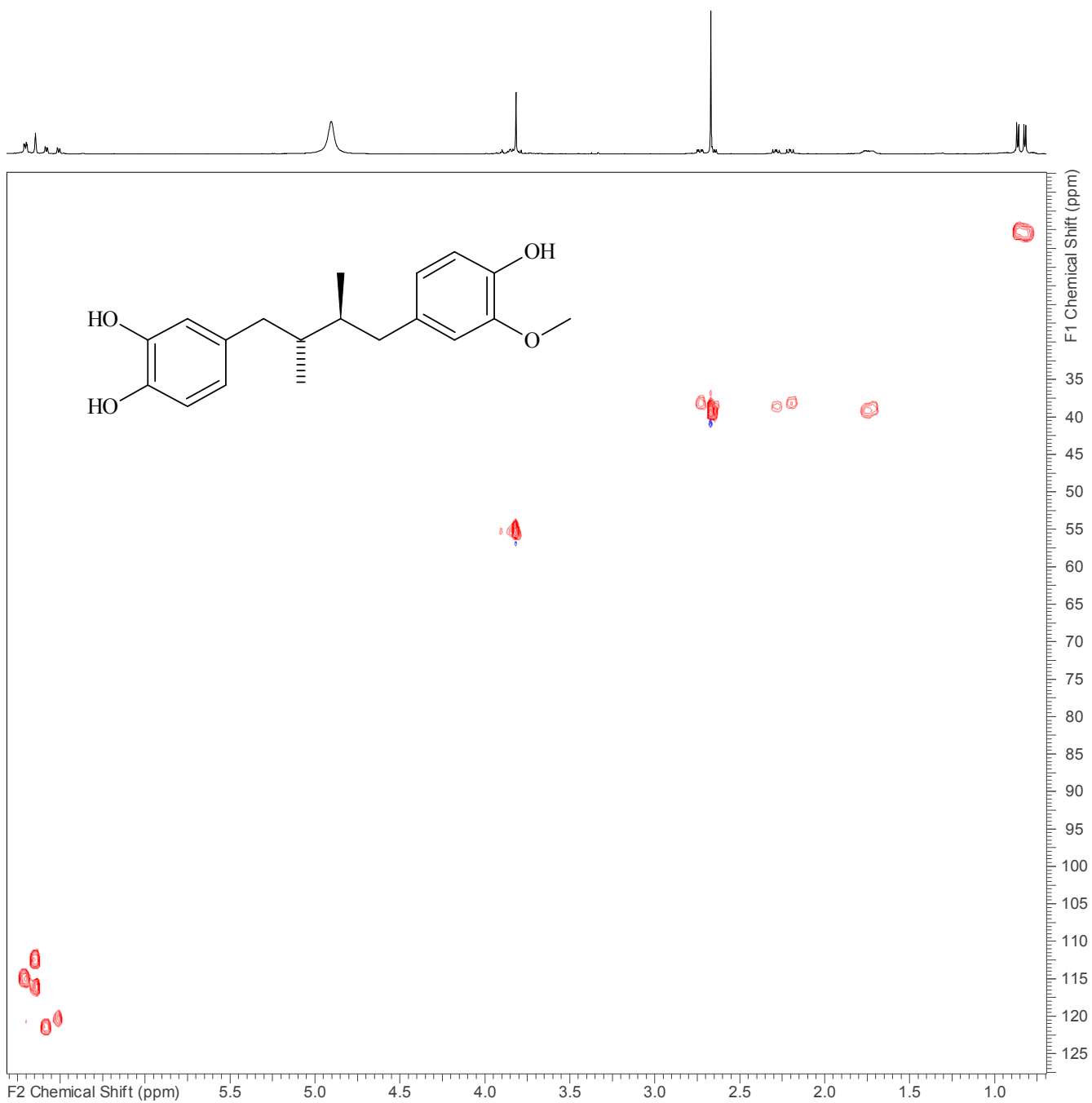




Figure S27. Expanded area of a ^1H - ^1H COSY spectrum of compound **9** in CD_3OD at 600 MHz.

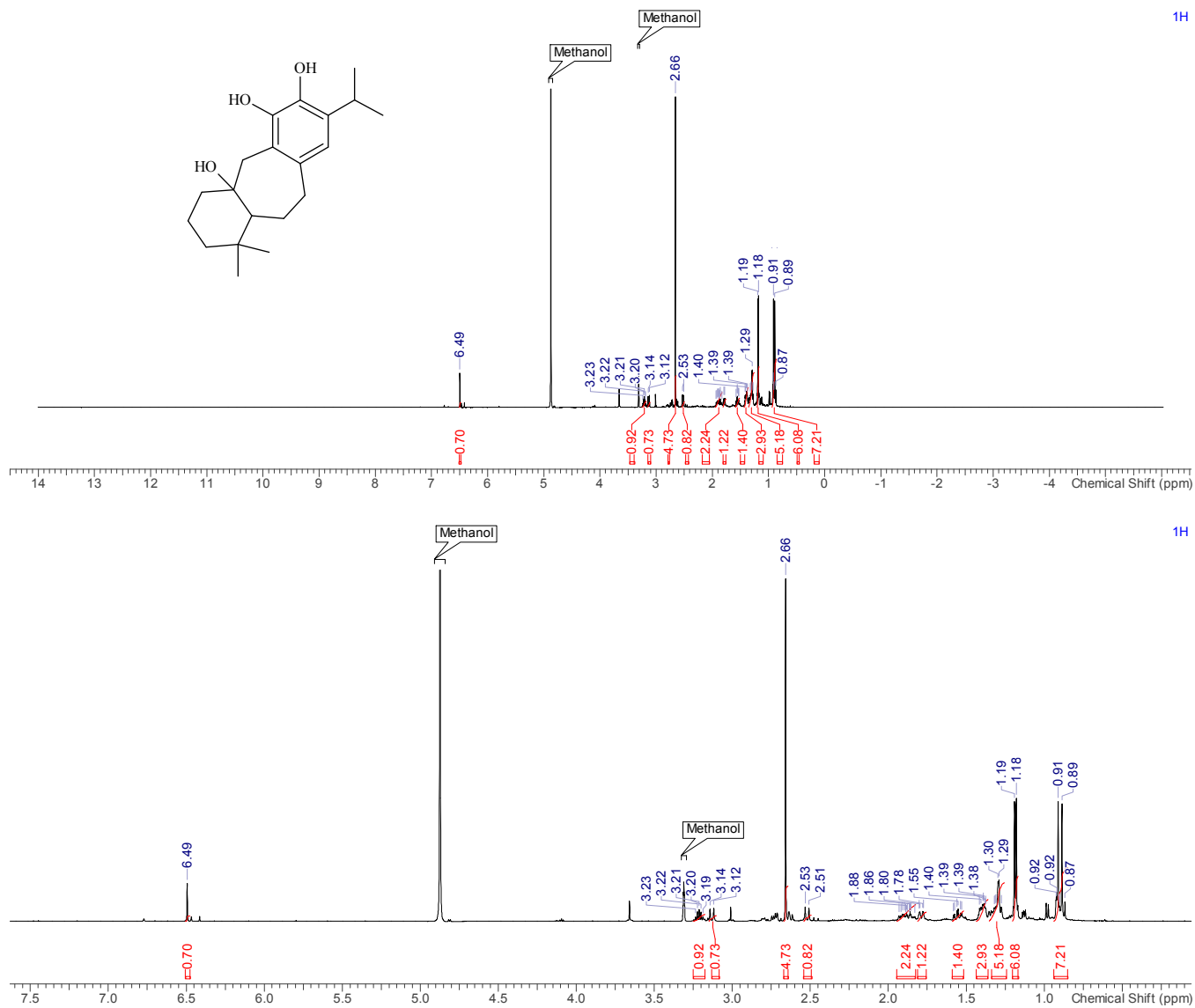


Figure S28. Proton NMR spectra (full and inset) of compound 10 in CD₃OD at 600 MHz.

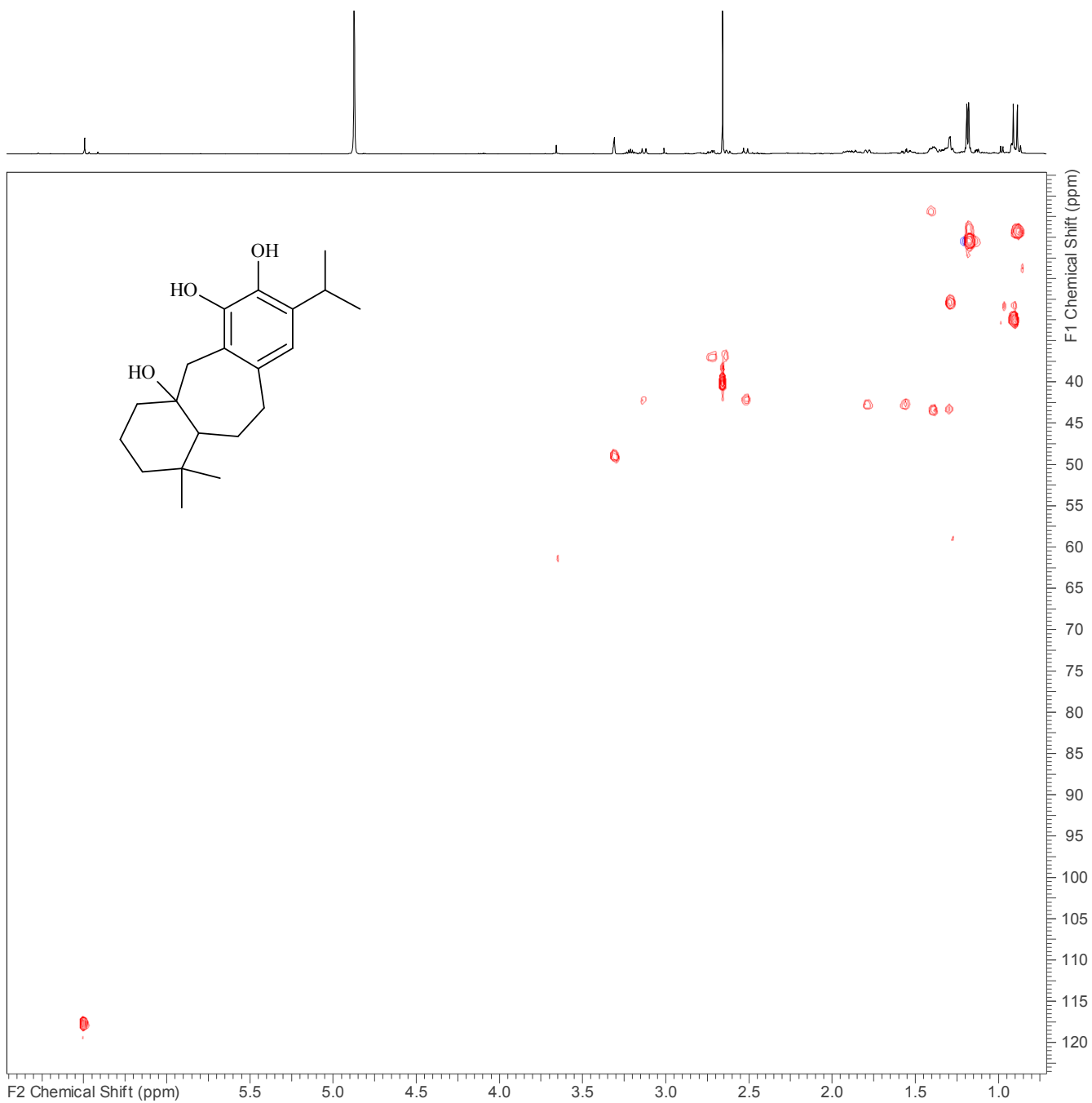


Figure S29. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **10** in CD_3OD at 600 MHz.

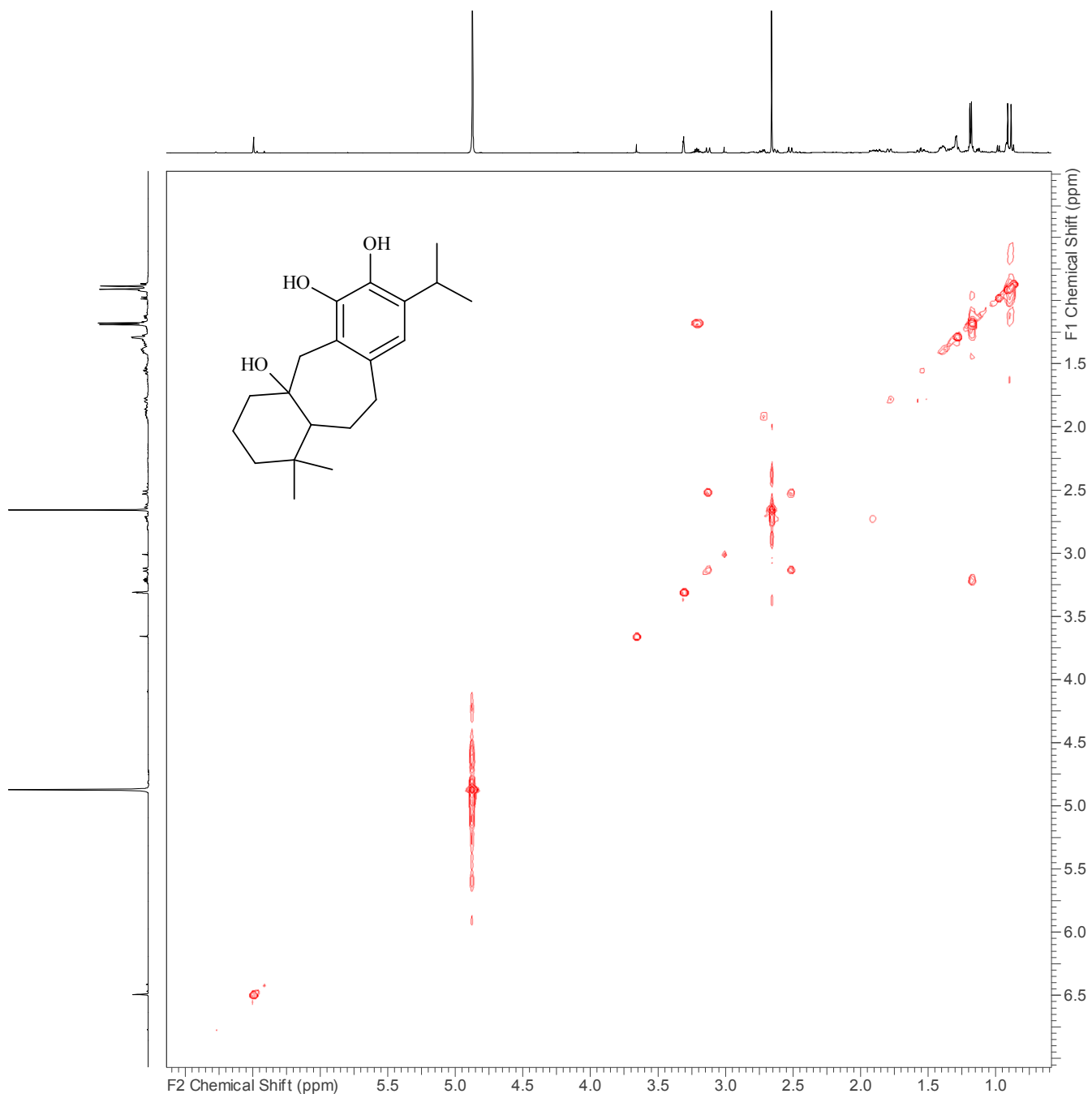


Figure S30. Expanded area of a ^1H - ^1H COSY spectrum of compound **10** in CD_3OD at 600 MHz.

