Supporting Information

Structural elucidation of bisulfite adducts to pseudouridine that result in deletion signatures during reverse transcription of RNA

Aaron M. Fleming,[†] Anton Alenko,[†] Jay P. Kitt,[†] Anita M. Orendt,[†‡] Peter F. Flynn,[†]

Joel M. Harris,[†] and Cynthia J. Burrows[†*]

[†]Department of Chemistry, University of Utah, Salt Lake City, UT 84112-0850, United States

[‡]Center for High Performance Computing, University of Utah, Salt Lake City, UT 84112-0190, United States

*To whom correspondence should be addressed burrows@chem.utah.edu

Additional Methods

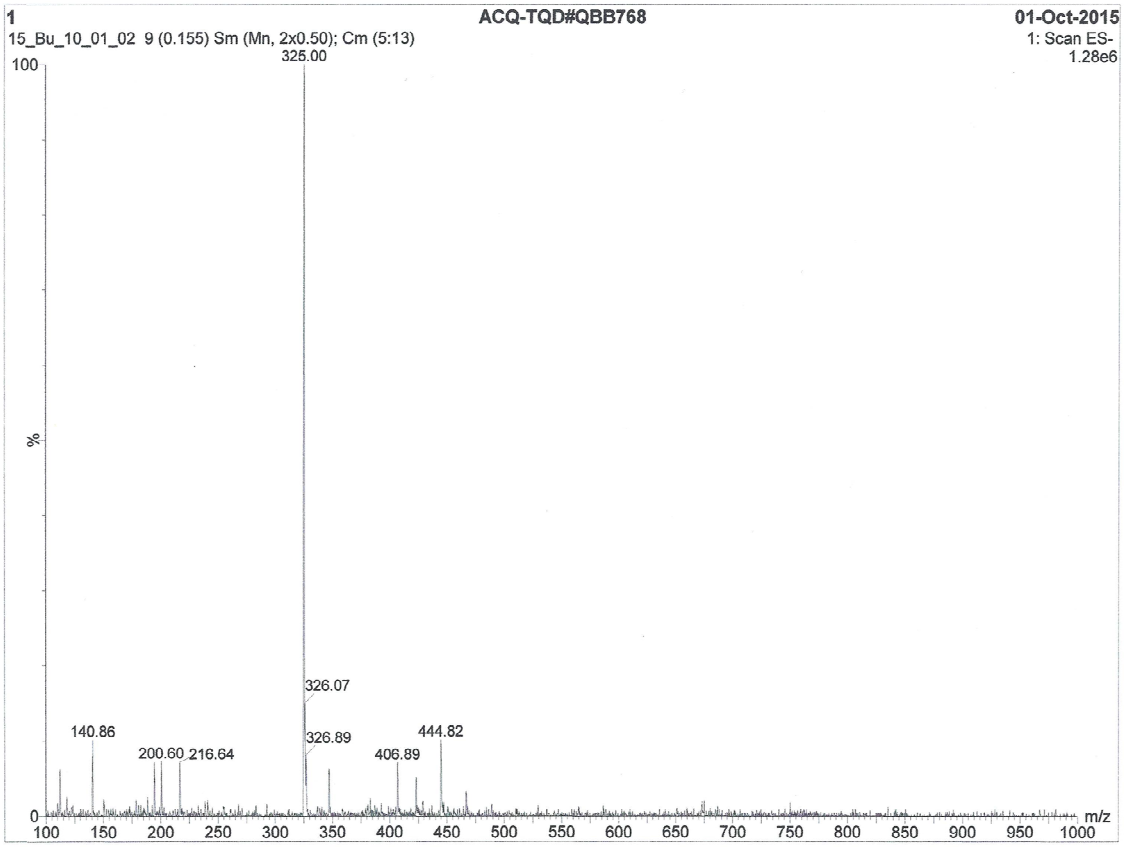
Labeling of the DNA primer. To monitor progression of primer extension by reverse transcriptases via polyacrylamide gel electrophoresis (PAGE), the DNA primer was 5' end-labeled with ^{32}P following a procedure adopted from the literature (Fleming, A. M., Muller, J. G., Dlouhy, A. C., and Burrows, C. J. (2012) *J. Am. Chem. Soc. 134*, 15091-15102) using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA.) and [γ-^{32}P] ATP (PerkinElmer, Waltham, MA.).

Primer extension assay. Before all primer extension assays, samples were annealed by heating 6.5 µL of aqueous solution containing reverse transcription buffer with no magnesium (75 mM KCl, 50 mM Tris-HCl pH 8.3 in a final volume of 10 µL), 1.17 pmol (117 nM in 10 µL) of HPLC purified RNA template before or after bisulfite treatment, and 1.06 pmol (106 nM in 10 µL) DNA primer including ~30,000 cpm of ^{32}P-labeled strand to 95 °C for 5 min, then incubating them at 55 °C for another 5 min followed by cooling at 15 °C for 10 min. Upon completion of annealing, stock solutions of $MgCl_2$ (3 mM in 10 µL), DTT (10 mM in 10 µL), and dNTPs (500 µM) were added to the reaction mixture to bring the volume to 9 µL. Then 1 µL of stock solution containing 100 U of SuperScript™ III (Invitrogen, Carlsbad, CA) in 50% glycerol was added to give a final volume of 10 µL. Reaction mixtures were incubated for 40 min at 35-50°C followed by 10 min at 93°C to facilitate degradation of enzyme and RNA template. Reaction mixture was then diluted with an equal volume of 2x gel loading buffer (8 M urea, 0.01% xylene cyanole, 0.01% bromphenol blue, x1 TBE buffer) and heated to 95 °C for 10 min. Resulting solutions were analyzed on 20% denaturing PAGE. Gels were stored with a storage phosphor screen for 12-18 hours which was then scanned using a phosphoimager. The resulting images were analyzed using ImageJ2 software. (Schindelin, J., Rueden, C. T., Hiner, M. C., and Eliceiri, K. W. (2015) Mol. Reprod. Dev. 82, 518-529)

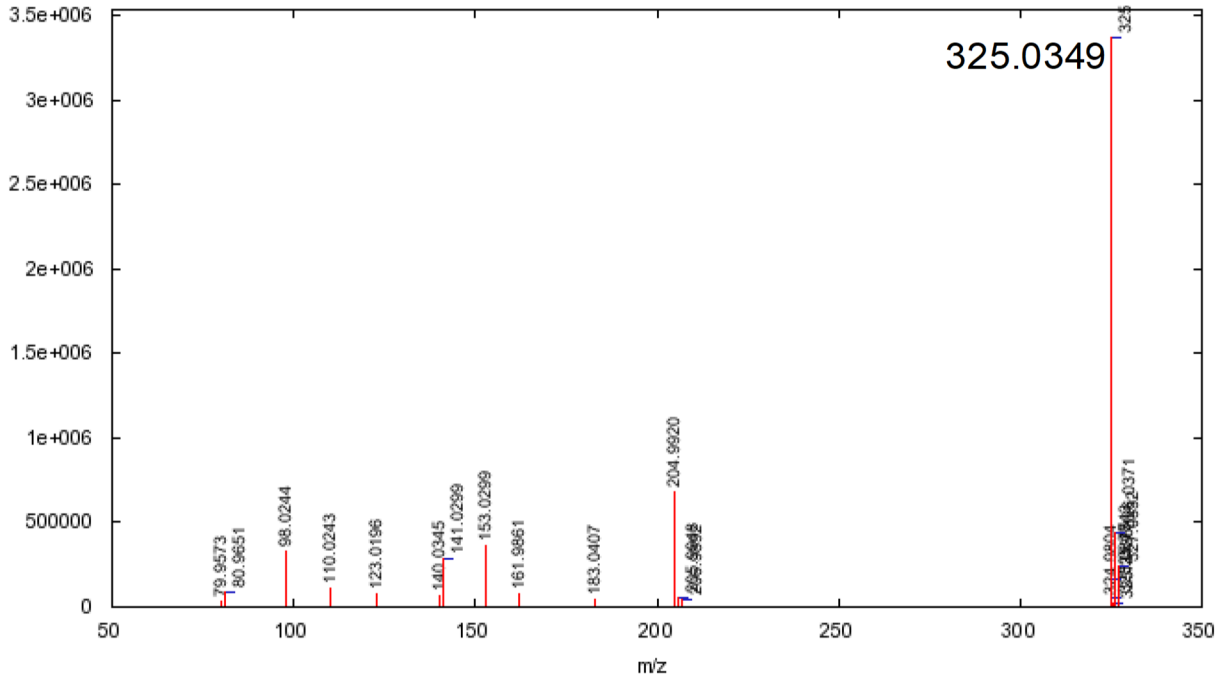
Raman spectroscopic characterization. Raman spectroscopy was utilized to characterize the the C-SO_3 connectivity between compounds 1 and 5. Raman measurements were carried out on a confocal Raman microscope. Details of the instrumentation have been published previously (Houlne, M. P.; Sjostrom, C. M.; Uibel, R. H.; Kleimeyer, J. A.; Harris, J. M. (2002) Anal. Chem. 74, 4311-9). Measurement of Raman spectra of compounds 1 and 5 were carried out in a microscopy well cell (Kitt, J.P., Harris, J.M. (2014) Anal. Chem. 86, 3, 1719-1725). A small volume of solution was placed in the well cell and the confocal probe was translated into solution. The Raman excitation source was a Kr^+ laser operating at 50 mW and spectra were collected for not more than two minutes. Three spectra were collected from each sample. The spectra were baseline corrected using a custom Matlab™ script which fits a 7^{th}-order polynomial to non-peak containing portions of the spectrum. Baseline corrected spectra were normalized to the C-H stretching region of the spectrum to correct for concentration differences and to highlight distinctive Raman bands in each sample. The spectra in Figure S12 are the average of the three normalized spectra from each sample. References for the C-S and S=O vibrational assignments are provided in the manuscript. The psuedouracil band assignments are from Ueda, T., Shinozaki, K., Ushizawa, K., Tsuboi, M. (1997) J. Raman Spectrosc. 28, 947-952.

Figure S1. ESI$^-$-MS and ESI$^-$-MS/MS spectra for the bisulfite adducts.

