

## SUPPORTING INFORMATION

### Exploring the Limits of DNA Size: Naphtho-homologated DNA Bases and Pairs

Alex H. F. Lee and Eric T. Kool\*

Department of Chemistry, Stanford University, Stanford CA 94305-5080

\*to whom correspondence should be addressed: kool@stanford.edu

#### CONTENTS:

Synthesis of 5'-DMT-3'-phosphoramidite derivatives of <b>1</b> and <b>2</b> .....	p. S1
Confirmation of anomeric geometry.....	p. S5
Figure S1. Semiempirical calculations of yyDNA bases and tautomers.....	p. S6
Oligonucleotide mass spectrometric characterization data.....	p. S7
Table S1. Pairing of single yyDNA bases in natural duplex DNA.....	p. S8
Figure S2. Thermal denaturation studies of parallel and mismatched sequences...	p. S9
Figure S3. Stoichiometric mixing plots of control sequences.....	p. S10
Figure S4. CD spectra.....	p. S11
Figure S5. Structures of dabcyI-DNA conjugates.....	p. S12
NMR spectra of nucleoside synthetic intermediates.....	p. S13

## Preparation of phosphoramidite derivative of nucleoside 1

**1'- $\beta$ -[7-(benzoquinazoline2,4-dione)]-5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-D-ribofuranose.** The nucleoside **1** (0.29 g, 0.72 mmol) was coevaporated with distilled pyridine (3  $\times$  5 mL), and the residue was dissolved in distilled pyridine (10 mL) under argon. 4,4'-dimethoxytrityl chloride (0.40 g, 1.10 mmol) was added to the solution in a portion. The mixture was stirred at room temperature for 12 h. The volatiles were removed *in vacuo*, and the residue was purified by silica column chromatography (EtOAc to EtOAc/ methanol 10:1) to afford 5'-DMT-**1** (0.45 g, 90%) as slightly yellow foam. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 400 MHz):  $\delta$  8.54 (s, 1H), 8.04 (s, 1H), 7.86 (d, 1H, *J* = 7.2 Hz), 7.61 (d, 1H, *J* = 7.2 Hz), 7.52 (s, 1H), 7.41 (d, 2H, *J* = 7.2 Hz), 7.28 (m, 7H), 6.87 (d, 4H, *J* = 8.8 Hz), 5.19 (dd, 1H, *J* = 10.4, 5.2 Hz), 4.20 (m, 1H), 3.99 (bs, 1H), 3.71 (s, 6H), 3.16 (m, 2H), 2.20 (m, 1H), 1.99 (m, 1H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 100 MHz):  $\delta$  158.05, 150.36, 145.00, 139.07, 136.49, 135.86, 135.70, 129.78, 128.93, 128.68, 128.16, 127.84, 126.92, 125.83, 113.17, 112.78, 110.29, 94.00, 85.40, 79.24, 72.56, 55.02. HRFABMS *m/z*: [M]<sup>+</sup> calcd for C<sub>38</sub>H<sub>34</sub>N<sub>2</sub>O<sub>7</sub>, 630.2366 found 630.2365.

**1'- $\beta$ -[7-(benzoquinazoline2,4-dione)]-5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-D-ribofuranose-3'-(2-cyanoethyl-N-diisopropylphosphoramidite).** To a solution of 5'-DMT-protected nucleoside **1** (0.43 g, 0.63 mmol) and *N, N*-diisopropylammonium tetrazolide (106 mg, 0.62 mmol) in freshly distilled CH<sub>2</sub>Cl<sub>2</sub> (15 mL) was added 2-cyanoethyl tetraisopropylphosphoramidite (0.30 mL, 0.98 mmol) in a portion. The reaction mixture was allowed to stir at room temperature for 6 h. The volatiles were removed *in vacuo* and the residue

was purified by silica column chromatography (hexanes/ EtOAc 1:1 initially, EtOAc subsequently) to yield 5'-DMT-3'-phosphoramidite **1** (0.45 g, 86%) as white foam containing two inseparable diastereomers due to the chiral center at phosphorus.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz):  $\delta$  8.67 (d, 1H,  $J = 4.4$  Hz), 7.94 (s, 1H), 7.75 (m, 1H), 7.65 (m, 1H), 7.50 (m, 3H), 7.38 (m, 4H), 7.20 (m, 3H), 6.83 (m, 4H), 5.32 (dd, 1H,  $J = 5.6, 4.8$  Hz) 4.59 (m, 1H), 4.31 (m, 1H), 3.85 (m, 6H), 3.60 (m, 2H), 3.36 (m, 2H), 2.60 (t, 1H,  $J = 6.4$  Hz), 2.49 (t, 1H,  $J = 6.4$  Hz), 2.16 (m, 1H), 1.20 (m, 9H), 1.10 (m, 3H).  $^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 161 MHz):  $\delta$  149.15, 149.07.  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta$  163.29, 158.32, 152.04, 144.77, 138.88, 136.45, 135.87, 135.09, 130.03, 129.76, 128.16, 128.12, 127.71, 127.12, 126.73, 126.13, 117.64, 117.52, 114.68, 112.99, 111.51, 86.04, 86.01, 85.77, 79.98, 77.32, 76.11, 75.52, 64.12, 60.31, 58.30, 58.11, 55.10, 45.46, 43.13, 43.08, 42.96, 25.13, 24.56, 24.49, 24.41, 24.35, 23.39, 22.73, 21.31, 20.96, 20.31, 20.24, 20.13, 20.07, 14.08. HRFABMS  $m/z$ :  $[\text{M}]^+$  calcd for  $\text{C}_{47}\text{H}_{47}\text{N}_4\text{O}_8\text{P}$ , 826.3120 found 826.3101.

### Preparation of phosphoramidite derivative of nucleoside 2

**1'- $\beta$ -{7-[N-(4-oxo-1,4-dihydrobenzoquinazolin-2-yl)]-isobutyramide}-5'-O-(4,4'-dimethoxytrityl)-2'-deoxyribofuranose.** The *N*-isobutyryl-nucleoside **16** (0.56 g, 1.41 mmol) was coevaporated with distilled pyridine ( $3 \times 5$  mL), and the residue was dissolved in distilled pyridine (12 mL) under argon. 4,4'-dimethoxytrityl chloride (0.60 g, 1.70 mmol) was added to the solution in a portion. The mixture was stirred at room temperature for 12 h. The volatiles were removed *in vacuo*, and the residue was purified by silica column chromatography (hexanes/ EtOAc 1:2 initially, EtOAc subsequently) to afford 5'-DMT-**16** (0.76 g, 78%) as a white foam.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz):  $\delta$  8.79 (s, 1H), 7.96 (s, 1H), 7.85 (d, 2H,  $J = 8.4$  Hz), 7.61 (d, 1H,  $J = 8.8$  Hz), 7.47 (d, 2H,  $J = 6.8$  Hz), 7.35 (d, 4H,  $J = 9.2$  Hz), 7.28 (m, 3H), 6.82 (d, 4H,  $J = 8.8$  Hz), 5.36 (dd, 1H,  $J = 5.6, 4.4$  Hz), 4.50 (m, 1H), 4.16 (m, 1H), 3.77 (s, 6H), 3.41 (m, 1H), 3.33

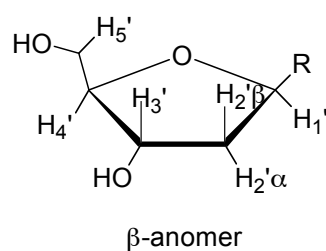
(m, 1H), 2.66 (bs, 1H), 2.35 (m, 1H), 2.15 (m, 1H), 1.26 (d, 6H,  $J = 7.2$  Hz).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta$  171.22, 158.38, 158.28, 149.19, 144.67, 139.52, 139.44, 136.33, 135.81, 135.78, 130.29, 129.96, 129.93, 129.07, 128.41, 128.05, 127.68, 127.65, 127.58, 127.29, 126.86, 126.66, 125.83, 123.83, 112.96, 112.94, 86.57, 86.09, 81.31, 79.79, 74.28, 64.45, 60.32, 55.02, 43.57, 36.23, 20.92, 18.89, 18.88, 14.03. HRFABMS  $m/z$ :  $[\text{M}]^+$  calcd for  $\text{C}_{42}\text{H}_{41}\text{N}_3\text{O}_7$ , 700.3017 found 700.2991.

**1'- $\beta$ -{7-[*N*-(4-oxo-1,4-dihydrobenzoquinazolin-2-yl)]-isobutyramide}-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyribofuranose-3'-(2-cyanoethyl-*N*-diisopropylphosphoramidite).**

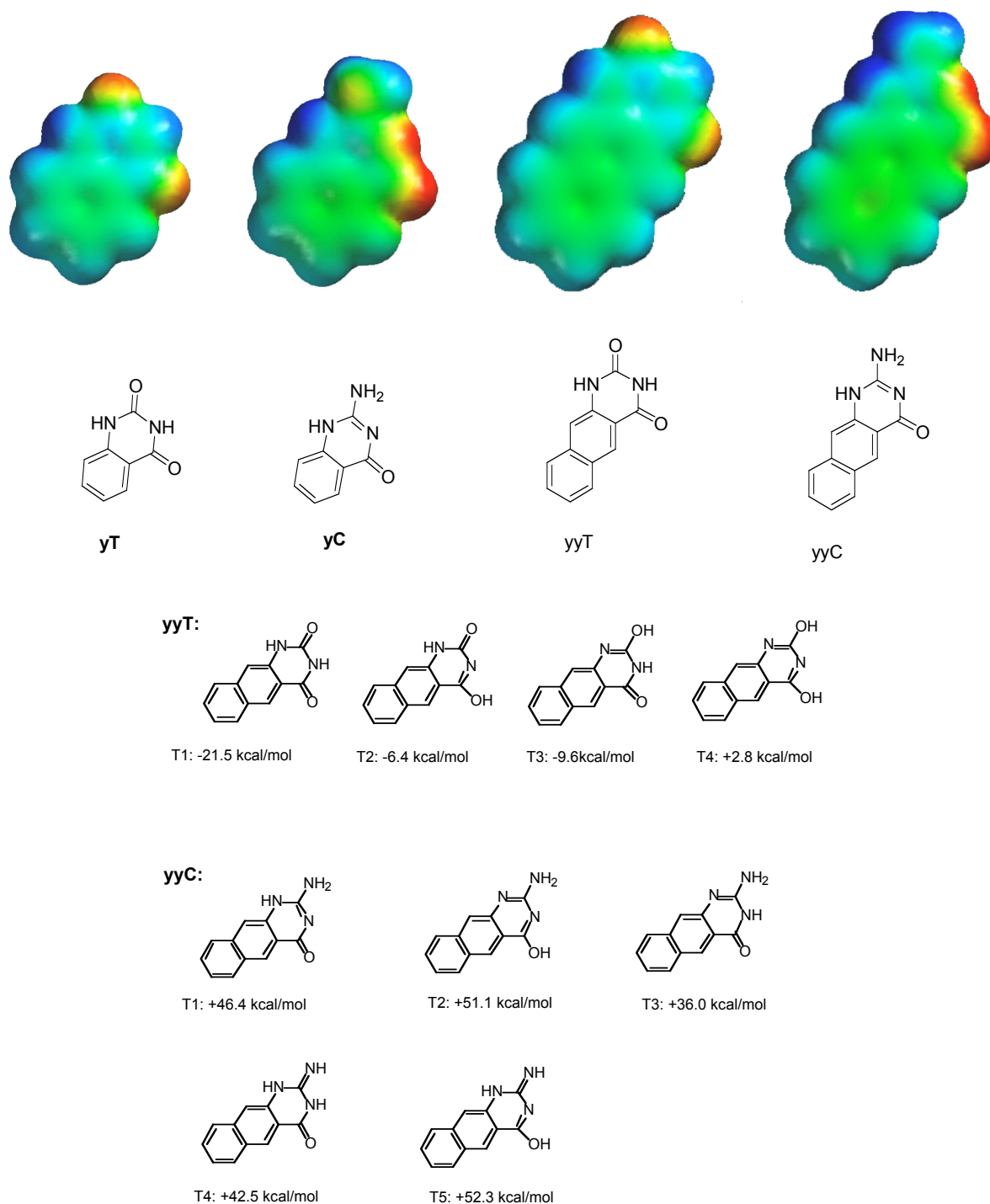
To a solution of 5'-DMT-protected nucleoside **16** (0.50 g, 0.72 mmol) and *N,N*-diisopropylammonium tetrazolide (140 mg, 0.80 mmol) in freshly distilled  $\text{CH}_2\text{Cl}_2$  (20 mL) was added 2-cyanoethyl tetraisopropylphosphoramidite (0.35 mL, 1.20 mmol) in one portion. The reaction mixture was allowed to stir at room temperature for 3.5 h. The volatiles were removed *in vacuo* and the residue was purified by silica column chromatography (hexanes/ EtOAc 1:1 initially, EtOAc subsequently) to yield 5'-DMT-3'-phosphoramidite **16** (0.45 g, 70%) as a white foam containing two inseparable diastereomers due to the chiral center at phosphorus.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz):  $\delta$  8.81 (s, 1H), 8.03 (d, 1H,  $J = 3.6$  Hz), 7.84 (dd, 2H,  $J = 12.8, 2.4$  Hz), 7.67 (d, 1H,  $J = 8.4$  Hz), 7.50 (m, 2H), 7.37 (m, 4H), 7.29 (m, 2H), 7.20 (m, 1H), 6.82 (m, 4H), 5.36 (m, 1H), 4.60 (m, 1H), 4.34 (m, 1H), 3.89 (m, 1H), 3.74 (m, 6H), 3.60 (m, 2H), 3.37 (m, 2H), 2.64 (t, 1H,  $J = 6.4$  Hz), 2.49 (t, 1H,  $J = 6.4$  Hz), 2.18 (m, 1H), 1.21 (m, 18H).  $^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 161 MHz):  $\delta$  149.12, 149.02.  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta$  158.36, 158.35, 144.75, 144.73, 135.88, 135.87, 135.86, 130.07, 130.03, 130.00, 128.17, 128.13, 127.71, 126.72, 126.69, 126.02, 117.55, 117.47, 112.98, 86.18, 86.07, 86.05, 85.85, 85.80, 80.12, 80.04, 75.63, 64.12,