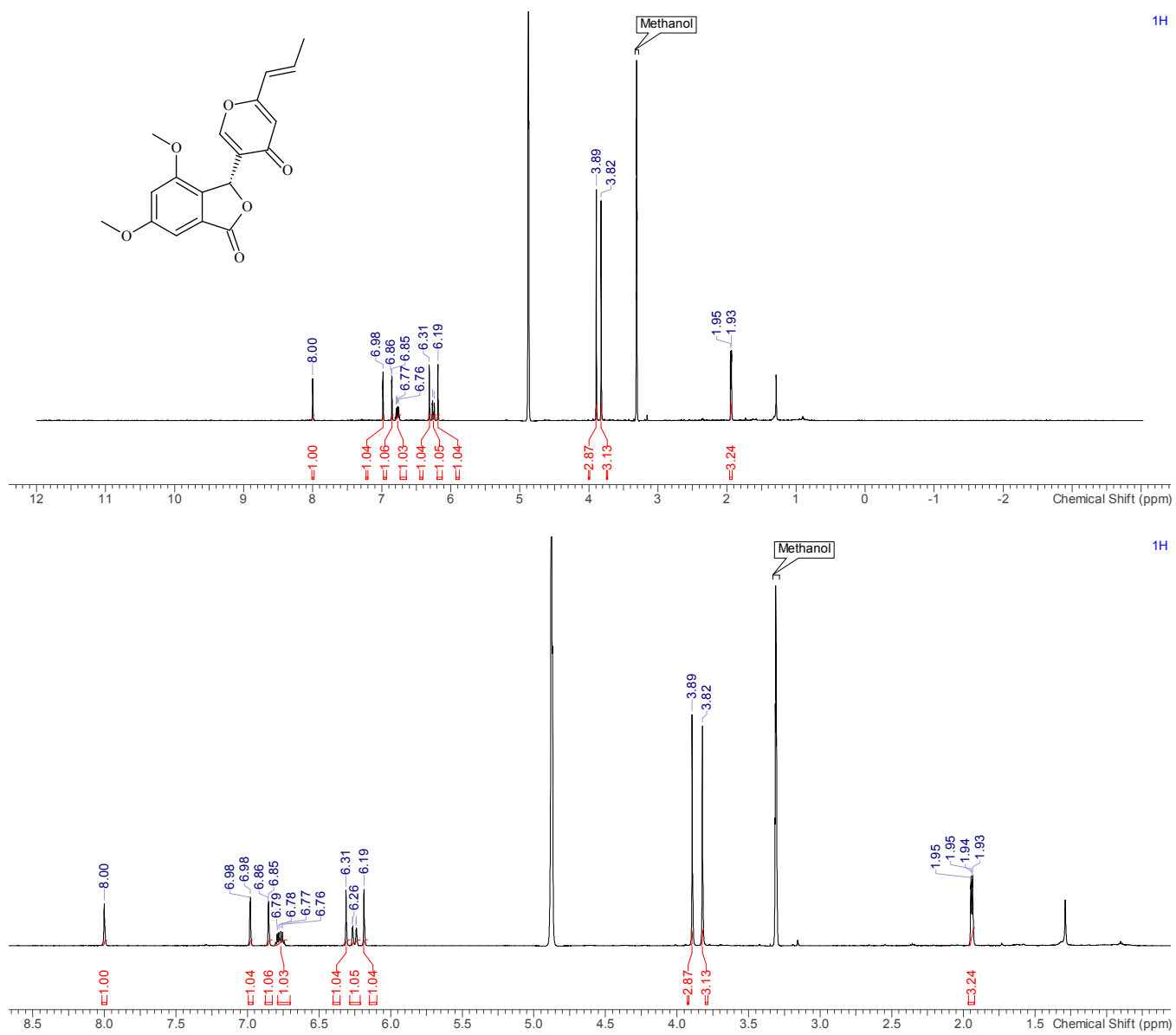


Figure S31. Proton NMR spectra (full and inset) of compound **11** in CD₃OD at 600 MHz.

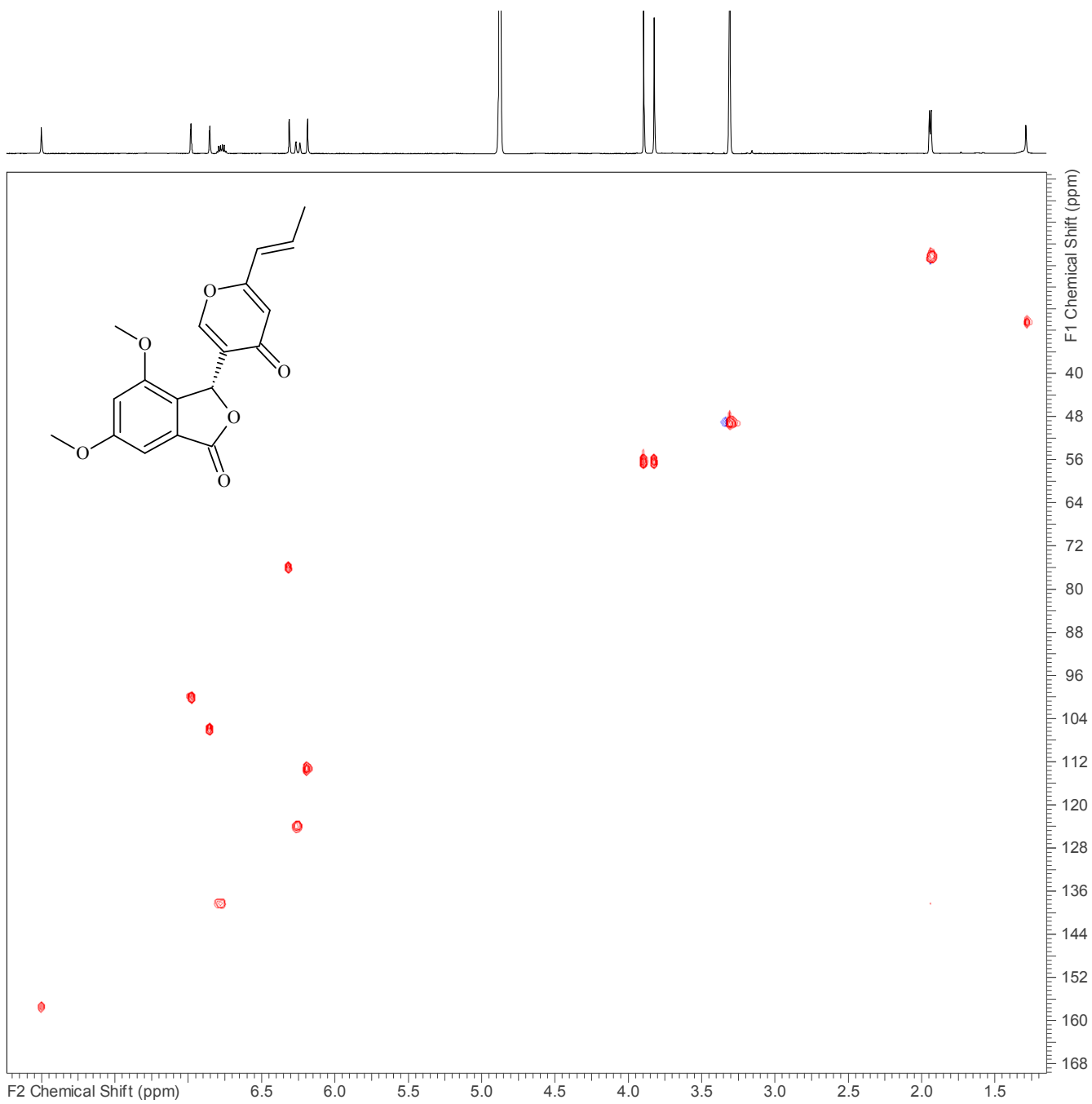


Figure S32. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **11** in CD_3OD at 600 MHz.

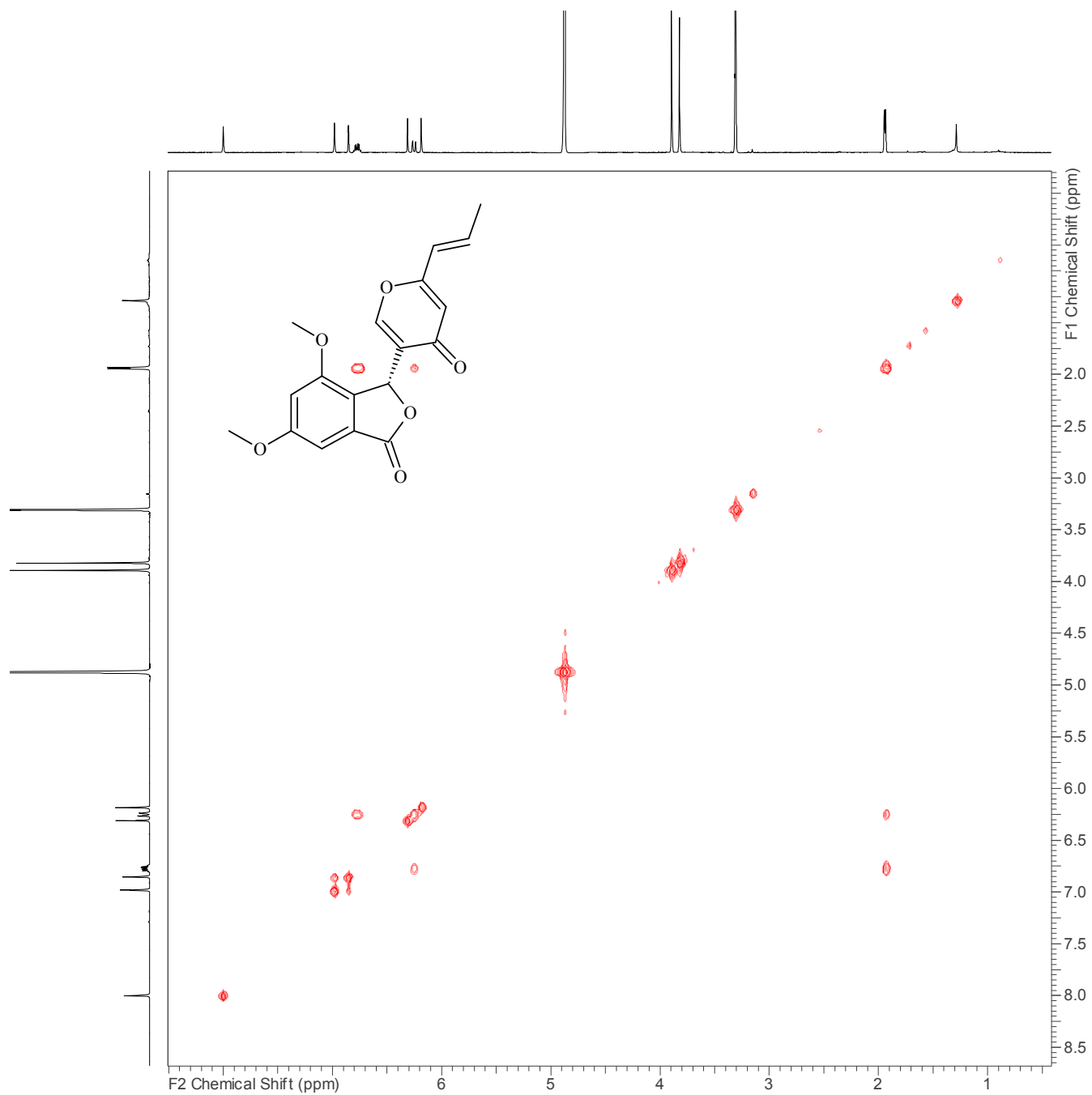


Figure S33. Expanded area of a ¹H-¹H COSY spectrum of compound **11** in CD₃OD at 600 MHz.

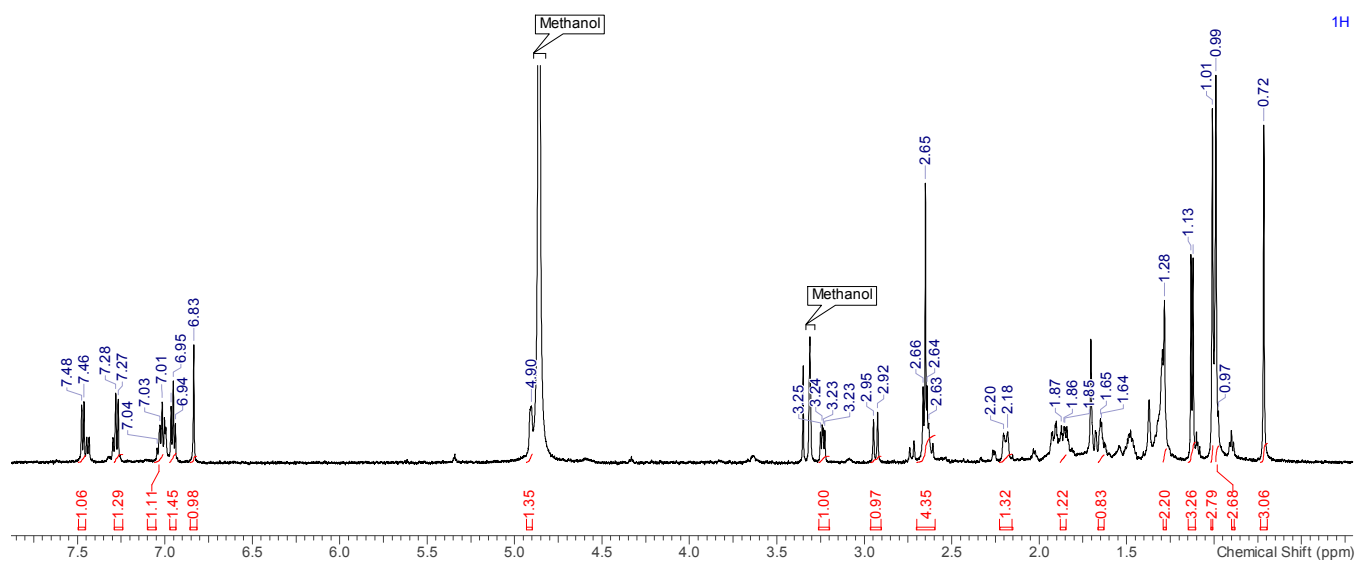
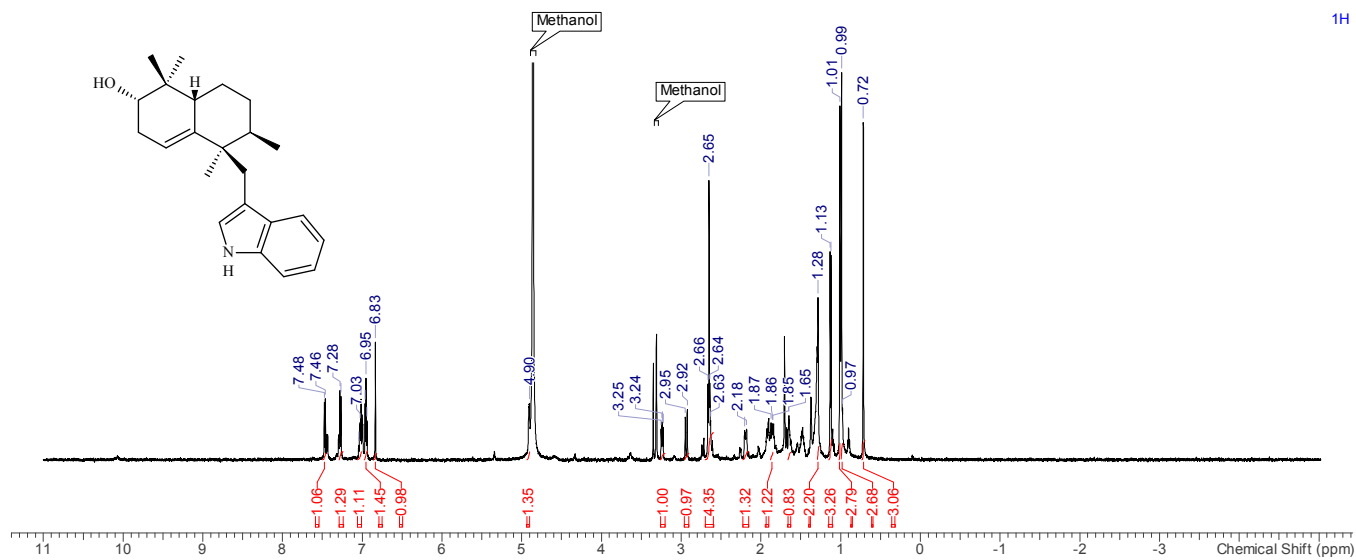


