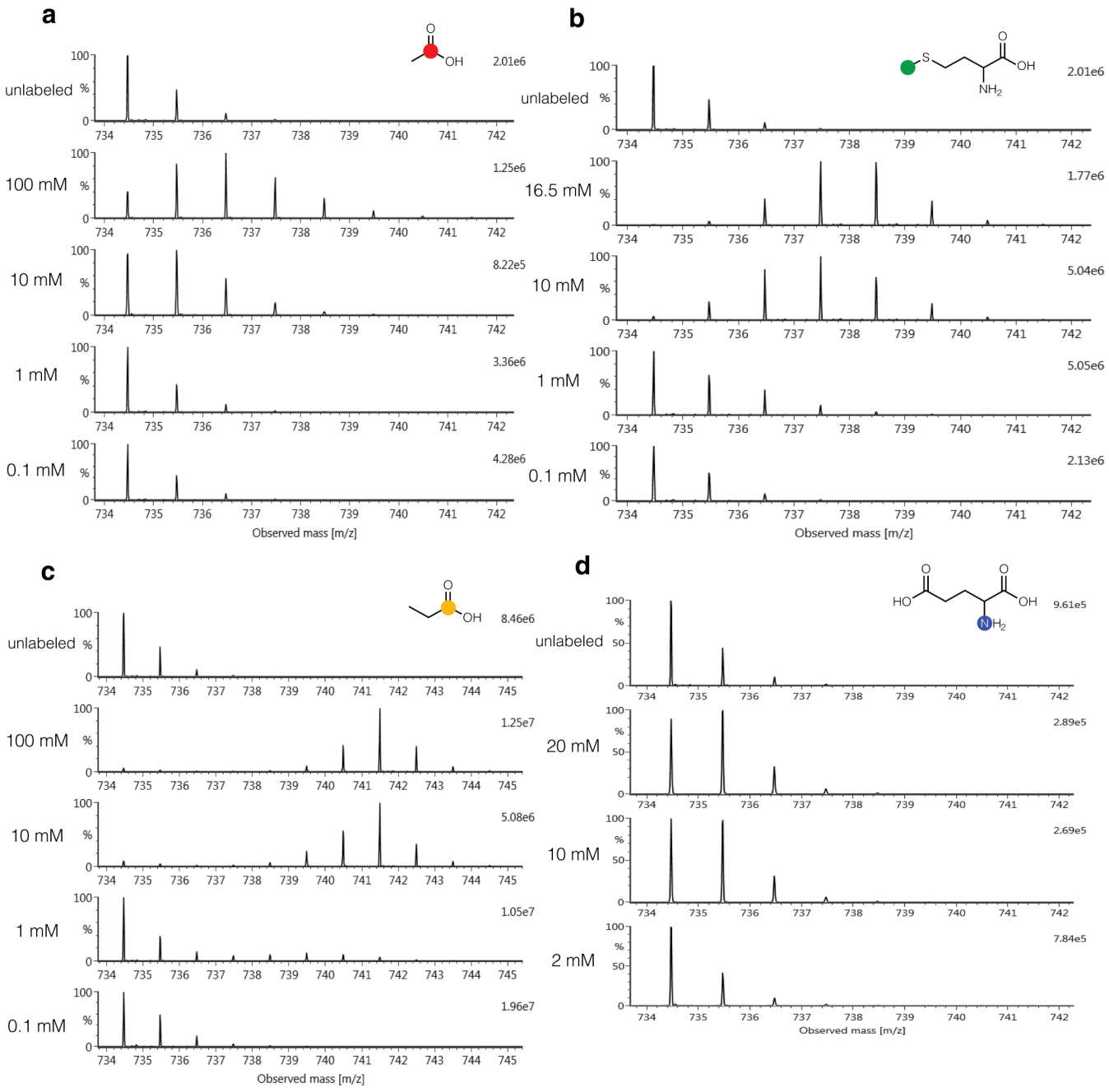
An Isotopic Labeling Approach Linking Natural Products with Biosynthetic Gene Clusters

*Catherine S. McCaughey[1], Jeffrey A. van Santen[1], Justin J. J. van der Hooft[2], Marnix H. Medema[2], Roger G. Linington[1,]**

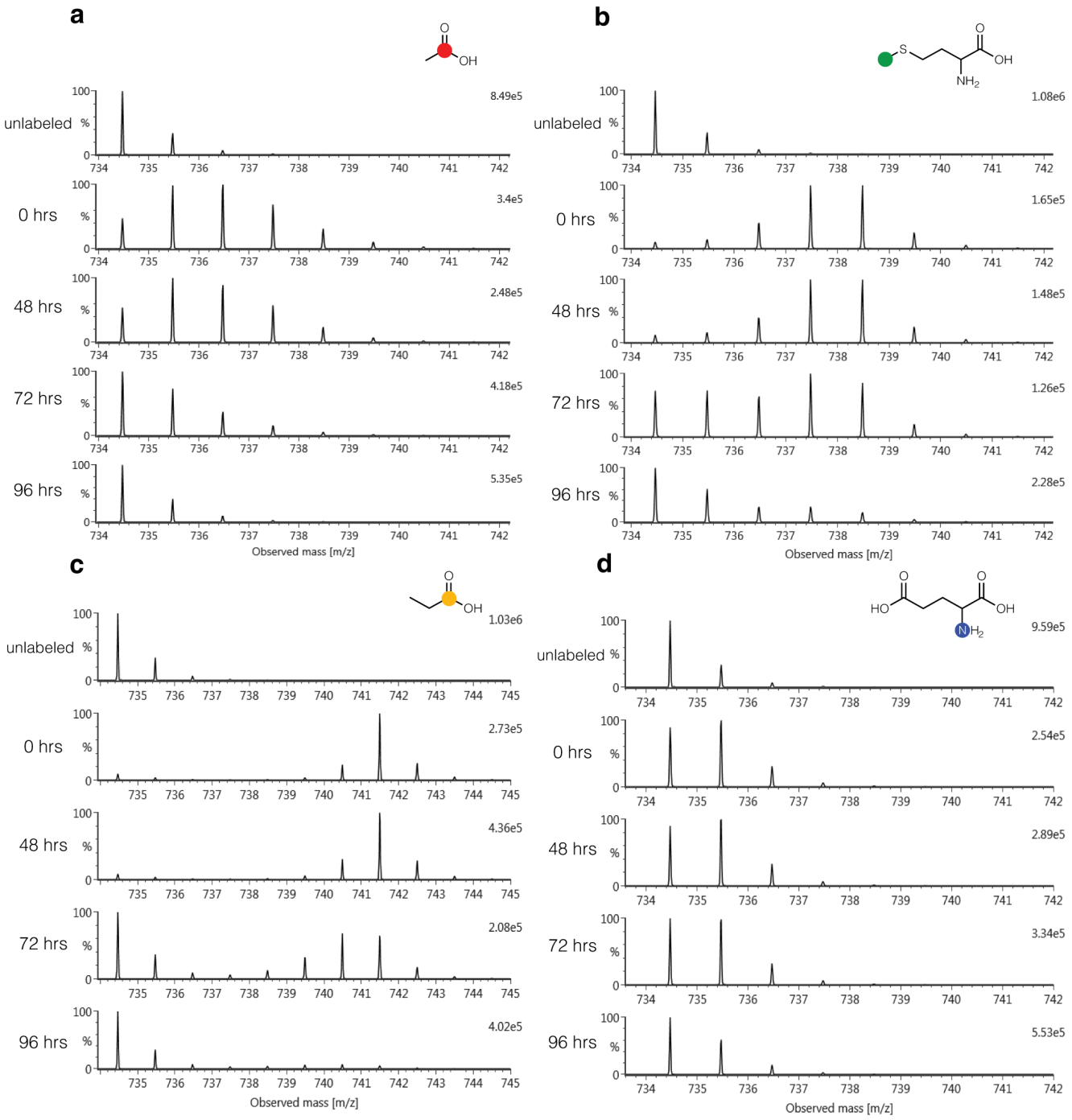
[1]Department of Chemistry, Simon Fraser University, Burnaby, Canada
[2]Bioinformatics Group, Wageningen University, Wageningen, The Netherlands

Supporting Information



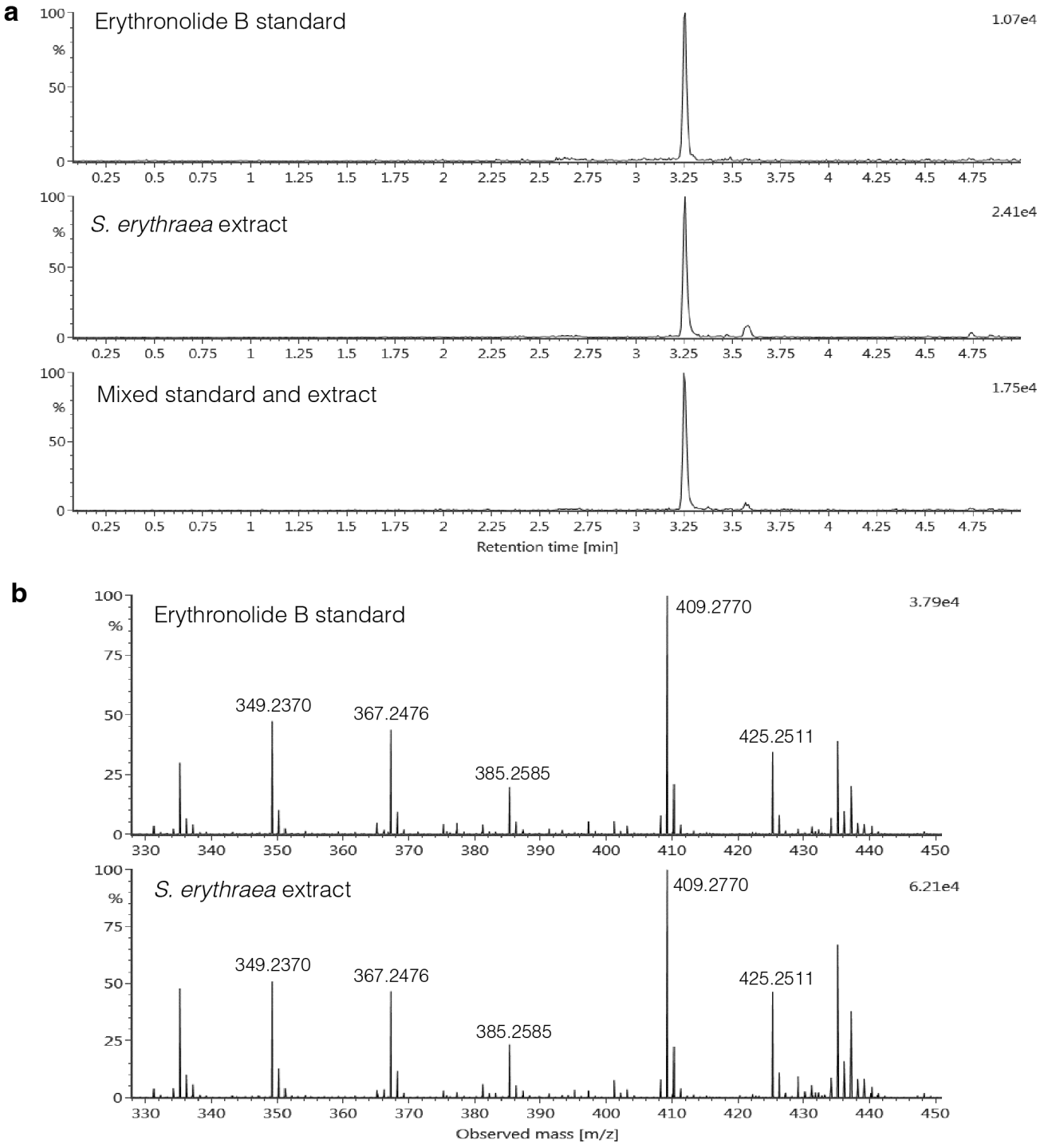
Supplementary Figure 1: Erythromycin A (**1**) SIL incorporation by concentration

Mass spectra of erythromycin A (**1**) from *S. erythraea* extracts supplemented with [1-^{13}C]acetate (a), [methyl-^{13}C]methionine (b), [1-^{13}C]propionate (c), and [1-^{15}N]glutamate (d) at the concentrations indicated. SIL feedstocks were added at the time of inoculation (0 hours) in all conditions.



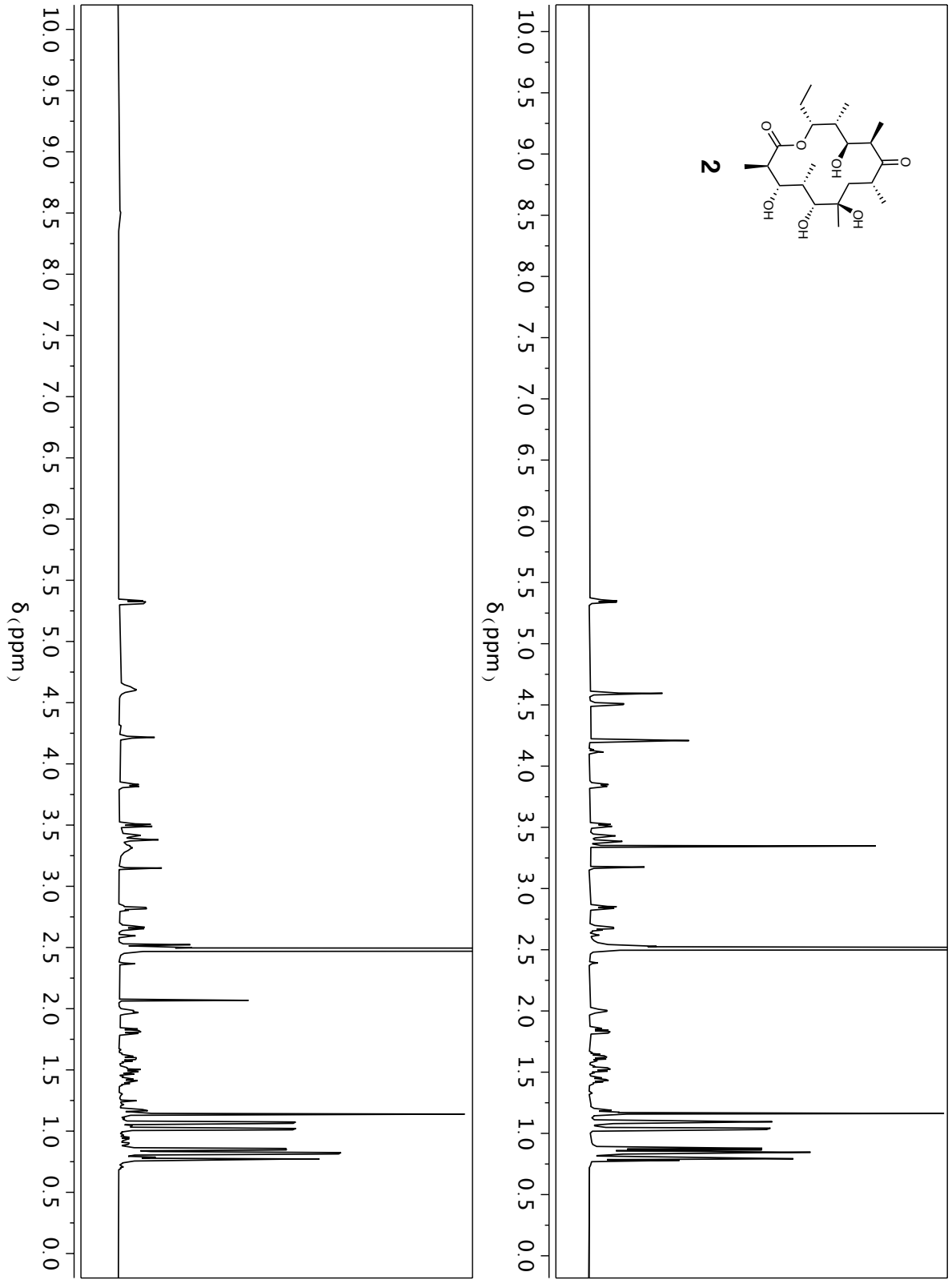
Supplementary Figure 2: Erythromycin A (**1**) SIL incorporation by day of SIL addition

Mass spectra of erythromycin A (**1**) from *S. erythraea* extracts supplemented with [1-^{13}C]acetate (30 Mm) (a), [methyl-^{13}C]methionine (5 mM) (b), [1-^{13}C]propionate (30 mM) (c), and [1-^{15}N]glutamate (10 mM) (d). Each SIL precursor was added to the culture at the hours indicated following inoculation. Addition of SIL at 0 hrs, or the same time as inoculation, consistently produced a higher extent of labeling.

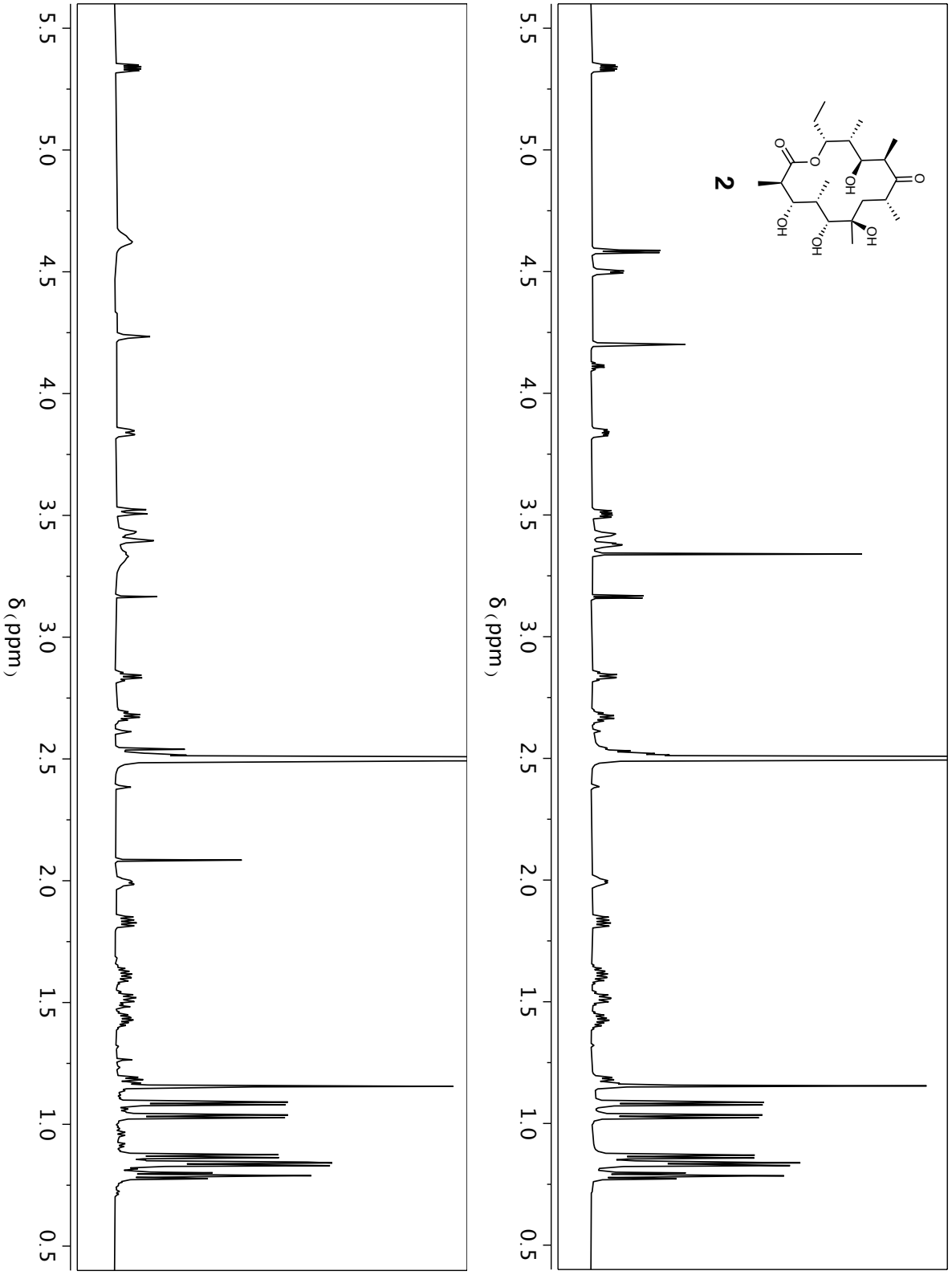


Supplementary Figure 3: Erythronolide B (**2**) standard co-injection and mass spectra

