

Supplementary Information

UGT74B1 from *Arabidopsis thaliana* as a versatile biocatalyst for the synthesis of desulfoglycosinolates

Sami Marroun^{a†}, Sabine Montaut^{b†}, Stéphanie Marquès^{c†}, Pierre Lafite^c, Gaël Coadou^a, Patrick Rollin^c, Guillaume Jousset^c, Marie Schuler^c, Arnaud Tatibouët^c, Hassan Oulyadi^a and Richard Daniellou^{*c}

^a Normandie Univ, COBRA, UMR 6014 et FR 3038; Univ Rouen; INSA Rouen; CNRS, IRCOF, 1 rue Tesnière, 76821 Mont Saint Aignan Cedex, France

^b Department of Chemistry and Biochemistry, Biomolecular Sciences Programme, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada

^c Université d'Orléans, CNRS, ICOA, UMR 7311, Orléans, France.

Table S1: Percentages of sequence identities for UGT74B1 and its templates.

Table S2: Percentages of sequence identities for UGT74B1 and its templates in the active site region.

Figure S1 : Sequence alignment for UGT74B1 and its closest homologues.

Figure S2 : ¹H and ¹³C NMR Spectra of compound **7**.

Figure S3 : HRMS Spectrum of compound **7**.

Figure S4 : ¹H and ¹³C NMR Spectra of compound **9**.

Figure S5 : HRMS Spectrum of compound **9**.

Figure S6 : ¹H and ¹³C NMR Spectra of compound **10**.

Figure S7 : HRMS Spectrum of compound **10**.

	74B1	2VCE	3HBF	2C1Z	2PQ6	4WHM	2ACW
74B1	100	85.65	87.53	88.12	86.55	86.61	84.47
2VCE	25.83	100	83.7	84.01	84.99	82.56	87.09
3HBF	23.26	24.58	100	91.09	85.14	87.88	83.17
2C1Z	23.48	24.79	48.03	100	85.82	89.27	83.46
2PQ6	25.93	28.01	23.86	22.82	100	84.4	83.53
4WHM	26.09	22.71	39.65	42.98	21.99	100	82.99
2ACW	22.15	29.79	23.44	22.8	24.27	20.86	100

Table S1: Percentages of sequence identities obtained by a dual structure-sequence multiple sequence alignment (green) and structural similarity scores (blue) obtained using CATH-SSAP for UGT74B1 and its templates. The maximum possible structural similarity score is 100.

Active site	74B1	2VCE	3HBF	2C1Z	2PQ6	4WHM	2ACW
74B1	100	51.02	47.92	45.83	60.42	45.83	45.83
2VCE		100	46.94	51.02	44.9	40.82	55.1
3HBF			100	81.25	43.75	68.75	50
2C1Z				100	45.83	64.58	45.83
2PQ6					100	43.75	41.67
4WHM						100	39.58
2ACW							100

Table S2: Percentages of sequence identities obtained by a dual structure-sequence multiple sequence alignment for UGT74B1 and its templates in the active site region.