Figure S34. Proton NMR spectra (full and inset) of compound **12** in CD₃OD at 600 MHz.

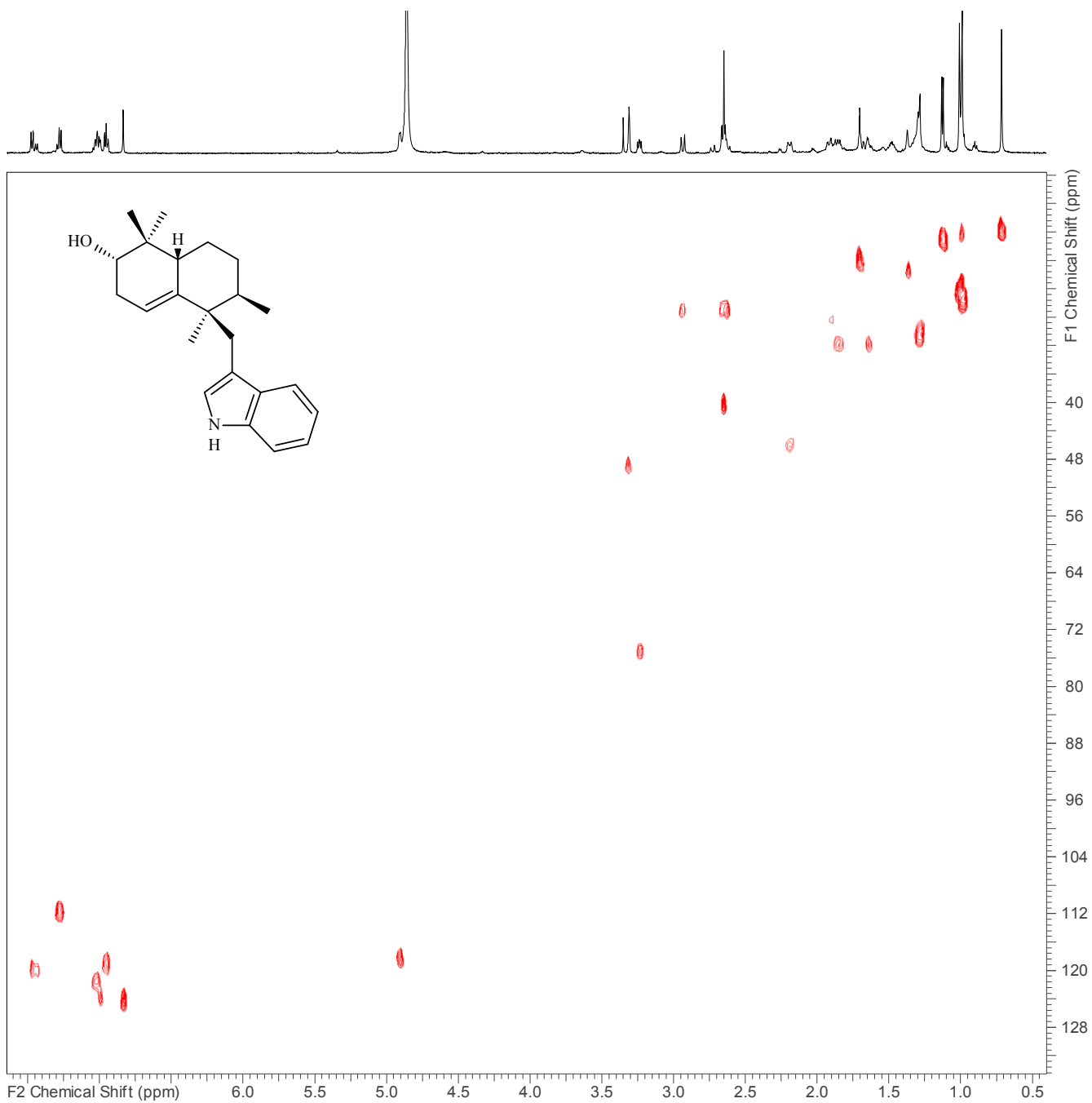


Figure S35. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **12** in CD_3OD at 600 MHz.

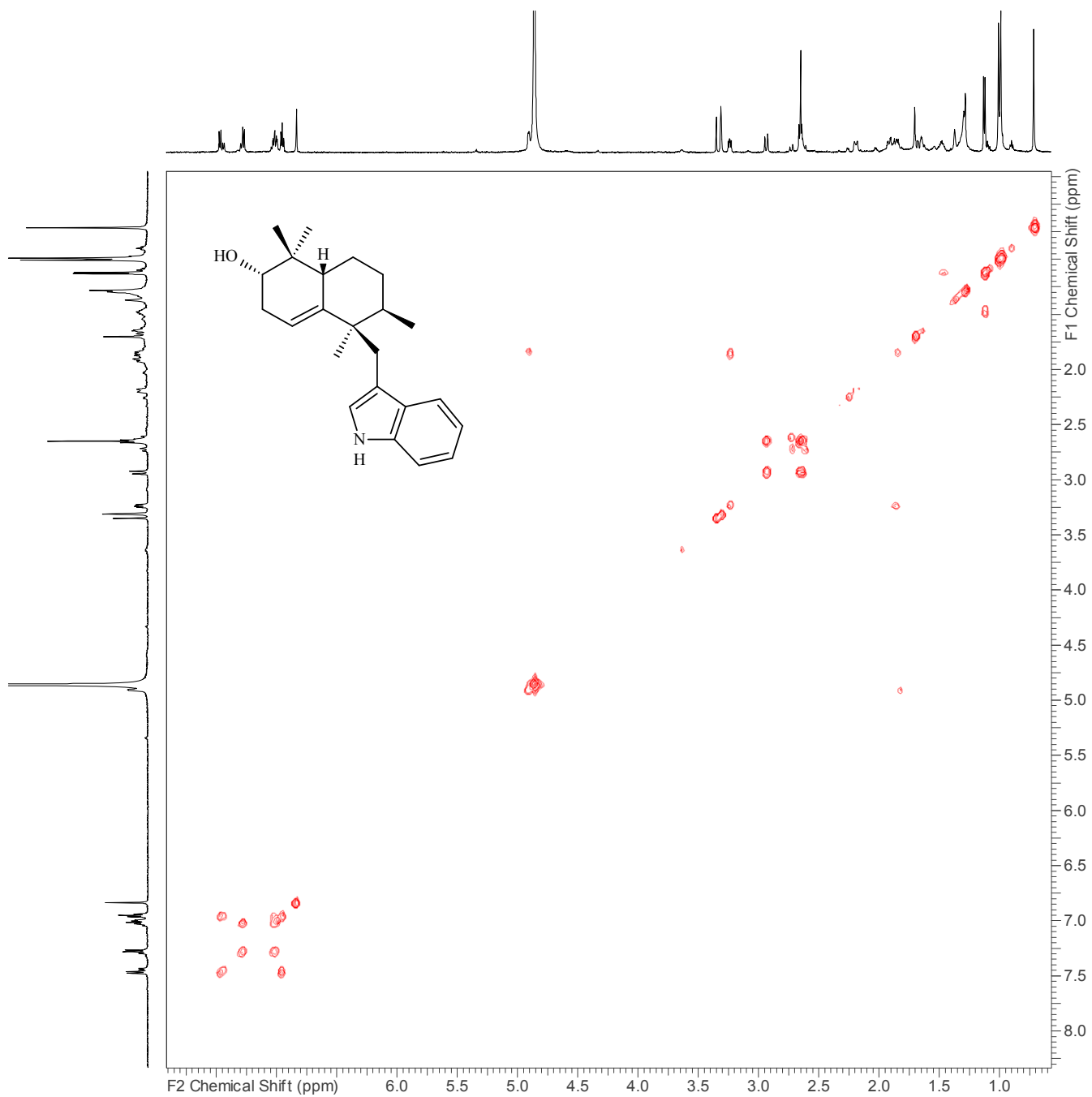
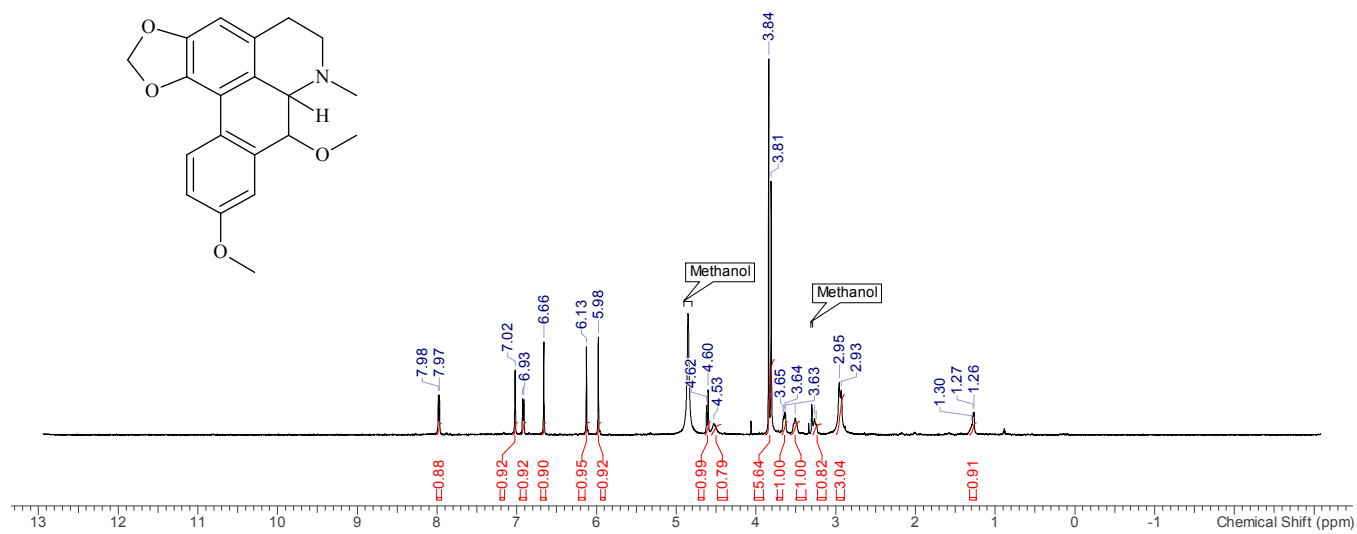


Figure S36. Expanded area of a ^1H - ^1H COSY spectrum of compound **12** in CD_3OD at 600 MHz.

1H



1H

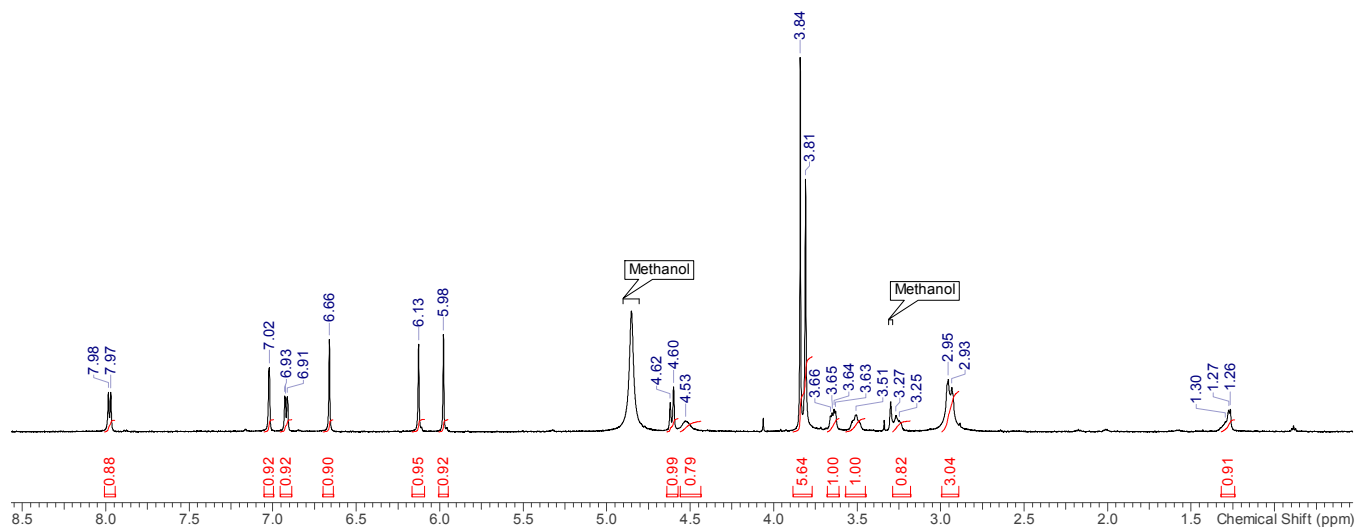


Figure S37. Proton NMR spectra (full and inset) of compound 13 in CD₃OD at 600 MHz.

