Supplemental Information: Integration of high-content phenotypic screening and untargeted metabolomics analysis for the comprehensive functional annotation of complex natural products libraries.

Kenji L. Kurita, Emerson Glassey, Roger G. Linington[†,*]

Department of Chemistry and Biochemistry, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

*Corresponding author. Email: rliningt@sfu.ca

[†]Current address: Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6

Supplemental Table of Contents:

Source and Mass Spectrometer Settings:

The mass spectrometer was run with a detector range from 100 to 1700 m/z. The ESI source was operated with a desolvation temperature of 350°C and a drying gas flow rate of 11 L/min. The fragmentor voltage was held at 135 V. In positive ESI mode, the capillary voltage was ramped from 2500 V at 0 min to 2750 V at 1 min, and to 3000 V at 3 min. In negative ESI mode, the capillary voltage was held at 2750 V. Each sample was run in both positive and negative ESI source modes and in both high resolution (4 GHz) and extended dynamic range (2 GHz) detector modes.

Data Processing and Analysis:

MassHunter Workstation Qualitative Analysis Settings:

Find by Molecular Feature: We used default settings for the find by molecular feature algorithm from MassHunter 5.0 with a minimum peak height cutoff of 700 counts.

Export Peak Lists: Peak lists are exported in XML file format and include compound, adduct, and isotope information. Unless otherwise stated, we use adduct information described as *m/z* features and adduct isotope information for isotope pattern matching.

Compound Activity Mapping:

Isotope Pattern Matching: Feature alignment between sample runs uses three different criteria to determine if two *m/z* features are in fact the same: retention time, high accuracy mass, and isotope pattern. Isotope pattern matching was done using the method described by Pluskal, *et al*. by finding the product of one minus the peak height differences within 20 ppm for each peak between two normalized isotope patterns(28). Unless specified, the tolerances used for *m/z* feature alignment are 7 ppm, 0.4 minutes, and 0.5 isotope similarity.

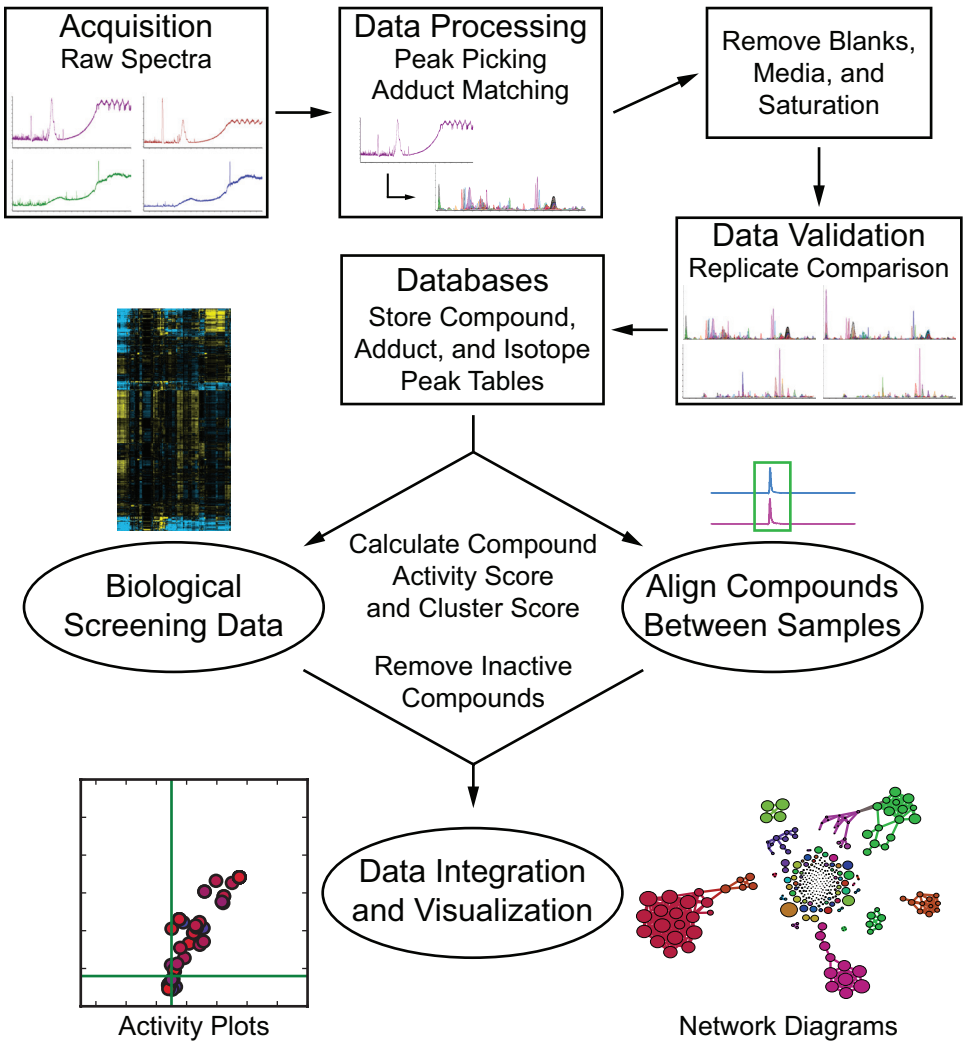
Blank Removal: In order to correct for retention time shift in HPLC experiments and to remove systematic signals, five blank runs were performed before every plate analyzed. The first two were used to clean and equilibrate the system, while the third through fifth are used for blank removal. Runs of fractionated starch/yeast/peptone (SYP) medium are also subtracted in the same step to remove systematic media contaminants. We align *m/z* features between blanks/SYP (20 ppm, 0.4 minutes, 0.5 isotope difference), condense the blank/media background runs into one *m/z* feature list, and remove all of those features from the 96-well plate data they preceed. Blanks and SYP are acquired in both high-resolution and high-dynamic range modes, and subtracted from their respective data modes.

Saturation Removal: Detector saturation causes a number of errors in data acquisition including loss of mass accuracy and signal ringing of false features around the real saturated feature. During initial data analysis, after exporting peak lists, we identify *m/z* features which are saturated, and remove all features within 0.4 minutes and 1 mass unit of the saturated feature. Because the features are saturated, we do not use isotope pattern matching, because isotope patterns are not reliable for saturated features.

The "D" prefractions: It was observed that the "D" or 80% methanol/water prefractions contained many nonspecific masses that were present in most of the "D" fractions, but absent

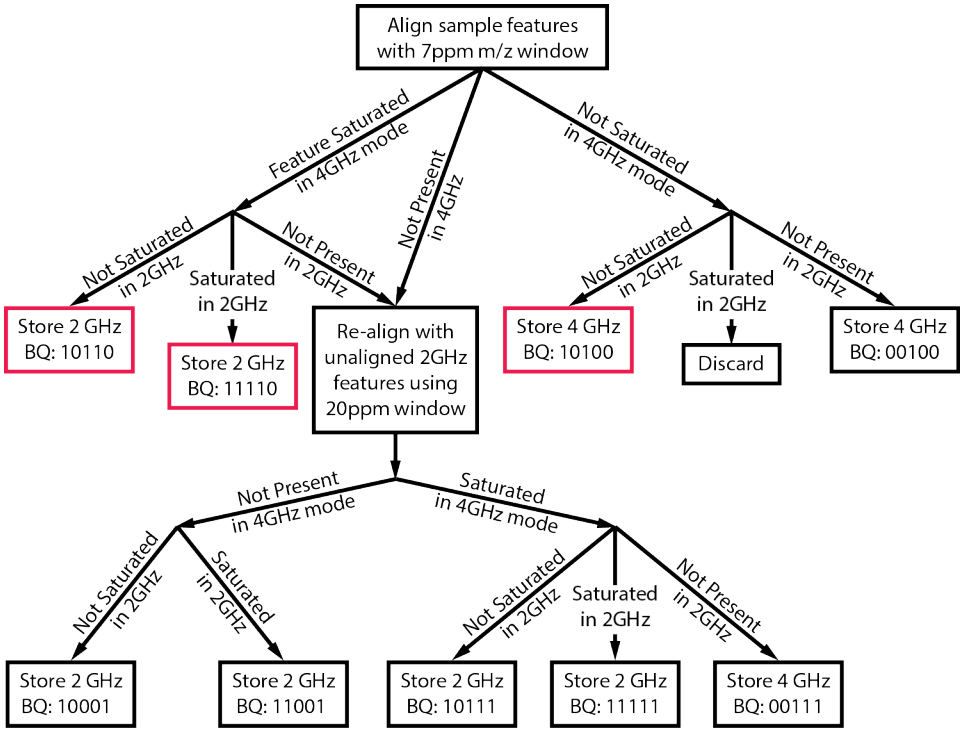
from all other fractions; therefore, any features that were only associated with "D" prefractions were omitted from network analysis in order to reduce non-specific clustering. These masses were likely due to the degradation of XAD-16 polymer resin used in the fermentation media to sequester organic molecules.

Supplemental Figure 1: Overview of the Compound Activity Mapping work flow from data acquisition in 4 GHz and 2 GHz modes in postive and negative ionization mode to data integration and vizualization.

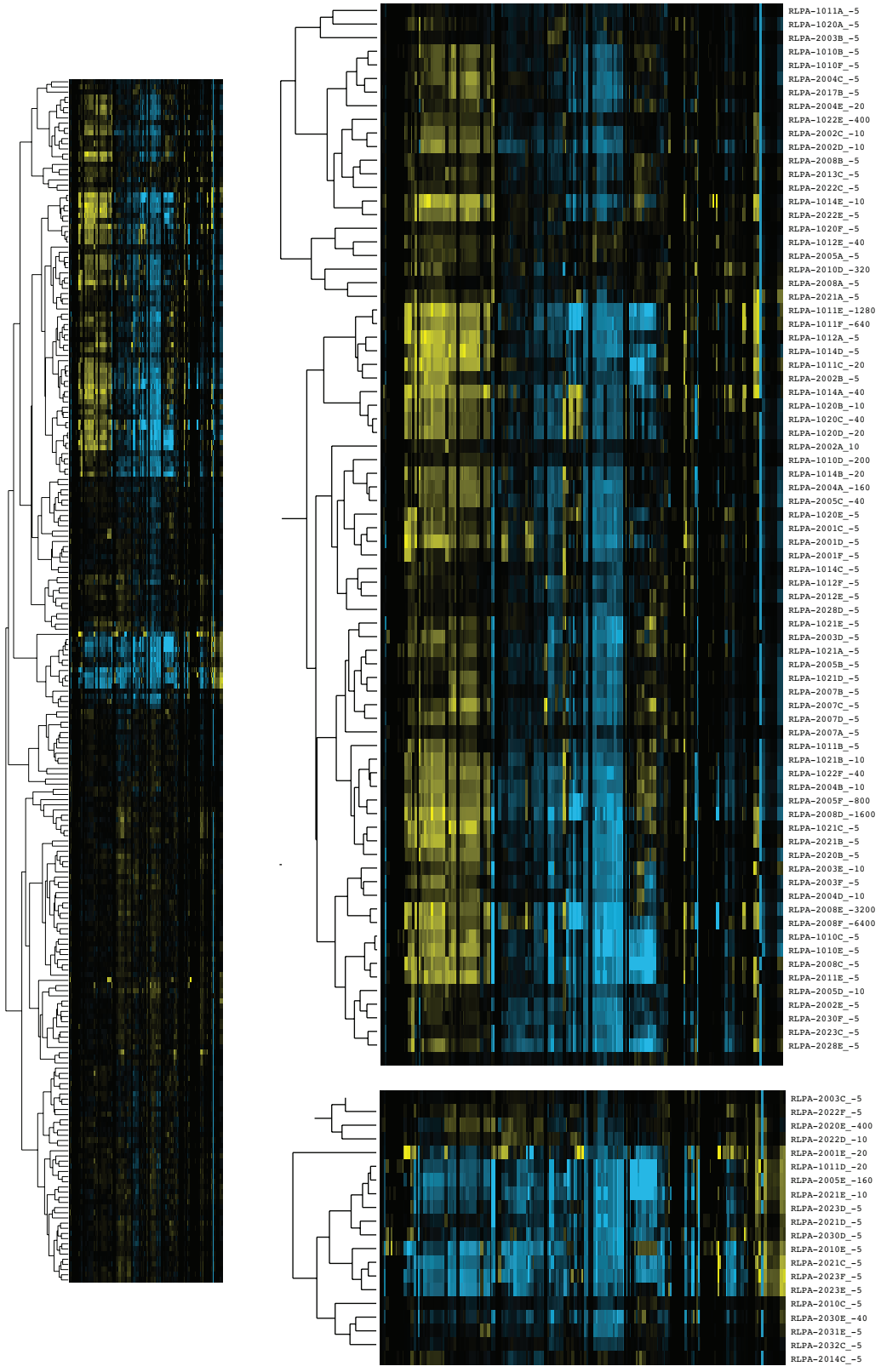


Feature Validation: In order to guarantee the fidelity of the *m/z* features used by our platform, we perform a variation of technical replicates during data acquisition. The Agilent Time of Flight mass spectrometer can acquire data in two modes: high-resolution (4 GHz) mode, which provides greater *m/z* resolution and increased sensitivity, but has low tolerance for concentration variances so the detector is more readily saturated, and high dynamic range (2 GHz), which sacrifices *m/z* resolution for greater detector range, reducing saturation. Our samples are microbial extracts with mixtures of unknown constitution and concentration, but we also require high mass accuracy. We acquire data in both modes and align features between the two modes (7 ppm or 20 ppm, 0.4 minutes, 0.5 isotope difference), storing data from the mode that is not saturated, with preference for high-resolution data (Supplemental Figure 2). Because we are dynamically changing the data we store, we assign a binary qualifier (bq) value to each *m/z* feature that specifies data origin. The bq value is 5 digit binary number that correspond to: peak present in 2 GHz, peak saturated in 2 GHz, peak present in 4 GHz, peak saturated in 4 GHz, aligned between modes with 7 ppm (assigned 0) or 20 ppm (assigned 1). This is shown visually in Supplemental Figure 2.

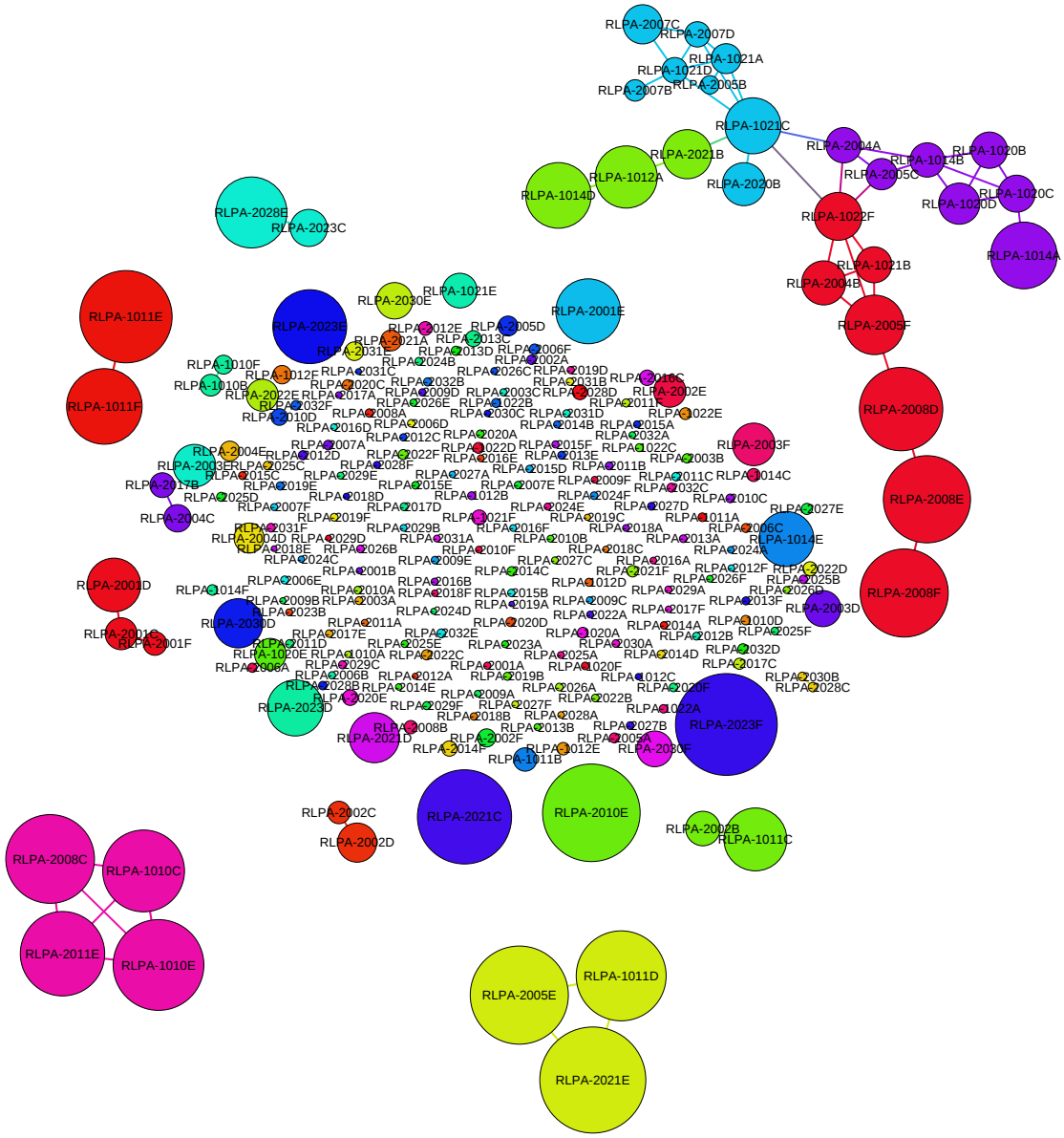
Supplemental Figure 2: The decision tree for *m/z* feature alignment and scoring displaying how peaks are compared across mass spectrometry experiments of the same prefraction in the same ionization mode.