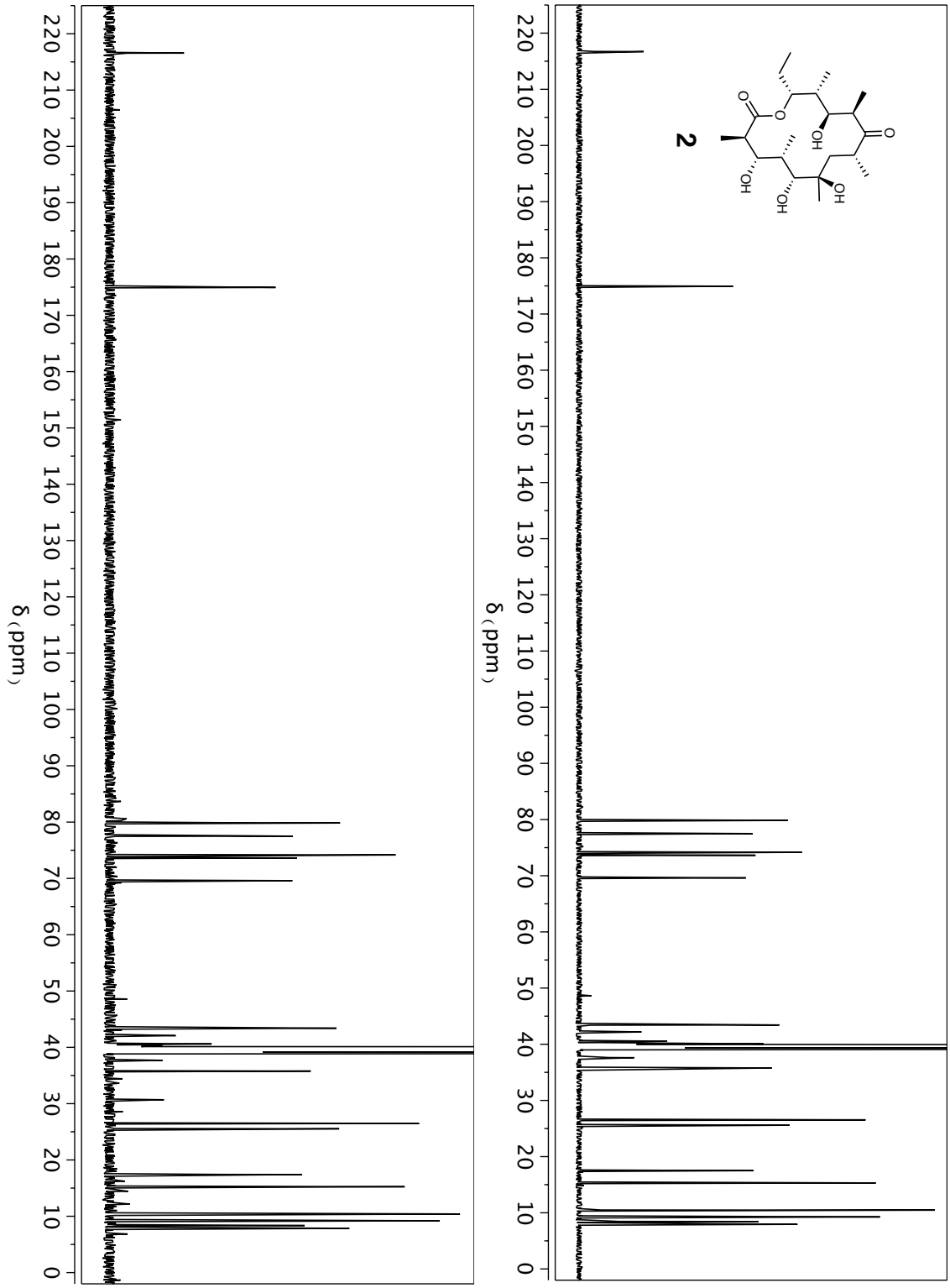
(a) Extracted ion chromatograms for **2** in commercial standard(top), *S. erythraea* extract (center), and mixed sample (bottom). (b) Mass spectra of a commercial standard of **2** (top) and in an *S. erythraea* extract (bottom).



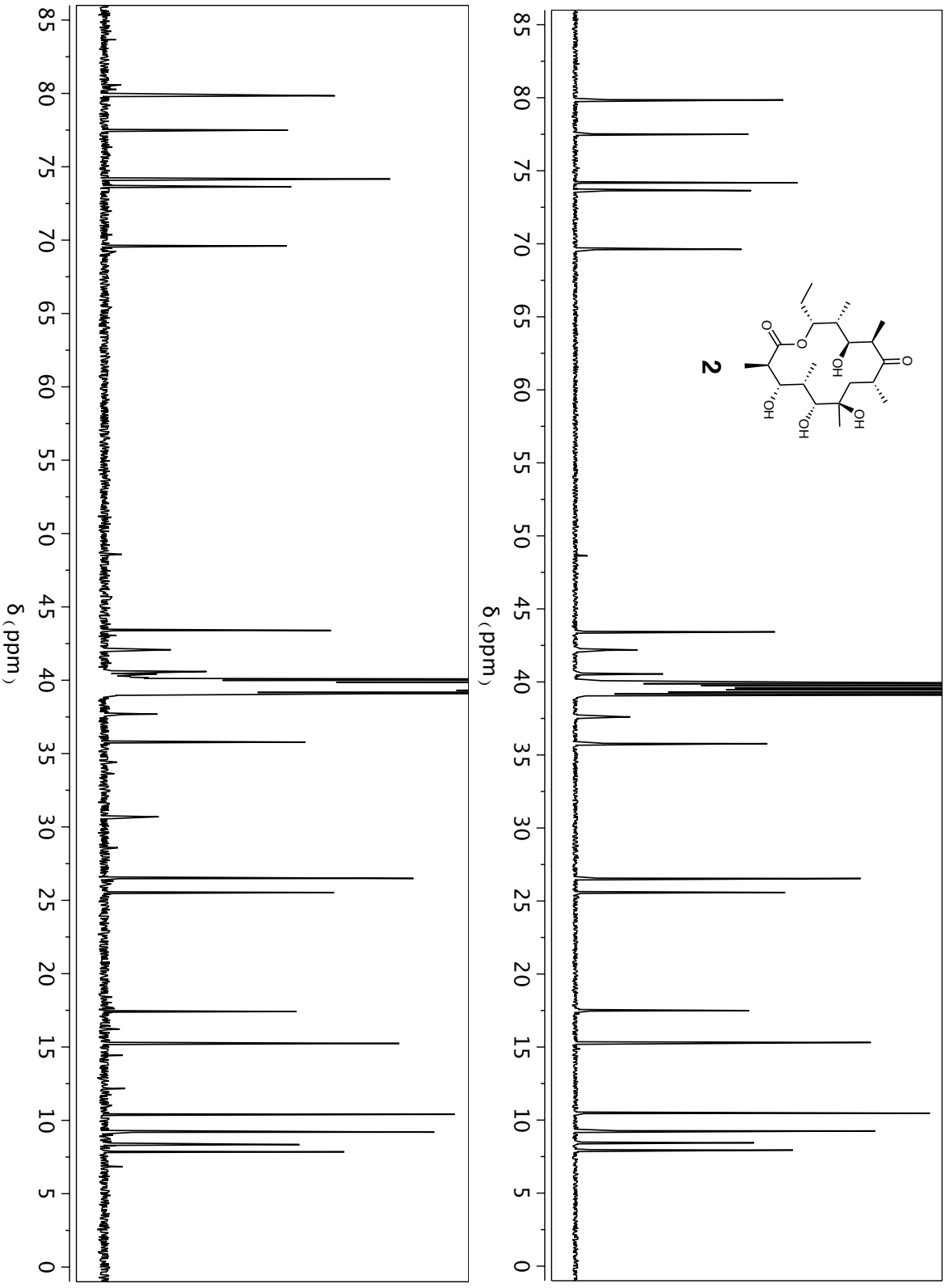
Supplementary Figure 4: ^{1}H NMR spectra of erythronolide B (**2)** acquired in DMSO-d_6 at 600 MHz Commercial standard (top) and isolated compound (bottom).



Supplementary Figure 5: Expansion of Supplementary Figure 4
Commercial standard (top) and isolated compound (bottom).

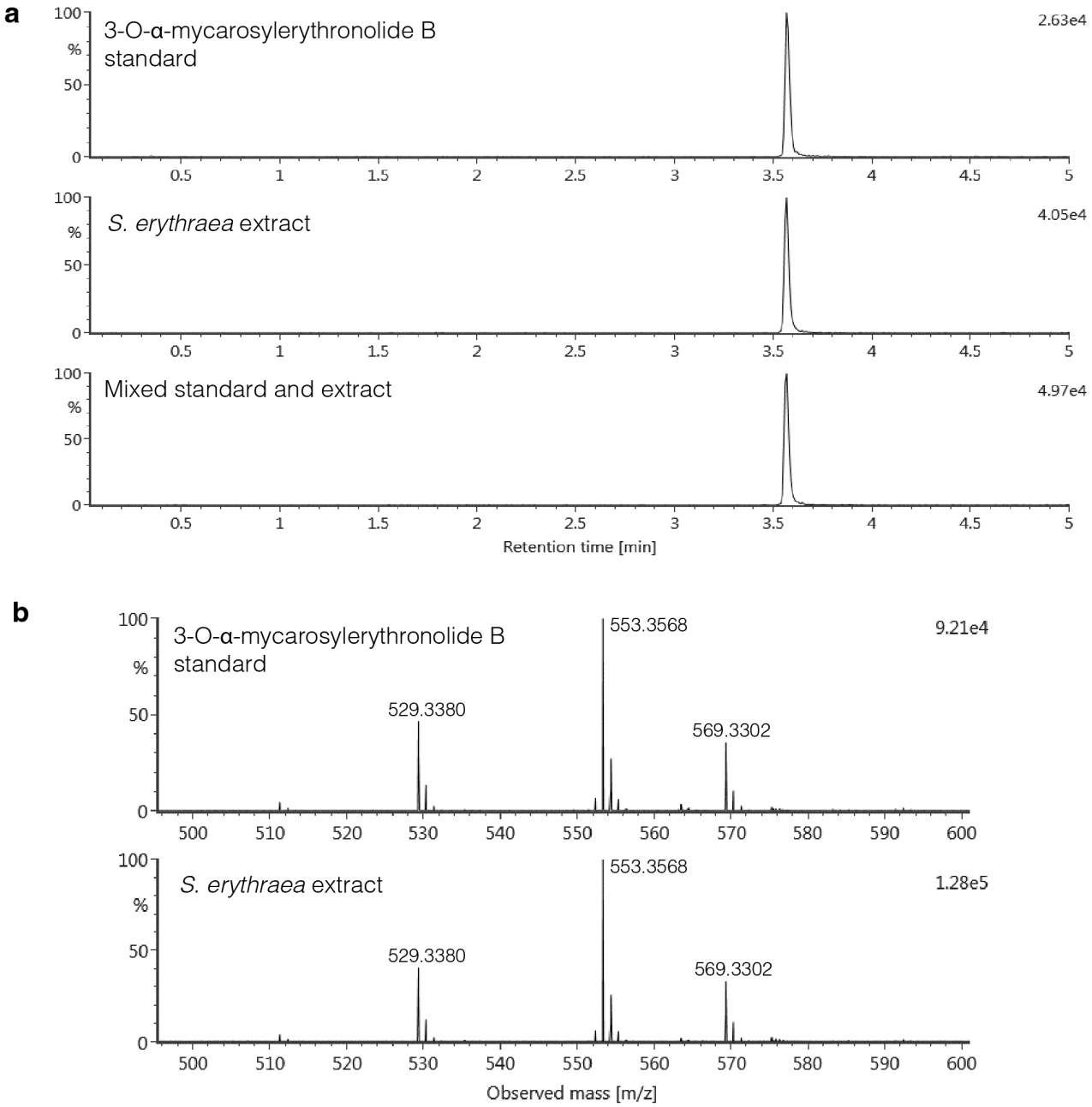


Supplementary Figure 6: ^{13}C NMR spectra of erythronolide B (**2)** acquired in DMSO-d6 at 150 MHz Commercial standard (top) and isolated compound (bottom).



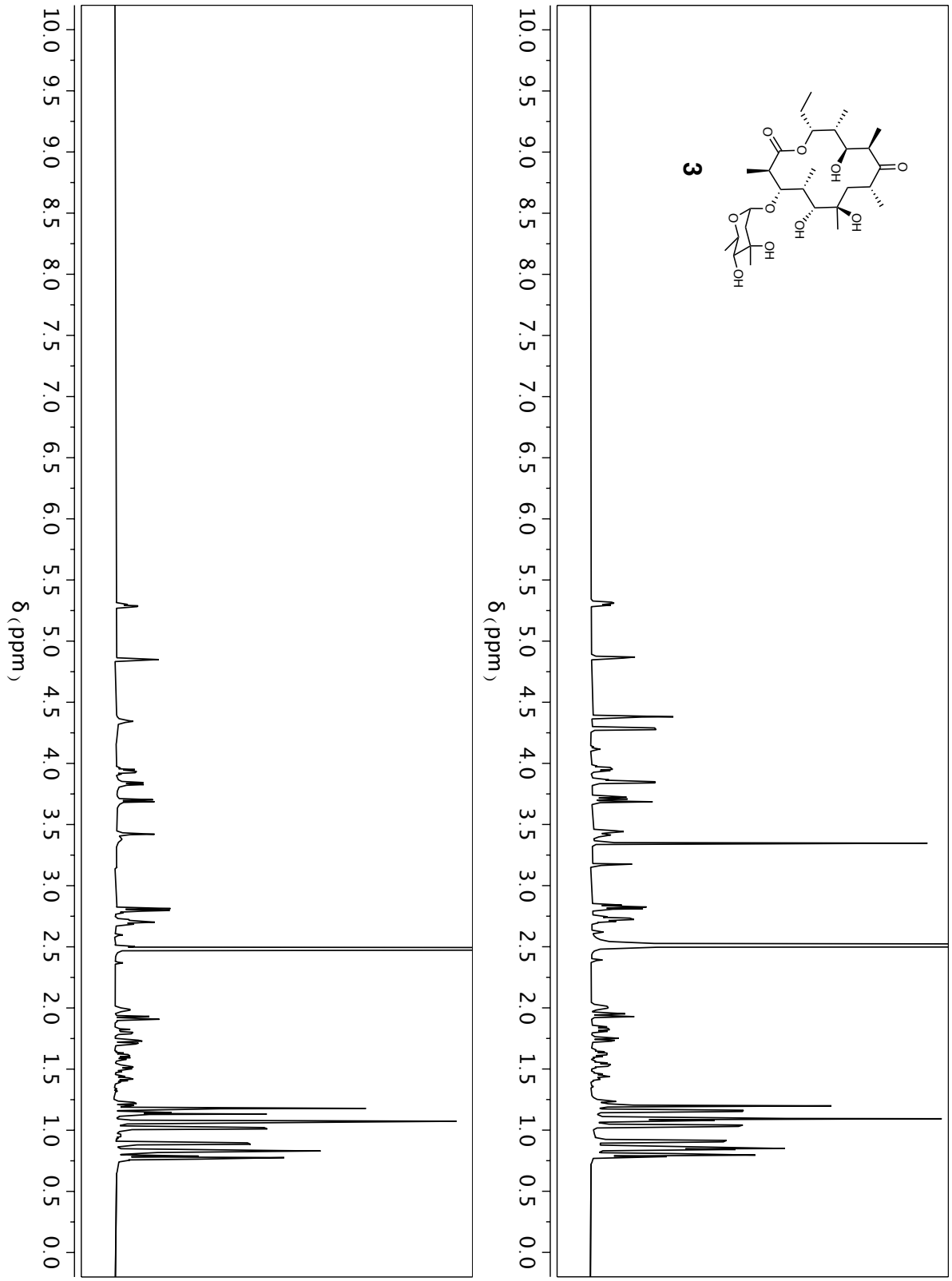
Supplementary Figure 7: Expansion of Supplementary Figure 6

Commercial standard (top) and isolated compound (bottom).



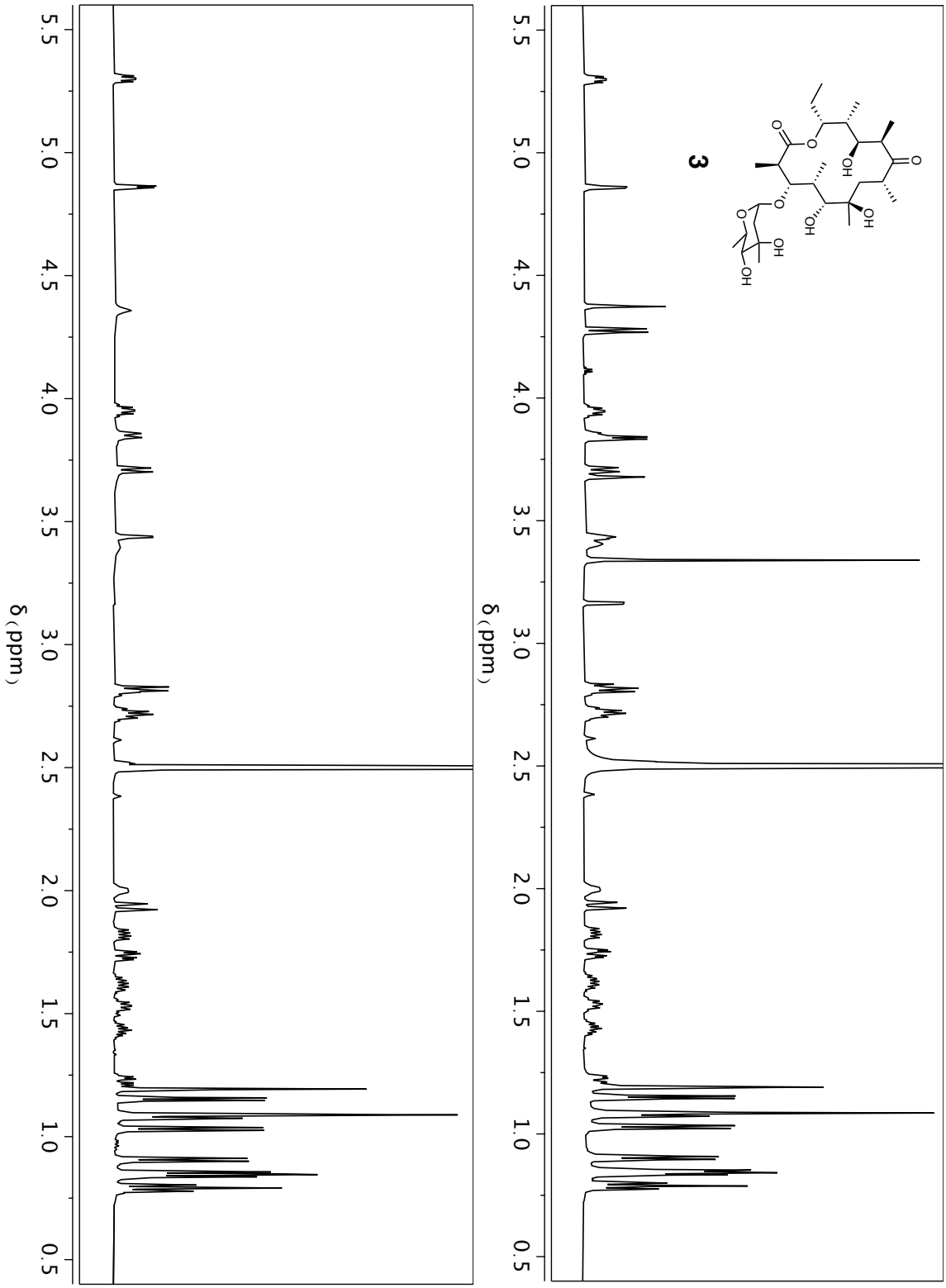
Supplementary Figure 8: 3-O-α-mycarosylerythronolide B (**3**) standard co-injection and MS

(a) Extracted ion chromatograms for **3** in commercial standard (top), *S. erythraea* extract (center), and mixed sample (bottom). (b) Mass spectra of standard of **3** (top) and in an *S. erythraea* extract (bottom).