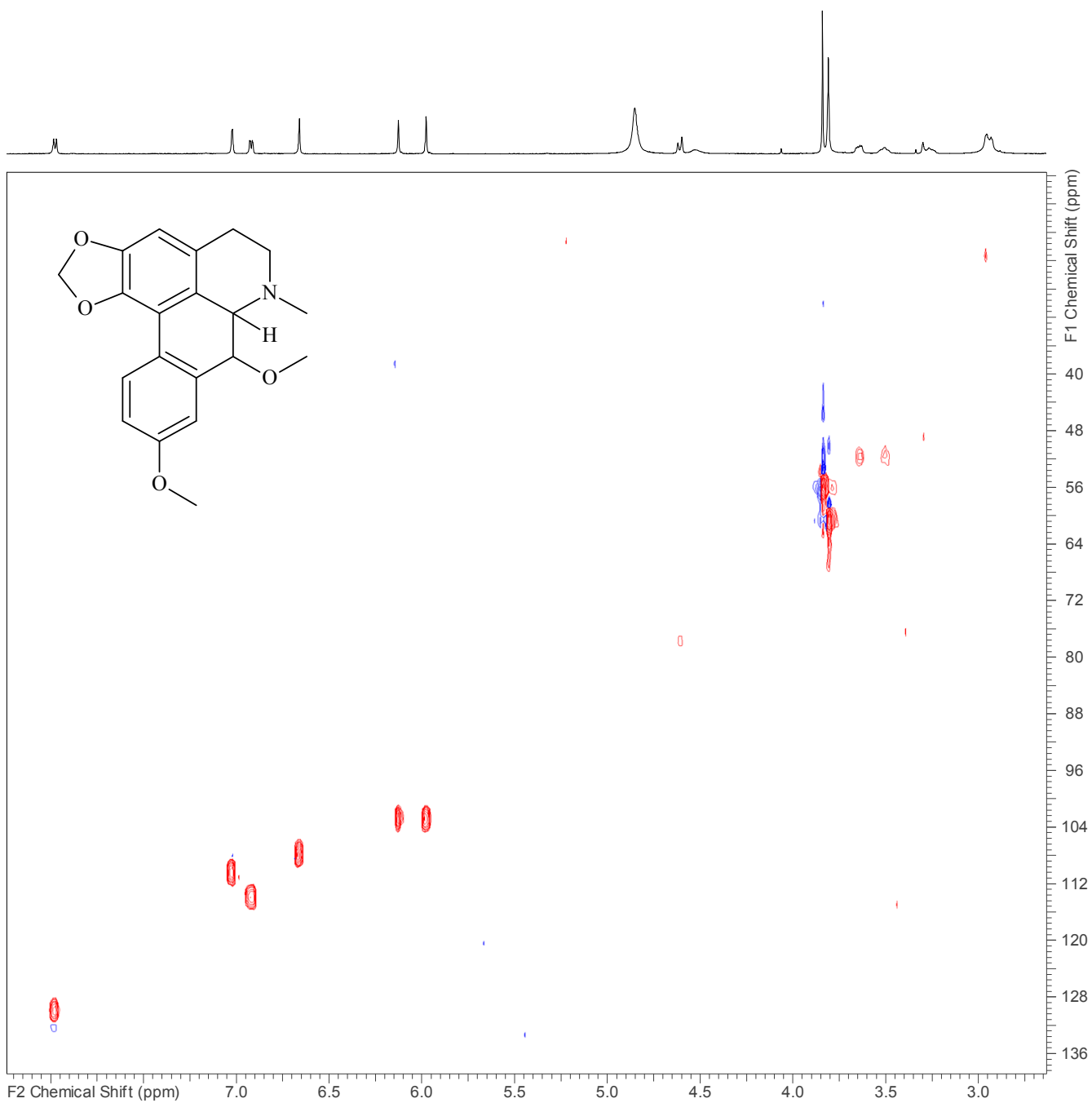


Figure S38. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **13** in CD_3OD at 600 MHz.

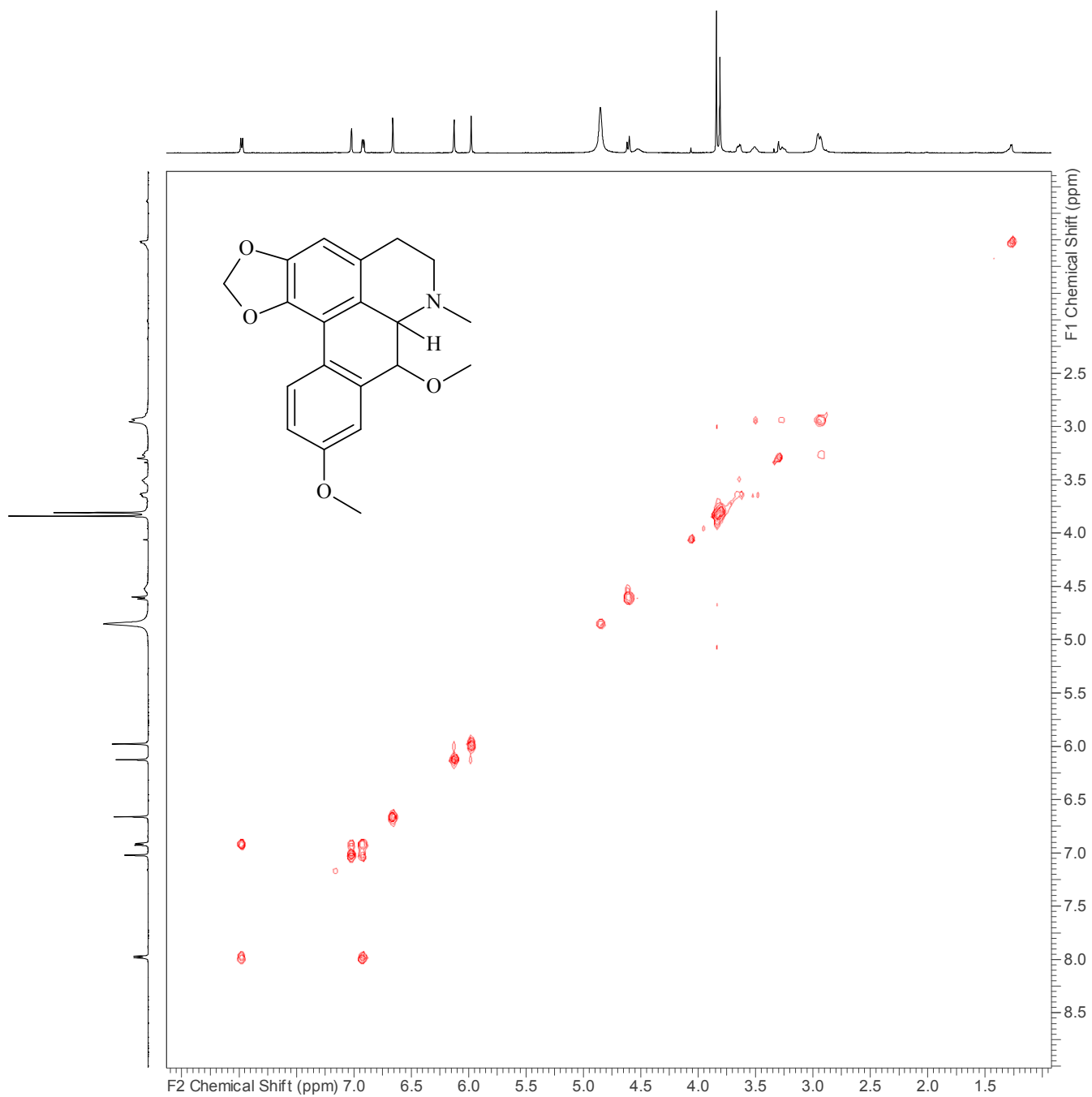
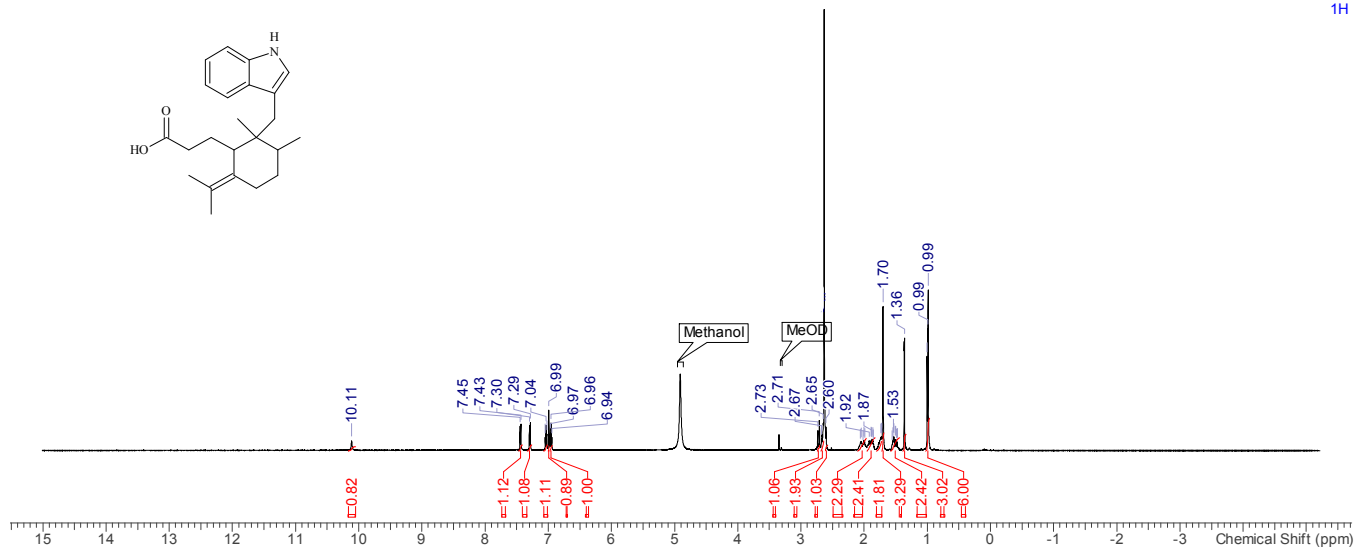


Figure S39. Expanded area of a ^1H - ^1H COSY spectrum of compound **13** in CD_3OD at 600 MHz.

1H



1H

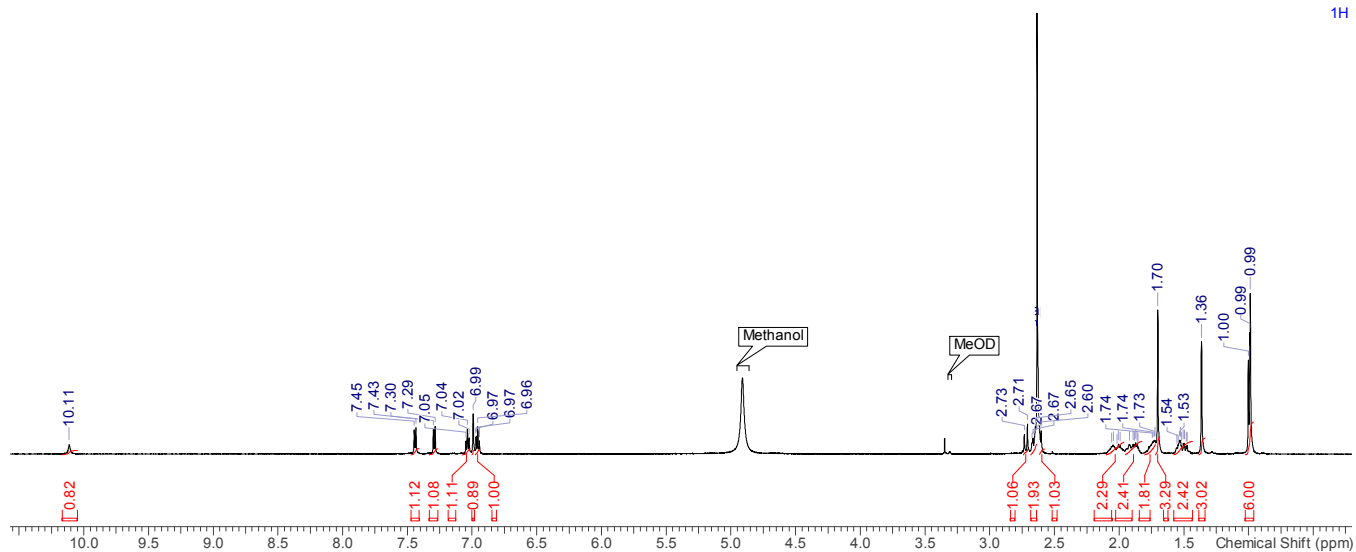


Figure S40. Proton NMR spectra (full and inset) of compound **14** in CD₃OD at 600 MHz.