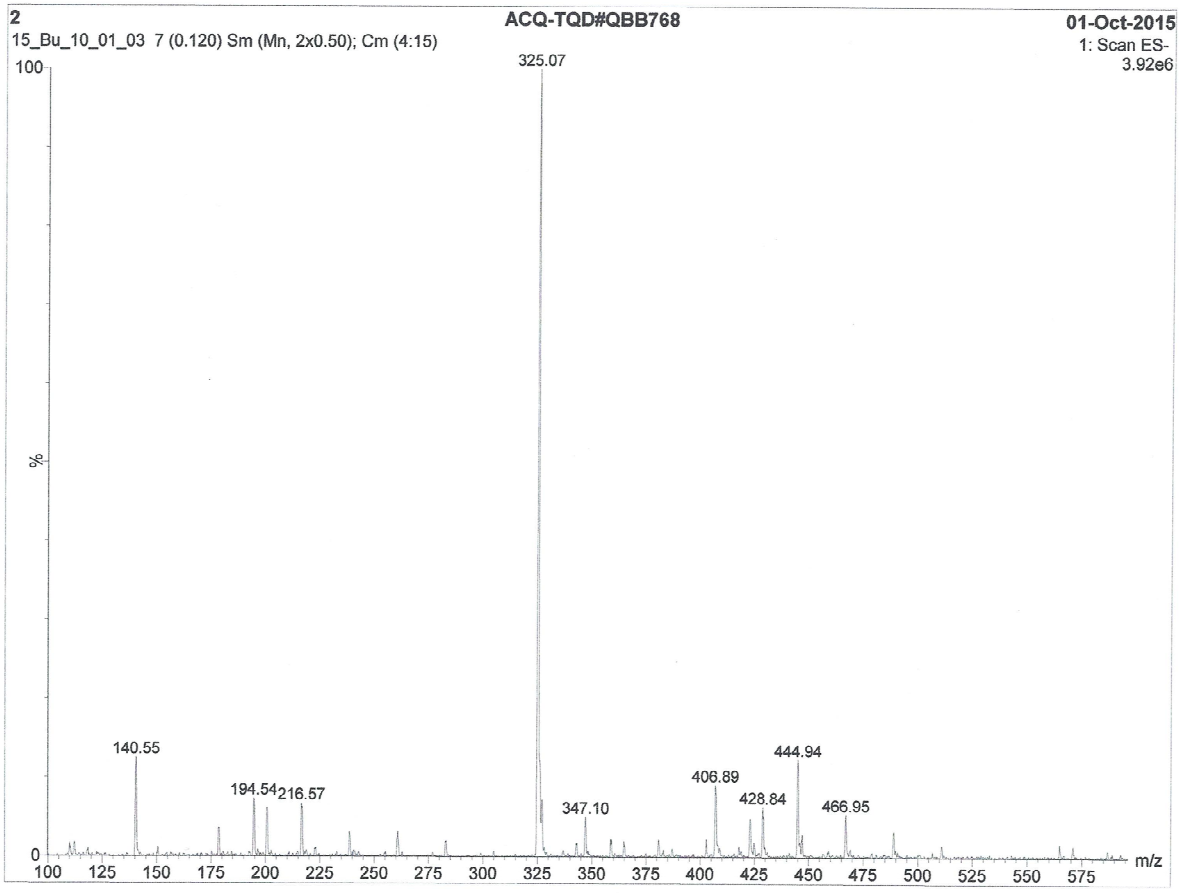
ESI$^-$-MS for **1**



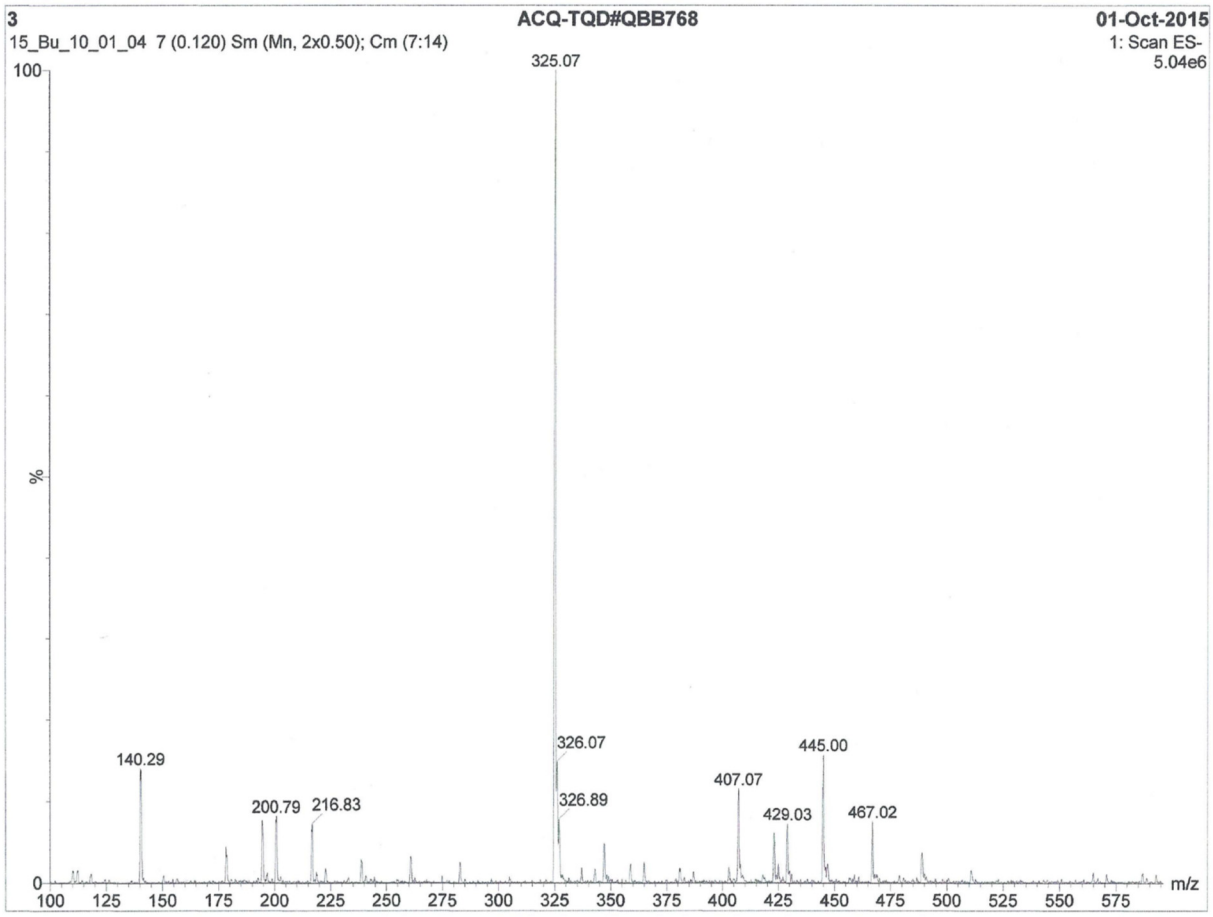
ESI$^-$-MS/MS for **1**



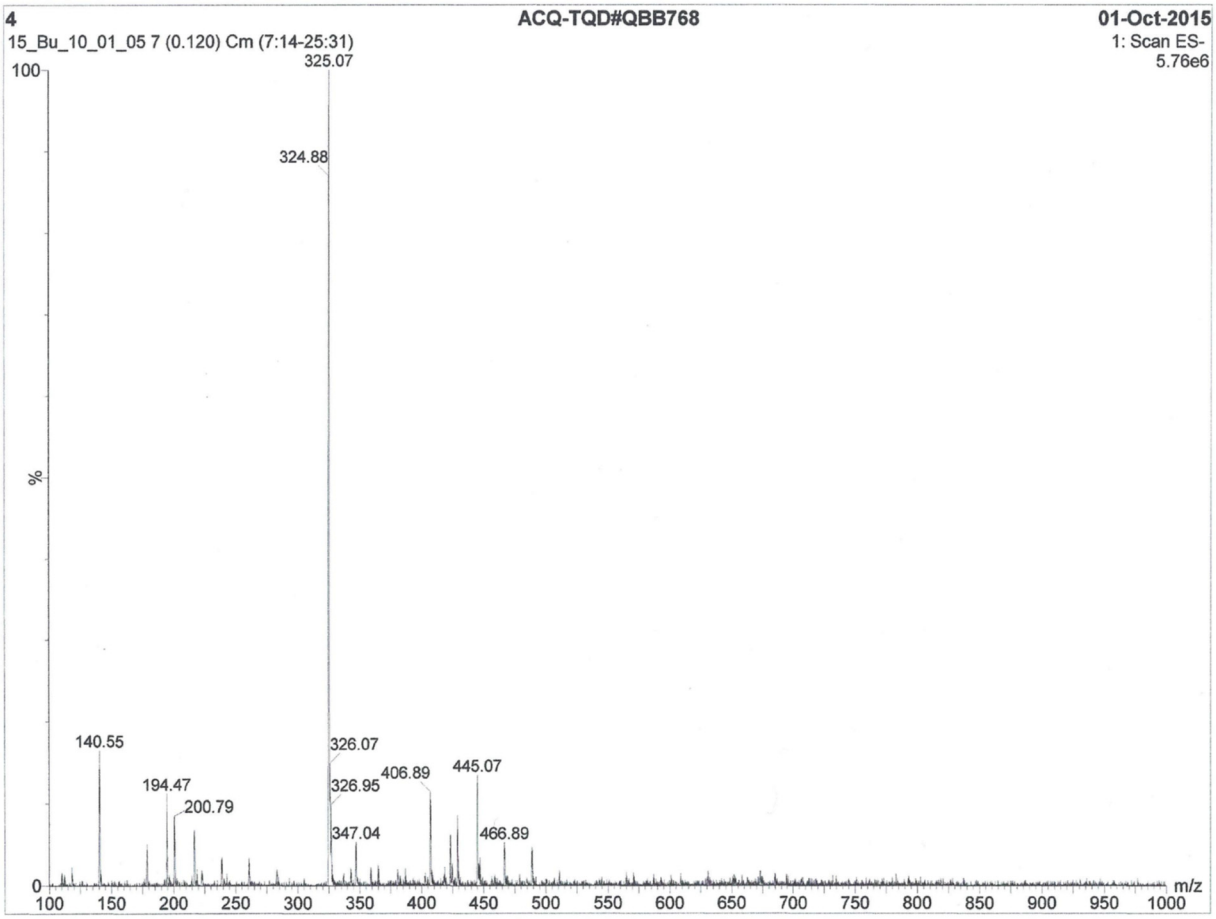
ESI$^-$-MS for **2**


2
ACQ-TQD#QBB768
01-Oct-2015
15_Bu_10_01_03 7 (0.120) Sm (Mn, 2x0.50); Cm (4:15)
1: Scan ES-
3.92e6
100
%
0
325.07
140.55
194.54
216.57
347.10
406.89
428.84
444.94
466.95
m/z
100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575

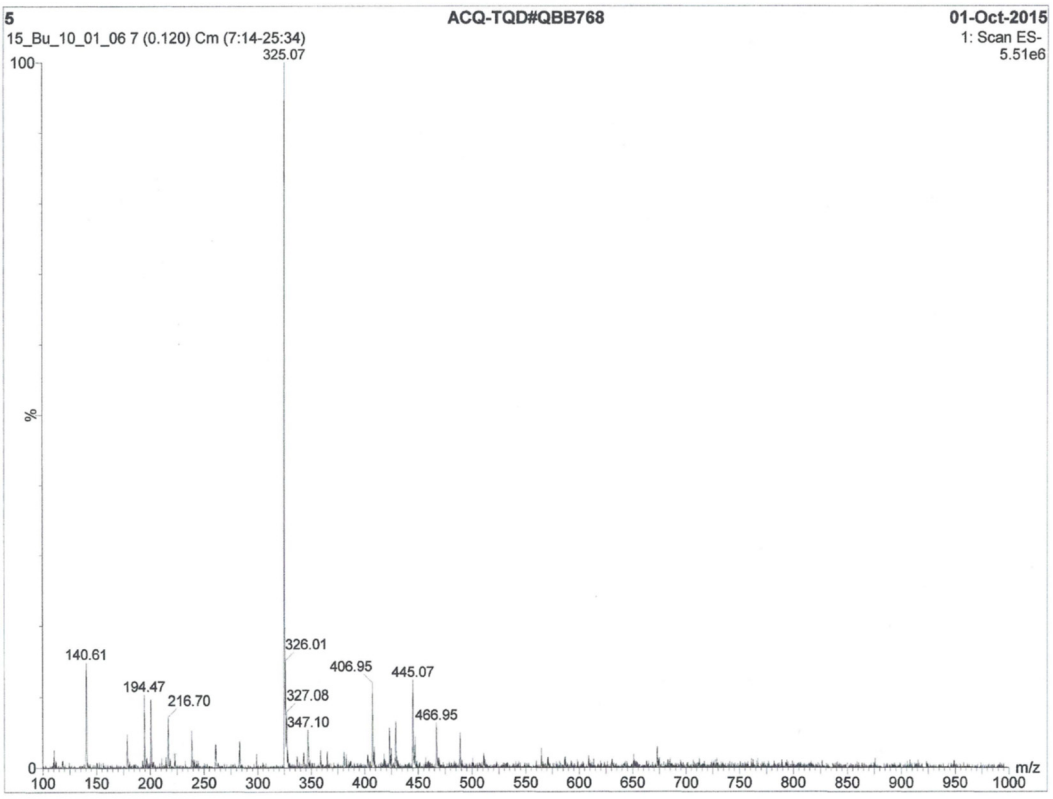
ESI$^-$-MS for **3**


3
ACQ-TQD#QBB768
01-Oct-2015
15_Bu_10_01_04 7 (0.120) Sm (Mn, 2x0.50); Cm (7:14)
1: Scan ES-
5.04e6
100
%
0
325.07
326.07
326.89
140.29
200.79
216.83
407.07
429.03
445.00
467.02
m/z
100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575