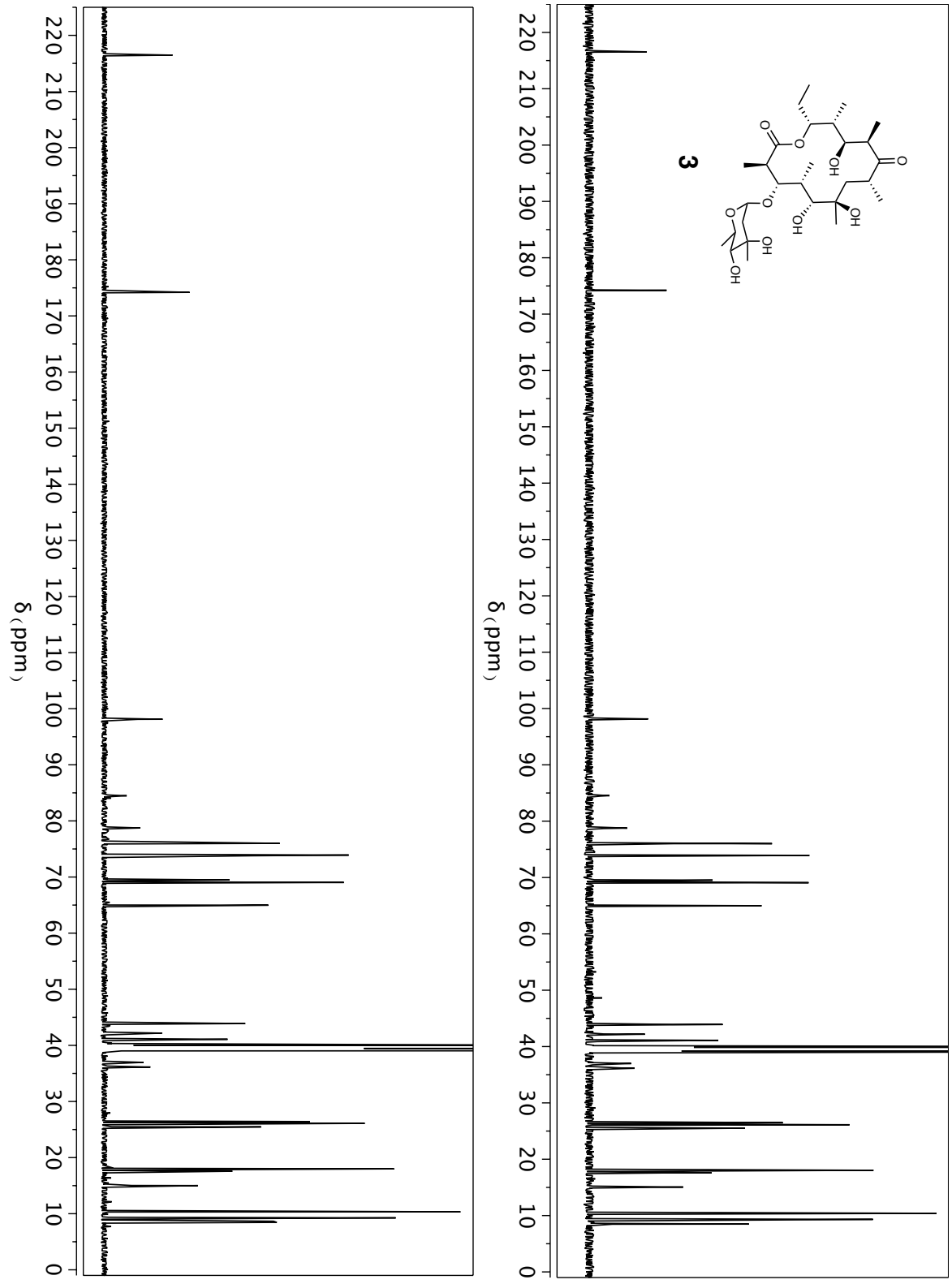


Supplementary Figure 9: ^{1}H NMR spectra of 3-O-α-mycarosylerythronolide B (**3**) acquired in DMSO-d_6 at 600 MHz.

Commercial standard (top) and isolated compound (bottom)

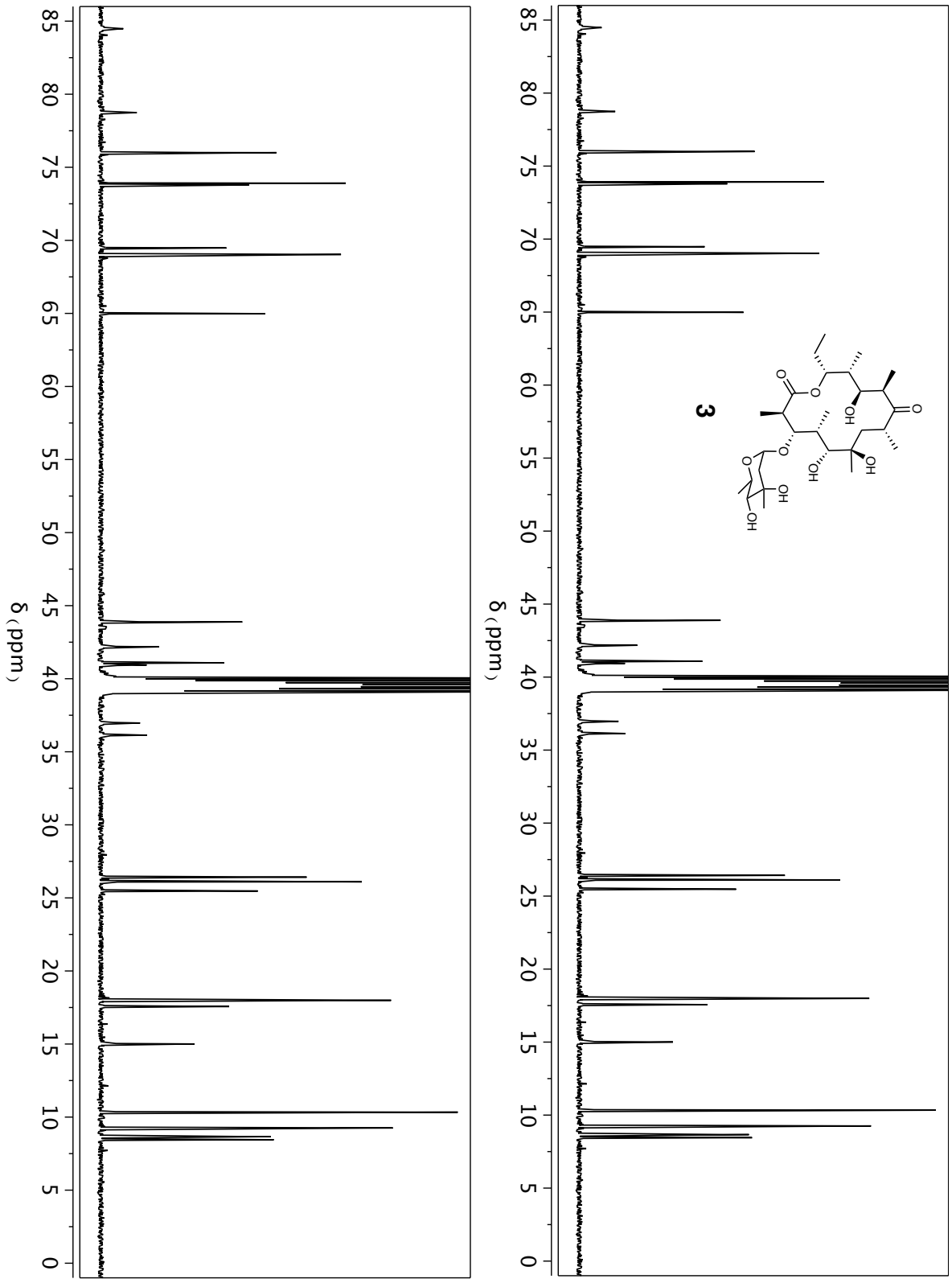


Supplementary Figure 10: Expansion of Supplementary Figure 9
Commercial standard (top) and isolated compound (bottom)



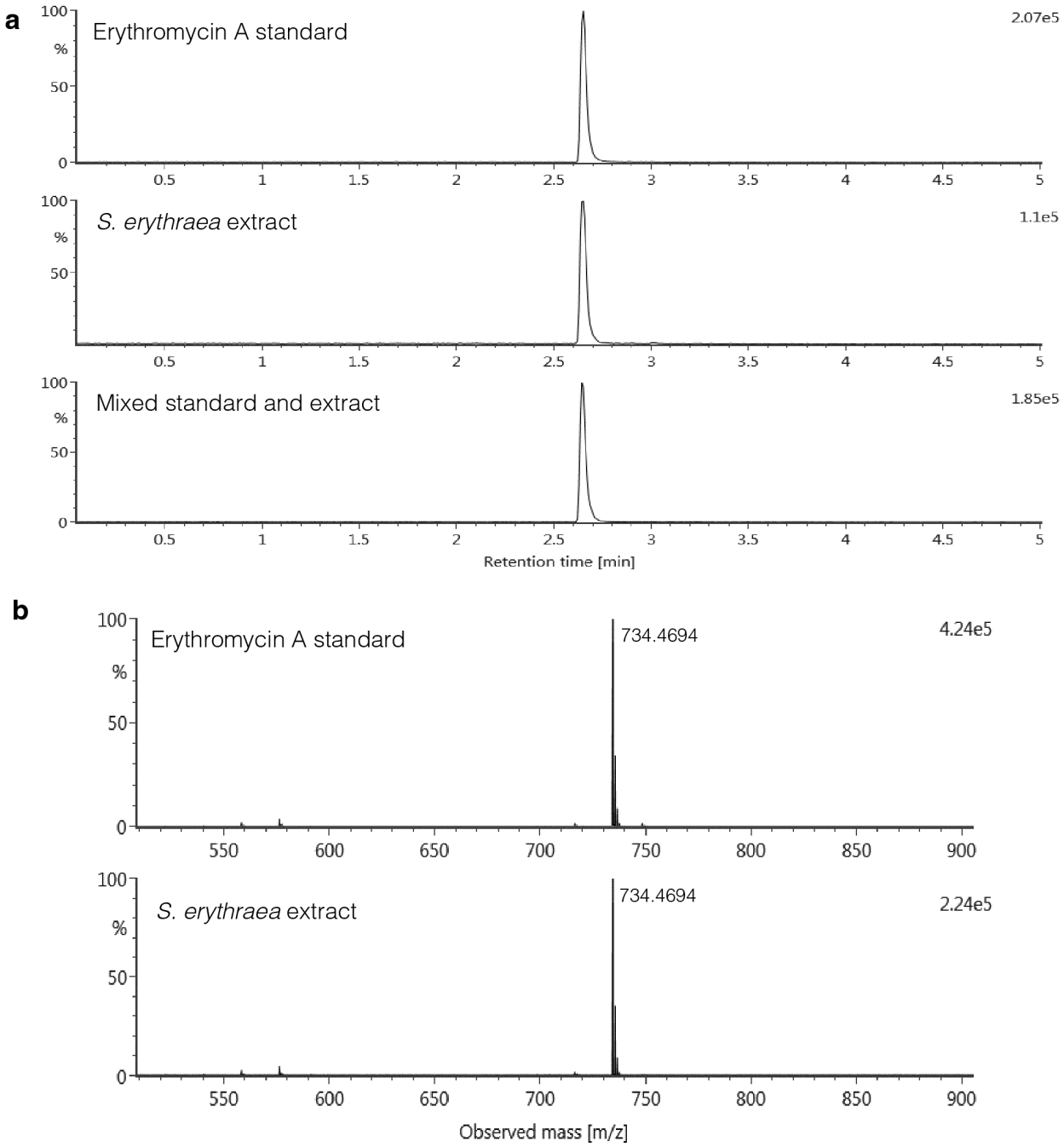
Supplementary Figure 11: ^{13}C NMR spectra of 3-O-α-mycarosylerythronolide B (**3**) acquired in DMSO-d_6 at 150 MHz.

Commercial standard (top) and isolated compound (bottom)



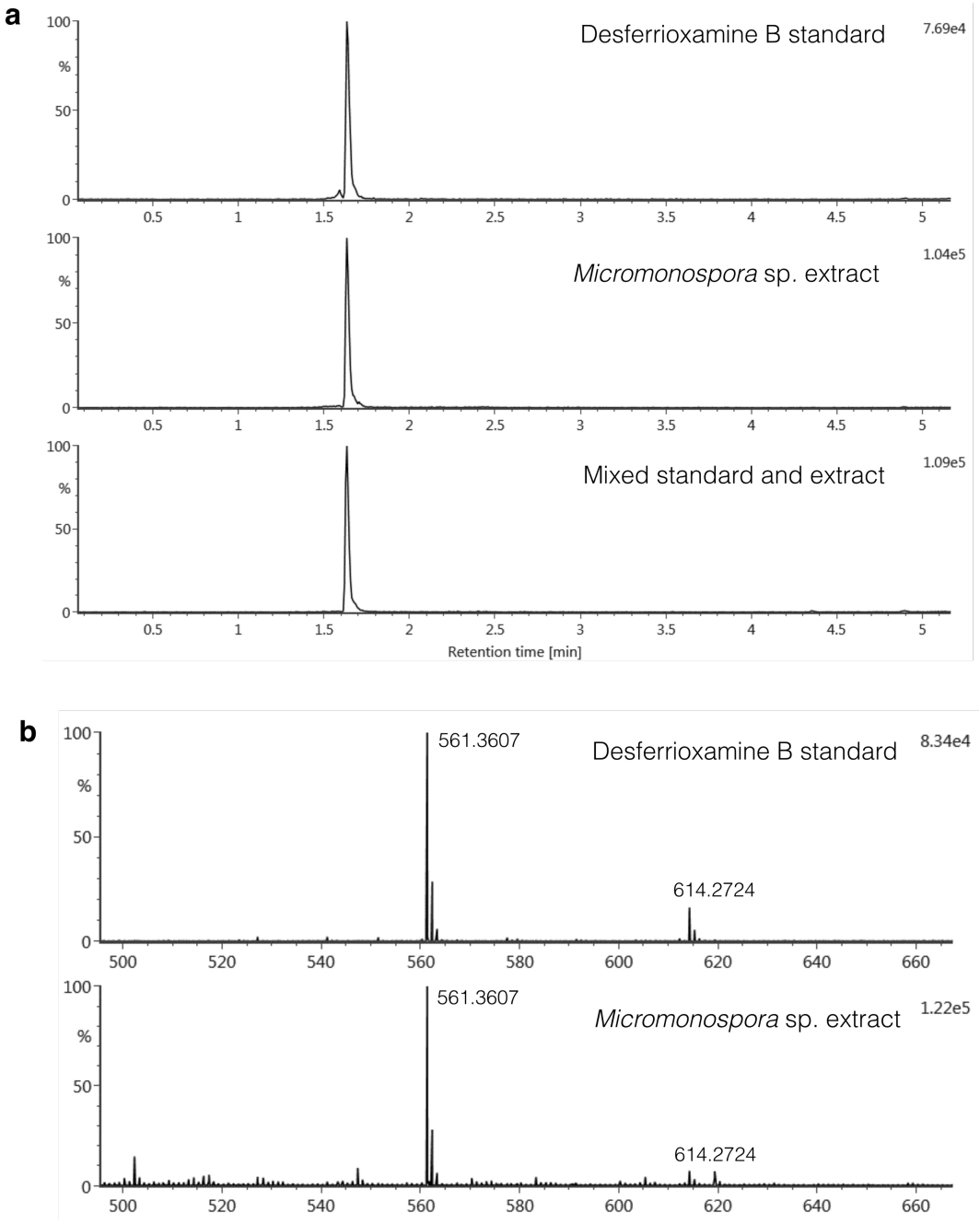
Supplementary Figure 12: Expansion of Supplementary Figure 11

Commercial standard (top) and isolated compound (bottom)



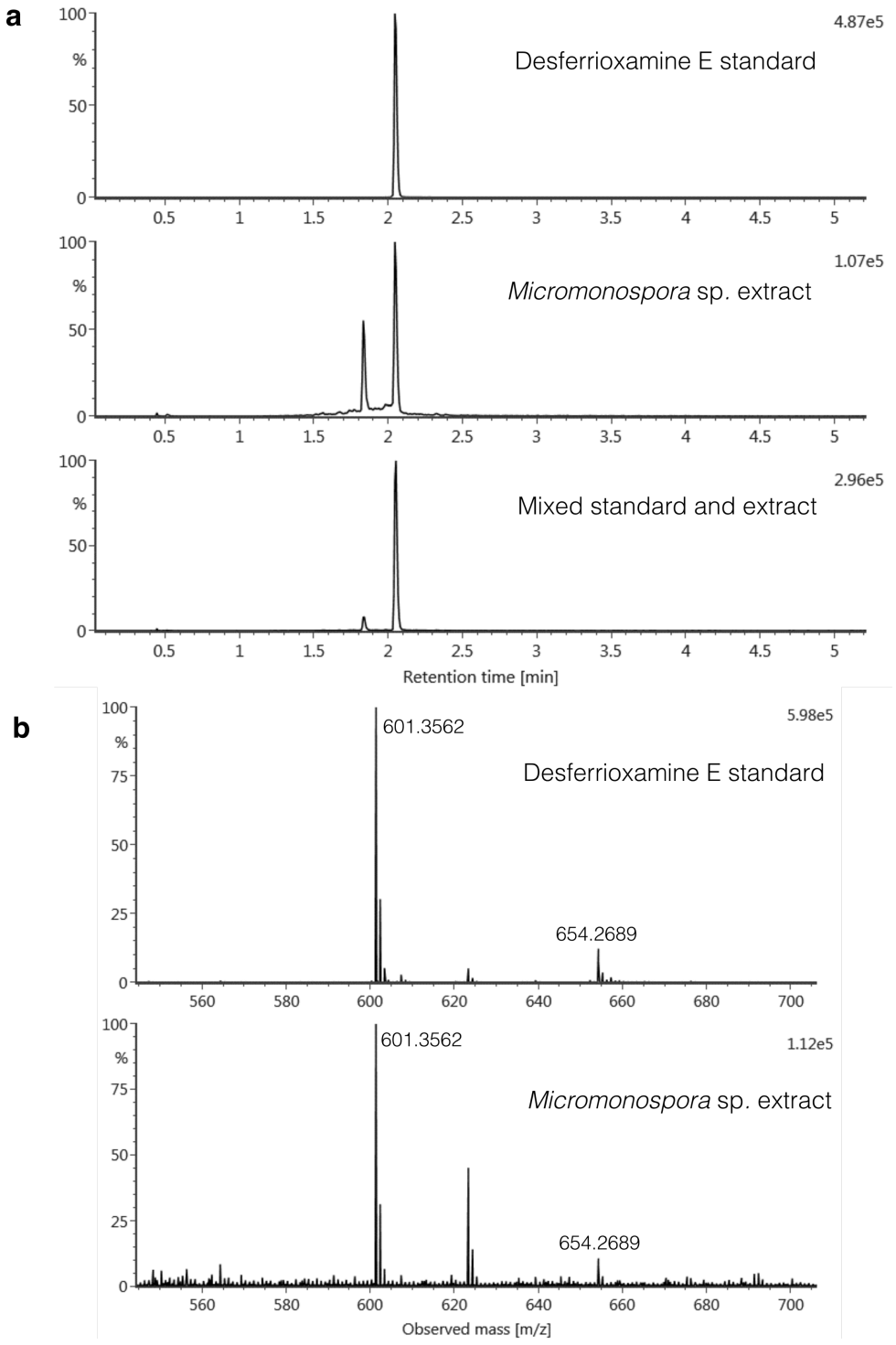
Supplementary Figure 13: Erythromycin A (**1**) standard co-injection and mass spectra

(a) Extracted ion chromatograms for **1** in commercial standard (top), *S. erythraea* extract (center), and mixed sample (bottom). (b) Mass spectra of standard of **1** (top) and in an *S. erythraea* extract (bottom).



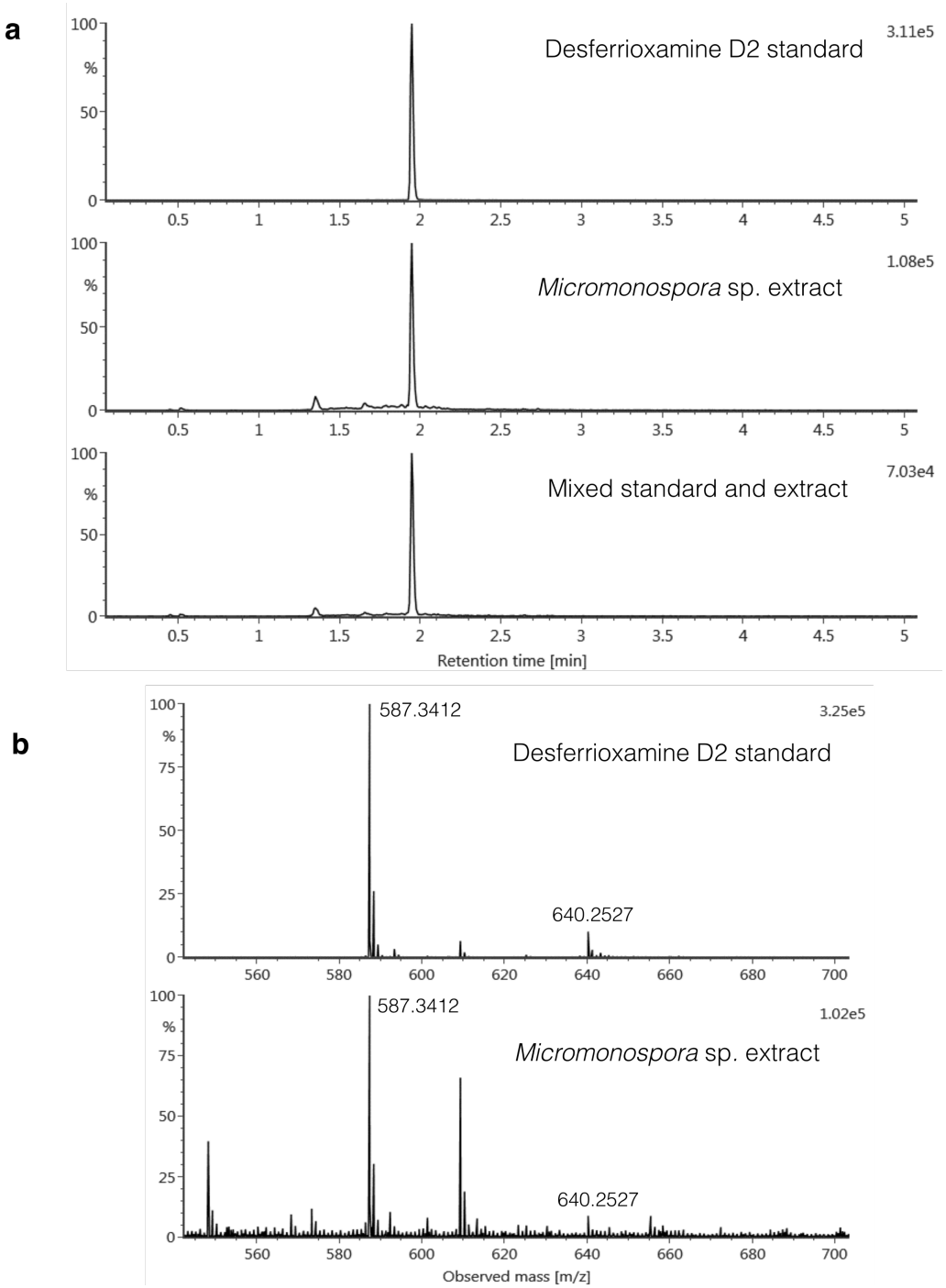
Supplementary Figure 14: Desferrioxamine B (**8**) standard co-injection and mass spectra

(a) Extracted ion chromatograms for **8** in commercial standard (top), *Micromonospora* sp. extract (center), and mixed sample (bottom). (b) Mass spectra of **8** standard (top) and in an *Micromonospora* sp. extract (bottom).



