Electronic Supplementary Material (ESI) for Green Chemistry.

Sustainable route to produce scytonemin precursor using *Escherichia coli*

Sailesh Malla[1], Morten O. A. Sommer[1,2]*

[1]*The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark*

[2]*Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Supplementary Data

Content

Protein Expression of ScyA, ScyB and ScyC.

For the analysis of *scyA*, *scyB* and *scyC* expression, the recombinant plasmids pCDF-ScyA, pACYC-ScyB and pCDF-ScyAC were transformed into *E. coli* BL21 (DE3) and protein isolation from the recombinant strains were carried out as described in materials and methods. Then the protein expression was analyzed in SDS-PAGE.

Gene name	Nucleotides (bp)	Amino acid (aa)	Protein size (kDa)
scyA	1875	625	68.7
scyB	1062	354	38.9
scyC	969	323	35.5

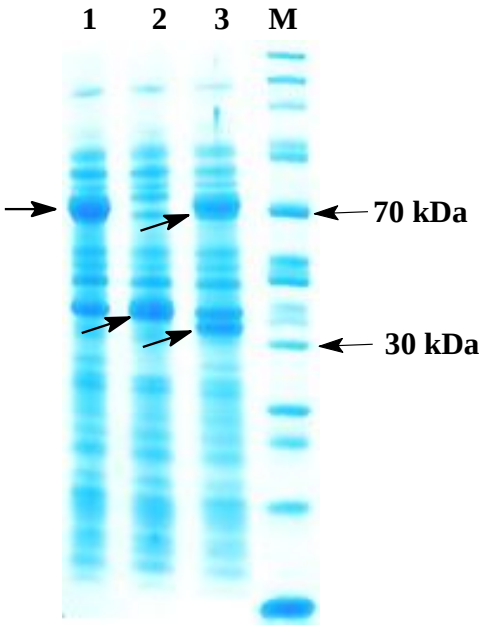


Figure S1: Protein expression. Lane 1: pCDF-ScyA, pACYC-ScyB, pCDF-ScyAC, and M: Marker.

Yellowish metabolites produced from *E. coli* SM2, SM3 and SM4 strains

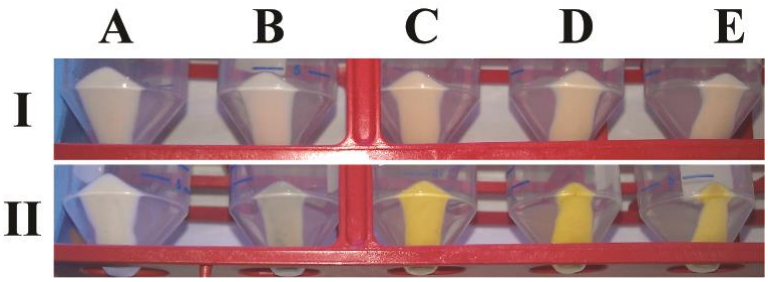


Figure S2: Cell pellets of *E. coli* recombinant strains and control strain cultured in 50 mL of M9 minimal media (I) without and (II) with supplementation of 1 mM of tryptophan and 1 mM of tyrosine at 5 days. A) *E. coli* BL21 (DE3), B) *E. coli* SM1, C) *E. coli* SM2, D) *E. coli* SM3, and E) *E. coli* SM4 strains.

UV absorbance spectra of compounds

The UV absorbance spectra of standard scytonemin, the monomer moiety of scytonemin, and other compounds were given as follows:

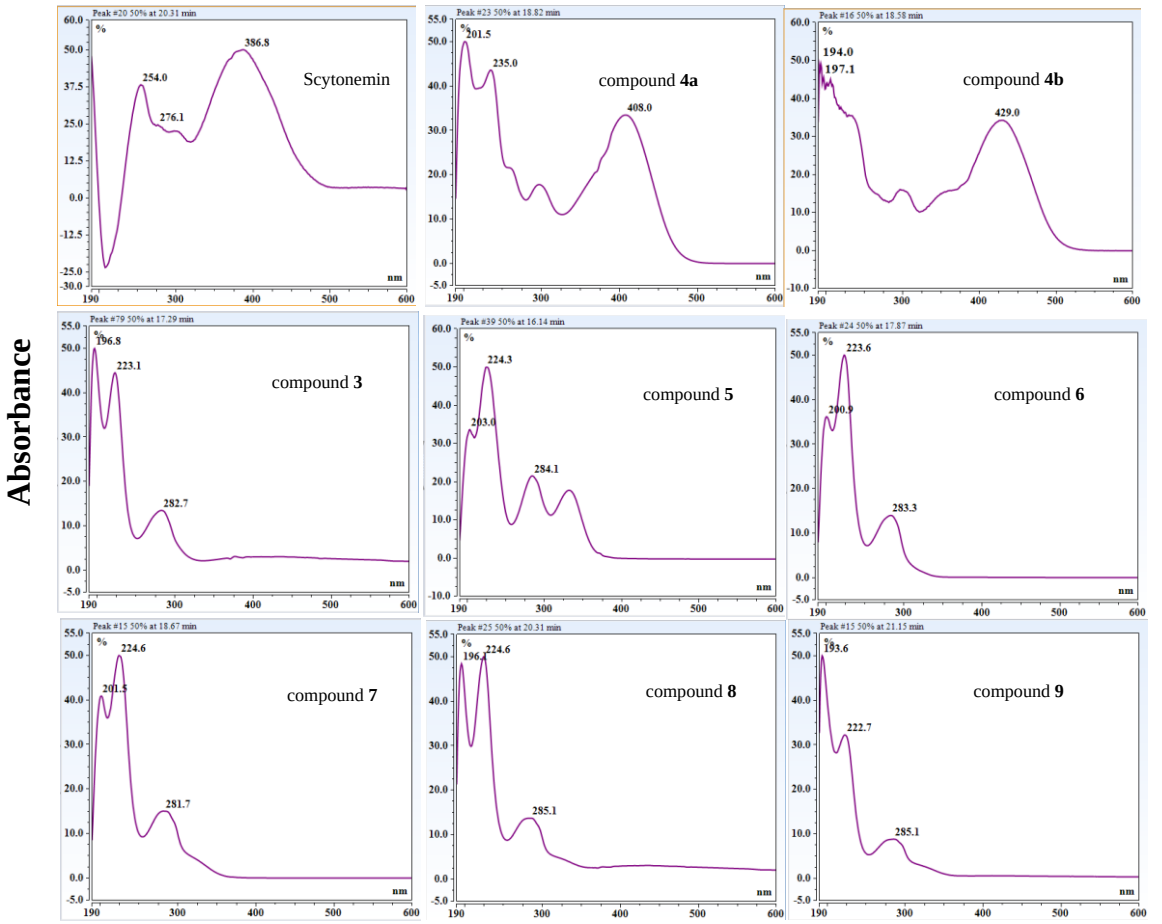


Figure S3: UV absorbance spectra of scytonemin standard, compounds **4a, 4b**, **3**, **5**, **6**, **7**, **8**, and **9**.

Analysis of biotransformation products of *E.coli* SM1

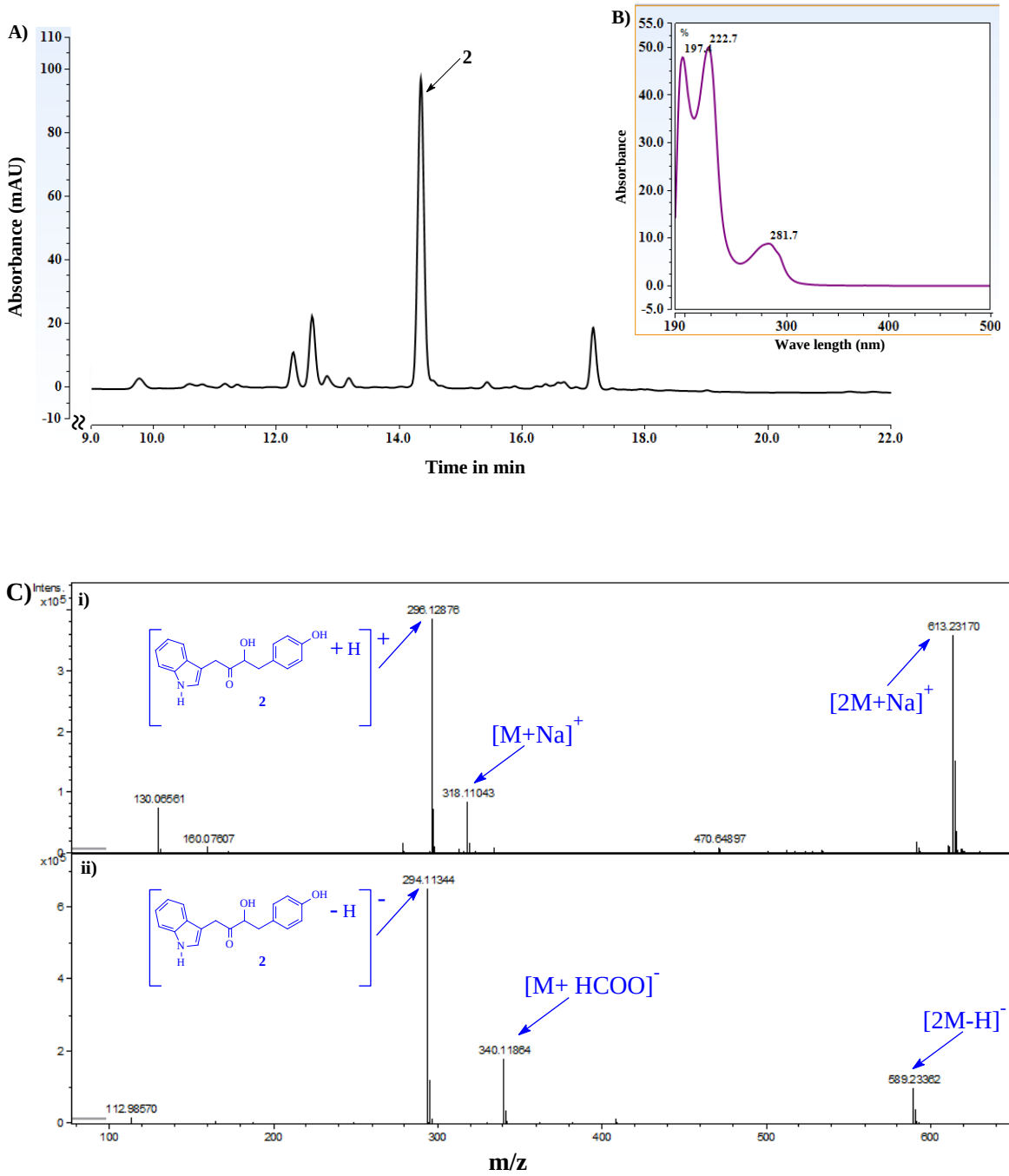


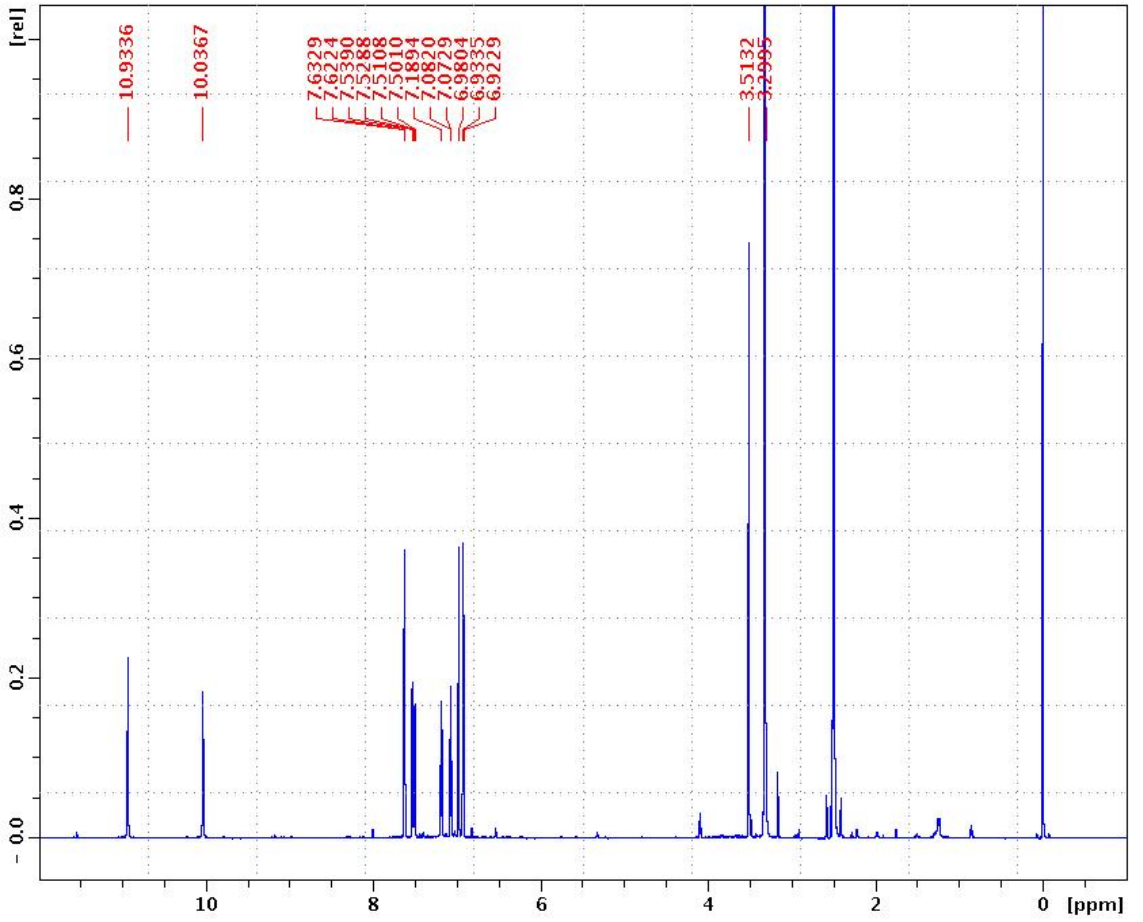
Figure S4: A) HPLC analysis of biotransformation products from *E. coli* SM1 supplemented with 1 mM of tryptophan and tyrosine. Compound **2** has retention time of 14.2 min. B) UV absorbtion spectra of compound **2** and C) Mass analysis of compound **2**.