58.31, 58.29, 58.12, 58.10, 55.12, 55.11, 45.49, 43.16, 43.12, 43.03, 42.99, 36.33, 24.58, 24.57, 24.49, 24.35, 23.40, 22.67, 21.37, 20.36, 20.26, 20.16, 20.93, 18.99. HRFABMS  $m/z$ :  $[M]^+$  calcd for  $C_{51}H_{58}N_5O_8P$  899.4009, found 899.3990.

### Confirmation of anomeric geometry



yyT <b>8</b> Irradiation at :	H1' (5.22 ppm)	NOE observed at H4' (4.01 ppm), H2'α (2.29 ppm)
yyC <b>16</b> Irradiation at:	H1' (5.18 ppm)	NOE observed at H4' (3.85 ppm), H2'α (2.15 ppm)



**Figure S1.** (top) Space-filling models of the yyT, yyC, yT and yC nucleobases with electrostatic potentials mapped on the surfaces. Natural T and C bases are shown for comparison. Calculated using AM1 with Spartan (Wavefunction, Inc.). (below) AM1-Calculated heats of formation in the gas phase (kcal/mol) for several tautomers of yyT and yyC.

**Oligonucleotide characterization.** Synthesized oligonucleotides containing modified deoxyribosides were characterized MALDI-TOF mass spectrometry. Data are given below.

Mass spectrometry data for oligonucleotides containing yyT and yyC deoxyribosides.

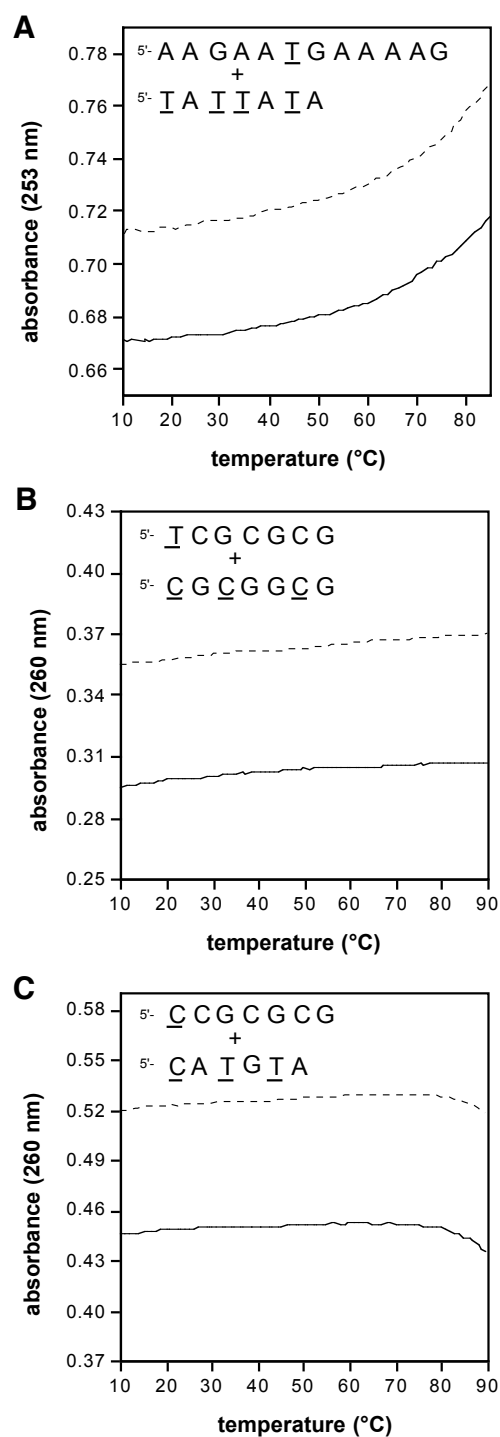
Oligonucleotide	Average MW calcd.	MW found
5'-yyTCGCGCG-3'	2183	2183
5'-yyCCGCGCG-3'	2183	2183
5'-CTTTTCyyTTTCTT-3'	3630	3631
5'-AAGAAyyTGAAAAG-3'	3822	3823
5'-CTTTTCyyCTTCTT-3'	3629	3631
5'-AAGAAyyCGAAAAG-3'	3821	3825
5'-yyCyyCyyTGAYyCG-3'	2468	2472
5'-yyCGyyTyyCAGG-3'	2408	2408
5'-yyCyyCGyyCGG-3'	2095	2098
5'-yyTAYyTyyTAYyTA-3'	2242	2238
5'-yyTAYyTAAyyTA-3'	2366	2369
5'-yyCGyyCyyCGyyCG-3'	2488	2490
5'-yyCGyyCGGyyCG-3'	2425	2429
5'-yyTAYyCAYyTG-3'	2064	2065
5'-yyCAYyTGyyTA-3'	2064	2064
5'-GyyTAYyTAYyTA-3'	2378	2379
5'-AYyTAYyTAYyTG-3'	2378	2379
5'-yyCyyCyyTyyTyyCyyTyyCyyC-3'	3055	3057

**Table S1.** Thermal melting data for 12mer DNA duplexes containing a size-expanded base pair or mismatch (X-Y) in a central position.<sup>a,b,e</sup>

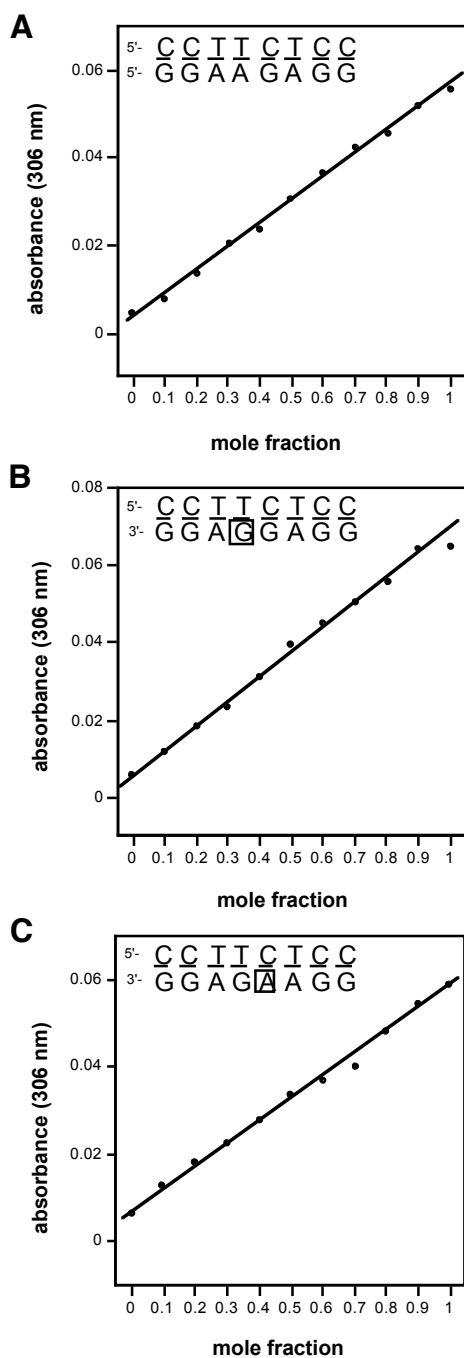
base pair <b>X - Y</b>	$T_m^c$ (°C)	$\Delta G_{37}^\circ^d$ (kcal/mol)	base pair <b>X - Y</b>	$T_m^c$ (°C)	$\Delta G_{37}^\circ^d$ (kcal/mol)
A - T	40.4 ± 0.5	-9.4 ± 0.1	T - A	41.2 ± 0.5	-9.4 ± 0.1
A - yyT	33.9 ± 0.5	-7.7 ± 0.1	yyT - A	34.3 ± 0.5	-7.8 ± 0.1
G - yyT	36.3 ± 0.5	-8.0 ± 0.2	yyT - T	30.8 ± 0.5	-7.1 ± 0.1
T - yyT	25.3 ± 0.5	-7.7 ± 0.3	yyT - G	34.4 ± 0.5	-7.7 ± 0.1
C - yyT	26.3 ± 0.5	-7.9 ± 0.2	yyT - C	32.6 ± 0.5	-7.4 ± 0.1
G - C	46.5 ± 0.5	-11.1 ± 0.4	C - G	43.0 ± 0.5	-9.7 ± 0.1
A - yyC	32.5 ± 0.5	-7.3 ± 0.1	yyC - G	30.8 ± 0.5	-7.0 ± 0.3
G - yyC	33.3 ± 0.5	-7.4 ± 0.2	yyC - C	29.7 ± 0.5	-7.1 ± 0.2
C - yyC	33.1 ± 0.5	-7.9 ± 0.2	yyC - A	29.8 ± 0.5	-6.9 ± 0.3
T - yyC	28.2 ± 0.5	-7.8 ± 0.2	yyC - T	27.8 ± 0.5	-6.5 ± 0.4
ϕ - T	21.0 ± 0.5	nd	T - ϕ	22.4 ± 0.5	nd
ϕ - C	20.6 ± 0.5	nd	C - ϕ	23.5 ± 0.5	nd
ϕ - yyT	31.1 ± 0.5	-7.7 ± 0.4	yyT - ϕ	32.7 ± 0.5	-7.7 ± 0.2
ϕ - yyC	35.0 ± 0.5	-8.6 ± 0.2	yyC - ϕ	30.9 ± 0.5	-7.2 ± 0.1

<sup>a</sup> Conditions: 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM PIPES•Na (pH 7.0).<sup>b</sup> Sequence is 5'-AAGAAXGAAAAG • 5'-CTTTTCYTTCTT.<sup>c</sup>  $T_m$  values are for 5.0 μM oligonucleotide.<sup>d</sup> Averages of values from van't Hoff and curve fitting methods.<sup>e</sup> "ϕ" is tetrahydrofuran abasic analogue.