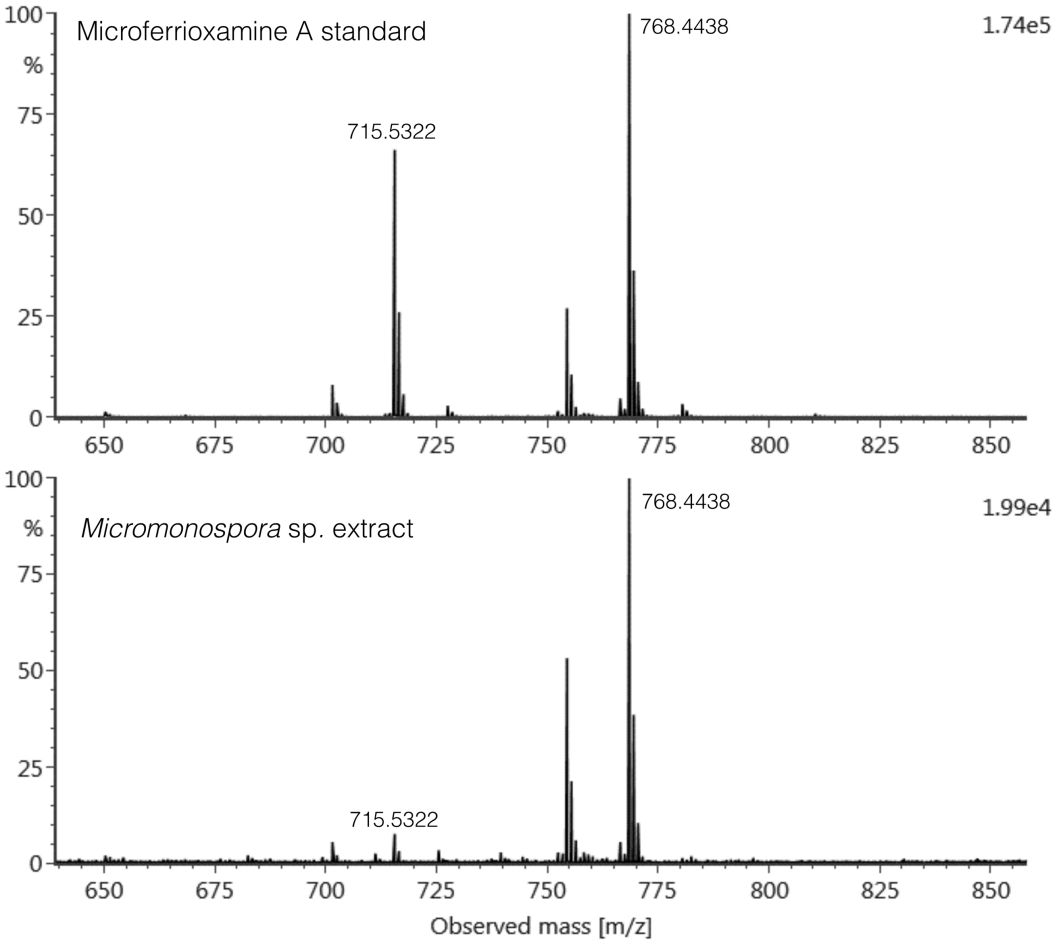
Supplementary Figure 15: Desferrioxamine E (**9**) standard co-injection and mass spectra

(a) Extracted ion chromatograms for **9** in standard (top), *Micromonospora* sp. extract (center), and mixed sample (bottom). (b) Mass spectra of **9** standard (top) and in an *Micromonospora* sp. extract (bottom).



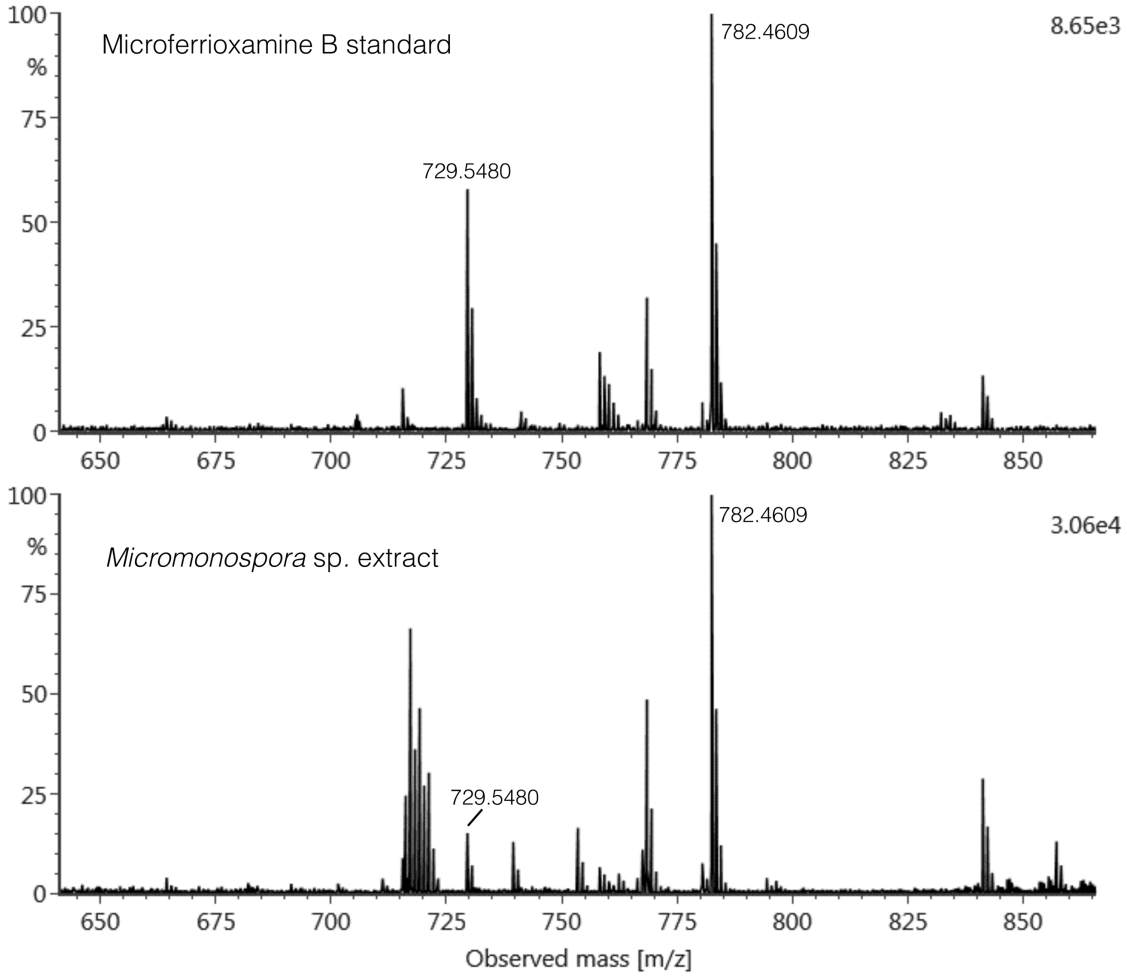
Supplementary Figure 16: Desferrioxamine D2 (**10**) standard co-injection and mass spectra

(a) Extracted ion chromatograms for **10** in standard (top), *Micromonospora* sp. extract (center), and mixed sample (bottom). (b) Mass spectra of **10** standard (top) and in an *Micromonospora* sp. extract (bottom).



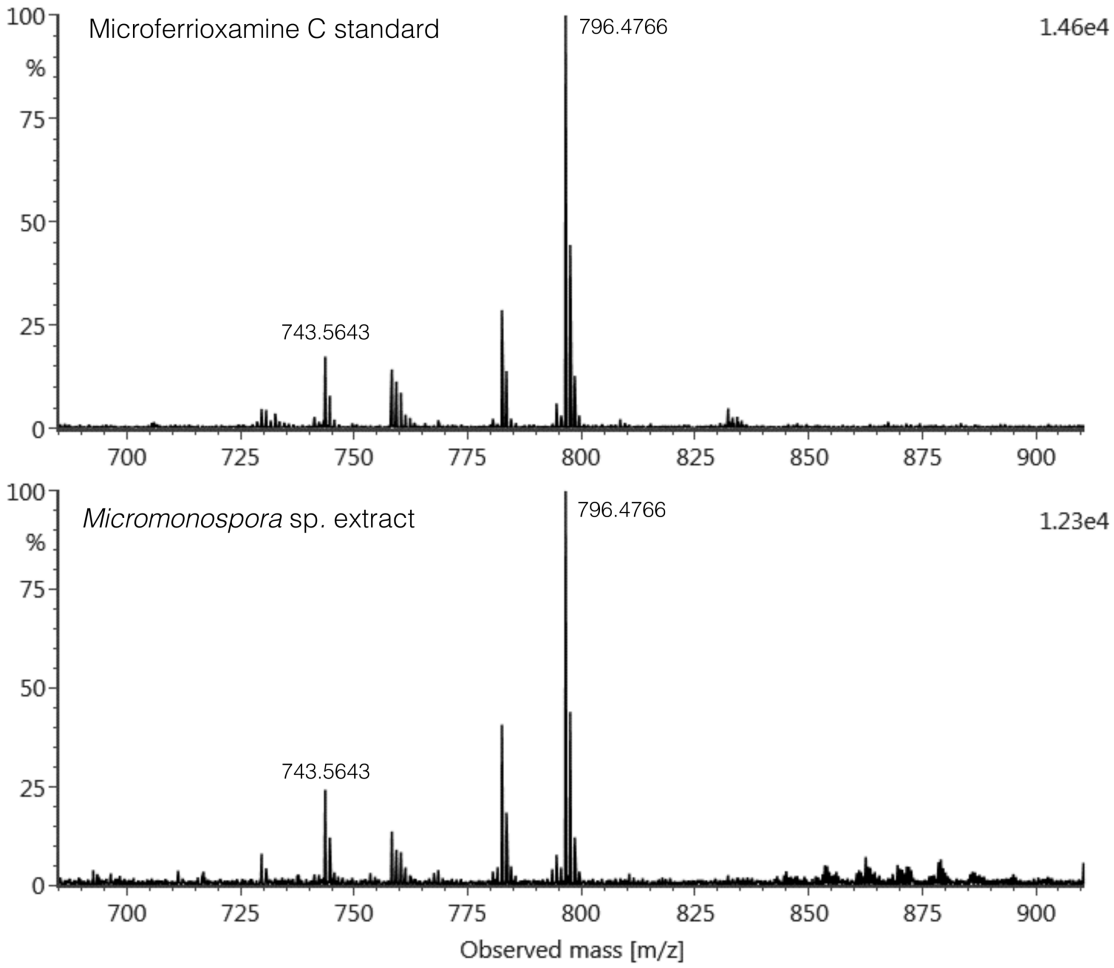
Supplementary Figure 17: Microferrioxamine A (**11**) mass spectra

Mass spectra of standard of **11** (top) and in an *Micromonospora* sp. extract (bottom).



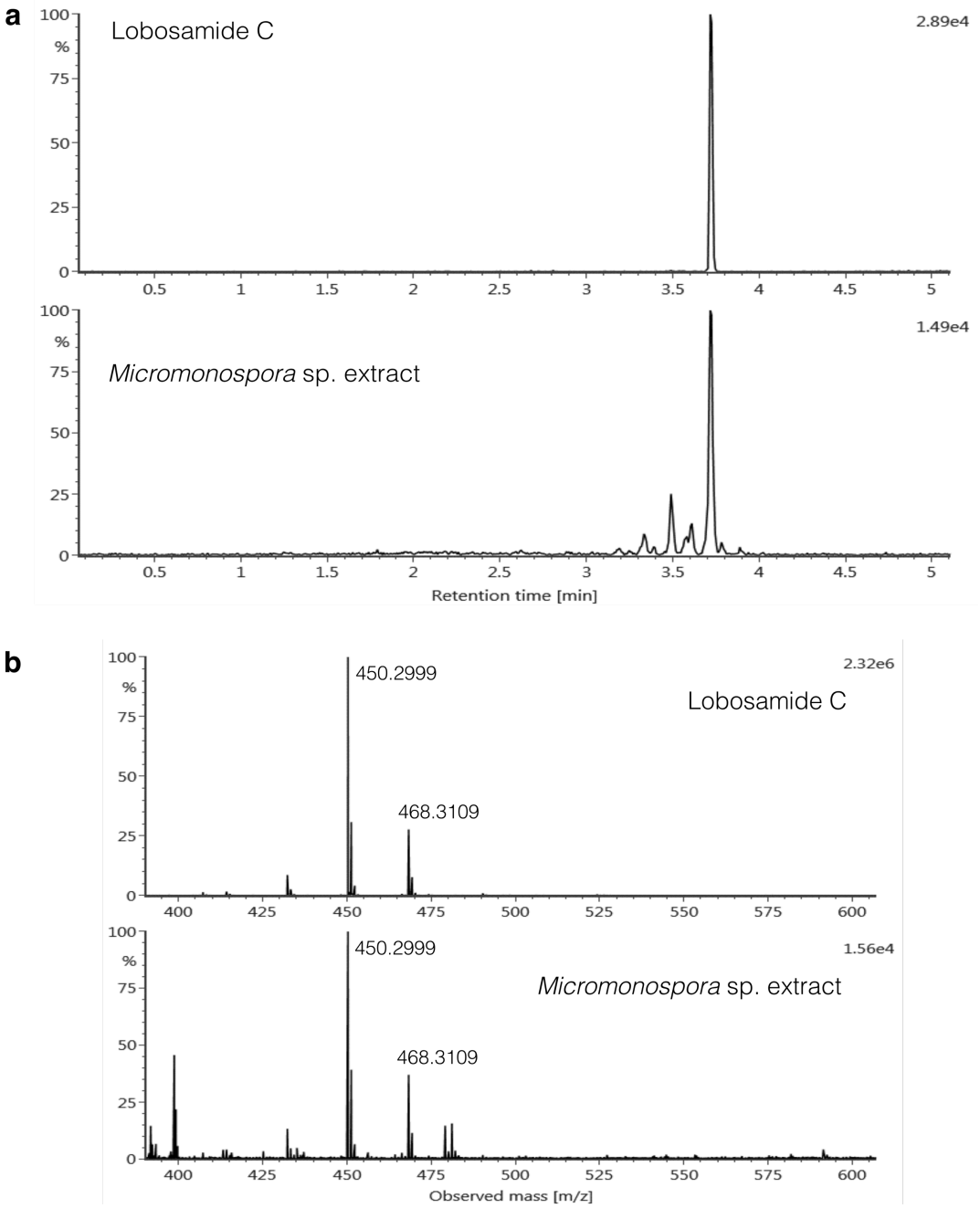
Supplementary Figure 18: Microferrioxamine B (**12**) mass spectra

Mass spectra of standard of **12** (top) and in an *Micromonospora* sp. extract (bottom).



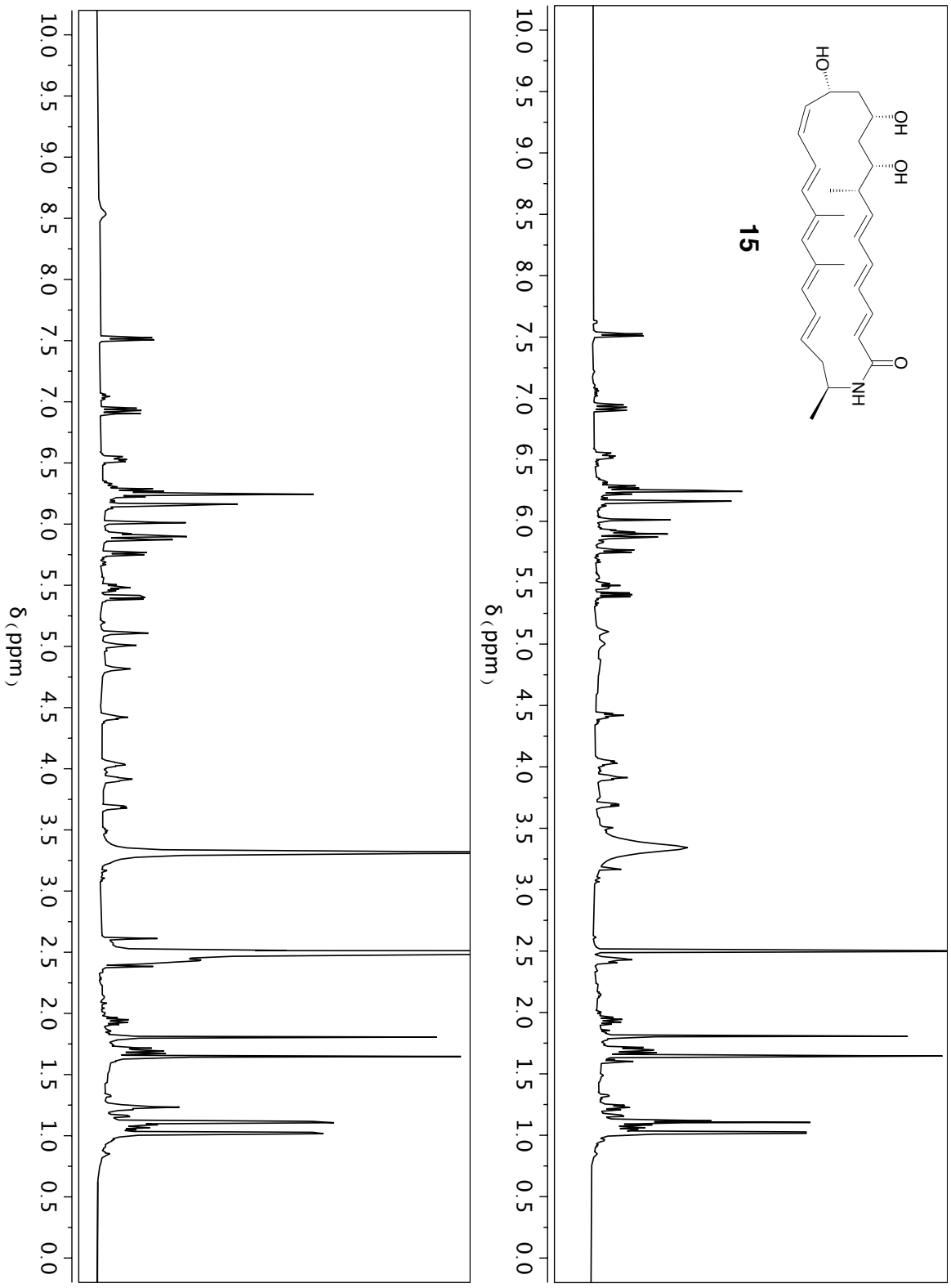
Supplementary Figure 19: Microferrioxamine C (**13**) mass spectra

Mass spectra of standard of **13** (top) and in an *Micromonospora* sp. extract (bottom).