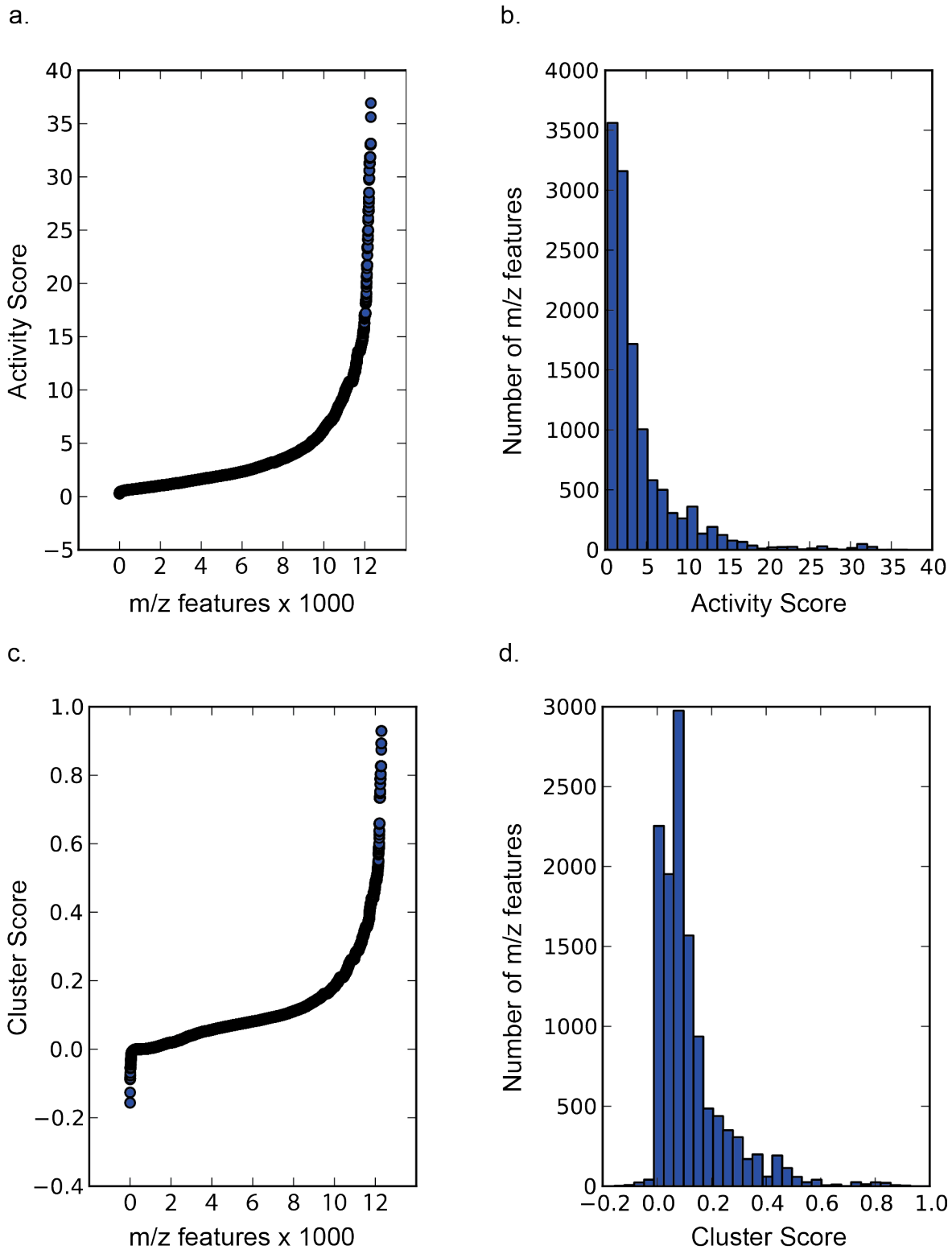
Supplemental Figure 3: Cytological Profiling heatmap output of diluted DMSO prefractions. These are the processed data from which the synthetic fingerprints for *m/z* features and thereby the activity and cluster scores are derived.



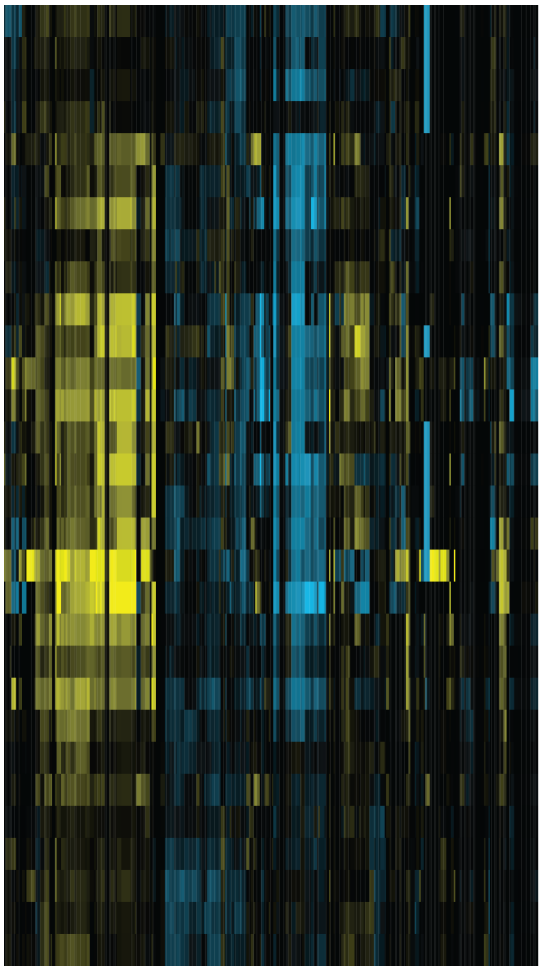
Supplemental Figure 4: A network representation of the data displayed as a heatmap in **Supplental Figure 3**. Prefractions that induce phenotypes with a pearson correlation greater than 0.875 are connected and colored using Gephi's modularity package.



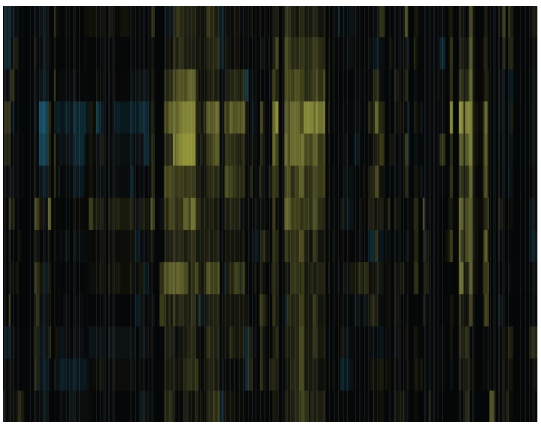
Supplemental Figure 5: Graphs displaying the activity and cluster score values of each *m/z* feature. (a and c) depict activity and cluster score respectively versus feature count. (b and d) histograms of number of *m/z* features versus activity and cluster scores respectively.



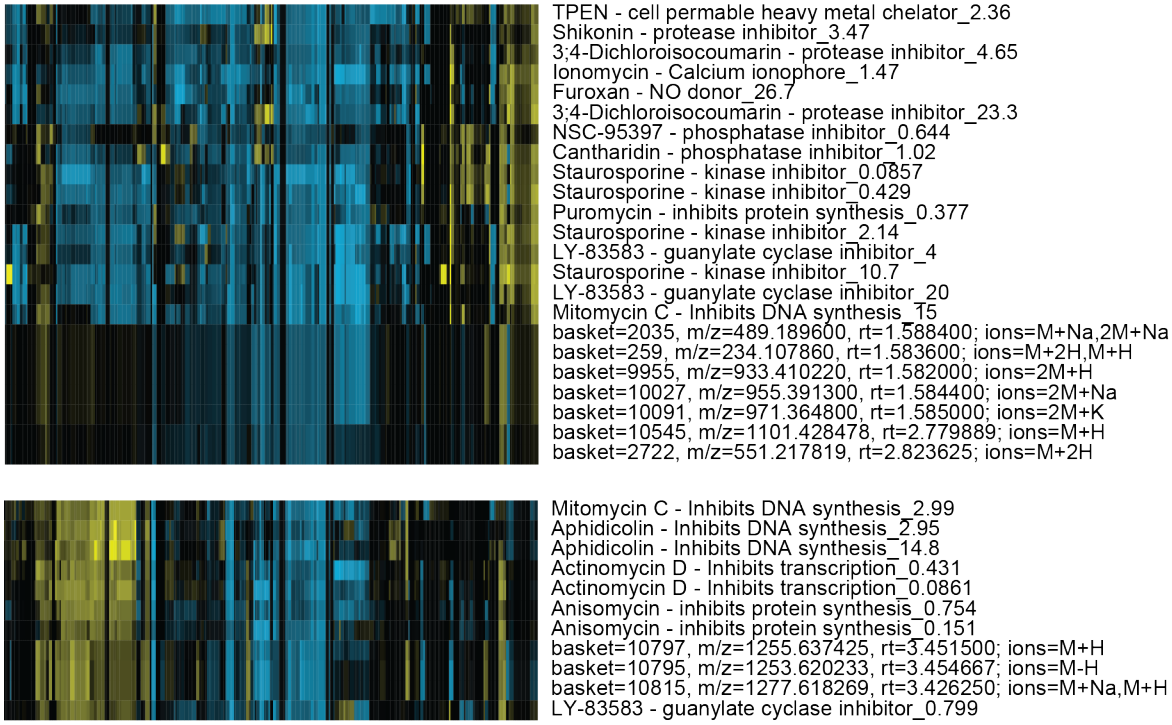
Supplemental Figure 6: The cytological profiles of the ENZO compound library clustered with the purified fluostatins C, D, and J. The compound name is followed by the in well µM concentration.



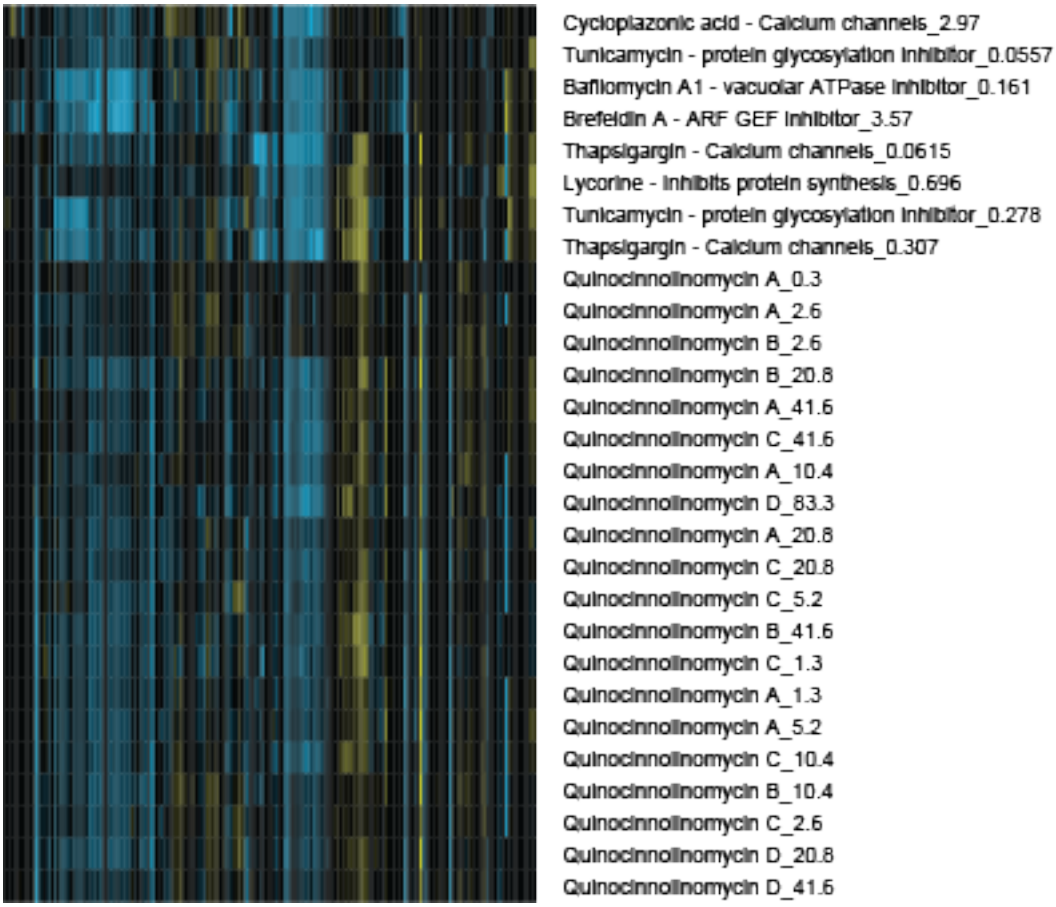
Supplemental Figure 7: The cytological profiles of the ENZO compound library with the purified rosaramicin. The compound name is followed by the in well µM concentration.



Supplemental Figure 8: The cytological profiles of the ENZO compound library with synthetic fingerprints (the predicted cytological profiles of *m/z* features). The first cluster contains *m/z* features corresponding to the compounds staurosporine and echinomycin while the second contains *m/z* features corresponding to actinomycin D.

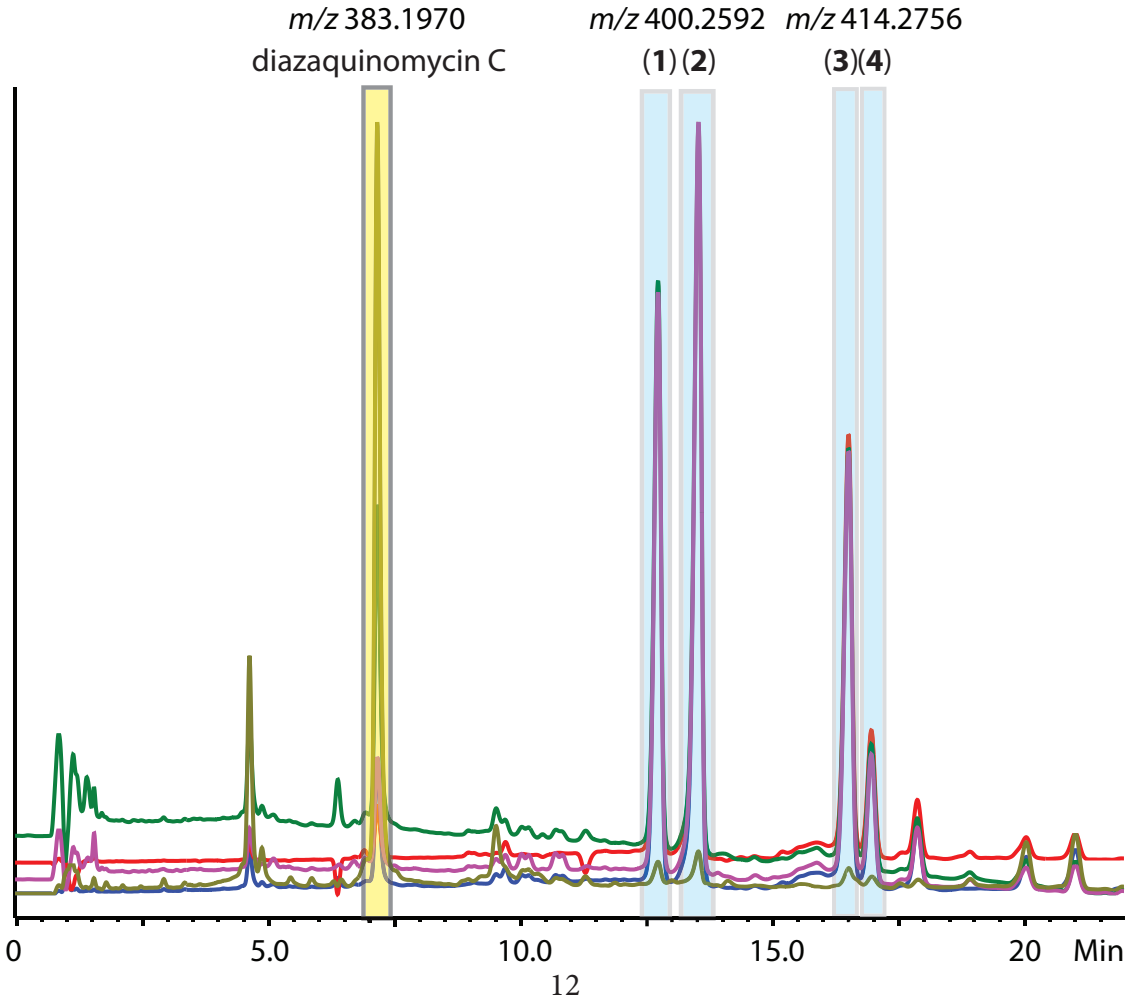


Supplemental Figure 9: The cytological profiles of the ENZO compound library clustered with the purified quinocinnolinomycins (**1-4**) in a dilution series. The compound name is followed by the in well μM concentration. The strong similarity of the cytological fingerprints of the quinocinnolinomycins (**1-4**) with compounds known to cause endoplasmic reticulum stress (thapsigargin, tunicamycin, lycorine, and brefeldin A) suggest that (**1-4**) have a similar mechanism of action.

