Supporting Information for:

A General NMR-Based Strategy for the In situ Characterization of Sugar Nucleotide-Dependent Biosynthetic Pathways

Shanteri Singh,[†*] Pauline Peltier-Pain,[‡#] Marco Tonelli,[§] Jon S. Thorson[†*]

[†]Center for Pharmaceutical Research and Innovation, College of Pharmacy, University of Kentucky, 789 South Limestone Street, Lexington, Kentucky 40536; [‡] Pharmaceutical Sciences Division, University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI 53705, USA. [#]Present address: Glycom A/S, Denmark.[§] National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA.

*To whom correspondence should be addressed.E-mail: jsthorson@uky.edu, ssingh555@uky.edu

Table of Contents

1. General materials and methods. Unless otherwise stated, all general chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) or New England Biolabs (Ipswich, MA, USA). [U-^{13}C]glucose was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). *E. coli* BL21(DE3) competent cells were purchased from Invitrogen (Carlsbad, CA). The pET28a *E. coli* expression vector was purchased from Novagen (Madison, WI). Primers were purchased from Integrated DNA Technology (Coralville, IA). Pfu DNA polymerase was purchased from Stratagene (La Jolla, CA). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). PD-10 column and Ni-NTA superflow columns were purchased from GE Healthcare (Piscataway, NJ).

The sequence of enzymes involved in this study can be found within the National Center for Biotechnology Information (NCBI) protein database and include: *Micromonospora echinospora calS1* (genebank accession number AAM94768.1); calS2 (genebank accession number AAM94769.1); *calS3* (genebank accession number AAM94770.1); *calS13* (genebank accession number AF505622.1); *calS11* (genebank accession number AAM70337.1) and *Escherichia coli wecE* (genebank accession number AAG58073.1). DNA sequencing was conducted by the University of Wisconsin Biotechnology Center with the primers T7 promoter (5'-TAATACGACTCACTATAGGG) or T7 terminator (5'-GCTAGTTATTGCTCAGCGG). All enzymes were over-expressed in Luria-Bertani (LB) media and agar were supplemented with 50 μg mL^{-1} kanamycin.

NMR experiments were carried on a 600 MHz Varian (Palo Alto, CA) VNMRS spectrometer equipped with a z-axis gradient 5 mm HCN cold probe at the National Magnetic Resonance Facility at Madison (NMRFAM). High resolution mass spectra (HRMS) were determined on a Bruker MaXis ultra-high resolution quadrupole time of flight mass spectrometer by negative ionization electrospray with a source potential of 2800 V, drying gas at 200 °C flowing at 4 L min^{-1} and a nebulizing gas pressure of 0.4 bar. Samples were infused at 3 μL min^{-1} and spectra collected for 2 min. Routine TLC analyses were performed on aluminum TLC plates coated with 0.2 mm silica gel (from Sigma-Aldrich, St. Louis, MO, USA) and visualized at 254 nm or using a 5% solution of H_2SO_4 in EtOH followed by heating. Flash column chromatography was achieved on 40 – 63 μm, 60 Å silica gel (from Silicycle, Quebec, Canada).

2. Cloning, expression and purification

2.1 Cloning and expression: Genes *calS1 calS2 calS3 calS11* and *calS13* from the calicheamicin producer, *Micromonospora echinospora* LL6600 were amplified from the genomic DNA and *E. coli* gene *wecE* was amplified by colony PCR from the cells of *E. Coli* BL21(DE3) using primer containing NdeI (forward) and EcoRI (reverse) restriction sites. The PCR products were digested with restriction enzymes and ligated into pET28a vector to provide expression plasmids for proteins bearing *N*-terminal-His$_6$ fusions. The genes were expressed in a corresponding *E. coli* BL21(DE3) host in the presence of 50 μg mL^{-1} of kanamycin at 37 °C to an OD_{600} of ~0.6, at which point the temperature was lowered to 25 ºC and the gene expression induction was accomplished with 0.5 mM IPTG. Cells were allowed to continue to grow at 25 ºC for approximately 18 hours at 220 rpm.

2.2 Protein purification: The cells were harvested by centrifugation (6,000 x*g* at 4 °C for 20 min) and resuspended in buffer A (20 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 7.8). The cells were lysed via incubation with 1 mg mL^{-1} lysozyme (~50,000 U mg^{-1}; Sigma-Aldrich, St. Louis, MO, USA) for 30 min on ice followed by sonication (VirSonic 475; Virtis, Gardiner, NY; 100 W, 4 x 30 s pulses, ~1 min between pulses) on ice. Cell debris were removed by

centrifugation (10,000 x*g* at 4°C for 45 min) and purified via affinity chromatography (5 mL HiTrap Ni-NTA chelating column, GE Healthcare, Piscataway, NJ) using a standard linear gradient (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0 with a linear 10-500 mM imidazole gradient) using an AKTA Purifier 10 (GE Healthcare). Buffer exchange of pooled fractions containing the purified protein was accomplished using a PD-10 column (GE Healthcare) eluted with 25 mM sodium phosphate buffer, pH 7.5. Final purified protein was drop frozen in liquid nitrogen and stored at -80°C. Protein purity was confirmed by SDS-PAGE and protein concentration was determined using the Bradford Protein Assay Kit (from Bio-Rad, Hercules, CA, USA) using BSA as a standard. Small aliquots of protein were thawed for experiments and used fresh as required. Overproduction and purification of variant of *Streptomyces antibioticus* OleD, TDP16 (P67T/S132F/A242L/Q268V) was carried out as described before.[6] All enzymes used for this study retained the N-terminal-His_6 affinity tag (herein referred to simply as TDP16, CalS1, CalS2, CalS3, CalS11, CalS13 and WecE).

3. NMR spectroscopy: Enzyme reactions (500-750 µL final volume) were performed in 50 mM sodium phosphate buffer, pH/pD 7.5 (10-20% D_2O). NMR spectra were collected at 25 ºC in 5 mm NMR tubes on a 600 or 800 MHz Varian (Palo Alto, CA) VNMRS spectrometers equipped with a z-axis gradient 5 mm HCN cold probe.The probe was tuned and shimmed before the addition of enzyme. Formation of new signals corresponding to product formation monitored by recording sensitivity enhanced 2D-^{1}H-^{13}C-HSQC (referred simply as HSQC henceforth) after the addition of the enzyme. For the rate measurement, the first spectrum was acquired beginning 4 min after the addition of enzyme to the reaction mixture, due to spectrometer setup requirements and a set of 40-80 sequential HSQC spectra (2048 × 256 complex data points) were acquired over a 7-18 h time period. At the completion or near completion of each enzyme reaction, ^{1}H-^{13}C 2D planes of HCCH COSY (1024 × 256 complex data points) and/or HCCH TOCSY(1024 × 256 complex data points; with a spin lock field of 7.5 kHz and 24 ms) were recorded for resonance assignments. Processing of the spectra were accomplished using the NMRPipe software package[15] or MestReNova (Santiago de Compostela, SPAIN) and the proton axis was referenced to the water resonance at 4.766 ppm. Integration of area under 2D HSQC signals to calculate the rate of the reactions were performed using the NMRPipe software package.[15] All putative products were also subsequently confirmed by HRMS (**see general methods**).

4. *In-vitro* enzyme reactions.

4.1. Preparation of *S*-[^{13}C-methyl]*S*-adenosyl-L-methionine ($^{13}CH_3$-SAM)*:*** Isotopically-labeled $^{13}CH_3$-SAM was prepared using 2.0 mM [^{13}C]methionine, 1.2 equivalent of ATP and 50 µg *E. coli* methionine adenosyltransferase,[11] in 25 mM NaH_2PO_4 buffer, 5 mM $MgCl_2$, 40 mM KCl, pH 8.0 at 37 °C for 1 h. Under these conditions, most of the substrate was converted to the desired product as observed by HSQC (**Figure S1**). The crude reaction mixture was lyophilized and dissolved in 50 µl water and directly used in the subsequent CalS11 reaction (**section 4.6.**).

4.2. TDP16 reaction. TDP16 reaction was initiated directly in a 5 mm NMR tube in 500 µl final volume (20 % D_2O) using standard assay conditions with 2.2 mM TDP and 2 mM **4**. A total of 80 sequential HSQC experiments were recorded for rate measurement following the addition of 0.5 µM of TDP16. At the end of 80 HSQC experiments, an additional 2 µM TDP16 was added and the reaction was incubated at 37 °C for 1 h to ensure completion of the reaction. Additional HSQC and 2D ^{1}H-^{13}C plane of HCCH -TOCSY experiments were recorded for the resonance assignments.

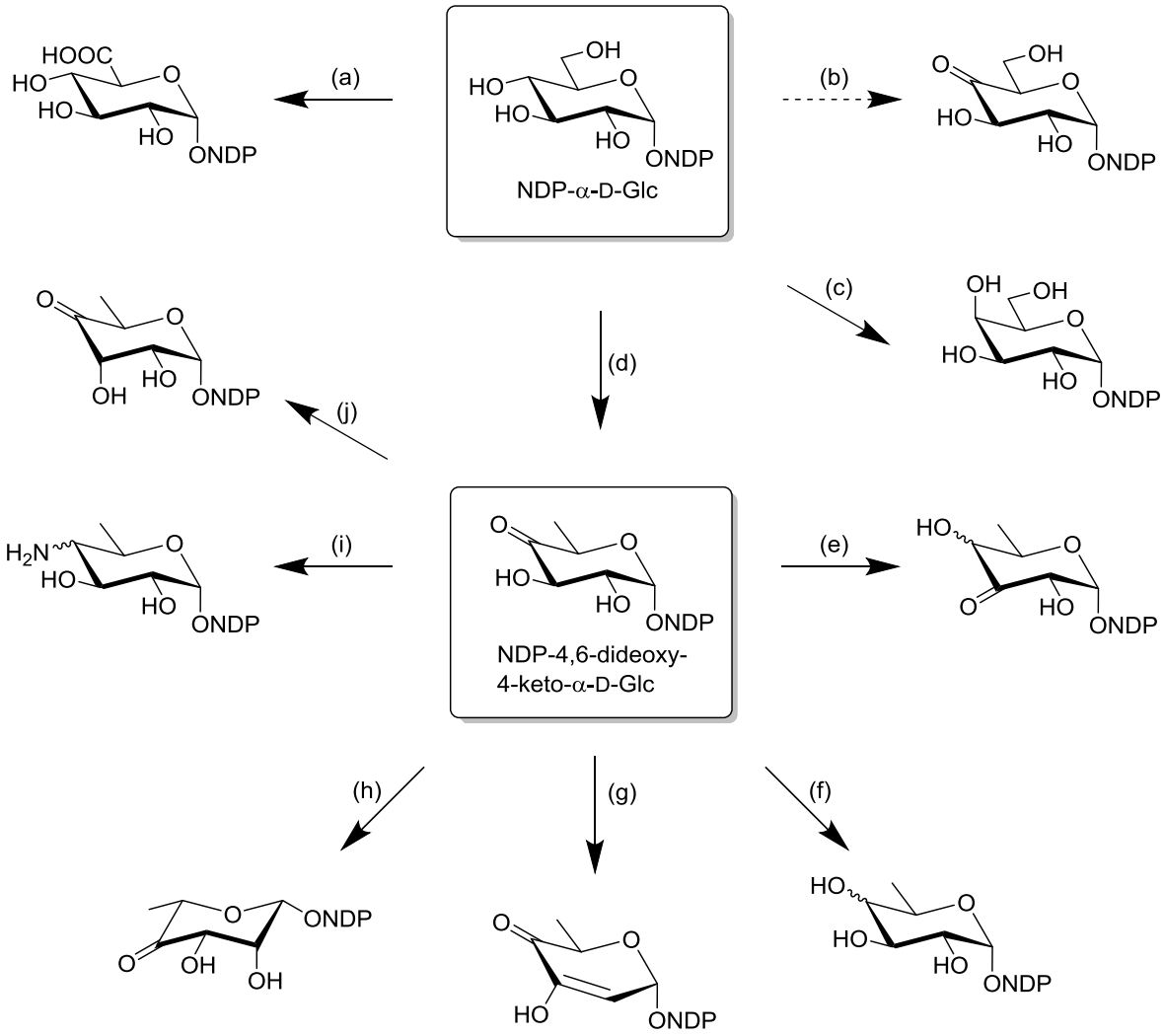
4.3. CalS3 reaction. To the reaction mixture from **section 4.2** was added 11.5 µM of CalS3 and 0.2 mM NAD^+ in a final reaction volume of 600 µl. A total of 40 sequential HSQC experiments were recorded for the rate measurement following the addition of CalS3. Overnight incubation of the reaction at 37 °C ensured ~90 % turnover to **6**. Additional HSQC and 2D ^{1}H-^{13}C plane of HCCH -TOCSY experiments were recorded for the resonance assignments.

4.4. CalS13 and WecE reaction. To the reaction mixture from **section 4.3** was added 200 µM PLP, 5 mM L-glutamate and 14 µM of CalS13 or WecE in a final volume of 700 µL directly in the 5 mm NMR tube, incubated the reaction at 37 °C for 4 h. The reaction catalyzed by CalS13 and WecE were partially complete (~50-70 %) even after incubating the reaction overnight at 37 °C. Additional HSQC and 2D ^{1}H-^{13}C plane of HCCH-TOCSY experiments were recorded for the resonance assignments.

4.5. CalS1 and CalS2 reaction. To the reaction mixture from **section 4.3** was added 9.5 µM of CalS1 (final volume 600 µL) incubated at 30 °C and three HSQC experiements were recorded at intervals of one hour. Since there was no change in the HSQC spectra after ~3 h of the addition of CalS1, 4 mM NADPH and 6.2 µM of CalS2 was added (final volume 650 µL) to the same reaction. HSQC experiments were recorded at regular intervals (1–2 h) to monitor the formation of product. The reaction was ~50 % complete after the overnight incubation at 30 °C as calculated by the ratio of the product to substrate signals from the HSQC. The 2D ^{1}H-^{13}C planes of HCCH-TOCSY and HCCH-COSY experiment were recorded for resonance assignments.