Structural elucidation

A) NMR spectra of Compound 4

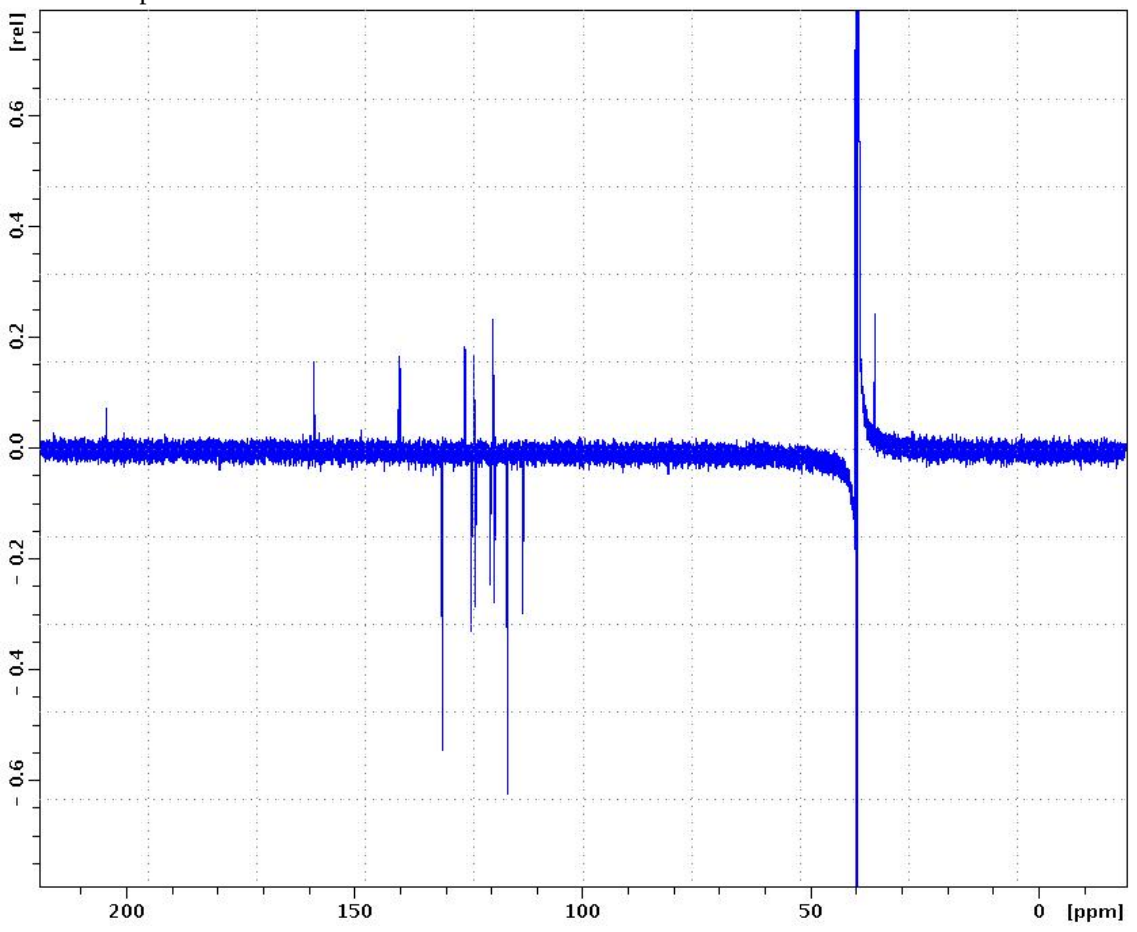
i) ^{1}H NMR spectra of 4

^{1}H NMR spectra of 4



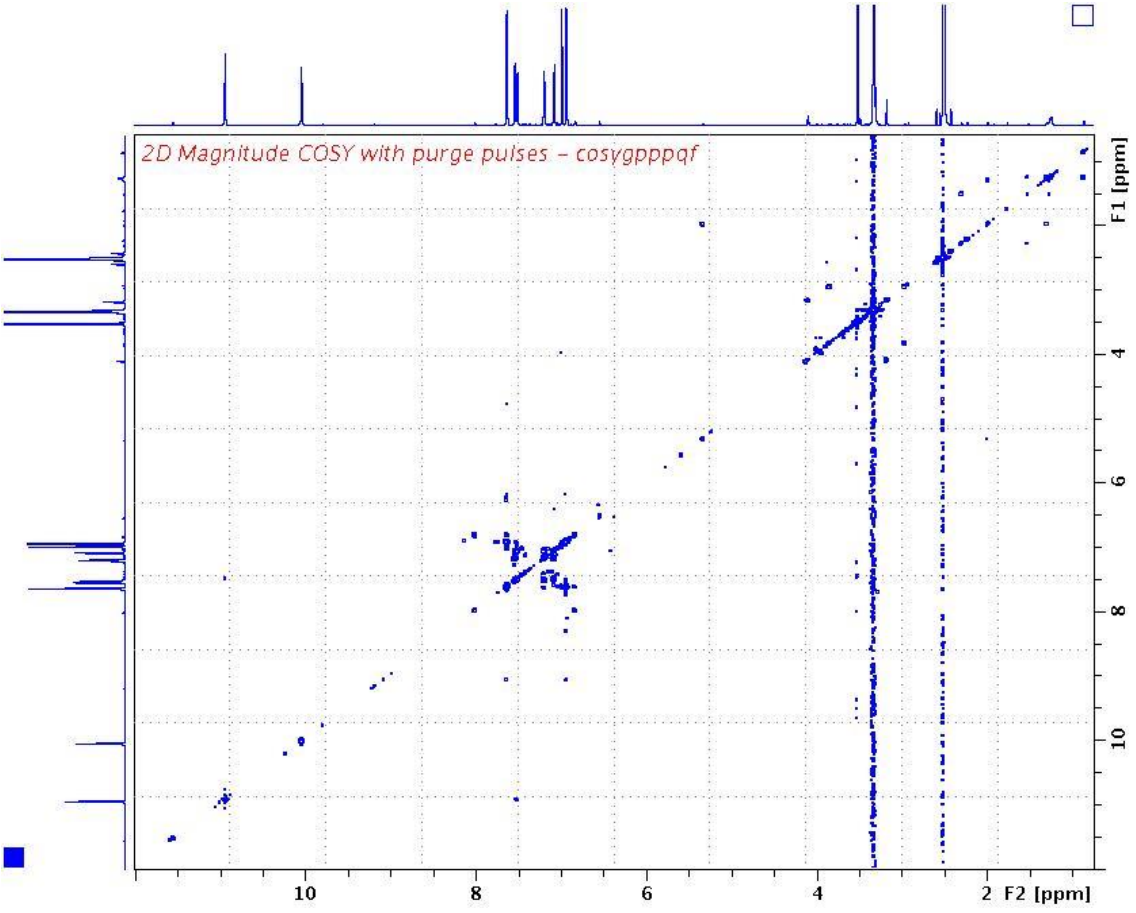
ii) ^{13}C NMR spectra of 4

^{13}C NMR spectra of 4



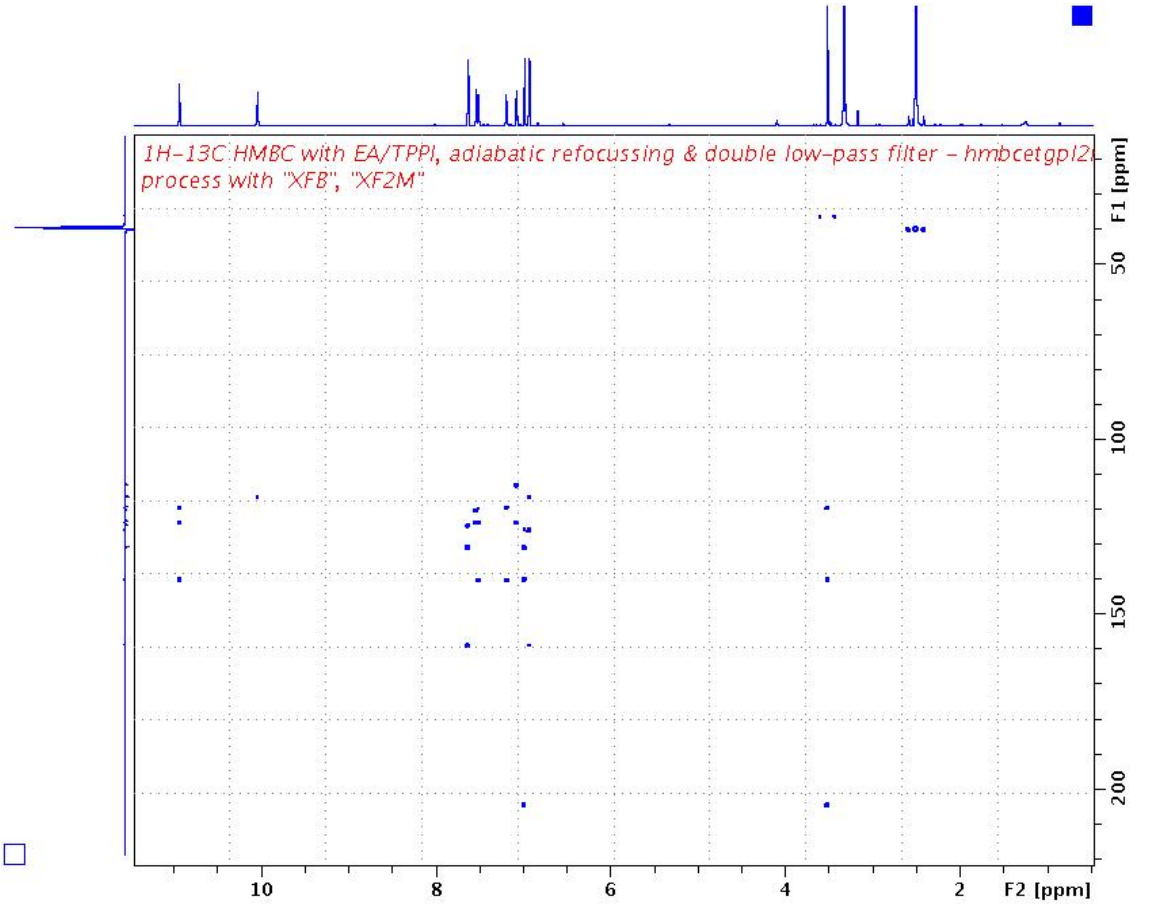
iii) COSY spectra of 4

COSY NMR spectra of 4



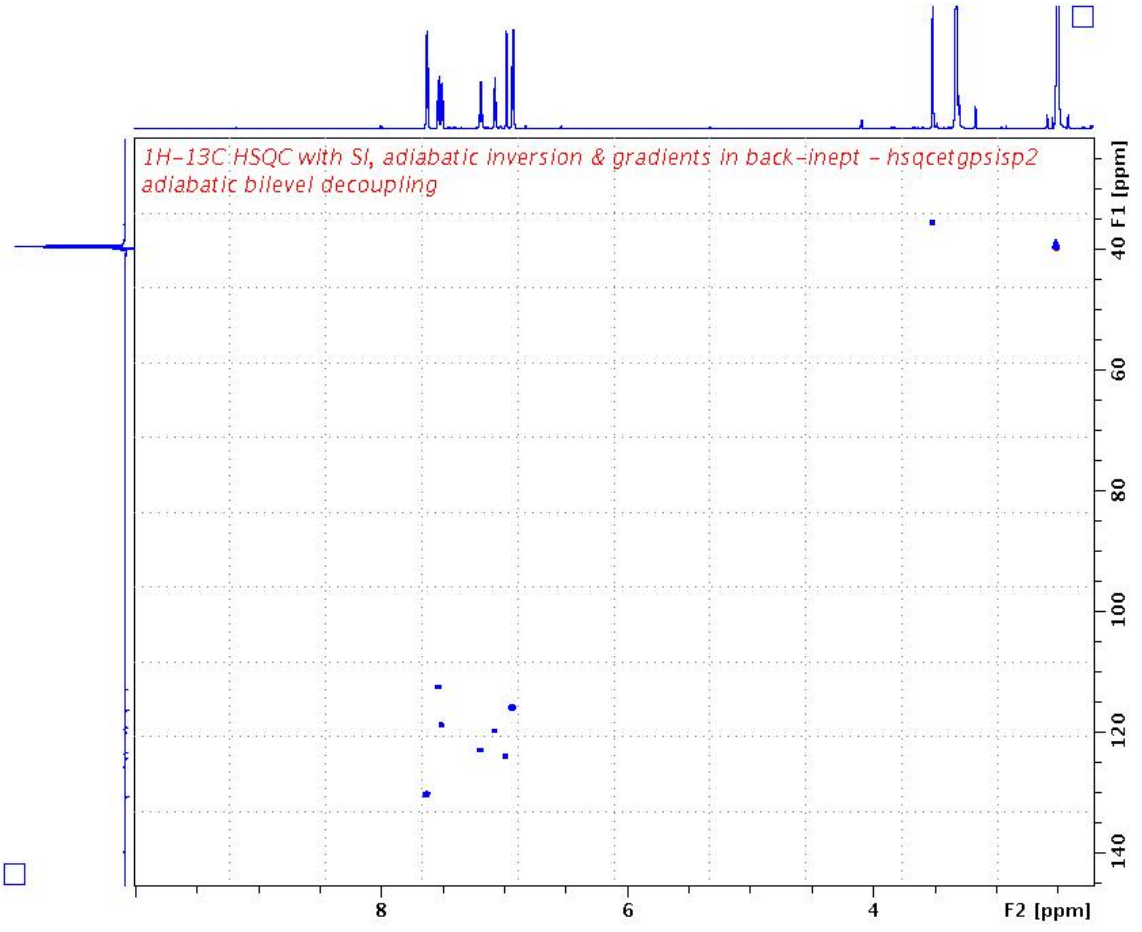
iv) HMBC NMR spectra of 4

HMBC NMR spectra of 4



v) HSQC spectra of 4

HSQC NMR spectra of 4



vi) NOE spectra of 4

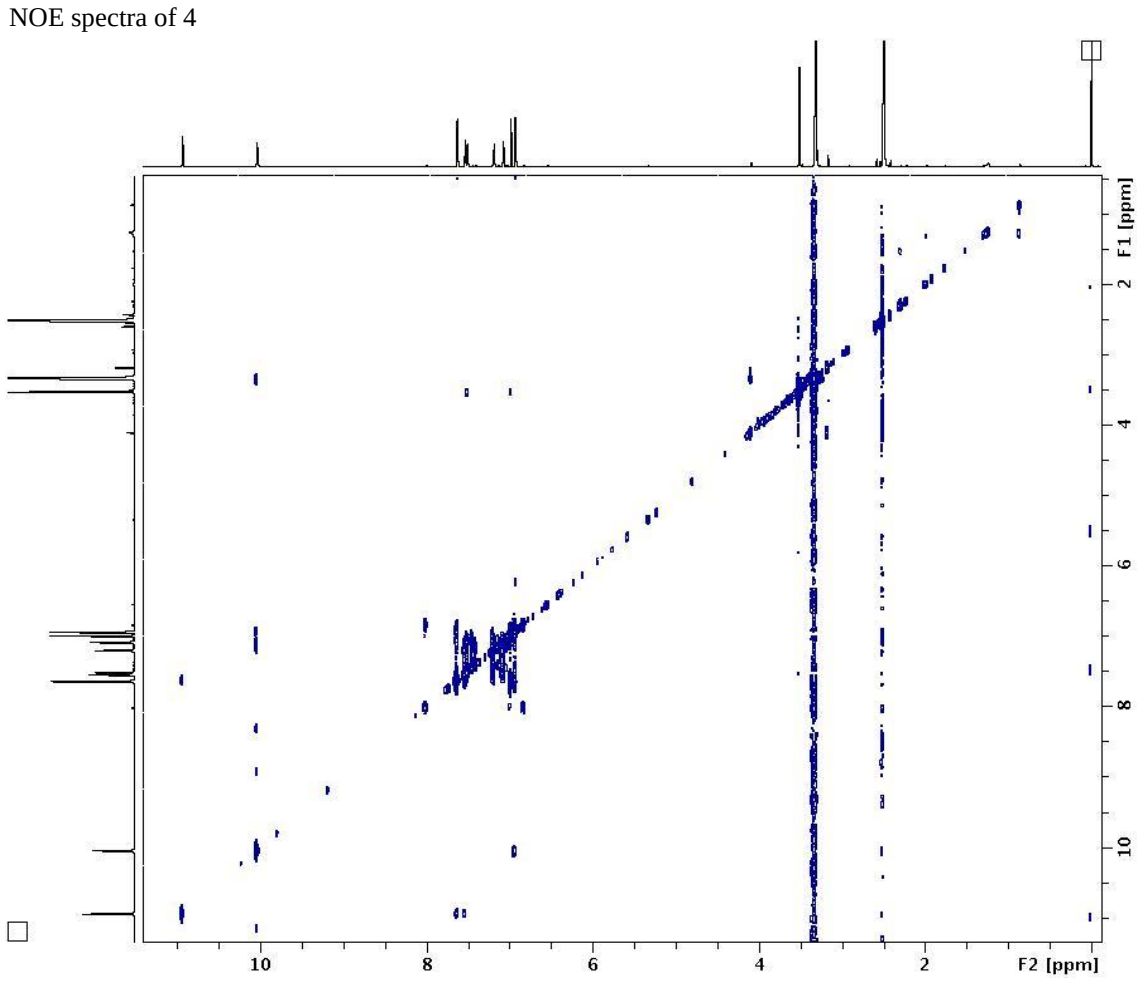
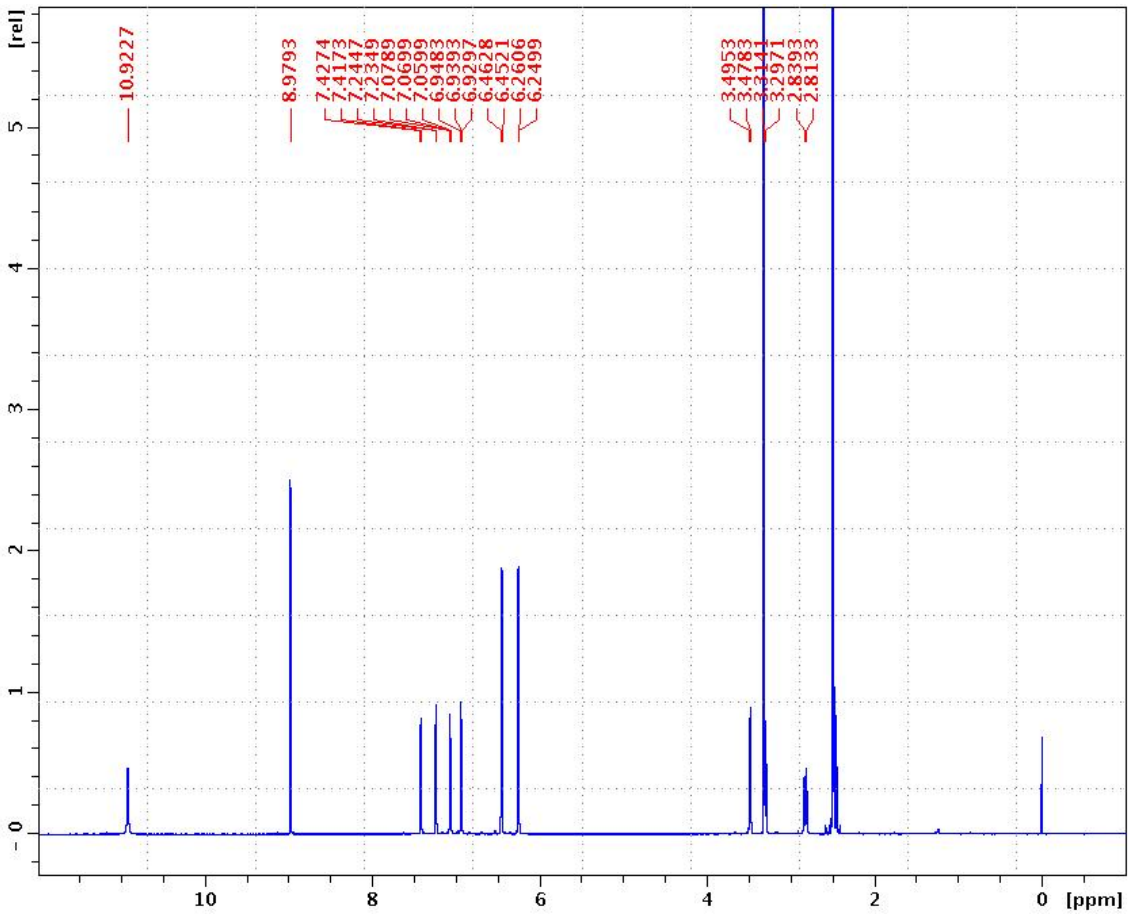


Figure S5: NMR spectra of compound **4**. i) ^{1}H-NMR, ii) ^{13}C-NMR, iii) COSY, iv) HMBC, v) HSQC and vi) NOE spectra of compound **4**.