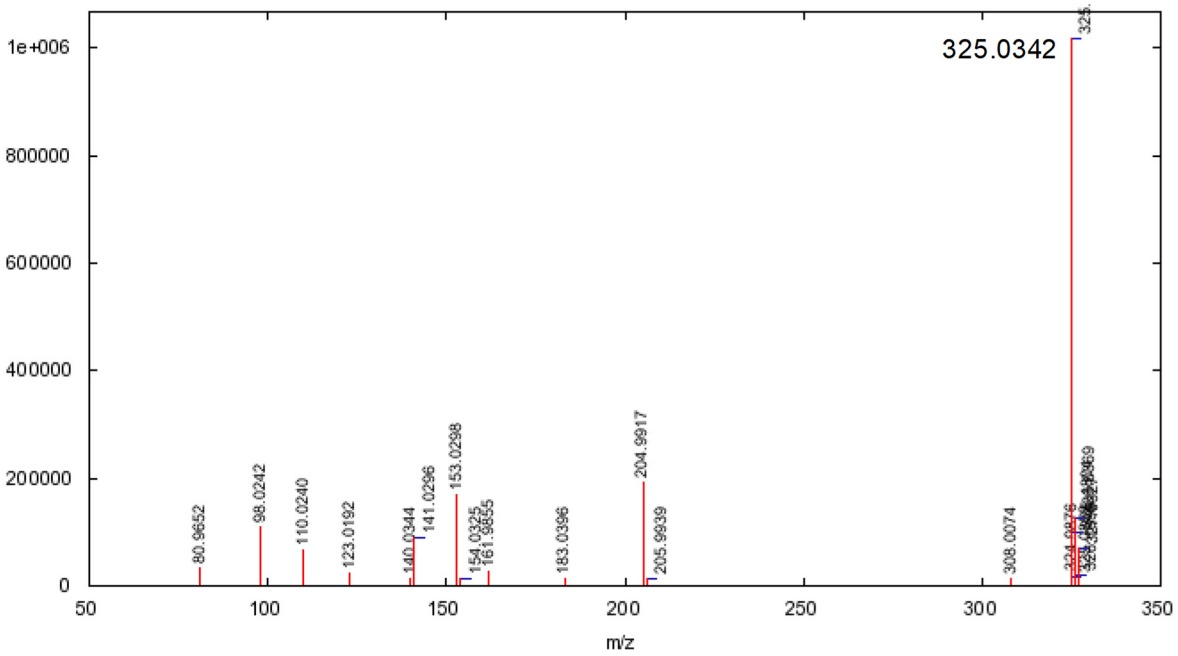
ESI$^-$-MS for **4**


4
ACQ-TQD#QBB768
01-Oct-2015
15_Bu_10_01_05 7 (0.120) Cm (7:14-25:31)
1: Scan ES-
5.76e6
100
%
0
325.07
324.88
140.55
326.07
194.47
326.95
406.89
445.07
200.79
347.04
466.89
m/z
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000

ESI$^-$-MS for **5**



ESI$^-$-MS/MS for **5**



ESI$^-$-MS for **6**

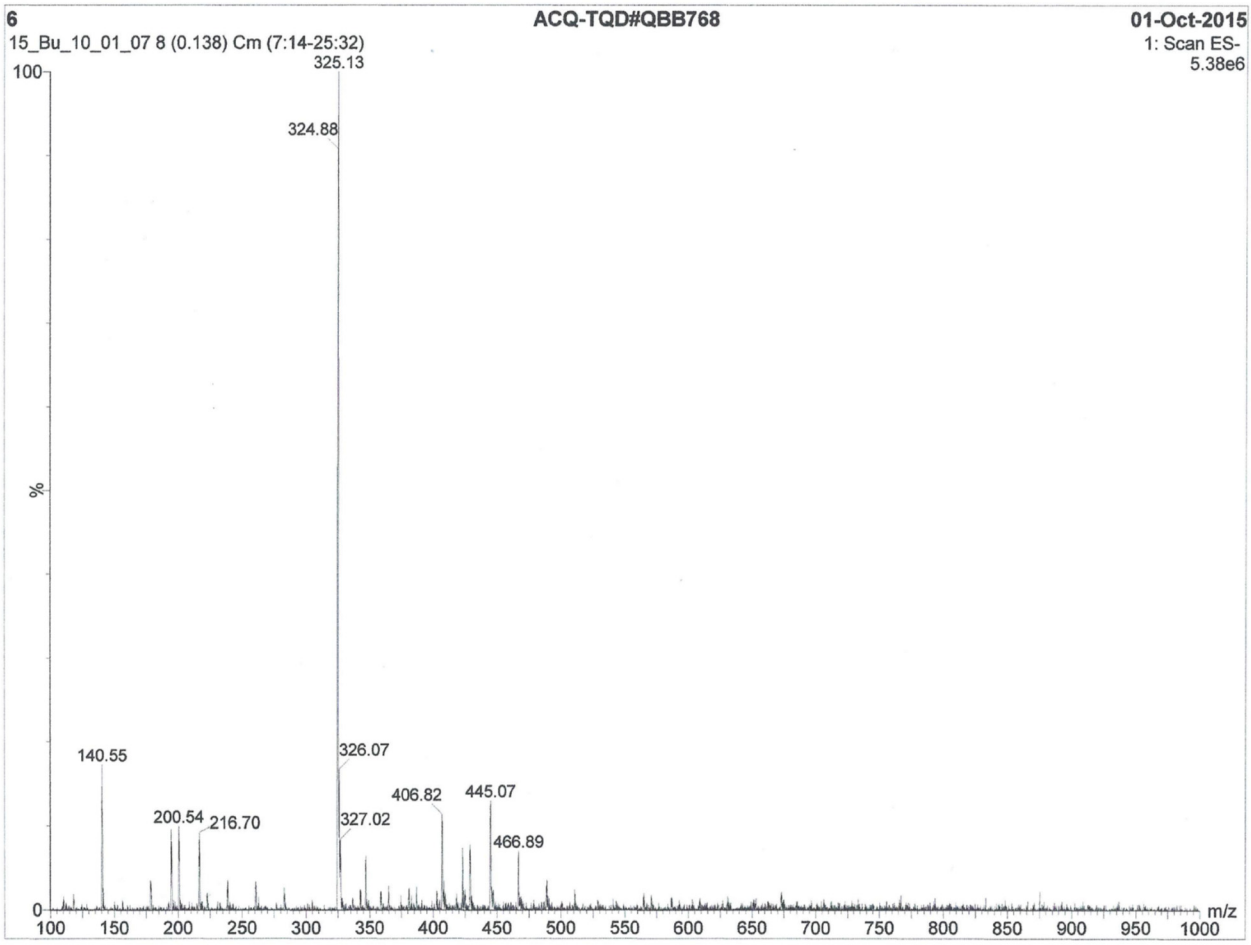

6
ACQ-TQD#QBB768
01-Oct-2015
15_Bu_10_01_07 8 (0.138) Cm (7:14-25:32)
1: Scan ES-
5.38e6
100
%
0
325.13
324.88
140.55
326.07
406.82
445.07
200.54
216.70
327.02
466.89
m/z
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000

Figure S2. ^{1}H-NMR for **1**.

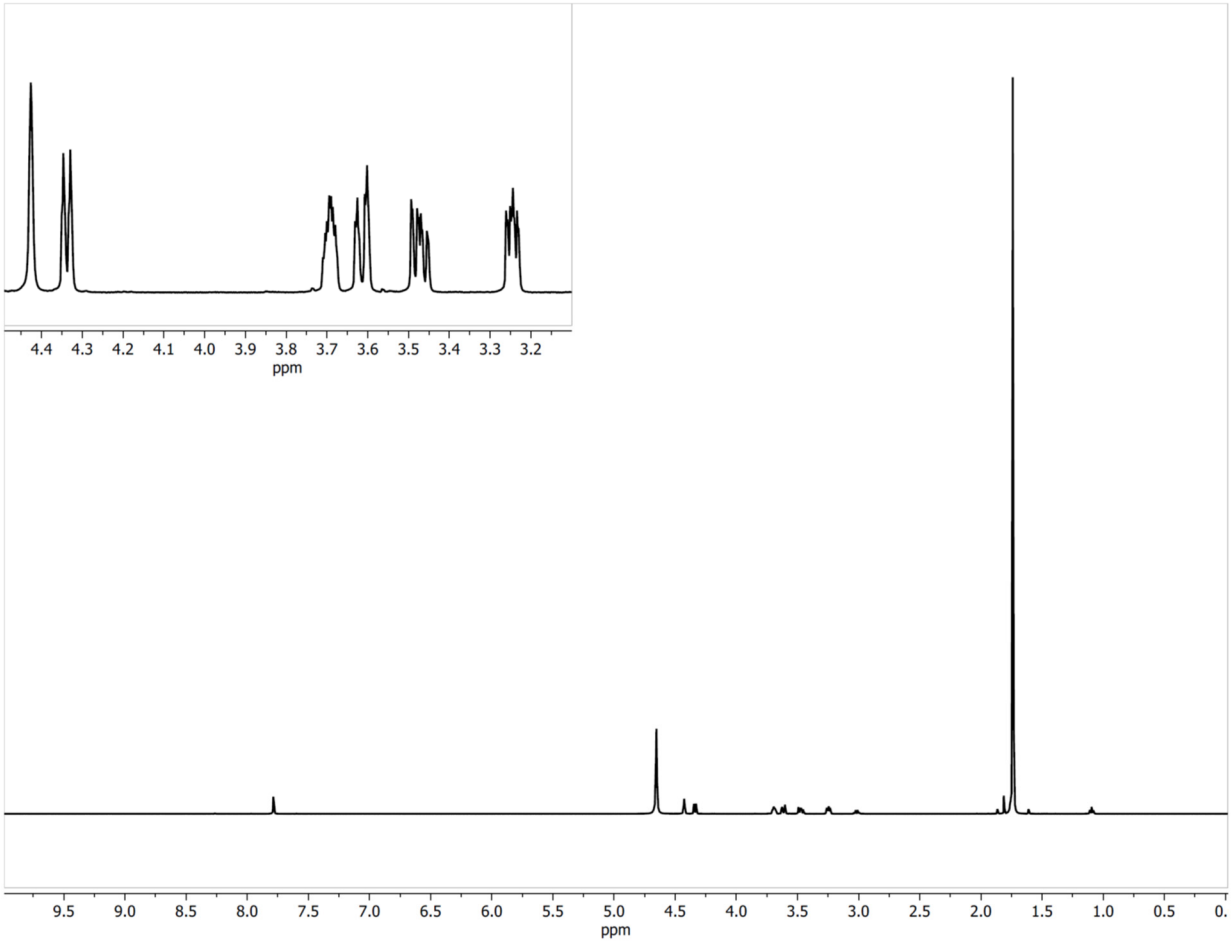


Table of ^{1}H-NMR Signals

Proton	δ ppm (D_2O)
1`	4.40
2`	4.31
3`	3.22
4`	3.67
5`A	3.61
5`B	3.45
6	7.76

There exists acetate as an impurity in the spectrum from the HPLC purification.

Figure S3. $^{1}H,^{1}H$-COSY for **1**.