4.6. CalS11 reaction. To the reaction mixture from **section 4.5** was added 50 µL $^{13}CH_3$-SAM (prepared as described in **section 4.1.**) and 9 µM CalS11 in a final volume of 750 µL and incubated at 37 °C for 6 h. The progress of the reaction was monitored by recording HSQC every 2 h. NMR chemical shift assignments were made using 2D ^{1}H-^{13}C planes of HCCH-TOCSY and HCCH-COSY.

4.7. Rate measurement of TDP16 and CalS3. Rate determination for TDP16 and CalS3 catalyzed reactions was performed in triplicate by recording a series of sequential HSQC experiments (number of transients = 2 and number of increments = 80). Data processing and peak integration were performed using the NMRPipe software.[15] A calibration curve to convert the intensity of integral to concentration units was generated using standards of **4** and **5**. Using the plot of the integral of the anomeric peak against the concentration of the standards (0.2, 0.4, 0.6, 1.2, 1.8 and 2 mM) (**Figure S2**), the concentrations of product formation in TDP16 and CalS3 catalyzed reactions were calculated. The slope obtained by linear regression of the product formation versus time was used to calculate the rate of the reaction (**Figure 3**).



Scheme S1. Representative sugar nucleotide dependent transformations.
(a) dehydrogenation (b) oxidation (c) epimerization (d) 4,6-dehydration (e) 3,4-ketoisomerization (f) ketoreduction (g) 2,3-dehydration (h) 3,5-epimerization (i) transamination (j) 3-epimerization.

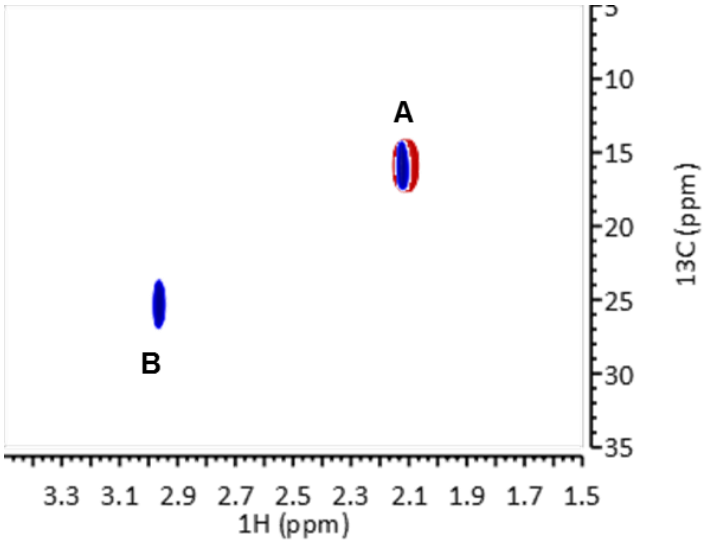


Figure S1. HSQC (600 MHz of ^{1}H) **of enzymatic formation of $^{13}CH_3$-SAM** . Overlay of region of 2D-^{13}C-^{1}H HSQC spectrum consisting of *S*-CH_3 groups of *S*[^{13}C-methyl]-L-methionine (**A**) and the corresponding ^{13}C-methylated product ($^{13}CH_3$-SAM, **B**).

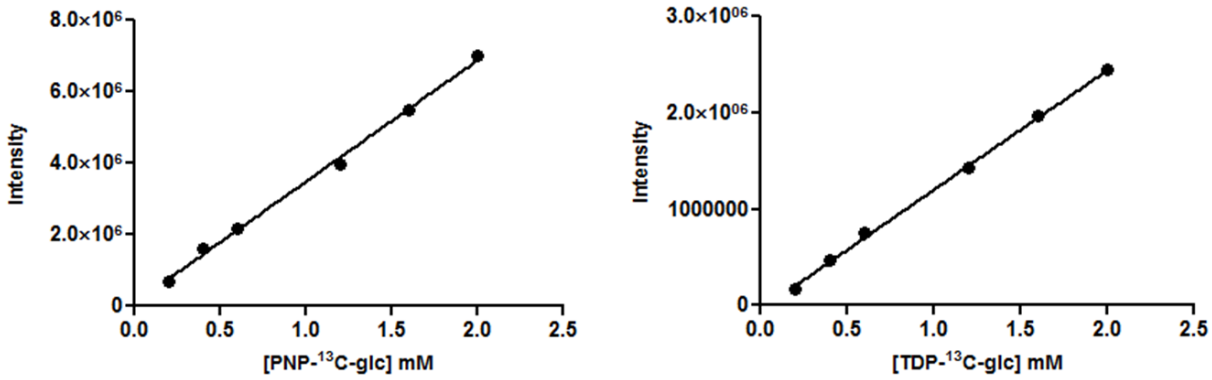
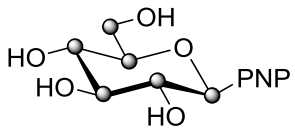


Figure S2. Calibration curve obtained from integrating the H1'-C1' peak from ^{1}H-^{13}C-HSQC of the standard concentrations of **4** (left) and **5** (right).

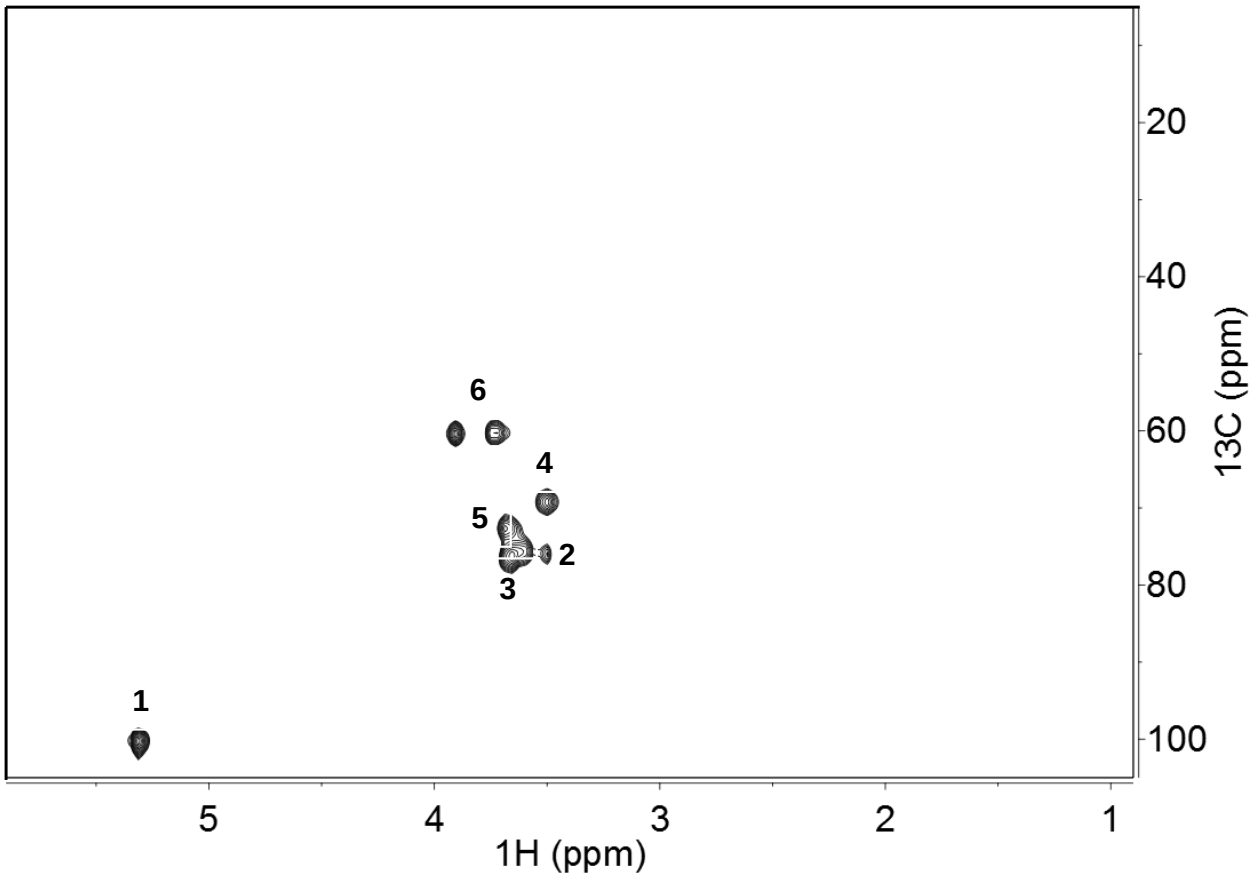
Table S1. ^{1}H and ^{13}C chemical shift values (in ppm) of **4**, **5** (product of reaction catalyzed by TDP16), **6** (product of reaction catalyzed by CalS3), **10** (product of reaction catalyzed by CalS13), **11** (product of reaction catalyzed by WecE), **8** (product of reaction catalyzed by CalS2), **9** (product of reaction catalyzed by CalS11) .[a]

Atom	4	5	6	10	11	8	9
H1', C1'	5.32, 101.95	5.57, 97.24	5.41, 98.09	5.38, 97.95	5.46, 97.89	5.07, 98.36	5.05, 98.27
H2', C2'	3.63, 77.21	3.49, 74.04	3.47, 73.41	3.46, 74.72	3.65, 75.47	3.94, 73.72	4.19, 69.20
H3', C3'	3.69, 78.95	3.76, 75.25	3.61, 76.10	3.70, 71.92	3.49, 70.47	3.49, 73.26	3.17, 84.02
H3", C3"[b]	NA	NA	NA	NA	NA	NA	3.26, 58.48
H4', C4'	3.52, 71.51	3.44, 71.49	NA	2.76, 60.01	3.44, 58.16	3.23, 75.20	3.23, 73.14
H5', C5'	3.71, 74.45	3.87, 75.45	3.94, 72.47	4.06, 69.18	4.00, 66.82	3.30, 75.67	3.28, 75.34
H6a',H6b', C6'	3.92, 3.75, 62.72	3.83, 3.77, 62.27	1.06, 14.03	1.17, 19.64	1.11, 18.14	1.16, 19.60	1.14, 19.22

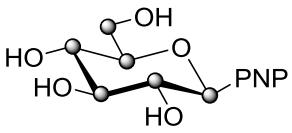
[a] Chemical shifts of only sugar part of the molecule
[b] H3", C3" correspond to methoxy group at position 3 of methoxy on rhamnose



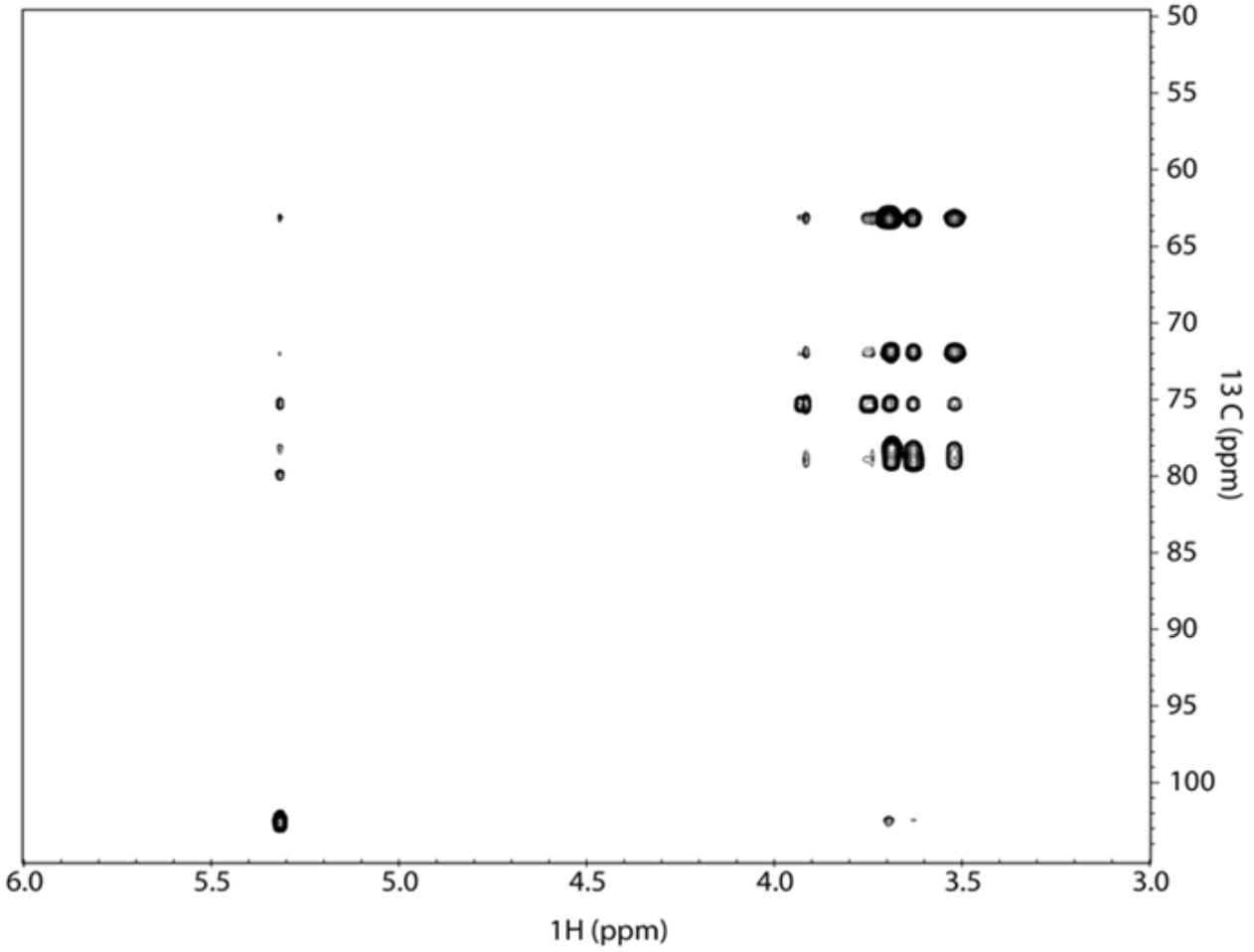
4



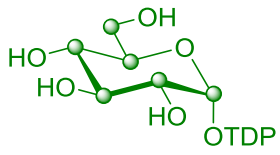
2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of **4** with resonance assignments.