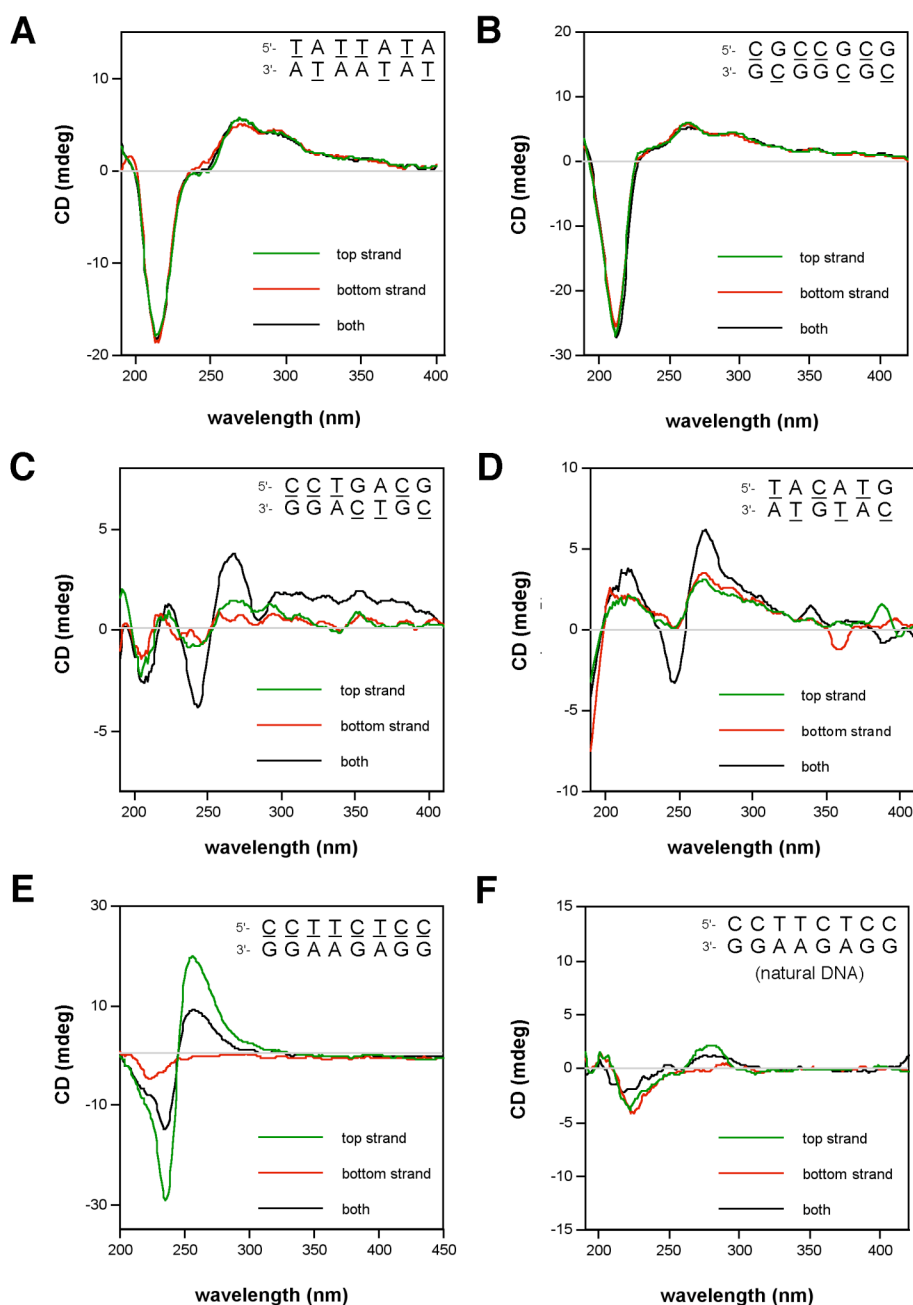




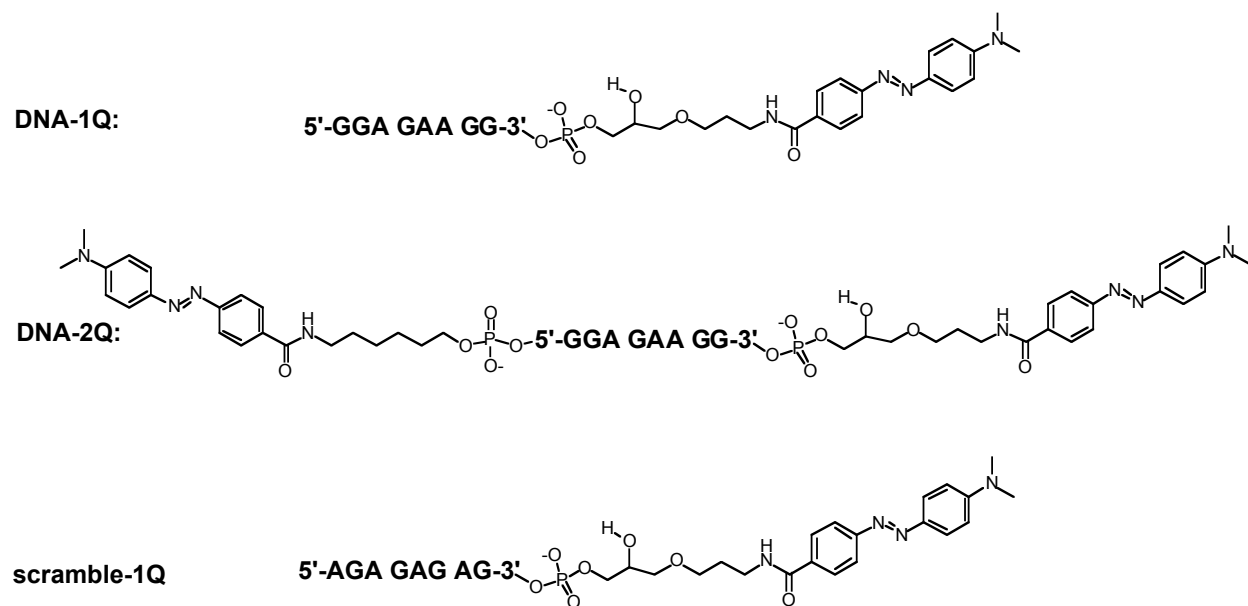
**Figure S2.** Thermal melting data for three noncomplementary sequences (A-C), showing little or no difference in curve shape between oligonucleotide mixtures (solid lines) and mathematical addition of data for single strands alone (dashed lines). Sequences are as shown in each plot, with yyDNA bases underlined. Compare to Figure 3, main text.



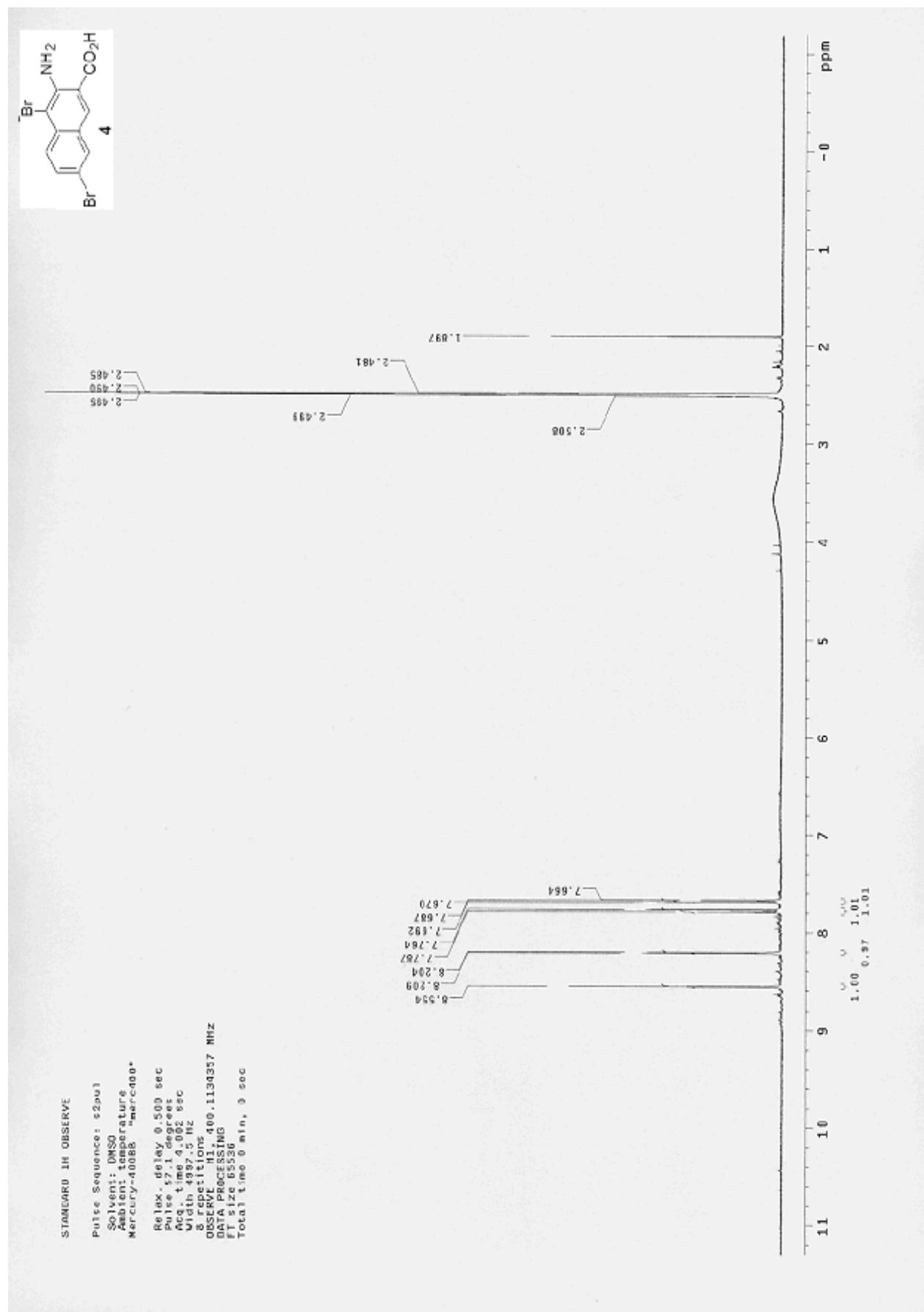
**Figure S3.** Mixing plots for three different yDNA sequences, testing strand orientation and sequence selectivity. All three cases are best fit by a single line, suggesting no complexation under the experimental conditions. (A) Parallel-oriented strands; (B) Strands containing a yT – G mismatch; (C) Strands containing a yC – A mismatch. The mixing plot for the antiparallel, fully complementary case is shown in Fig. 4A (main text). See Table 3 (main text) for all sequences. Doublewide nucleotides are underlined.