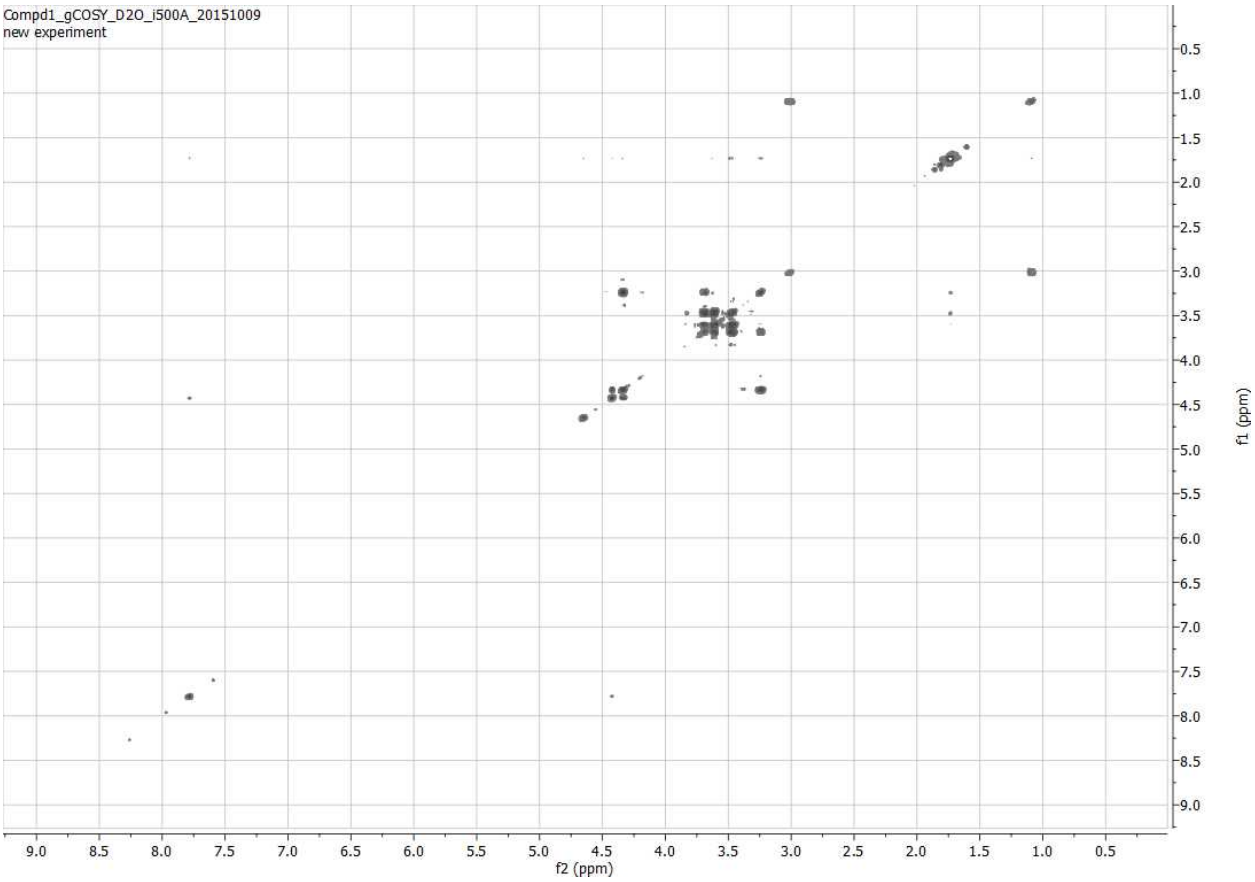


Figure S4. $^{1}H,^{1}H$-TOCSY for **1**.

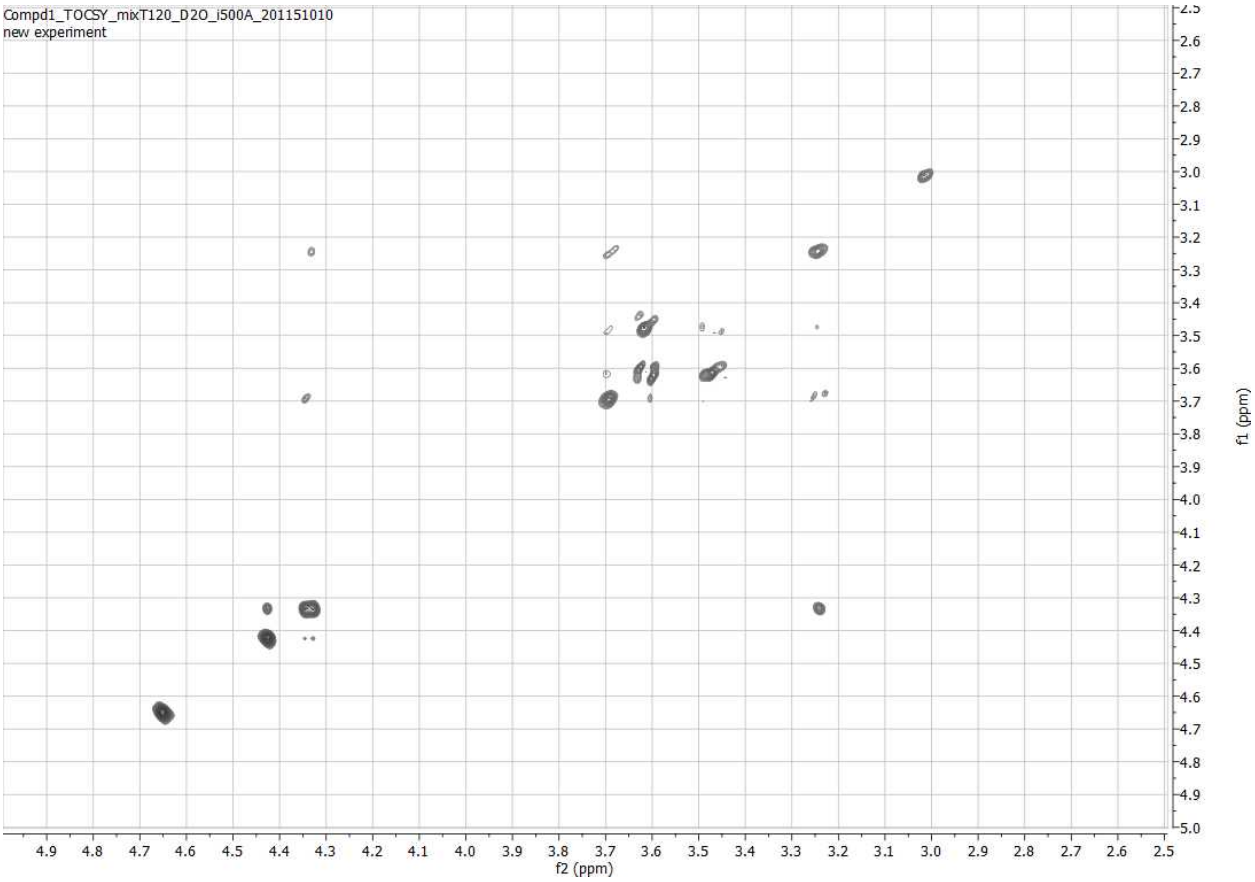


Figure S5. ^{13}C-NMR for **1**.

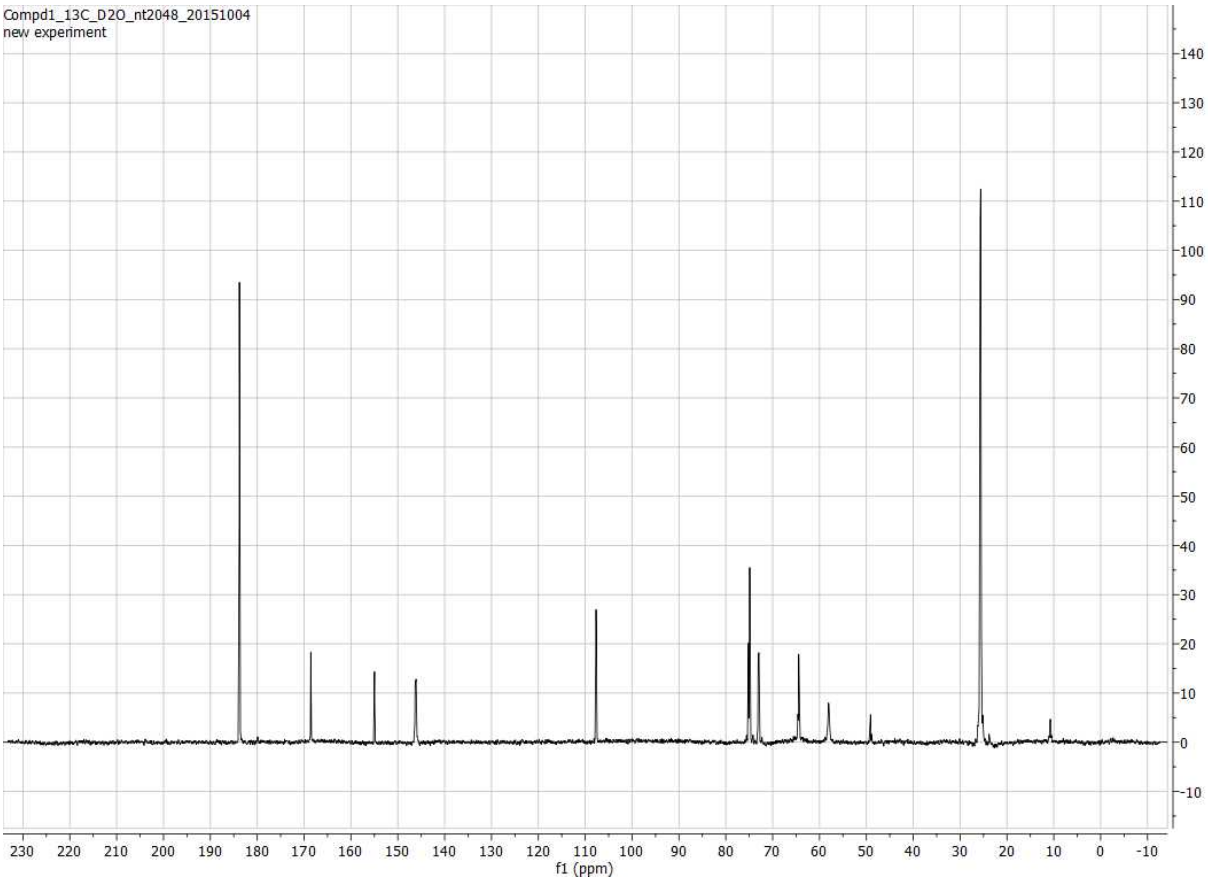


Table of ^{13}C-NMR Shifts

Proton	δ ppm (D_2O)
1`	58.1
2`	73.0
3`	75.2
4`	74.9
5`	64.5
2	154.9
4	168.6
5	107.7
6	146.1

There exists acetate as an impurity in the spectrum from the HPLC purification.

Figure S6. $^{1}H,^{13}C$-HSQC for **1**.

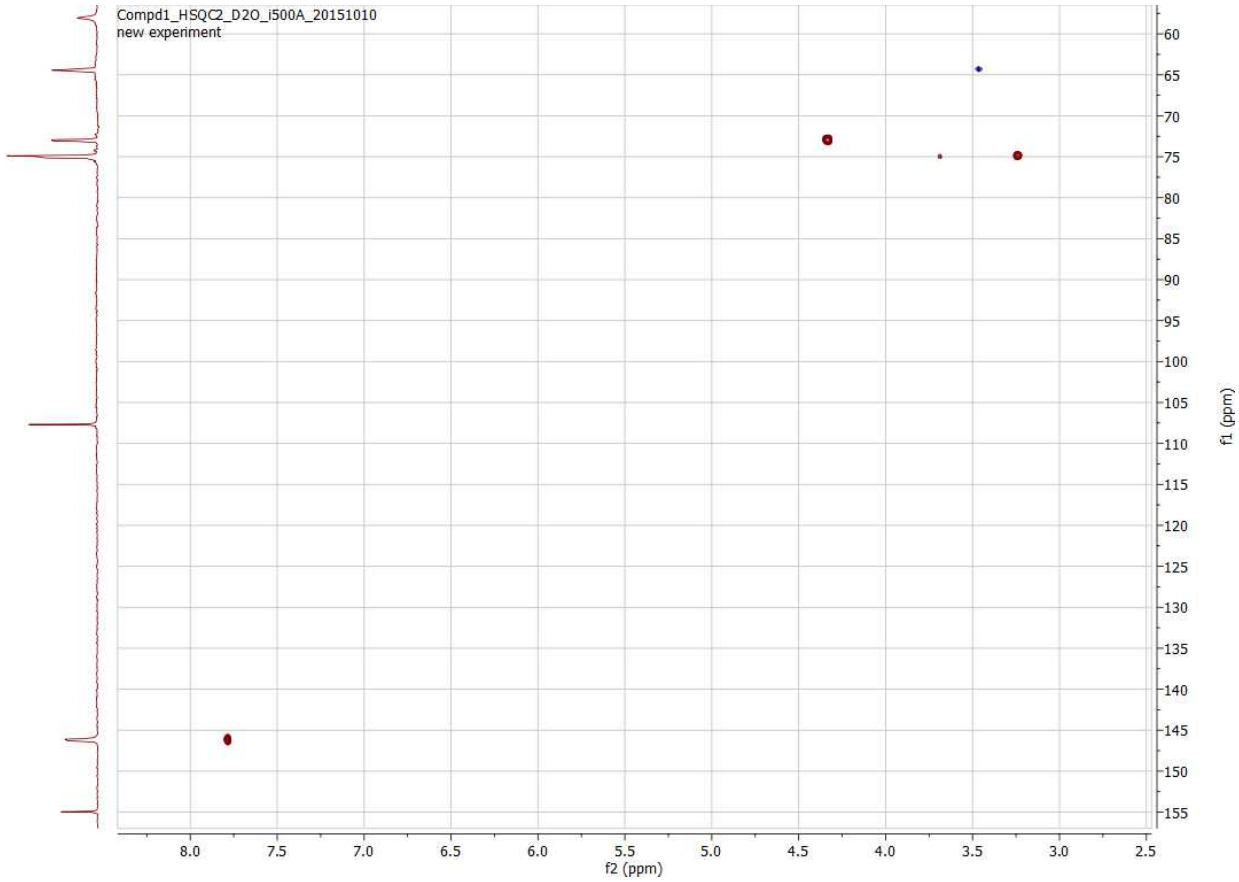


Figure S7. ^{1}H-NMR for **5**.