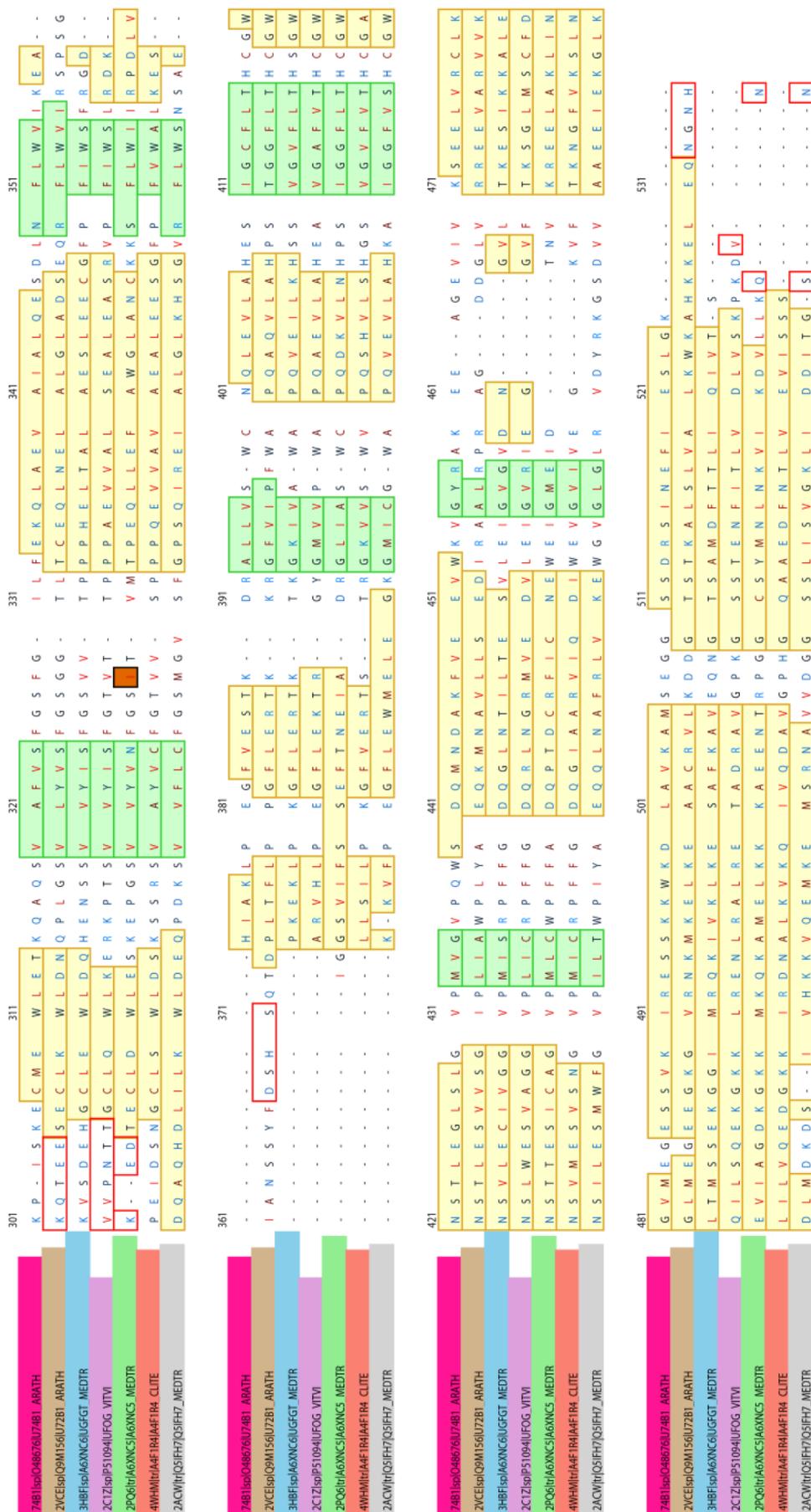


Figure S1 (continued): Sequence alignment for UGT74B1 and its closest homologues using the TCOFFEE accurate mode. Secondary structures are highlighted in green for sheets and in yellow for helices. Secondary structure assignment for UGT74B1 is based on the homology model obtained using the MODELLER program.