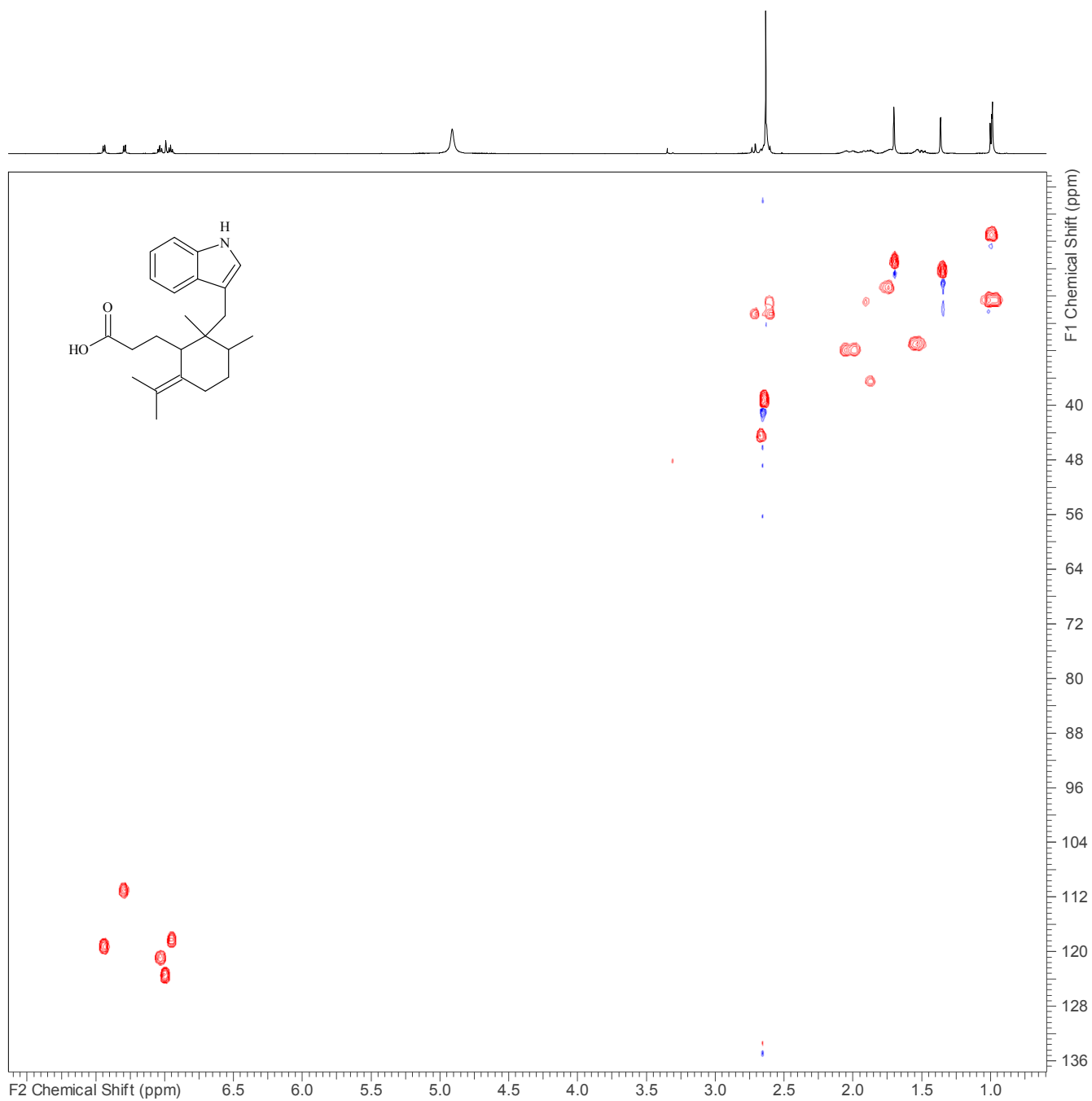


Figure S41. Expanded area of a ^1H - ^{13}C HMQC spectrum of compound **14** in CD_3OD at 600 MHz.

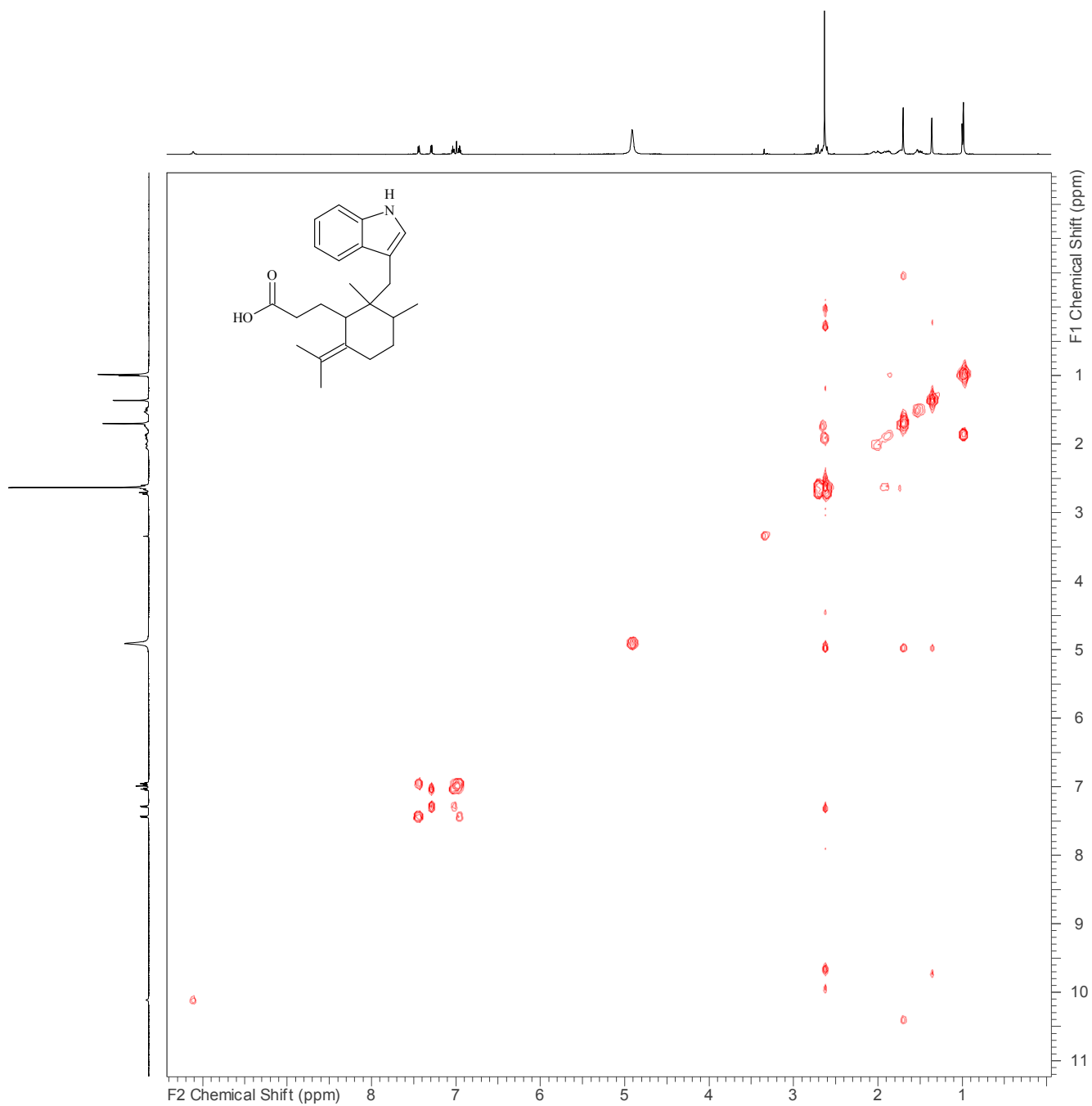


Figure S42. Expanded area of a ^1H - ^1H COSY spectrum of compound **14** in CD_3OD at 600 MHz.

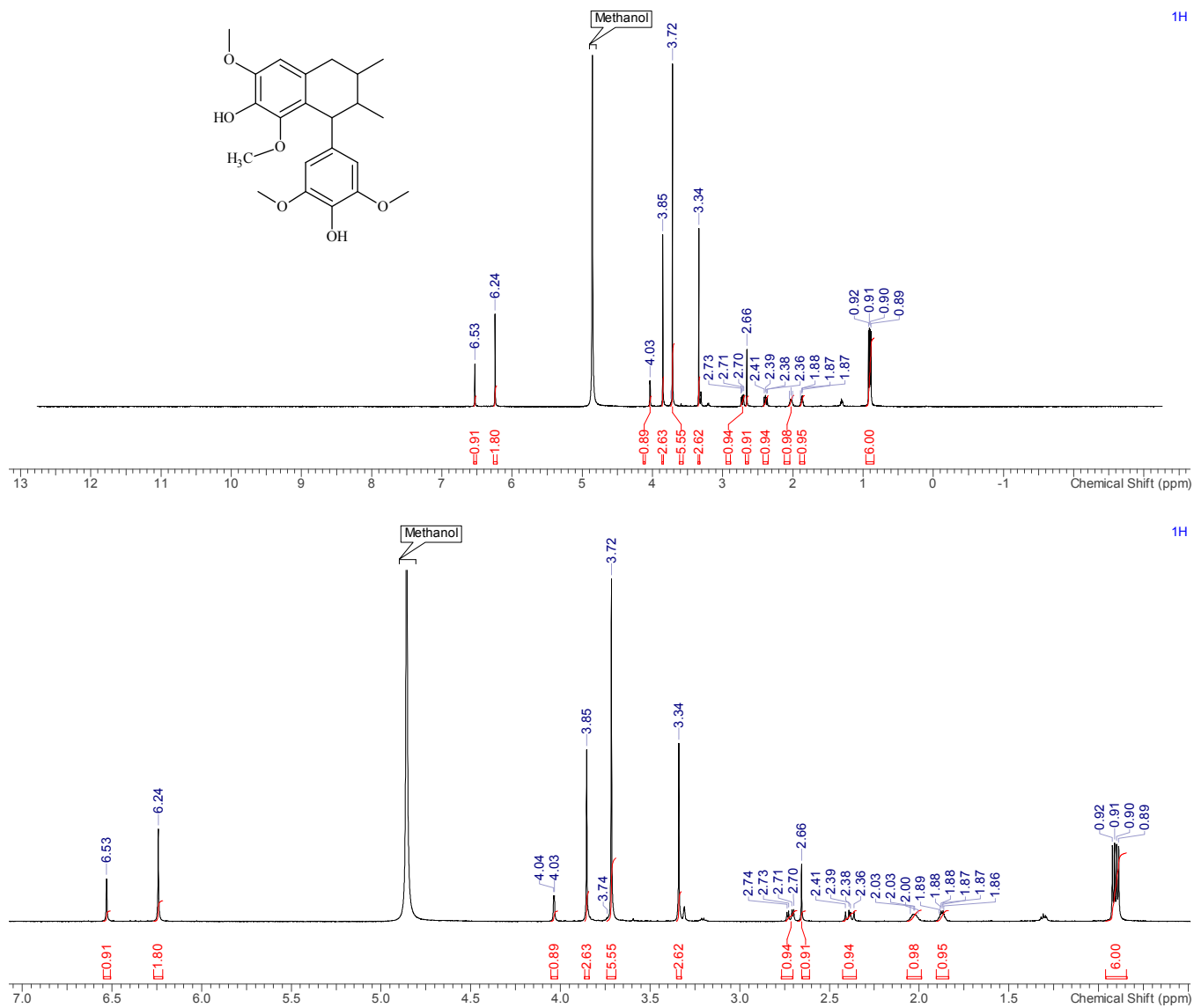


Figure S43. Proton NMR spectra (full and inset) of compound **15** in CD₃OD at 600 MHz.

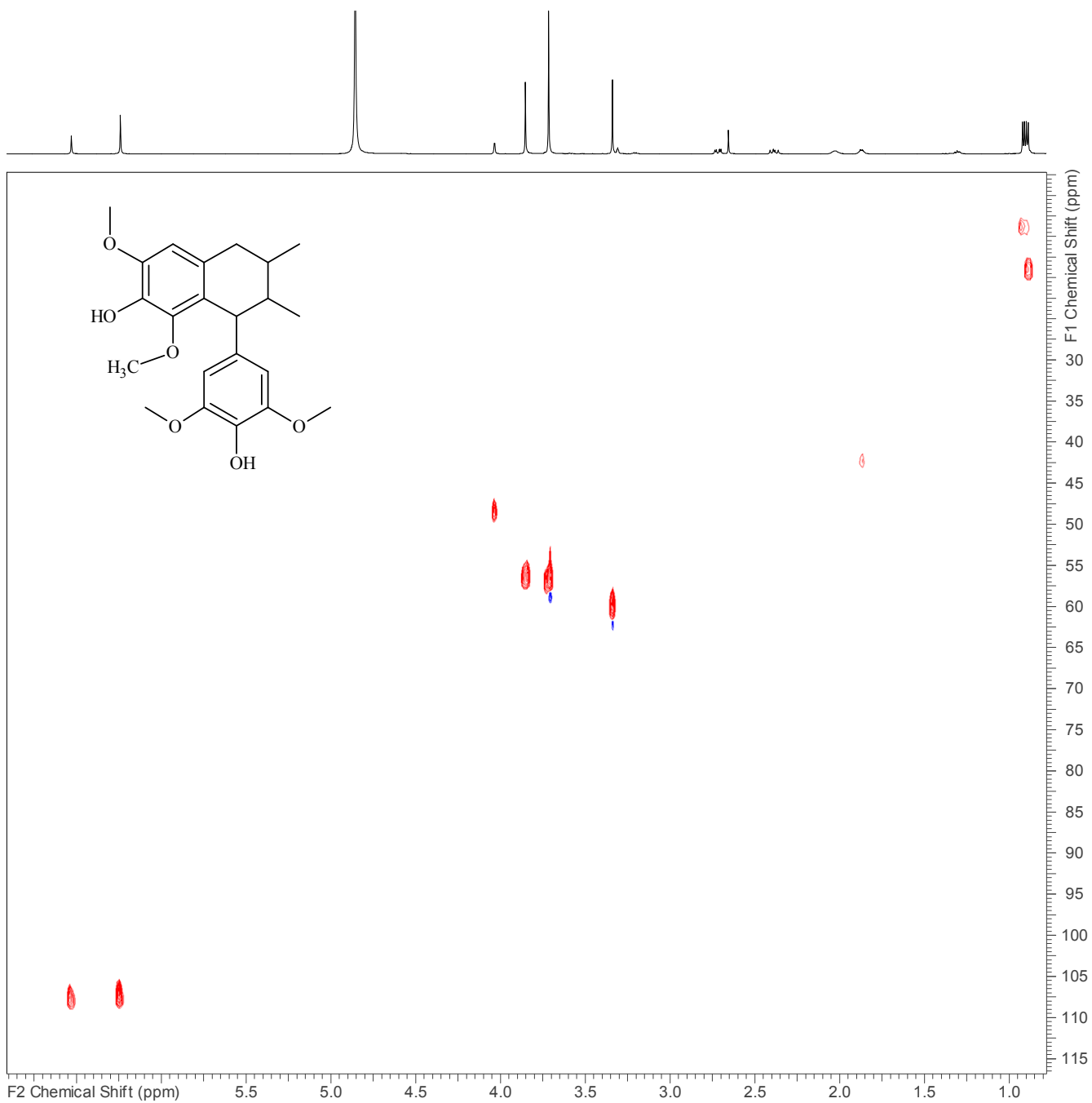


Figure S44. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **15** in CD_3OD at 600 MHz.