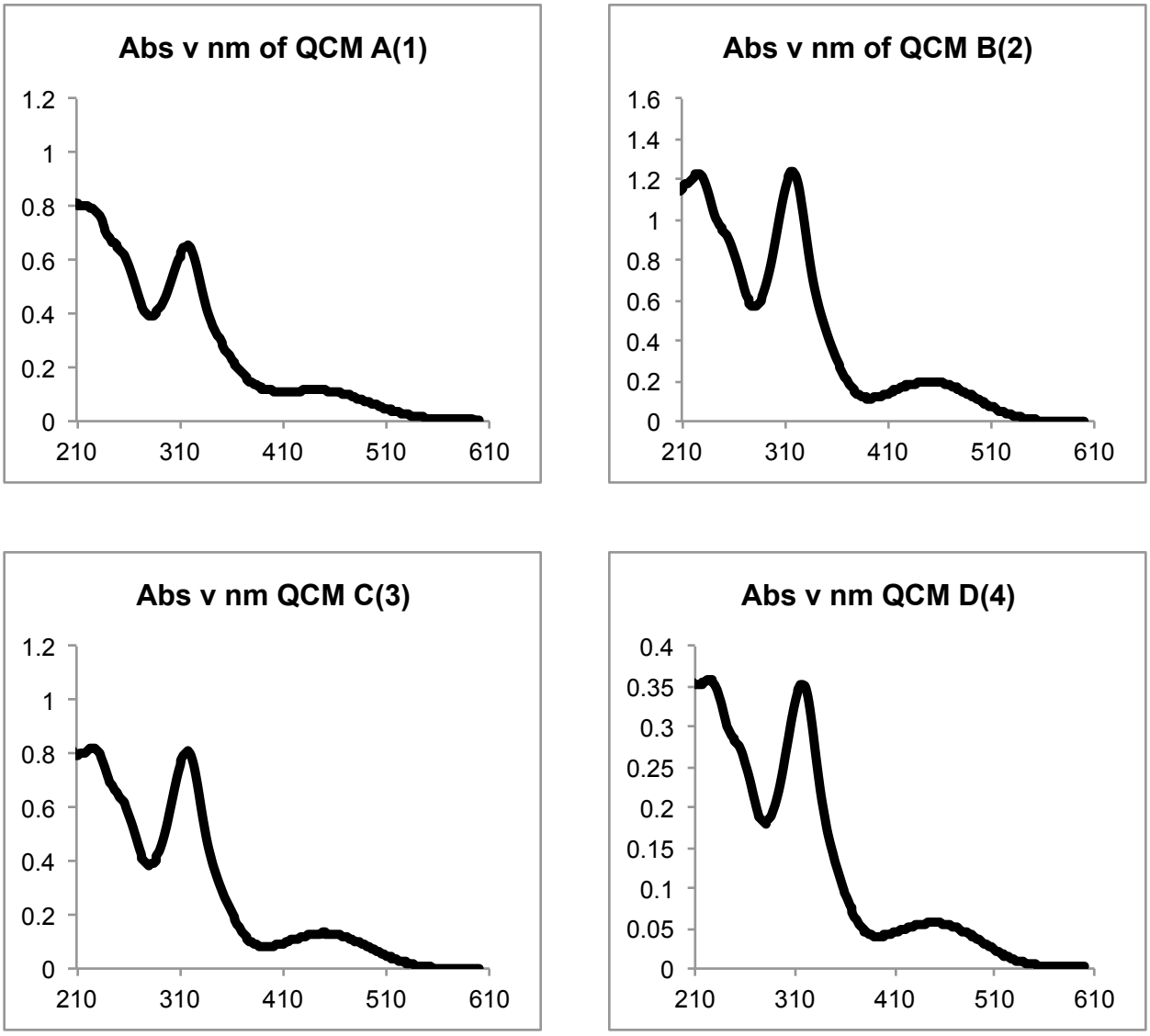


Quinocinnolinomycin growth and isolation: Bacterial frozen stock was struck out on solid media (DIFCO™ Marine Broth 37.4 g and 18.0 g of agar). Colonies were inoculated into a capped 40 mL culture tube with 7.0 ml of liquid media containing 31.2 g of instant ocean, 10.0 g of Soluble Starch, 4.0 g of yeast extract, and 2.0 g of peptone per liter of water. All liquid media cultures were maintained at r.t. and shaken at 210 r.p.m. After 4 days 6 mL from the small-scale culture was used to inoculate 60 mL of the same media in a 250 mL wide neck Erlenmeyer flask with a 1 cm diameter metal spring coil and milk filter top. After 4 days 45 mL of this medium-scale culture were inoculated into 1 L of media in a 2.8 L wide mouth Fernbach flask containing a large spring coil and then topped with a milk filter. This culture was grown for 5 days. The cells were then filtered using a glass filter, washed with sterile water, transferred to a 1.0 L Erlenmeyer flask, and extracted with 250 mL of 1:1 dichloromethane in methanol. The cell debris was filtered off and the extract solution was evaporated under vacuum. This dried extract was prefractionated using the eluotropic series of methanol water described in the methods section. The 80% methanol fraction was dried under vacuum and resuspended in minimal methanol and centrifuged. The supernatant was purified by HPLC on a (Phenomonex Kinetix 2.6 µm XB-C18 100 x 4.6 mm) using a gradient of MeCN:H_2O + 0.02% formic acid (50% MeCN for 2 min, 50%-65% MeCN over 20 min) at a flow rate of 2 ml min^{-1}. The peaks at minutes 8, 14, 15, 17 and 18 (diazaquinomycin C, and 1-4 respectively)were collected separately and dried under vacuum.

Supplemental Figure 10: HPLC-DAD trace of RLPA-2003D with peaks labeled.



Supplemental Figure 11: UV-Vis absorbance spectra for quinocinnolinomcyins A-D (1-4).



Supplemental Table 1: Table of extinction coefficients for quinocinnolinomycins A-D (1-4) at 318 and 450 nm in $M^{-1}cm^{-1}$.

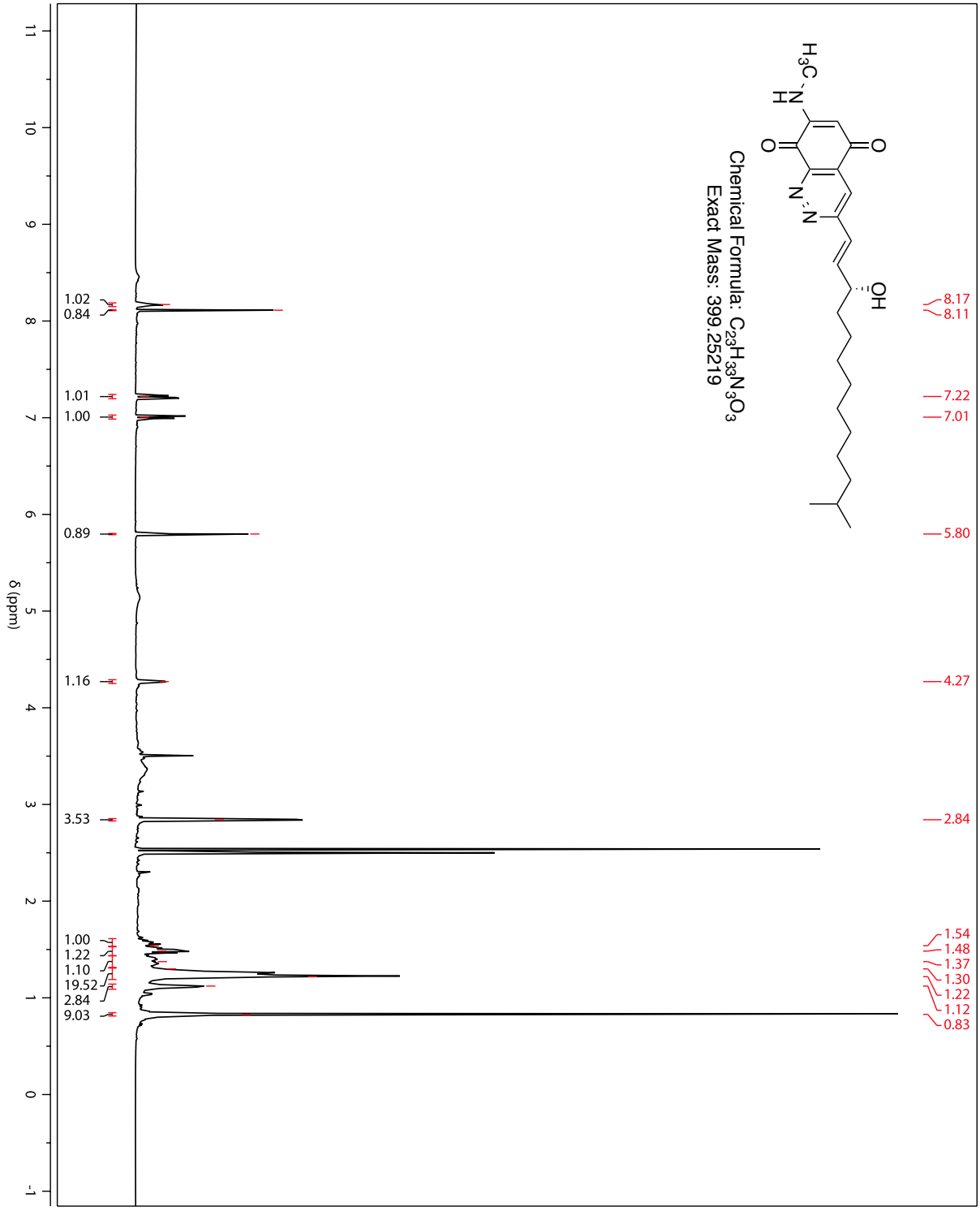
	(log ε) 318	(log ε) 450
QCM A (**1**)	4.00	3.20
QCM B (**2**)	3.72	2.96
QCM C (**3**)	3.47	2.67
QCM D (**4**)	3.82	3.05

Supplemental Table 2: Tabulated NMR data from **1-4**. All spectra were acquired in DMSO-d6 at 600 MHz and 150 MHz for ^{1}H and ^{13}C respectively. The structure of the core of the quinocinnolinomycins and each of the different tails displayed and numbered for clarity.

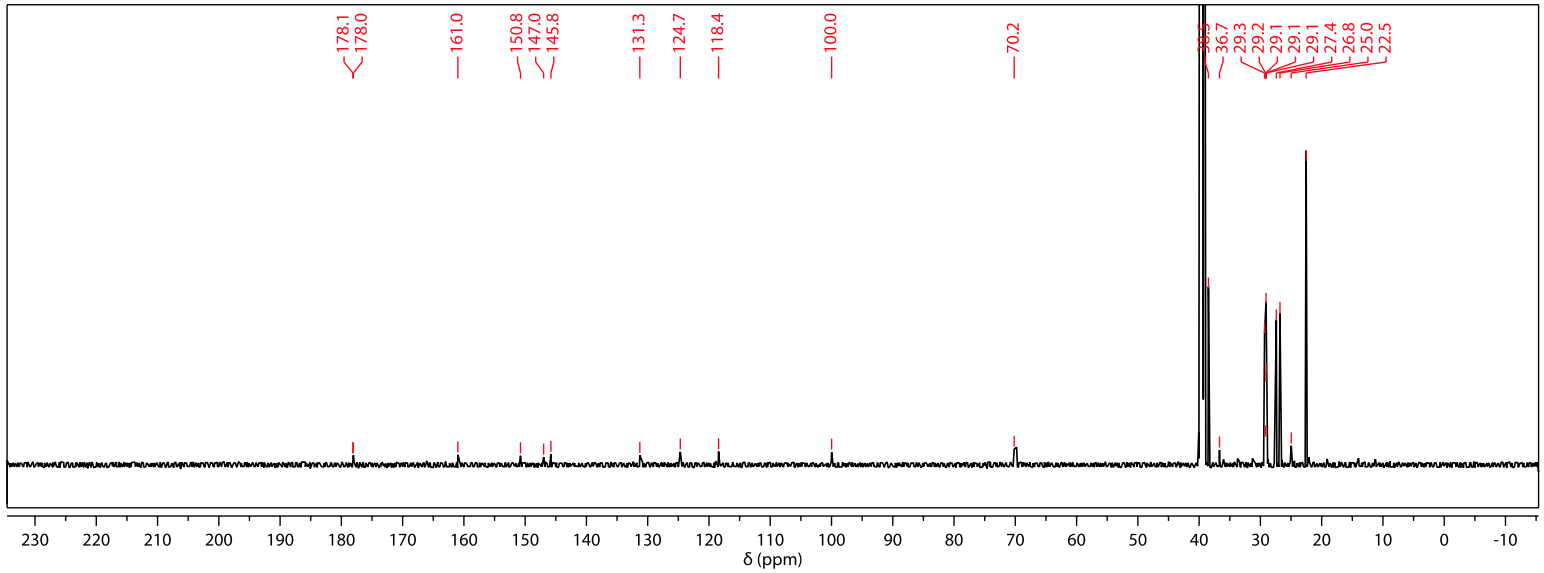
Position	(**1**) δ_C	(**1**) δ_H(J in Hz)	(**2**) δ_C	(**2**) δ_H(J in Hz)	(**3**) δ_C	(**3**) δ_H(J in Hz)	(**4**) δ_C	(**4**) δ_H(J in Hz)
3	161.0		161.0		161.0		161.0	
4	118.2	8.11, s	118.4	8.11, s	118.4	8.11, s	118.4	8.12, s
4a	131.3		131.3		131.3		131.3	
5	178.1		178.1		178.0		178.1	
6	100.0	5.8, s	100.0	5.78, s	100.0	5.80, s	100.0	5.80, s
7	150.8		150.8		150.8		150.8	
8	178.0		178.0		178.1		178.1	
8a	147.0		147.0		147.0			
9	124.5	7.01, dd (16.0, 1.4)	124.7	7.01, dd (16.0, 1.5)	124.7	7.01, dd (16.0, 1.6)	124.7	7.01, dd (16.0, 1.6)
10	145.7	7.22, dd (16.0, 4.9)	145.8	7.22, dd (16.0, 4.9)	145.8	7.22, dd (16.0, 5.0)	145.8	7.22, dd (16.0, 5.0)
11	69.8	4.27, m	70.2	4.271, m	70.2	4.27, m	70.2	4.27, m
12	36.7	1.57, m	36.7	1.57, m	36.7	1.57, m	36.7	1.57, m
		1.51, m		1.51, m		1.51, m		1.51, m
13	25.0	1.37, m	25.0	1.38, m	24.9	1.4, m	24.9	1.38, m
14	29.3-29.1	1.30-1.22, m	29.2-29.0	1.32-1.20, m	29.4-29.0	1.33-1.19, m	29.4-29.0	1.37-1.22, m
15	29.3-29.1	1.30-1.22, m	29.2-29.0	1.32-1.20, m	29.4-29.0	1.33-1.19, m	29.4-29.0	1.37-1.22, m
16	29.3-29.1	1.30-1.22, m	29.2-29.0	1.32-1.20, m	29.4-29.0	1.33-1.19, m	29.4-29.0	1.37-1.22, m
17	29.3-29.1	1.30-1.22, m	29.2-29.0	1.32-1.20, m	29.4-29.0	1.33-1.19, m	29.4-29.0	1.37-1.22, m
18	26.8	1.22, m	29.2-29.0	1.32-1.20, m	29.4-29.0	1.33-1.19, m	29.4-29.0	1.37-1.22, m
19	38.5	1.12, m	29.2-29.0	1.32-1.20, m	36.0	1.25, m	26.8	1.22, m
						1.05, m		
20	27.4	1.48, m	29.2-29.0	1.32-1.20, m	33.7	1.27, m	38.5	1.12, dt (13.7, 6.4)
21	22.1	0.83, d (6.6)	31.3	1.32-1.20, m	28.9	1.33-1.19, m	27.4	1.47, m
						1.08, m		
22			14.0	0.84, t (7.0, 7.0)	11.2	0.82, t (7.4, 7.4)	22.5	0.83, m
23	29.1	2.84, d (4.9)	29.2	2.84, s	19.1	0.81, d (6.2)		
24					29.2	2.84, d (5.1)	29.2	2.84, d (5.0)
NH		8.17, q (4.7)		8.16, s		8.17, q (5.2)		8.16, q (4.8)
OH				5.12, d (4.9)		5.13, s		5.12, d (4.7)

Structure Elucidation Details for quinocinnolinomycin A (1): The molecular formulae $C_{23}H_{33}N_3O_3$ and $C_{24}H_{35}N_3O_3$ were determined based on the strong consensus between the $[M + H]^+$ and $[M + Na]^+$ m/z features for each set of two constitutional isomers. The earliest eluting compound with the molecular formula C23H33N3O3 was solved by NMR analysis, using a combination of ^1H, ^{13}C, gCOSY, gHSQC, gHMBC, and 1D-TOCSY spectra (Fig. 5). Consideration of the ^1H-NMR spectrum indicated the presence of two vinylic and two aromatic signals, an aryl amine, one N-methyl doublet, two methyl doublets, and multiple overlapping resonances in a methylene envelope at 1.20 – 1.32 ppm (Table S2). Interpretation of the ^{13}C and gHSQC spectra confirmed the presence of two ketones and four aromatic/quaternary carbons. The planar structure of the tail of the molecule was assigned from either side of the methylene region using gCOSY and HMBC correlations (Fig. 5B). 1D-TOCSY from the H11 proton to the H21 was used to confirm that the allyl oxymethine was connected to the tail through the methylene region.