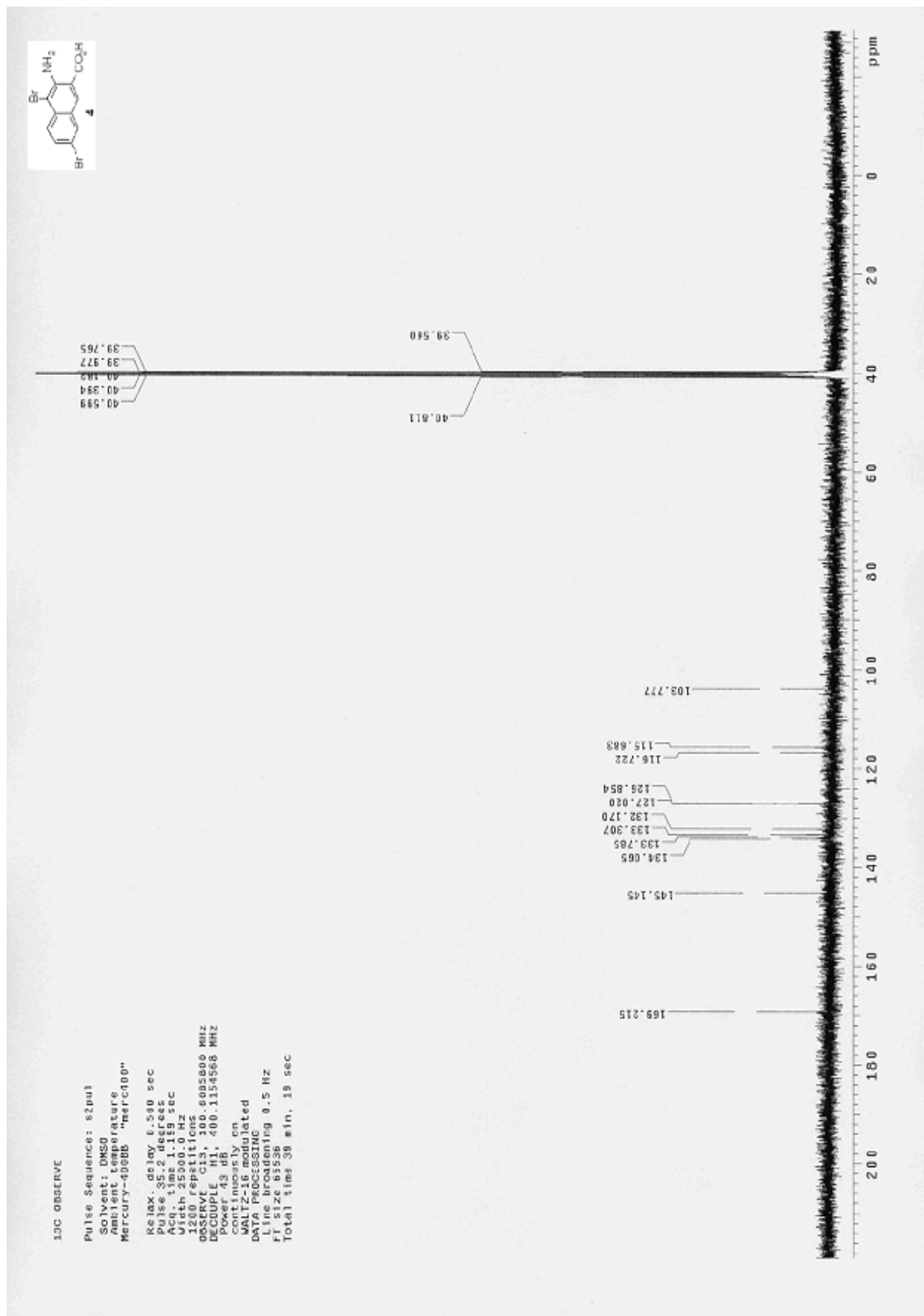


**Figure S4.** CD spectra of yyDNA strands and putative duplexes. Sequences are as shown; yyDNA nucleotides are underlined. (A,B) related sequences containing only one type of yyDNA base; (C,D) mixed sequences containing four different base pairs; (E) sequence in which all yyDNA bases are segregated to one strand; (F) all-DNA control for (E). Conditions: 2  $\mu$ M [DNA]; 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM Na•PIPES (pH 7.0).

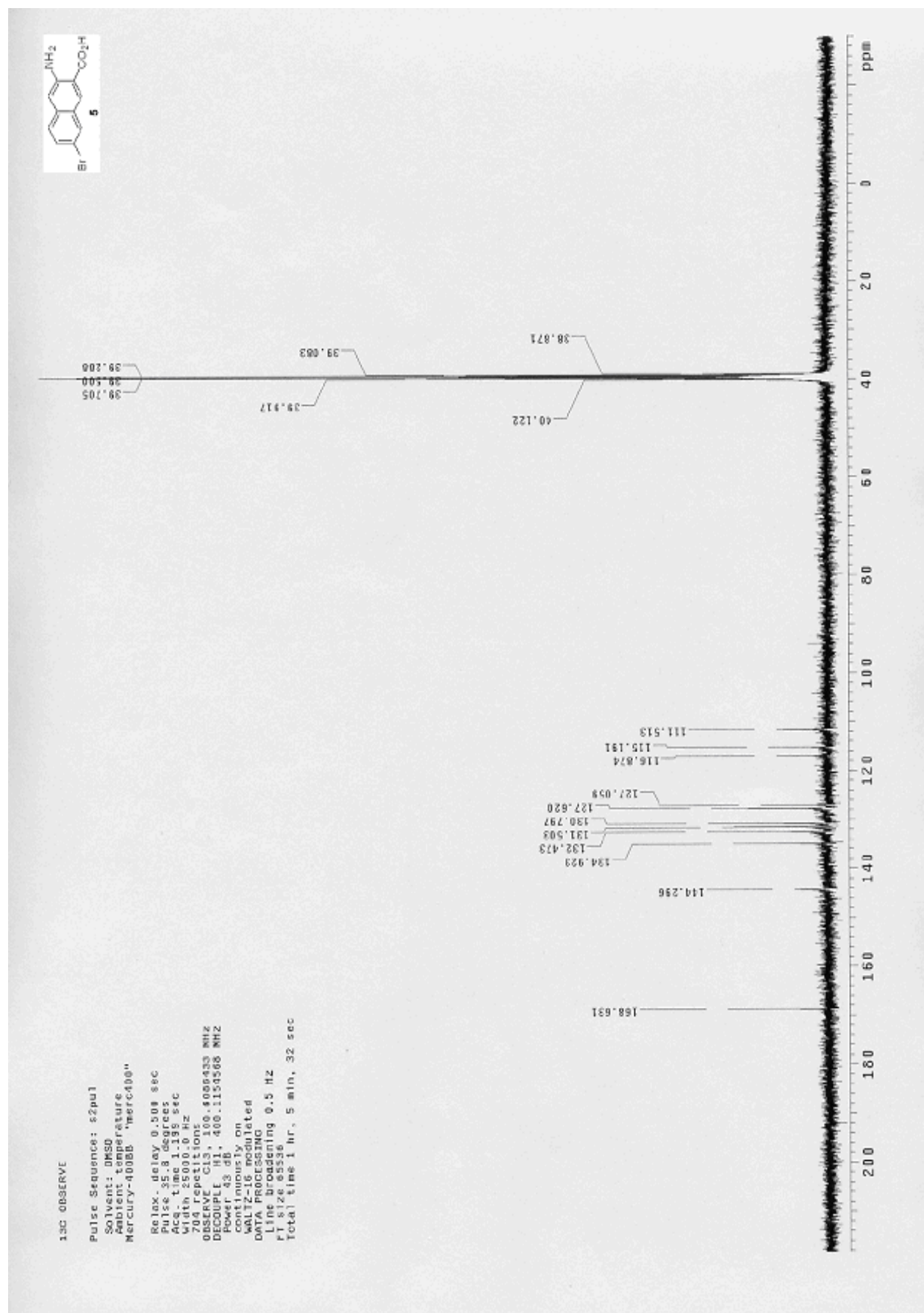


**Figure S5.** Structures of quencher-oligonucleotide conjugates used in experiments to test hybridization and quenching of a yyDNA strand. See Fig. 6 (main text) for fluorescence data. 5' and 3' Dabcyl conjugates were added during oligonucleotide synthesis, using commercial reagents (Glen Research).