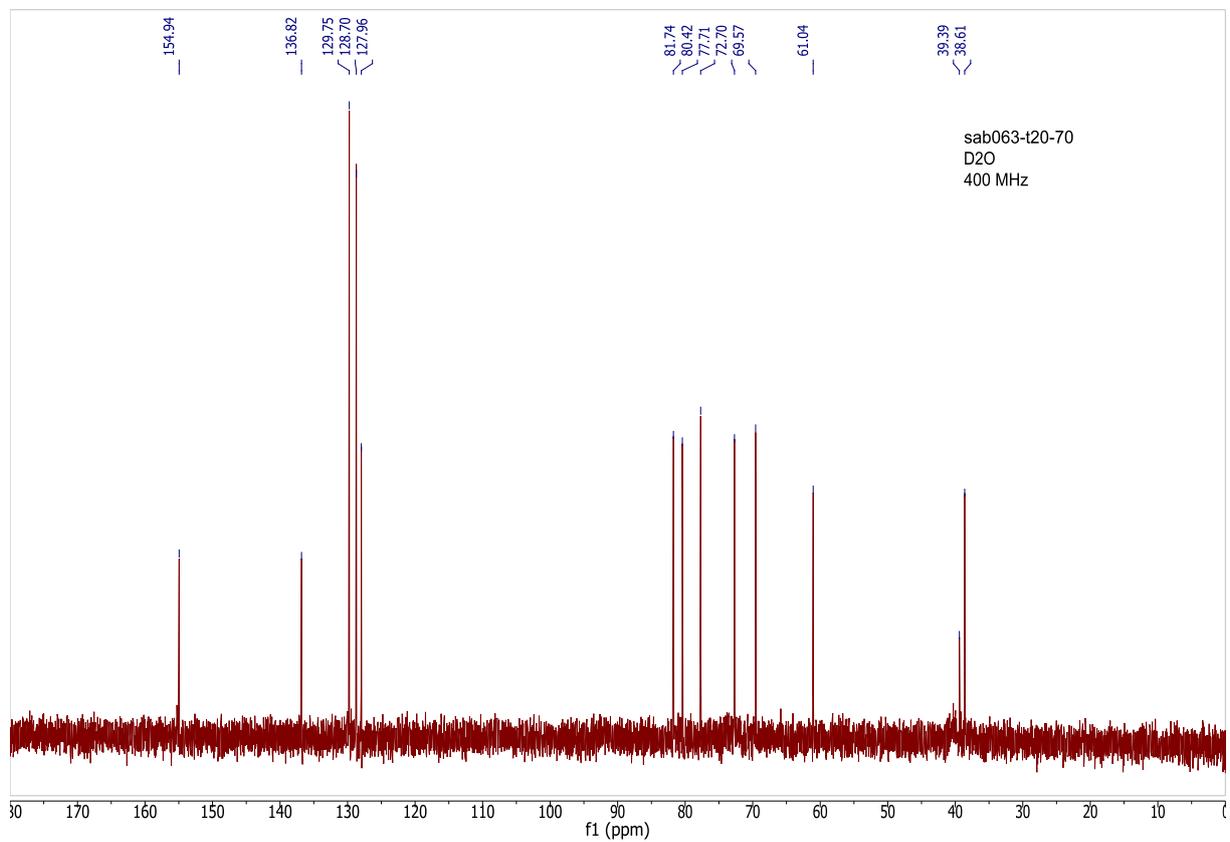
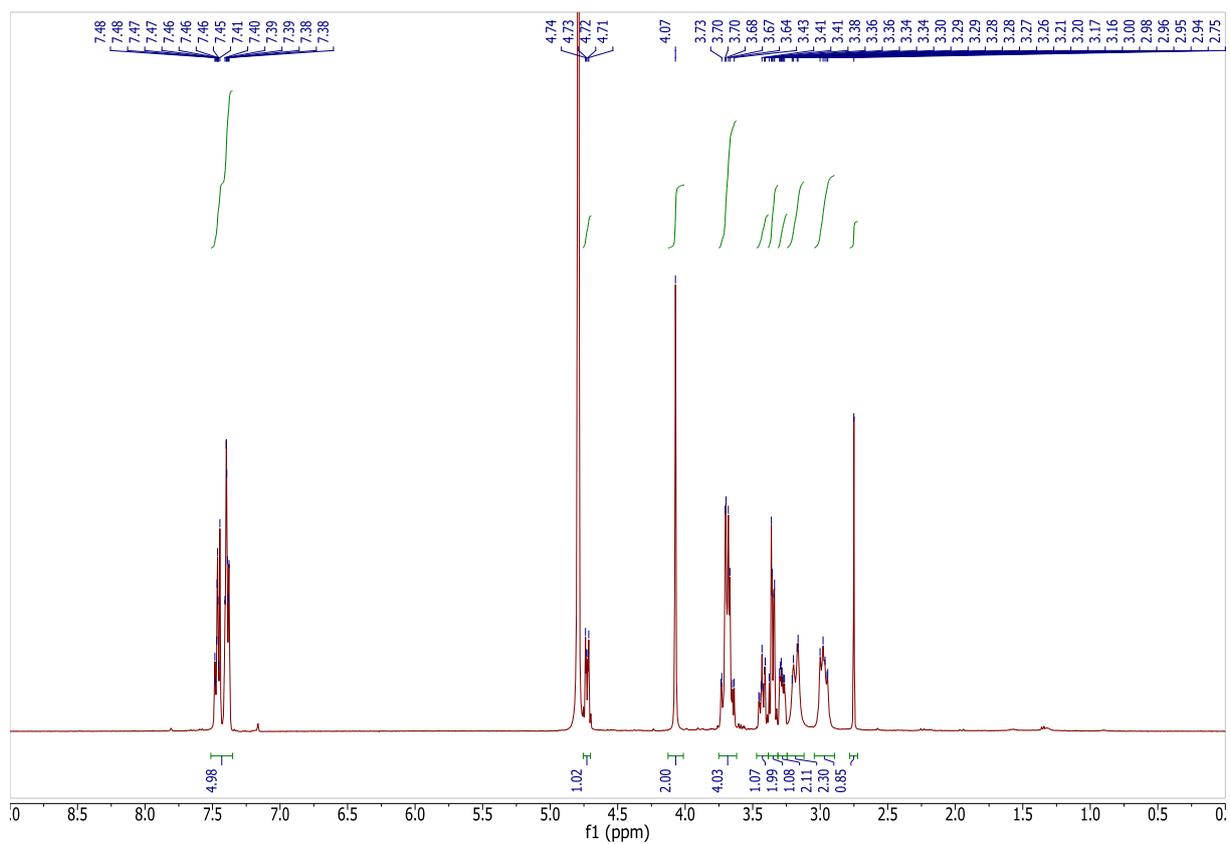


Figure S2 : ^1H (top) and ^{13}C (bottom) NMR Spectra of compound 7.