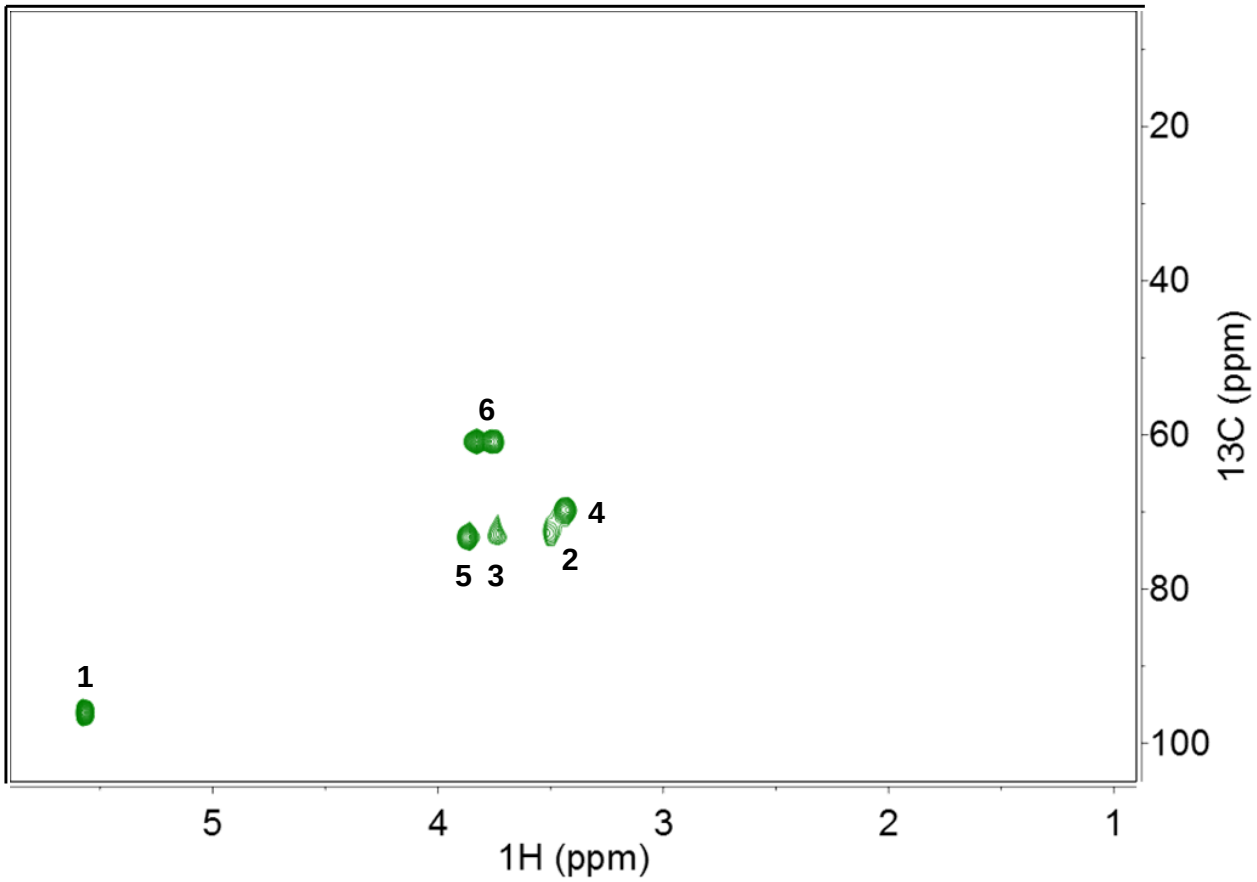
4



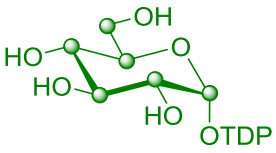
^{1}H-^{13}C 2D plane of HCCH -TOCSY (600 MHz of ^{1}H) of **4**.



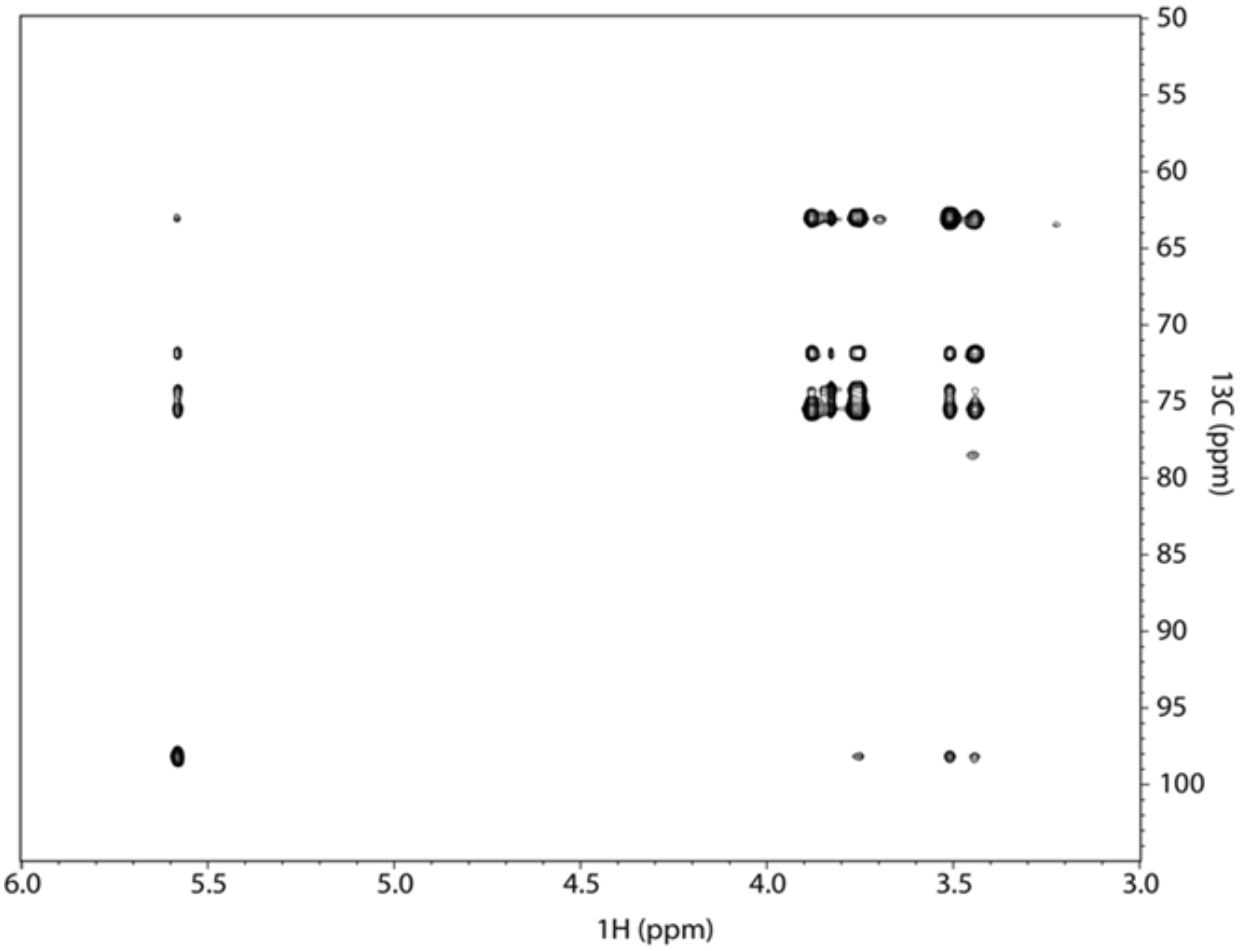
5



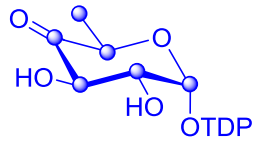
2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H)of **5** with resonance assignments.



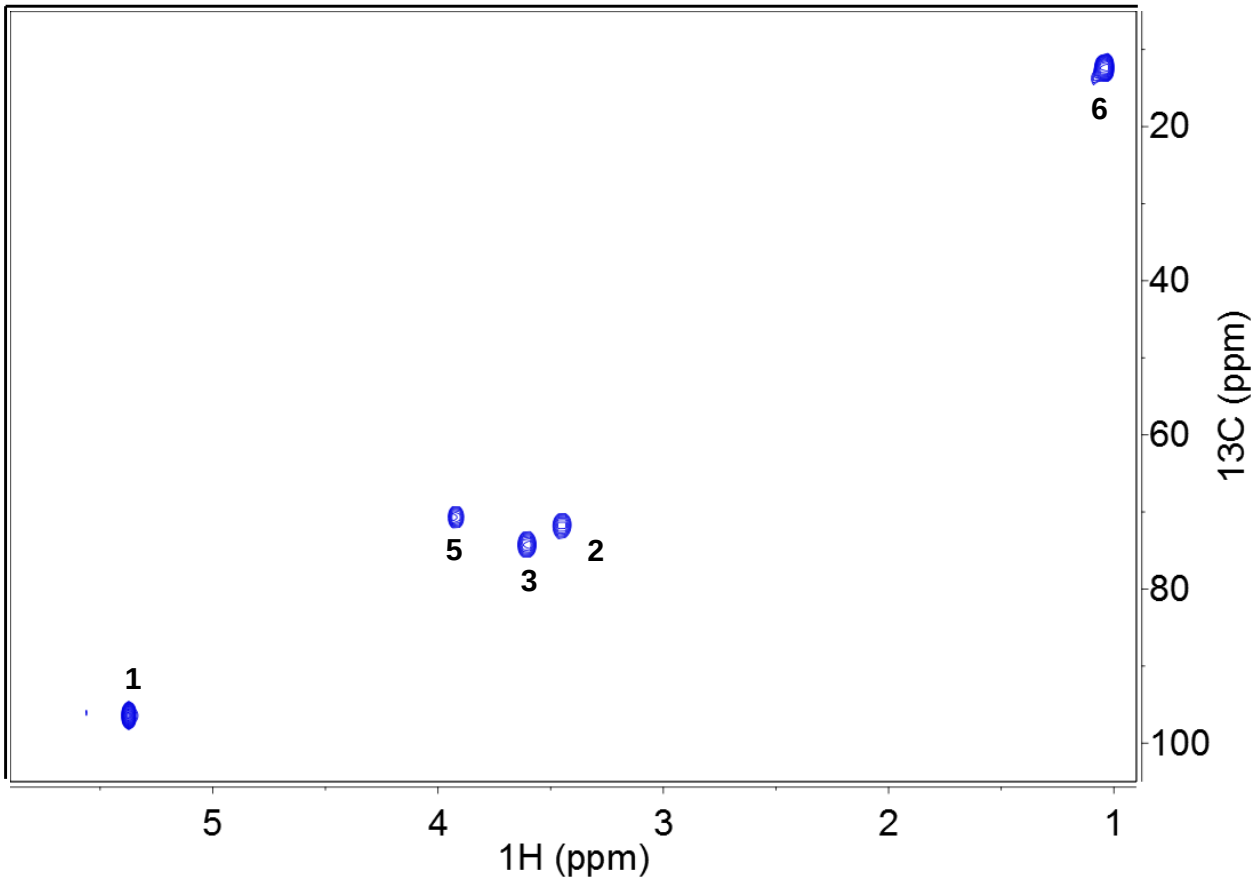
5



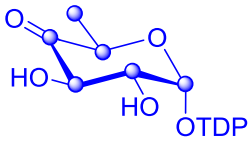
^{1}H-^{13}C 2D plane of HCCH -TOCSY (600 MHz of ^{1}H) of **5**.



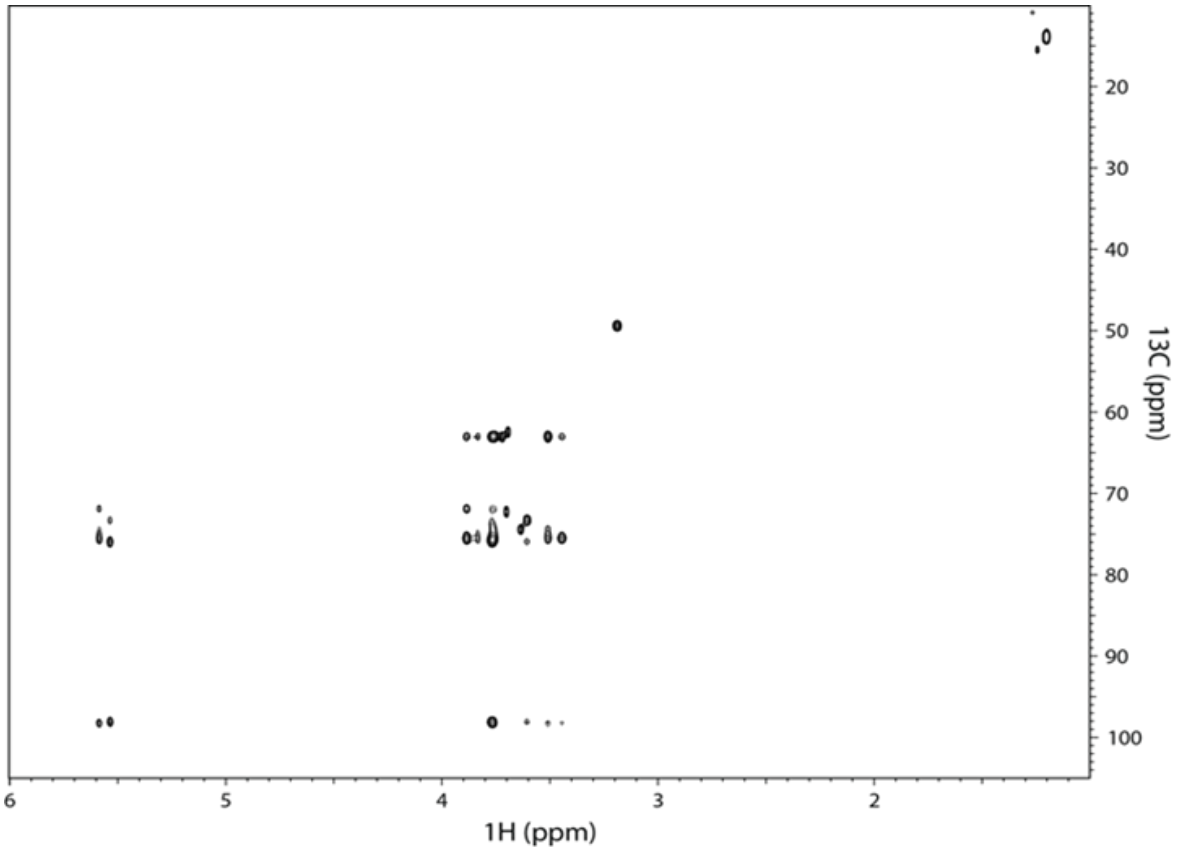
6



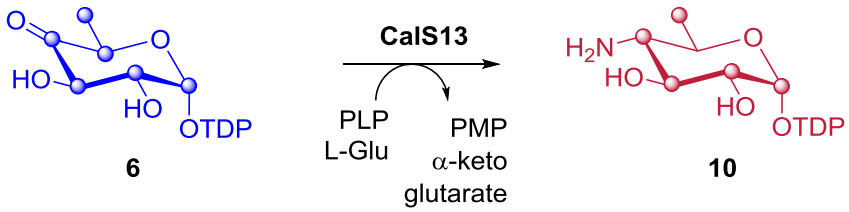
2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of **6** with resonance assignments.