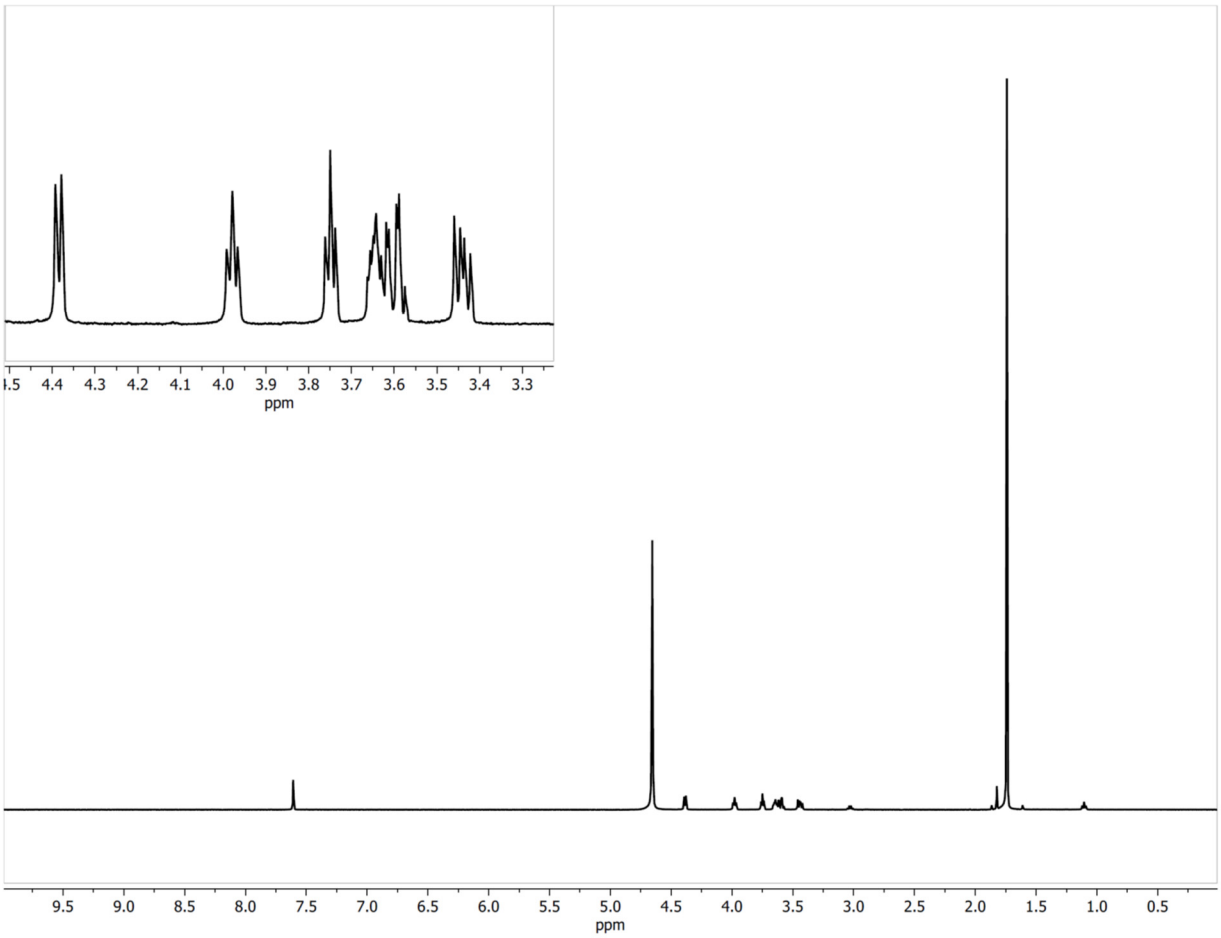


Table of ^{1}H-NMR Signals

Proton	δ ppm (D_2O)
1`	4.39
2`	3.98
3`	3.75
4`	3.65
5`A	3.58
5`B	3.44
6	7.61

There exists acetate as an impurity in the spectrum from the HPLC purification.

Figure S8. $^{1}H,^{1}H$-COSY for **5**.

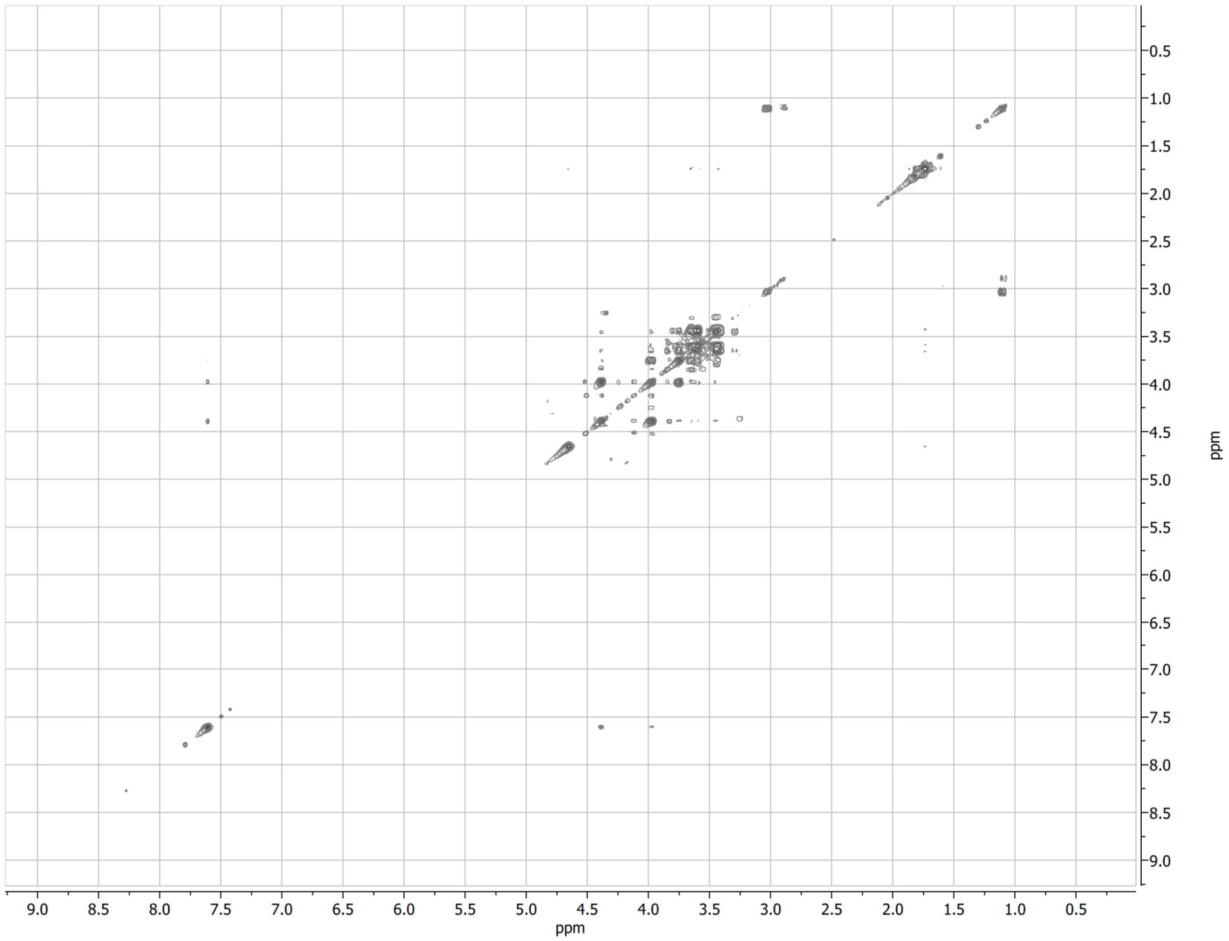


Figure S9. $^{1}H,^{1}H$-TOCSY for **5**.

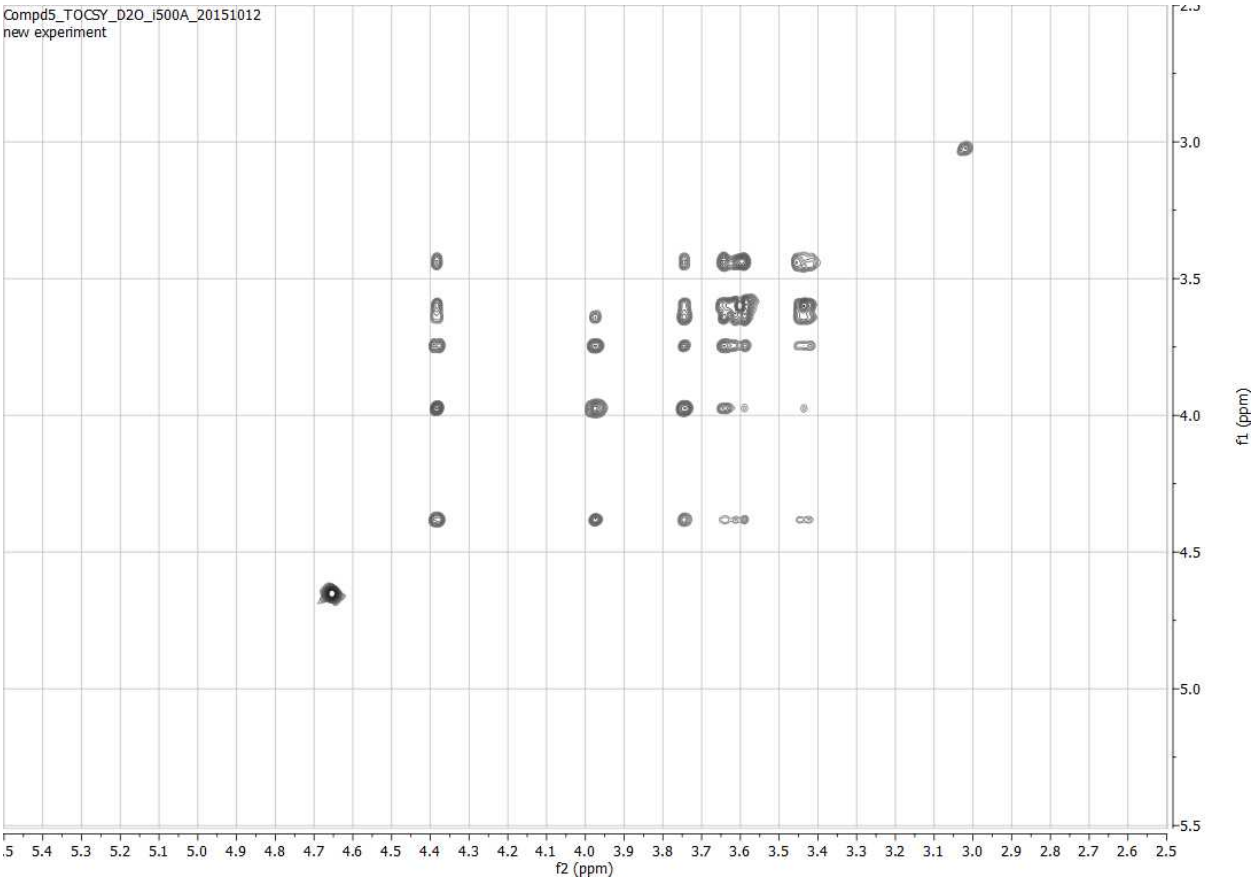


Figure S10. ^{13}C-NMR for **5**.