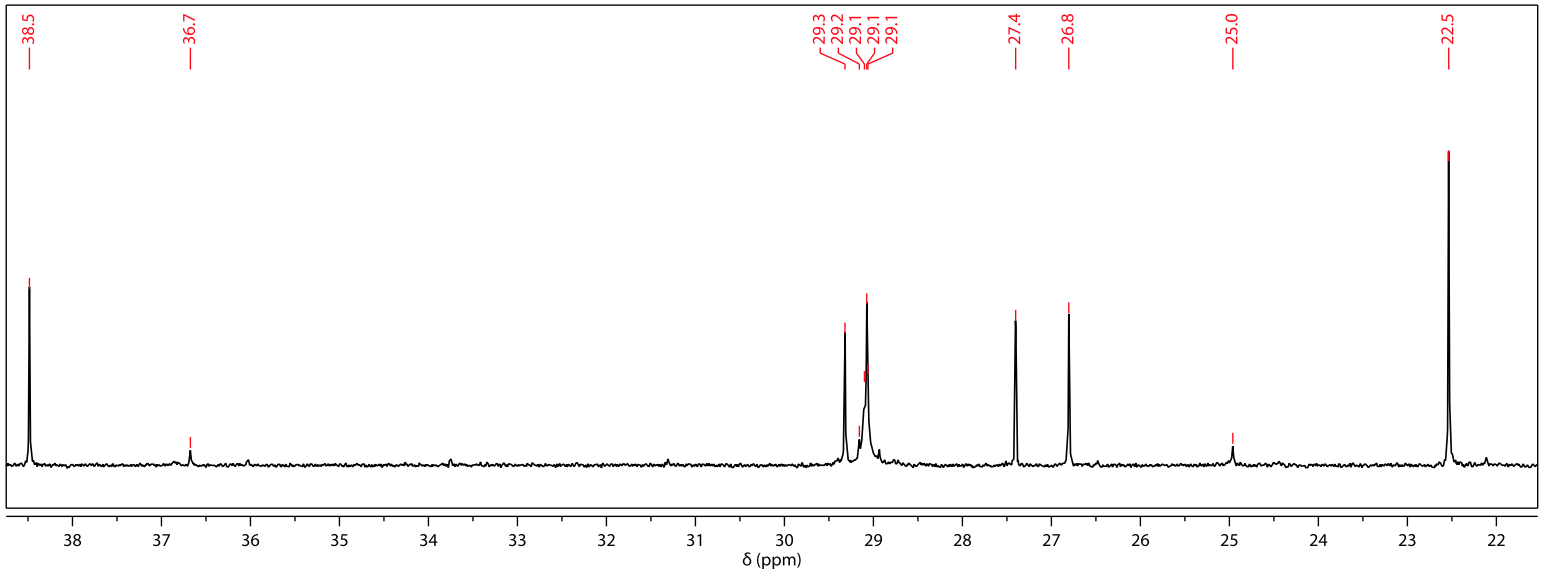
The remaining $C_9N_3O_2$ belong to the headgroup. The attachment to the tail was assigned by gHMBC correlations from the vinylic protons at δ 7.01 and 7.22 to the carbons at positions 3 and 4 and gHMBC correlations from the aromatic proton on position 4 at δ 8.11 to the carbon at position 9. One of the ketones could be placed at position 5 based on gHMBC correlations from the aromatic resonance at δ 8.11 (Fig. 5B). The resonance at δ 5.80 could be assigned to the vinylic proton at position 6 between the two ketones that form the quinone based on gHMBC correlations to the carbons at positions 4a, 5, 7, and 8. The gHMBC correlation from the N-methyl at δ 2.84 to the carbons at positions 6, 7, and 8 placed the relative orientation of the vinylic amine (Fig. 5B). Finally, the incorporation of the remaining atoms (CN_2) could only satisfy the requirement for three additional degrees of unsaturation by inclusion of the quaternary carbon at position 8a and two heteroaromatic nitrogens in positions 1 and 2, thus completing the structural assignment (Fig. 5B). The 'R' stereochemistry of quinocinnolinomycin A was determined using Mosher's ester method (Fig. 5C), and this assignment extended to quinocinnolinomycins B-D based on their common biosynthetic origin (fig. S12).


¹H-NMR of quinocinnolinomycin A (1) at 600 MHz in DMSO-d6
Chemical Formula: C23H33N3O3
Exact Mass: 399.25219
δ (ppm)
1.02
0.84
1.01
1.00
0.89
1.16
3.53
1.00
1.22
1.10
19.52
2.84
9.03
8.17
8.11
7.22
7.01
5.80
4.27
2.84
1.54
1.48
1.37
1.30
1.22
1.12
0.83

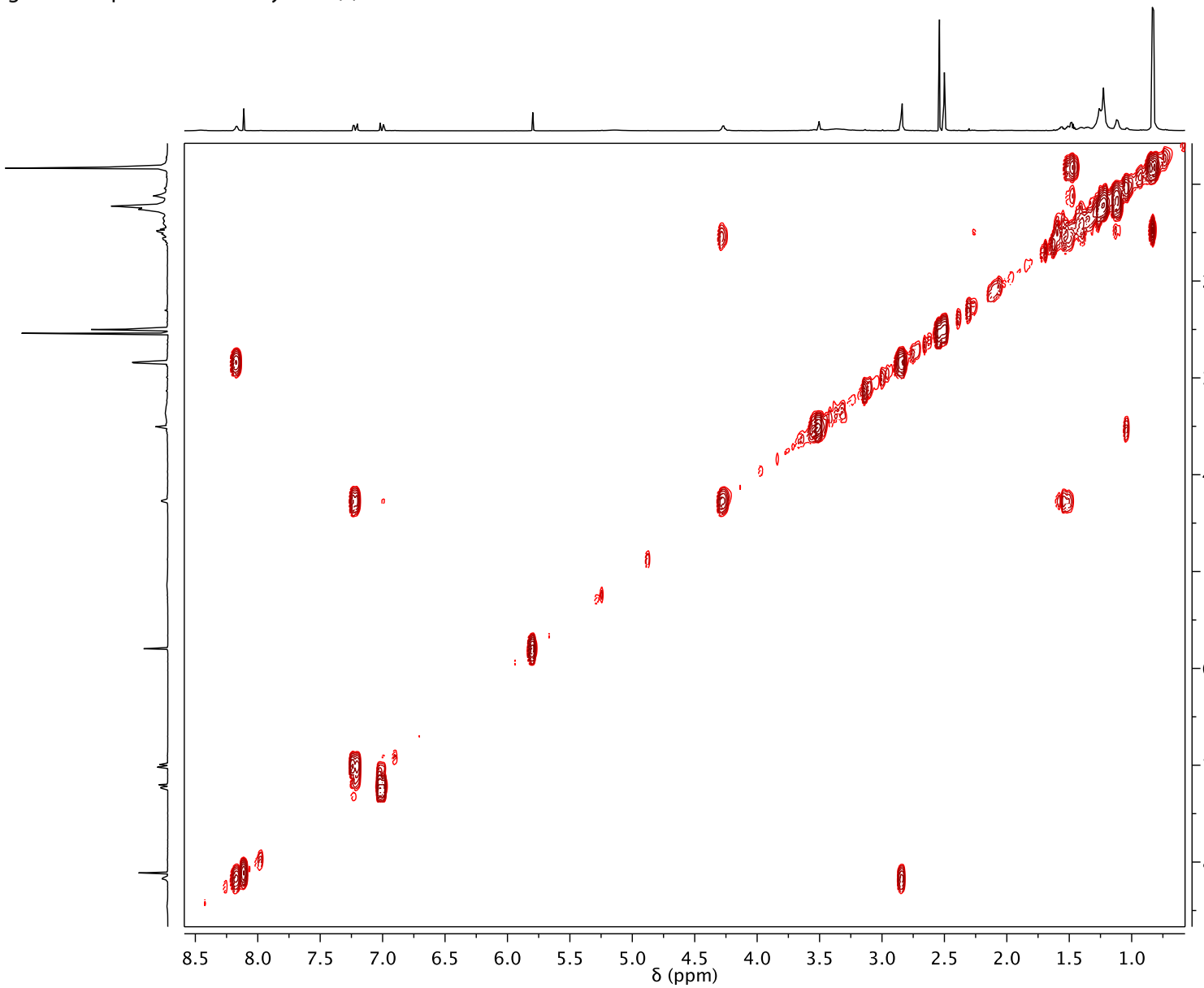
^{13}C-NMR of quinocinnolinomycin A (**1**) at 150 MHz in DMSO-d6



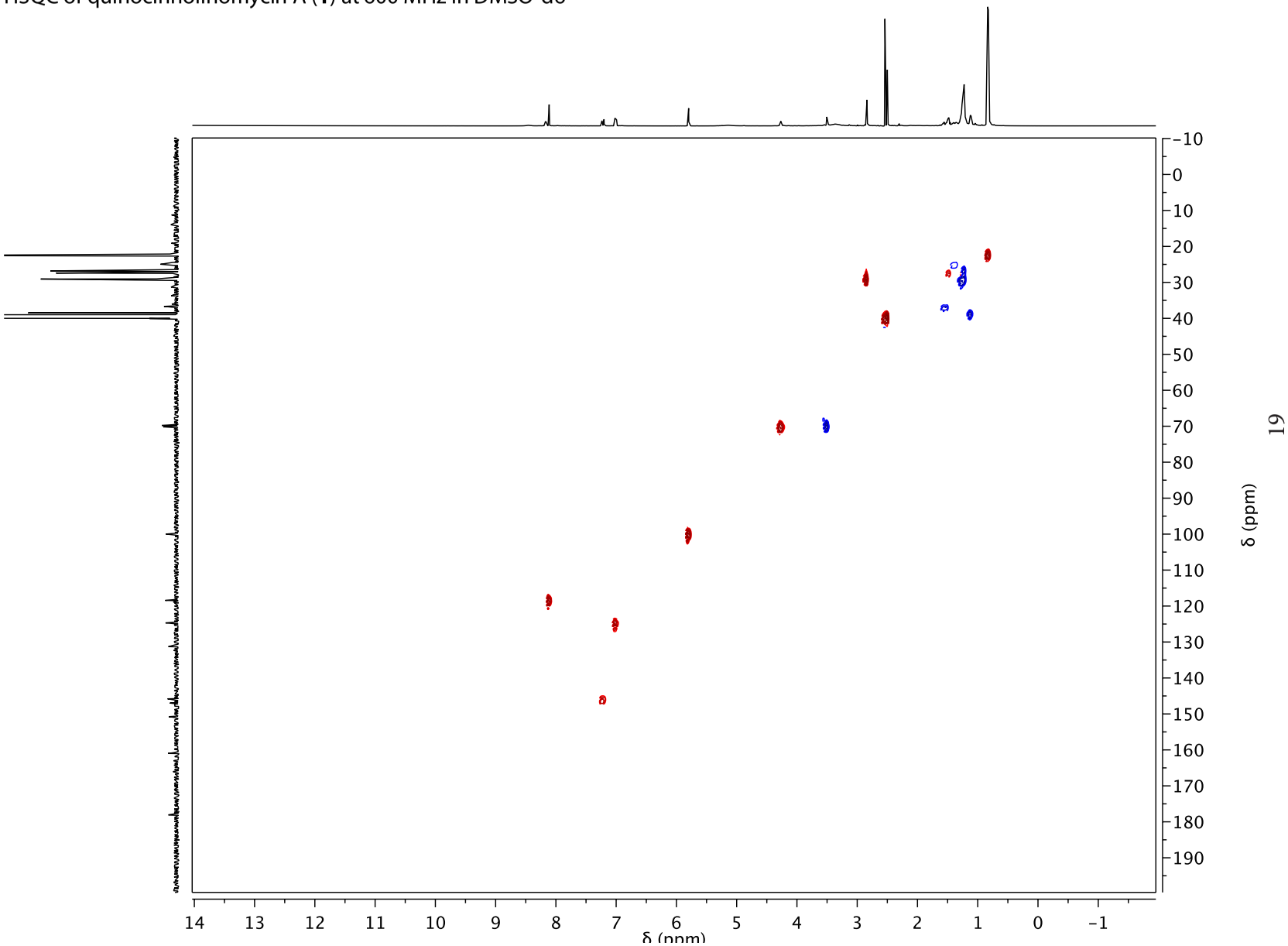
^{13}C-NMR of quinocinnolinomycin A (**1**) at 150 MHz in DMSO-d6



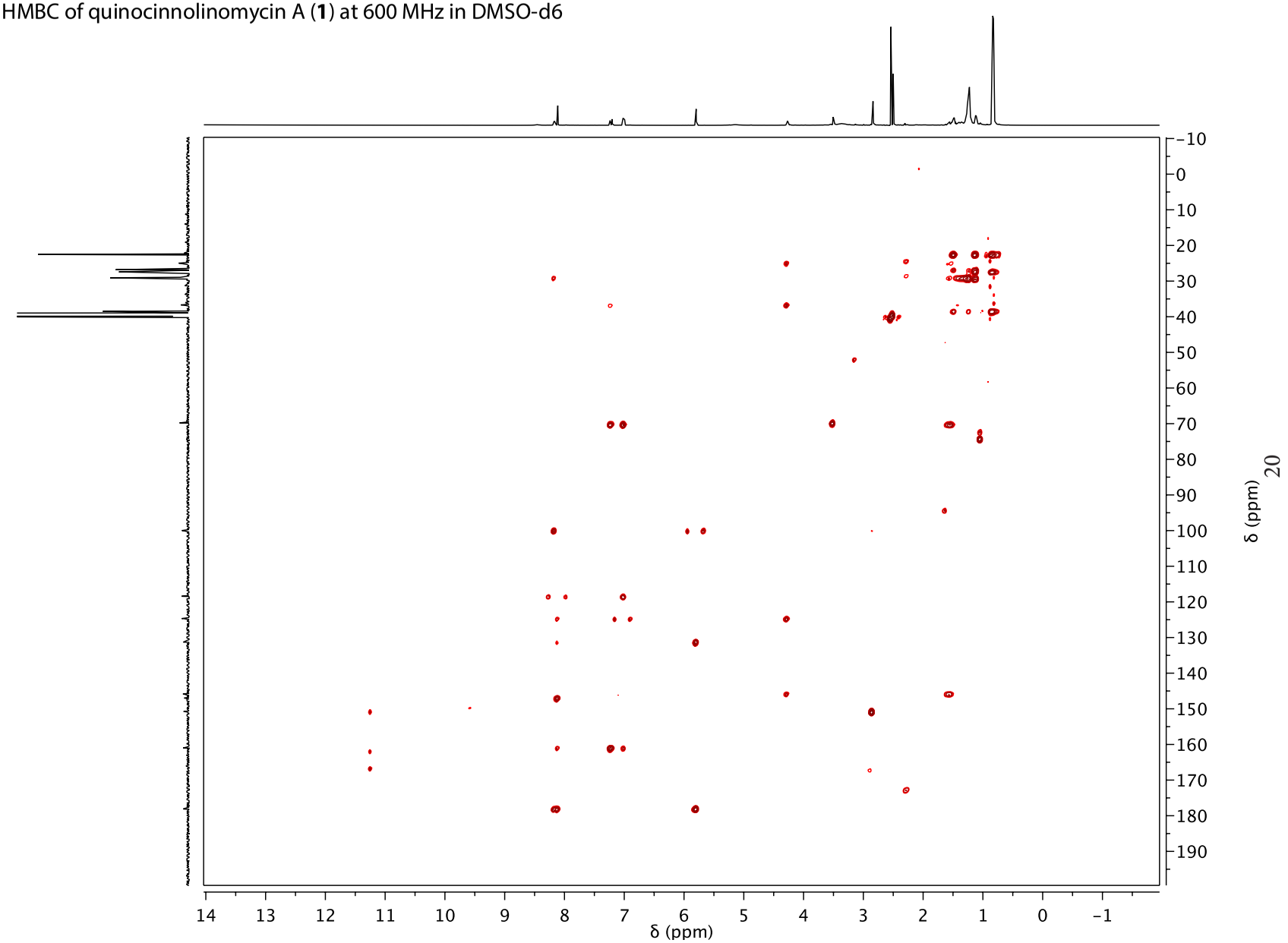
gCOSY of quinocinnolinomycin A (**1**) at 600 MHz in DMSO-d6