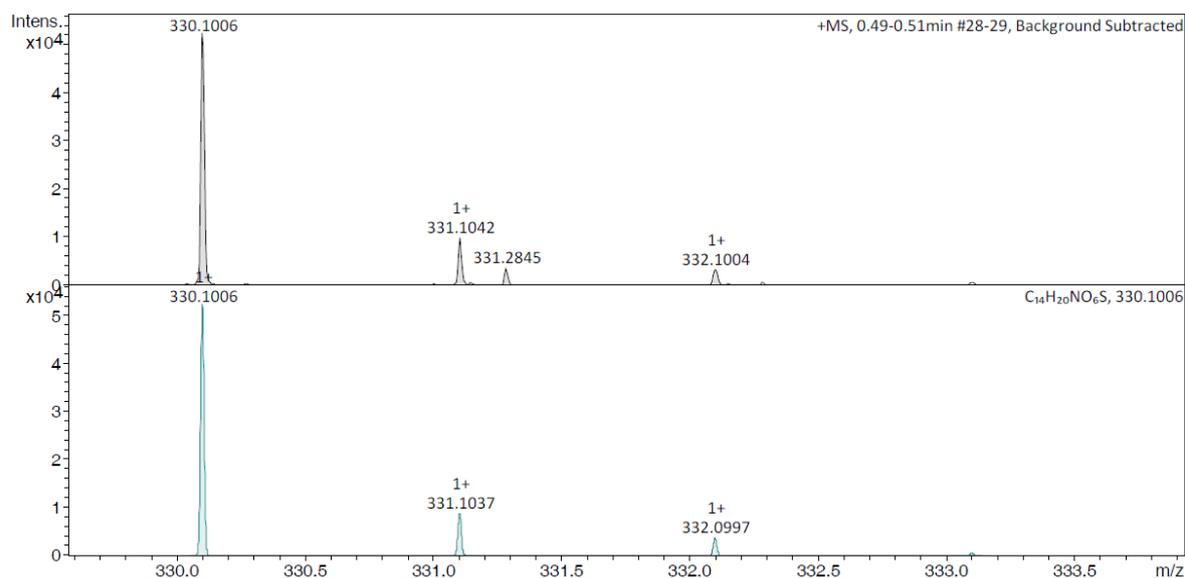
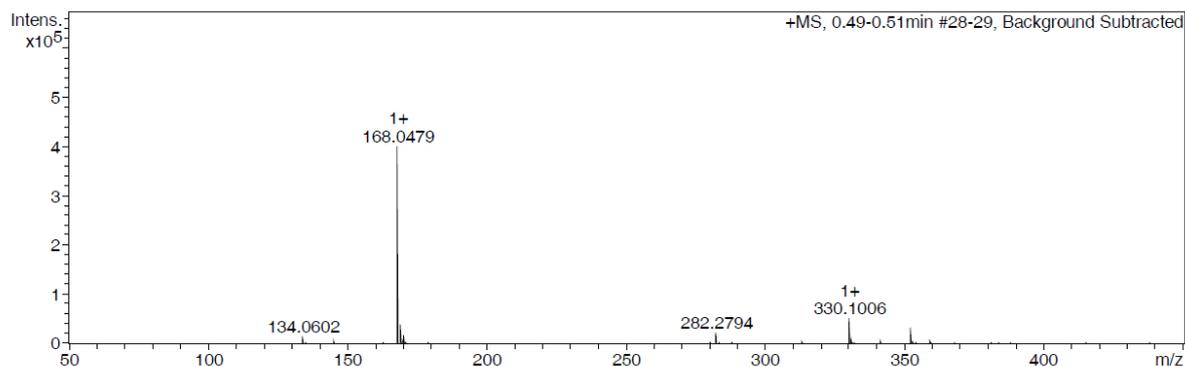
Analysis Info

Sample Name **SAB063-T20-T70**
 Analysis Name X017023CYC_24837.d

Acquisition Date 03/12/2014 16:09:30
 Instrument / Ser# maXis 255552.00086
 Method Positif.m

Acquisition Parameter

Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.6 Bar
Scan Begin	50 m/z	Set Capillary	4500 V	Set Dry Heater	200 °C
Scan End	3000 m/z	Set Collision Cell RF	500.0 Vpp	Set Dry Gas	7.0 l/min



Meas. m/z	z	#	Ion Formula	m/z	err [ppm]	mSigma	rdb	e ⁻	Conf
168.047942	1+	1	C8H10NOS	168.047761	1.1	4.4	4.5	even	
330.100620	1+	1	C14H20NO6S	330.100585	0.1	11.7	5.5	even	

Figure S3 : HRMS Spectrum of compound 7.