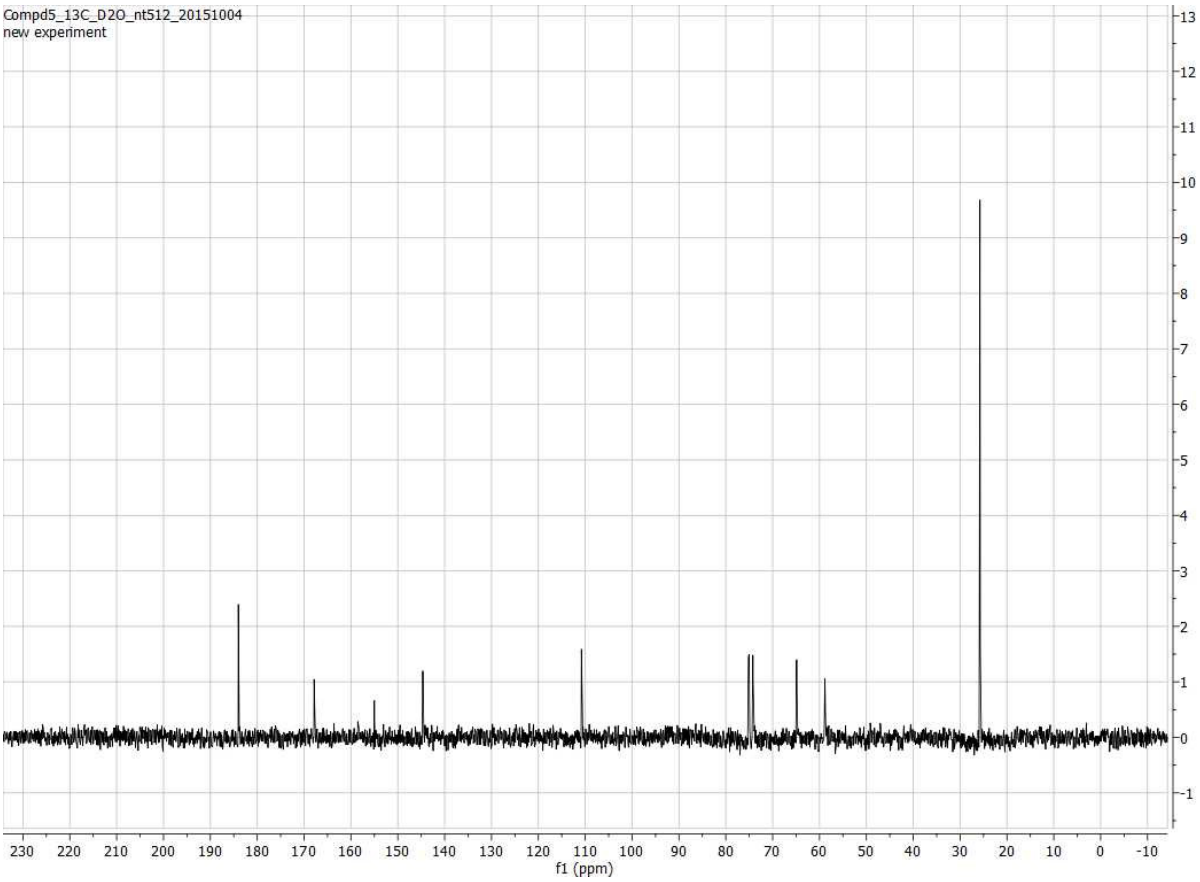


Table of ^{13}C-NMR Shifts

Proton	δ ppm (D_2O)
1`	58.8
2`	75.0
3`	75.1
4`	74.2
5`	64.8
2	154.9
4	167.9
5	110.8
6	144.7

There exists acetate as an impurity in the spectrum from the HPLC purification.

Figure S11. $^{1}H,^{13}C$-HSQC for **5**.

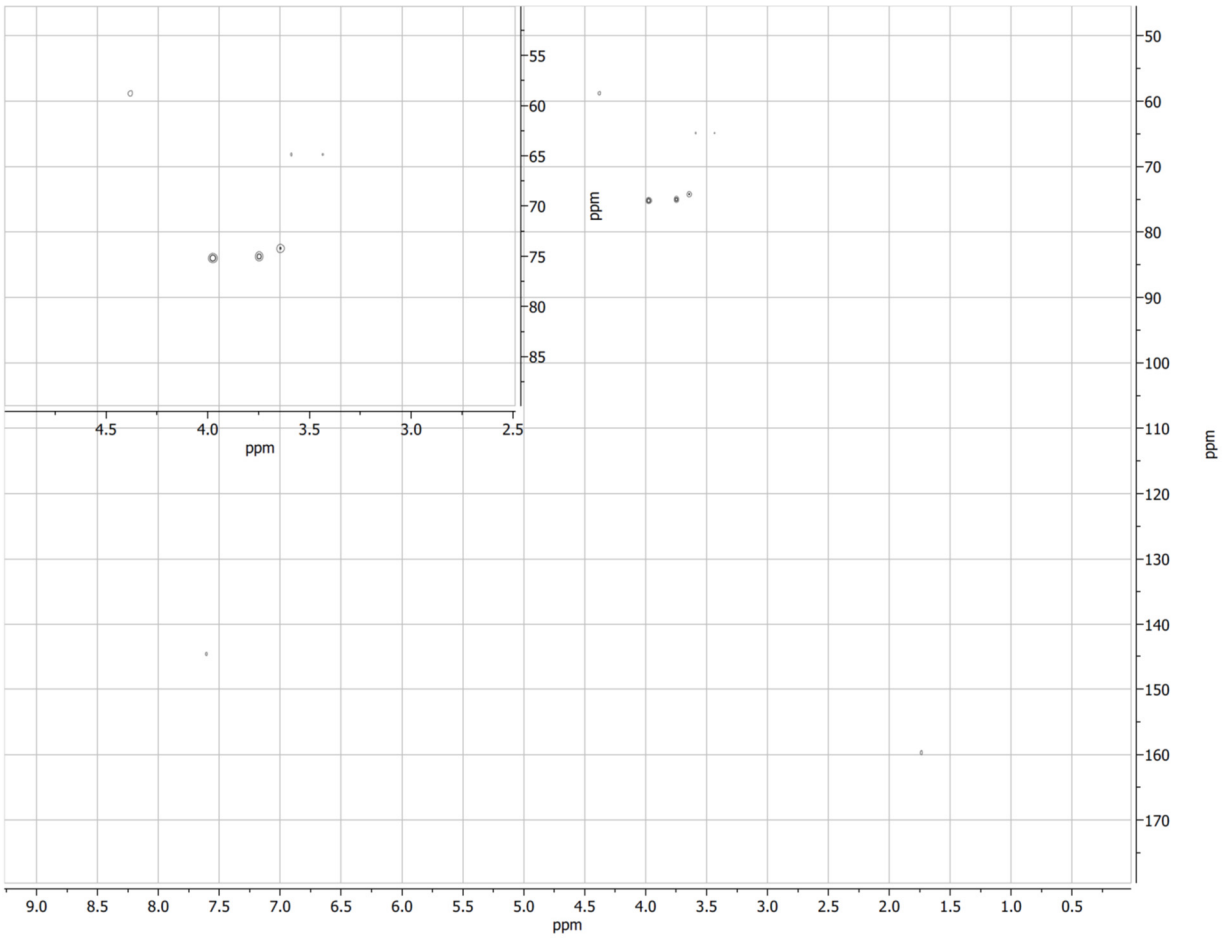


Figure S12. Raman spectra for **1** (C-S adduct, red) and **5** (C-O adduct, black).

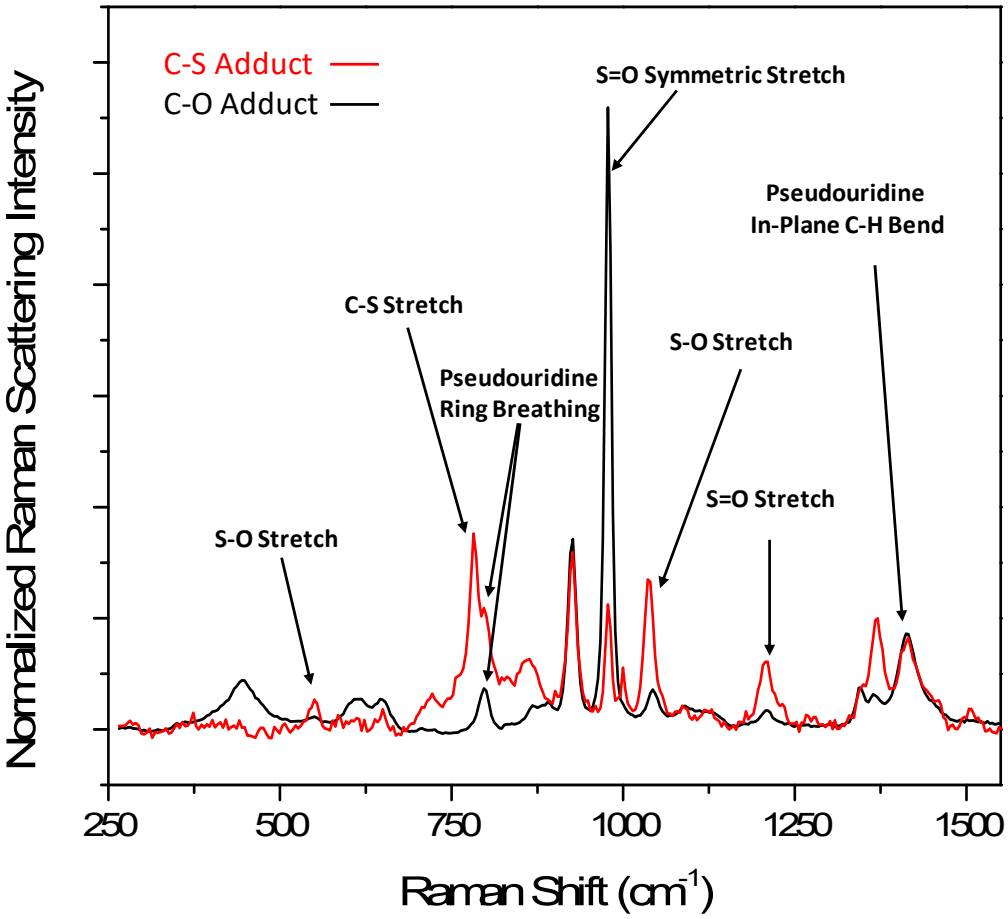
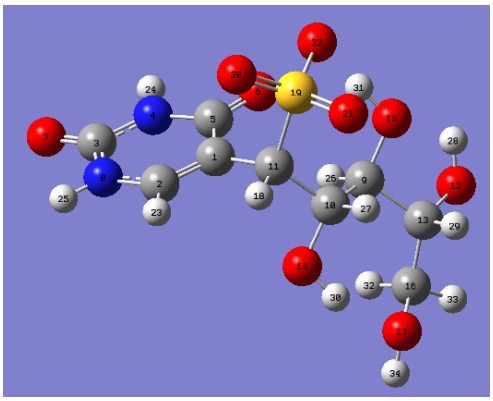


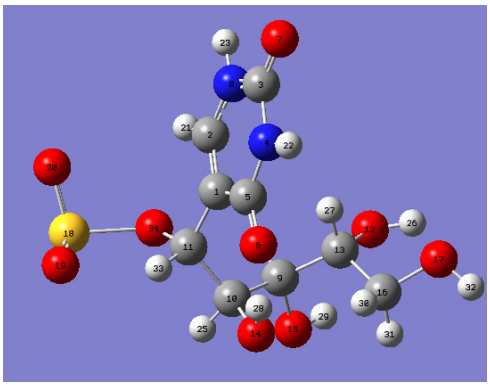
Figure S13. DFT optimized energy and coordinates for the (*S*) isomer of the *S* adduct