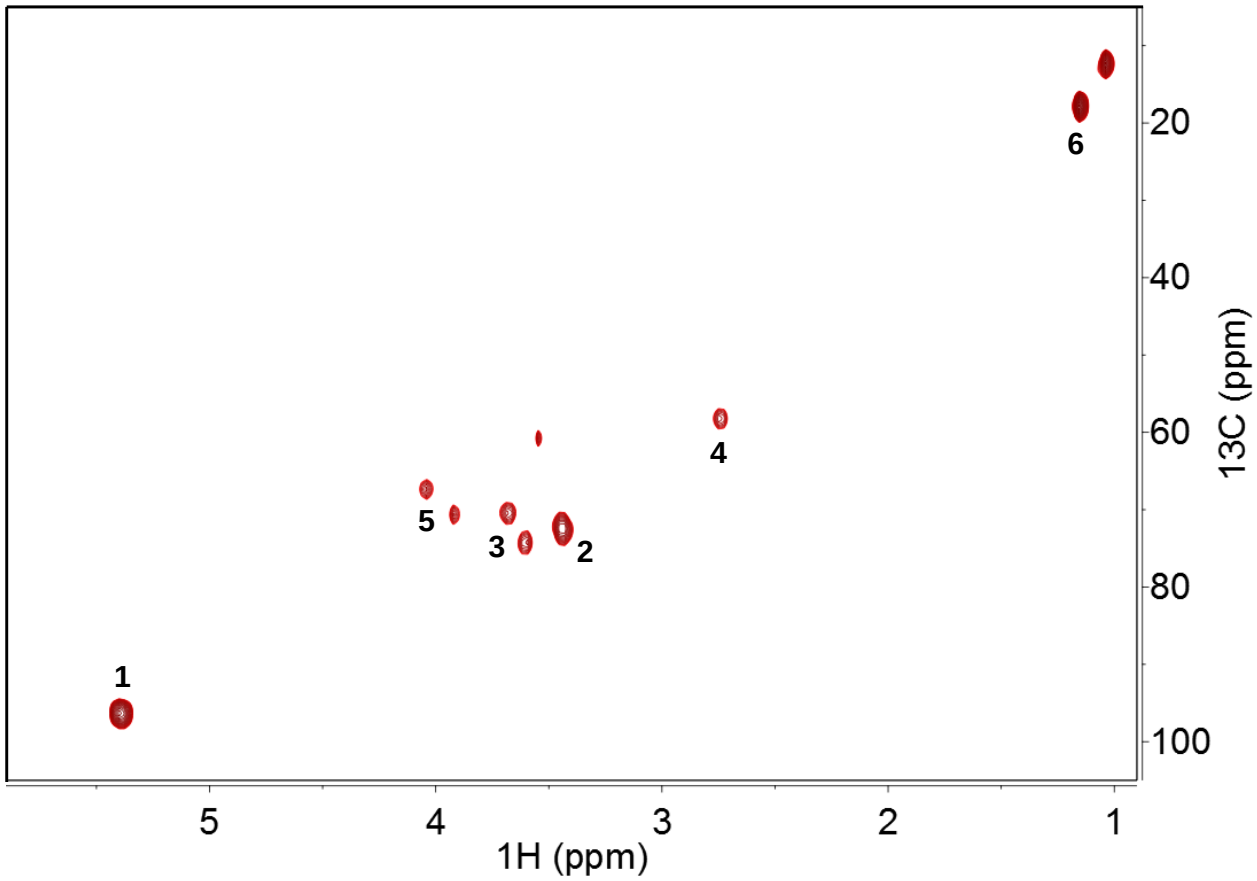


6

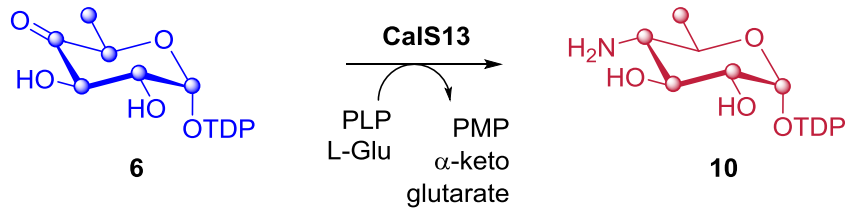


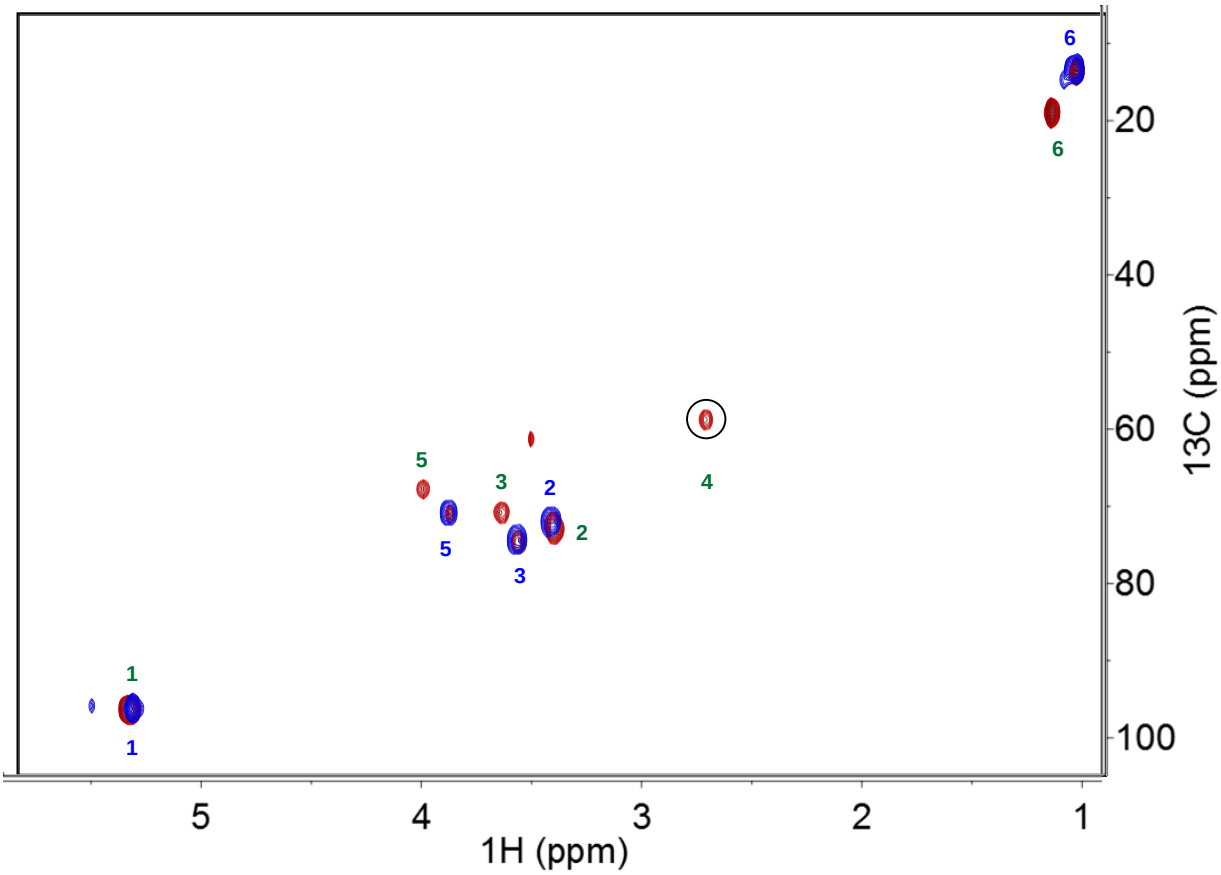
^{1}H-^{13}C 2D plane of HCCH -TOCSY(600 MHz of ^{1}H) of reaction mixture containing **5** and **6**.



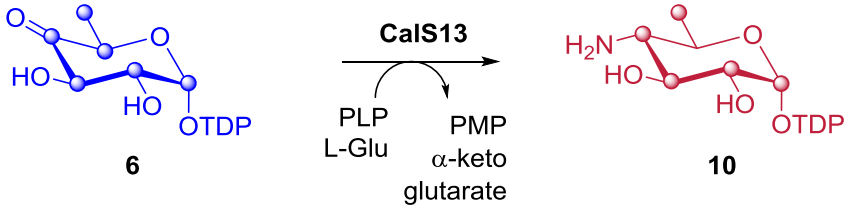


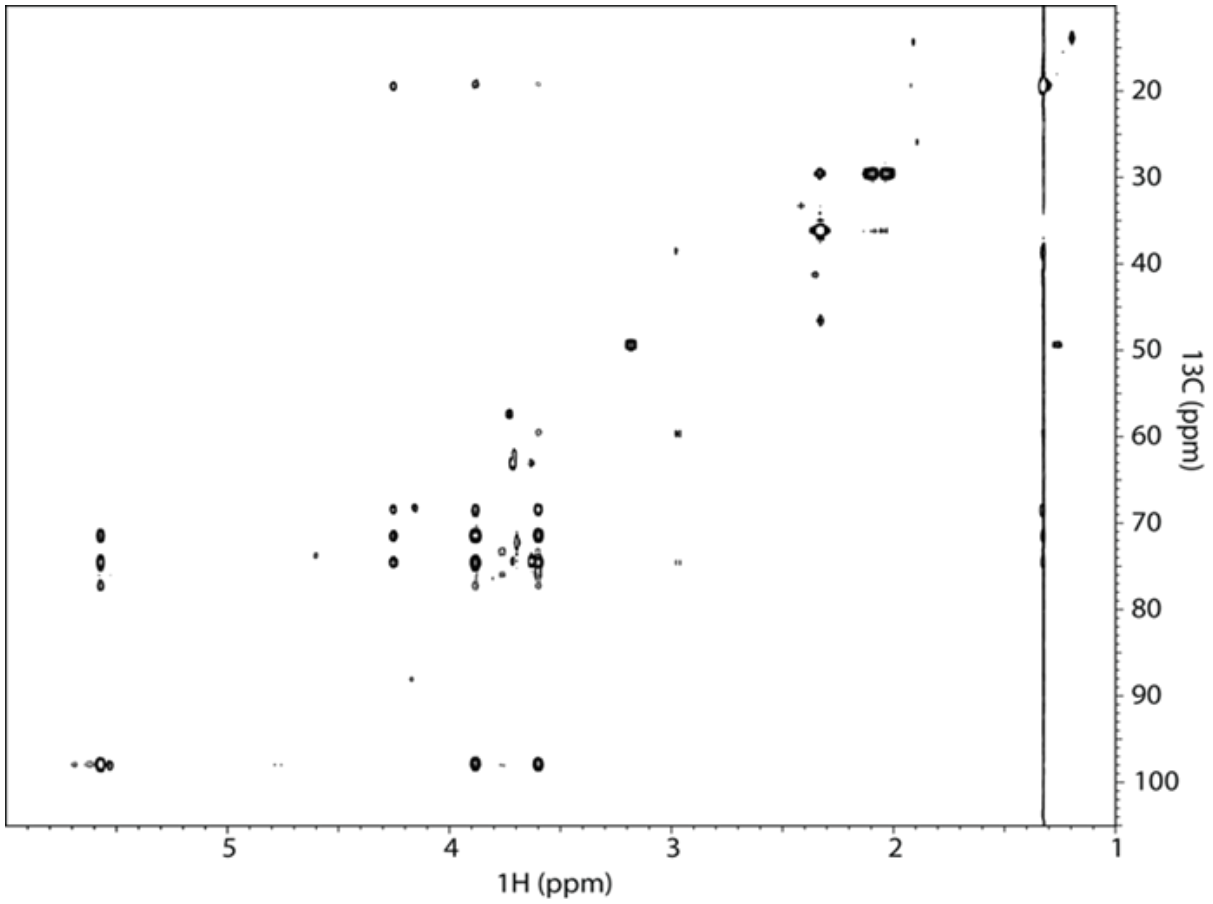
Resonance assignment of **10** in 2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of reaction mixture containing **6** and **10**.