B) NMR spectra of Compound 7

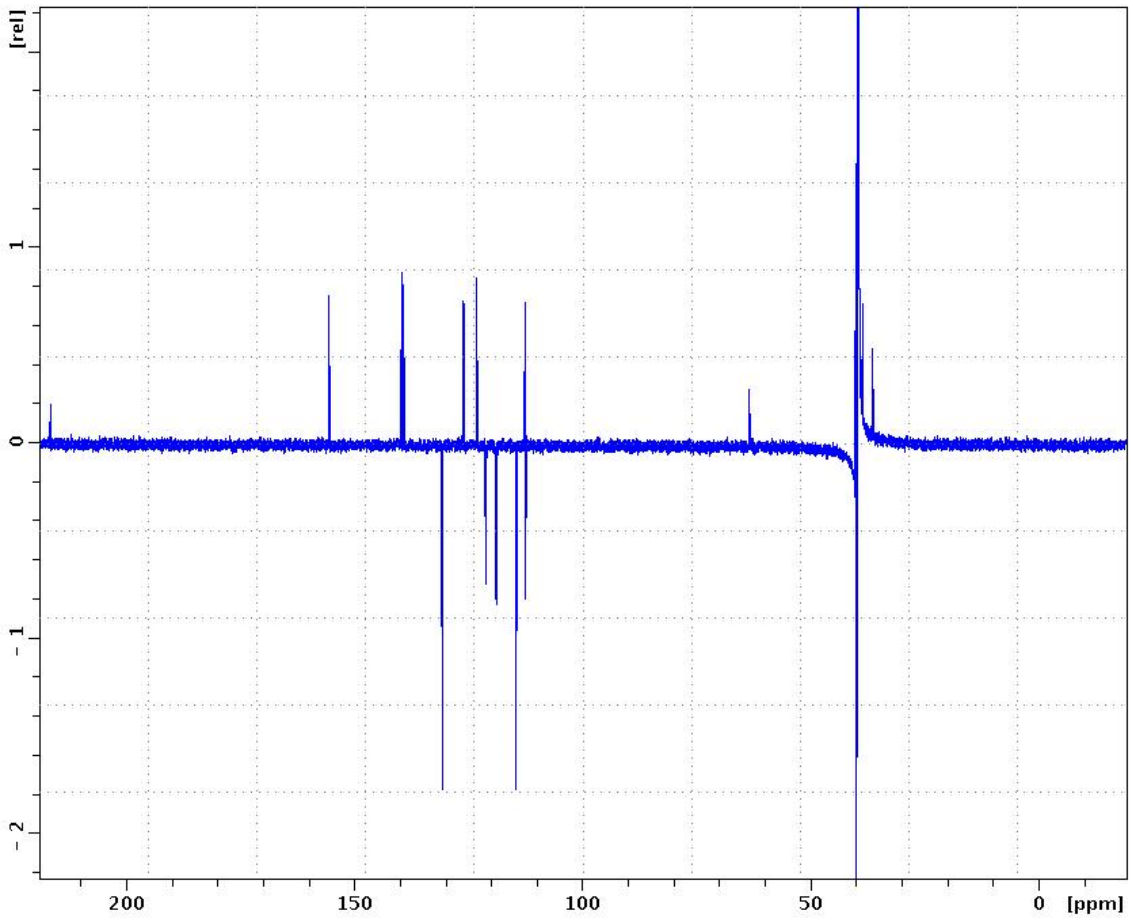
i) ^{1}H NMR spectra of 7

^{1}H NMR spectra of 7



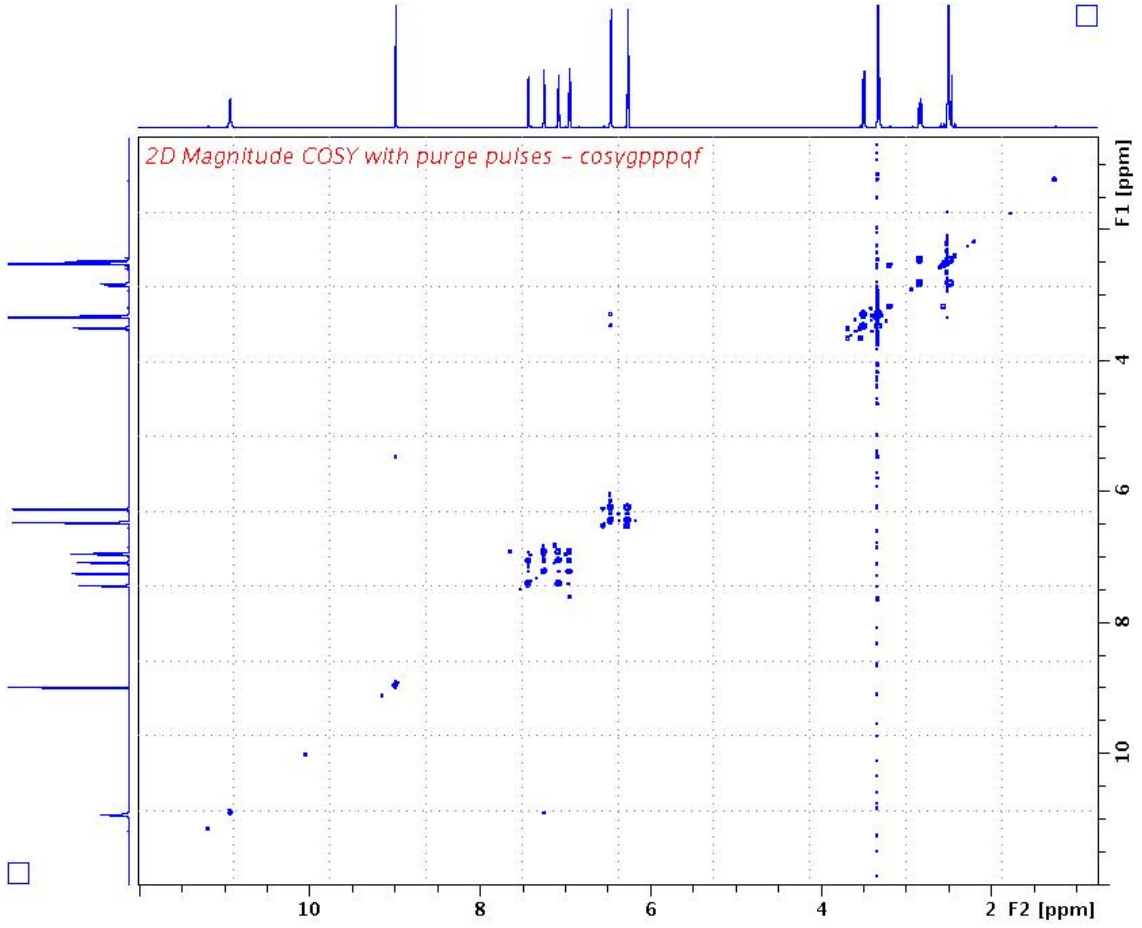
ii) ^{13}C NMR spectra of 7

^{13}C- NMR spectra of 7



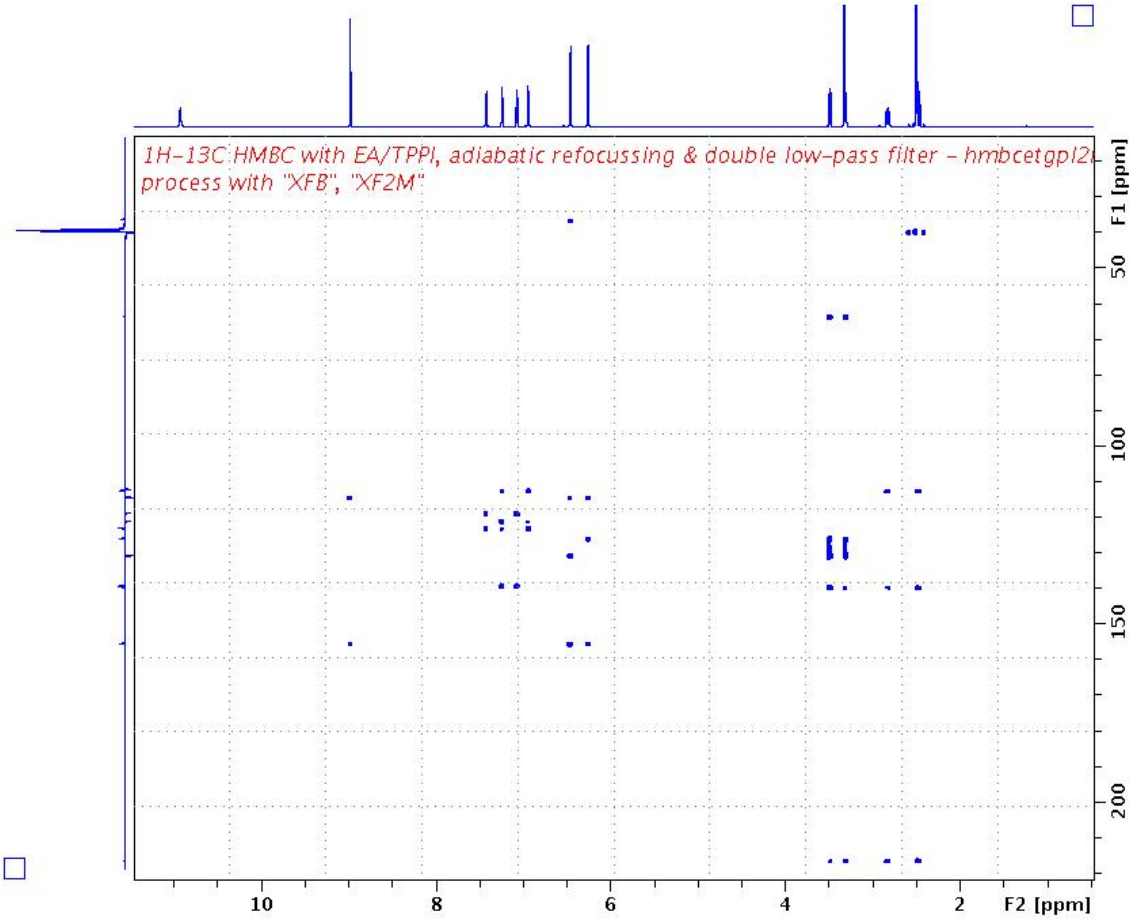
iii) COSY spectra of 7

COSY NMR spectra of 7



iv) HMBC spectra of 7

HMBC NMR spectra of 7



v) HSQC spectra of 7

HSQC NMR spectra of 7

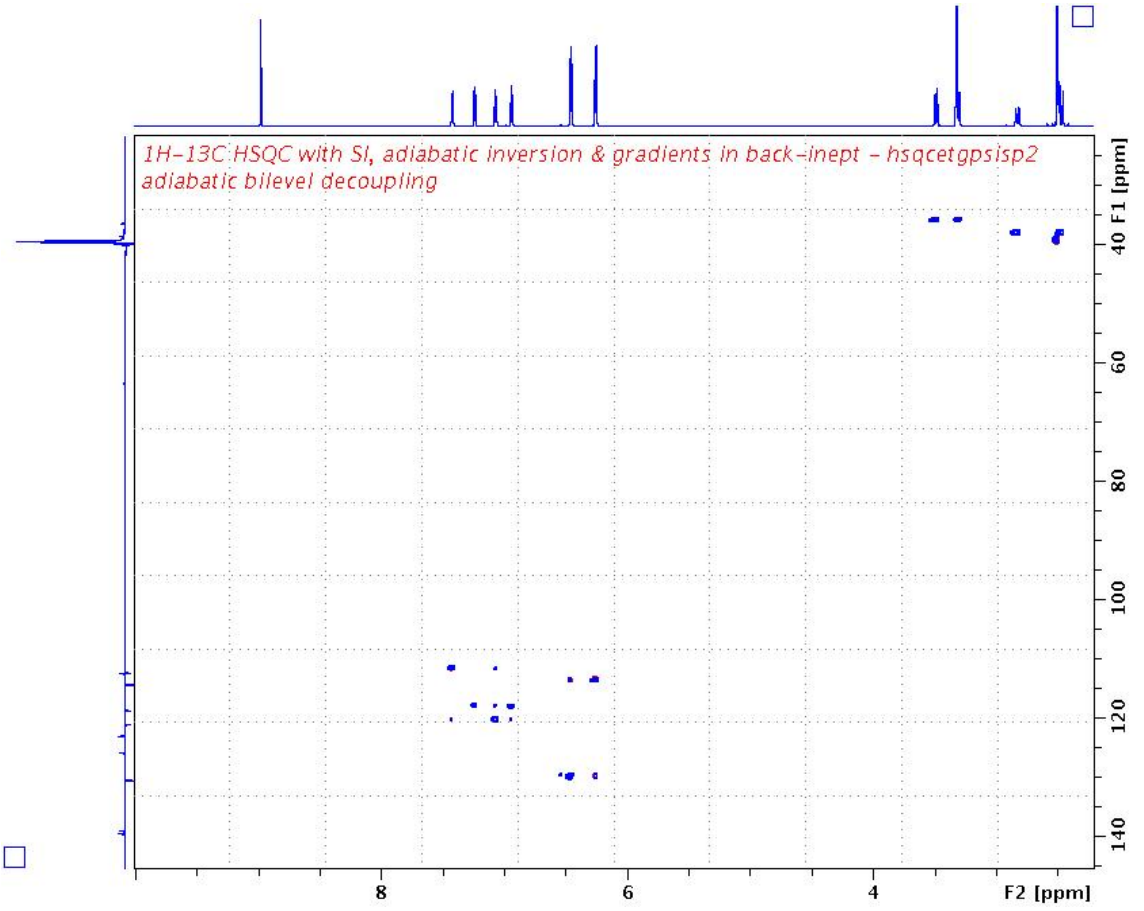


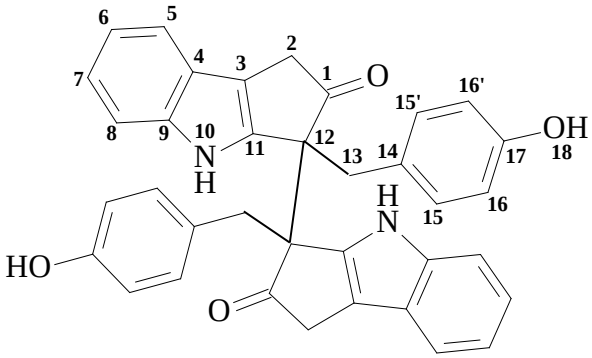
Figure S6: NMR spectra of compound **7**. i) ^{1}H-NMR, ii) ^{13}C-NMR, iii) COSY, iv) HMBC, and v) HSQC spectra of compound **7**.

Table S1. Assignment of ^{1}H and ^{13}C signals for compound **7**

	Chemical shift (ppm)	
position	^{13}C	^{1}H
1	216.36	
2	38.51	2.47, d (J = 20.88 Hz); 2.83, d (J = 20.88 Hz)
3*	112.30	
4*	139.0	
5	118.60	7.24, d (J = 7.97 Hz)
6	118.70	6.94, t (J = 7.42 Hz)
7	121.00	7.07, t (J = 7.42 Hz)
8	112.30	7.42, d (J = 8.24 Hz)
9*	122.98	
10		10.92, s
11*	139.50	
12	63.36	
13	36.34	3.30, d (J = 13.46 Hz); 3.48, d (J = 13.46 Hz)
14	125.87	
15	130.55	6.45, d (J = 8.79 Hz)
16	114.31	6.25, d (J = 8.79 Hz)
17	155.34	
18		8.98, s