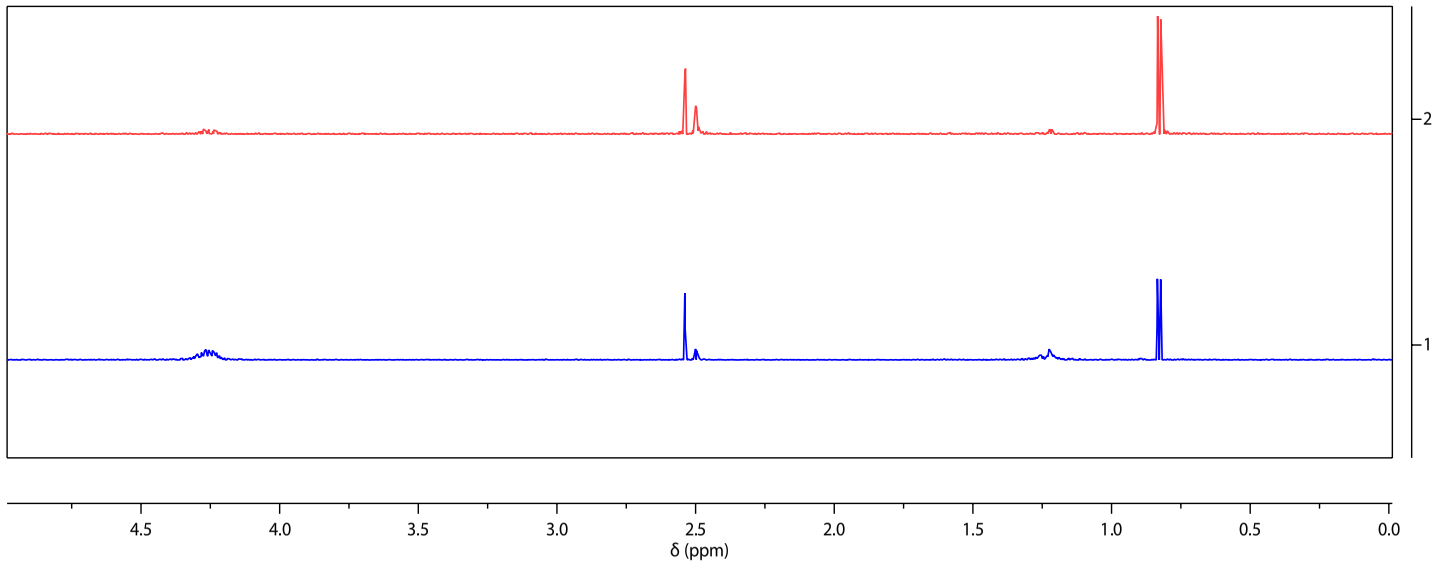
HSQC of quinocinnolinomycin A (**1**) at 600 MHz in DMSO-d6



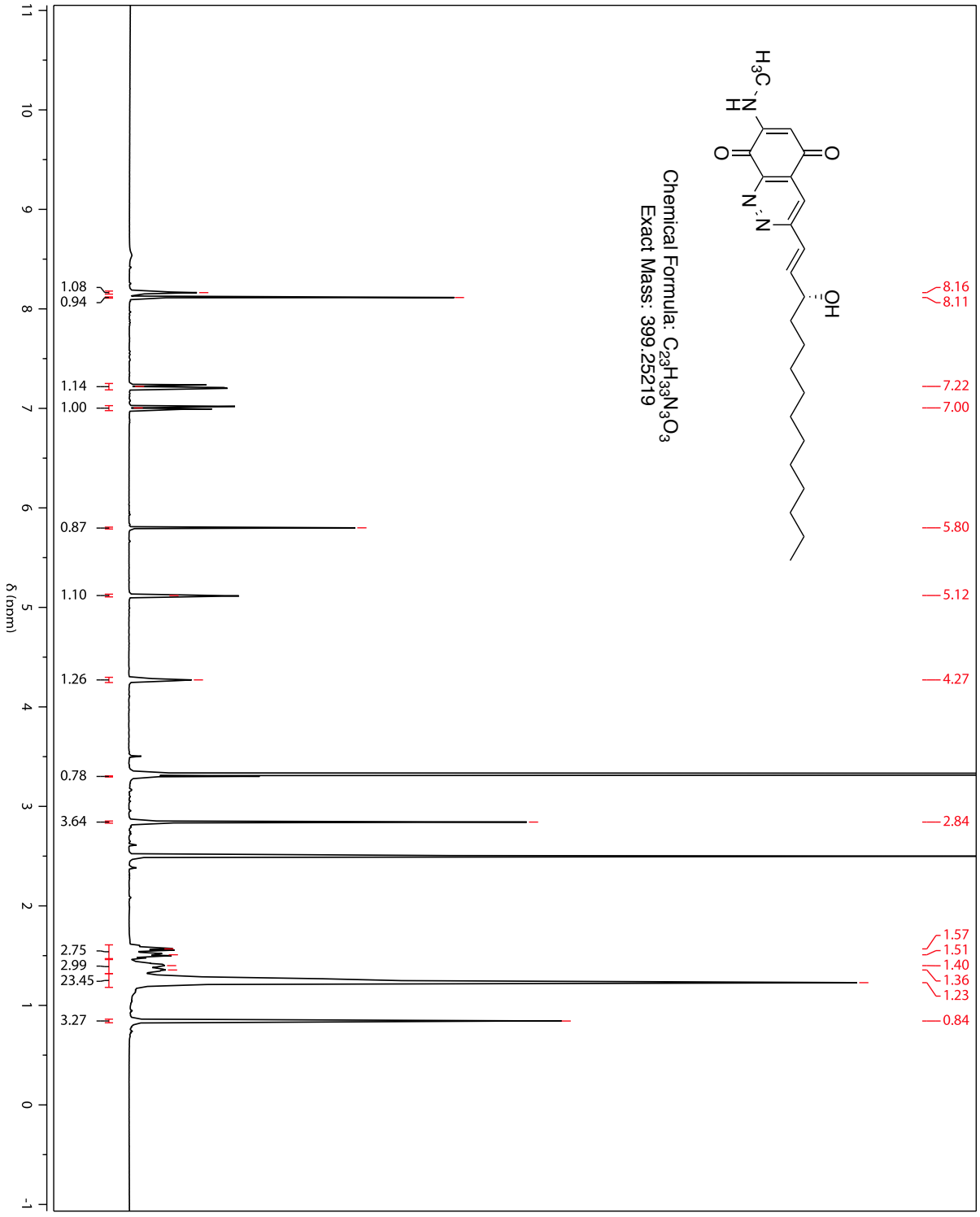
HMBC of quinocinnolinomycin A (**1**) at 600 MHz in DMSO-d6



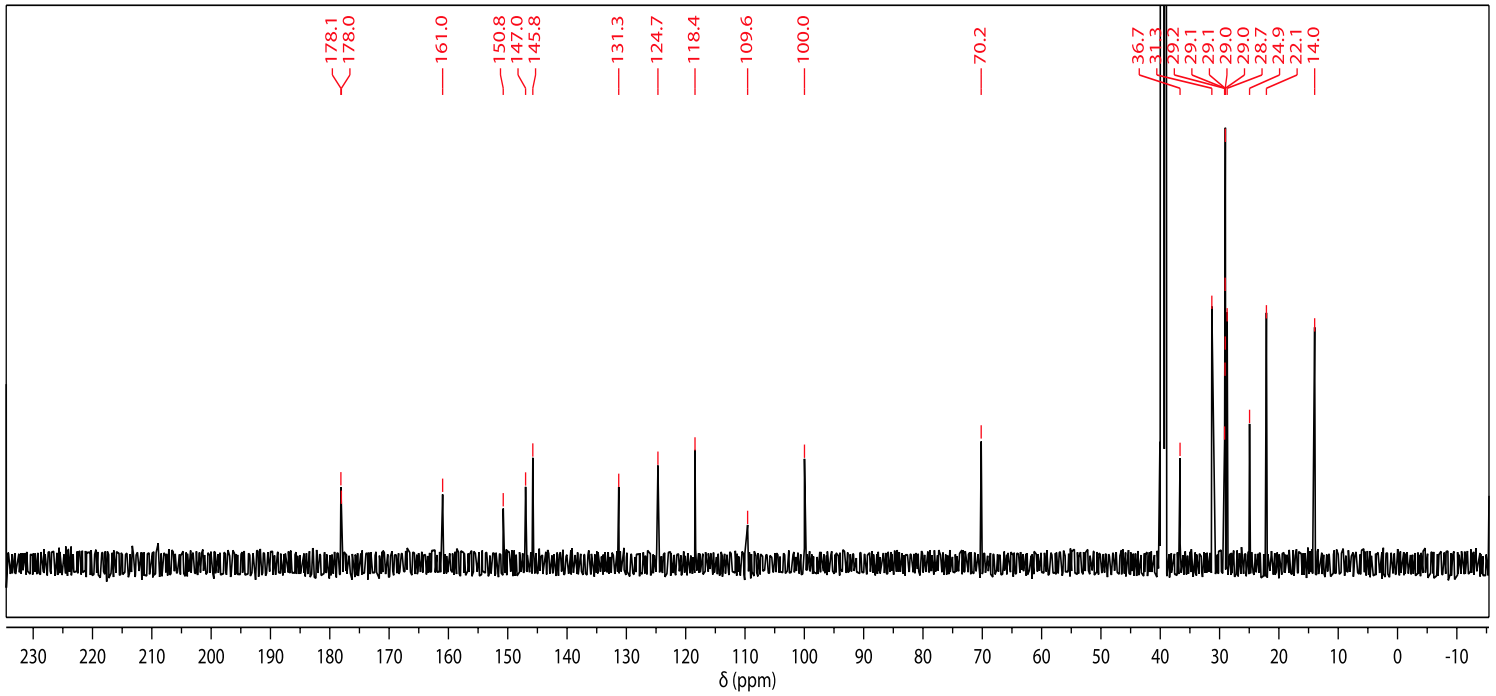
1D TOCSY of quinocinnolinomycin A (**1**) at 600 MHz in DMSO-d6 excited at 4.27 ppm.



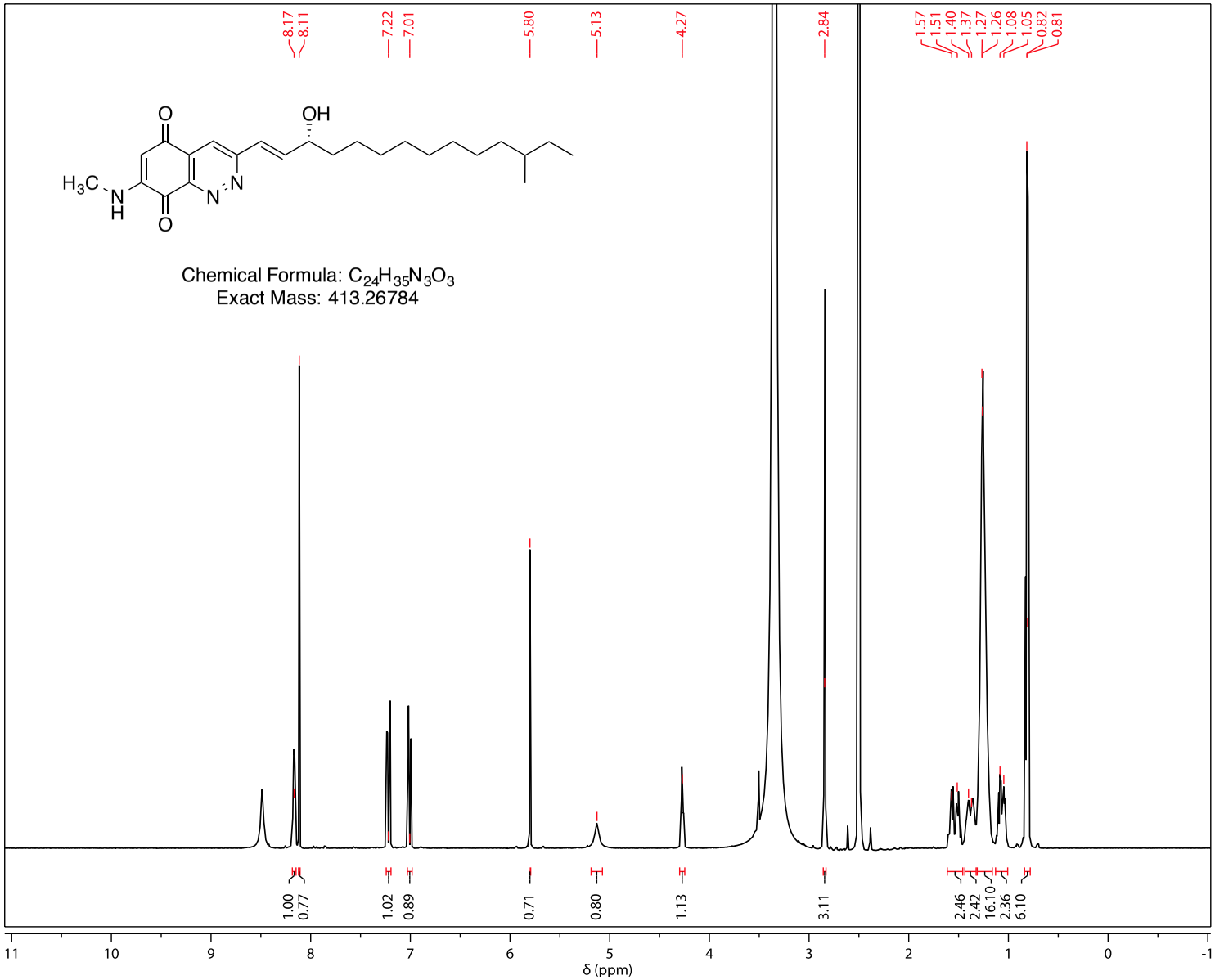
^{1}H-NMR of quinocinnolinomycin B (**2**) at 600 MHz in DMSO-d6