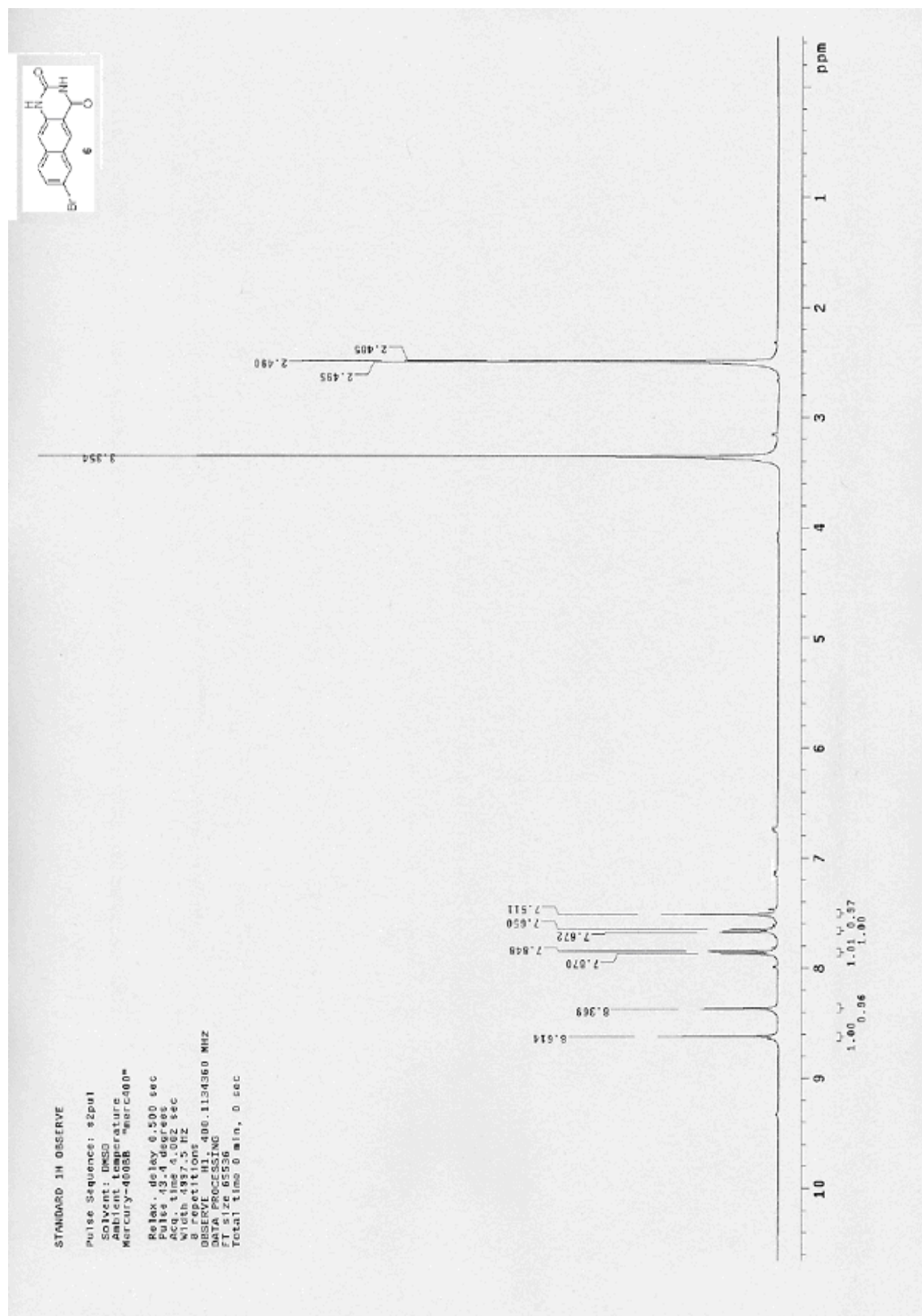


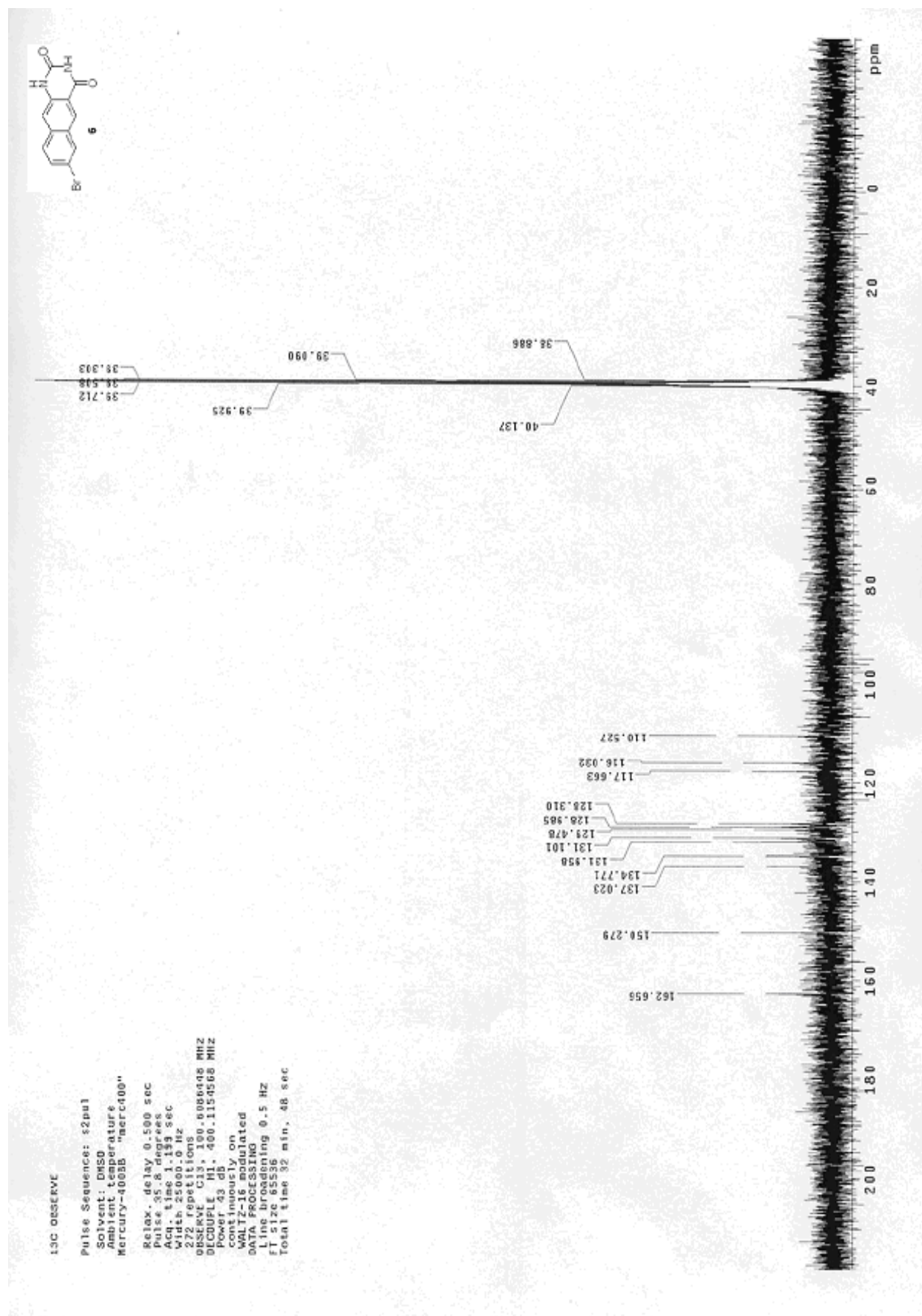


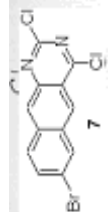






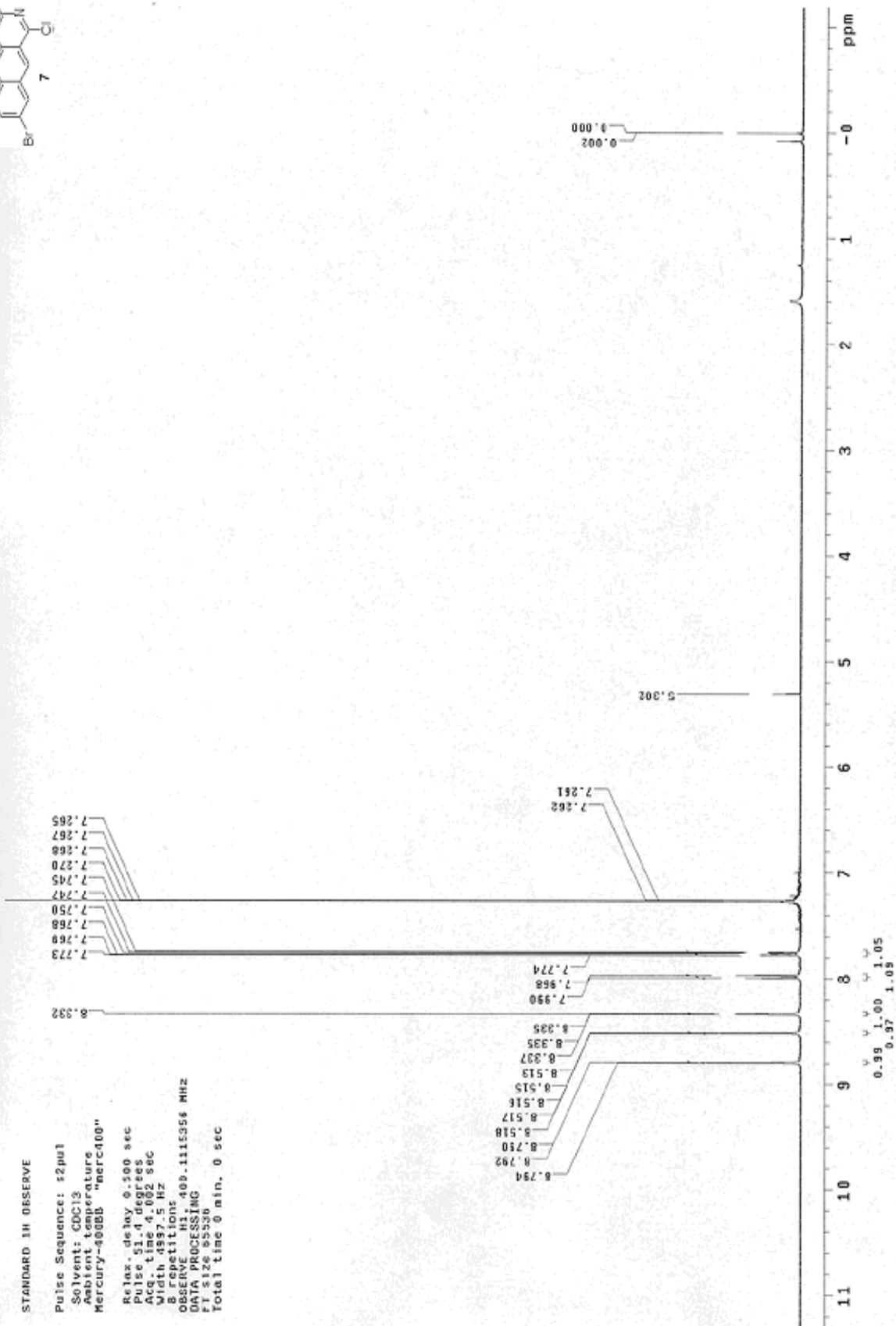


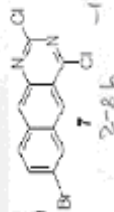




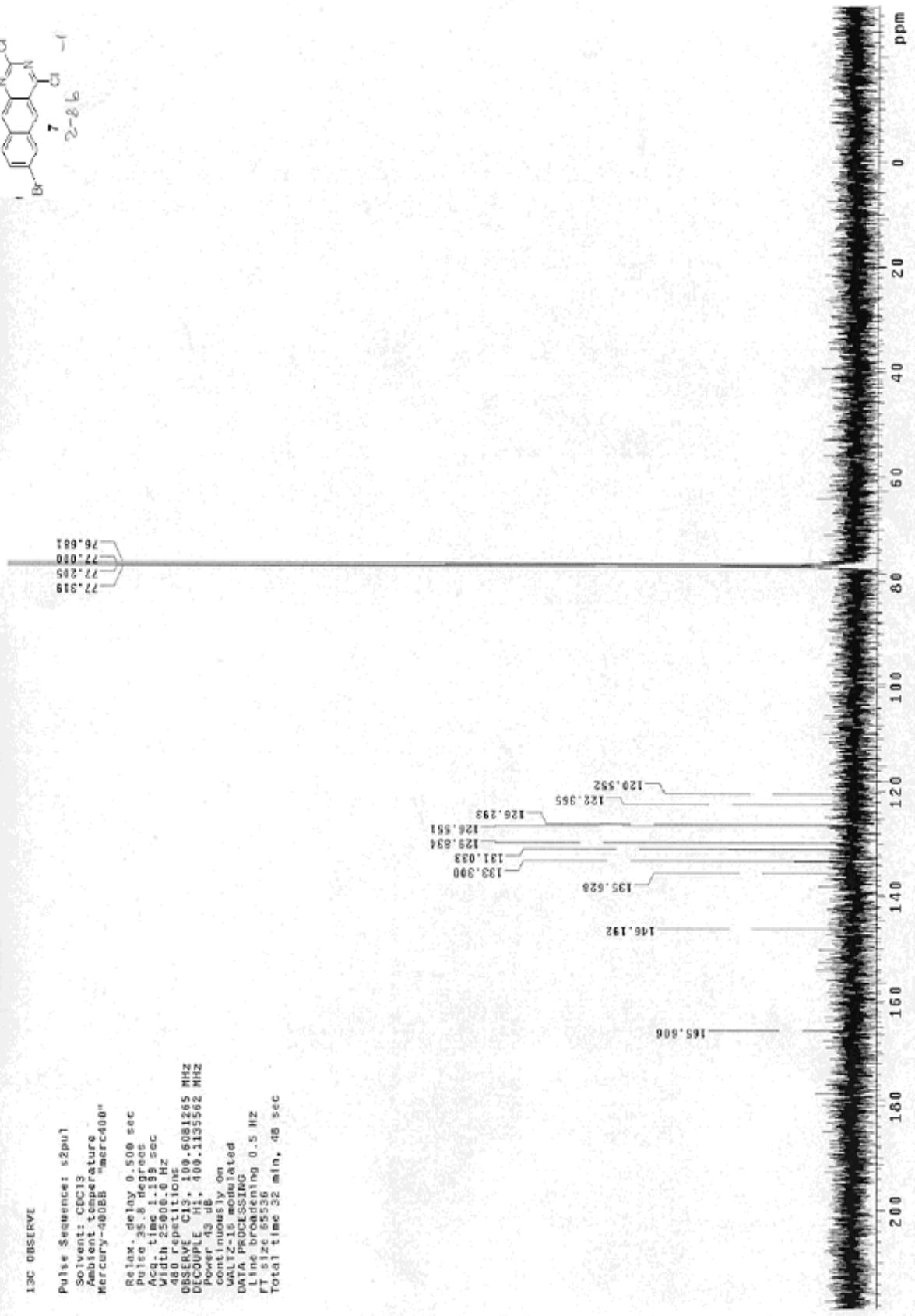
STANDARD IN OBSERVE

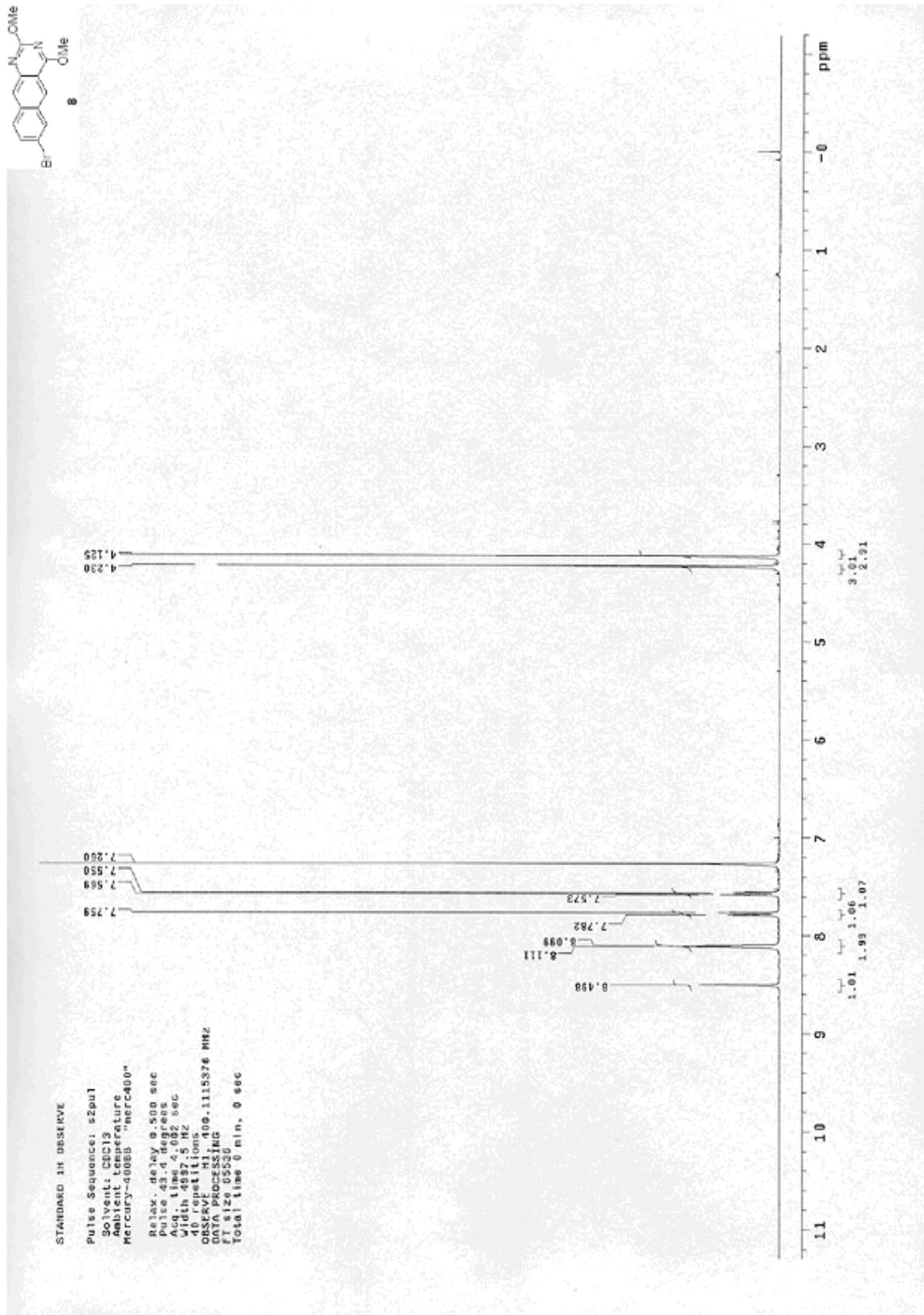
Pulse Sequence: s2pul  
 Solvent: CDC13  
 Ambient temperature  
 Mercury-400BB "merc100"  
 Relax delay 0.500 sec  
 Pulse 51.4 degrees  
 Acq. time 4.062 sec  
 Width 4997.5 Hz  
 8 repetitions  
 OBSERVE H1 400.1115356 MHz  
 DATA PROCESSING  
 FI size 85536  
 Total time 0 min. 0 sec





13C OBSERVE  
 Pulse Sequence: s2pu1  
 Solvent: CDCl3  
 Ambient Temperature  
 Mercury-400BB "merc400"  
 Relax. delay 0.500 sec  
 Pulse 35.0 mcg/sec  
 Width 25006.8 Hz  
 480 repetitions  
 OBSERVE C13 100.6081265 MHZ  
 DECOUPLE H1 400.1135562 MHZ  
 Power 43 dB  
 continuously on  
 MALTZ-13 modulated  
 DATA PROCESSING  
 F1 136.0000000 MHz  
 Total time 32 min, 48 sec