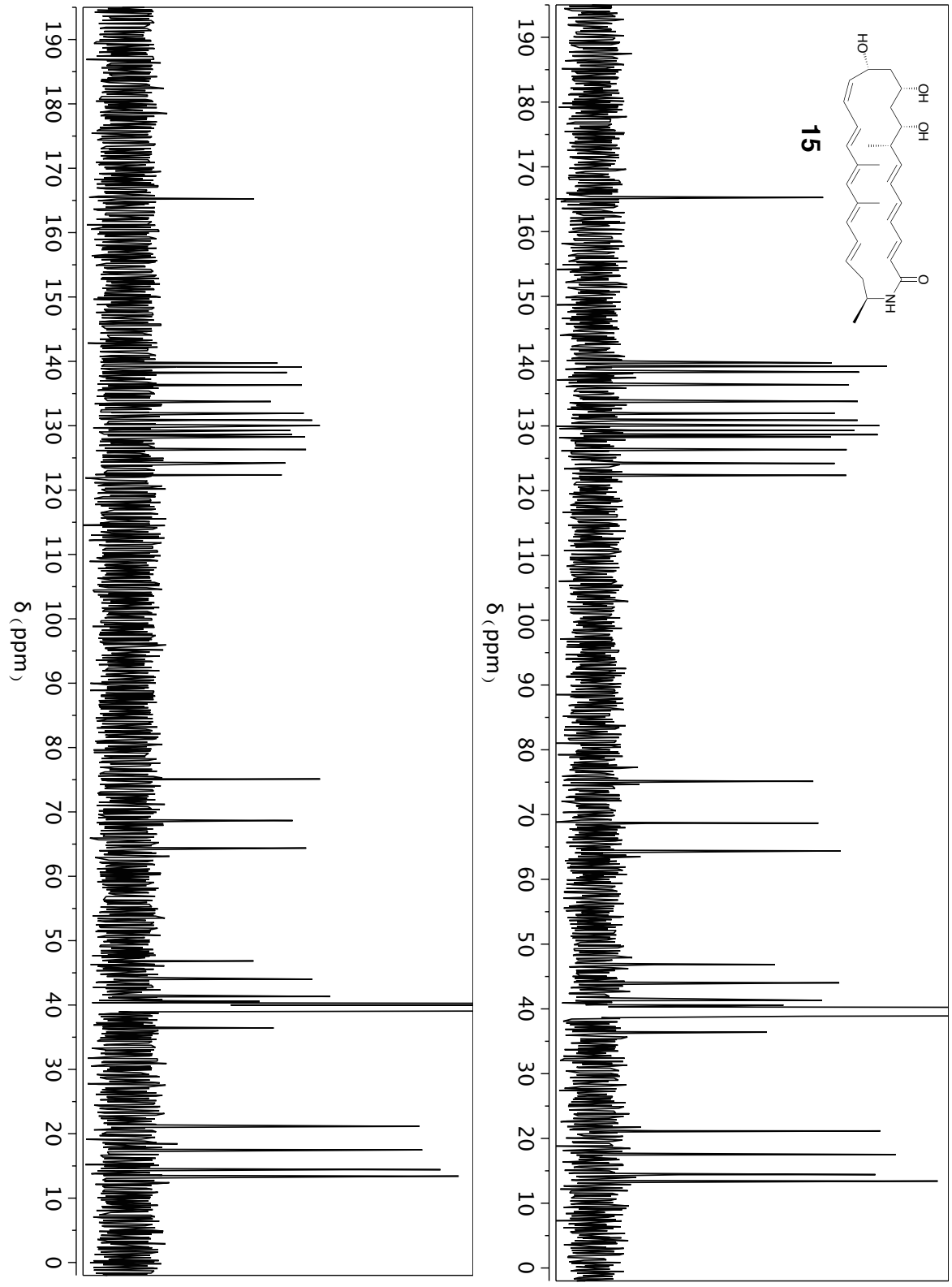


Supplementary Figure 20: Lobosamide C (**15**) chromatogram and mass spectra

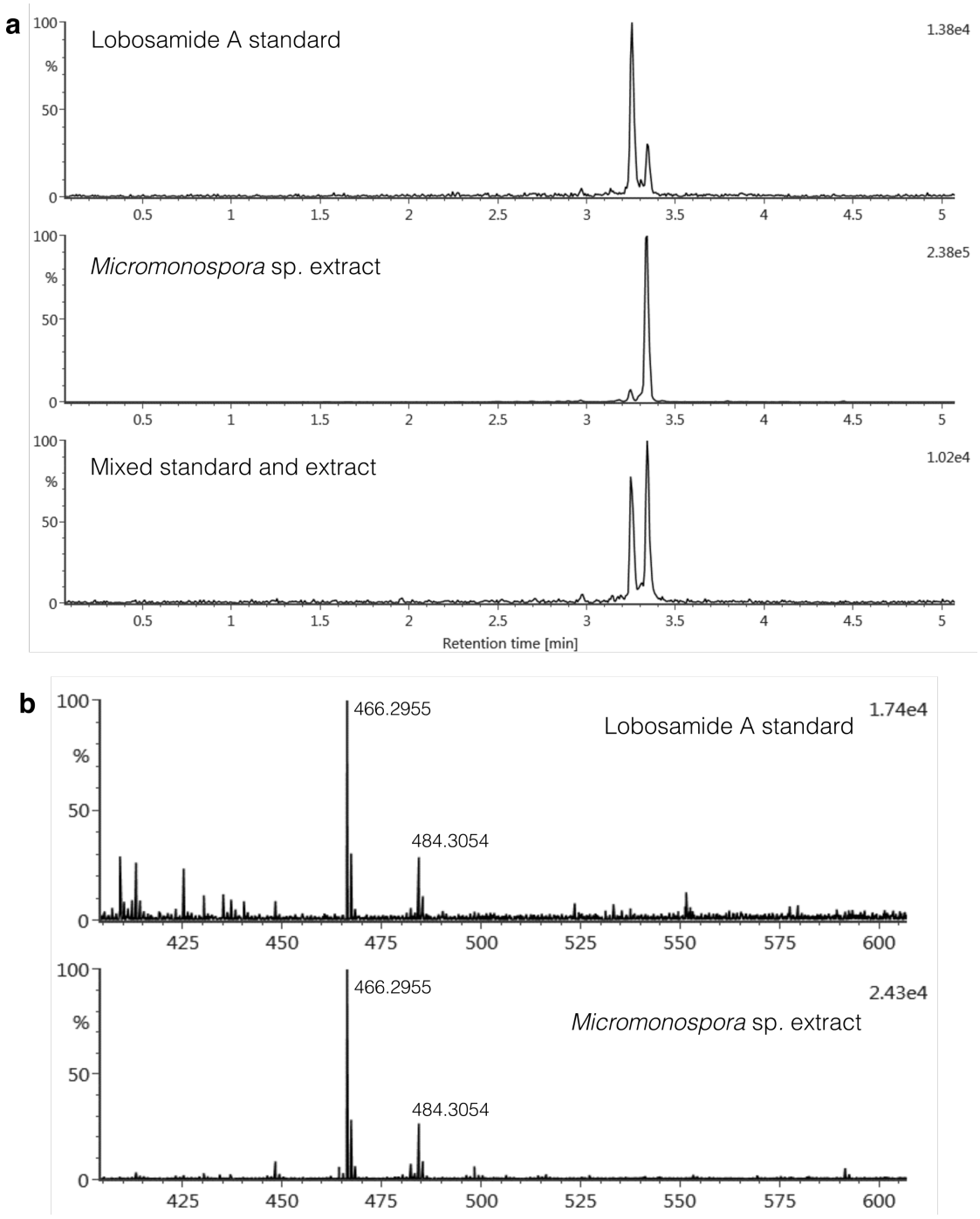
Mass spectra of isolated standard of **15** (top) and in the *Micromonospora* sp. extract (bottom).



Supplementary Figure 21: ^{1}H NMR spectra of lobosamide C (**15**) acquired in DMSO-d_6 at 600 MHz. Standard (top) and isolated compound (bottom)

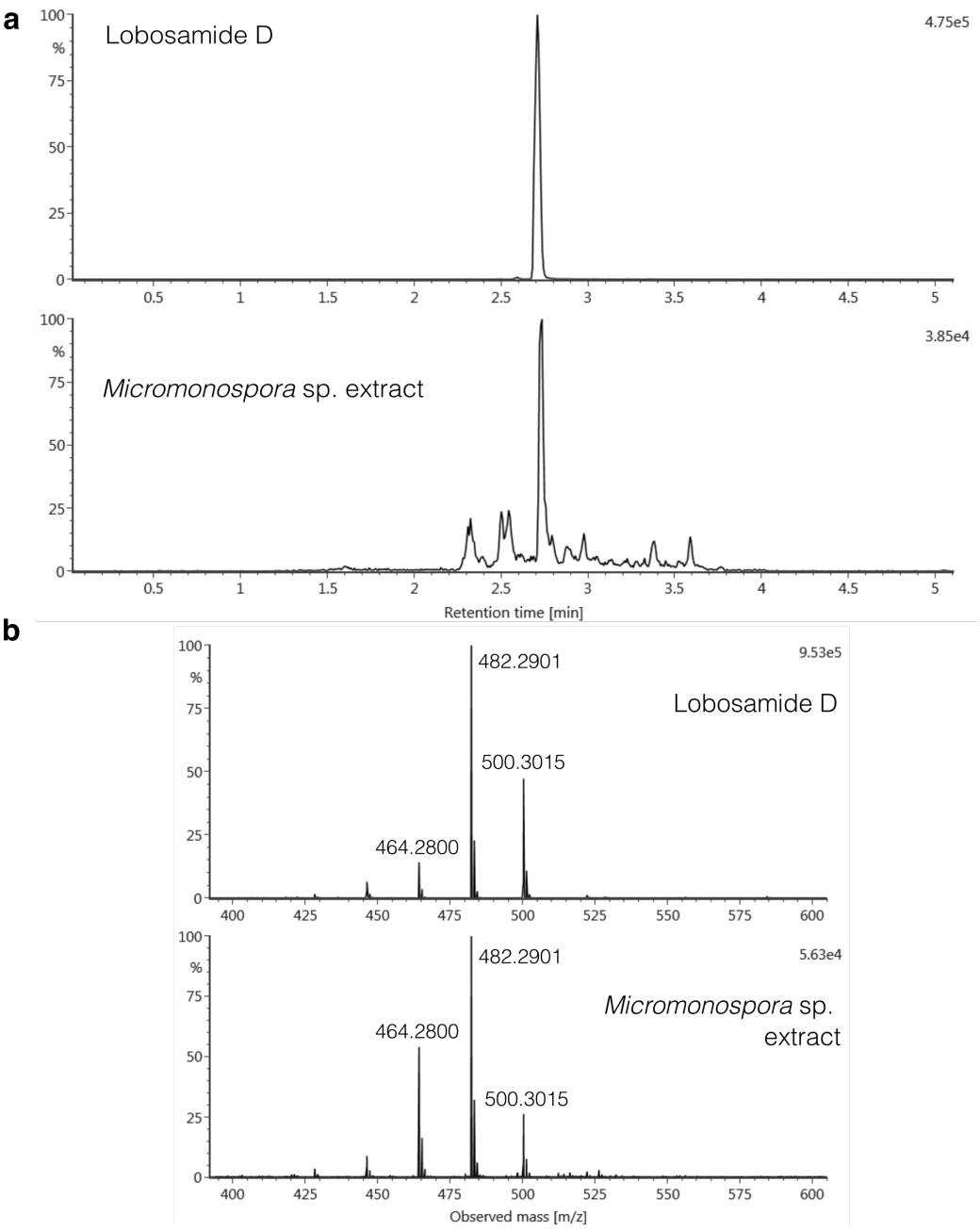


Supplementary Figure 22: ^{13}C NMR spectra of lobosamide C (**15**) acquired in DMSO-d_6 at 150 MHz. Commercial standard (top) and isolated compound (bottom)



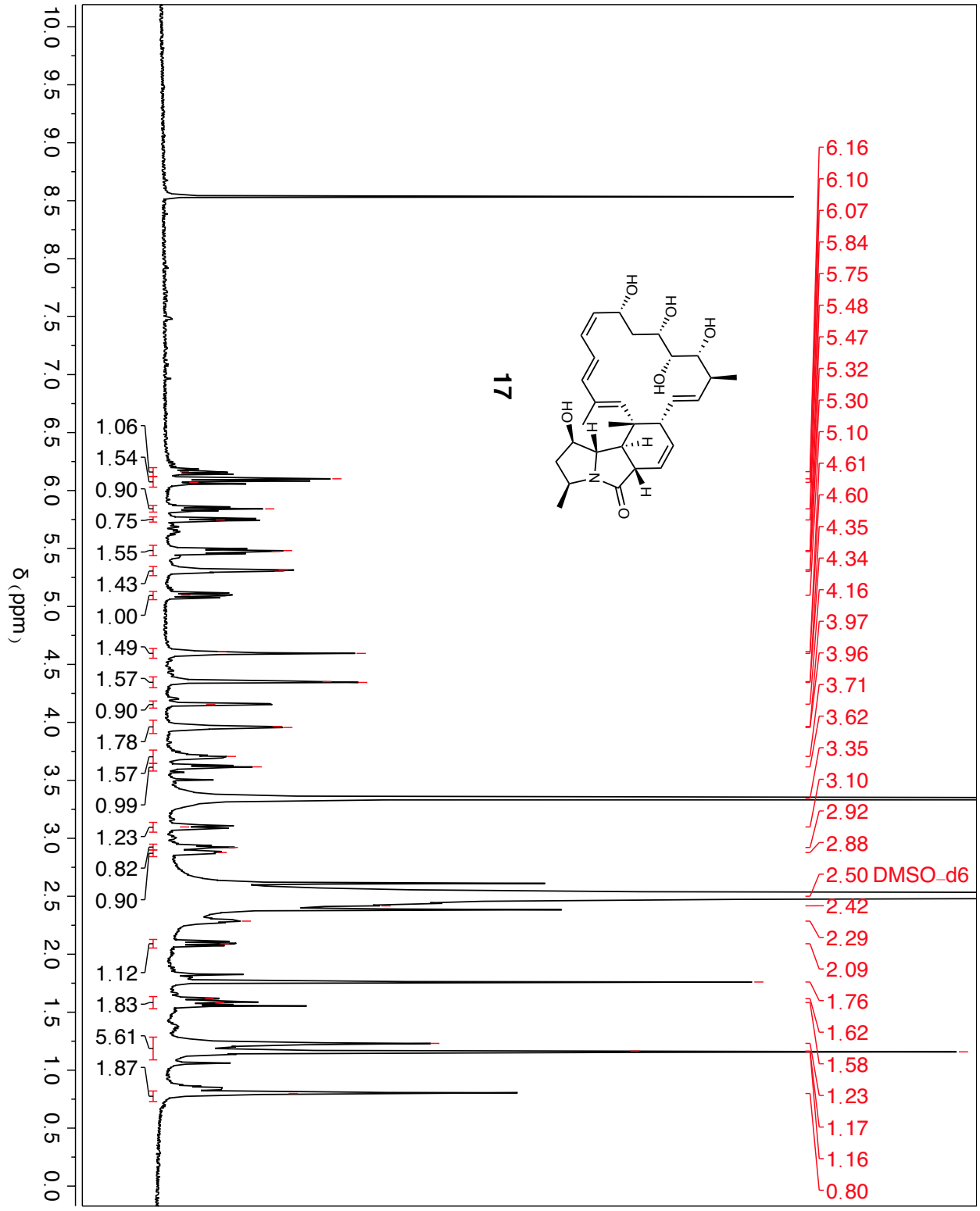
Supplementary Figure 23: Lobosamide A (**16**) mass spectra

Mass spectra of standard of **16** (top) and in the *Micromonospora* sp. extract (bottom).

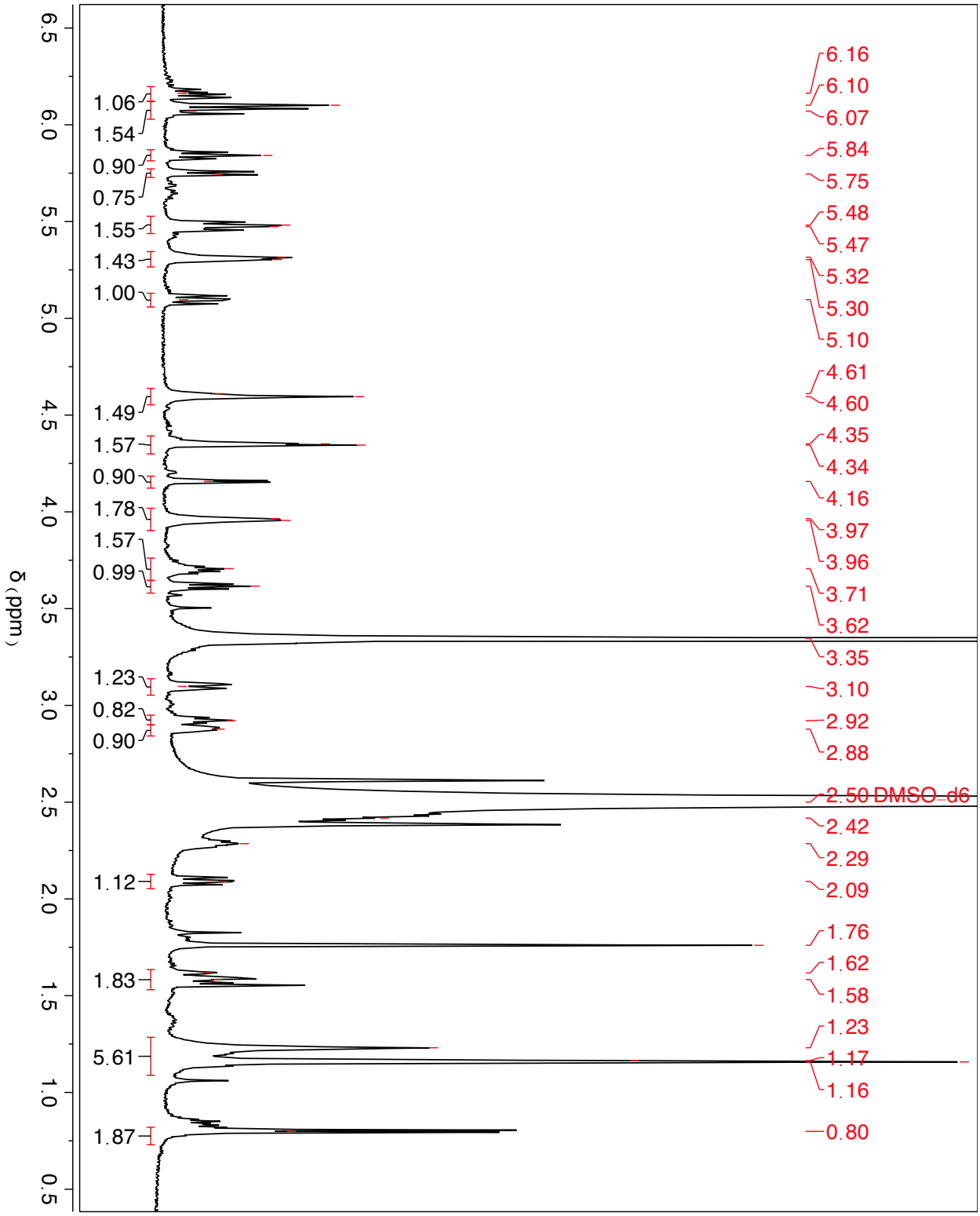


Supplementary Figure 24: Lobosamide D (**17**) chromatogram and mass spectra

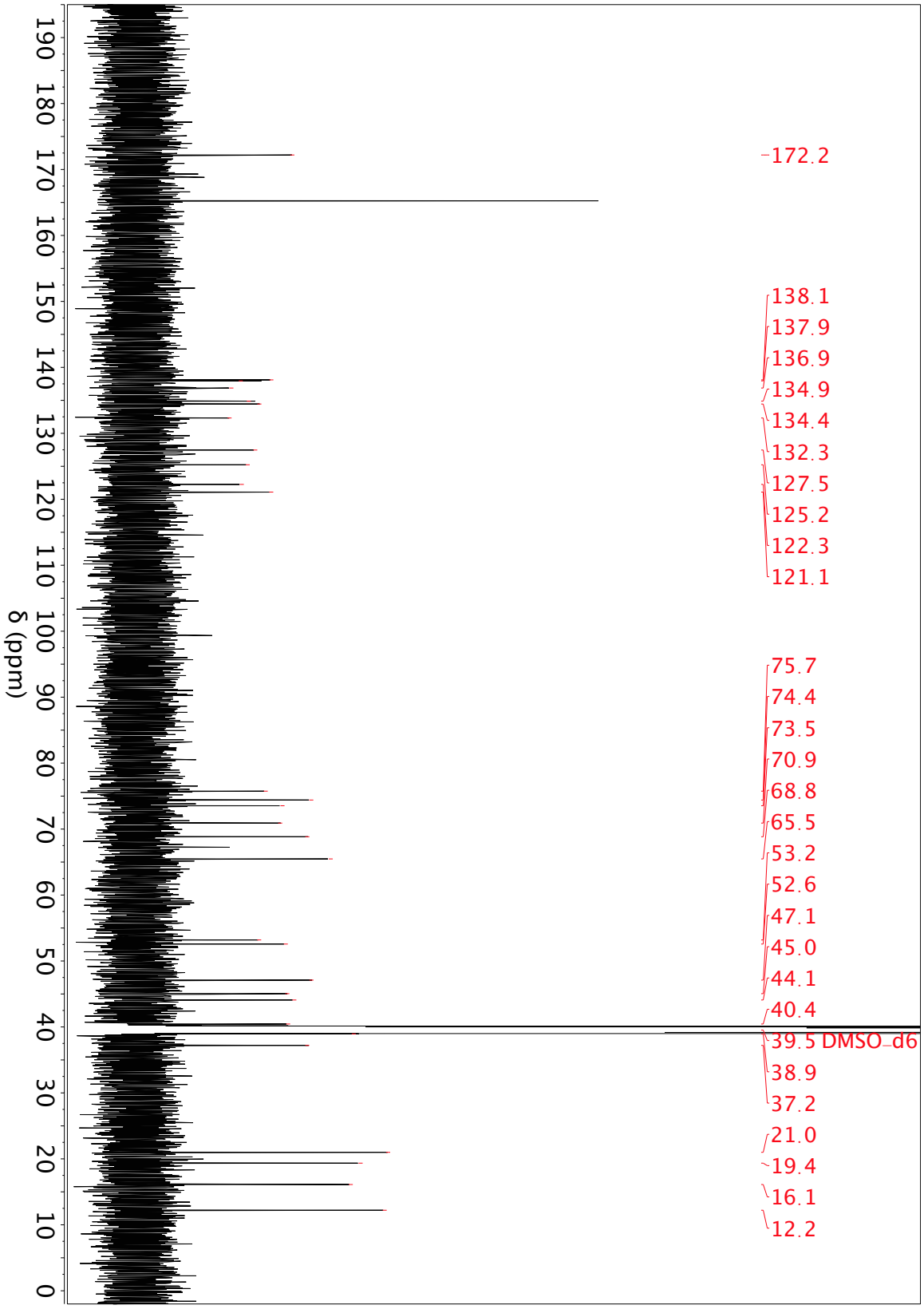
(a) Extracted ion chromatograms for **17** (top) and *Micromonospora* sp. extract (bottom) (b) Mass spectra of **17** (top) and in an *Micromonospora* sp. extract (bottom).