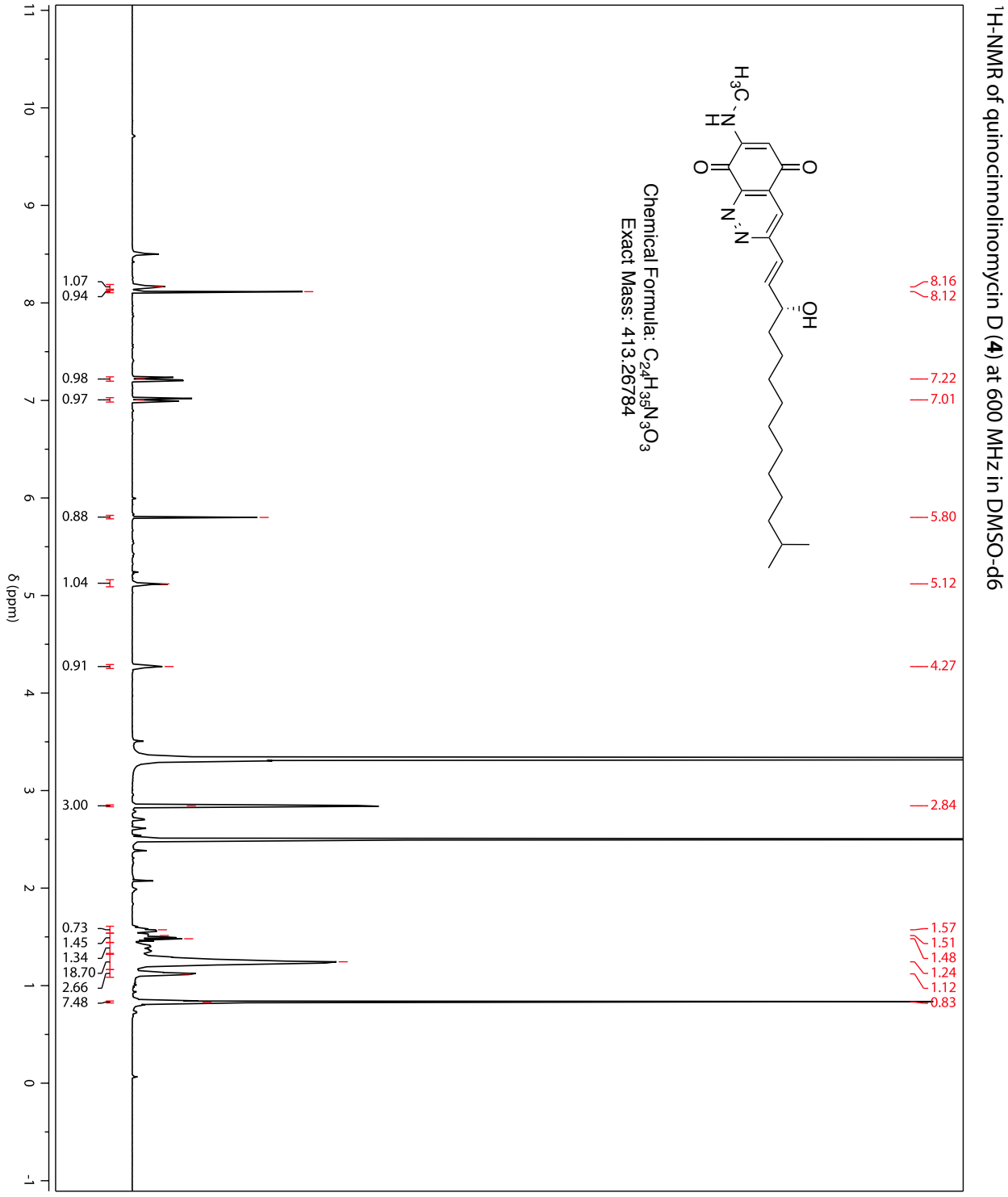


^{13}C-NMR of quinocinnolinomycin B (**2**) at 150 MHz in DMSO-d6



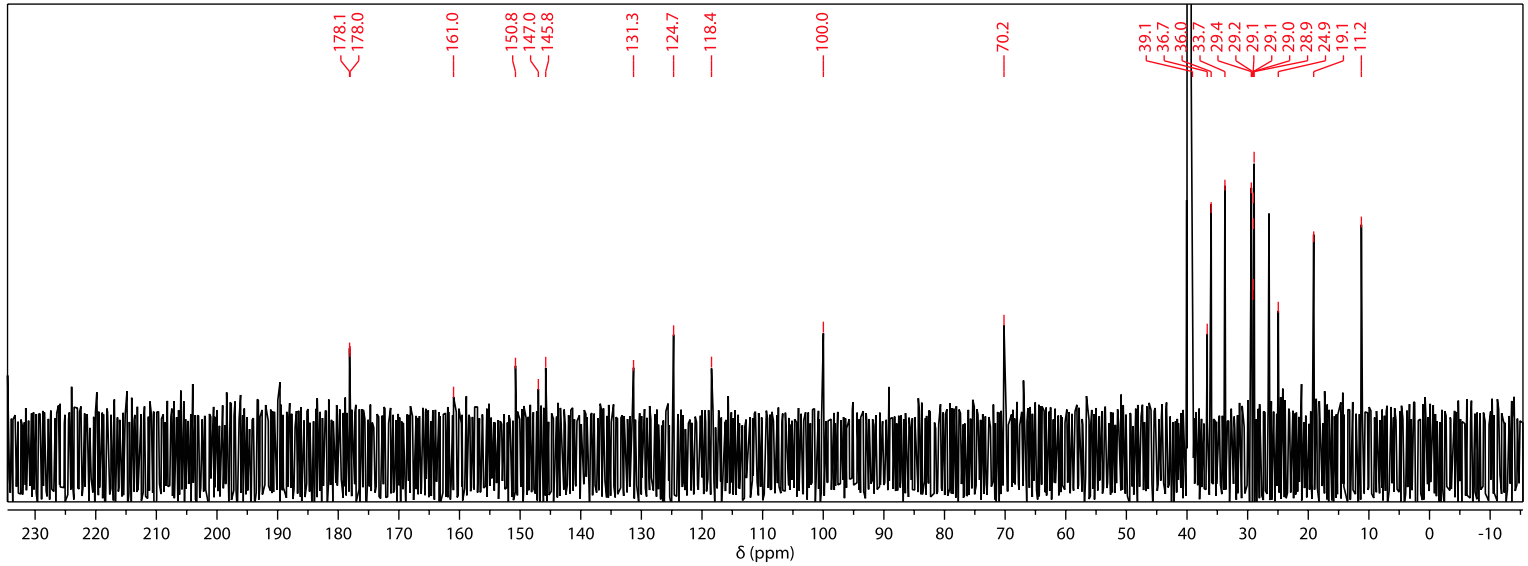
[1]H-NMR of quinocinnolinomycin C (**3**) at 600 MHz in DMSO-d6



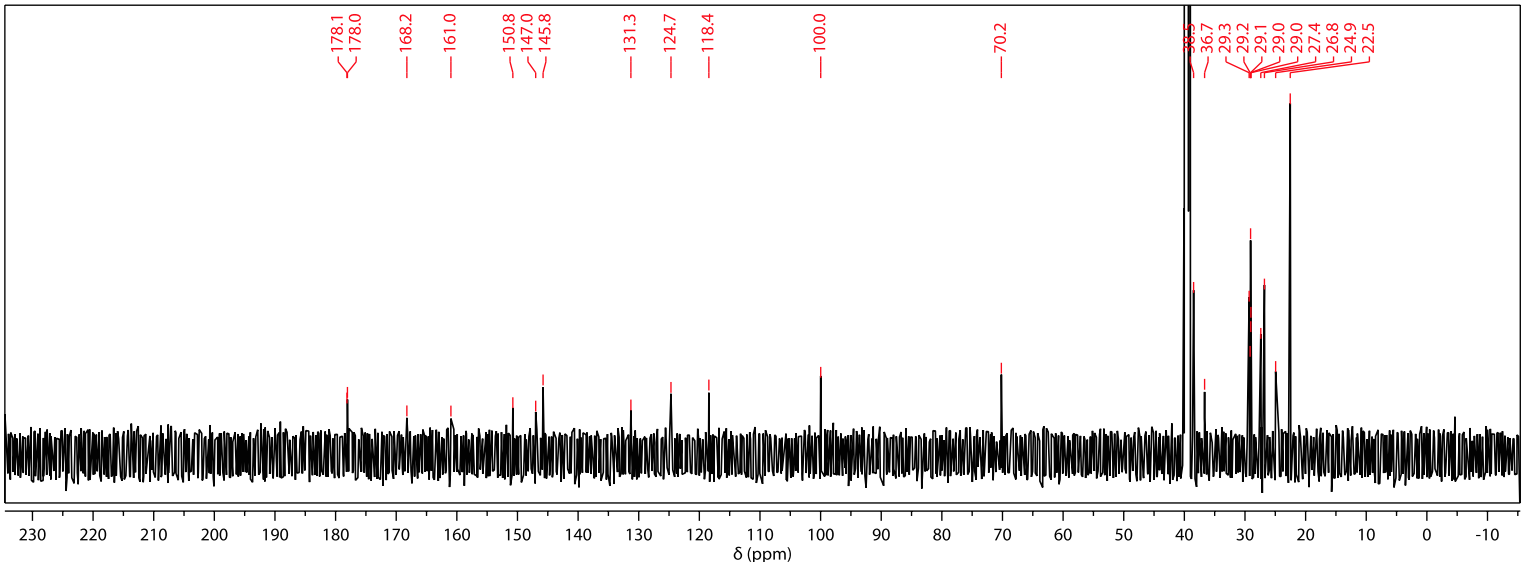


^{1}H-NMR of quinocinnolinomycin D (**4**) at 600 MHz in DMSO-d6

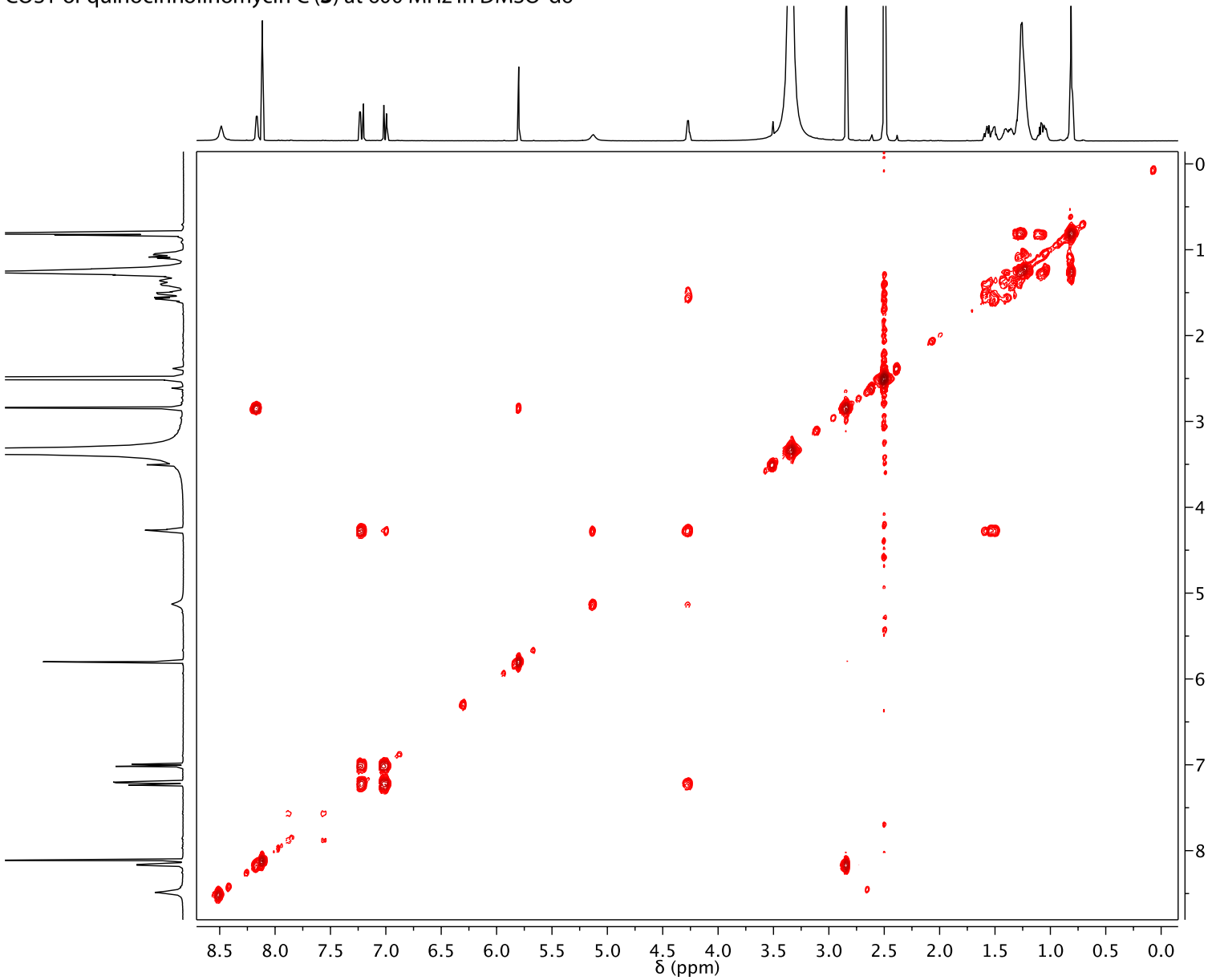
^{13}C-NMR of quinocinnolinomycin C (**3**) at 150 MHz in DMSO-d6



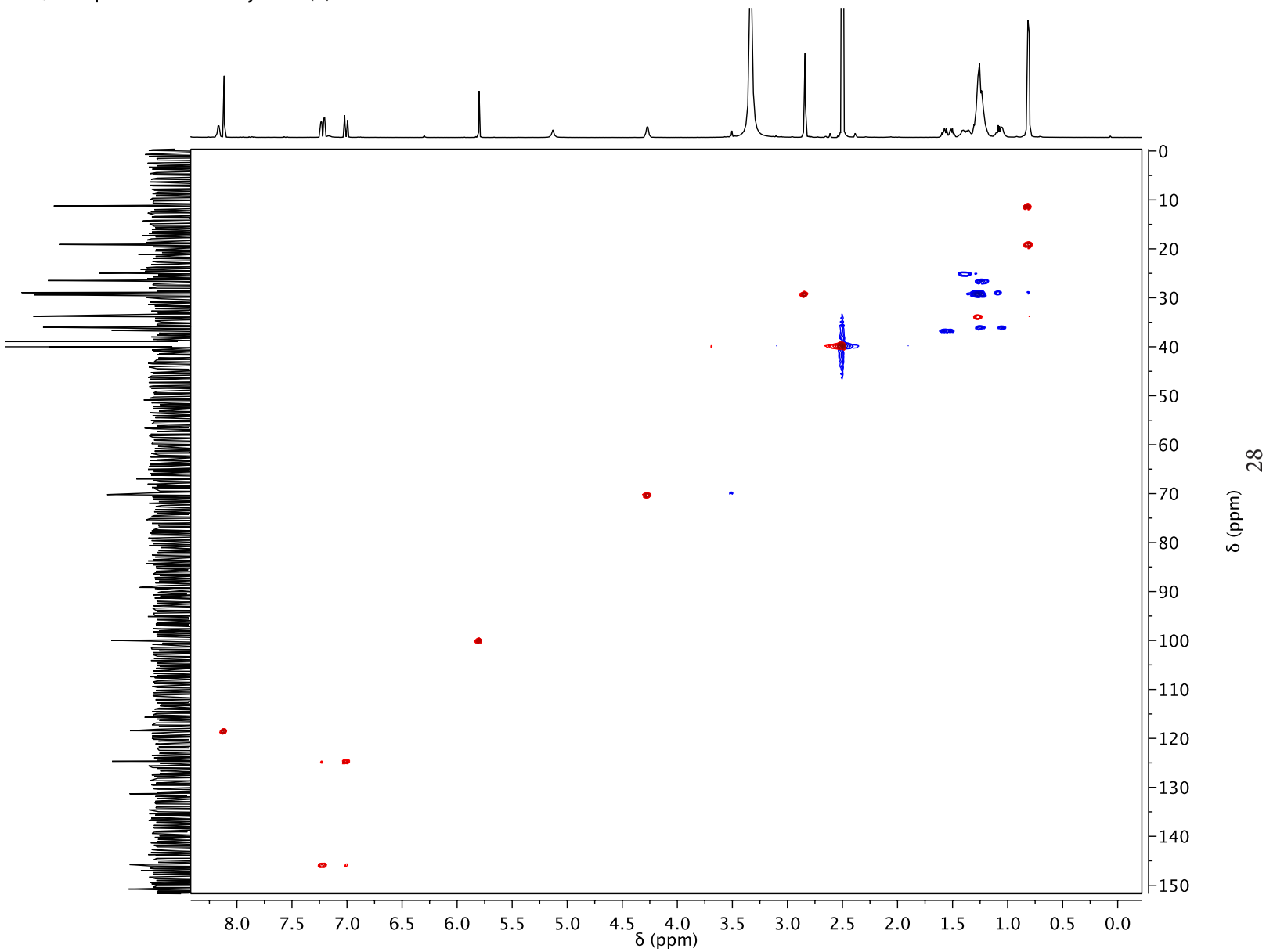
^{13}C-NMR of quinocinnolinomycin D (**4**) at 150 MHz in DMSO-d6