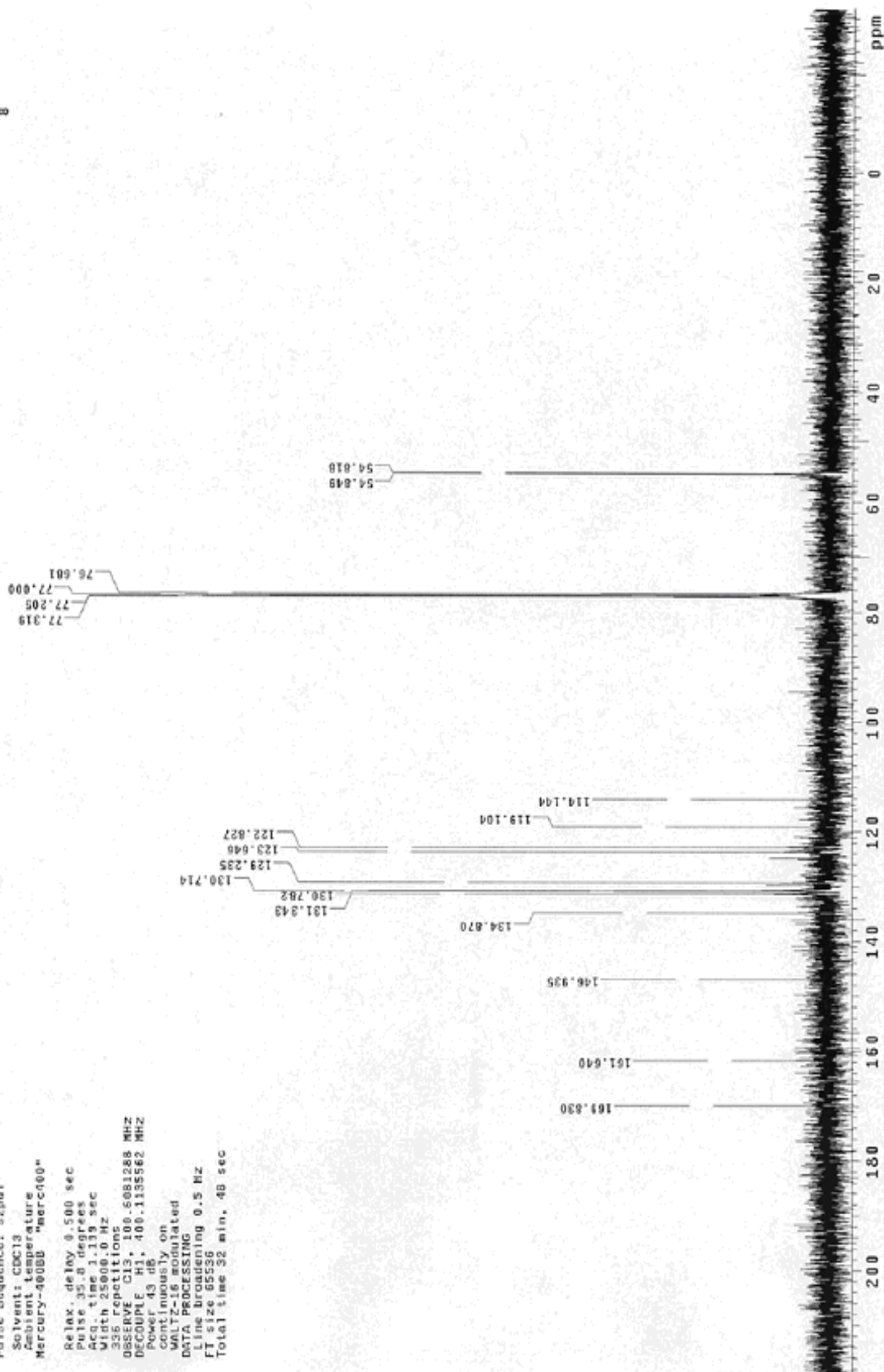


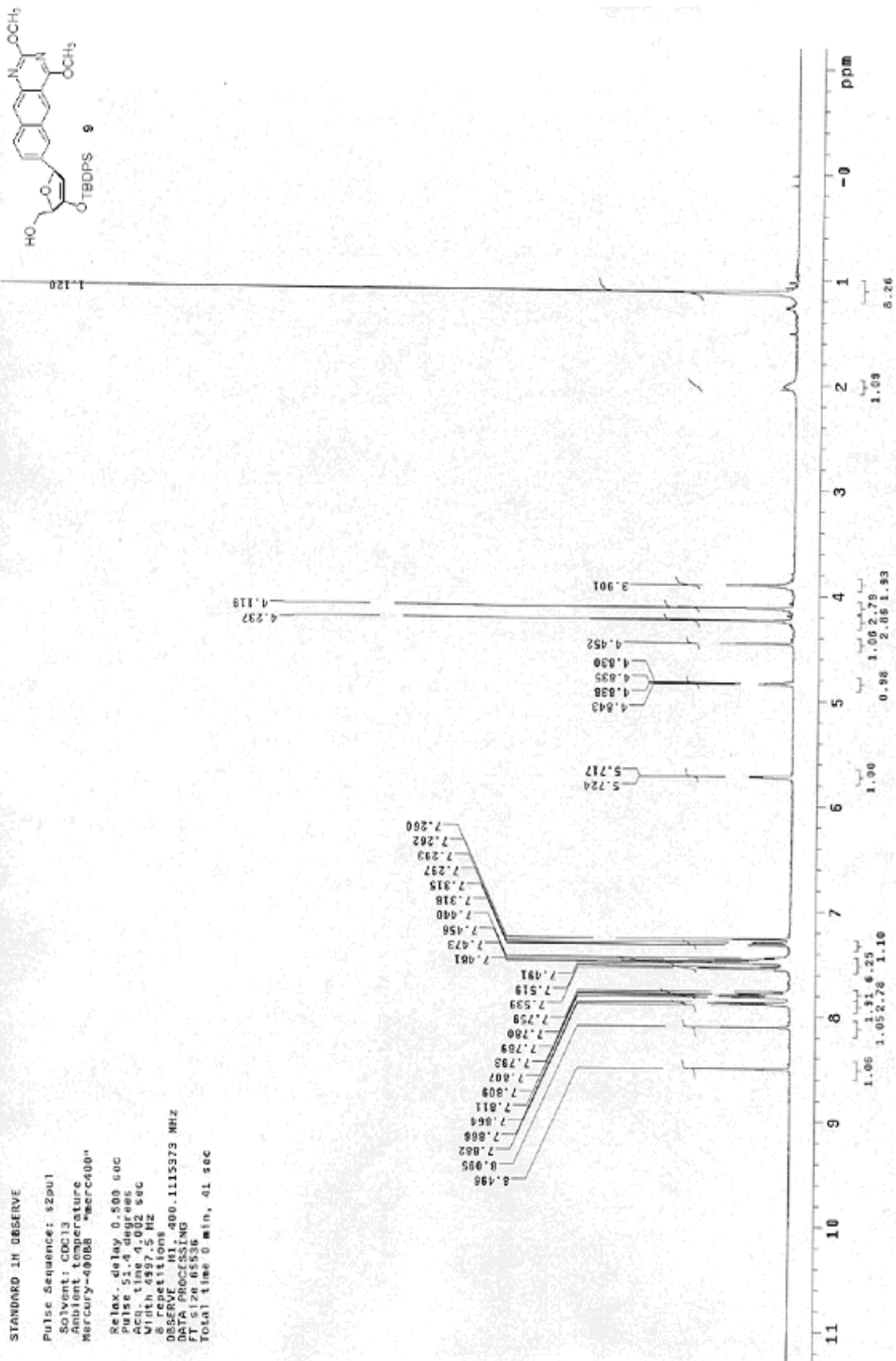


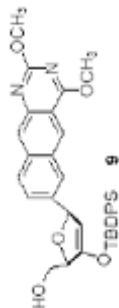


13C OBSERVE

Pulse Sequence: s2pul  
 Solvent: CDCl3  
 Ambient temperature  
 Mercury-400BB "merc100"  
 Relax delay 0.500 sec  
 pulse 35.0 degrees  
 Acq. time 1.119 sec  
 Width 25000.0 Hz  
 356 repetitions  
 OBSERVE C13, 100.6081288 MHz  
 DECOUPLE H1, 400.1155562 MHz  
 Power 43 dB  
 continuously on  
 waltz16c controlled  
 DMLT PROCESSED  
 Line Broadening  
 FI size 65536 ng 0.5 N2  
 Total time 32 min, 48 sec