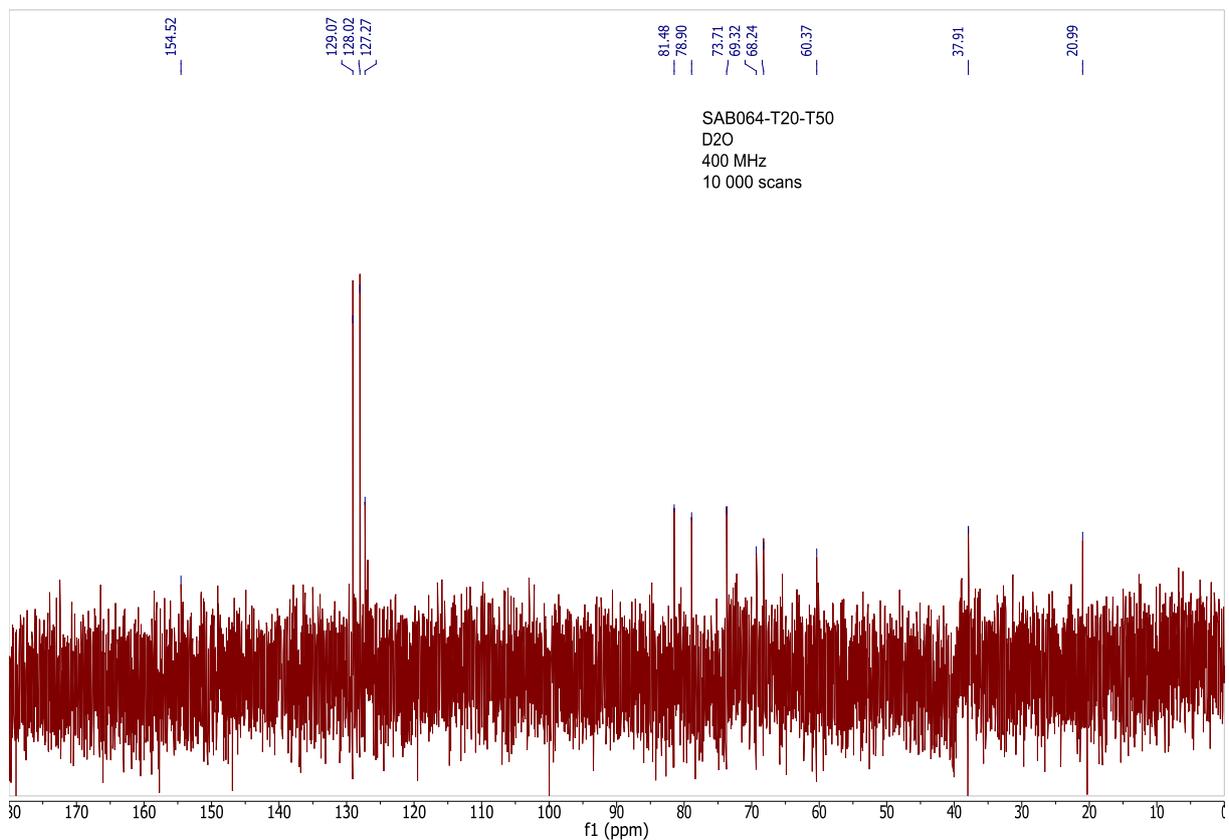
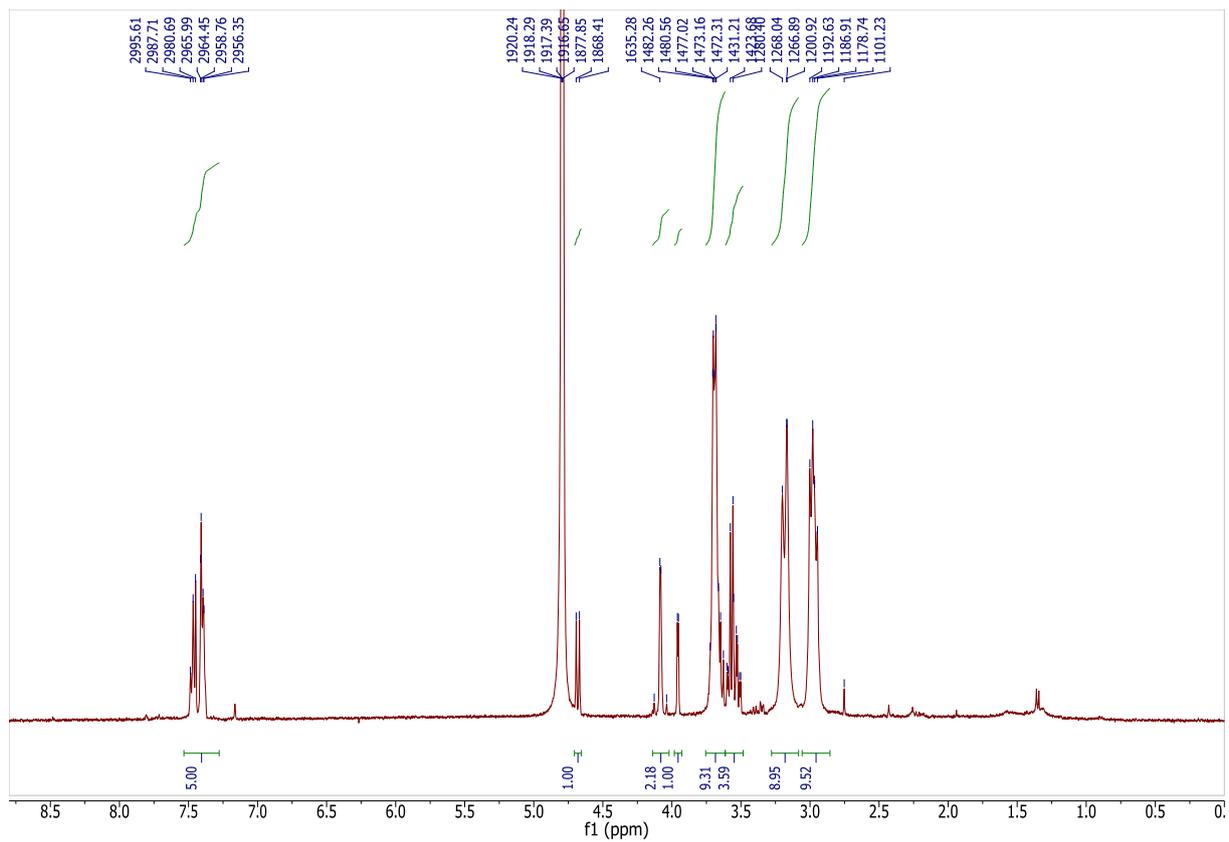


Figure S4: ^1H (top) and ^{13}C (bottom) NMR Spectra of compound **9**.