*assignments of carbon signals were assisted by chemical shift estimation using chemdraw software.

The ^{13}C spectrum confirmed the presence of a symmetric molecule, since only 16 carbon signals is present in a 36 carbon molecule (16 carbon signal plus 2 carbon signal from the symmetric phenyl ring times 2 gives 36 carbon). The indole and phenyl structures are readily apparent, since typical proton signals multiplicities and COSY correlations are present. A ketone group resonating at 216.36 ppm is also present. The HSQC spectrum detects one aliphatic quaternary carbon and two aliphatic CH_2 groups (one attached to the phenyl group and another one in the ring). The only possible explanation is that molecule **3** dimerized via the carbon attached to the phenyl group as depicted in Figure S6.1. All the acquired NMR data is consistent with the structure depicted in Figure S6.1 and all of the carbon and proton signals were assigned in Table S1.



104

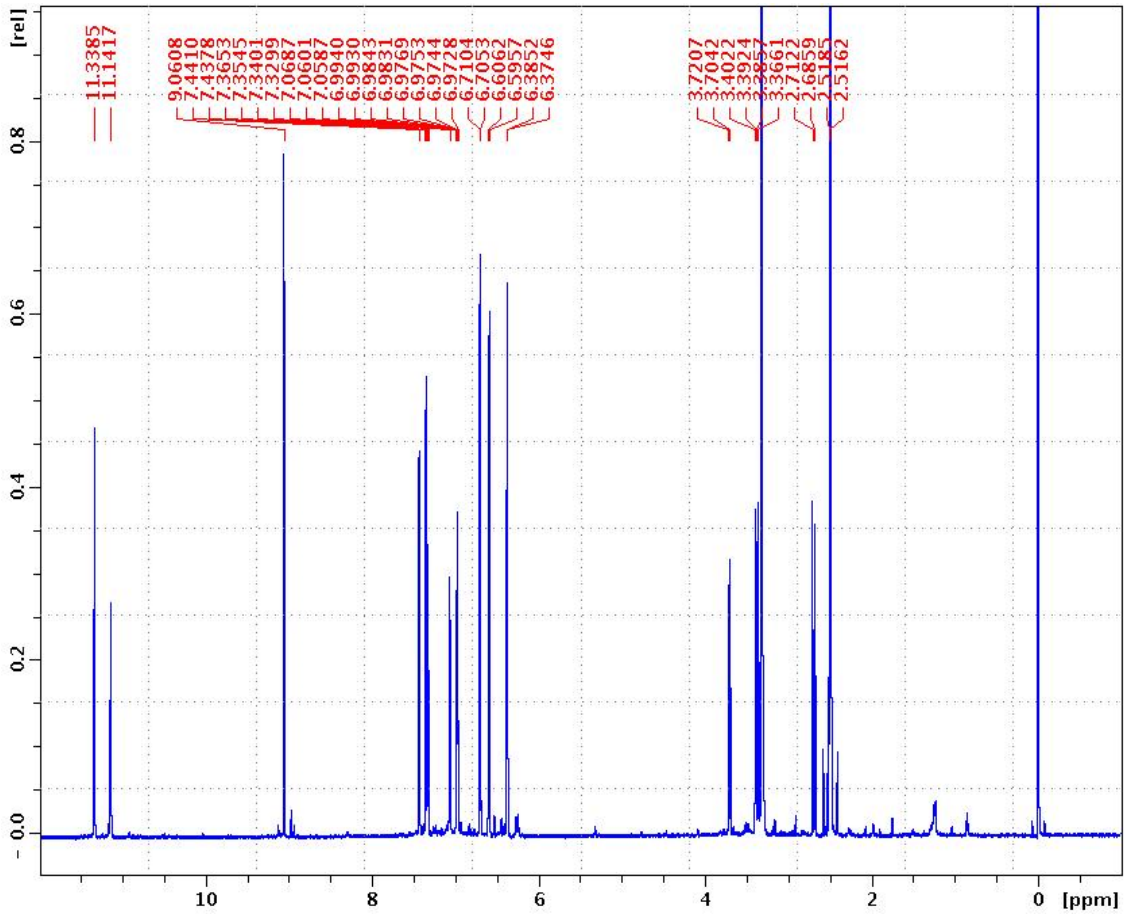
105 **Figure S6.1.** Structure of the compound **7** with atom numbering

106

C) NMR spectra of Compound 6

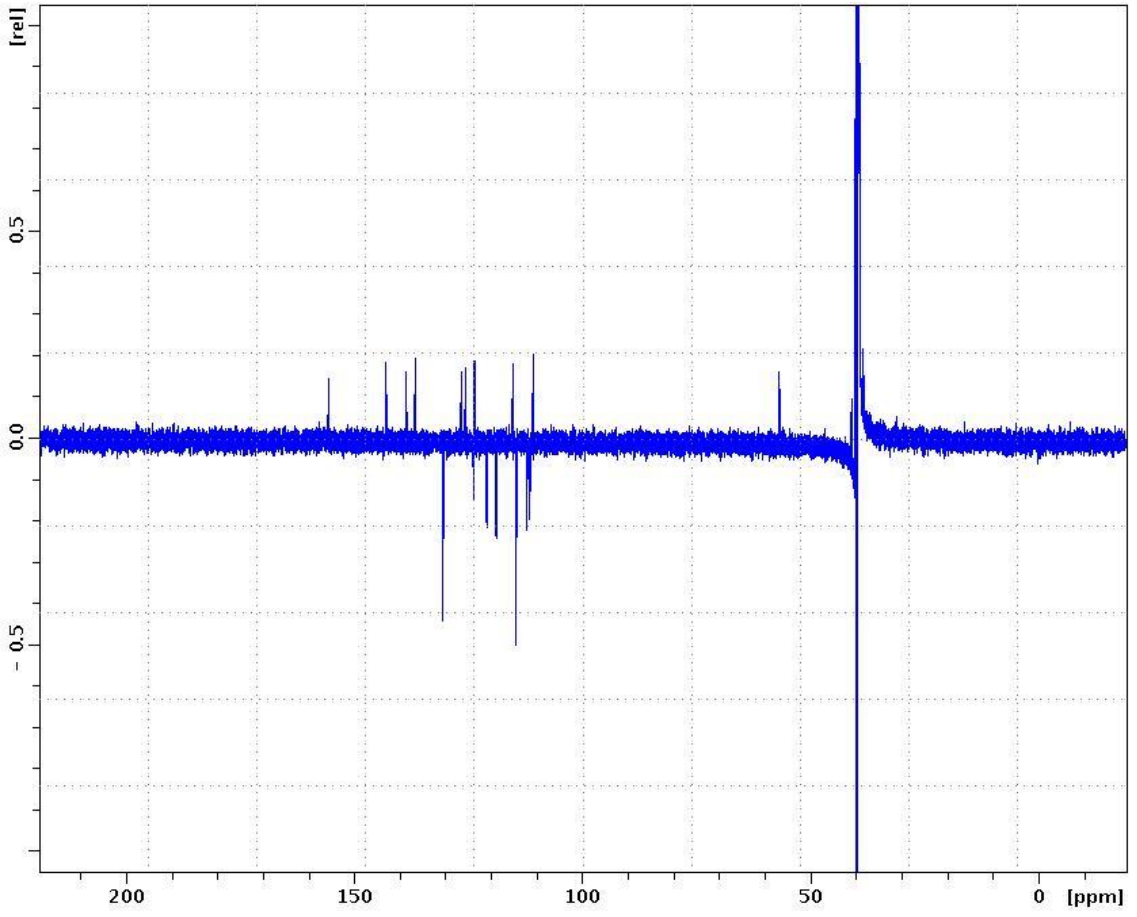
i) ^{1}H NMR spectra of compound 6.