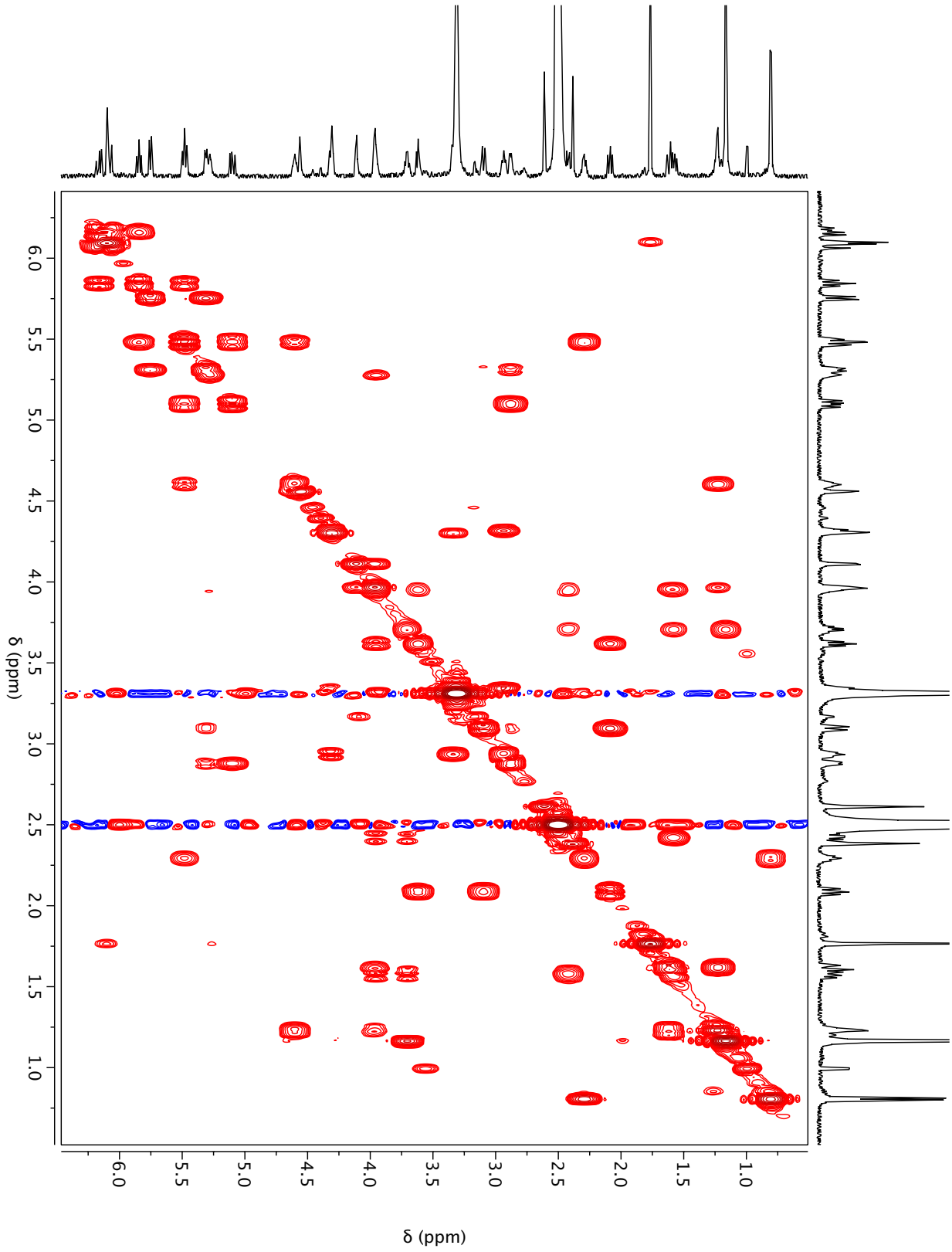
Supplementary Figure 25: ^{1}H NMR spectrum of lobosamide D (**17**) acquired in DMSO-d_6 at 600 MHz.



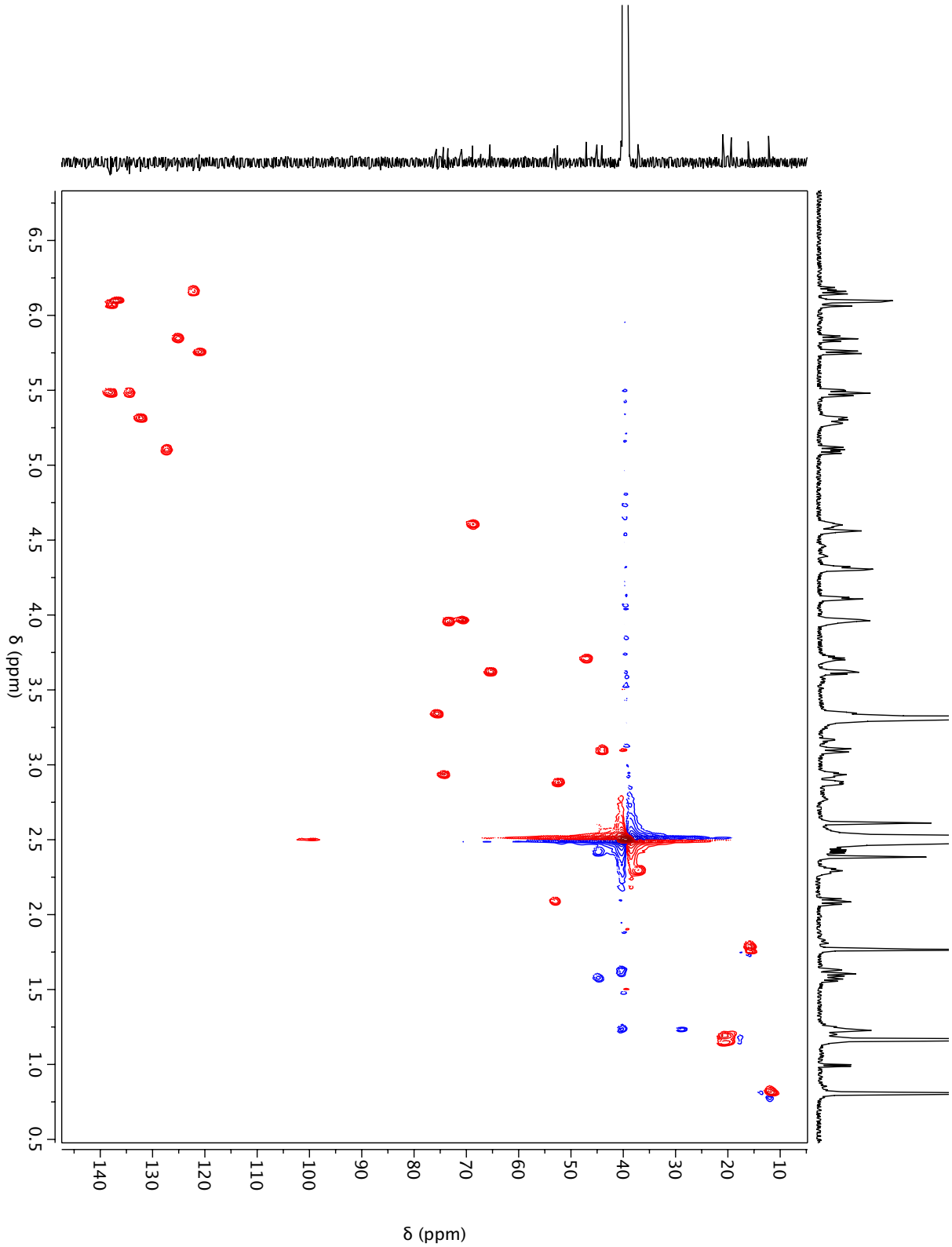
Supplementary Figure 26: Expansion of Supplementary Figure 25



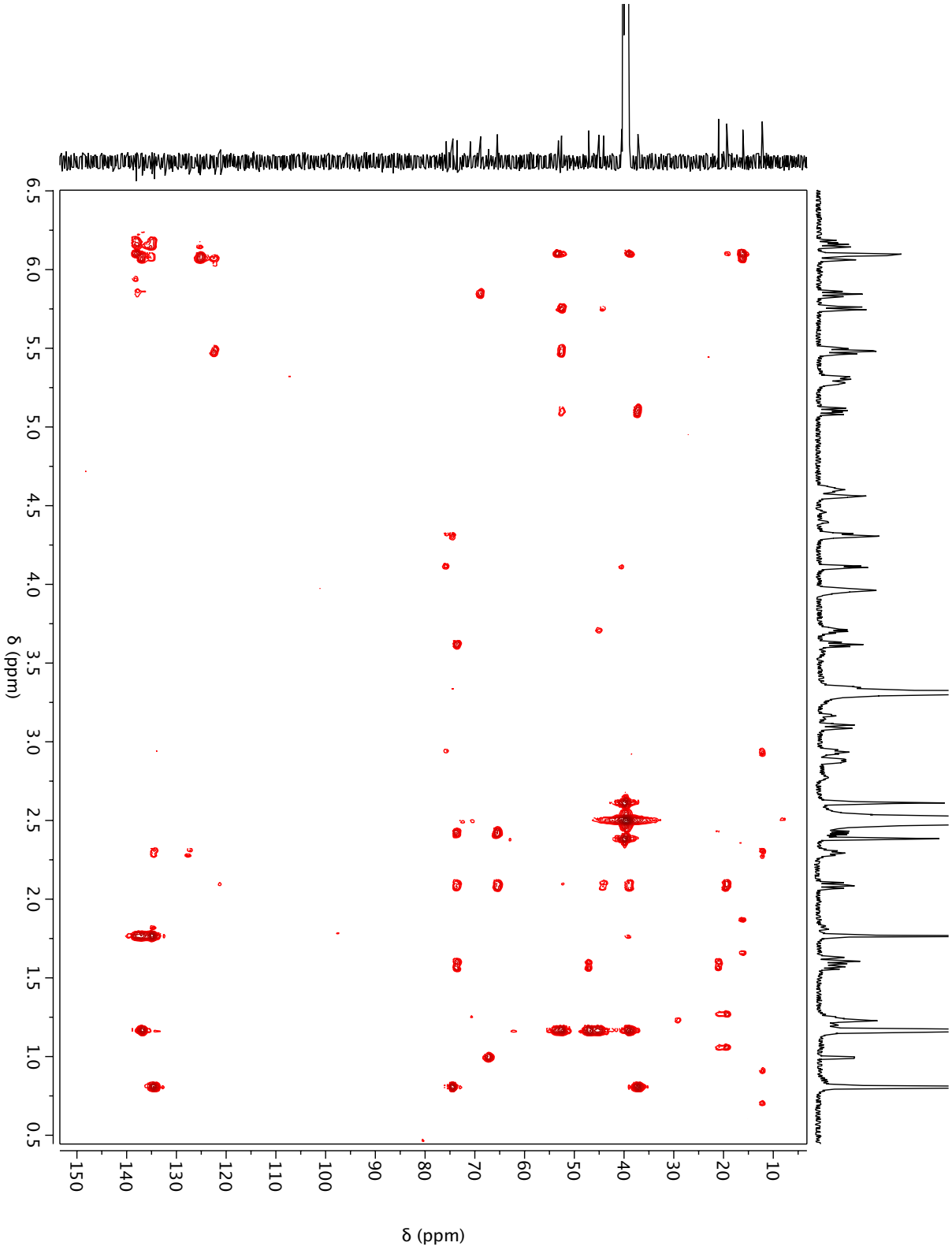
Supplementary Figure 27: ^{13}C NMR spectrum of lobosamide D (**17**) acquired in DMSO-d_6 at 150 MHz.



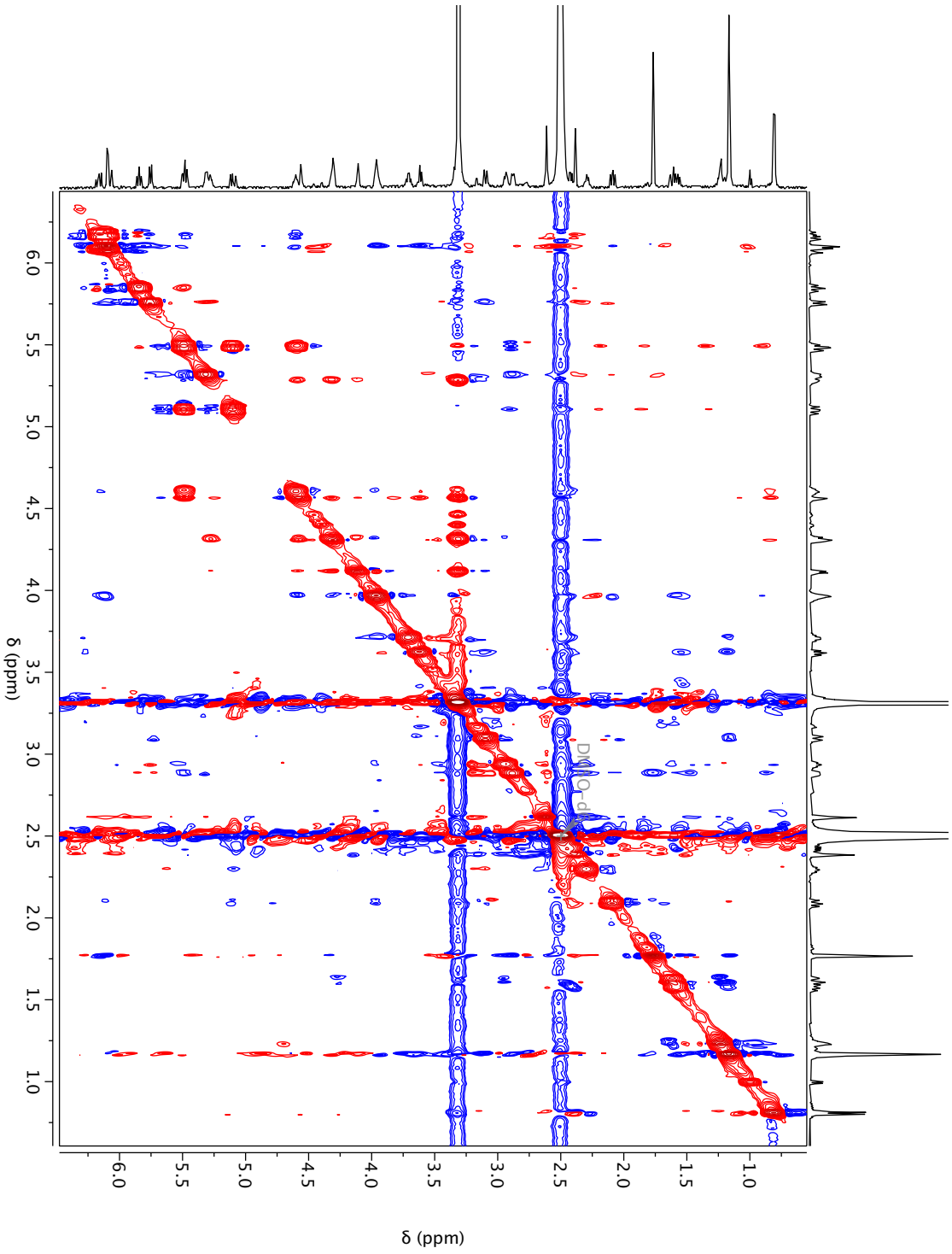
Supplementary Figure 28: COSY of lobosamide D (**17**) in DMSO-d_6 at 600 MHz



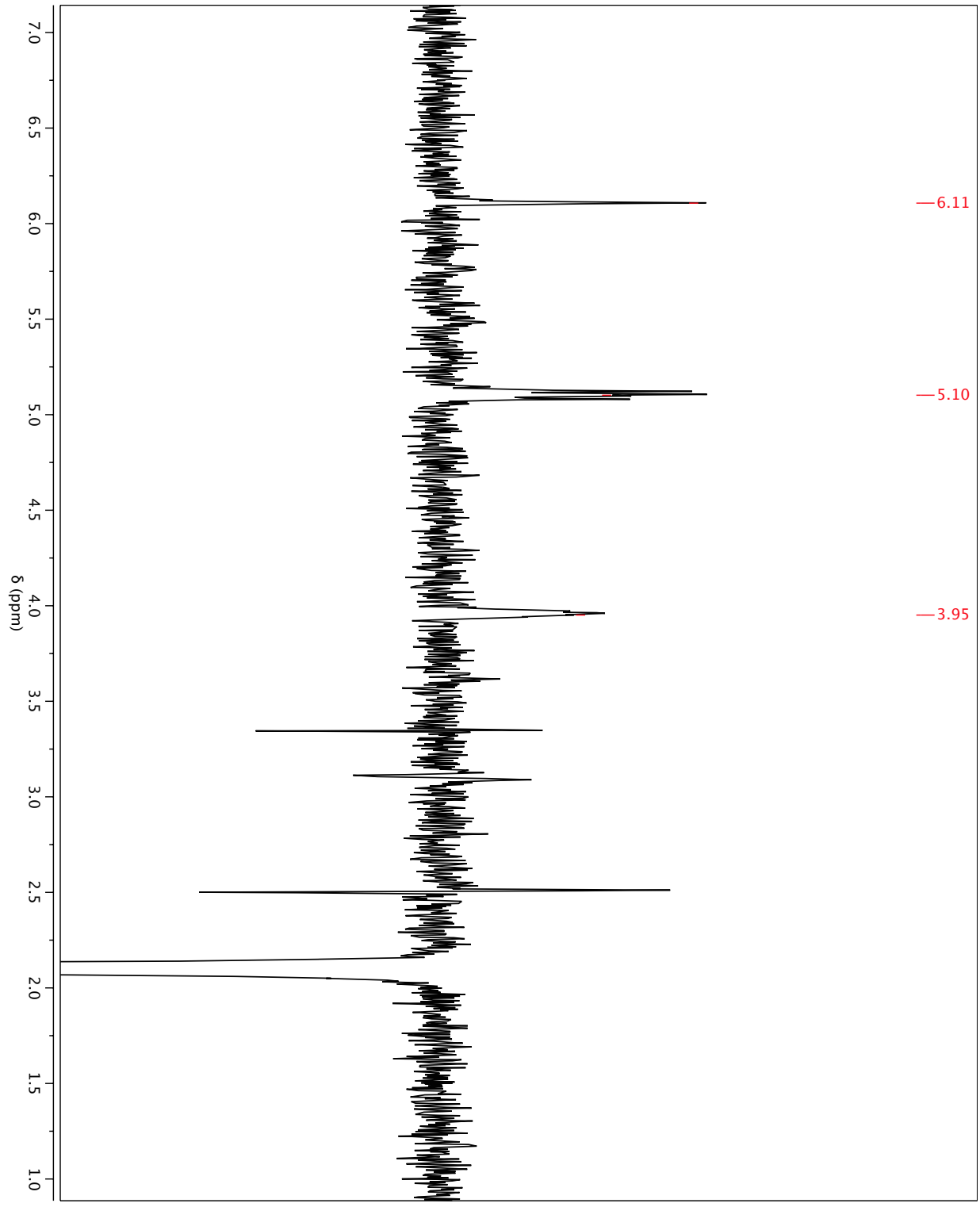
Supplementary Figure 29: HSQC of lobosamide D (**17**) in DMSO-d_6 at 600 MHz