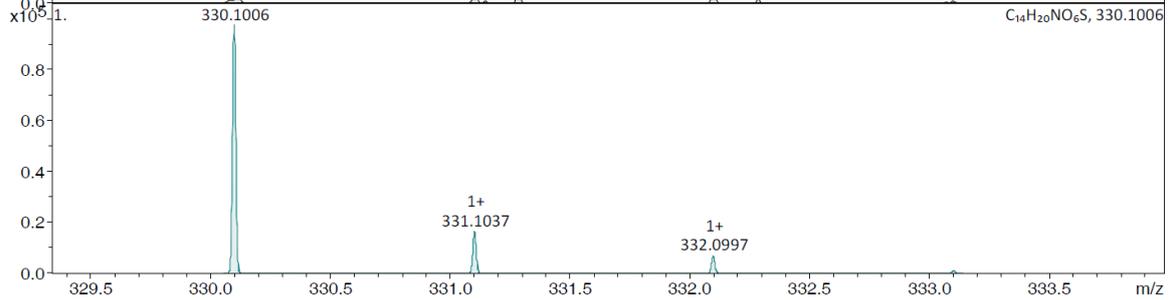
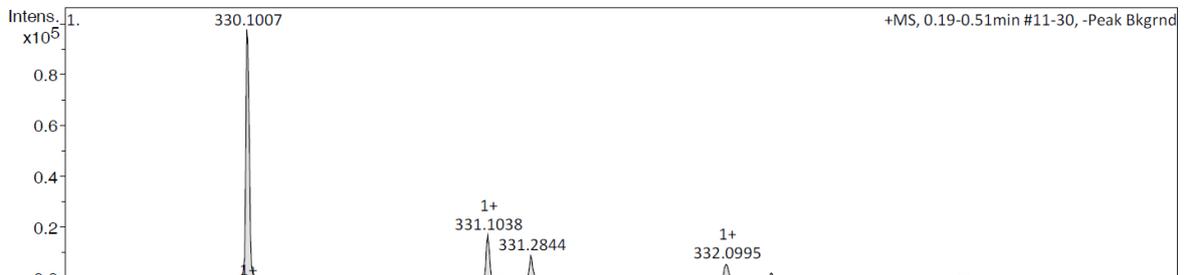
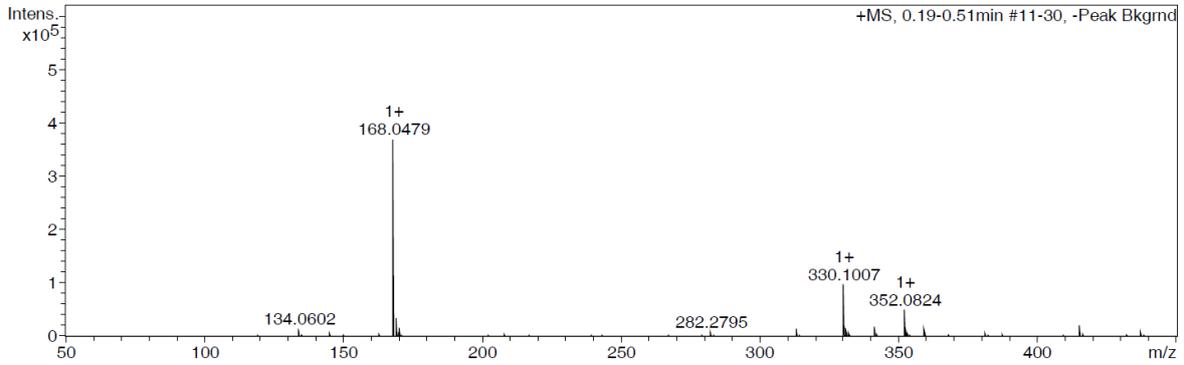
Analysis Info

Sample Name **SAB064-T20-T50**
 Analysis Name X017024CYC_24838.d

Acquisition Date 03/12/2014 16:11:23
 Instrument / Ser# maXis 255552.00086
 Method Positif.m

Acquisition Parameter

Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.6 Bar
Scan Begin	50 m/z	Set Capillary	4500 V	Set Dry Heater	200 °C
Scan End	3000 m/z	Set Collision Cell RF	500.0 Vpp	Set Dry Gas	7.0 l/min



Meas. m/z	z	#	Ion Formula	m/z	err [ppm]	mSigma	rdb	e ⁻	Conf
168.047920	1+	1	C ₈ H ₁₀ NOS	168.047761	0.9	4.6	4.5	even	
330.100677	1+	1	C ₁₄ H ₂₀ NO ₆ S	330.100585	0.3	6.6	5.5	even	
352.082400	1+	1	C ₁₄ H ₁₉ NNaO ₆ S	352.082529	-0.4	8.0	5.5	even	

Figure S5 : HRMS Spectrum of compound 9.