Energy = -1535.99655417 Hartees

Tag	Symbol	X	Y	Z
1	C	1.36687	-0.14771	-0.39613
2	C	2.30487	-0.47872	-1.32119
3	C	3.64951	-1.80327	0.1916
4	N	2.69777	-1.4618	1.12767
5	C	1.5481	-0.67174	0.95559
6	O	0.81166	-0.4935	1.91643
7	O	4.619	-2.50495	0.43333
8	N	3.38962	-1.26708	-1.0531
9	C	-1.82919	-0.30199	0.51349
10	C	-1.21068	0.17764	-0.81945
11	C	0.23363	0.75793	-0.81373
12	O	-3.80621	-0.94427	1.73728
13	C	-3.30788	-0.75474	0.41005
14	O	-1.11813	-0.89197	-1.76174
15	O	-1.89676	0.76214	1.46312
16	C	-3.59758	-2.0695	-0.30752
17	O	-3.63886	-1.83054	-1.72569
18	H	0.39614	0.96936	-1.87272
19	S	0.4173	2.49442	-0.15185
20	O	1.78523	2.90124	-0.57411
21	O	-0.66401	3.26918	-0.82188
22	O	0.27048	2.45297	1.33778
23	H	2.24869	-0.12992	-2.34424
24	H	2.84058	-1.8326	2.05923
25	H	4.05141	-1.48357	-1.78525
26	H	-1.23902	-1.13015	0.91724
27	H	-1.85161	0.9741	-1.2157
28	H	-3.40895	-0.24129	2.27463
29	H	-3.88026	0.04286	-0.0855
30	H	-2.01751	-1.21194	-1.97498
31	H	-1.01089	1.13209	1.64288
32	H	-2.84409	-2.82691	-0.06902
33	H	-4.57256	-2.42632	0.032
34	H	-3.84559	-2.65423	-2.18359

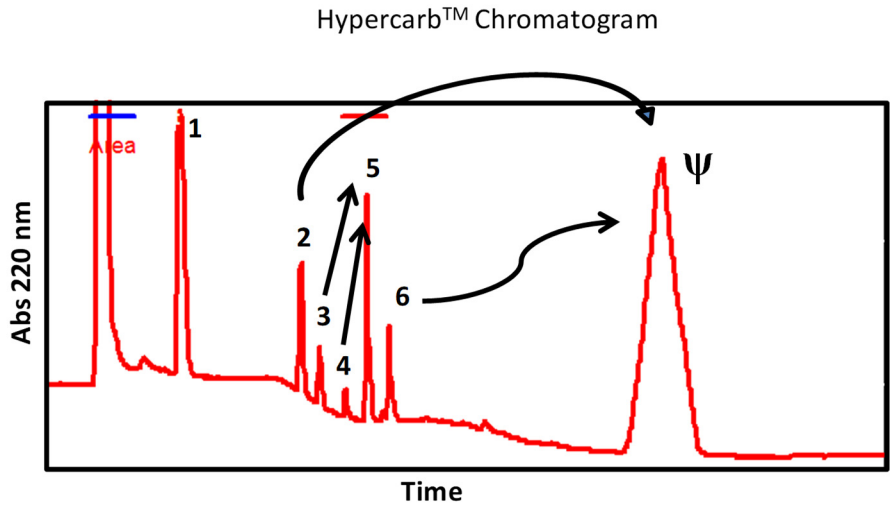
Figure S14. DFT optimized energy and coordinates for the (*R*) isomer of the *O* adduct



Energy = -1535.444450302 Hartrees

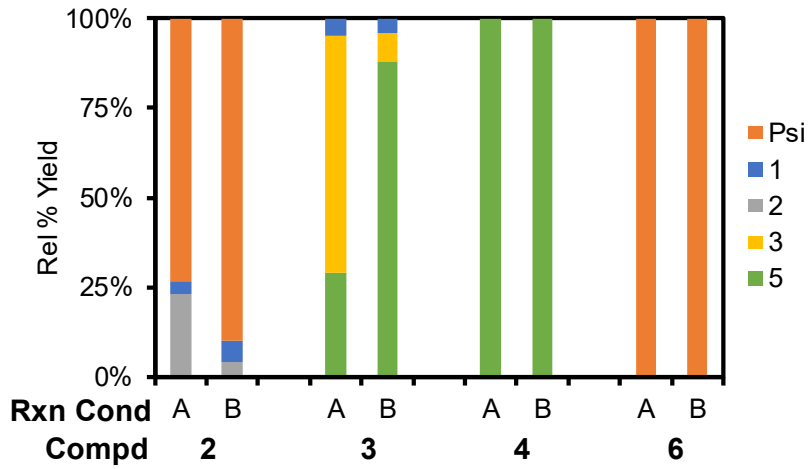
Tag	Symbol	X	Y	Z
1	C	-0.55819	0.77706	0.16462
2	C	-0.94657	1.38898	-0.98029
3	C	-0.14343	3.57497	-0.31315
4	N	0.23224	2.9408	0.85741
5	C	0.08755	1.59262	1.17489
6	O	0.50674	1.18766	2.26607
7	O	0.03214	4.7654	-0.51766
8	N	-0.73364	2.72091	-1.2195
9	C	1.26252	-1.76539	-0.53838
10	C	0.28925	-1.67581	0.6476
11	C	-0.8983	-0.68962	0.43256
12	O	2.99859	-1.00942	-1.9318
13	C	2.40252	-0.74339	-0.64787
14	O	0.92493	-1.51164	1.91439
15	O	1.80048	-3.0975	-0.5461
16	C	3.48719	-0.81671	0.42285
17	O	4.54774	0.0421	-0.03959
18	S	-3.49756	-1.54964	-0.2682
19	O	-3.58322	-1.31737	1.21648
20	O	-4.16577	-0.44965	-1.04239
21	H	-1.46369	0.84553	-1.75783
22	H	0.67691	3.52504	1.55547
23	H	-1.04518	3.1345	-2.08734
24	H	0.65618	-1.63298	-1.4433
25	H	-0.19069	-2.65742	0.68029
26	H	3.87225	-0.59068	-1.9242
27	H	1.98321	0.26946	-0.6599
28	H	0.84194	-0.58129	2.19784
29	H	2.42727	-3.14264	-1.28194
30	H	3.10988	-0.47388	1.38589
31	H	3.8519	-1.84169	0.52692
32	H	5.3541	-0.15062	0.45211
33	H	-1.471	-0.72035	1.37148
34	O	-1.69008	-1.1942	-0.61521

Figure S15. Decomposition pathways for **2**, **3**, **4**, and **6**.



- Peaks **2**, **3**, **4**, and **6** disappear on heating at pH 9 and 70 °C
- **2** and **6** lose bisulfite to give **Ψ**
- **3** and **4** go to **5**
- **1**, **5** and **Ψ** are the final products

Analysis of the desulfonation and Mg^{2+} dependency in intermediates **2**, **3**, **4**, and **6** degradation in the nucleoside context.

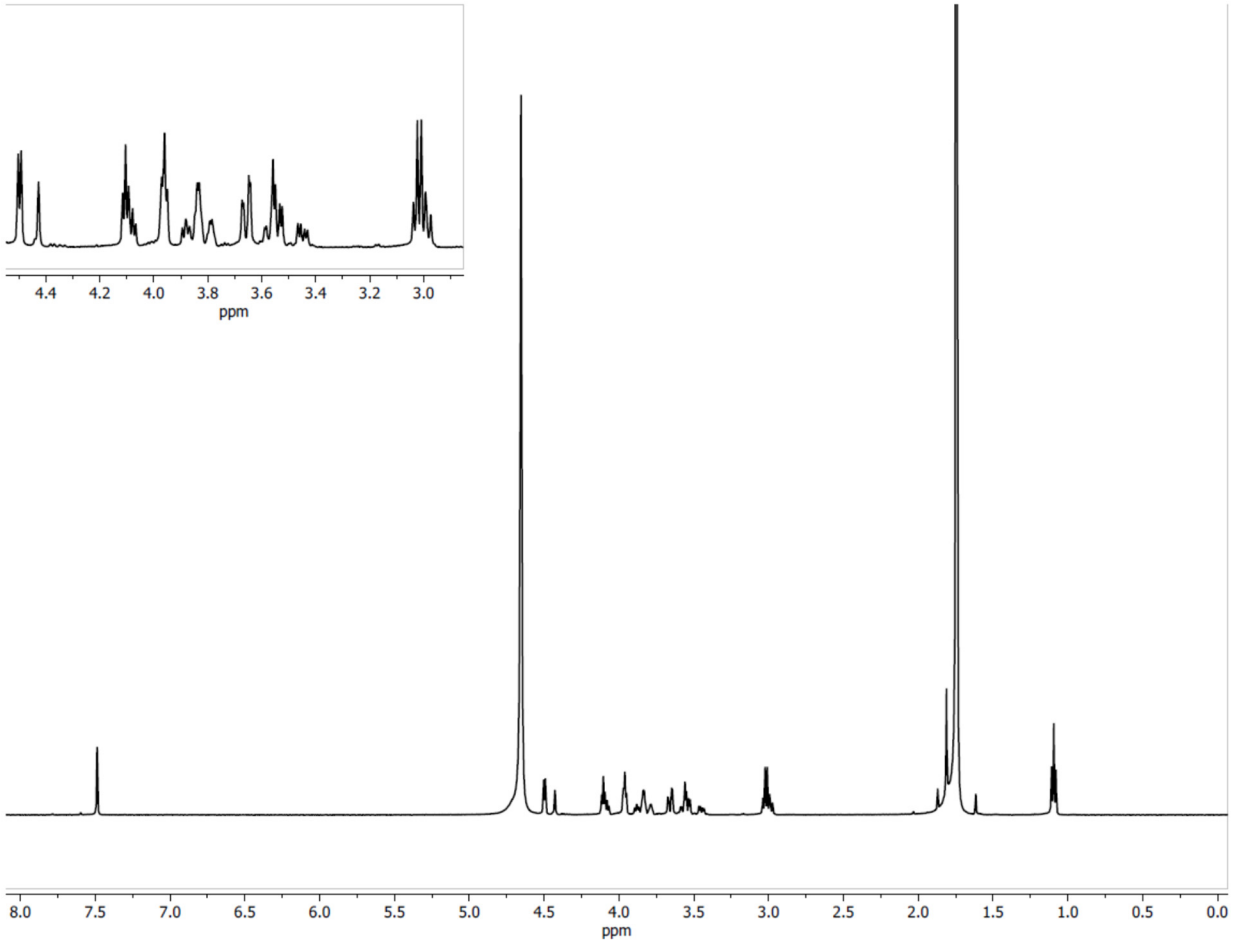


Reaction Conditions-

A = 70 mM Tris (pH 7.4), 10 mM $MgCl_2$, 37 °C, 2 h

B = 1 M Tris (pH 9), 37 °C, 2 h (established desulfonation conditions).

Figure S16. ^{1}H-NMR for **2**.