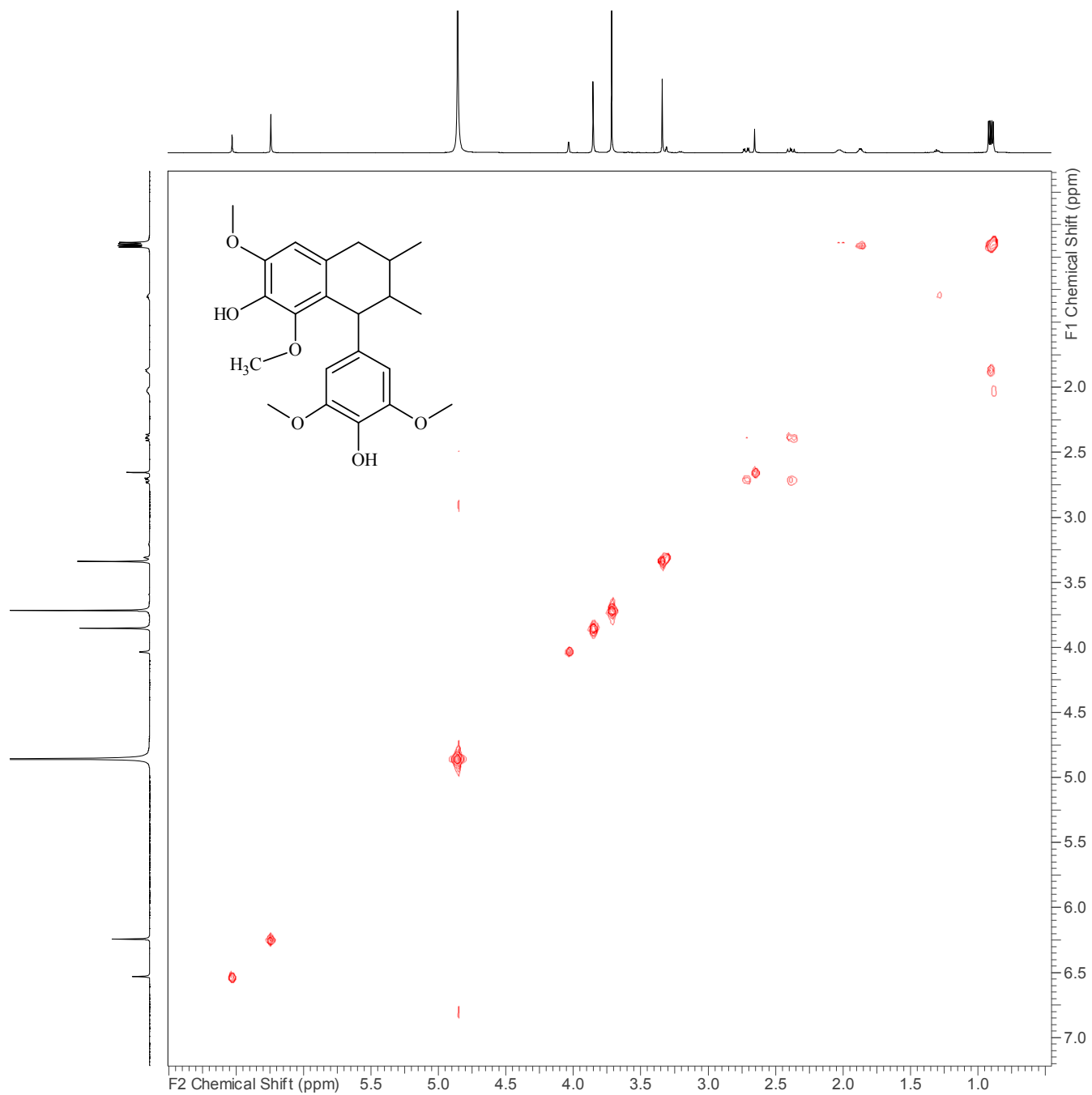
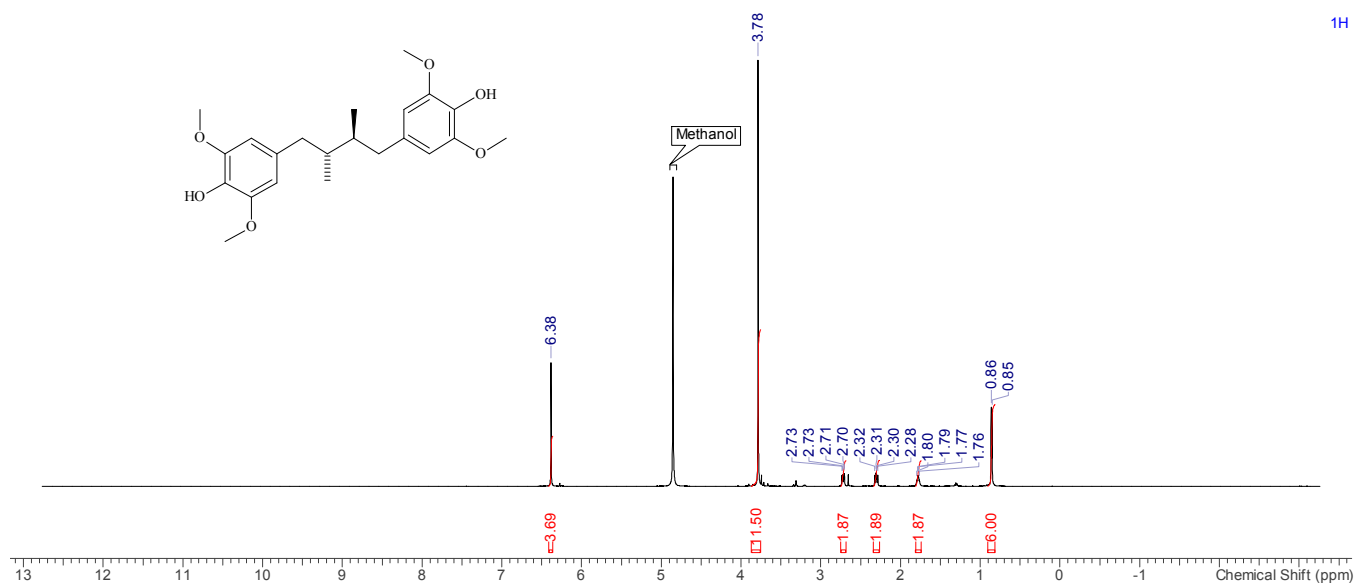


Figure S45. Expanded area of a ¹H-¹H COSY spectrum of compound **15** in CD₃OD at 600 MHz.

1H



1H

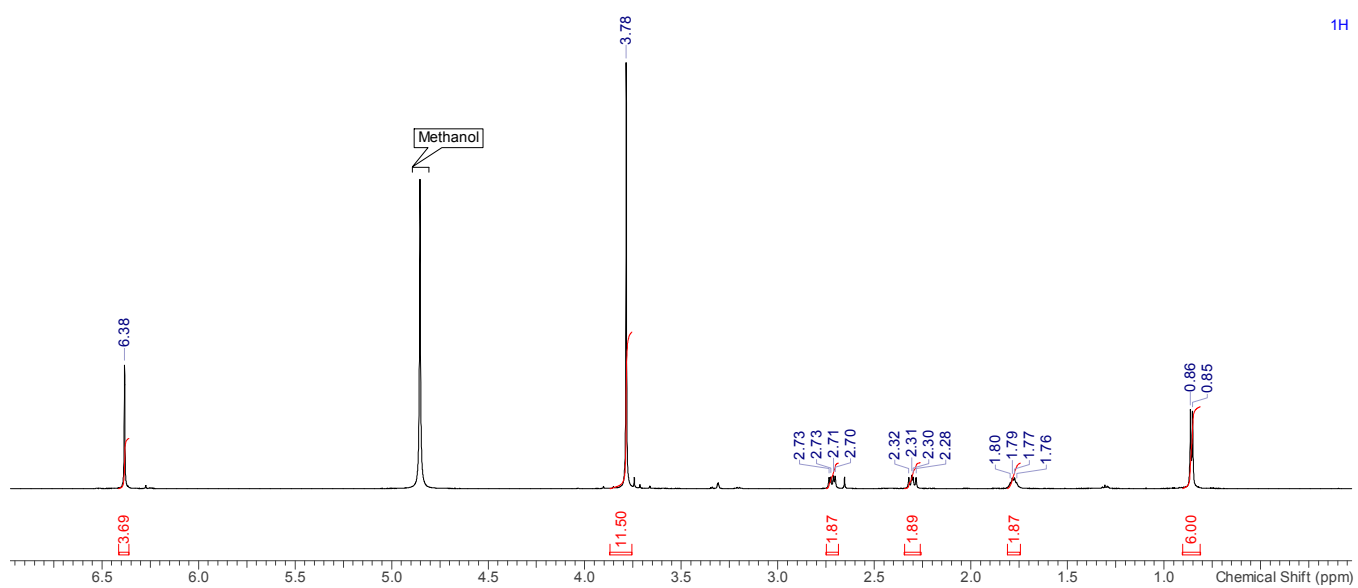


Figure S46. Proton NMR spectra (full and inset) of compound **16** in CD₃OD at 600 MHz.

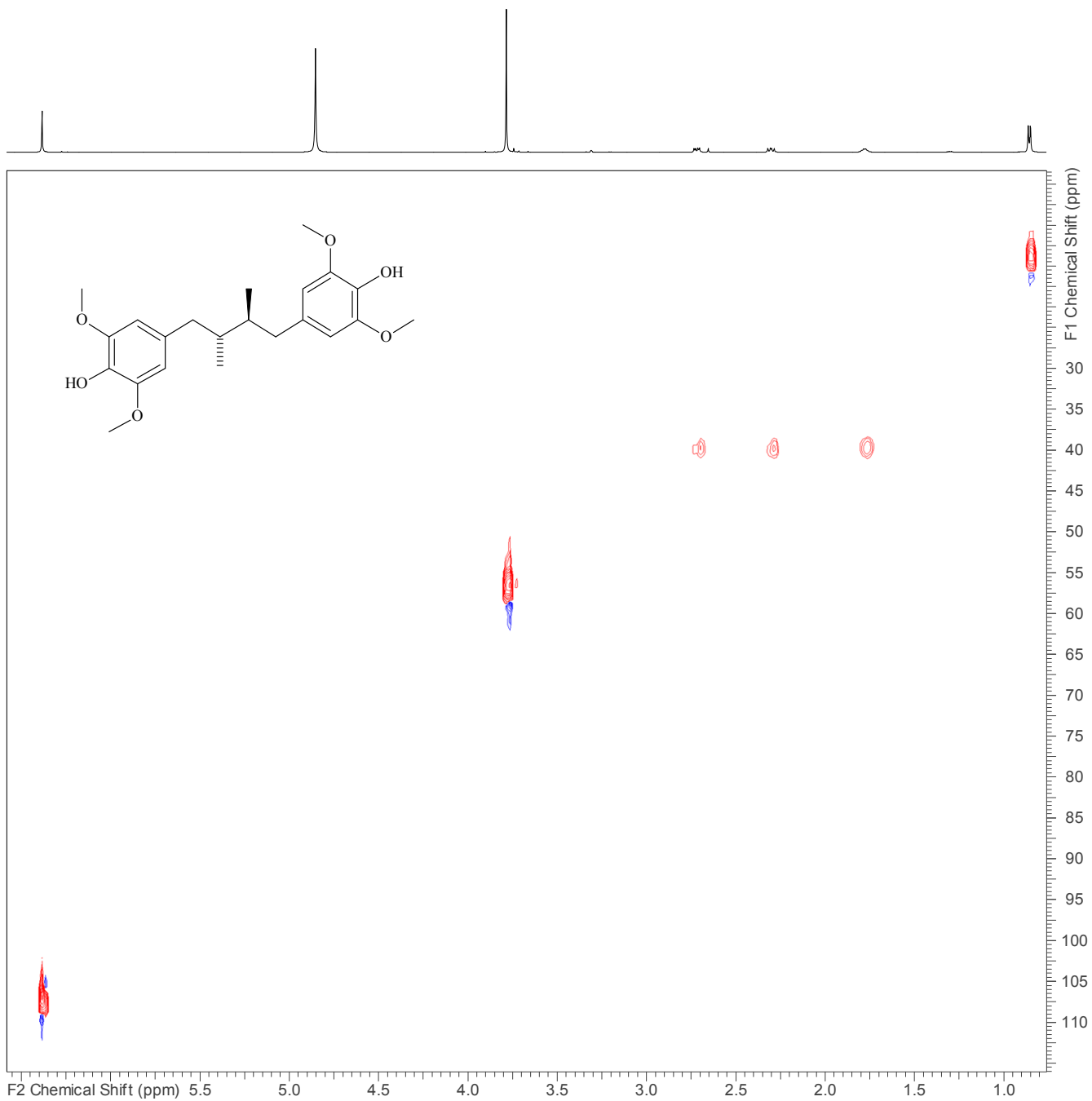


Figure S47. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **16** in CD_3OD at 600 MHz.

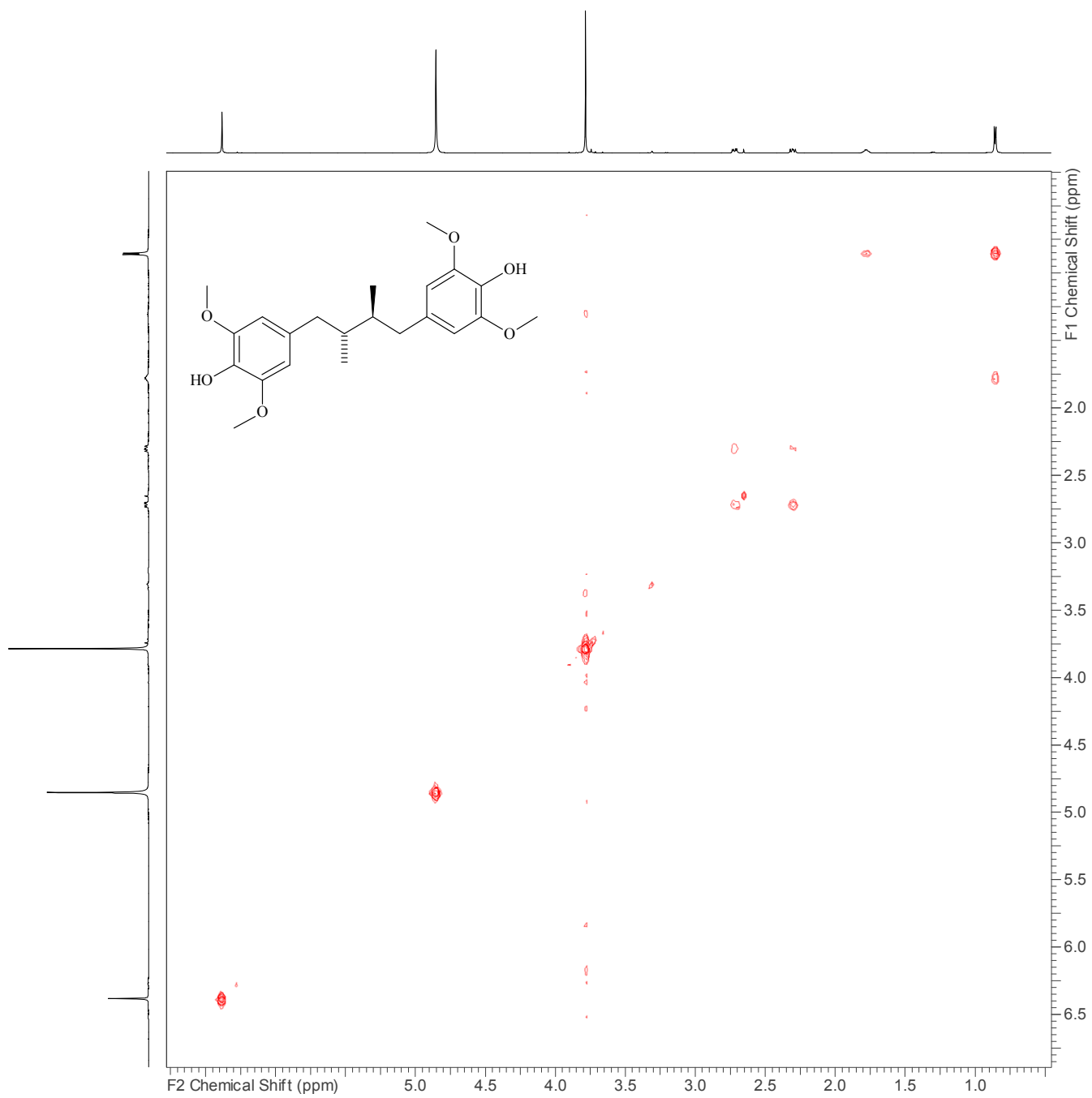


Figure S48. Expanded area of a ^1H - ^1H COSY spectrum of compound **16** in CD_3OD at 600 MHz.