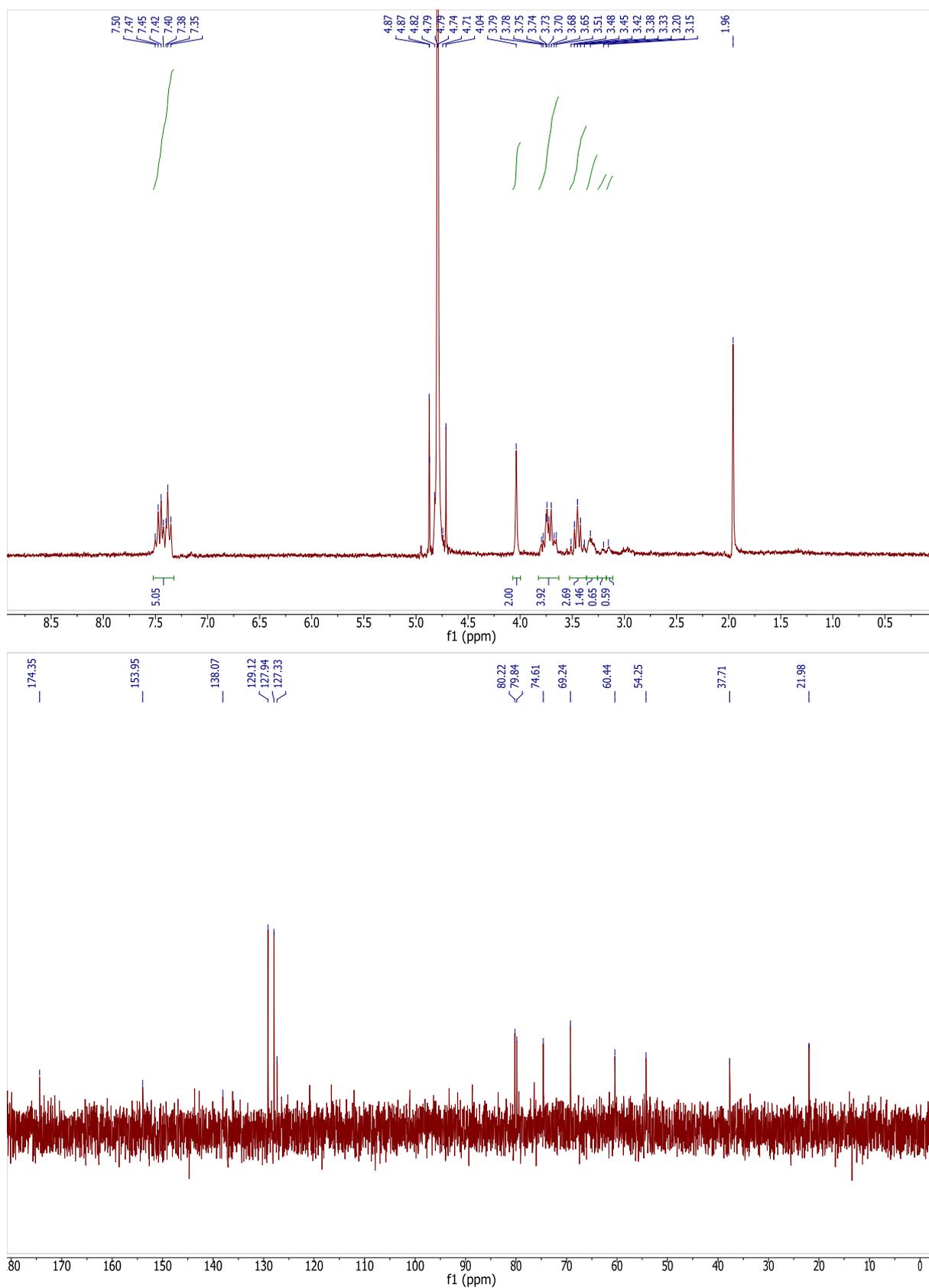


Figure S6 : ^1H (top) and ^{13}C (bottom) NMR Spectra of compound 10.