^{1}H-NMR spectra of 6



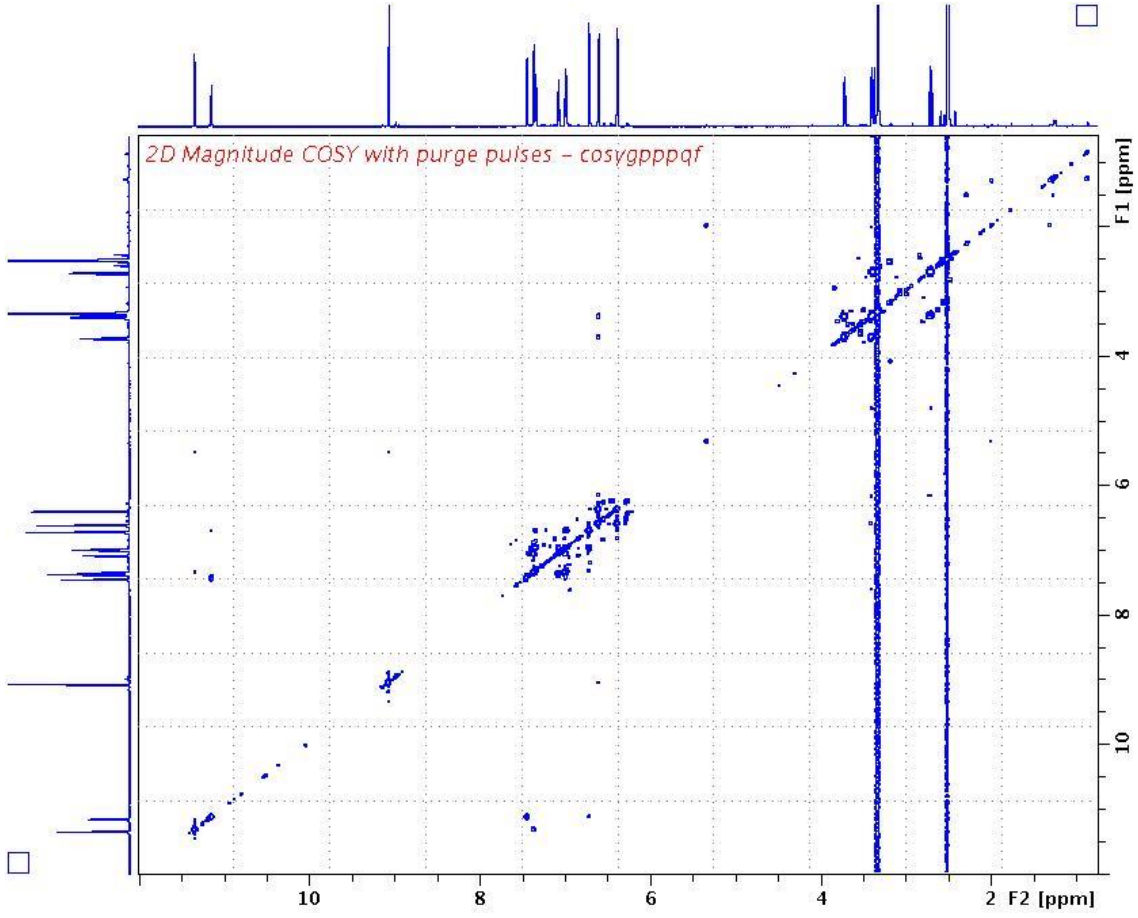
ii) ^{13}C-NMR spectra of compound 6.

^{13}C-NMR spectra of 6



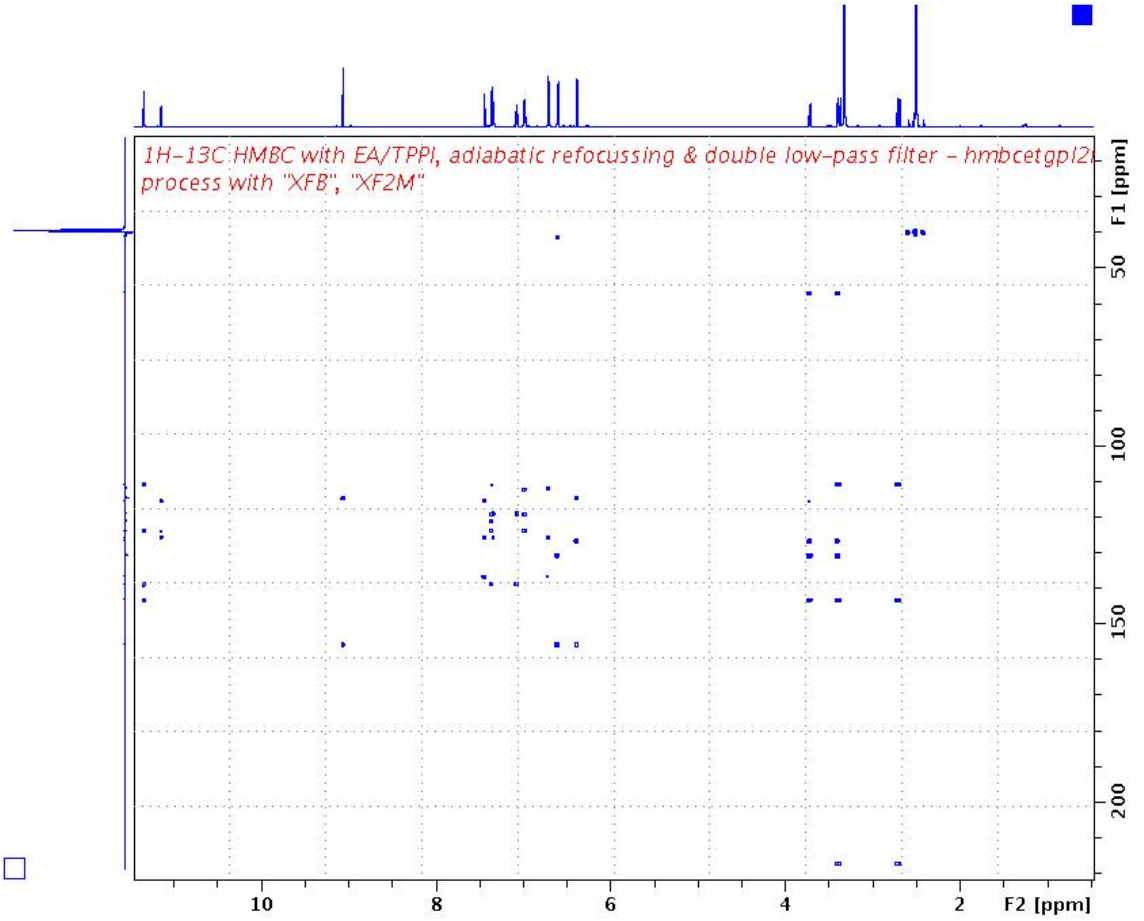
iii) COSY, spectra of compound 6.

COSY NMR spectra of 6



iv) HMBC spectra of compound 6.

HMBC NMR spectra of 6



v) HSQC spectra of compound 6.

HSQC NMR spectra of 6

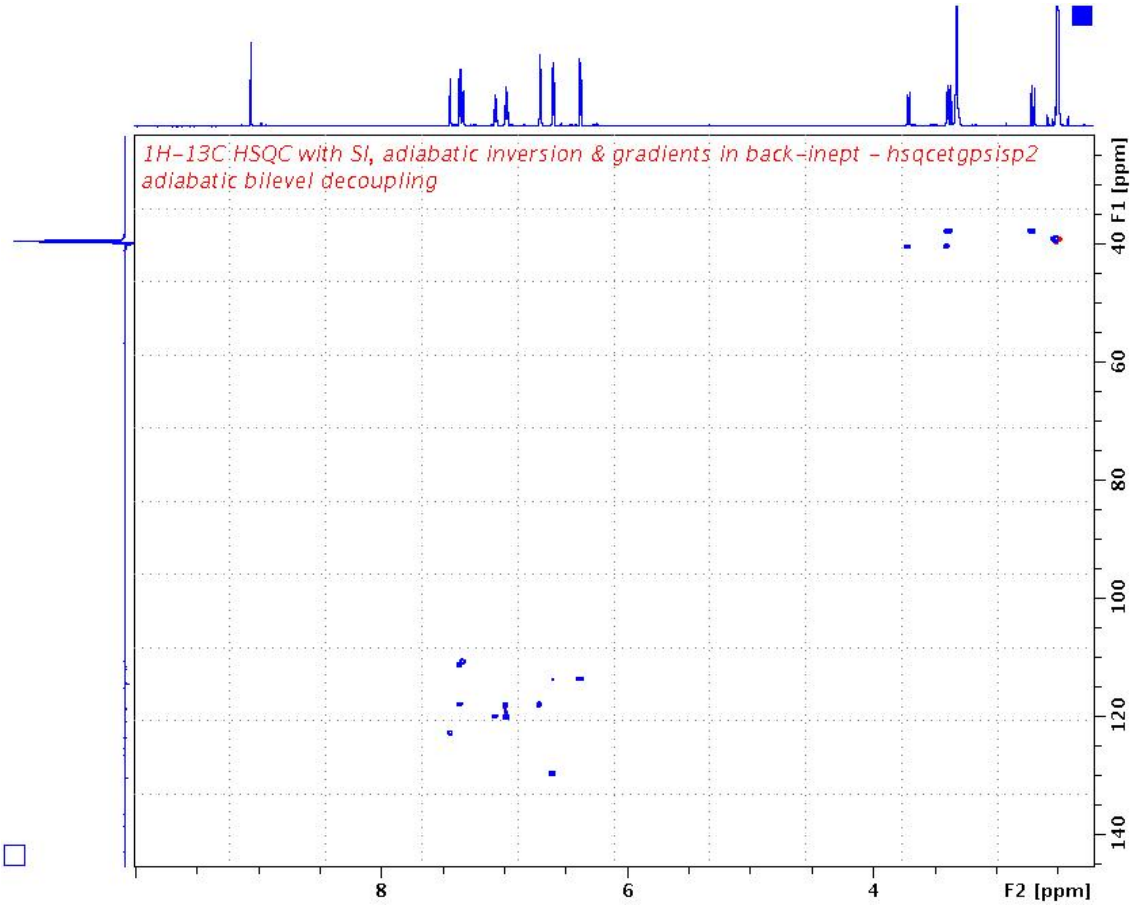


Figure S7: NMR spectra of compound **6**. i) ^{1}H-NMR, ii) ^{13}C-NMR, iii) COSY, iv) HMBC, and v) HSQC spectra of compound **6**.

Table S2. Partial assignment of ^{1}H and ^{13}C signals for compound **6**

	Chemical shift (ppm)	
position	^{13}C	^{1}H
1	217.10	
2	38.33	3.38, d (J = 20.81 Hz); 2.70, d (J = 20.81 Hz)
3	na	
4	na	
5	118.70	7.36, d (J = 8.21 Hz)
6	118.82	6.98, t (J =7.86 Hz)
7	120.76	7.07, t (J = 7.69 Hz)
8	118.70	7.36, d (J = 8.21 Hz)
9	na	
10		11.34, s
11	na	
12	56.62	
13	40.89	3.71, d (J = 13.32 Hz); 3.39, d (J = 13.32 Hz)
14	126.40	
15	130.39	6.60, d (J = 8.39 Hz)
16	114.45	6.38, d (J = 8.46 Hz)
17	155.40	
18		9.06, s
19		11.14, s
20	123.57	7.44, d (J = 2.51 Hz)
21	na	
22	na	
23	118.70	6.71, nd
24	118.70	6.71, nd
25	120.88	6.97, nd
26	111.46	7.33, d (J = 8.21 Hz)
27	na	

na − Not assigned

nd − can not determine (due to overlap)

NMR data (1H, ^{13}C APT, COSY, HSQC and HMBC) was acquired. The spectral data showed signals that can be interpreted as two indole like structures, one phenyl group, one ketone, two CH_2 groups and one quaternary aliphatic carbon. These structures with the respective carbon and proton information are shown in Figure S7.1.

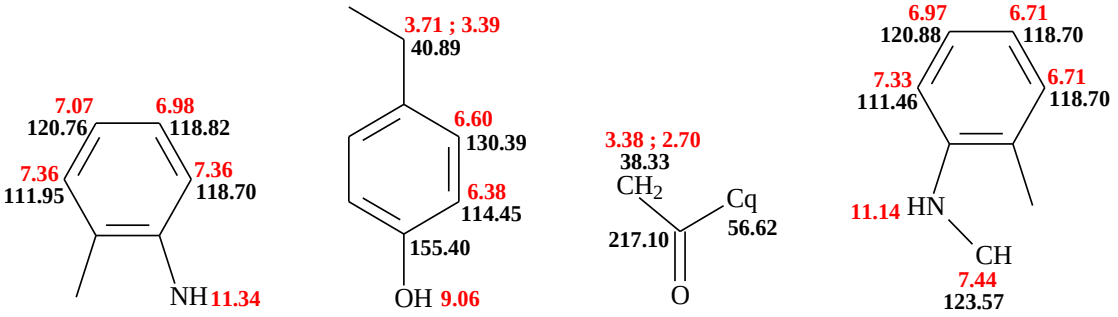


Figure S7.1. Partial structures that appear to be present in compound **6** with carbon and proton chemical shift information. Carbon chemical shifts are in black font, proton in red font. Cq – aliphatic quaternary carbon atom.

It can be observed from the chemical shift values that there is considerable signal overlap both in carbon and proton (noticeably three resonances at 118.70 ppm in carbon, two resonances at 7.36 and at 6.71 in proton). The relatively high proportion of quaternary carbons in these molecules does not allow for a straightforward sequential assignment, and the only tool available to assign those resonances and to link the several pieces of the structure comes from the HMBC data. Since the HMBC can detect connectivity 2, 3, 4 (and sometimes even 5) bonds apart, in cyclic molecules like these the information becomes ambiguous. Considering the available information the structure depicted in Figure S7.2 is consistent with the data.