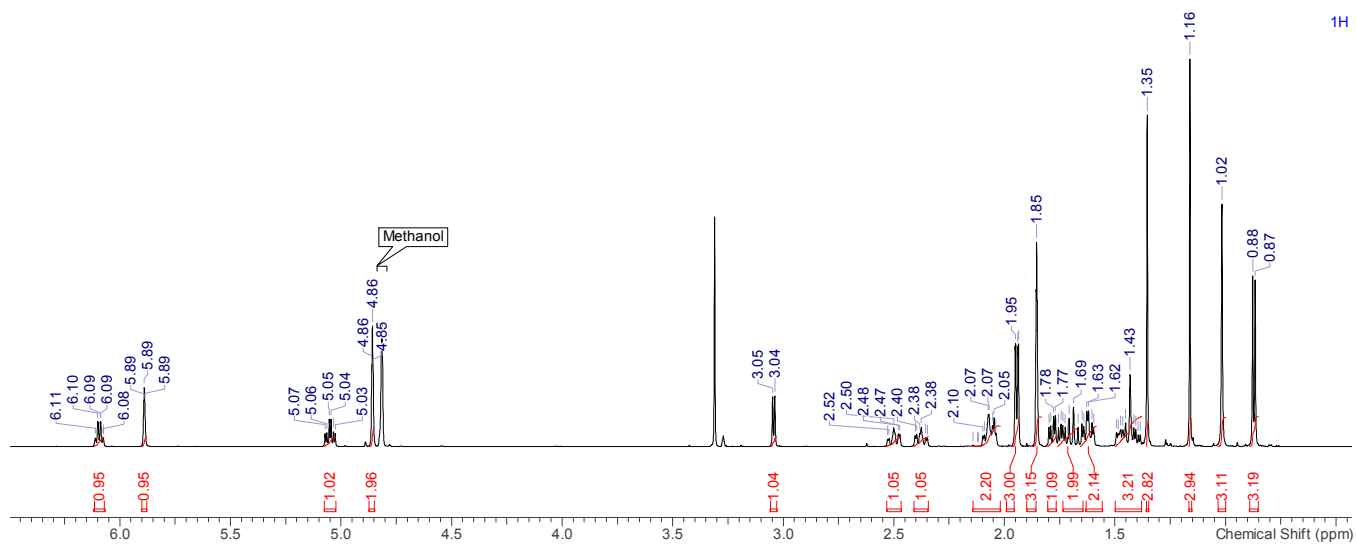
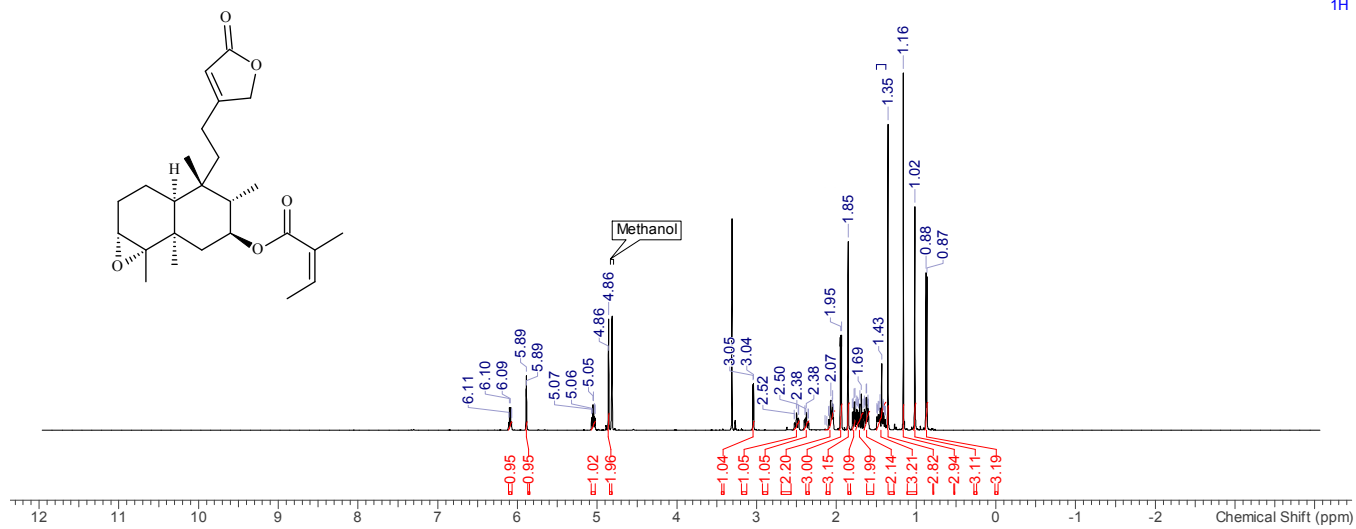


Figure S49. Proton NMR spectra (full and inset) of compound **17** in CD₃OD at 600 MHz.

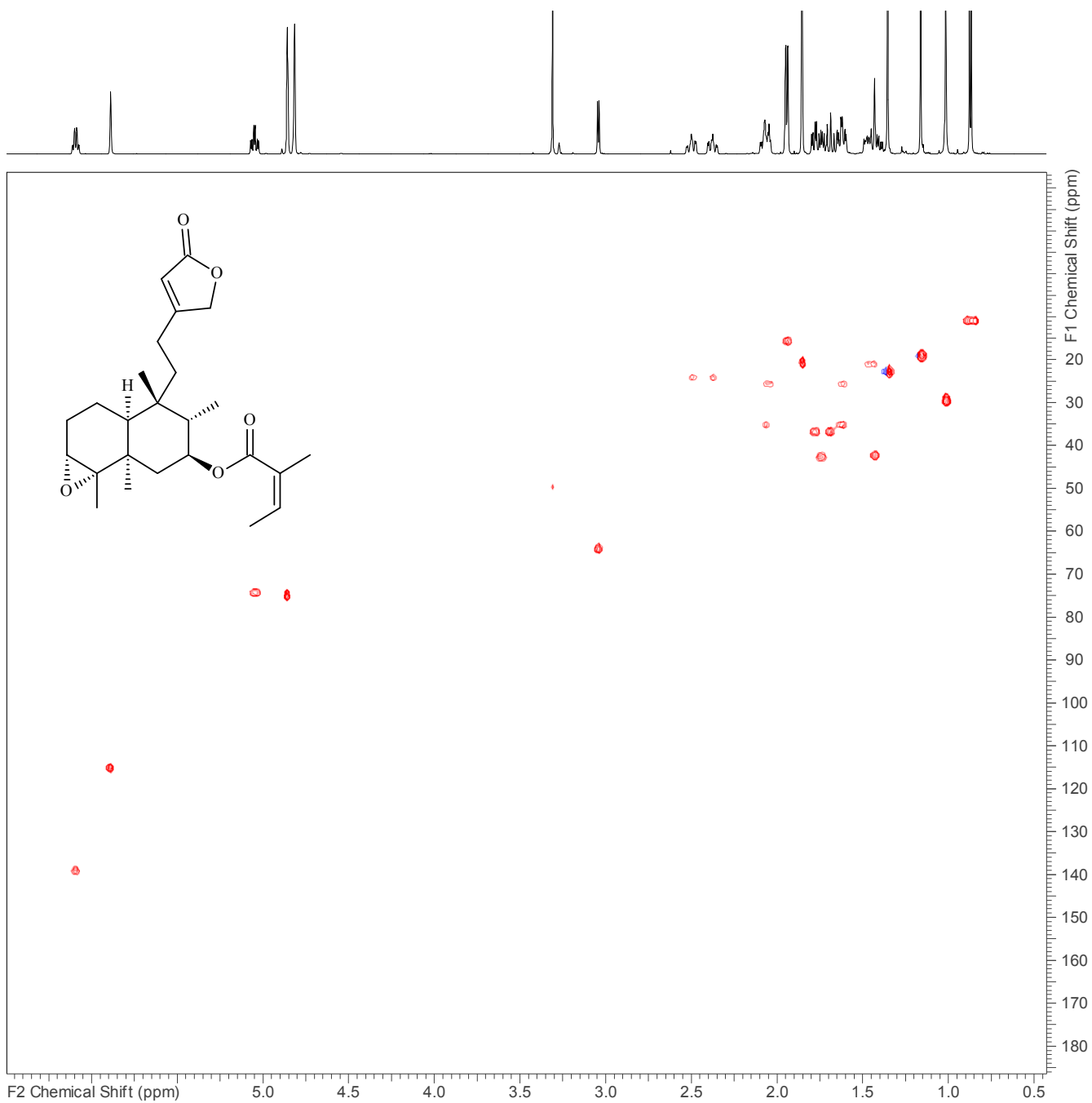


Figure S50. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **17** in CD_3OD at 600 MHz.

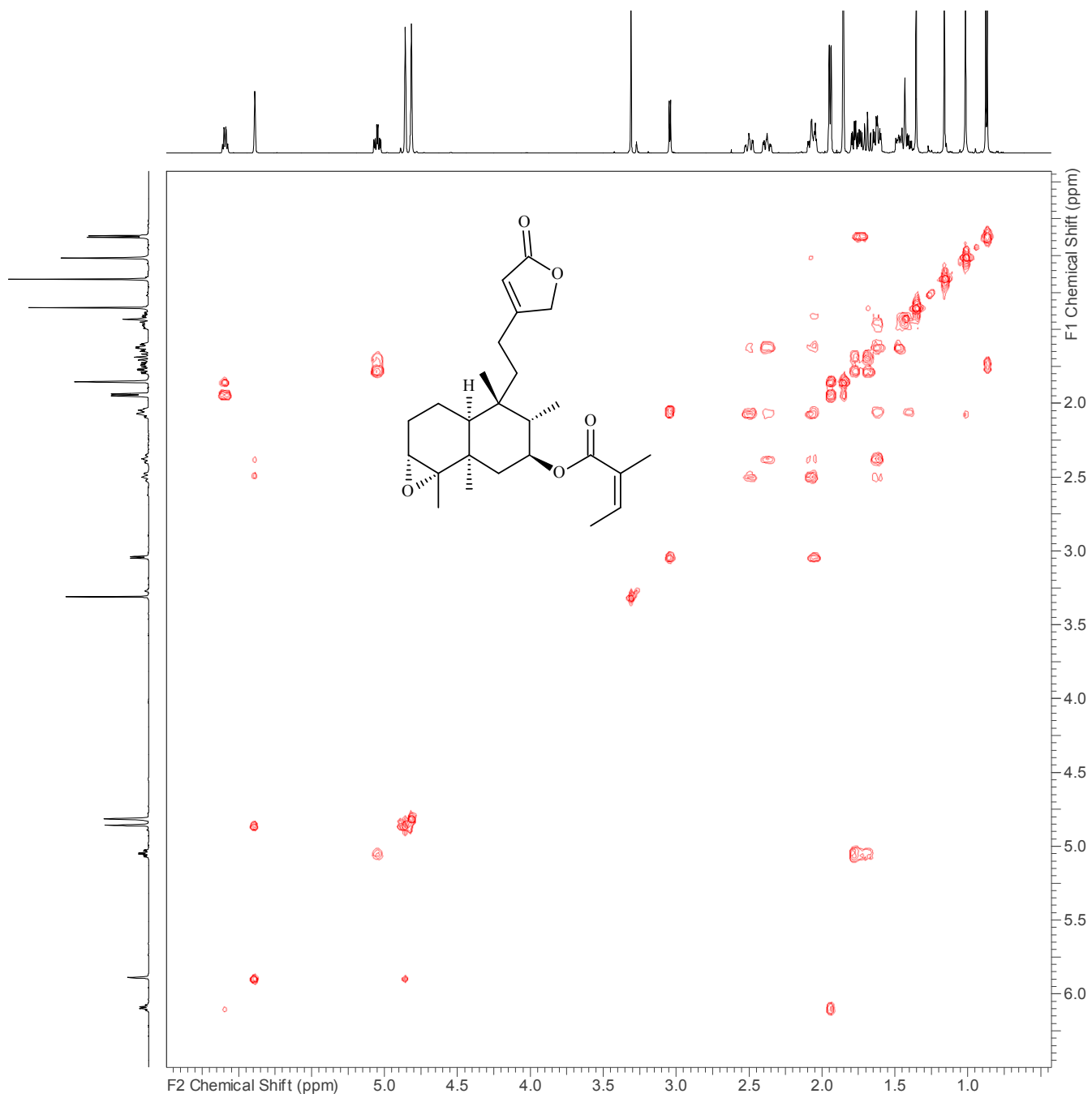


Figure S51. Expanded area of a ^1H - ^1H COSY spectrum of compound **17** in CD_3OD at 600 MHz.