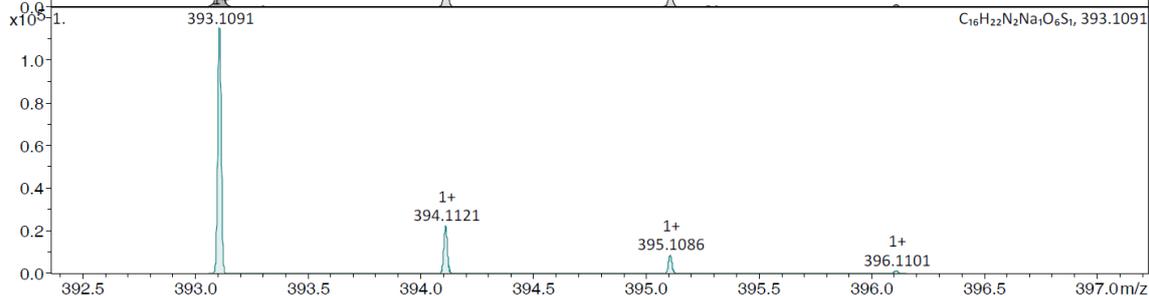
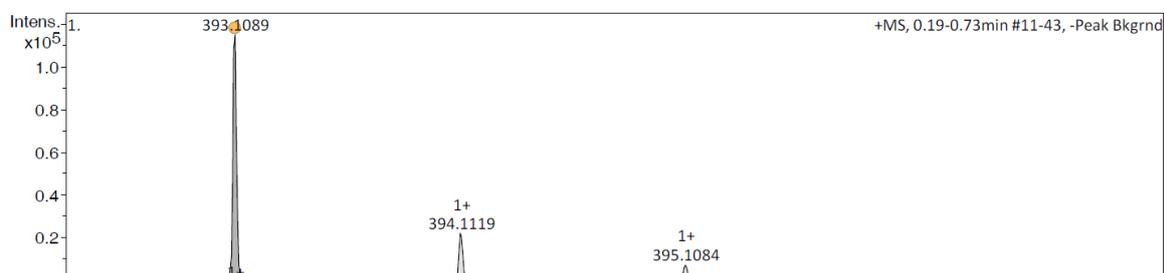
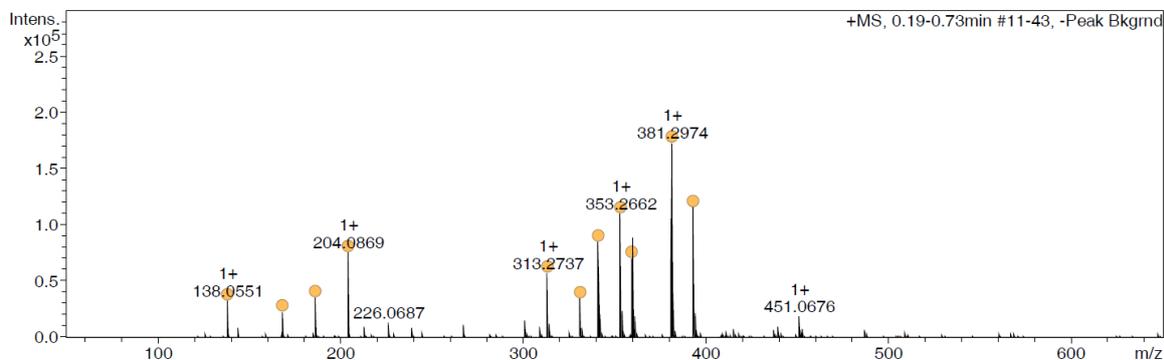
Analysis Info

Sample Name **SAB065-T15-4-T30**
 Analysis Name X017245CYC_Na.d

Acquisition Date 10/12/2014 22:50:05
 Instrument / Ser# maXis 255552.00086
 Method Positif.m

Acquisition Parameter

Source Type ESI Ion Polarity Positive Set Nebulizer 0.6 Bar
 Scan Begin 50 m/z Set Capillary 4500 V Set Dry Heater 200 °C
 Scan End 3000 m/z Set Collision Cell RF 500.0 Vpp Set Dry Gas 7.0 l/min



Meas. m/z	z	#	Ion Formula	m/z	err [ppm]	mSigma	rdb	e ⁻ Conf
138.055068	1+	1	C7H8NO2	138.054955	0.8	1.4	4.5	even
168.065466	1+	1	C8H10NO3	168.065520	0.3	2.1	4.5	even
186.076283	1+	1	C8H12NO4	186.076084	1.1	1.4	3.5	even
204.086877	1+	1	C8H14NO5	204.086649	1.1	26.1	2.5	even
313.273719	1+	1	C19H37O3	313.273721	-0.0	5.4	1.5	even
331.284332	1+	1	C19H39O4	331.284286	0.1	19.1	0.5	even
341.304892	1+	1	C21H41O3	341.305022	-0.4	5.5	1.5	even
353.266170	1+	1	C17H33N6O2	353.265951	0.6	2.1	4.5	even
	1+	1	C19H38NaO4	353.266230	0.2	0.4	0.5	even
359.315518	1+	1	C21H43O4	359.315586	-0.2	407.6	0.5	even
381.297360	1+	1	C19H37N6O2	381.297251	-0.3	3.4	4.5	even
	1+	1	C21H42NaO4	381.297531	0.4	2.8	0.5	even
393.108913	1+	1	C14H17N8O4S	393.108798	0.3	6.0	10.5	even
	1+	2	C22H13N6O2	393.109450	-1.4	38.5	19.5	even
	1+	1	C16H22N2NaO6S	393.109078	-0.4	8.2	6.5	even

FigureS7 : HRMS Spectrum of compound 10.