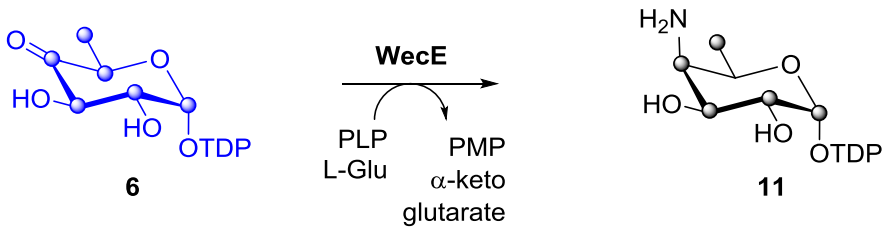


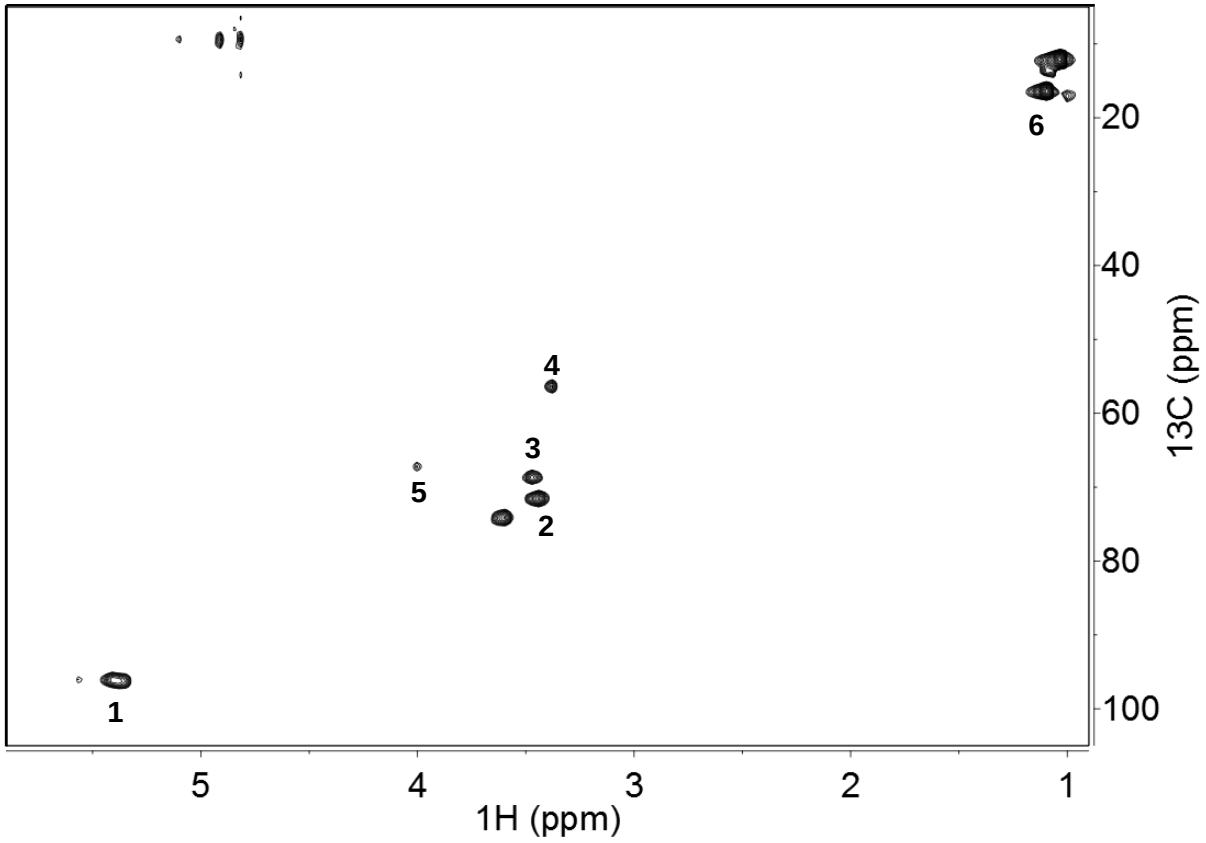
Overlay of 2D-^{1}H-^{13}C-HSQC spectra (600 MHz of ^{1}H) of **6** and CalS13 catalyzed reaction containing mixture of **6** and **10** where the blue and green labels correspond to resonance assignments of **6** and **10**, respectively. The signal corresponding to C4' – H4' of **10** is highlighted.



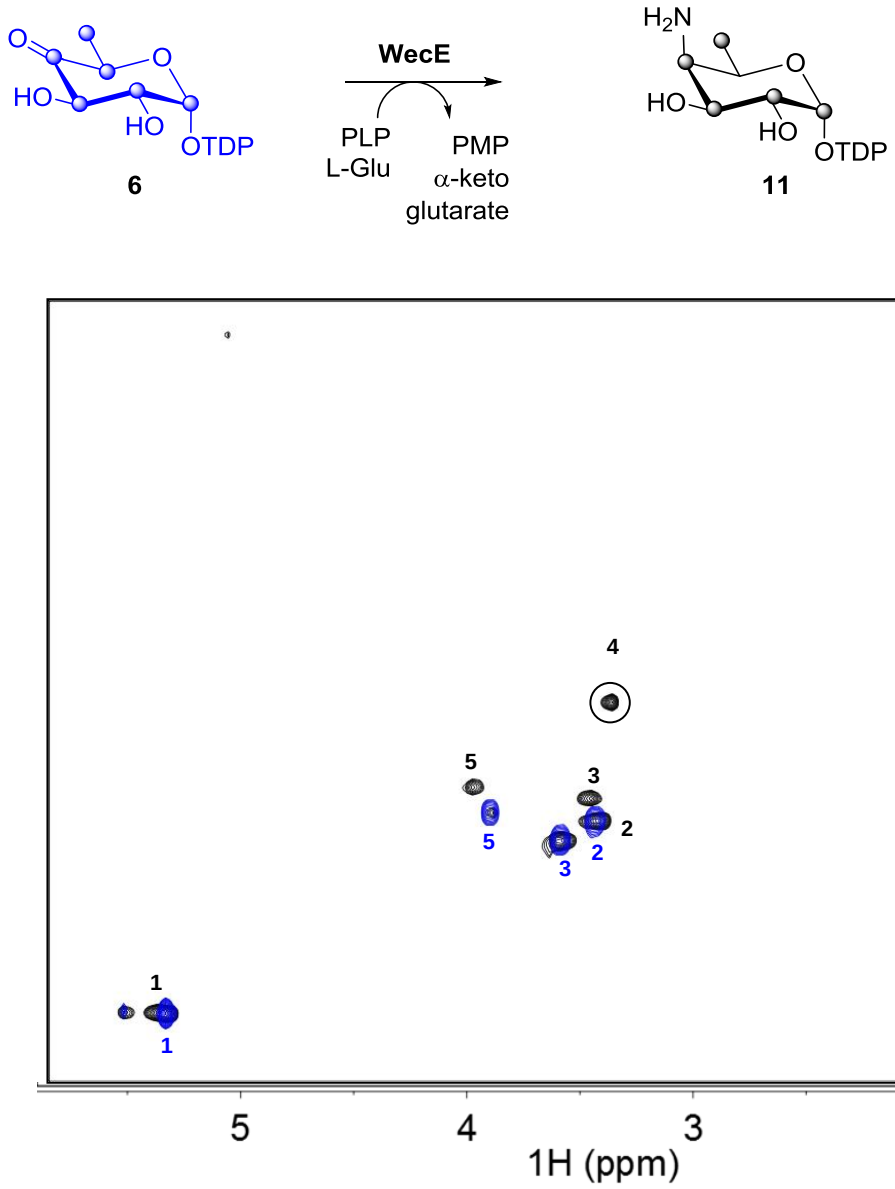


^{1}H-^{13}C 2D plane of HCCH–TOCSY (600 MHz of ^{1}H) of reaction containing mixture of **6** and **10**.

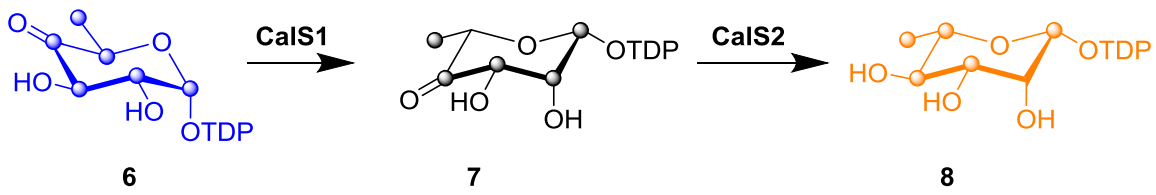




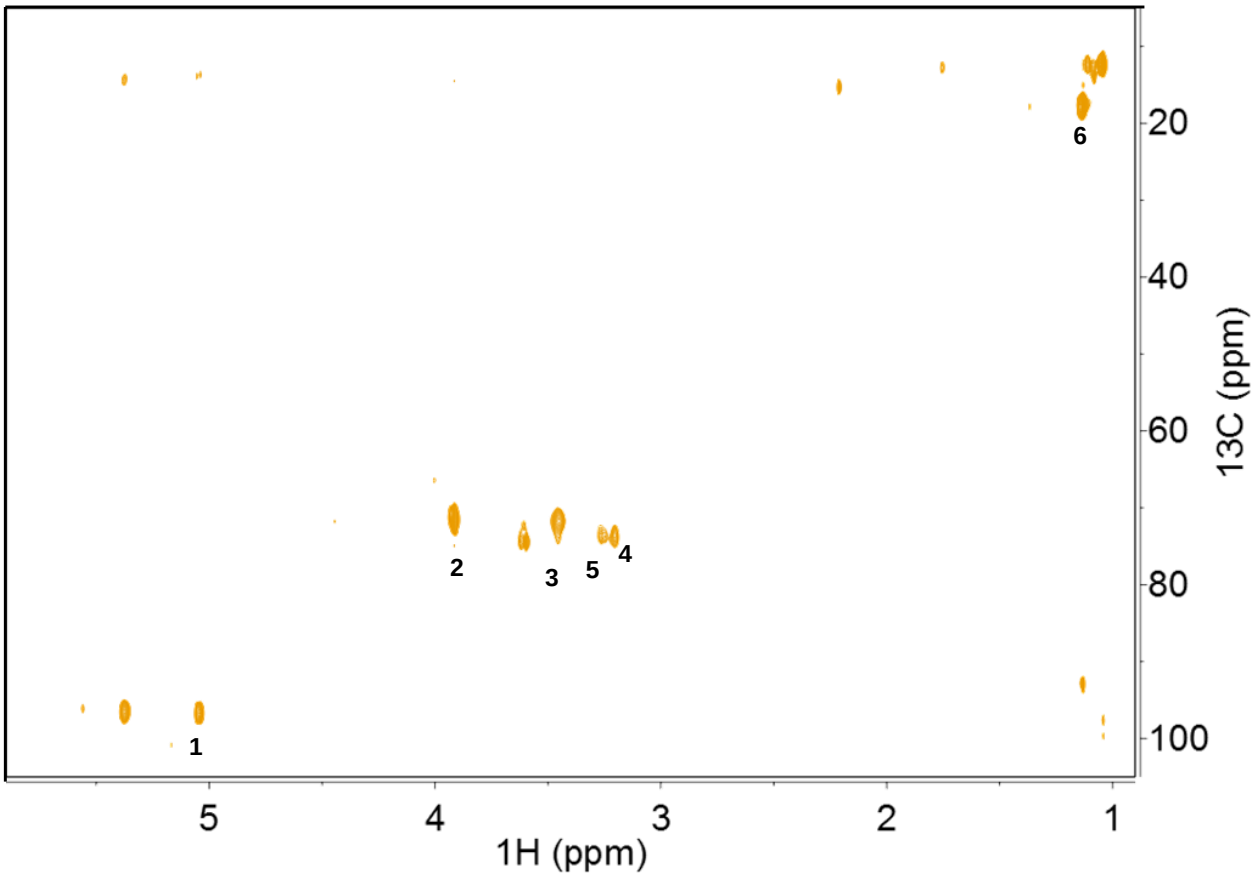
Resonance assignment of **11** in 2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of reaction containing mixture of **6** and **11**.



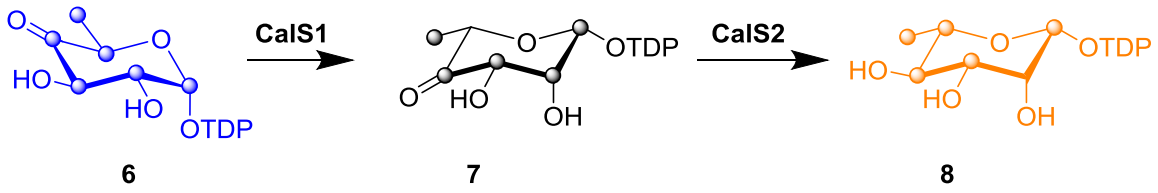
Overlay of 2D-^{1}H-^{13}C-HSQC spectra (600 MHz of ^{1}H) of **6** and WecE catalyzed reaction containing mixture of **6** and **11** where the blue and black labels correspond to resonance assignments of **6** and **11**, respectively. The signal corresponding to C4' – H4' of **11** is highlighted.