1H

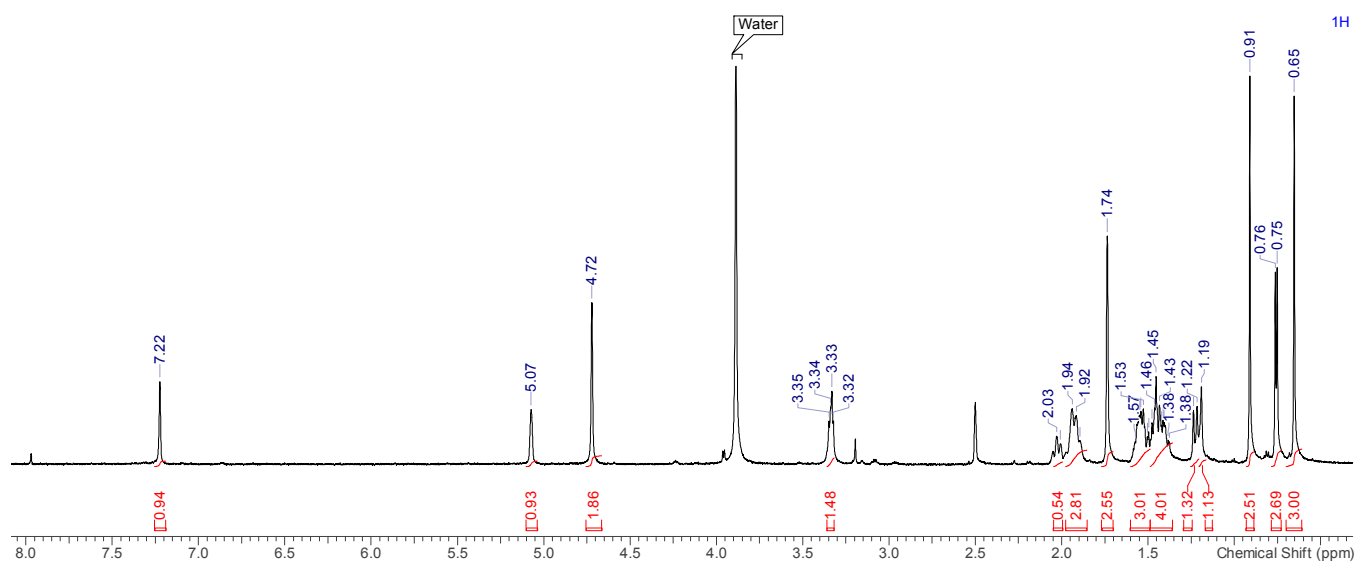
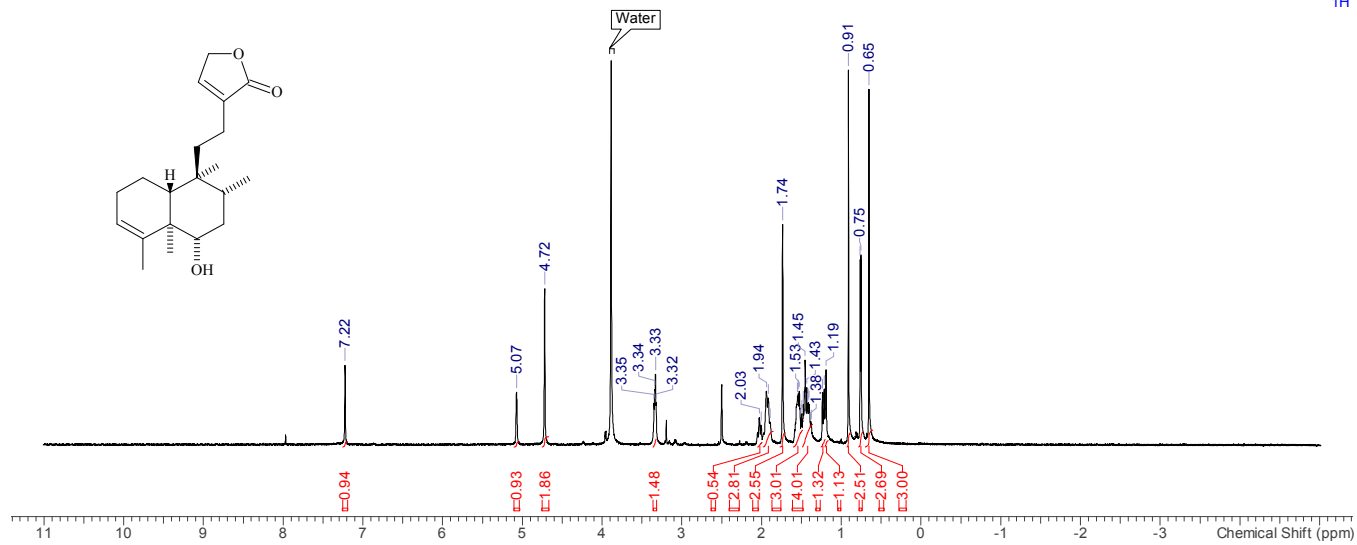


Figure S52. Proton NMR spectra (full and inset) of compound **18** in DMSO-*d*₆ at 600 MHz.

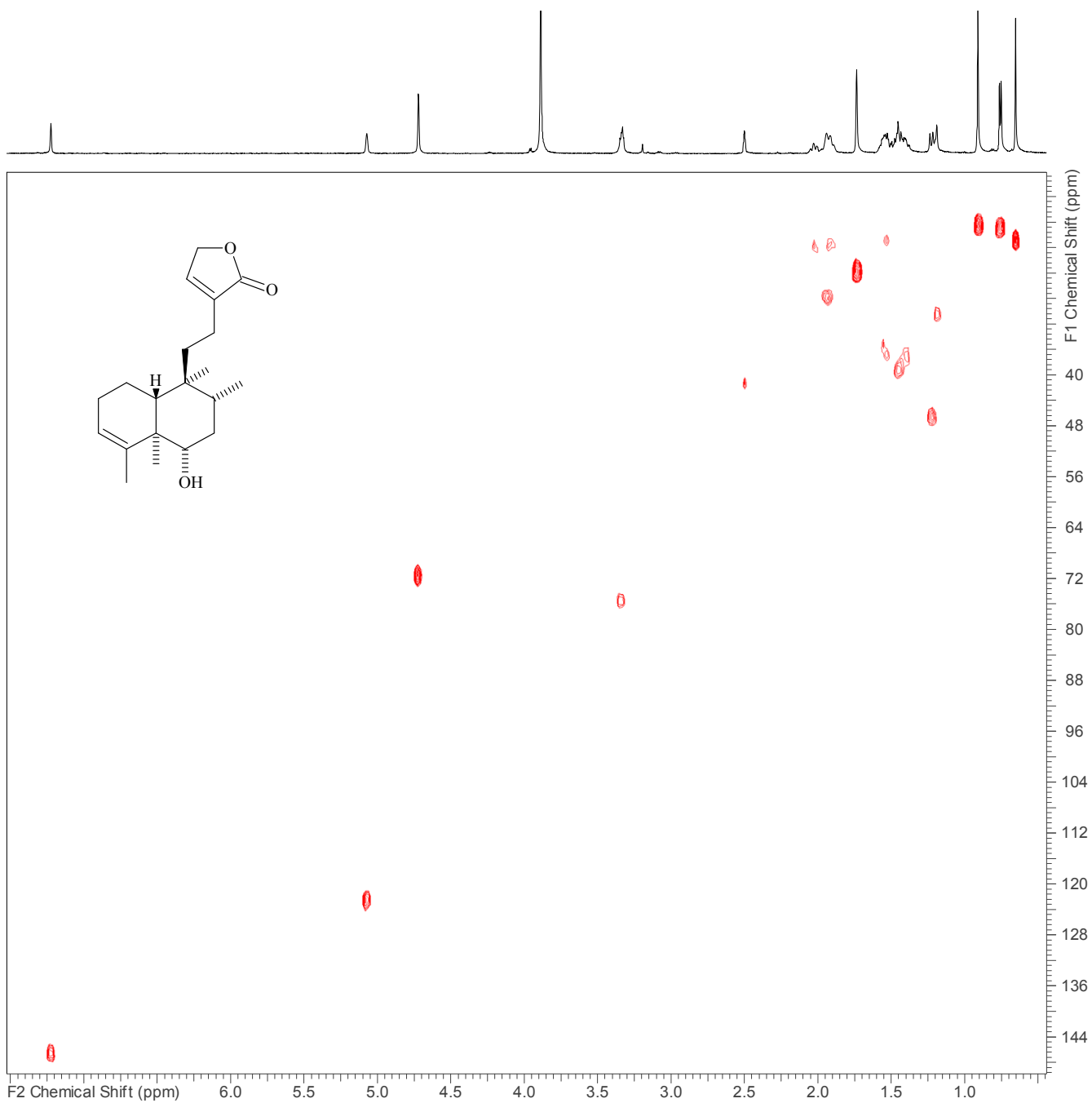


Figure S53. Expanded area of a ^1H - ^{13}C HSQC spectrum of compound **18** in $\text{DMSO-}d_6$ at 600 MHz.

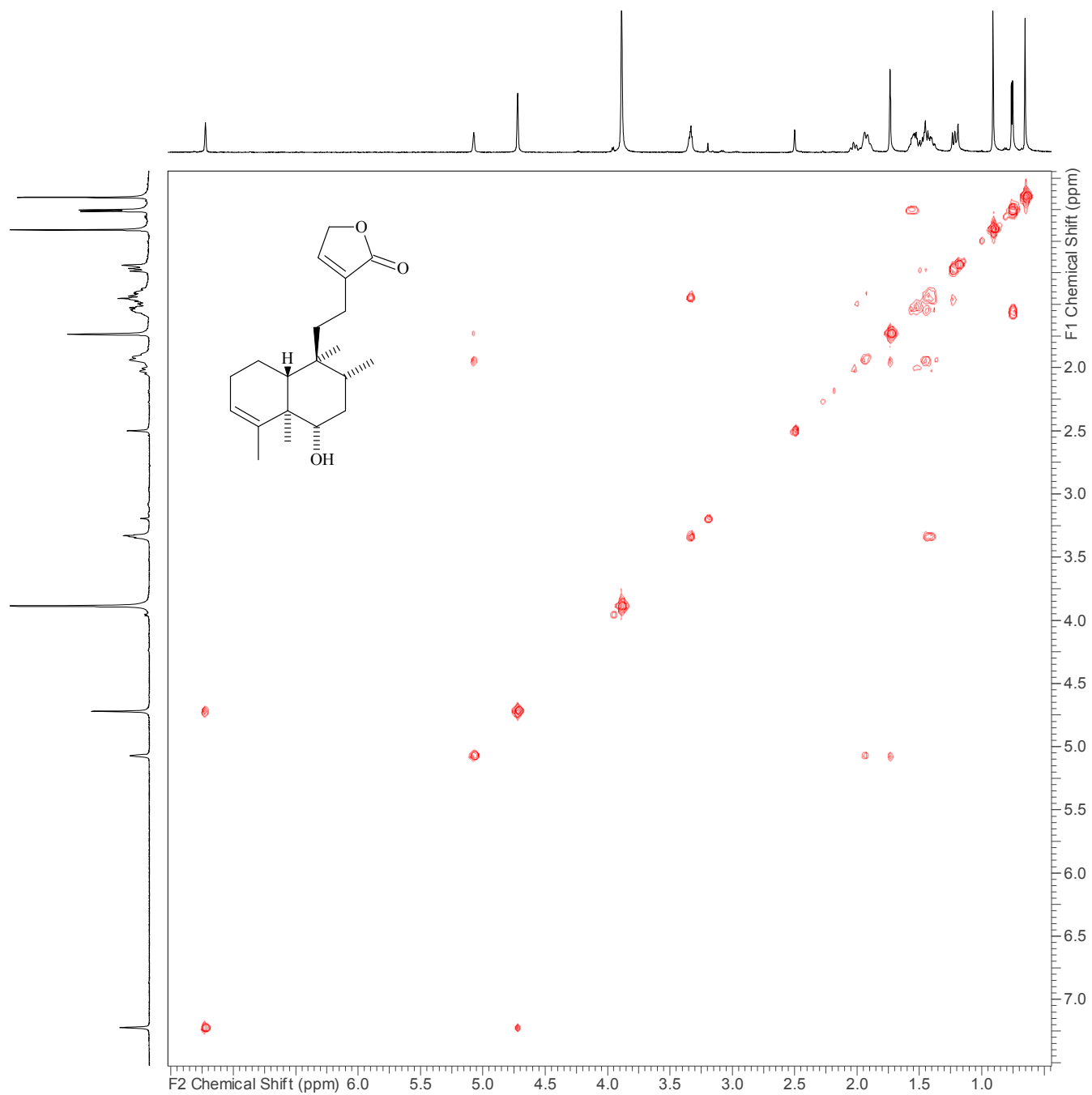


Figure S54. Expanded area of a ^1H - ^1H COSY spectrum of compound **18** in $\text{DMSO-}d_6$ at 600 MHz.

#	MF	1H+MF	1H+m/z	1H+fC	1H+substr+fC	1H+2 substr+fC	1H+IC	1H+substr+IC	1H+2 substr+IC	m/z	m/z+substr	m/z+2 substr	MF	MF+substr	MF+2 substr
1	C15H18O5	29	29	27387	135	64	22753	116	63	712	19	19	696	19	19
2	C16H24O4	9	9	14443	583	72	14151	565	71	846	83	37	822	83	37
3	C16H10O5	1	1	7655	122	102	849	22	22	227	117	1	104	84	1
4	C16H12O5	21	21	29379	94	4	26007	92	4	565	205	6	424	194	6
5	C17H12O5	1	1	47245	24	4	27983	19	4	419	111	2	253	81	2
6	C18H20O4	9	9	30421	157	5	22893	148	5	1467	19	5	1323	19	5
7	C19H24O3	10	10	438345	4740	219	300670	4392	218	975	60	10	905	60	10
8	C18H22O4	265	270	1304282	875	76	1049164	676	76	910	5	5	874	5	5
9	C19H24O4	22	22	68176	3104	80	67127	3082	80	564	39	2	509	36	2
10	C20H30O3	113	113	19607	2327	164	17095	2201	161	1260	112	23	1213	105	23
11	C18H16O6	112	112	111041	929	89	89813	684	89	725	6	5	624	6	5
12	C23H31N1O1	6	6	45838	3014	221	37827	2247	214	383	5	1	378	5	1
13	C20H21N1O4	40	40	27145	465	39	25073	420	39	3391	1254	31	3374	1253	31
14	C23H31N1O2	7	7	11872	478	15	8963	299	9	625	7	3	621	7	3
15	C22H28O6	15	15	45856	109	78	20517	97	75	481	9	1	402	9	1
16	C22H30O6	10	14	1581430	1066	19	91219	251	15	366	5	2	349	5	2

#	MF	HSQC+MF	HSQC+m/z	HSQC+fC	HSQC+substr+fC	HSQC+2 substr+fC	HSQC+IC	HSQC+substr+IC	HSQC+2 substr+IC
1	C15H18O5	15	15	1420	29	19	1413	29	19
2	C16H24O4	8	8	128	43	42	127	43	42
3	C16H10O5	4	4	7972	725	72	7598	687	72
4	C16H12O5	5	5	574	10	5	568	10	5
5	C17H12O5	5	5	4467	309	103	4426	306	103
6	C18H20O4	no HSQC available							
7	C19H24O3	242	242	261209	12327	894	248003	12245	894
8	C18H22O4	139	141	431645	1562	250	416586	1557	250
9	C19H24O4	6	6	2357	765	173	2353	765	173
10	C20H30O3	14	14	397	116	69	389	111	69
11	C18H16O6	8	8	3470	52	25	3430	52	25
12	C23H31N1O1	2	2	141	37	3	140	37	3
13	C20H21N1O4	2	2	243	15	4	242	15	4
14	C23H31N1O2	5	5	299	78	8	294	78	8
15	C22H28O6	1	1	151	106	30	140	95	29
16	C22H30O6	132	135	731876	1605	59	264297	1114	54

Figure S55. Table of ^1H and ^1H - ^{13}C HSQC NMR spectra and m/z and MF search results for NPs 1 to 16. MF = molecular formula, m/z tolerance = +/- 0.001 Da, fC = full composition (C0-100 H0-100 O0-20 N0-10), IC= limited composition (C10-30 H10-40 O0-15 N0-5) or (C10-30 H25-40 O0-15 N0-5) and substr = substructure. Note compound #14 used an HMQC.