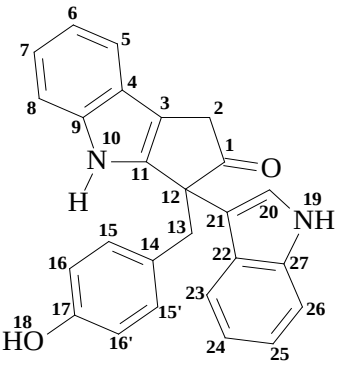


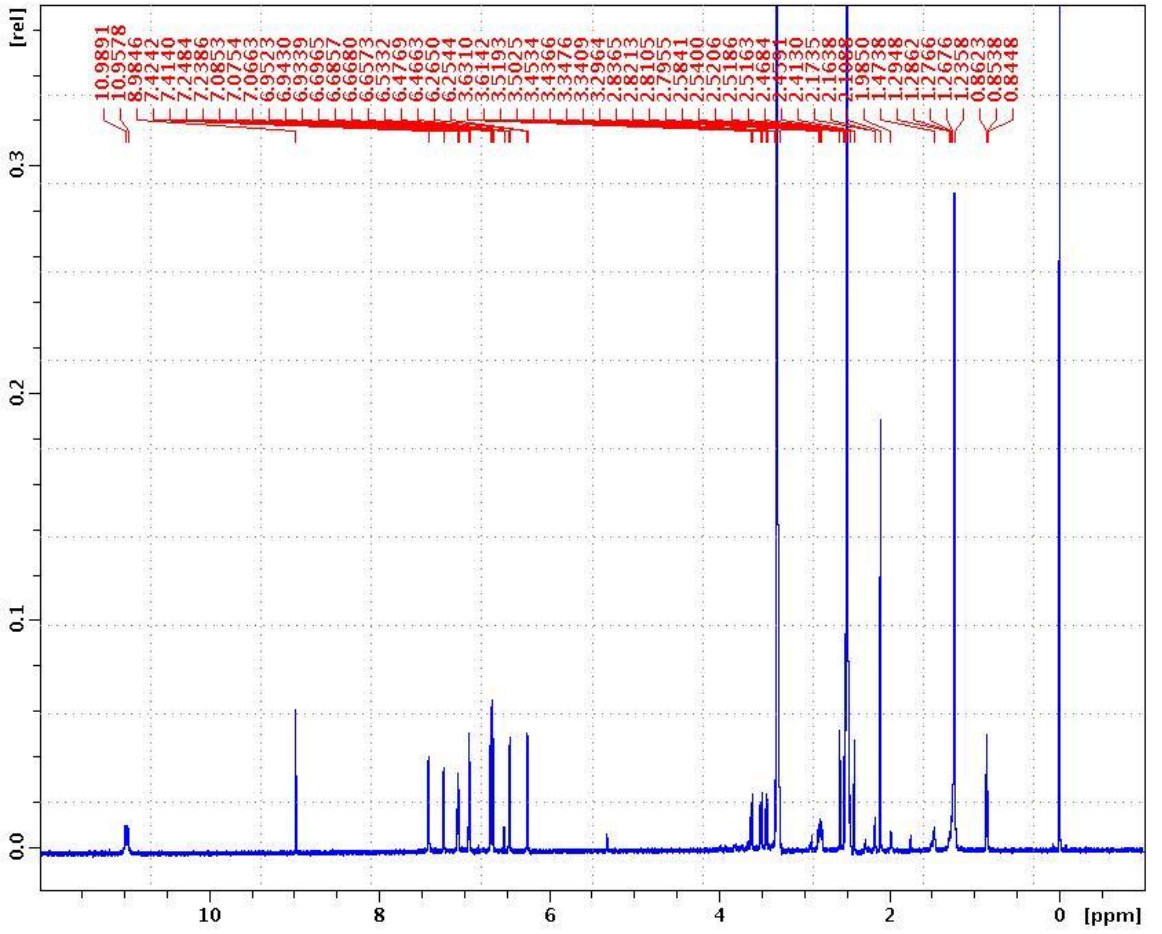
Figure S7.2. Possible structure of the compound 6 with atom numbering

The acquired NMR data is consistent with the structure depicted in Figure S7.2, however, signal overlap precluded the complete analysis of the HMBC spectrum and therefore the complete assignment of the carbon signals was not possible.

D) NMR spectra of Compound 8

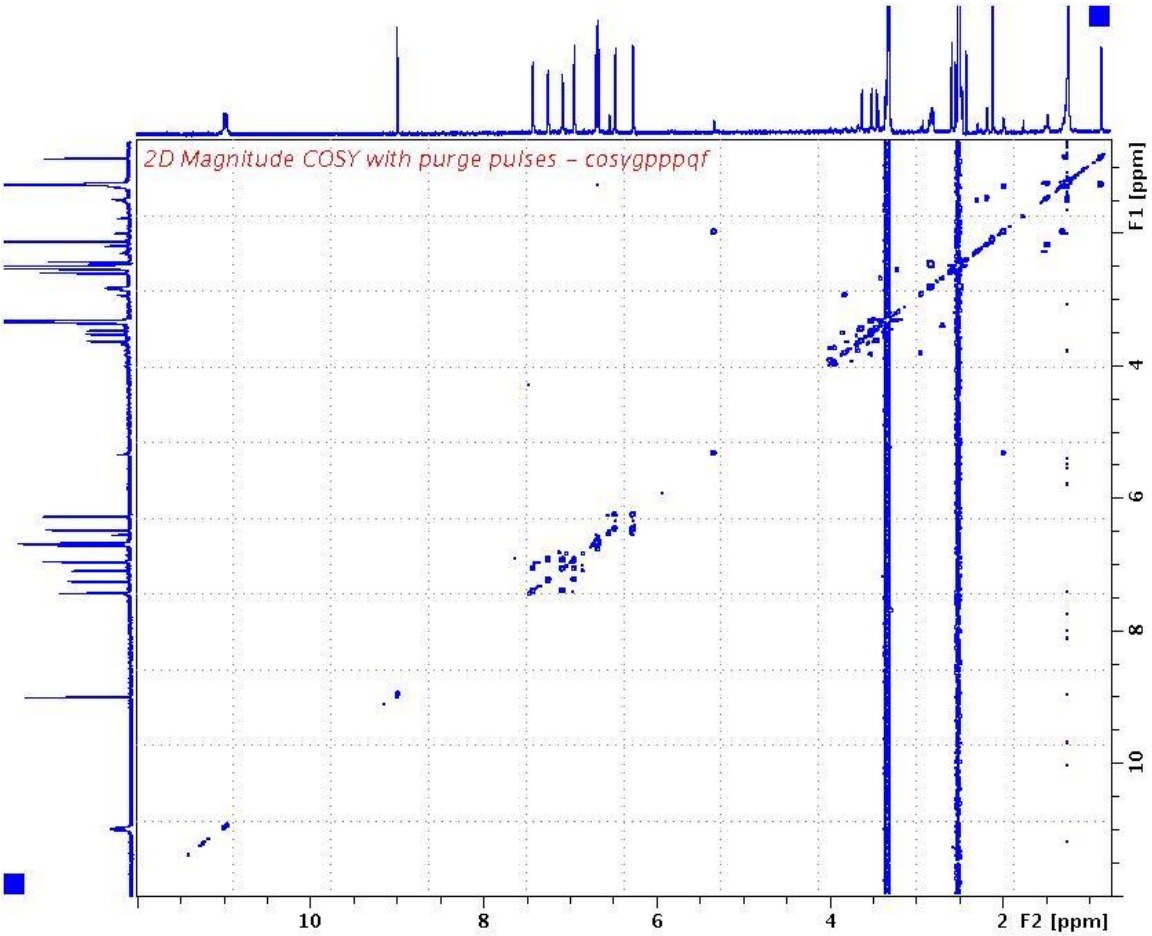
i) ^{1}H NMR spectra of compound 8.

^{1}H-NMR spectra of 8



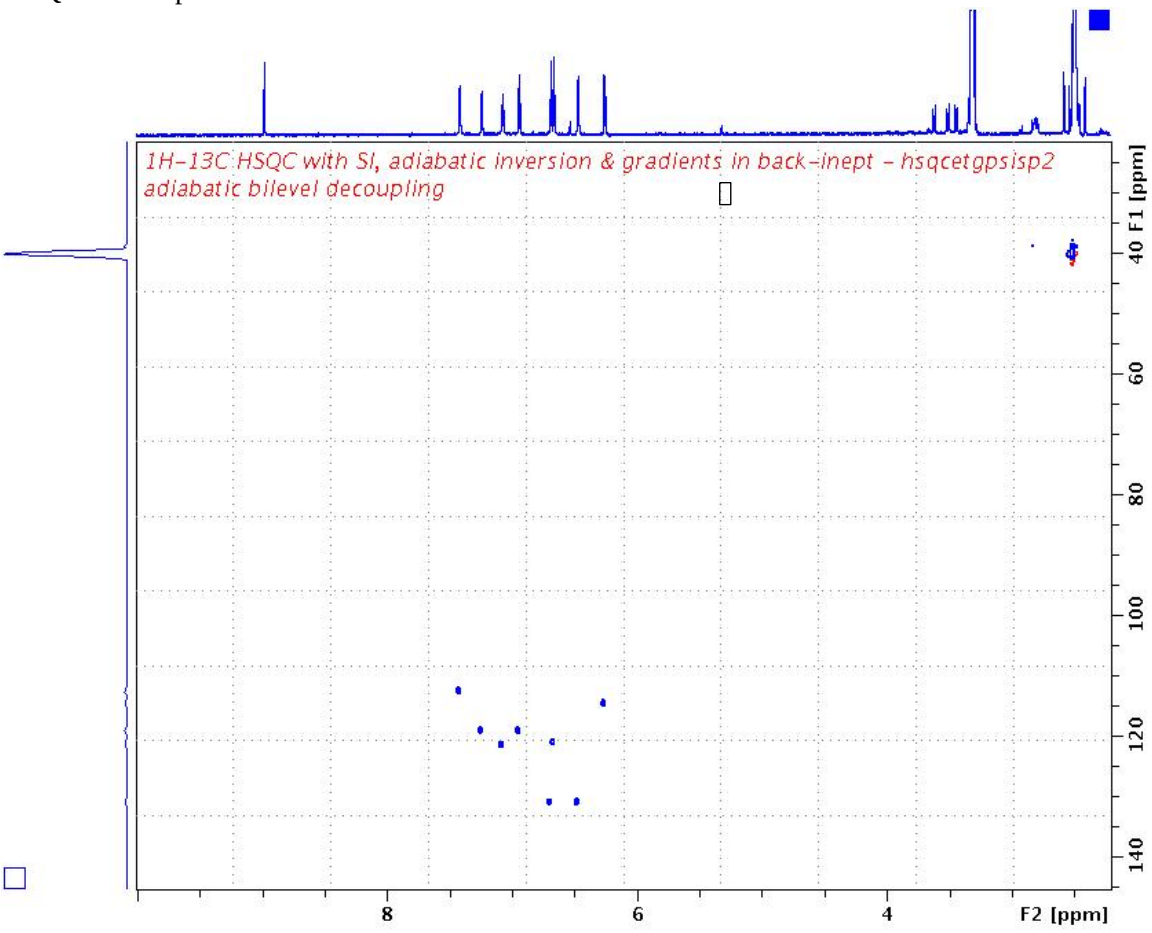
ii) COSY spectra of compound 8

COSY NMR spectra of 8



iii) HSQC spectra of compound 8

HSQC NMR spectra of 8



iv) HMBC spectra of compound 8

HMBC NMR spectra of 8

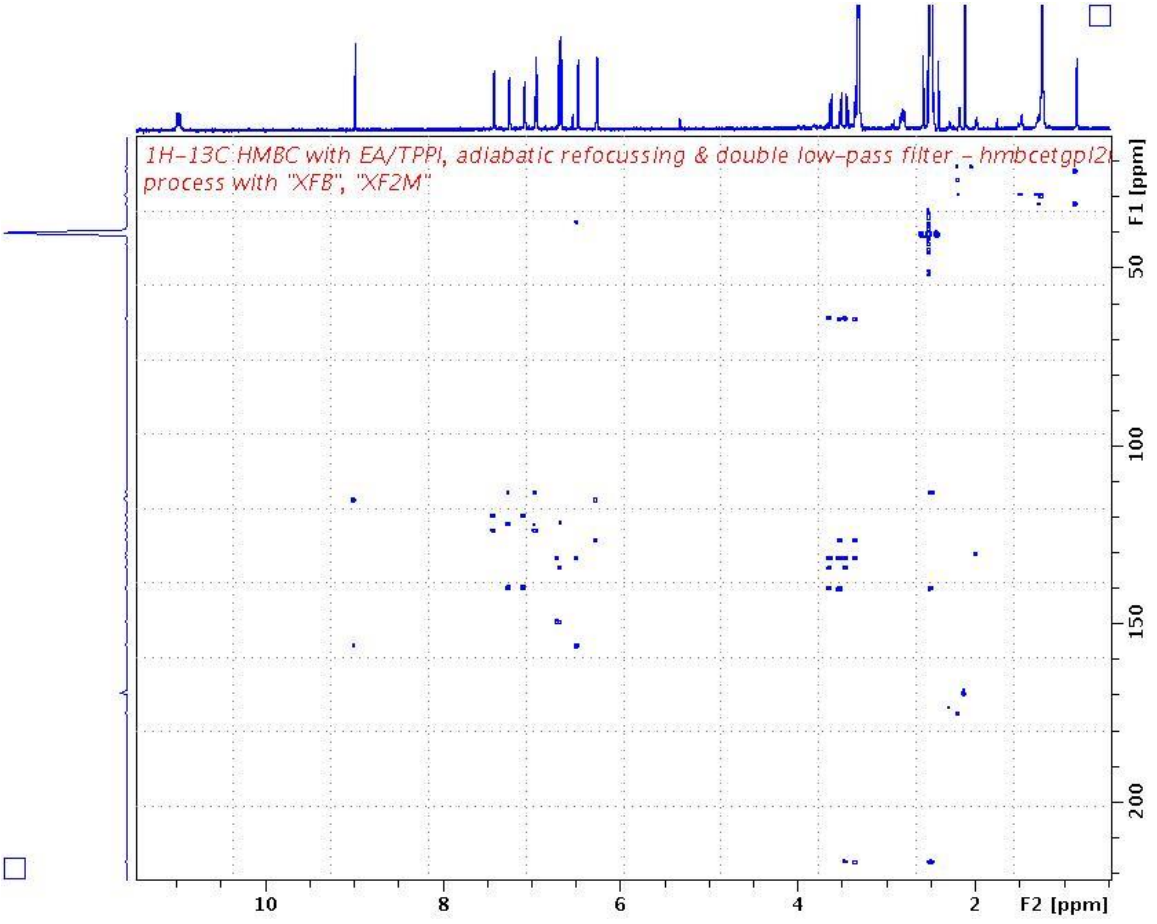


Figure S8: NMR spectra of compound **8**. i) ^{1}H-NMR, ii) COSY, iii) HMBC, and iv) HSQC spectra of compound **8**.

Table S3: Assignment of ^{1}H and ^{13}C signals for compound number 8

	Chemical shift (ppm)				
position	^{13}C	^{1}H	position	^{13}C	^{1}H
1	215.98		1’	216.16	
2	38.37	2.82, d (J= 12.51 Hz) ; 2.47, (#)	2’	38.37	2.82 d (J= 12.51 Hz) ; 2.47 (#)
3*	112.25		3’*	112.25	
4*	139.05		4’*	139.05	
5	118.72	7.24, d (J= 7.52 Hz)	5’	118.72	7.24, d (J= 7.52 Hz)
6	118.88	6.94, t (J= 7.52 Hz)	6’	118.88	6.94, t (J= 7.52 Hz)
7	121.07	7.08, t (J= 7.52 Hz)	7’	121.07	7.08, t (J= 7.52 Hz)
8	112.39	7.42, d (J= 7.88 Hz)	8’	112.39	7.42, d (J= 7.88 Hz)
9*	122.85		9’*	122.85	
10		10.99 s	10’		10.96, s
11*	139.30		11’*	139.00	
12	63.39		12’	63.18	
13	36.13	3.51, d (J=13.57 Hz) ; 3.33, (#)	13’	36.22	3.62, d (J = 13.57 Hz); 3.44, d (13.57 Hz)
14	125.75		14’	133.43	
15	130.50	6.47, d (J= 8.51 Hz)	15’	130.50	6.96, d (J= 8.60 Hz)
16	114.32	6.26, d (J= 8.51 Hz)	16’	120.75	6.66, d (J= 8.60 Hz)
17	155.24		17’	148.61	
18		8.98, s	18’		
			19’	168.71	
			20’	20.66	2.11, s

*assignments of carbon signals were assisted by chemical shift estimation using chemdraw software.
Peak structure cannot be determined because of overlap with solvent lines

A similar analysis was performed with this sample resorting to ^{1}H, COSY, HSQC and HMBC. In this case the 1D ^{13}C APT spectrum was not acquired because of lack of signal, the sample is not concentrated enough to allow direct carbon detection. However, carbon data was obtained from the analysis of the indirect detection spectra (HSQC and HMBC). This was possible partly because of the structural similarities between this compound and the previously analysed sample (compound **7**). From the gathered data an indole and phenyl structures are apparent, with typical proton signals multiplicities and COSY correlations. A ketone and a methyl signals are also present. As with compound 7, there is a CH_2 and a quaternary carbon in the five member ring and the link between the two major portions of the molecule is done via that quaternary carbon. The NMR data is consistent with the structure depicted in Figure S8.1.