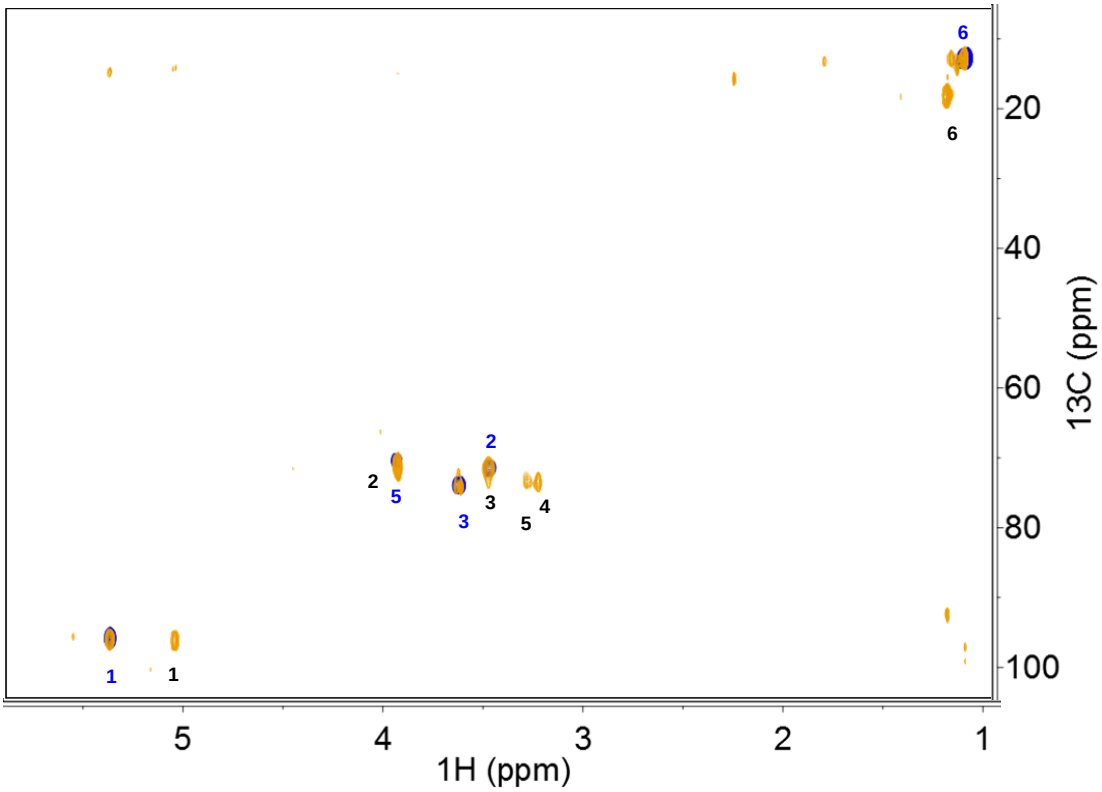
6 **7** **8**



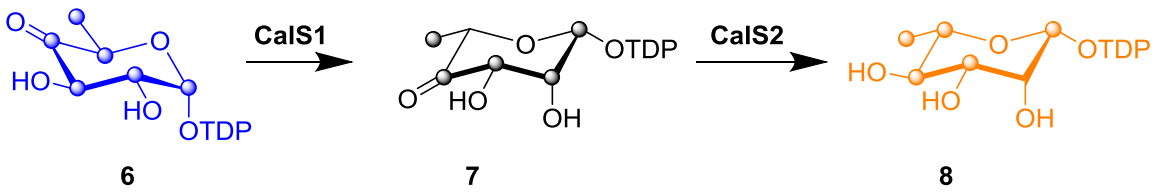
Resonance assignment of **8** in 2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of reaction containing mixture of **6** and **8**.



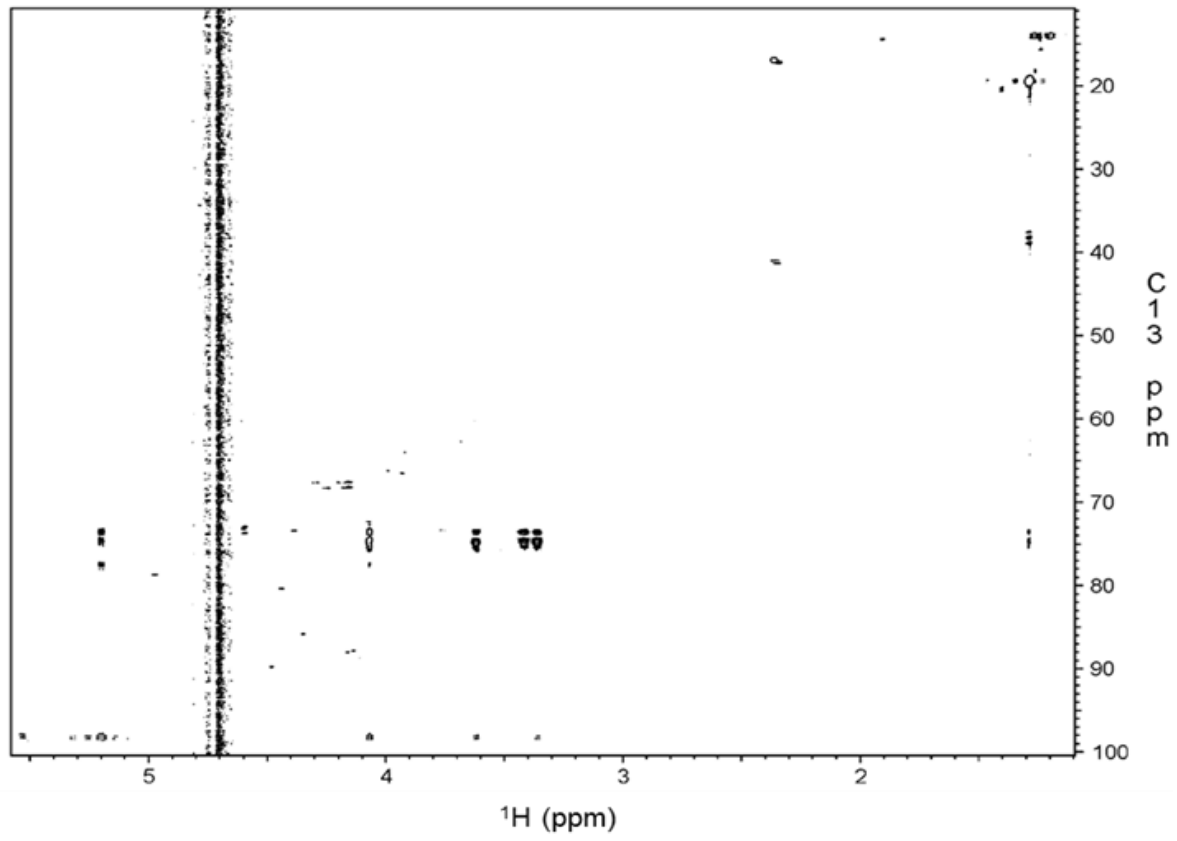
6 **7** **8**



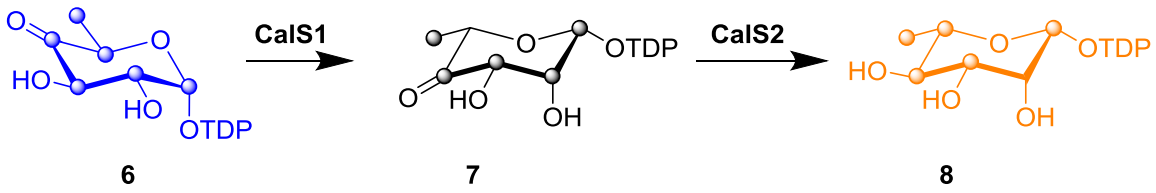
Overlay of 2D-^{1}H-^{13}C-HSQC spectra (600 MHz of ^{1}H) of **6** and CalS1, CalS2 catalyzed reaction containing mixture of **6** and **8** where the blue and black labels correspond to resonance assignments of **6** and **8**, respectively.



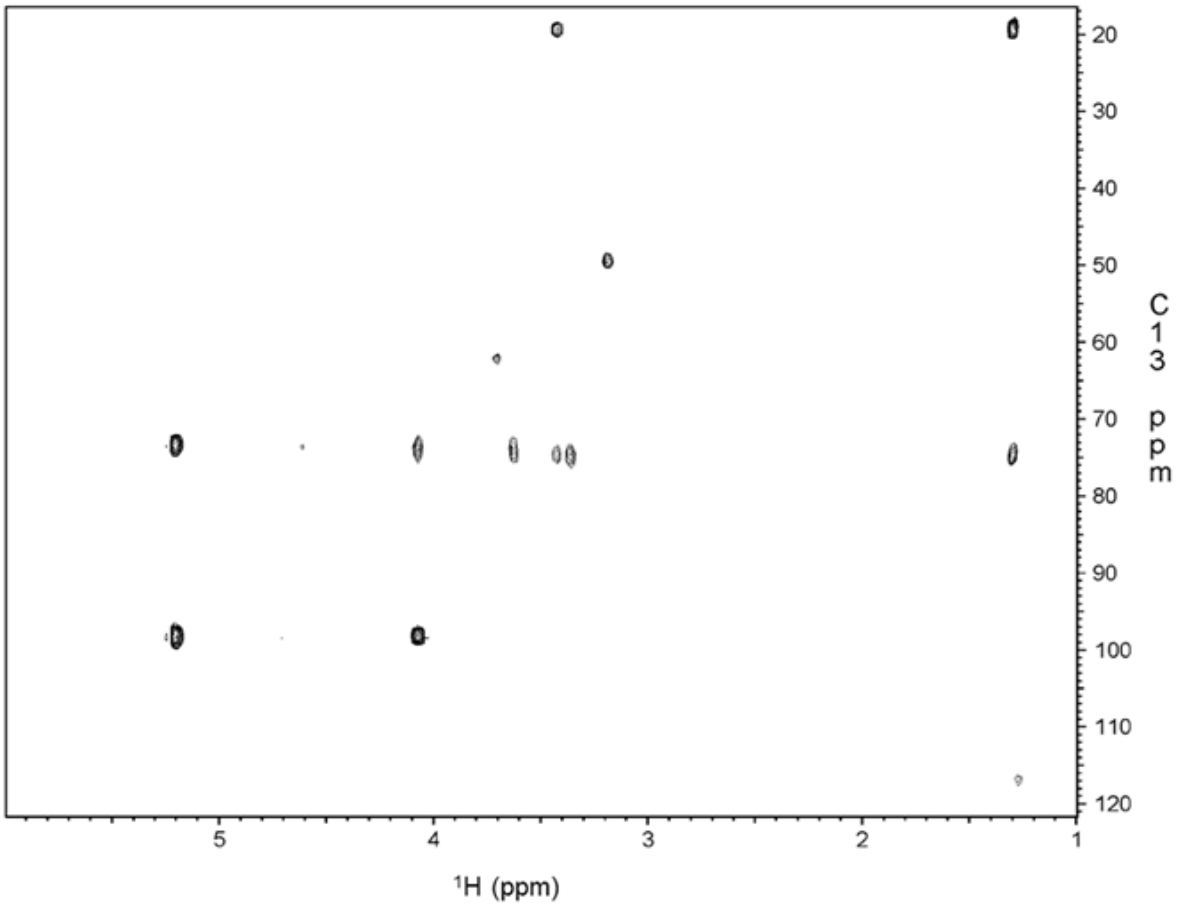
6 **7** **8**



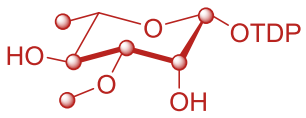
^{1}H-^{13}C 2D plane of HCCH–TOCSY (600 MHz of ^{1}H) of reaction containing mixture of **6** and **8**.



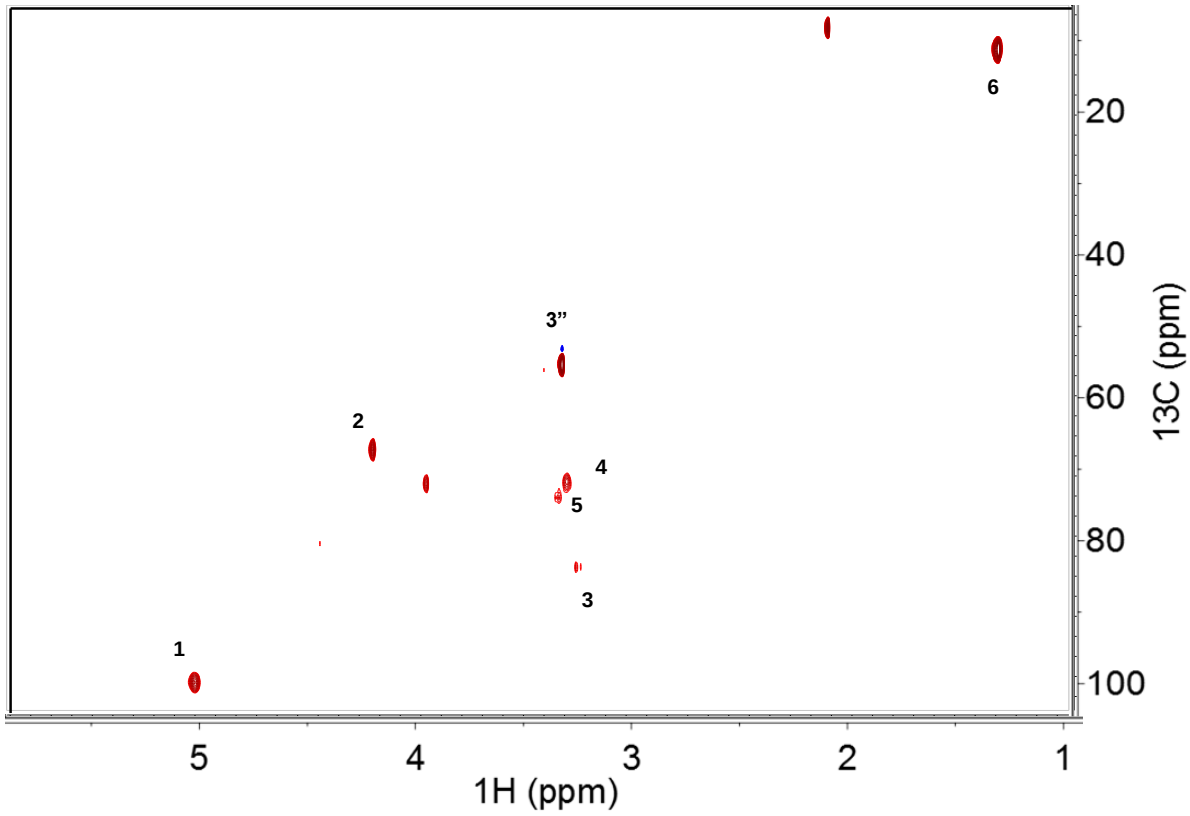
6 **7** **8**



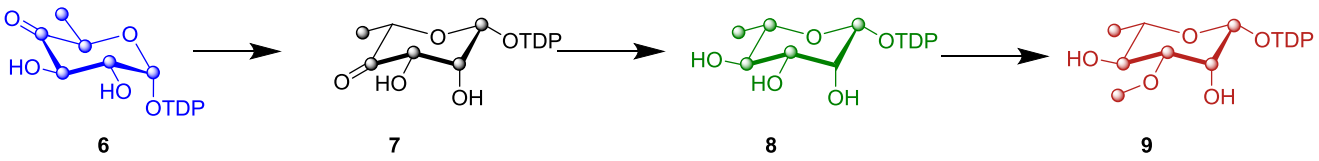
^{1}H-^{13}C 2D plane of HCCH–COSY (600 MHz of ^{1}H) of reaction containing mixture of **6** and **8**.