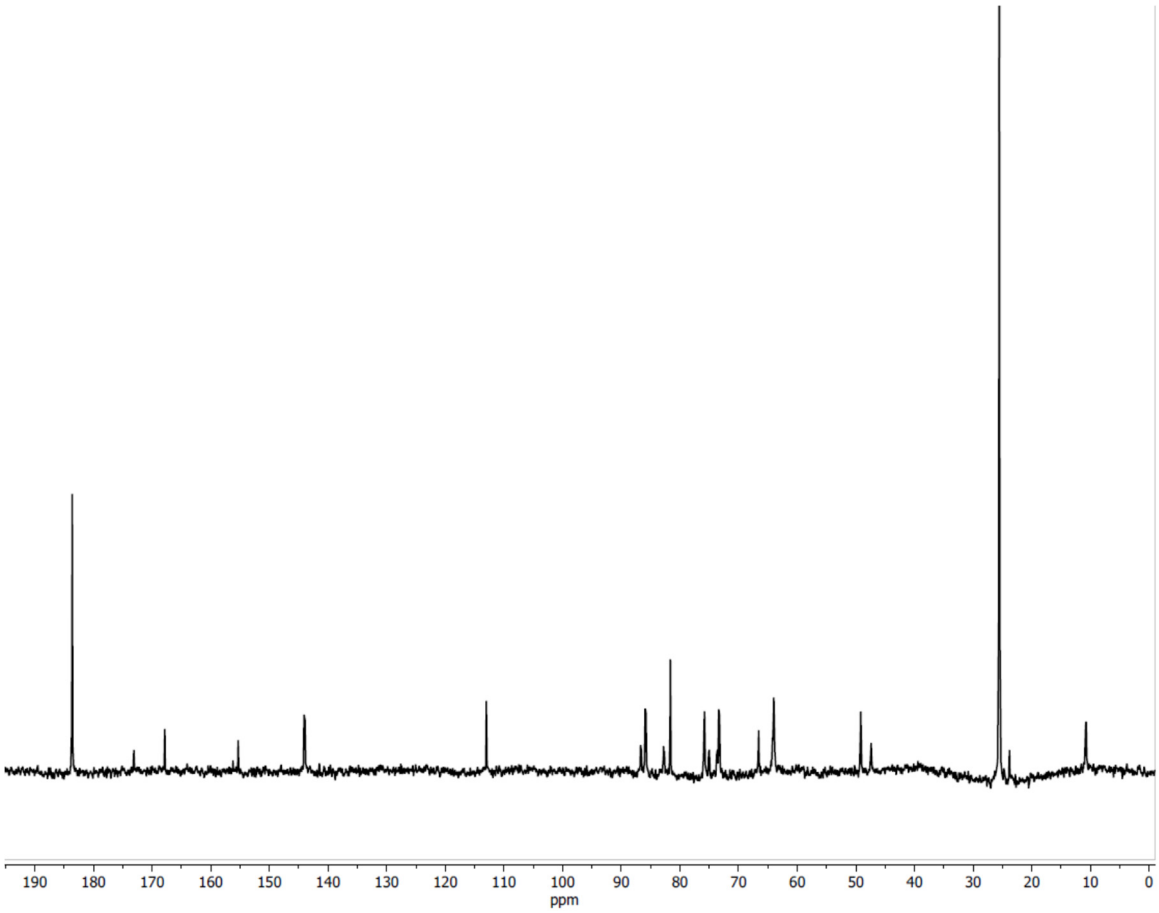


***This spectrum represents a mixture of **2** and its decomposition product **Ψ**. There exists acetate as an impurity in the spectrum from the HPLC purification.

Proton Resonances	
Pseudouridine	Compound 2
4.51	4.43
4.13	3.97
4.07	3.89
3.84	3.79
3.67	3.46
3.58	n.d.

Resonances for **Ψ** were reported in Wishard, D.S.; et al. *Nucleic Acids Res*. **2009**, *37*, D603, in addition to recording the spectra on our spectrometer. Lastly, n.d. = not detected that results from this peak most likely overlapped with a peak from **Ψ**.

Figure S17. ^{13}C-NMR for **2**.



*This spectrum represents a mixture of **2** and its decomposition product **Ψ**. There exists acetate as an impurity in the spectrum from the HPLC purification.

Carbon Resonances	
Pseudouridine	Compound 2
167.8	173.1
155.3	143.9
144.1	86.7
113.0	82.6
85.8	75.0
81.6	73.6
75.8	73.2
73.4	66.6
64.0	64.1

Resonances for **Ψ** were reported in Wishard, D.S.; et al. *Nucleic Acids Res*. **2009**, *37*, D603, in addition to recording the spectra on our spectrometer.

Figure S18. Anion-exchange HPLC analysis of bisulfite treated RNA.

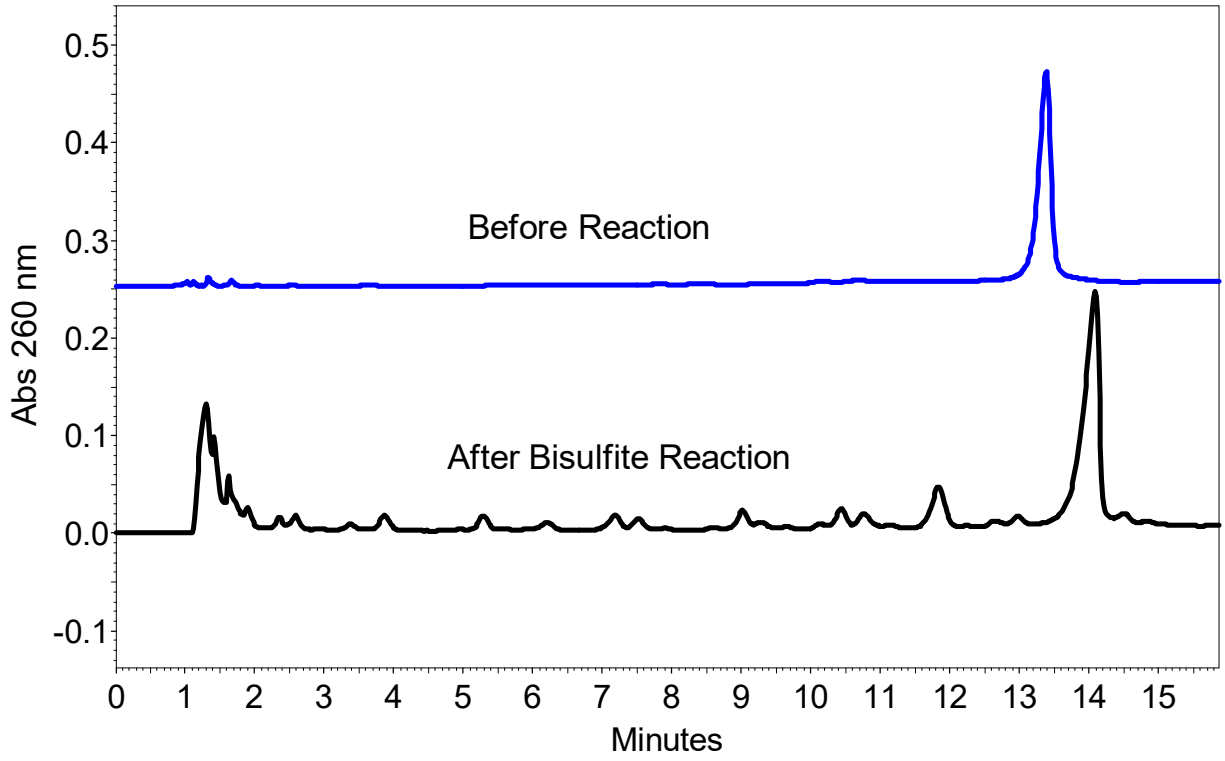
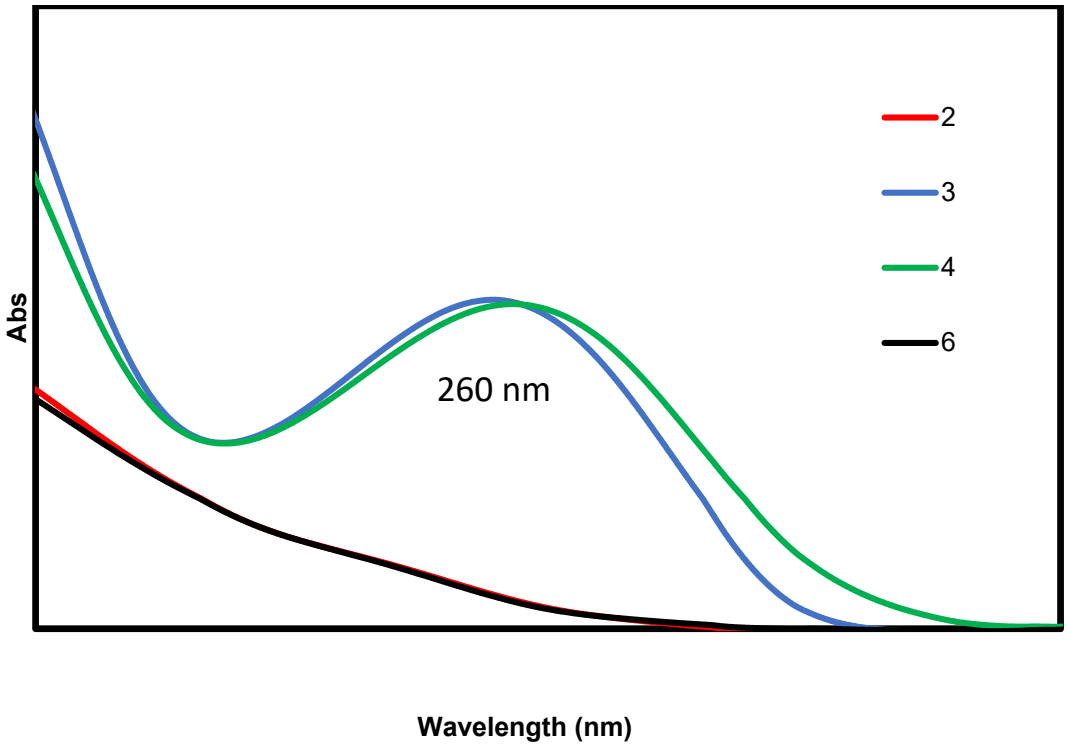
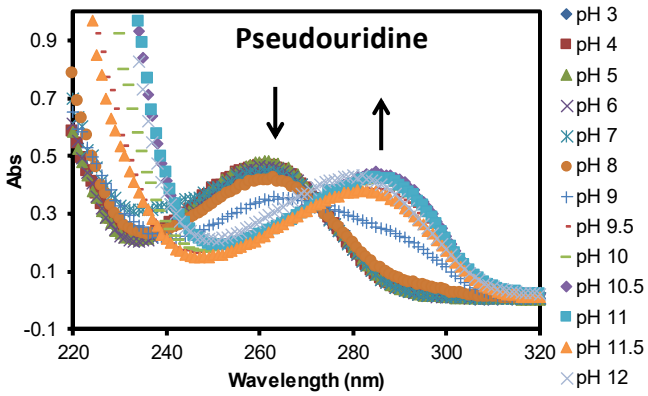


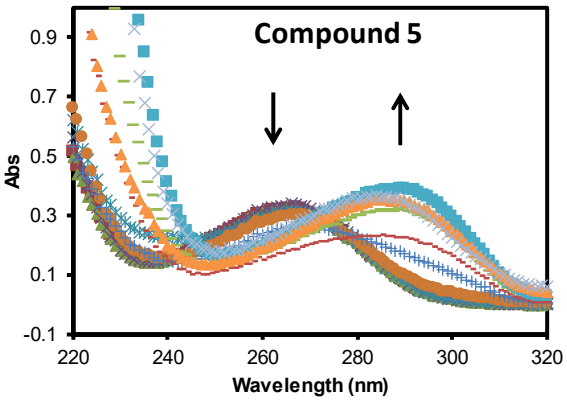
Figure S19. UV-vis spectra for **2**, **3**, **4**, and **6.**

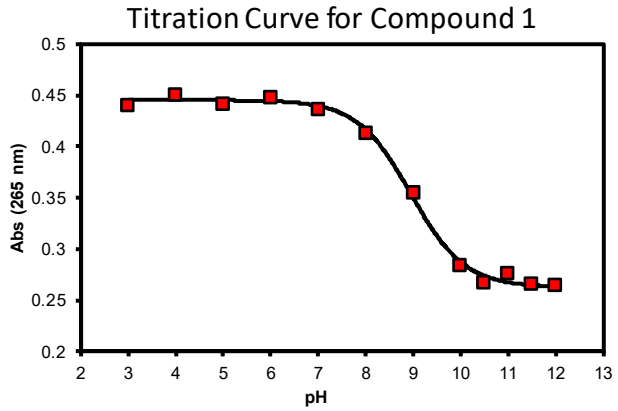


The UV-vis spectra were recorded during the HPLC analysis running the A line as 20 mM NH_4OAc (pH 7) and the B line as MeOH. The HPLC method and retention times are reported in the methods section for the compounds for readers that are interested in the exact solvent composition each compound was recorded.

Figure S20. pH titrations of **Ψ**, **1**, and **5** monitored by UV-vis.







Compound	pK_a (UV-vis)
Ψ	9.0 ± 0.1
Compound 1	9.1 ± 0.2
Compound 5	9.7 ± 0.4