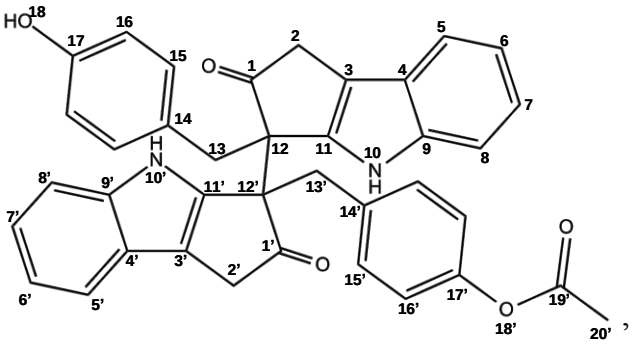


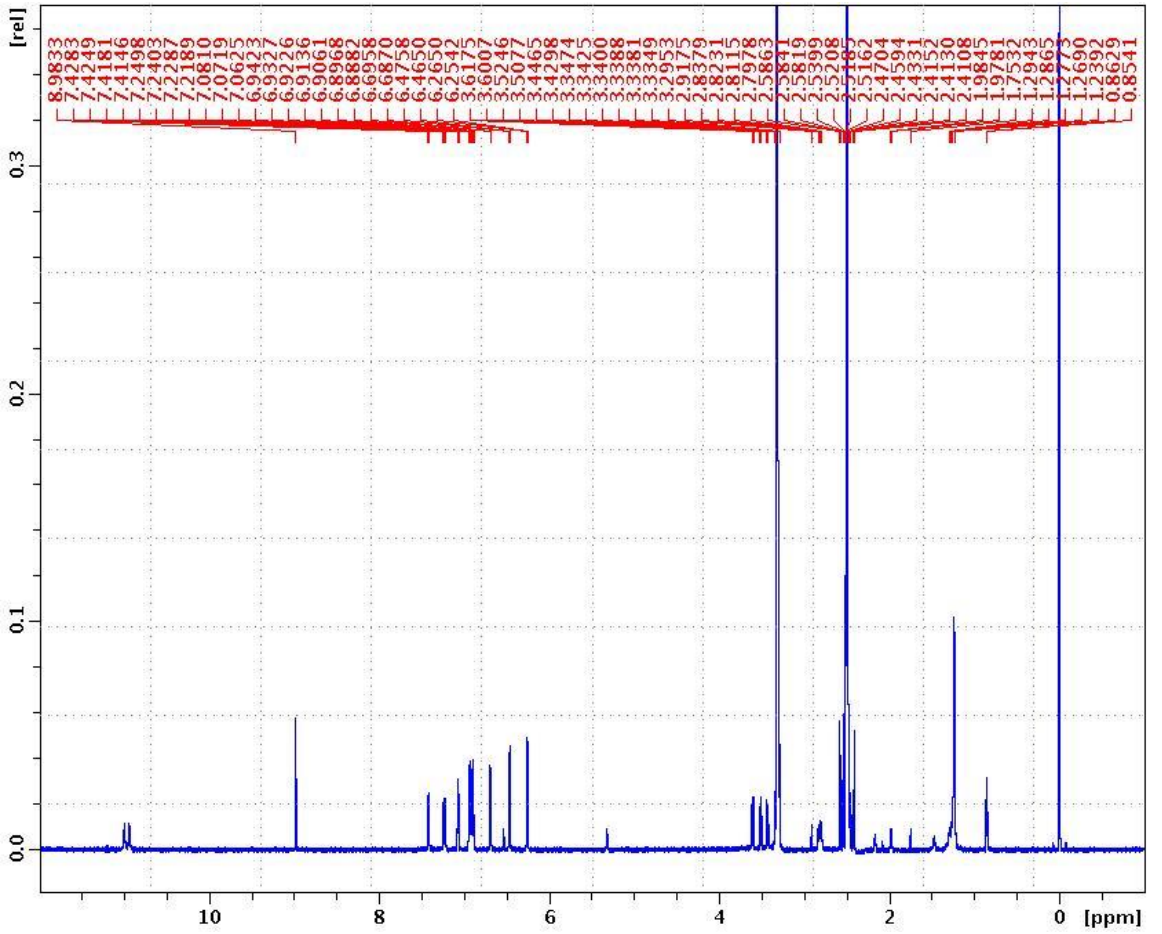
Figure S8.1. Structure of the compound **8** with atom numbering

As expected from the structural similarity, the spectra of compound **8** present features which are similar to those of compound **7**; however, in this case the symmetry of the molecule is broken. Some signals (like the ones coming from the indole group) overlap completely while others (like 15 and 15' or 16 and 16') are well differentiated and others still (like 1 and 1' or 12 and 12') are so close that it is almost impossible to distinguish them. Nevertheless, the analysis of the HSQC and HMBC together with the gathered knowledge from the assignment of compound 7 allowed to assign the proton and carbon resonances (Table S3).

E) NMR spectra of Compound 9

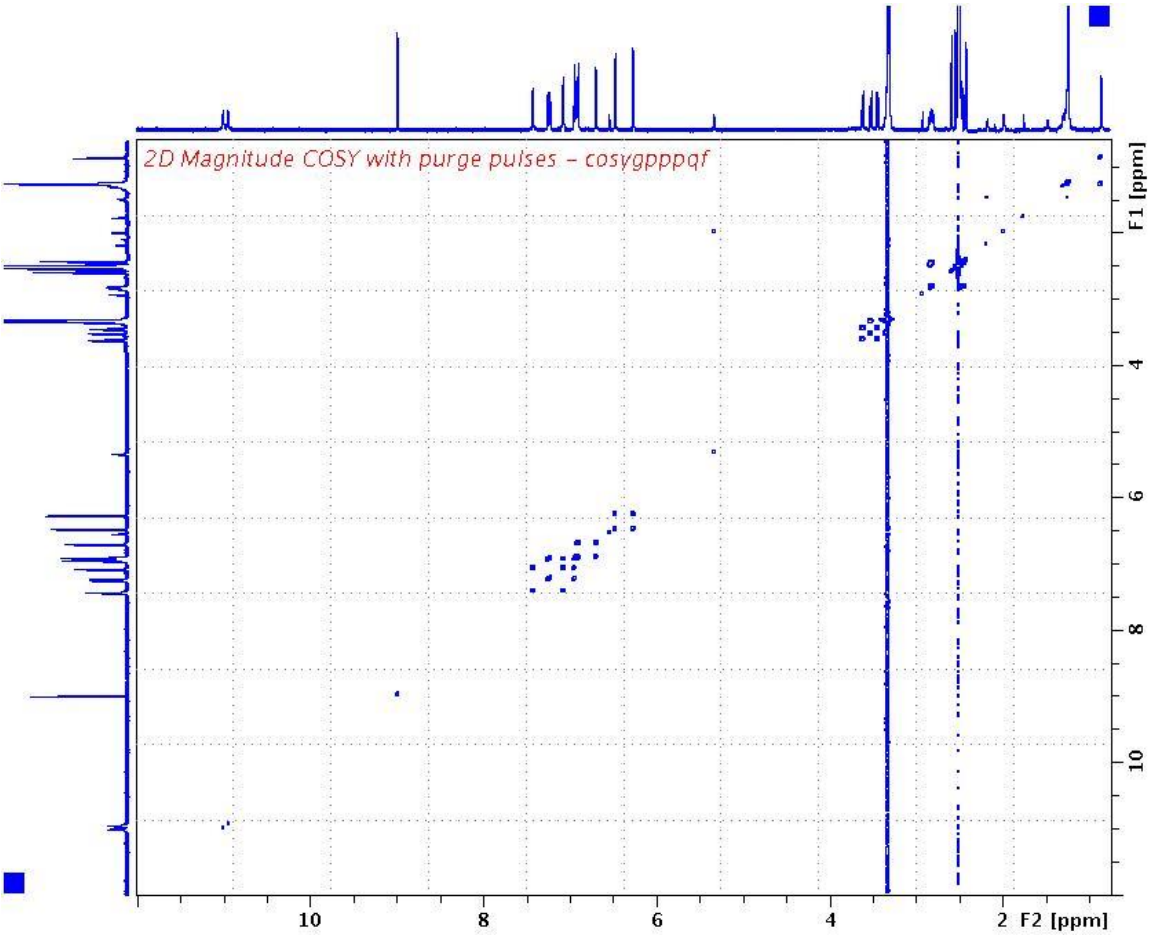
i) ^{1}H NMR spectra of compound 9.

^{1}H-NMR spectra of 9



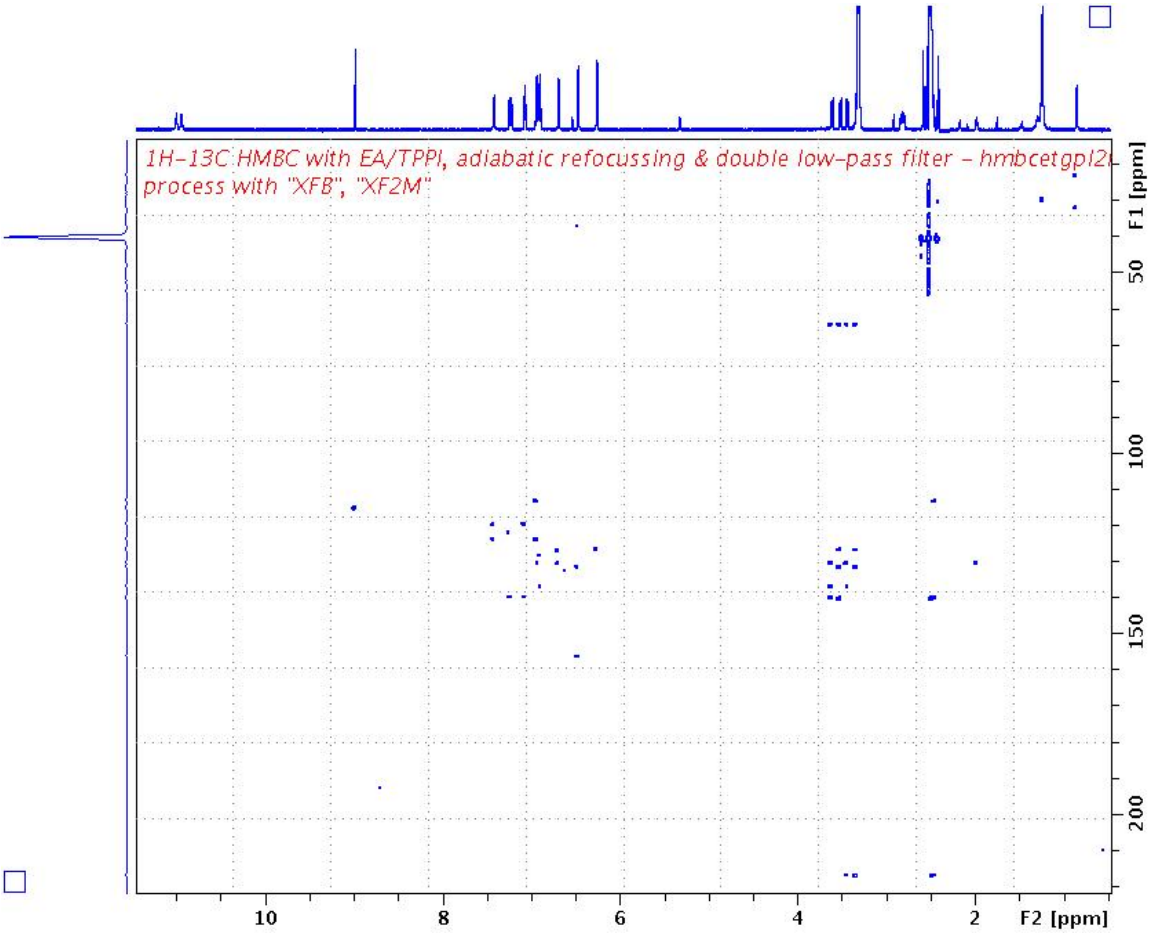
ii) COSY NMR spectra of compound 9.

COSY NMR spectra of 9



iii) HMBC NMR spectra of compound 9.

HMBC NMR spectra of 9



iv) HSQC NMR spectra of compound 9.