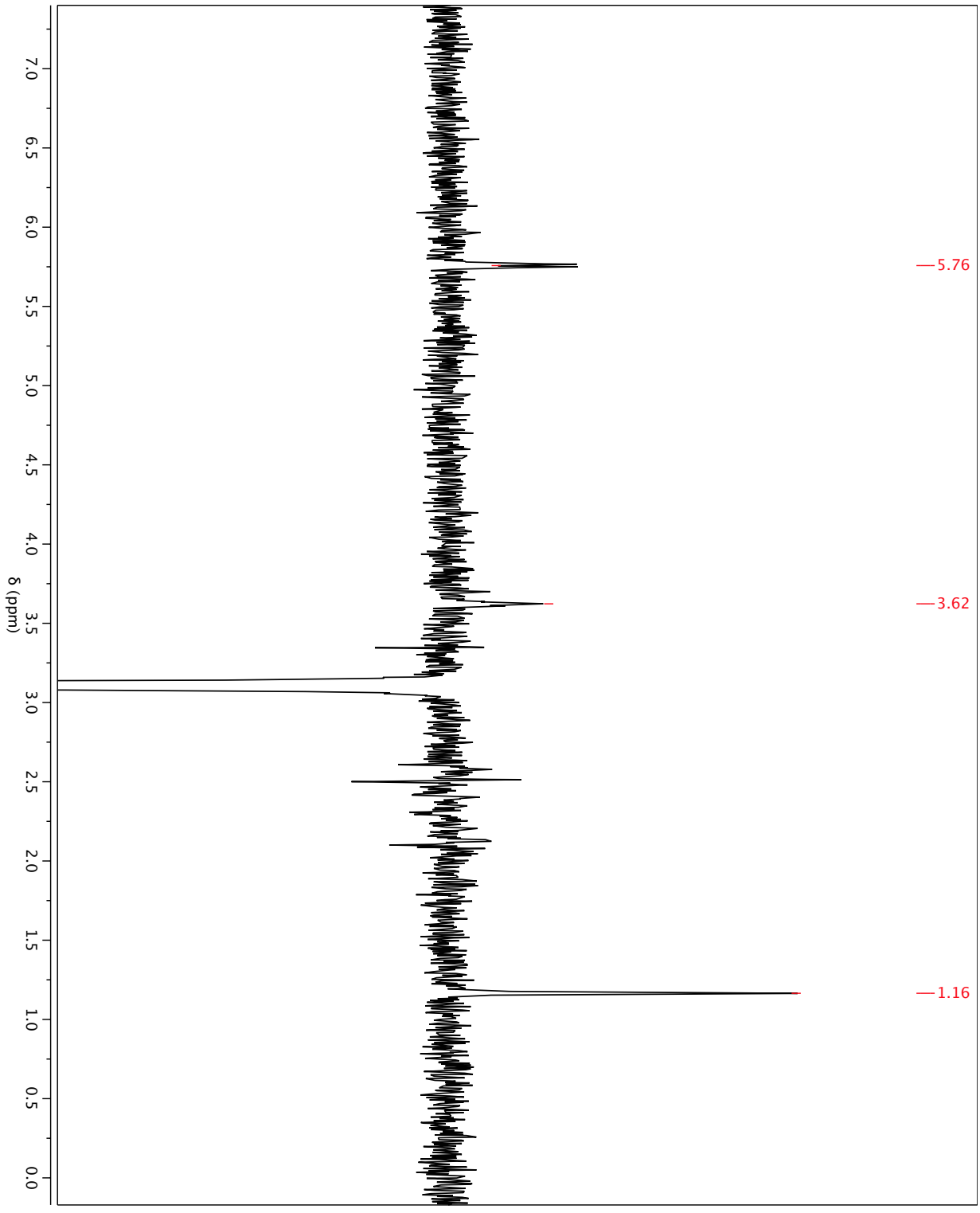
Supplementary Figure 30: HMBC of lobosamide D (**17**) in DMSO-d_6 at 600 MHz



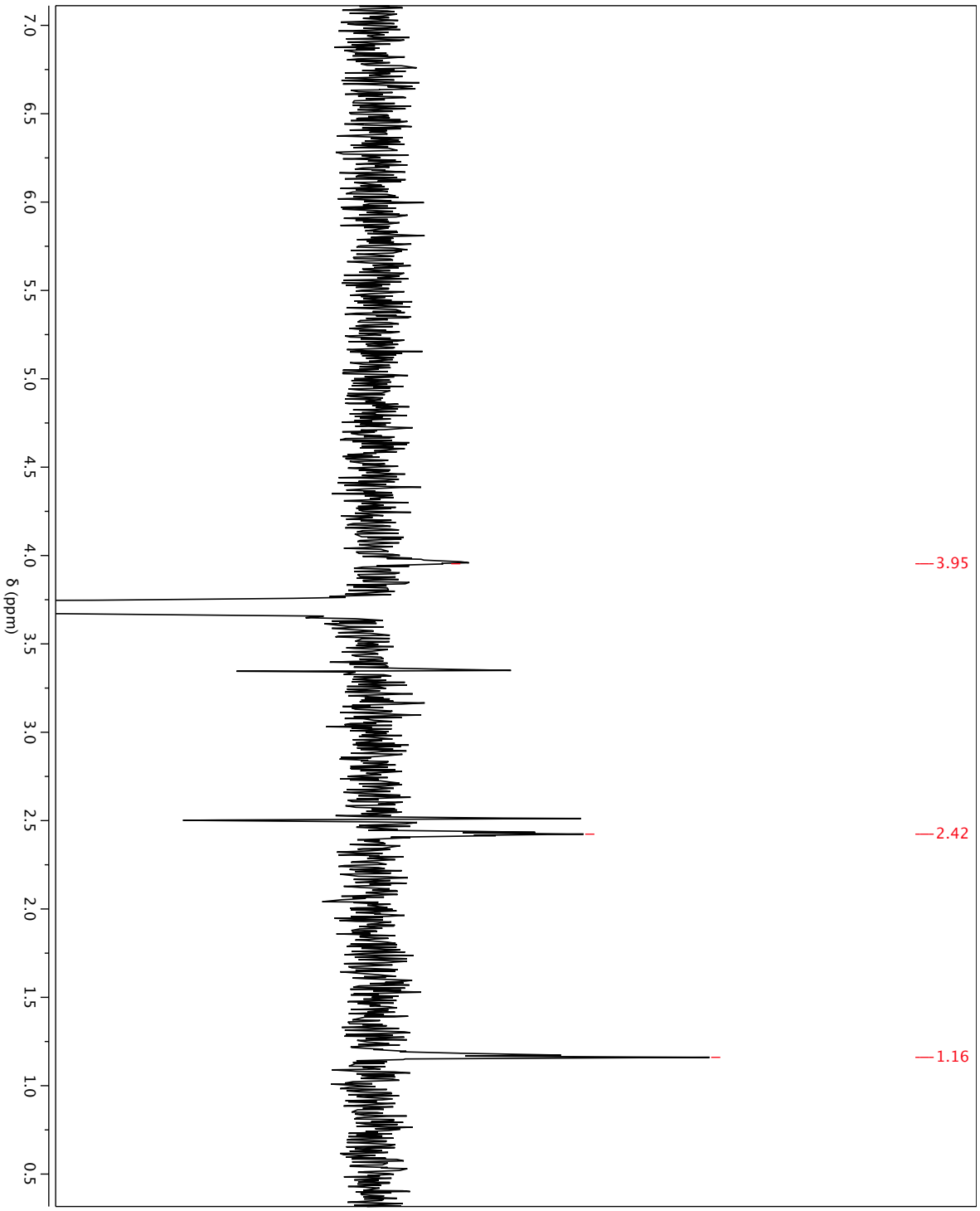
Supplementary Figure 31: ROESY of lobosamide D (**17**) acquired in DMSO-d_6 at 600 MHz



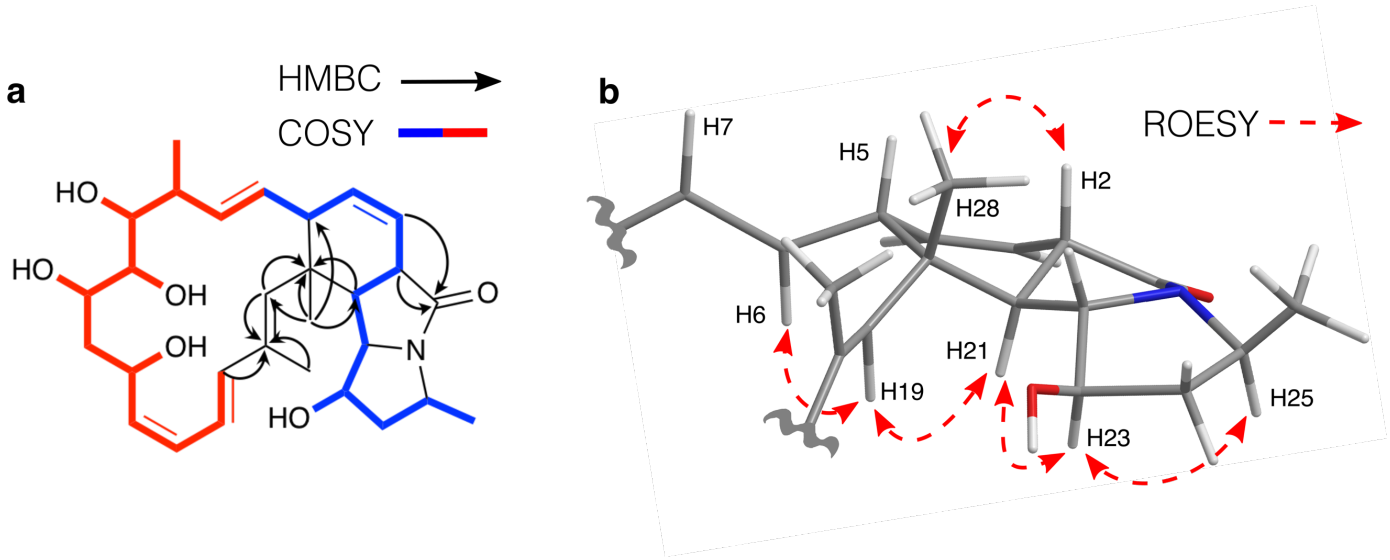
Supplementary Figure 32: Selective 1D ROESY for H21 (2.09 ppm) of **17** acquired in DMSO-d_6 at 600 MHz.



Supplementary Figure 33: Selective 1D ROESY for H2 (3.10 ppm) of **17** acquired in DMSO-d_6 at 600 MHz

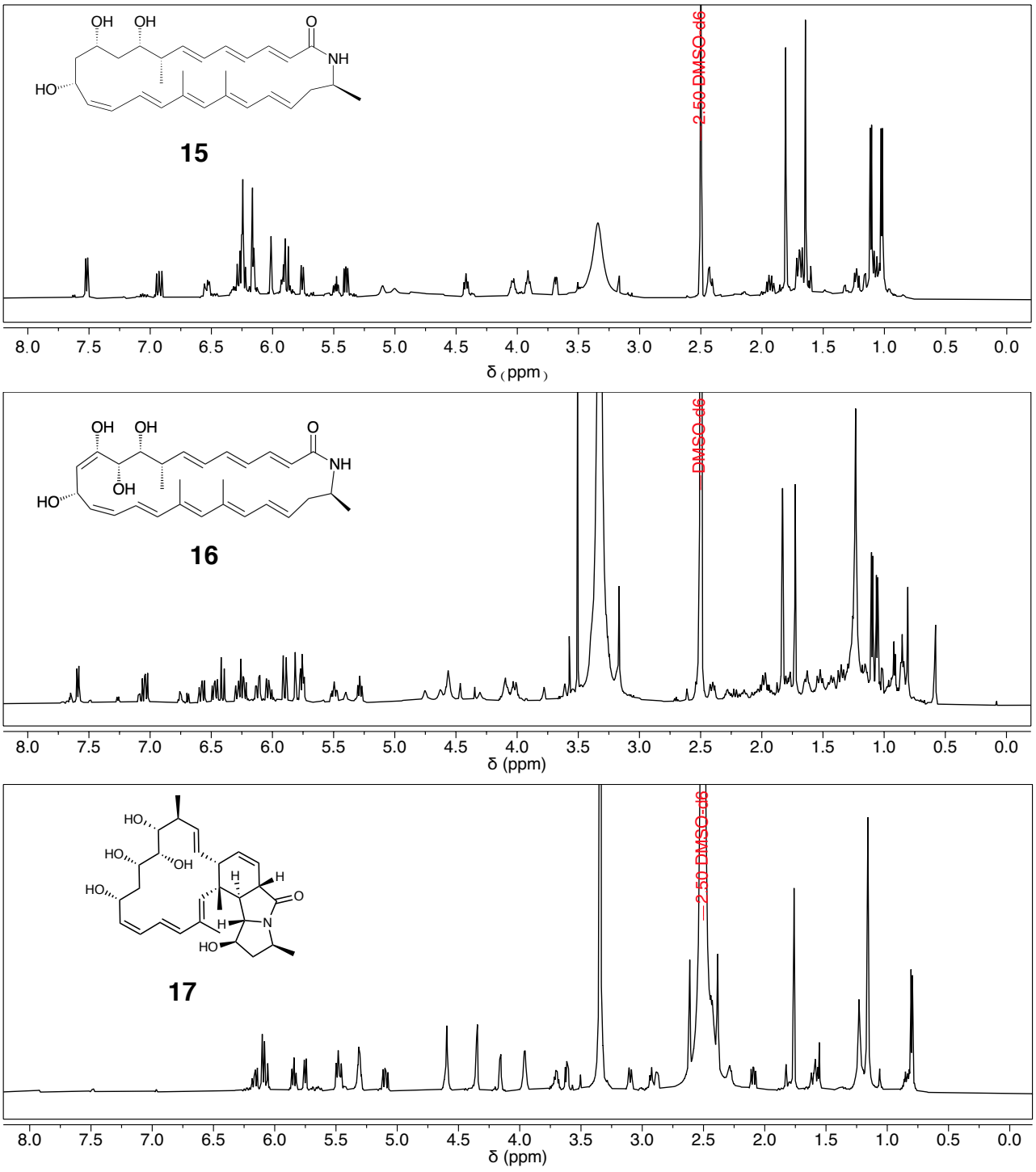


Supplementary Figure 34: Selective 1D ROESY for H25 (3.71 ppm) of **17** acquired in DMSO-d_6 at 600 MHz

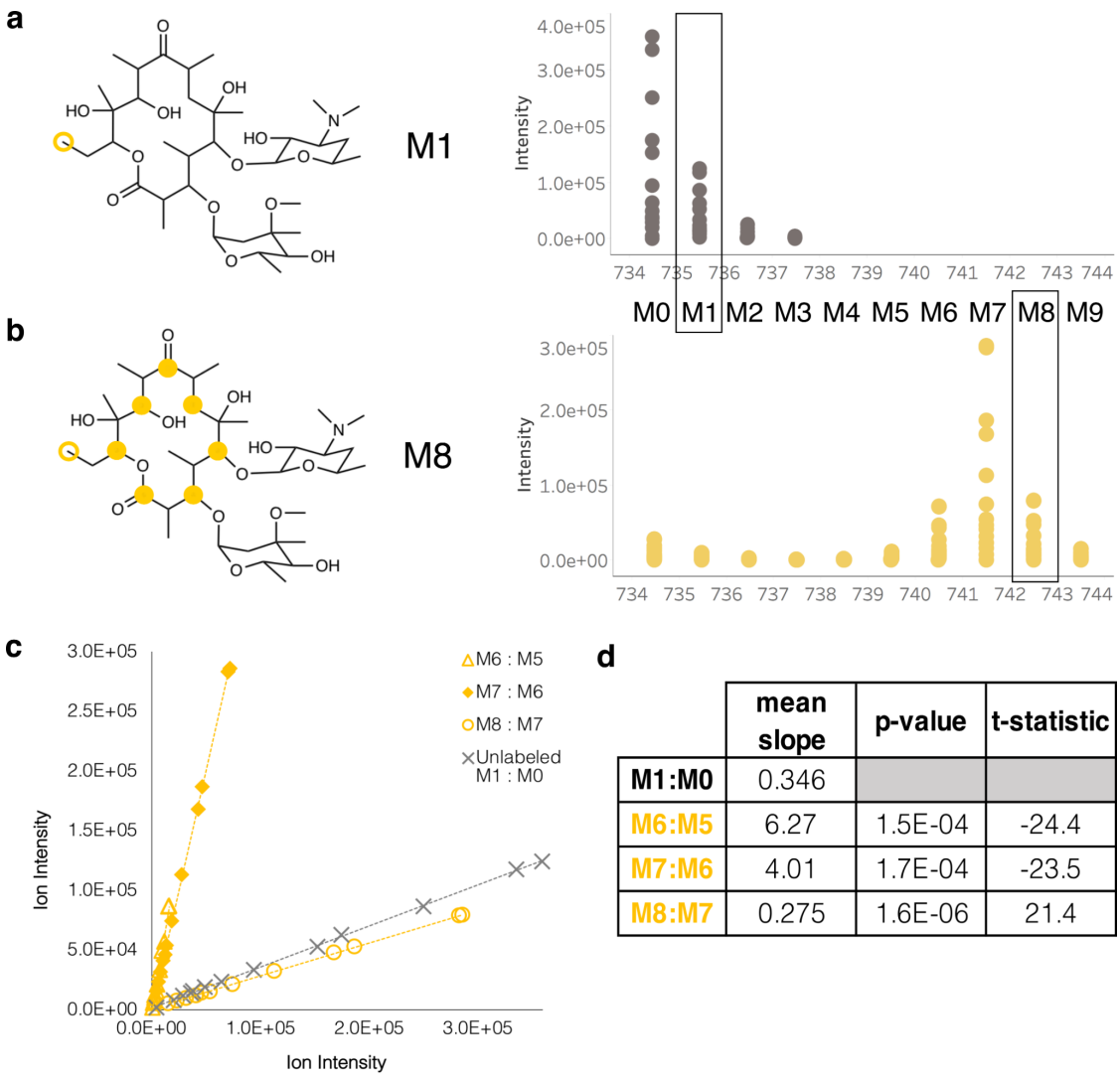


Supplementary Figure 35: Lobosamide D (**17**) Structure Elucidation Diagrams

(a) COSY correlations indicated by bolded bonds in either blue or red, and key HMBC correlations used in structure determination on **17**. (b) 3-D depiction of **17** showing key ROESY correlations used in complete configurational assignment.



Supplementary Figure 36: Comparison of Lobosamides C (top), A (middle), and D (bottom) 1H NMR spectra acquired in DMSO-d_6 at 600 MHz



	mean slope	p-value	t-statistic
M1:M0	0.346		
M6:M5	6.27	1.5E-04	-24.4
M7:M6	4.01	1.7E-04	-23.5
M8:M7	0.275	1.6E-06	21.4

Supplementary Figure 37: Isotopologue ratios for erythromycin A labeled by [1-^{13}C]propionate

(a) Structure of the M1 isotopologue of erythromycin A in the unlabeled sample. The open yellow circle represents ^{13}C derived from the natural abundance of ^{13}C (placed in an arbitrary position on the structure). (b) Structure of the M8 isotopologue of erythromycin A. Filled yellow circles represent ^{13}C derived from [1-^{13}C]propionate and the open yellow circle represents ^{13}C from the natural abundance of ^{13}C. (c) Isotopologue ratio plots from an unlabeled control and [1-^{13}C]propionate labeled sample. (d) Table showing the mean slope and Welch's two-tailed t-test results for the three heaviest isotopologue ratios of erythromycin A labeled by [1-^{13}C]propionate.

Supplementary Table 1: Substrate Labeling Table

Substrate	ACE	PROP	MET	GLU
Glu	2			1
Gln	2			2
Arg	2			4
Pro	2			1
Orn	2			2
Asp	1			1
Met	1		1	1
Thr	1			1
Ile	1			1
Asn	1			2
Lys	1			2
Ala				1
Leu				1
Val				1
Phe				1
Tyr				1
Trp				2
Ser				1
Gly				1
Cys				1
His				3
Unknown Amino Acid				1+
Acetyl or Malonyl-CoA	1			
Propionyl-CoA		1		
Methylmalonyl-CoA	1	1		
Methoxymalonate			1	
Hydroxymalonate				
Methyl			1	
IPP[a]	2			
DMAPP[a]	2			
GPP[a]	4			
FPP[a]	6			
Succinyl-CoA	1	1		
Fatty Acid	1+			
Amino-saccharide				1
Amino group				1

[a] mevalonate pathway only

Supplementary Table 2: Biosynthetic gene cluster labeling prediction for *Micromonospora* sp.