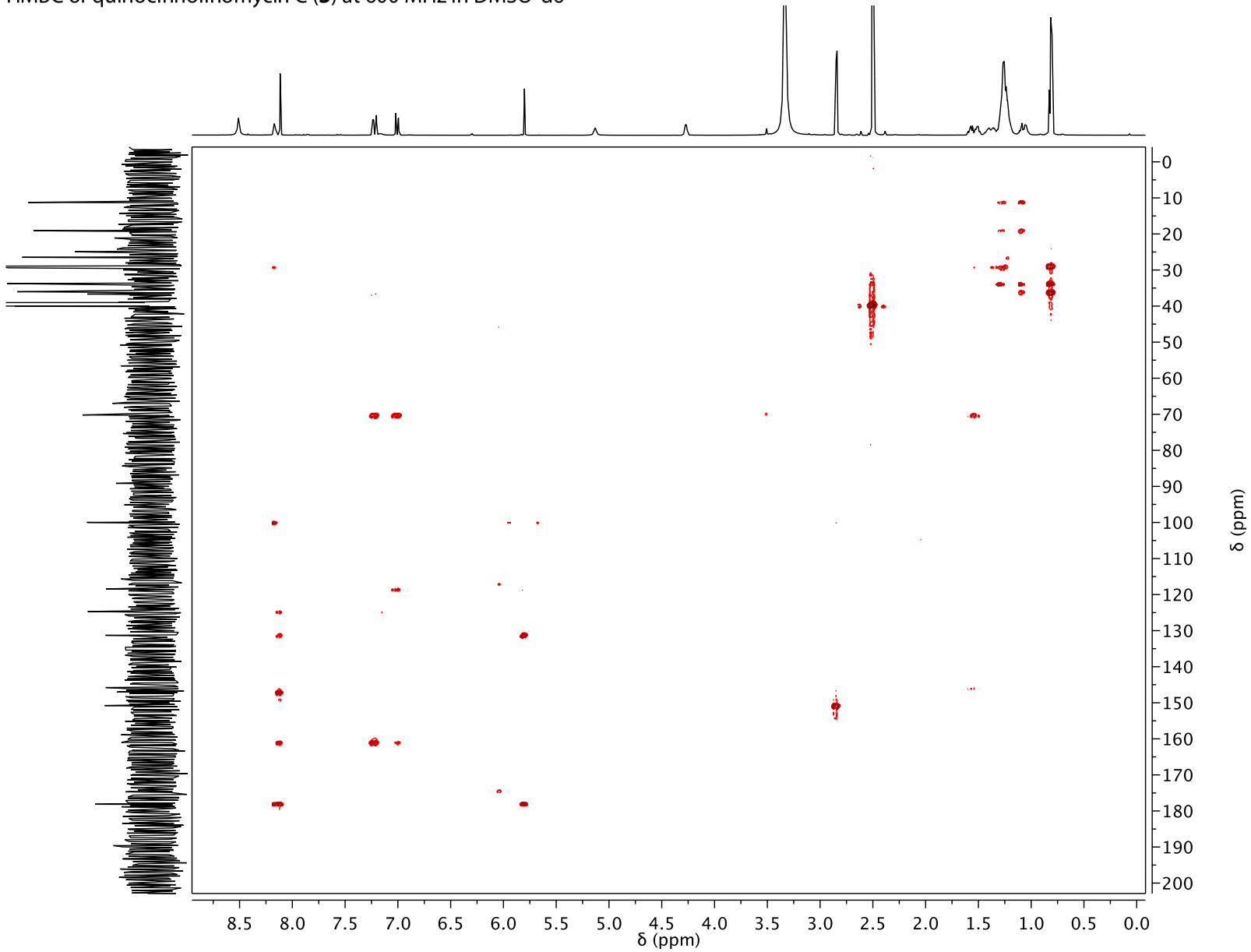
COSY of quinocinnolinomycin C (**3**) at 600 MHz in DMSO-d6



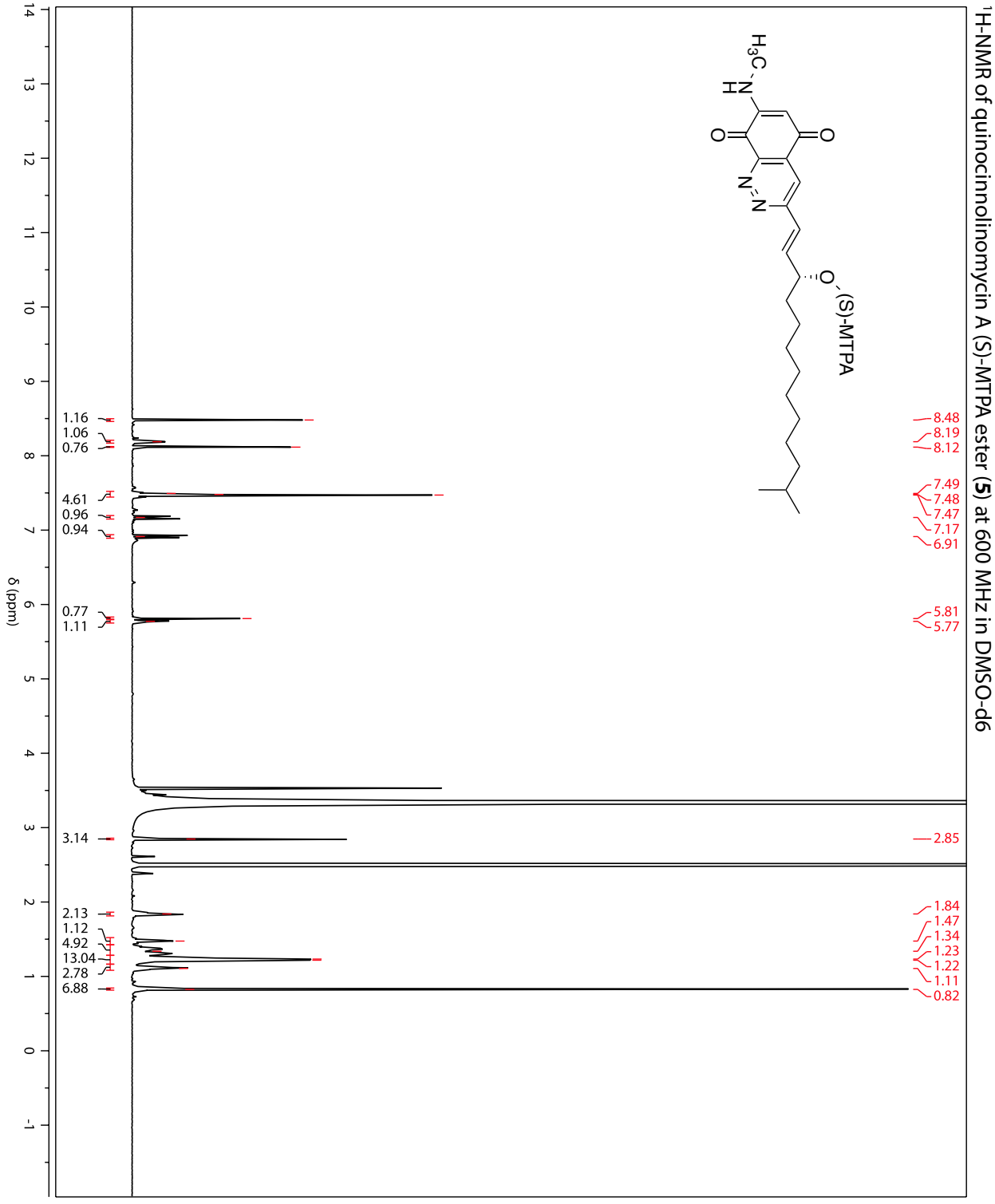
HSQC of quinocinnolinomycin C (**3**) at 600 MHz in DMSO-d6



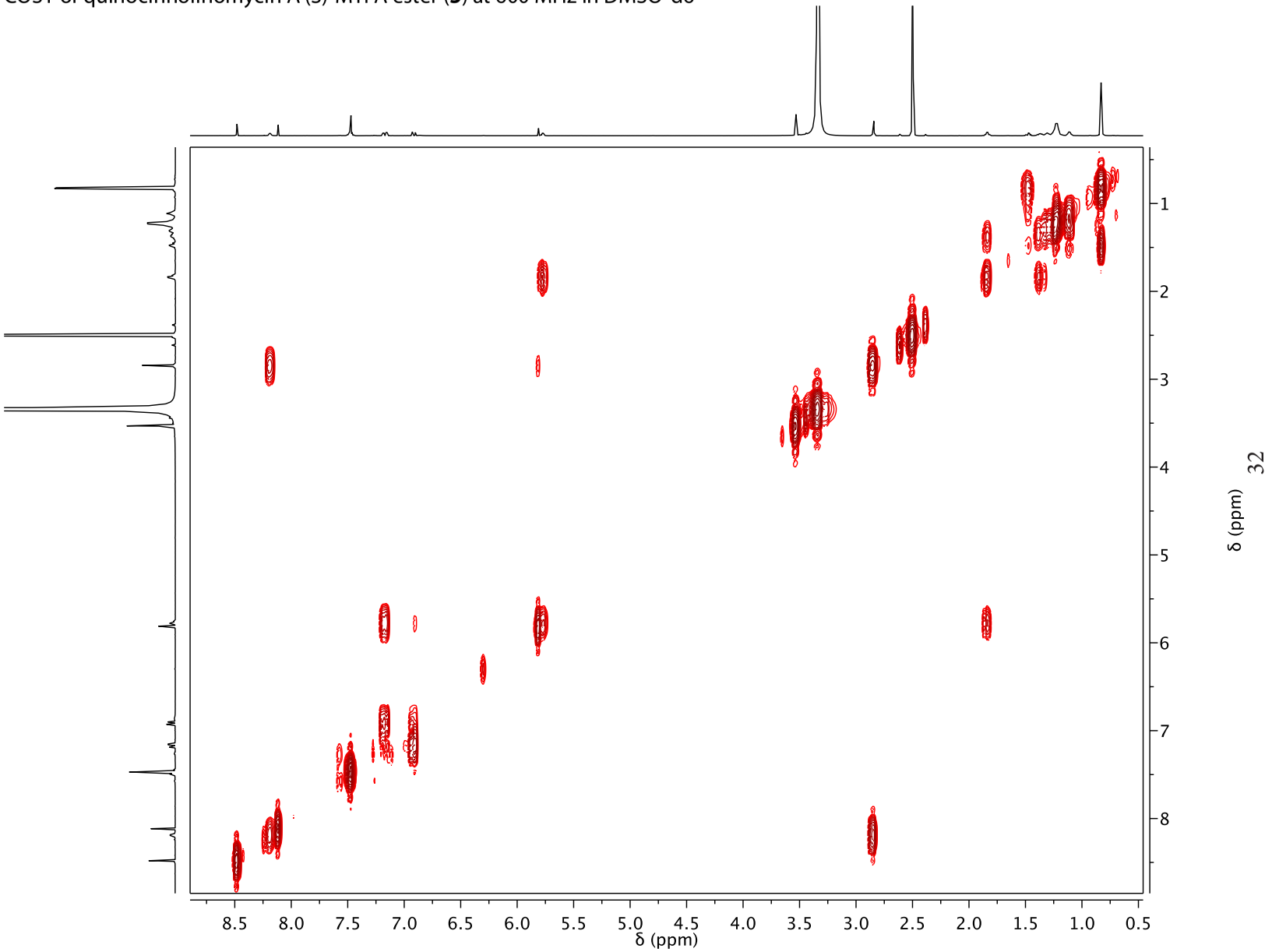
HMBC of quinocinnolinomycin C (**3**) at 600 MHz in DMSO-d6

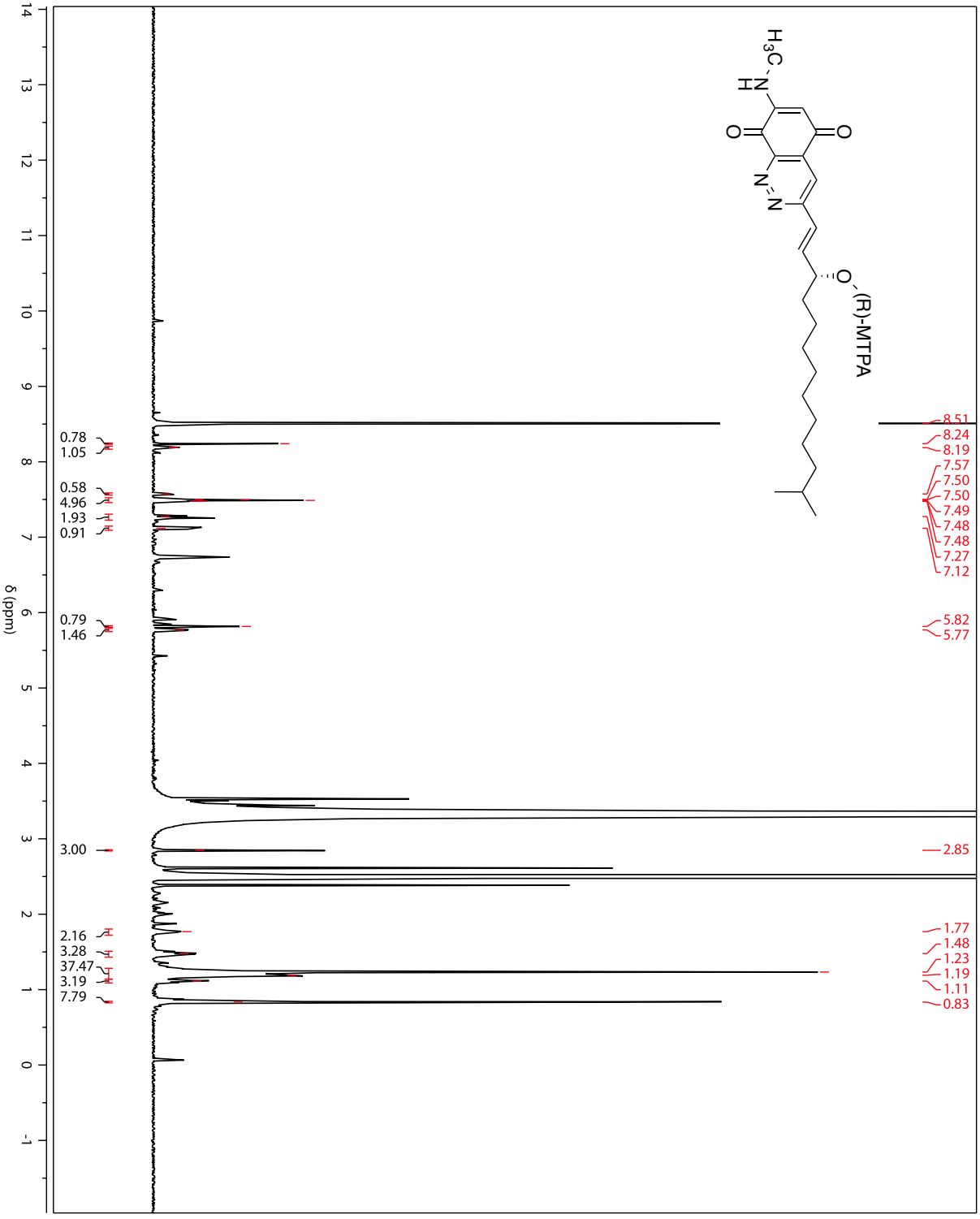


Synthesis of (*S*) and (*R*)-MTPA esters (5, 6) of quinocinnolinomycin A (1): In two vials with teflon septa pierced by a blunt needle containing a small stir bar, 1.0 mg of compound 1 was dried under vacuum. The vessels were flushed with argon and 0.200 mL of dry pyridine added to each. The *R* and *S* MTPA-Cl were added in 6-fold excess, one to each vial, and the reaction was run for 1 h at RT. The reaction was quenched by addition of a drop of methanol. The products were run through silica plugs, eluted with dichlormethane, and dried under vacuum. The subsequent products were purified by HPLC by (Eclipse XDB-C18 5 μm 4.6 x 150 mm) reverse phase column on a gradient of using a gradient of MeOH:H_2O + 0.02% formic acid (60%-100% MeOH over 16 min) at a flow rate of 2 mL min^{-1}. The *S*-ester eluted at 11 minutes while the *R*-ester eluted at 15 minutes.