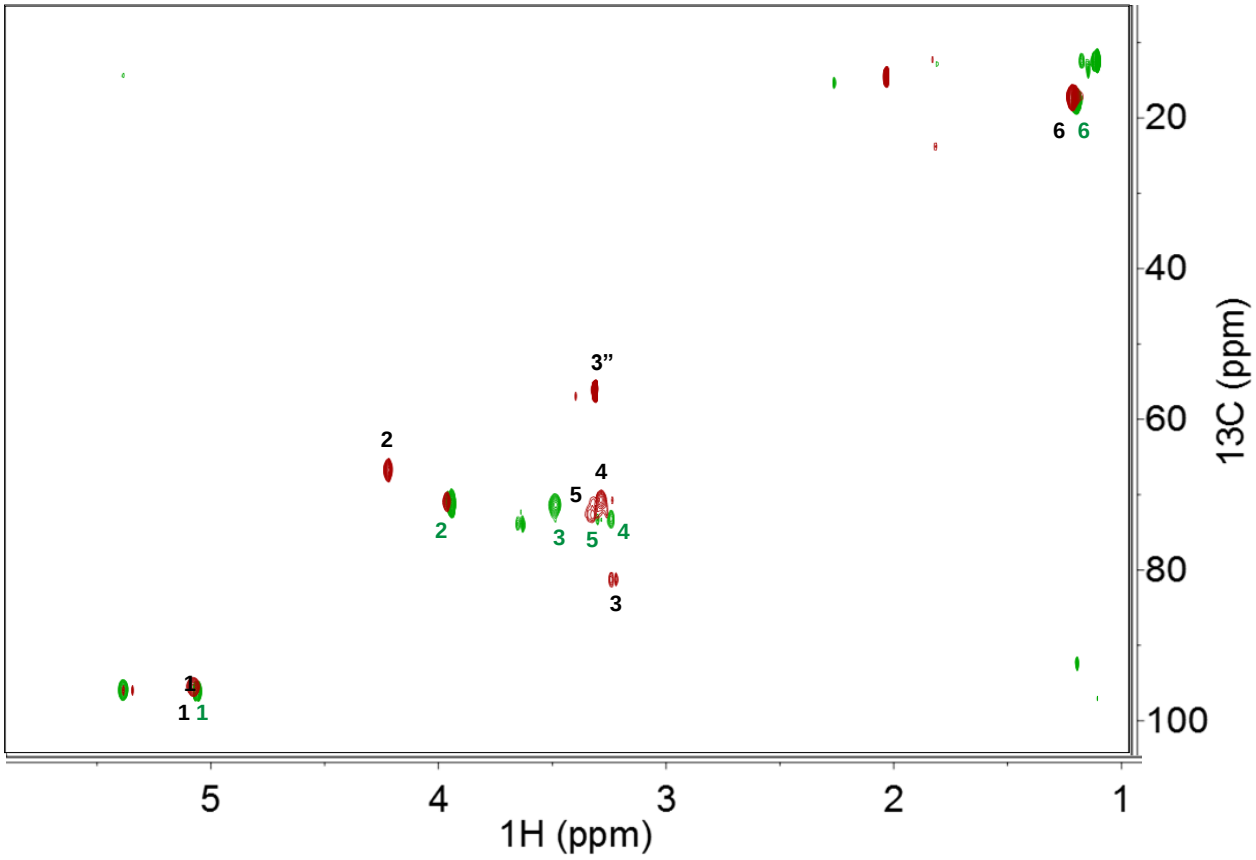


9

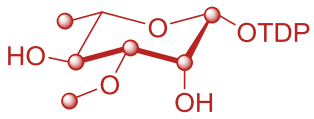


Resonance assignment of **9** in 2D-^{1}H-^{13}C-HSQC spectrum (600 MHz of ^{1}H) of reaction containing mixture of **8** and **9**.

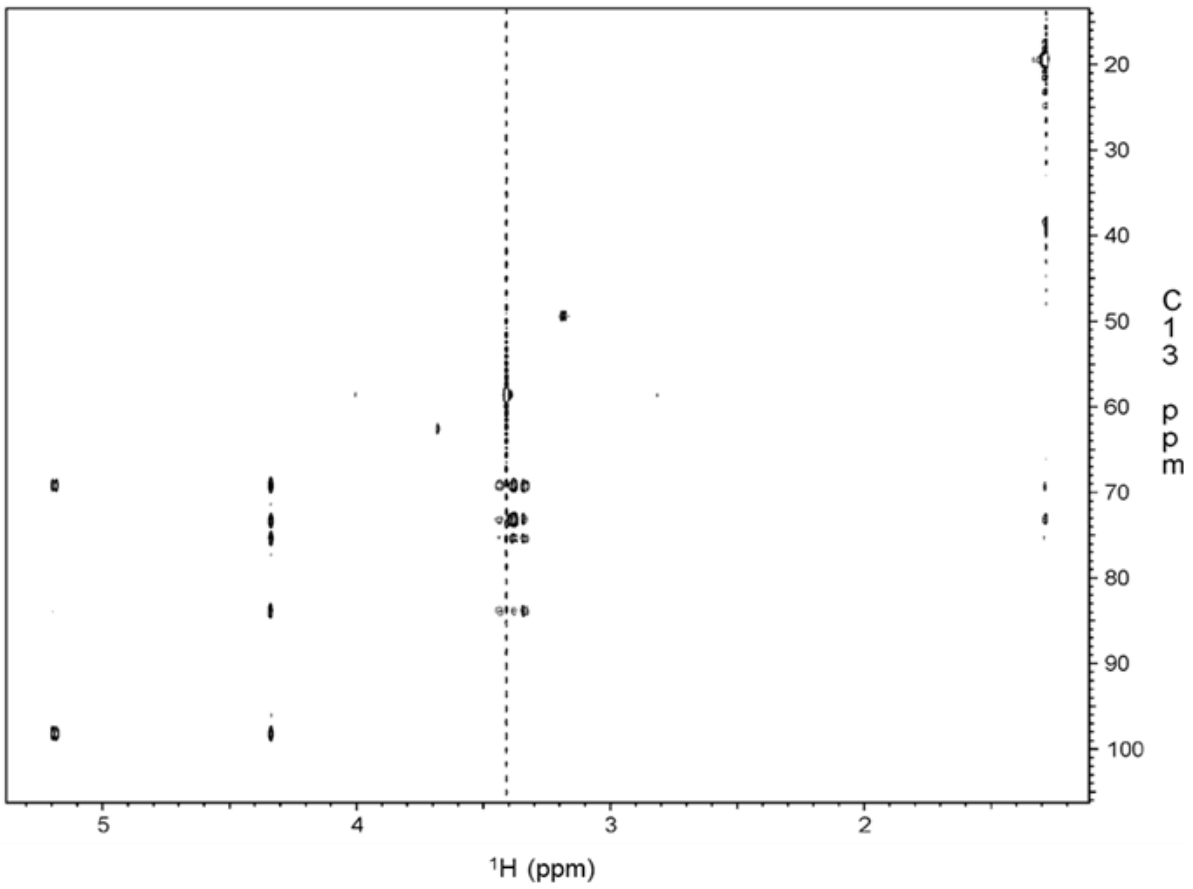




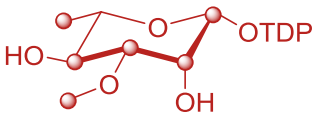
Overlay of 2D-^{1}H-^{13}C-HSQC spectra (600 MHz of ^{1}H) of reaction mixture containing mixture of **6, 8** and **9** where the green and black labels correspond to resonance assignments of **8** and **9**, respectively.



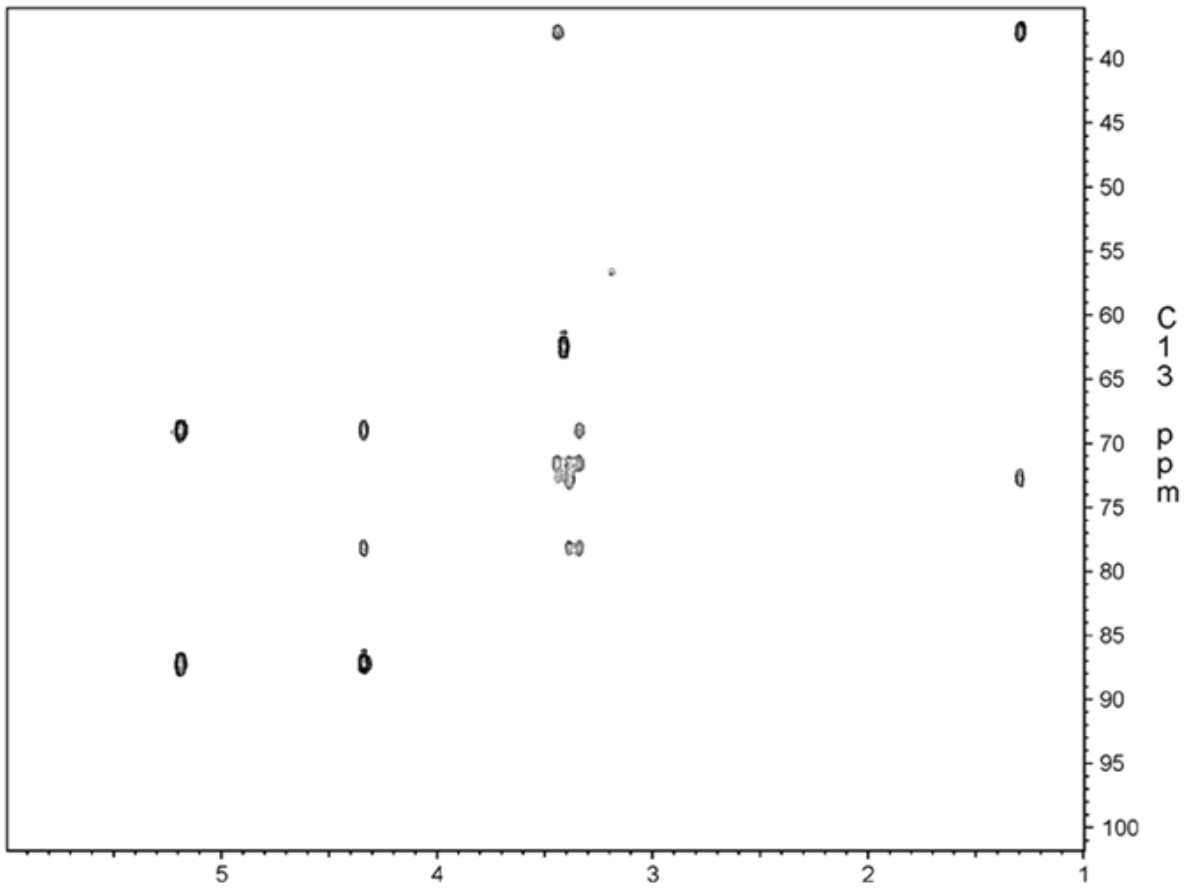
9



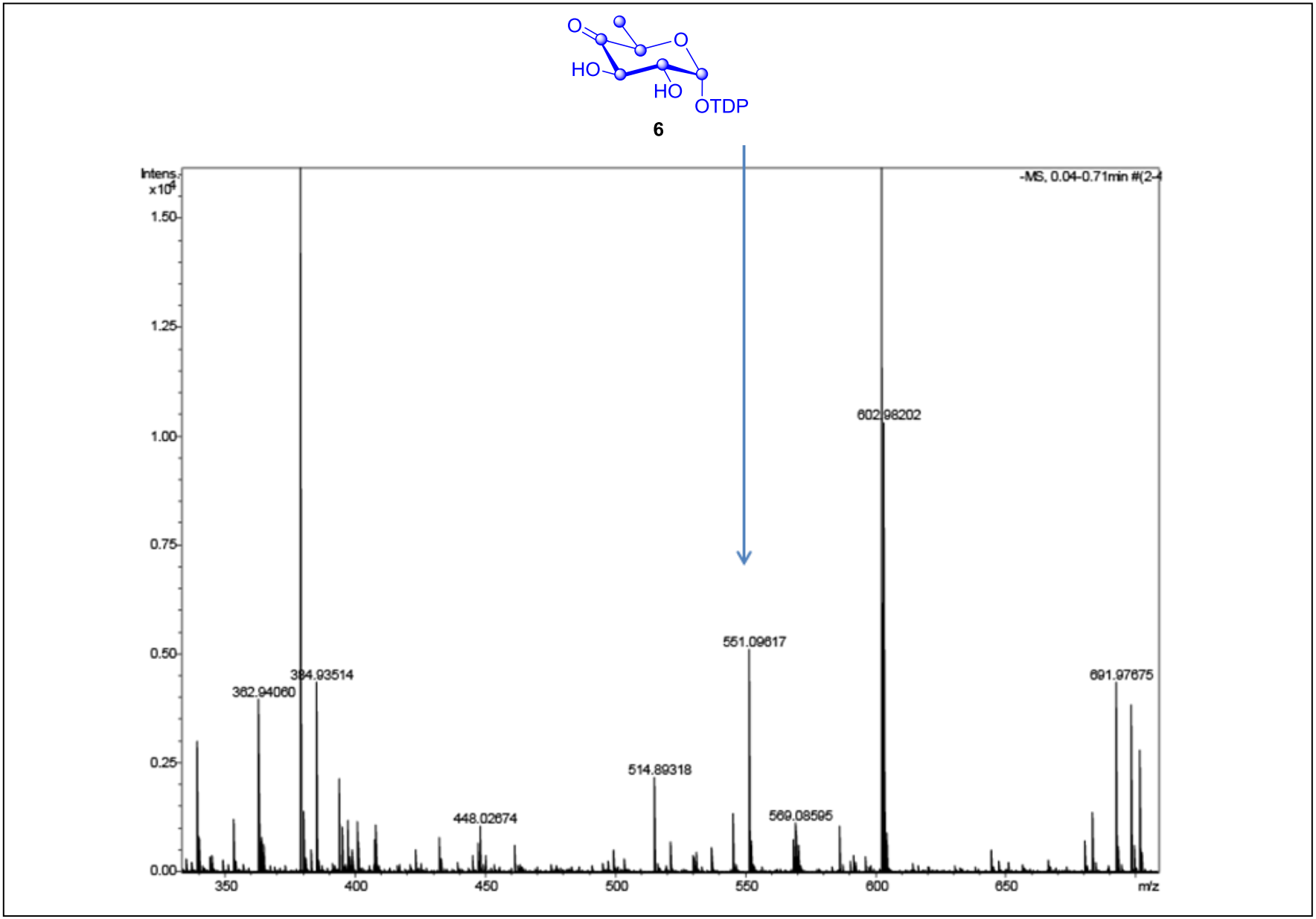
^{1}H-^{13}C 2D plane of HCCH–TOCSY (600 MHz of ^{1}H) of reaction containing **9**.