Cluster	Type	A	P	M	G
8	lantipeptide	7			27
28	lantipeptide	8			19
11b	lasso peptide	7			20
1b	RiPP	16			44
5	RiPP				1+
17	RiPP-like				
29	RiPP-like				
6	**siderophore**	**6**	**2**		**6**
12a	NRPS	1+			3
13	NRPS	4			4
16	NRPS	7+			13
1a	NRPS				1+
25	NRPS				1+
26	NRPS				2+
19	NRPS-like	1+			
4	PKS-NRPS	2		1	4
30a	PKS-NRPS	10	5		1
11a	PKS-NRPS	1+			
10	type I PKS	14+			1
9	type II PKS	12+			
20	type III PKS	15		1	1
2	PKS	4	4	4	
3	PKS	6+	3+	1	
12b	PKS	2	1		1
14	PKS	9	1		
30c	**PKS**	**11**	**3**		**1**
22	PKS-like				
7 & 21	terpene	16			
18	terpene	6+			
24	terpene	2+			
30b	terpene	2+			
23	indole	3			1
15	phenazine			1	3
27	NAGGN	5			5

[†]BGCs highlighted in light gray and dark gray meet one and two criteria respectively as indicated in text

Supplementary Table 3: NMR signals for lobosamide D (**17**)

Position	δ H (ppm)	m	*J*(Hz)	δ C
1				172.2
2	3.10	d	12.3	44.1
3	5.75	d	9.8	121.1
4	5.32	m		132.3
5	2.88	dd	8.4,3.1	52.6
6	5.10	dd	9.8,14.6	127.5
7	5.47	m		134.4
8	2.29	m		37.2
9	2.92	dd	8.0,8.0	74.4
10	3.35			75.7
11	3.97	m		70.9
12a	1.23	m		40.4
12b	1.62	m		
13	4.61	m		68.8
14	5.48	m		138.1
15	5.84	dd	10.1,10.1	125.2
16	6.16	dd	10.1,15.7	122.3
17	6.07	d	15.7	137.9
18				134.9
19	6.10	s		136.9
20				38.9
21	2.09	dd	9.3, 13.0	53.2
22	3.62	dd	6.6, 9.3	65.5
23	3.96	m		73.5
24a	1.58	m		45.0
24b	2.42	m		
25	3.71	m		47.1
26	0.80	d	6.7	12.2
27	1.76	s		16.1
28	1.17	s		19.4
29	1.16	d		21.0
OH(9)	4.35	m		
OH(10)	4.34	m		
OH(11)	4.16	d	5.4	
OH(13)	4.60	m		
OH(23)	5.30	m		

Supplementary Table 4: Power analysis for [1-^{13}C]acetate incorporation in erythromycin A

	Unlabeled	[1-^{13}C]acetate					
	M1:M0	**M1:M0**	**M2:M1**	**M3:M2**	**M4:M3**	**M5:M4**	**M6:M5**
Mean slope	0.356	2.07	1.03	0.664	0.446	0.323	0.268
RSD (%)	2.81	2.25	3.71	1.68	4.06	7.57	4.65
p-value		1.6E-4	6.5E-4	2.2E-6	5.0E-3	0.14	6.9E-4
t-statistic		-62.7	-29.9	-37.8	-7.76	2.18	10.2
N	4	3	3	3	3	3	3
SD	0.010	0.047	0.038	0.011	0.018	0.024	0.012
Pooled SD		0.030	0.025	0.010	0.014	0.017	0.011
Effect Size		56.3	26.6	29.3	6.49	-1.89	-8.03
Power		1.0	1.0	1.0	1.0	0.52	1.0

Supplementary Table 5: Cost comparison of SIL tracers in different size cultures

	2 mL	**7 mL**	**60 mL**	**1 L**
[1-^{13}C]acetate (30 mM)	0.22	0.78	6.73	112.09
[1-^{13}C]propionate (30 mM)	0.87	3.06	26.20	436.73
[methyl-^{13}C]methionine (5 mM)	0.34	1.19	10.18	169.73
[1-^{15}N]glutamate (20 mM)	1.37	4.81	41.24	687.28

Supplementary Note 1: Description of Statistical Analysis

Untargeted metabolomics software for working with SIL samples such as X^{13}CMS,[1] geoRge,[2] mzMatch-ISO,[3] and NTFD[4] all use statistical tests to compare the relative isotopologue intensity between unlabeled and labeled samples. The heavy isotopologue peaks are expected to be more intense in the SIL condition than in the unlabeled condition regardless of the overall isotopologue distribution. These methods use standard peak detection approaches, such as XCMS,[5] to detect isotopologue peaks and compare the relative isotopologue peak intensities between the labeled and unlabeled conditions. Most of these programs also offer some kind of visualization of the results which show the relative isotopologue intensities of compounds which are determined to be labeled and these isotopologue distributions can then be visually compared for biological interpretation. However, these methods do not use any statistical test to determine if the heaviest isotopologue peak detected is actually the heaviest isotopologue peak with SIL enrichment in a given group of isotopologues. Our method provides a different output which does not indicate the entire isotopologue distribution, but rather the heaviest isotopologue peak which has reliably detectable enrichment from the SIL tracer.

Although these approaches would certainly identify SIL incorporation in the same *m/z* features that are detected by IsoAnalyst, the results from these analyses do not answer our core question. With the aim of comparing the number of SIL tracers incorporated into a particular compound with the predicted number of substrates from the BGC, only the integer value of the heaviest enriched isotopologue is necessary. This information is not available using other approaches. By identifying the heaviest isotopologue peak with SIL enrichment, we estimate the integer number of SIL precursors incorporated into the structure of a given compound. Heavier isotopologues which are not statistically significant in our method represent isotopologue peaks which have additional contribution from naturally occurring ^{13}C. The true isotopologue distribution of an ion with SIL incorporation is based on overlapping binomial distributions, representing the incorporation derived from the SIL tracer and the incorporation derived from naturally occurring ^{13}C, which have different probabilities of occurring. These binomial distributions rely on the number of carbon positions present in the structure and the probability of ^{13}C incorporation.

The purpose of relying on the M1:M0 ratio from the unlabeled feature as the null hypothesis in our statistical analysis is to ask the question of whether the heavy isotopologue ratio in the labeled condition is larger, the same, or smaller than the natural ^{13}C ratio of the same molecule where no positions are enriched. Supplementary Figure 37 compares the isotopologue distributions of erythromycin A in the unlabeled control (Supplementary Fig. 37a) and when labeled by [1-^{13}C]propionate (Supplementary Fig. 37b). In Supplementary Figure 37a, the M1 isotopologue peak represents erythromycin A with a single ^{13}C derived from naturally occurring ^{13}C (open circle). Similarly, in Supplementary Figure 37b, the M8 isotopologue peak represents an erythromycin molecule with seven ^{13}C positions derived from [1-^{13}C]propionate (filled circles) and one position derived from naturally occurring ^{13}C (open circle). Because there are less ^{13}C positions remaining in the M8 isotopologue which can have naturally occurring ^{13}C compared to the M1 isotopologue, the M8:M7 ratio in the SIL condition is expected to be statistically smaller than the M1:M0 ratio in the unlabeled condition. These ratios are plotted in Supplementary Figure 37c and the mean slope and t-test results are shown in Supplementary Figure 37d.

Supplementary Note 2: Protocol for antiSMASH-based annotation of chemical substrates associated with BGCs

1. Run the latest version of antiSMASH on the complete genome sequence (either offline or through the web server) and open the HTML output. Make sure to run it including the KnownClusterBlast options to retrieve matches to validated and known gene clusters. Now, proceed with the following steps for each predicted BGC product separately, one after the other, until the last predicted BGC product.
2. To identify (homologues of) BGCs producing known metabolites, examine the KnownClusterBlast output.
    a. For BGCs for which (partial) similarity is detected to known gene clusters from MIBiG with >50% similarity at the gene content level, examine the gene cluster alignments. For other BGCs, proceed to step 3.
    b. If considerable alignment is found with an MIBiG cluster, examine the underlying literature to identify whether all enzyme-coding genes required for the production of the known molecule have homologues within the query BGC.
    c. If all genes are conserved with high sequence identity (typically >50-70% amino acid identity), it can be assumed that the BGC will encode the production of the same or a highly similar molecule. In this case, study the literature and the chemical structure of the products to identify or infer known or likely substrates from the experimentally characterized pathways and enter these into the table. For type I NRPS and PKS enzymes, also check the substrate specificity predictions of adenylation / acyltransferase domains to make sure these correspond with the known structure of the product of the reference BGC; if not, check for sequence changes in the active site and adjust the substrates for these when needed. Skip steps 3 and 4 and proceed with the next BGC.
    d. If only one or a few specific enzyme-coding genes are missing from the query BGC compared to the reference BGC from MIBiG (or if additional enzyme-coding genes are present in the query compared to the reference), check the biochemical functions of these based on HMMer protein domain analysis with Pfam and BLAST searches against all MIBiG proteins (and/or Uniprot if needed), and add/subtract substrates from the set of substrates as inferred purely from the MIBiG reference pathway. Skip steps 3 and 4 and proceed with the next BGC.
    e. If only partial gene content similarity is detected, proceed with step 3.
3. There are a number of generic enzymatic biotransformations that can be directly linked to substrates:
    a. Tailoring enzymes: SAM-dependent methyltransferases, aminotransferases
    b. Glycosyltransferases
    c. Sets of enzymes for the biosynthesis of precursors such as modified sugars or non-proteinogenic amino acids, mainly annotated through Subclusterblast.

    In case such enzymes are annotated in the BGC, add a methyl group or amino group (a), sugar (b, c) or amino acid (c) to the substrate table. In some cases, the identity of the monomer can be further annotated (i.e., an amino sugar through the presence of an

aminotransferase or even a specific sugar when in Subclusterblast there is sufficient homology to a known sugar), whereas in other cases the full identity remains unknown.
Proceed with step 4.

4. The following step in the procedure depends on the antiSMASH-annotated biosynthetic class of each BGC:
    a. Type I PKS (modular):
        i. For *cis*-AT PKSs, count the number of modules. For each module, check the predicted substrate specificities for the acyltransferase domains (malonyl-CoA, methylmalonyl-CoA, methoxomalonyl-CoA) and add the corresponding substrates into the substrate table.
        ii. For *trans*-AT PKSs, check for the presence of the *trans*-acyltransferase(s) in the BGC. If only one is found, it is generally safe to assume this is a malonyl-specific enzyme, and one malonyl-CoA monomer can be added for each AT-less PKS module. If more than one is found, look at the MIBiG protein BLAST results for each of them to check for similarity to acyltransferases specific to other substrates, such as ethyl-malonyl-specific ones.
        iii. For PKS systems with just 1-3 modules, oligomerization is likely; therefore, in this case write, in line with the number of modules, '1+', '2+' or '3+' for the substrates.
        iv. For hybrid *cis*/*trans*-AT PKSs (e.g., kirromycin-like), combine the above instructions.
        v. Check for tailoring domains inside the PKS assembly lines, such as methyltransferases and aminotransferases, and add the methyl or amino groups to the table when needed.
    b. Type II / III PKS
        i. For type II PKSs, check if antiSMASH was able to make predictions of the starter unit and/or the number of elongation cycles. If so, add the corresponding (numbers of) substrates to the table. If not, indicate '1+' for malonyl-CoA.
        ii. For type III PKSs, check whether significant similarity to known enzymes exists by studying the MIBiG protein BLAST results (>50-60% amino acid identity). If yes, 'copy' the starter unit and number of acetate units added through elongation from the known enzyme. If no, add '1+' (minimally 1) for acetyl-CoA to the table.
        iii. For non-canonical PKSs not belonging to types I-III, add '1+' (minimally 1) for acetyl-CoA to the table.
    c. Type I NRPS
        i. Check the substrate specificities of the adenylation domains in the NRPS modules predicted by antiSMASH. Add the reliably predicted amino acid substrates to the substrate list.
        ii. In case no reliable amino acid specificity could be predicted by antiSMASH, add an "unknown amino acid" to the substrate list.
        iii. Multiply the numbers of substrates for BGCs homologous to BGCs whose products are known to be oligomerized (e.g., in the case of enterobactin and

valinomycin). For NRPS systems with just 1-3 modules, oligomerization is likely; therefore, in this case, in line with the number of modules, write '1+', '2+' or '3+' for the substrates.

iv. Check for tailoring domains inside the PKS assembly lines, such as methyltransferases and aminotransferases, and add the methyl or amino groups to the table when needed.

d. RiPP
   i. For those RiPP classes for which cleavage sites are predicted by antiSMASH (i.e., lanthipeptides), start counting the predicted amino acids in the precursor peptide sequence there.
   ii. Add one amino acid substrate for each predicted amino acid. Note that some can occur multiple times.

e. Terpene
   i. Enzyme type. If an enzyme class can reliably be established based on Pfam domains and/or similarity observed in the MIBiG protein BLAST results, fill out the following substrates for these enzyme classes:
      1. Monoterpene synthases: 1 for GPP (or 1 IPP + 1 DMAPP)
      2. Sesquiterpene synthases: 1 for FPP (or 2 IPP + 1 DMAPP)
      3. Diterpene synthases: 1 for GGPP (or 3 IPP + 1 DMAPP)
      4. Sesterterpene synthases: 4 IPP + 1 DMAPP
      5. Triterpene synthases (oxidosqualene cyclases): 2 for FPP (or 2 DMAPP + 4 IPP)
      6. Lycopene cyclases / phytoene synthases: 7 IPP + 1 DMAPP
   ii. If no enzyme class can be reliably established, fill out '1+' (minimally 1) for IPP.

f. Others
   i. It is typically difficult to reliably assign substrates to this type of BGCs, unless significant similarity to MIBiG BGCs is found for at least part of the biosynthetic gene cluster. Continue with the next one.

Continue with the next BGC product. When all BGCs are processed, continue with step 5.

5. The antiSMASH BGC-substrate specificity table is finished. Proceed with the IsoAnalyst workflow.

Supplementary Note 3: Structure Elucidation of Lobosamide D

Planar Structure Elucidation of Lobosamide D

The molecular formula $C_{29}H_{41}NO_6$ was calculated from the $[M+H]^+$ ion detected at *m/z* 500.3012 (calcd. 500.3007, Δppm = 1.0), indicating 10 degrees of unsaturation. Review of the ^1H and phase-sensitive HSQC spectra revealed 4 x CH_3, 2 x CH_2, and 20 x CH including 9 olefinic signals. Further evaluation of the ^{13}C and HMBC spectra identified a further 3 x qC signals, including one olefinic carbon, and one amide carbonyl. These components accounted for all carbons, 36 of 41 protons, one oxygen and one nitrogen from the molecular formula. The remaining five oxygens and five protons were assigned as hydroxy groups, based on the presence of five broad exchangeable signals in the proton spectrum and five signals in the HSQC (^1H 3.35 - 4.61 ppm, ^{13}C 65.5 - 75.7 ppm) consistent with oxygenated methines. This completed the detection of all the elements in the molecular formula in the spectral data.

To solve the planar structure the COSY spectrum was used to identify one large spin system that incorporated all but two of the protonated carbons in the molecule. Starting from an olefinic doublet at 6.07 ppm (H17) sequential COSY correlations to 6.16 (H16), 5.84 (H15) and 5.48 (H14) ppm identified a diene motif. COSY signals from 5.48 (H14) sequentially to 4.61 (H13), diastereotopic methylene protons at 1.23 and 1.62 (H12), and oxygenated methine protons at 3.97 (H11), 3.35 (H10), and 2.92 (H9) indicated a polyhydroxylated region which in turn was connected to another olefin through COSY correlations sequentially from 2.29 (H8) to 5.47 (H7) and from 5.47 (H7) to 5.10 (H6). This completed the linear portion of the spin system highlighted in red in Supplementary Figure 35a.

The next section of the structure elucidation was complicated by the presence of a large number of aliphatic methine signals, suggestive of a complex fused ring system. From the last olefinic proton (H6) a series of sequential COSY signals connected protons at 2.88 (H5), 5.32 (H4), 5.75 (H3), 3.10 (H2), 2.09 (H21), 3.62 (H22), 3.96 (H23), a diastereotopic methylene at 1.58 and 2.42 (H24), 3.71 (H25) and a terminal methyl signal at 1.16 ppm. This completed the second component of the spin system illustrated in blue in Supplementary Figure 35a.

Assembly of the tetracyclic ring system in lobosamide D required extensive use of HMBC data. This was complicated by severe signal overlap for two of the methyl signals (H28 and H29). Key HMBC signals from H2 and H3 to the amide carbonyl carbon (C1) at 172.2 placed the carbonyl at C2. This left a total of five carbons; 1 x aliphatic CH_3, 1 x vinylic CH_3, 1 x olefinic CH, 2 x qC. This was suggestive of a trisubstituted olefin, which was supported by the presence of a methyl singlet at ^1H 1.76 ^{13}C 16.1 ppm (H27) which showed HMBC correlations to the olefinic methine (C19) and a quaternary carbon at 134.9 (C18). In addition, this vinyl methyl signal (H27) showed a strong HMBC correlation to C17, placing it at the terminus of the major spin system. The olefinic methine signal (H19) showed reciprocal HMBC correlations to C17, C18 and C27, confirming this assignment. H19 possessed two additional HMBC correlations, one to the remaining quaternary carbon at 38.9 ppm (C20) and the other to the remaining methyl carbon at 19.4 ppm (C28). Closer examination of the quaternary carbon at 38.9 (C20) revealed an HMBC correlation from H21, closing the macrocyclic ring. Finally, a strong HMBC correlation from methyl H28 to C20 placed the remaining methyl group on carbon 20. Additional HMBC correlations from this methyl group to C5 and C21 confirmed the presence of the 6, 5 fused ring system, and completed all of the carbon-carbon bond connections in the molecule.

The five hydroxyl groups were located on carbons 9, 10, 11, 13 and 23 based on COSY correlations from each broad OH proton to its associated methine proton (Supplementary Figure 28). Finally, the remaining nitrogen atom on the amide functional group was connected to the two remaining open positions in the molecule (C22 and C25). This assignment was supported by the chemical shifts for

these two carbons (65.5 and 47.1 respectively) and completed the planar structure assignment, accounting for all of the degrees of unsaturation.

Configurational Analysis of Lobosamide D

Lobosamide D was isolated from the same strain, *Micromonospora* sp. RL09-050-HVF-A, which was previously published by our laboratory as the producing organism of lobosamides A-C. To determine the complete absolute configurations of lobosamides A-C we previously obtained a full genome sequence for this strain and identified the *lob* PKS biosynthetic gene cluster. This sequence data, in conjunction with extensive dipolar coupling NMR experiments, defined the absolute configuration at every position in the molecule.

Given the common polyketide core precursor between lobosamides A-C and lobosamide D and the absence of any other relevant polyketide BGCs in the genome of the producing organism we hypothesize that this same BGC also produces Lobosamide D. The absolute configurations of positions 9, 11, and 13 of lobosamides A-C were previously determined by genetic analyses of three ketoreductase (KR) domains responsible for these hydroxyls. The relative configurations of the methyl at position 8 and the hydroxyl at position 10 were determined experimentally and the configurations were fully assigned through comparison to the hydroxyl stereocenters (9,11, and 13) produced by KR domains. The stereocenter at position 25 in all of the lobosamides and related compounds including salinilactam and micromonolactam, is derived from a 3-aminobutyrate starter unit. The enzymatic mechanism that produces this starter unit was characterized for incednine and demonstrated stereospecificity in producing (*S*)-3-aminobutyrate. This stereospecificity helped assign positions 25*S* for lobosamides A-C, and this conserved stereospecificity for compounds containing the (S)-3-aminobutyrate starter unit has been shown experimentally in other related natural products (mirilactam, micromonolactam, etc). The same absolute configuration is assigned to position 25 in lobosamide D. Based on these previously established data, we determined the absolute configuration of the following positions of lobosamide D to be 8*S*, 9*R*, 10*R*, 11*S*, 13*R*, 25*S*.

The double bond configurations in lobosamide D are also expected to match the configurations in the lobosamides. Double bond configurations were consistent in lobosamides A-C except for the olefin at C14-C15, which is 14E in lobosamide B and 14Z in lobosamides A and C. Evaluation of key coupling constants between olefinic protons in lobosamide D confirmed that it matched the configuration of lobosamide A in all positions including C14 ($^3J_{H14\text{-}H15}$ = 10.1 Hz). The full double bond configurations for lobosamide D are 3Z, 6E, 14Z, 16E, and 18E ($^3J_{H3\text{-}H4}$ = 9.8 Hz, $^3J_{H6\text{-}H7}$ = 14.6 Hz, $^3J_{H14\text{-}H15}$ = 10.1 Hz, and $^3J_{H16\text{-}H17}$ = 15.7 Hz) The double bond at position 18E was corroborated by ROESY correlations between H19/H21 and H5/H27, indicating that H19 and H27 are on opposite sides of the 6-membered ring.

The full absolute configuration of the 5-5-6 fused ring system was determined using ROESY correlations. When possible, ROESY correlations from 2D ROESY experiments were confirmed by selective 1D ROESY experiments. ROESY correlations between H21/H6, H21/H19, and H21/H23 indicated that protons H6, H19, H21, and H23 were all located on the same face of the ring system (Supplementary Figs. 31, 32, 35b). A ROESY correlation between H2/H28 and the trans relationship between H2 and H21 ($^3J_{H2\text{-}H21}$ = 13.0 Hz) indicated that H2 and H28 are located on the same face of the six membered ring (Supplementary Figs. 31, 35b).

The absolute configuration at position 25*S* was previously determined by BGC analysis which indicated the starter unit (*S*)-3-aminobutyrate. However, no correlations could be observed in the 2D ROESY spectra between H23 and H25 or H29 in order to establish the relationship between the fused ring system and the known configuration at position 25. To address this, we performed a 1D ROESY experiment selectively irradiating H25 and observed a correlation to H23 (Supplementary Figure 34). This connected the known absolute configuration at position 25 to the relative configurations in the 6,5,5

ring system determined by ROESY experiments and allowed us to assign the full absolute configuration of lobosamide D as 2*R*, 5*S*, 8*S*, 9*R*, 10*R*, 11*S*, 13*R*, 20*R*, 21*S*, 22*S*, 23*R*, 25*S*.

1. Huang, X. *et al.* X 13 CMS: Global Tracking of Isotopic Labels in Untargeted Metabolomics. *Anal. Chem.* **86**, 1632–1639 (2014).
2. Capellades, J. *et al.* GeoRge: A Computational Tool to Detect the Presence of Stable Isotope Labeling in LC/MS-Based Untargeted Metabolomics. *Anal. Chem.* **88**, 621–628 (2016).
3. Chokkathukalam, A. *et al.* mzMatch–ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics* **29**, 281–283 (2013).
4. Hiller, K., Metallo, C. M., Kelleher, J. K. & Stephanopoulos, G. Nontargeted Elucidation of Metabolic Pathways Using Stable-Isotope Tracers and Mass Spectrometry. *Anal. Chem.* **82**, 6621–6628 (2010).
5. Gowda, H. *et al.* Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* **86**, 6931–6939 (2014).