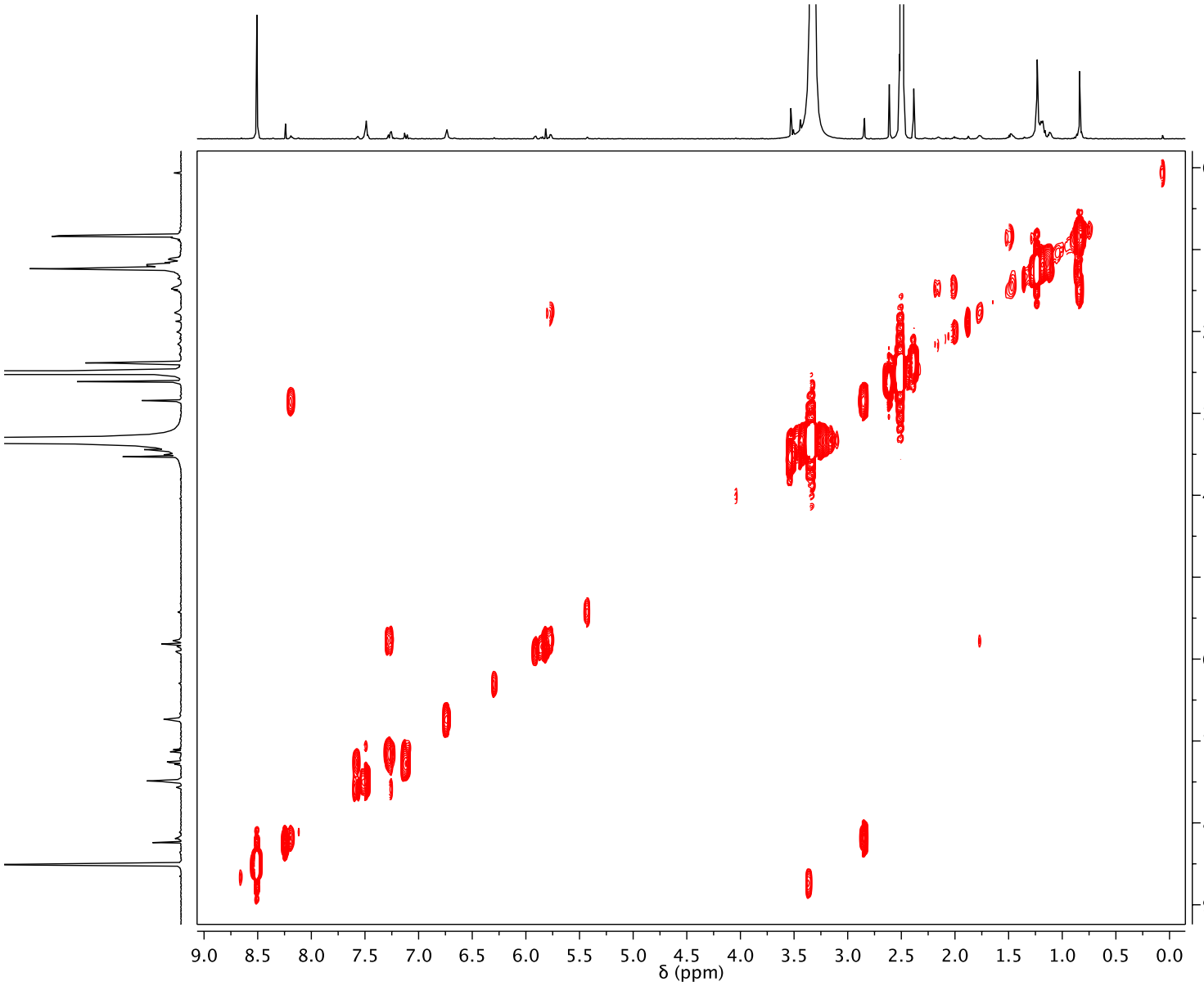

1H-NMR of quinocinnolinomycin A (S)-MTPA ester (5) at 600 MHz in DMSO-d6
(S)-MTPA
δ (ppm)
1.16
1.06
0.76
4.61
0.96
0.94
0.77
1.11
3.14
2.13
1.12
4.92
13.04
2.78
6.88
8.48
8.19
8.12
7.49
7.48
7.47
7.17
6.91
5.81
5.77
2.85
1.84
1.47
1.34
1.23
1.22
1.11
0.82

COSY of quinocinnolinomycin A (S)-MTPA ester (**5**) at 600 MHz in DMSO-d6

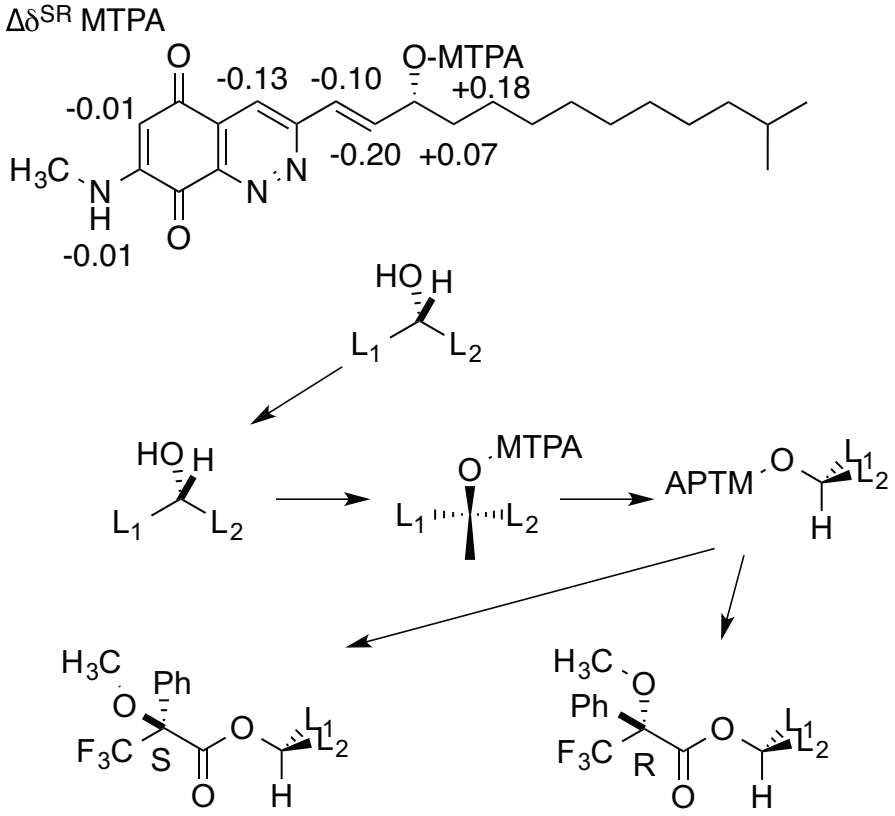



¹H-NMR of quinocinnolinomycin A (R)-MTPA ester (6) at 600 MHz in DMSO-d6
(R)-MTPA
δ (ppm)
0.78
1.05
0.58
4.96
1.93
0.91
0.79
1.46
3.00
2.16
3.28
37.47
3.19
7.79
8.51
8.24
8.19
7.57
7.50
7.50
7.49
7.48
7.48
7.27
7.12
5.82
5.77
2.85
1.77
1.48
1.23
1.19
1.11
0.83

COSY of quinocinnolinomycin A (R)-MTPA ester (**6**) at 600 MHz in DMSO-d6



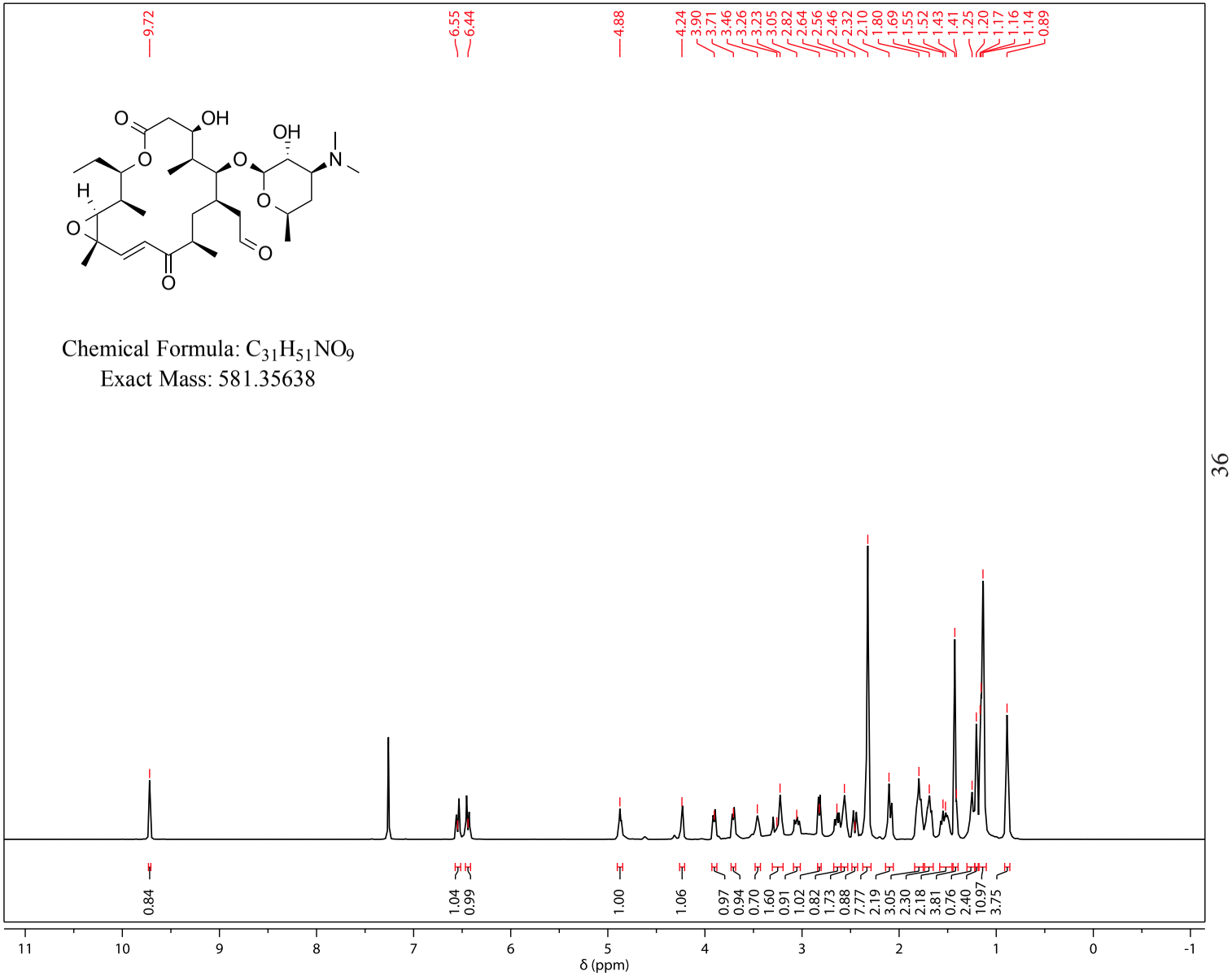
Supplemental Figure 12: Quinocinnolinomycin A (1) is displayed with the $\Delta\delta^{SR}$ values for the modified Mosher's ester method. Shielding from in the phenyl ring in the suggested major comformer displayed below causes the affected protons to be shifted upfield for that particular diastereomer.

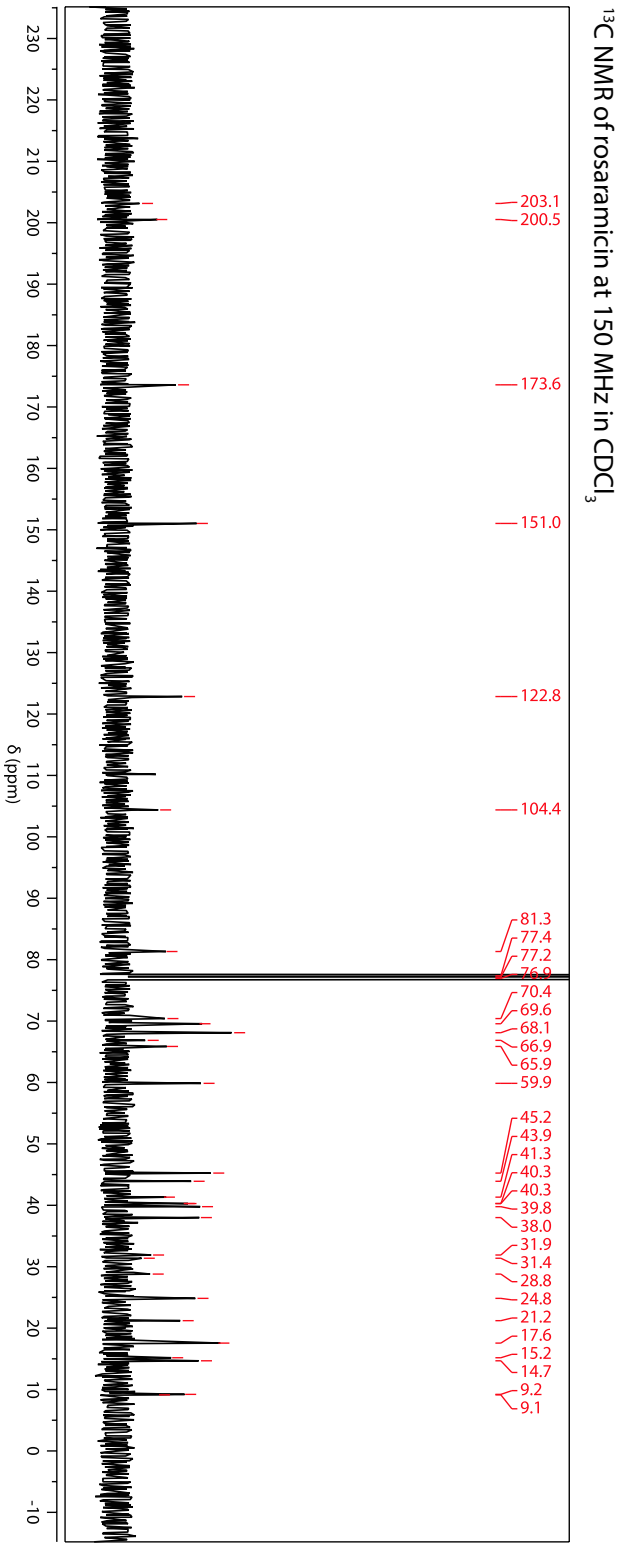


Supplemental Table 3: Tabulated ^{1}H NMR data for (**1**, **5**, and **6**). All spectra were acquired in DMSO-d6 at 600 MHz.

Position	(**1**) δ_H	S-Ester (**5**) δ_H	R-Ester (**6**) δ_H	$\Delta\delta^{SR} = \delta^S - \delta^R$
4	8.11	8.11	8.24	-0.13
4a				
5				
6	5.8	5.81	5.82	-0.01
7				
8				
8a				
9	7.01	6.91	7.12	-0.2
10	7.22	7.17	7.27	-0.1
11	4.27	5.77	5.77	0
12	1.54	1.84	1.77	0.07
13	1.37	1.37	1.19	0.18

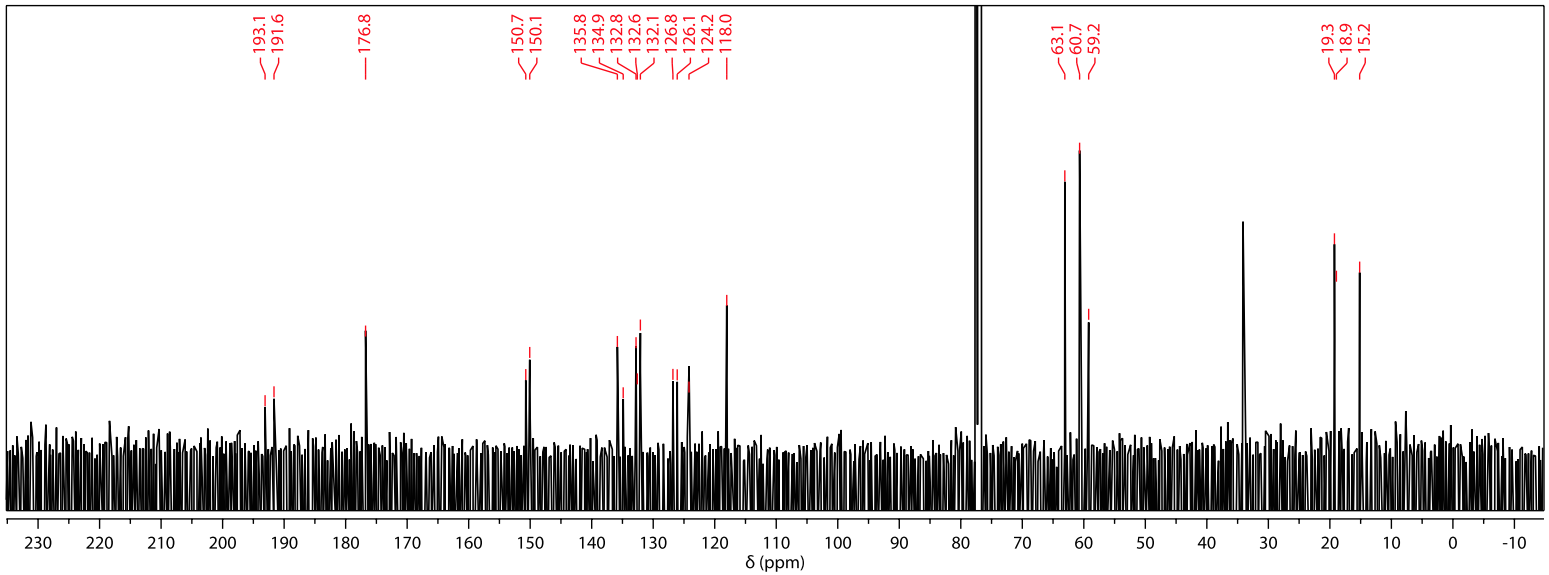
^{1}H NMR of rosaramicin at 600 MHz in $CDCl_3$





^{13}C NMR of rosaramicin at 150 MHz in $CDCl_3$

^{13}C NMR of fluostatin D at 150 MHz in $CDCl_3$



^{13}C NMR of fluostatin J at 150 MHz in $CDCl_3$