HSQC NMR spectra of 9

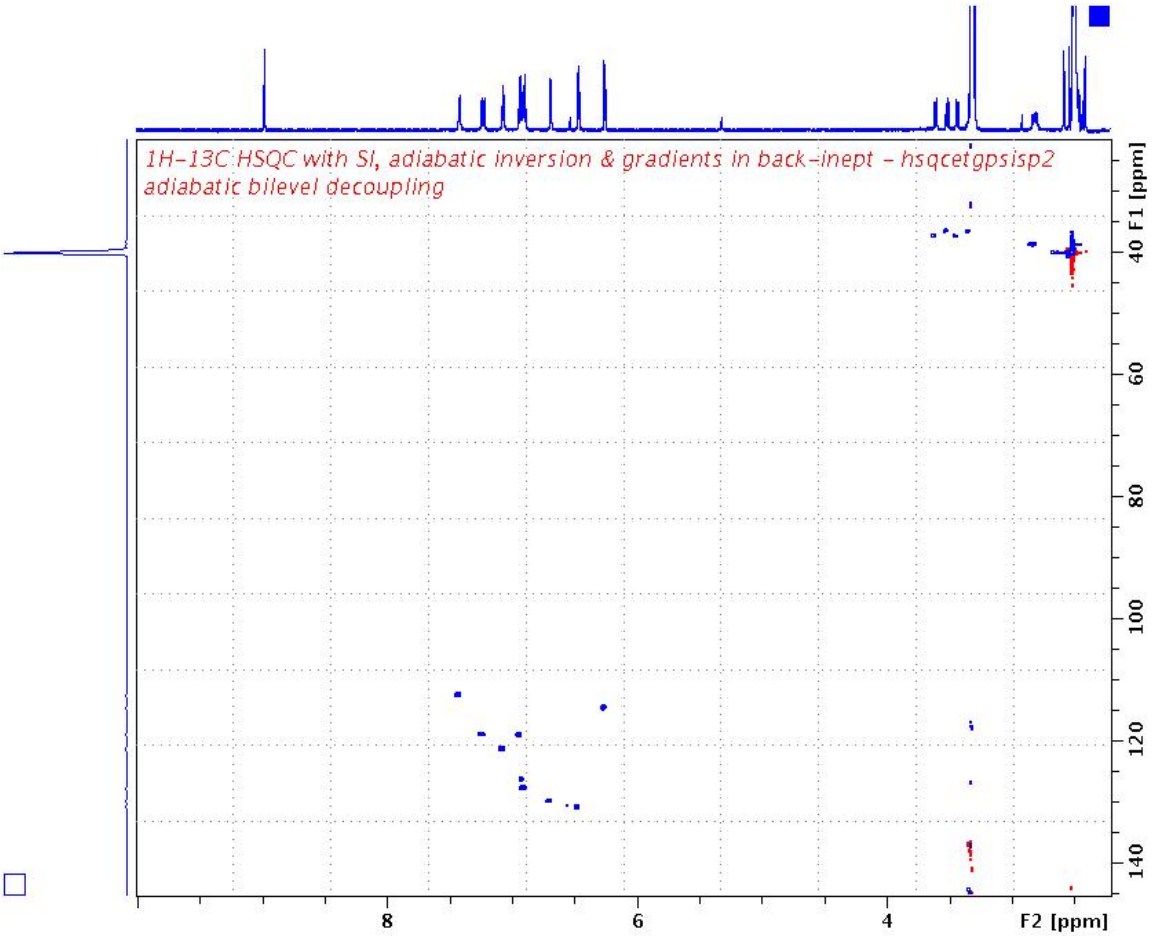


Figure S9: NMR spectra of compound 9. i) ^{1}H-NMR, ii) COSY, iii) HMBC, and iv) HSQC spectra of compound **9**.

Table S4 Assignment of ^{1}H and ^{13}C signals for compound 9

	Chemical shift (ppm)				
position	^{13}C	^{1}H	position	^{13}C	^{1}H
1	216.24		1’	216.02	
2	38.48	2.82, d (J=12.51Hz) ; 2.45, (#)	2’	38.48	2.81, d (J=12.51Hz); 2.45, (#)
3*	112.20		3’*	112.20	
4*	138.98		4’*	138.98	
5	118.72	7.23, d (J=7.52 Hz)	5’	118.72	7.23, d (J=7.52 Hz)
6	118.82	6.94, t (J=7.52 Hz)	6’	118.82	6.94, t (J=7.52 Hz)
7	121.08	7.07, t (J=7.52 Hz)	7’	121.08	7.07, t (J=7.52 Hz)
8	112.27	7.42, d (J=7.88 Hz)	8’	112.27	7.42, d (J=7.88 Hz)
9*	122.87		9’*	122.87	
10		11.00, s	10’		10.94, s
11*	139.29		11’*	139.18	
12	63.34		12’	63.27	
13	36.17	3.51, d (J=13.57 Hz); 3.34 (#)	13’	36.92	3.61, d (J=13.57 Hz); 3.44, d (J=13.57 Hz)
14	125.69		14’	136.04	
15	130.63	6.47, d (J=8.60 Hz)	15’	129.65	6.70, d (J=8.60 Hz)
16	114.34	6.26, d (J=8.60 Hz)	16’	127.51	6.90, d (J=8.60 Hz)
17	155.29		17’	126.08	6.91, d (J=8.60 Hz)
18		8.98,s			

*assignments of carbon signals were assisted by chemical shift estimation using chemdraw software.
peak structure cannot be determined because of overlap with solvent lines

A similar analysis was performed with compound **9** resorting to ^{1}H, COSY, HSQC and HMBC. Also, in this case the 1D ^{13}C APT spectrum was not acquired because of lack of signal and carbon data was obtained from the analysis of the indirect detection spectra (HSQC and HMBC). The spectral features of this compound were extremely similar to those of compound **8** with the exception that the acetyl group is not present and that a benzene ring like structure is visible. All the other characteristics, like the CH_2 group and the quaternary carbon in the five member ring, are same. The NMR data is consistent with the structure depicted in Figure S9.1.

Like compound **8**, the compound **9** is also not symmetrical. Some signals (like the ones coming from the indole group) overlap completely while others (like 15 and 15’ or 16 and 16’) are well differentiated and others still (like 1 and 1’ or 12 and 12’) are so close that it is almost impossible to distinguish them. Nevertheless, the analysis of the HSQC and

HMBC together with the gathered knowledge from the assignment of compound **7** and **8** allowed to assign the proton and carbon resonances (Table S4).

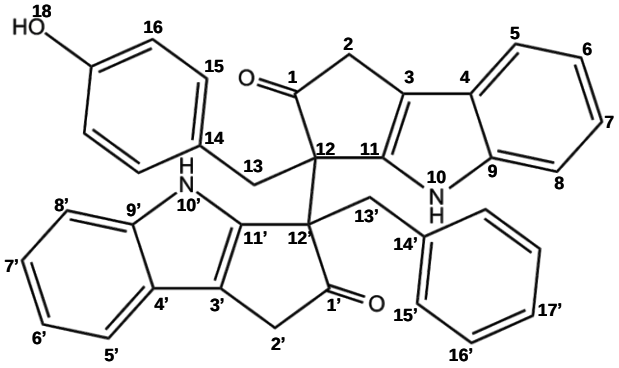
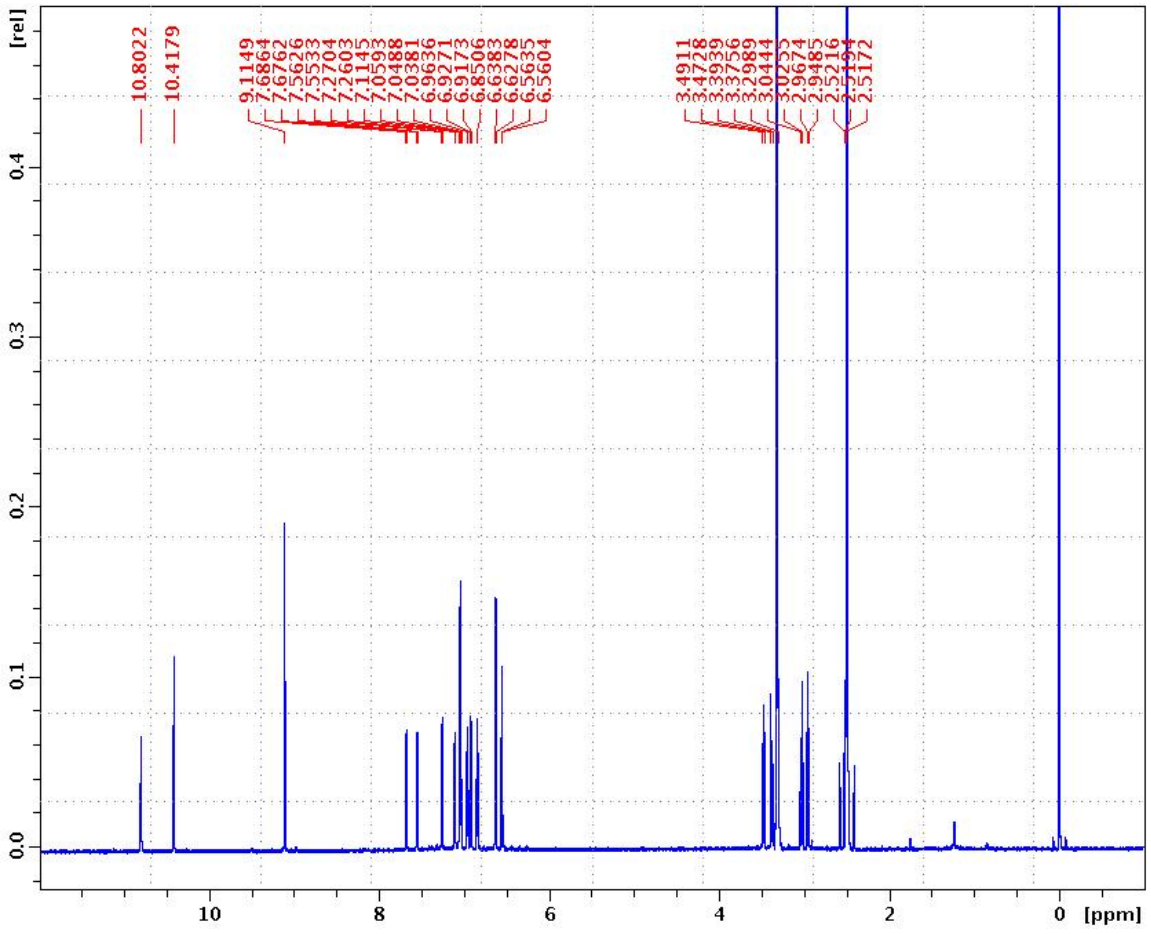


Figure S9.1. Structure of the compound **9** with atom numbering

F) NMR spectra of Compound 5

i) ^{1}H NMR spectra of compound 5.

^{1}H-NMR spectra of 5



i) COSY NMR spectra of compound 5.

COSY NMR spectra of 5

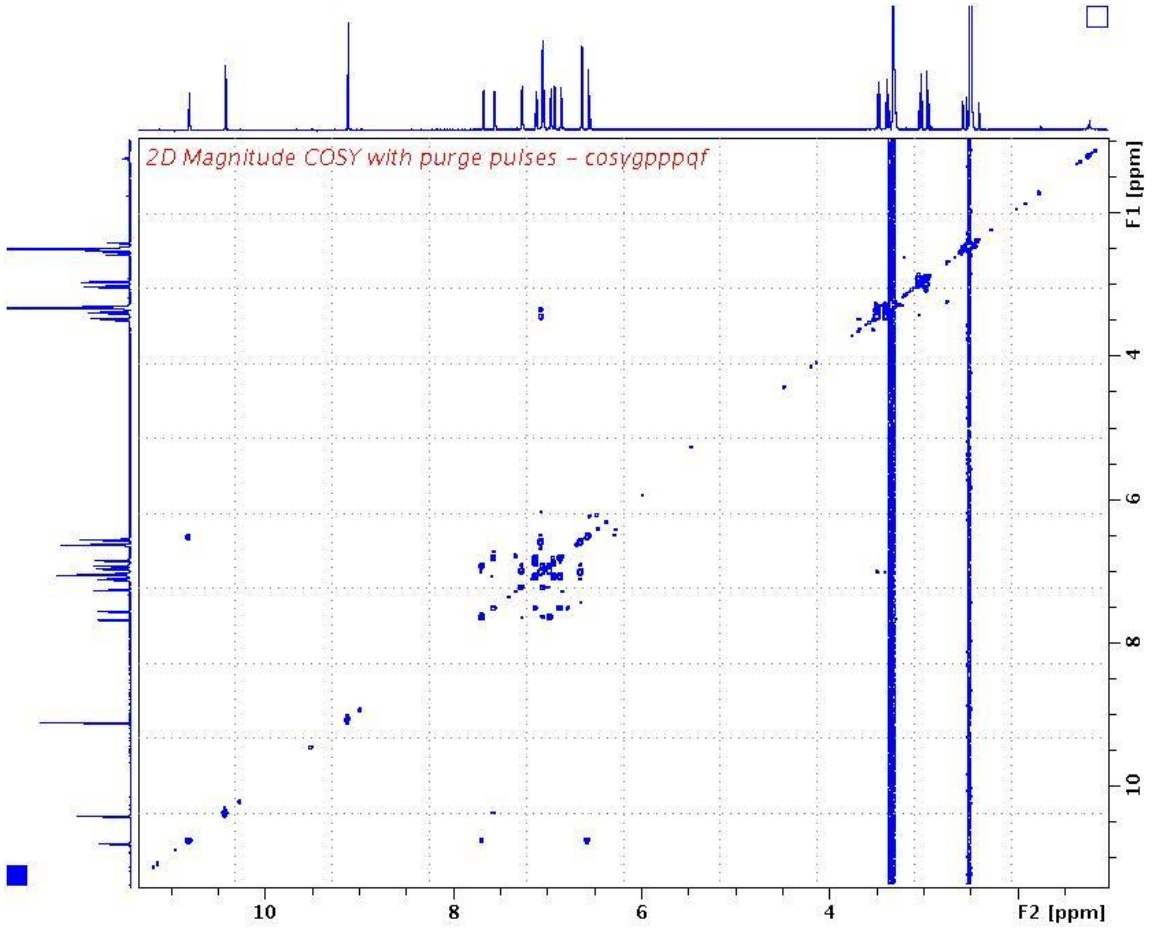


Figure S10: NMR spectra of compound **5**. i) ^{1}H NMR, and ii) COSY spectra of compound **5**.

The NMR data showed signals that can be interpreted as two indole like structures, one phenyl group, one ketone, two CH_2 groups and one quaternary aliphatic carbon. These structures with the respective carbon and proton information are shown in figure S 9.1.

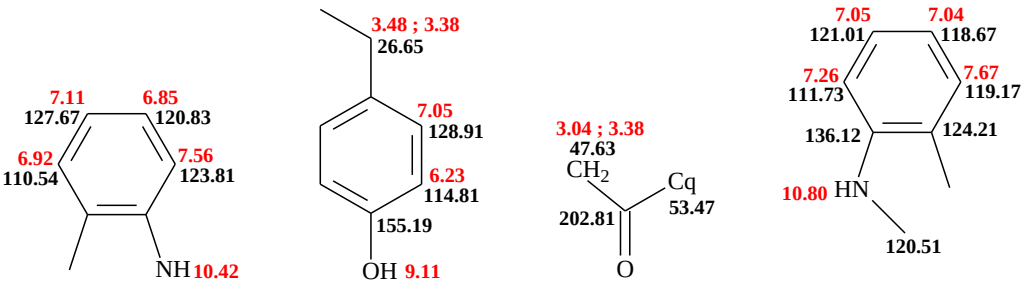


Figure S10.1. Partial structures that appear to be present in compound 5 with carbon and proton chemical shift information. Carbon chemical shifts are in blue, proton in red. Cq – aliphatic quaternary carbon atom.

These partial structures are similar to those of compound **6**. Although both compounds **5** and **6** have same mass, they differ in retention time in HPLC and UV absorption spectra. Based upon these information, we proposed the structure of compound **5** could be the isomer of compound **6** shown in Figure S10.2.

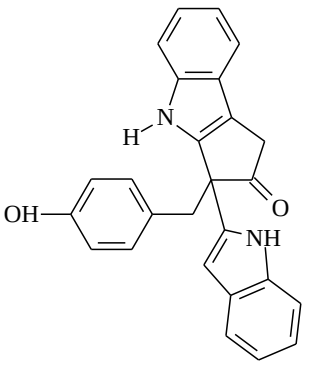


Figure S10.2. Proposed Structure of the compound **5**.