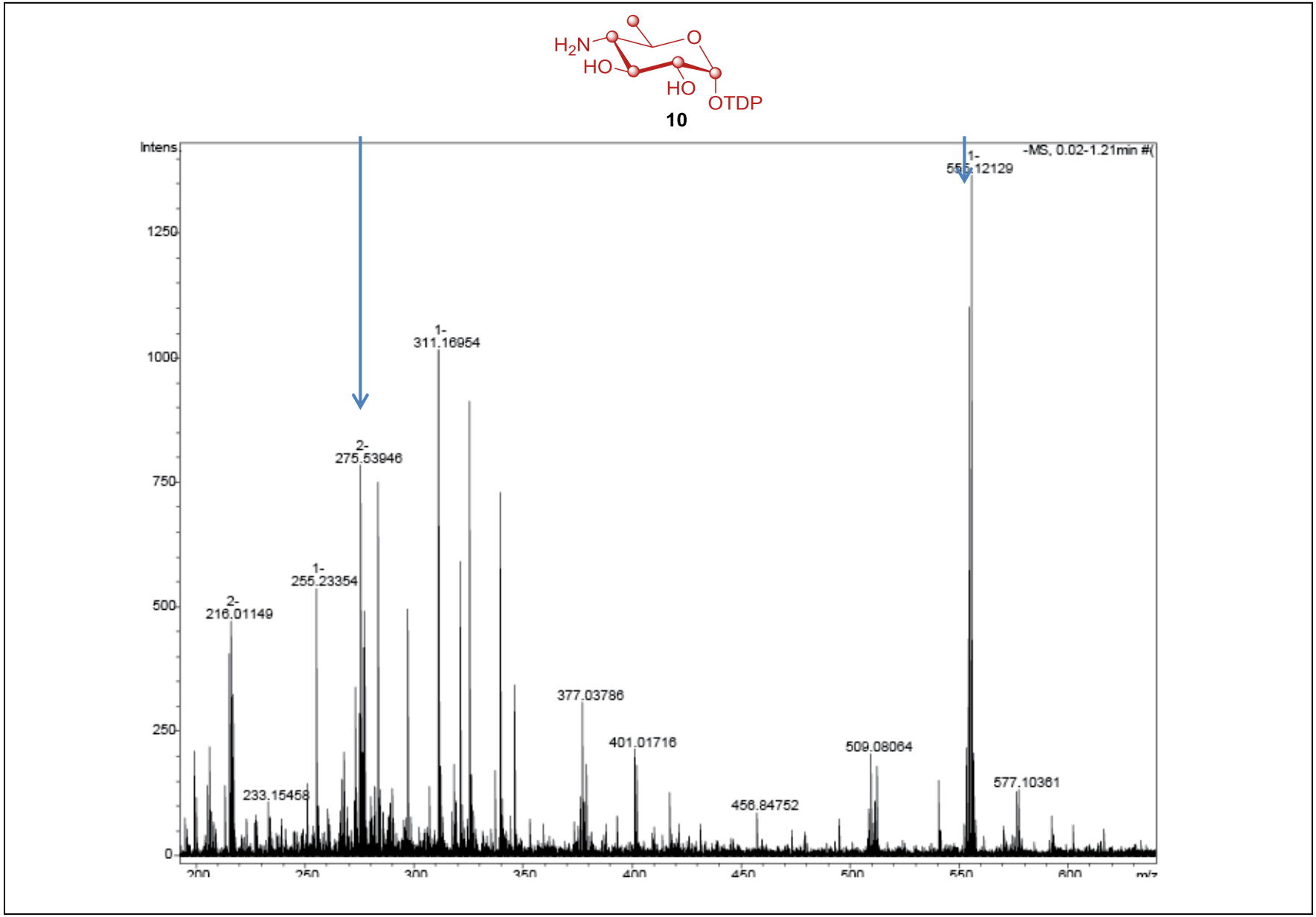


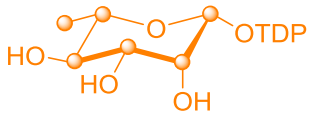
9



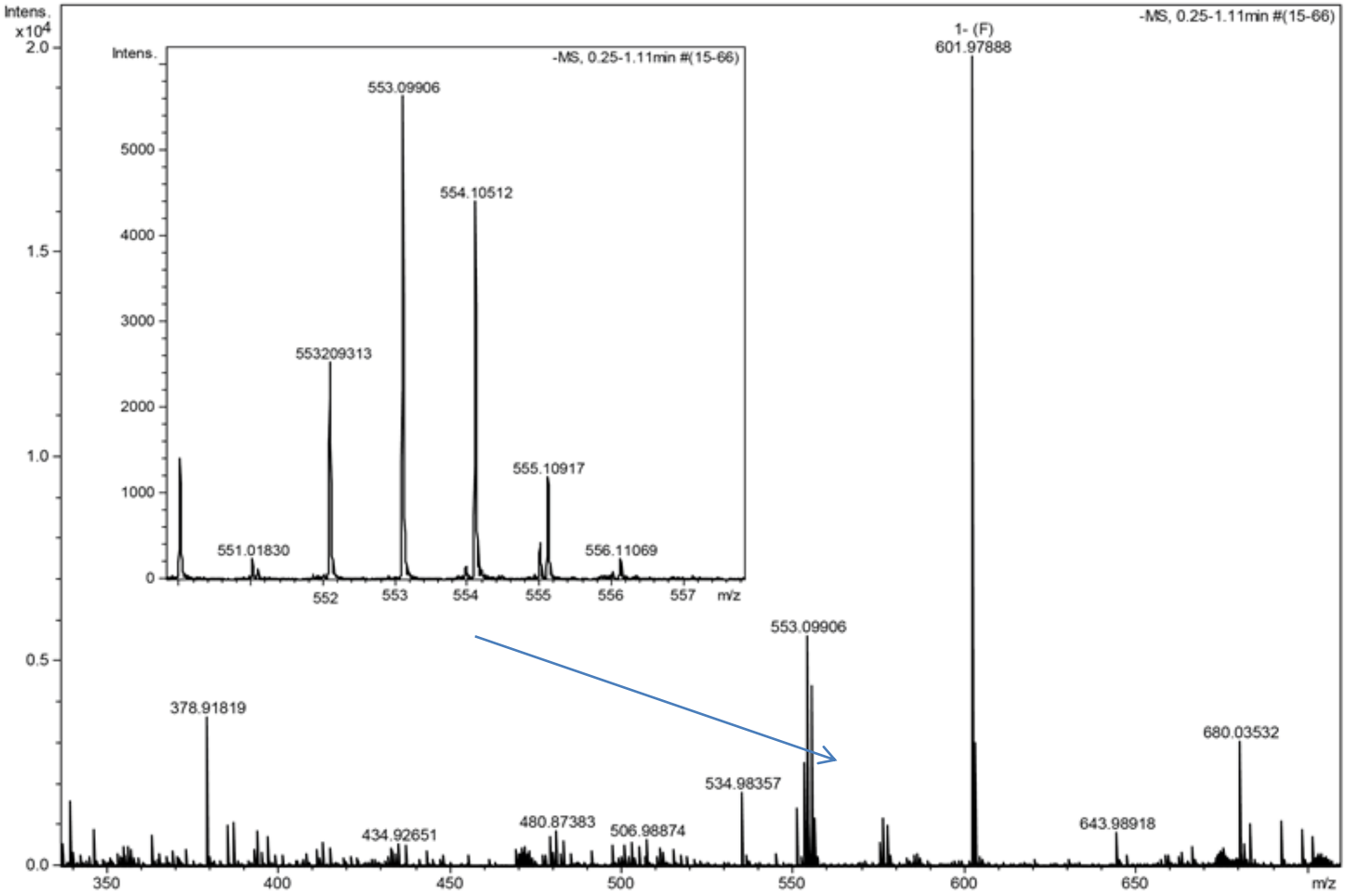
^{1}H-^{13}C 2D plane of HCCH–COSY (600 MHz of ^{1}H) of reaction containing **9**.

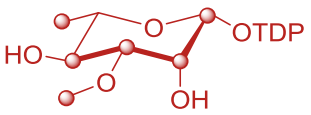

O
O
HO
HO
OTDP
6
Intens.
x10^4
1.50
1.25
1.00
0.75
0.50
0.25
0.00
-MS, 0.04-0.71min #(2-4
602.98202
551.09617
384.93514
382.94060
691.97675
514.89318
448.02674
569.08595
350
400
450
500
550
600
650
m/z


H2N
HO
O
HO
OTDP
10
Intens
-MS, 0.02-1.21min #(
1-
555.12129
1-
311.16954
2-
275.53946
1-
255.23354
2-
216.01149
377.03786
401.01716
509.08064
577.10361
233.15458
456.84752
1250
1000
750
500
250
0
200
250
300
350
400
450
500
550
600
m/z

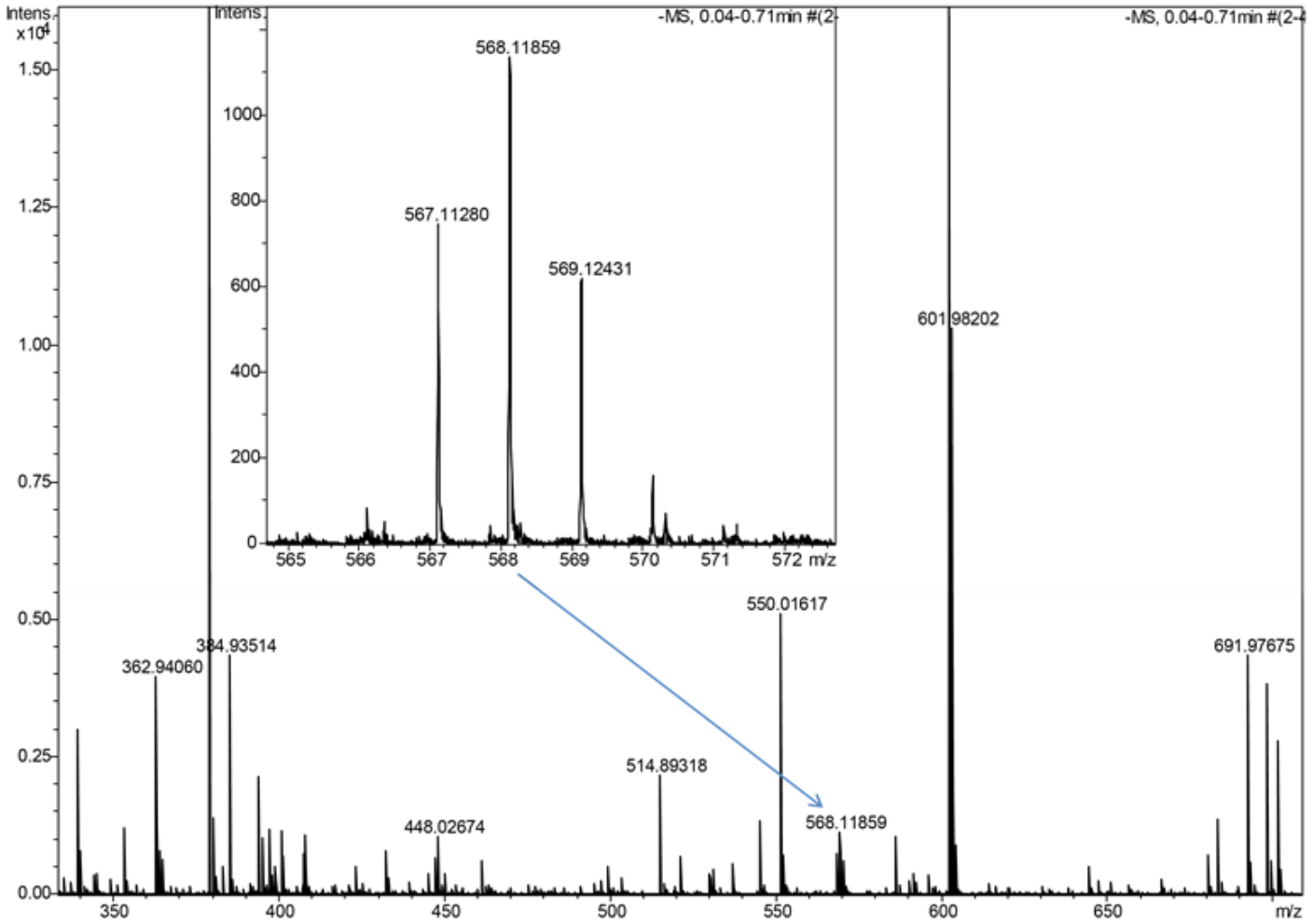

HO
O
OTDP
HO
OH

8


Intens.
x10^4
2.0
1.5
1.0
0.5
0.0
-MS, 0.25-1.11min #(15-66)
1- (F)
601.97888
553.09906
378.91819
680.03532
534.98357
643.98918
434.92651
480.87383
506.98874
350
400
450
500
550
600
650
m/z
Intens.
-MS, 0.25-1.11min #(15-66)
553.09906
5000
554.10512
4000
3000
553209313
2000
555.10917
1000
551.01830
556.11069
0
552
553
554
555
556
557
m/z


HO
O
OTDP
O
OH

9


Intens.
x10^4
1.50
1.25
1.00
0.75
0.50
0.25
0.00
350
400
450
500
550
600
650
m/z
362.94060
384.93514
448.02674
514.89318
550.01617
568.11859
601.98202
691.97675
-MS, 0.04-0.71min #(2-
Intens.
1000
800
600
400
200
0
565
566
567
568
569
570
571
572 m/z
568.11859
567.11280
569.12431