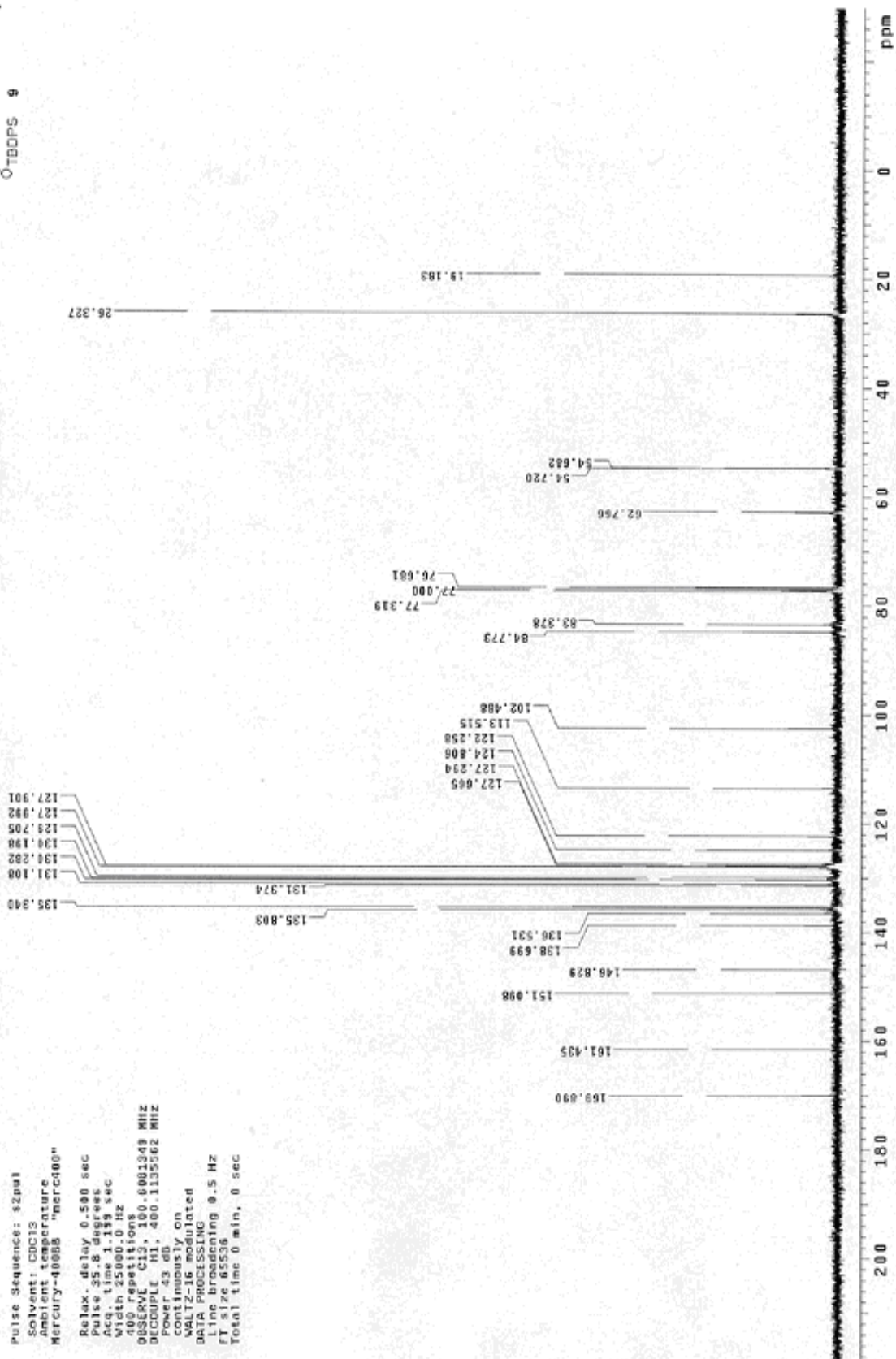




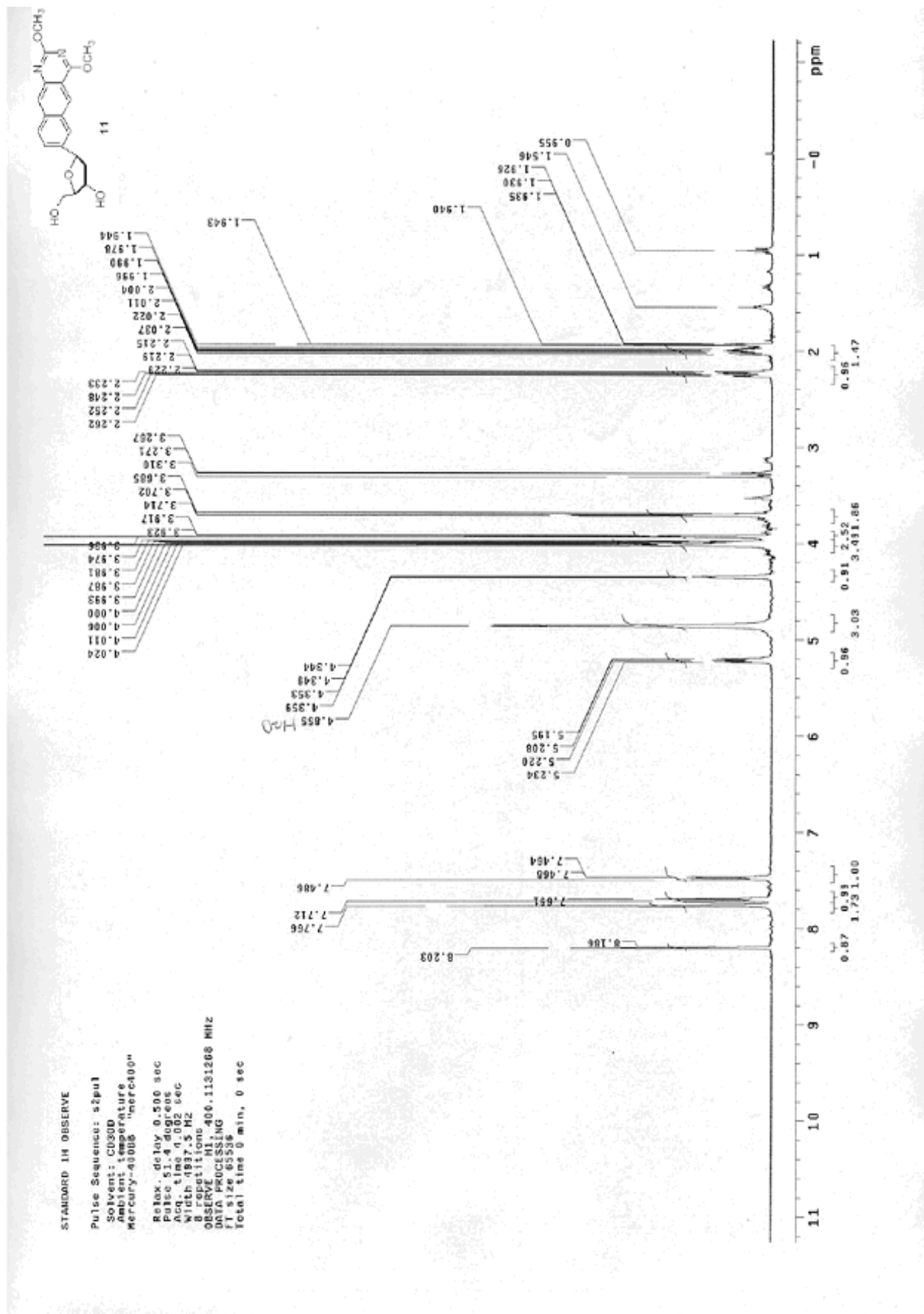


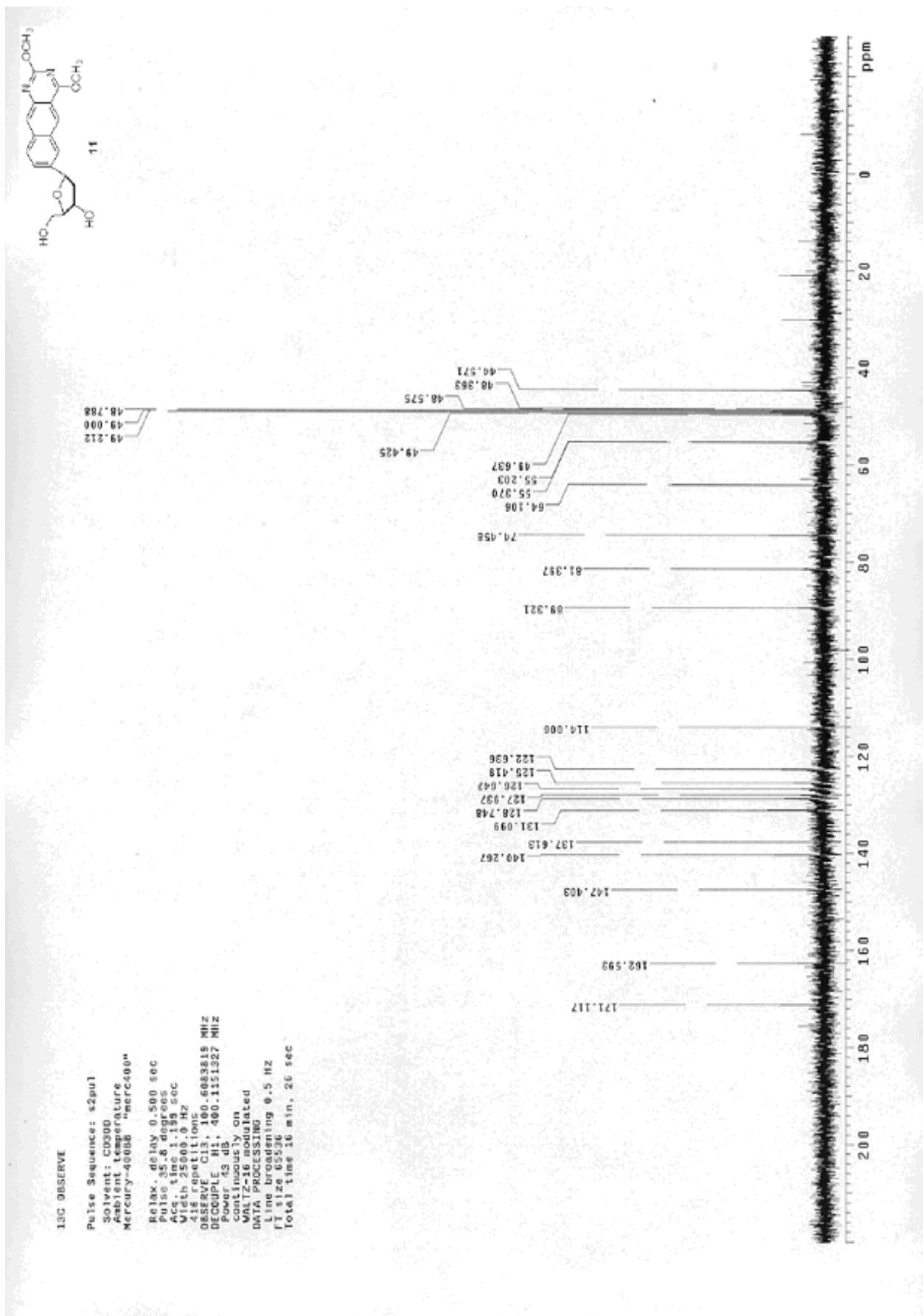
13C OBSERVE

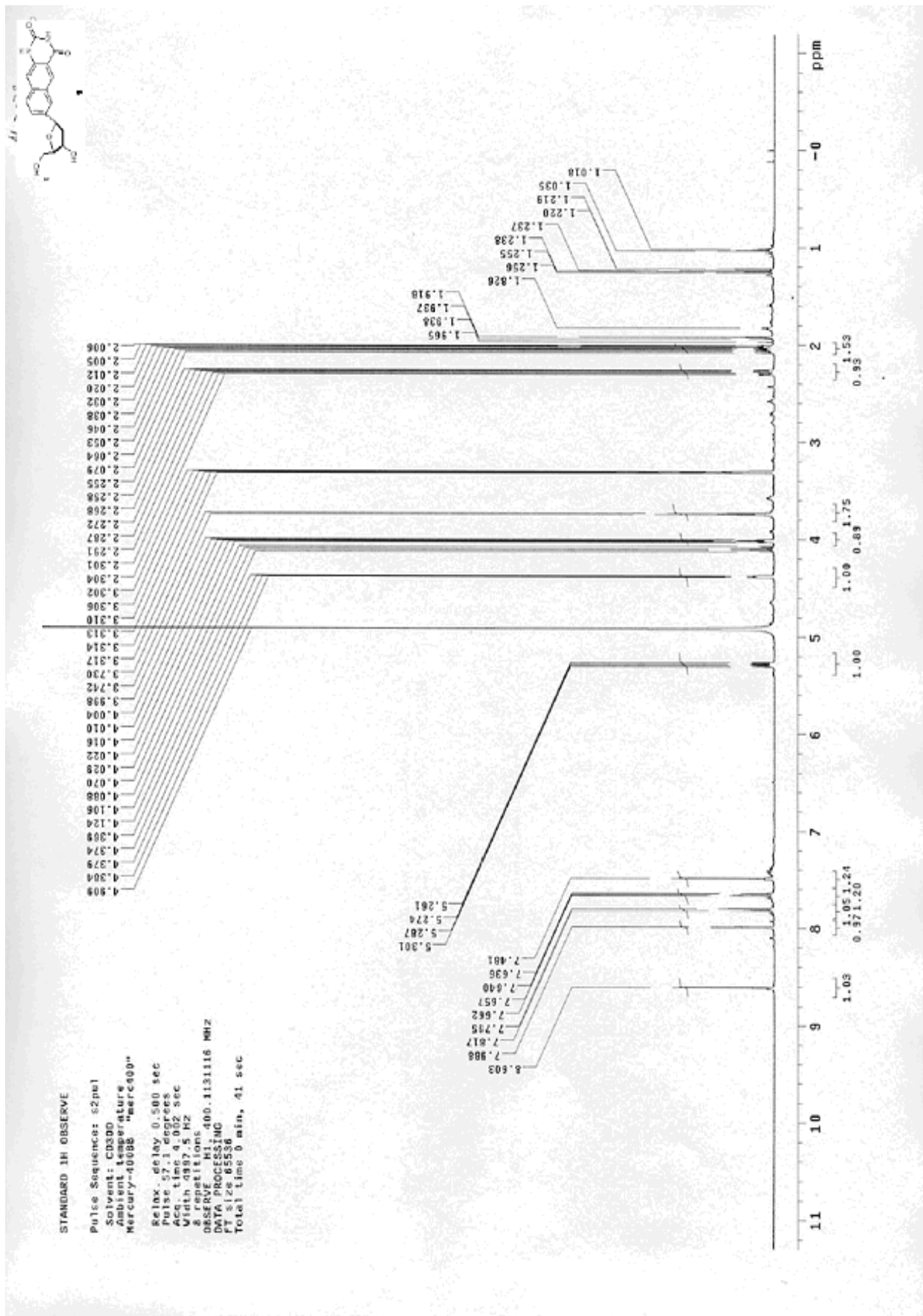
Pulse Sequence: s2pul  
 Solvent: CDCl3  
 Ambient temperature  
 Mercury-400MB "merc400"  
 Relax. delay 0.500 sec  
 Pulse 35.8 degrees  
 Acq. time 1.139 sec  
 Width 45000.0 Hz  
 400 repetitions  
 OBSERVE CH1, 100.6003049 MHz  
 OBSERVE CH2, 400.1135562 MHz  
 Power 43 dB, continuously on  
 WALTZ-16 modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FT size 65536  
 Total time 0 min, 0 sec

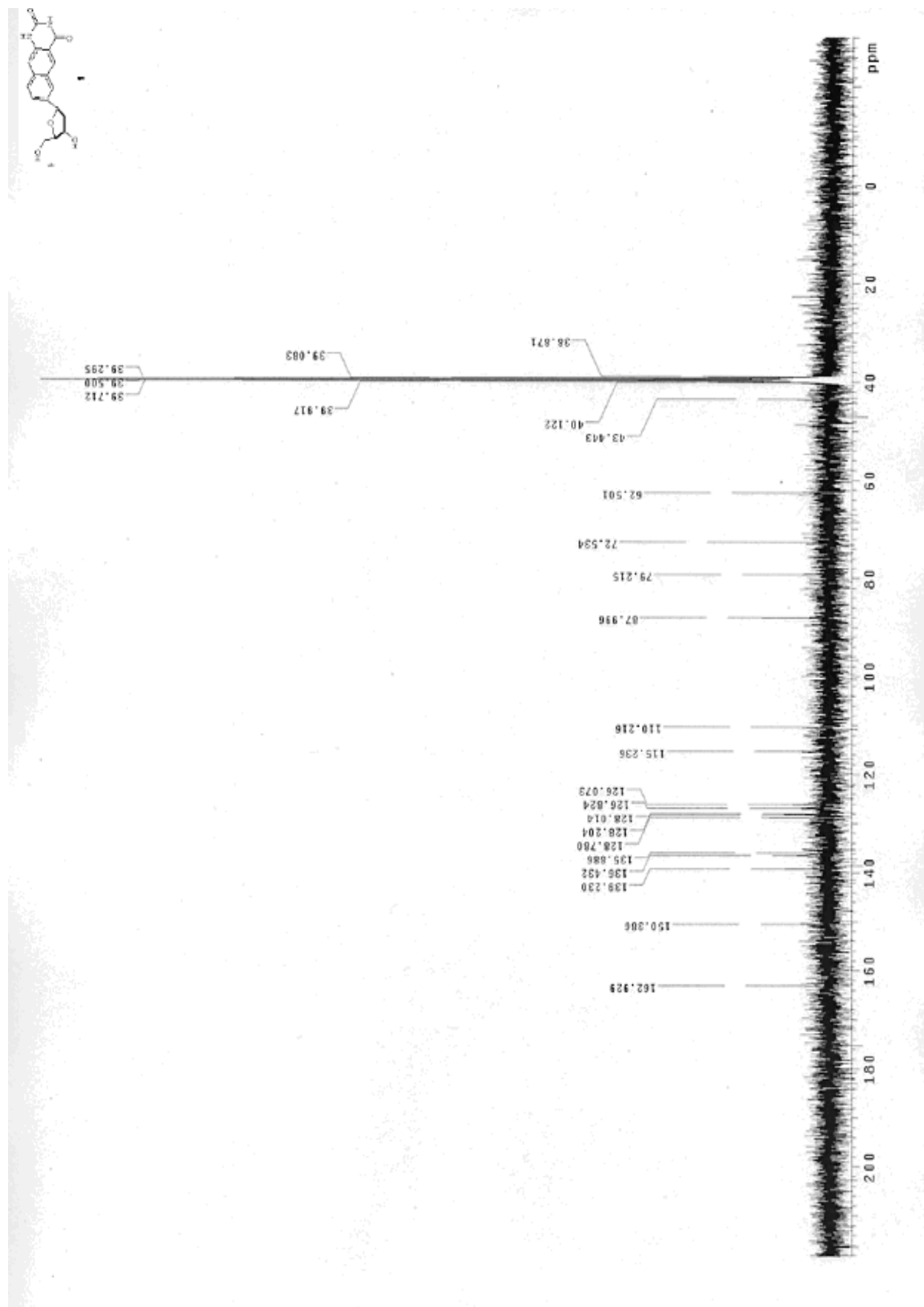


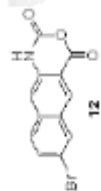






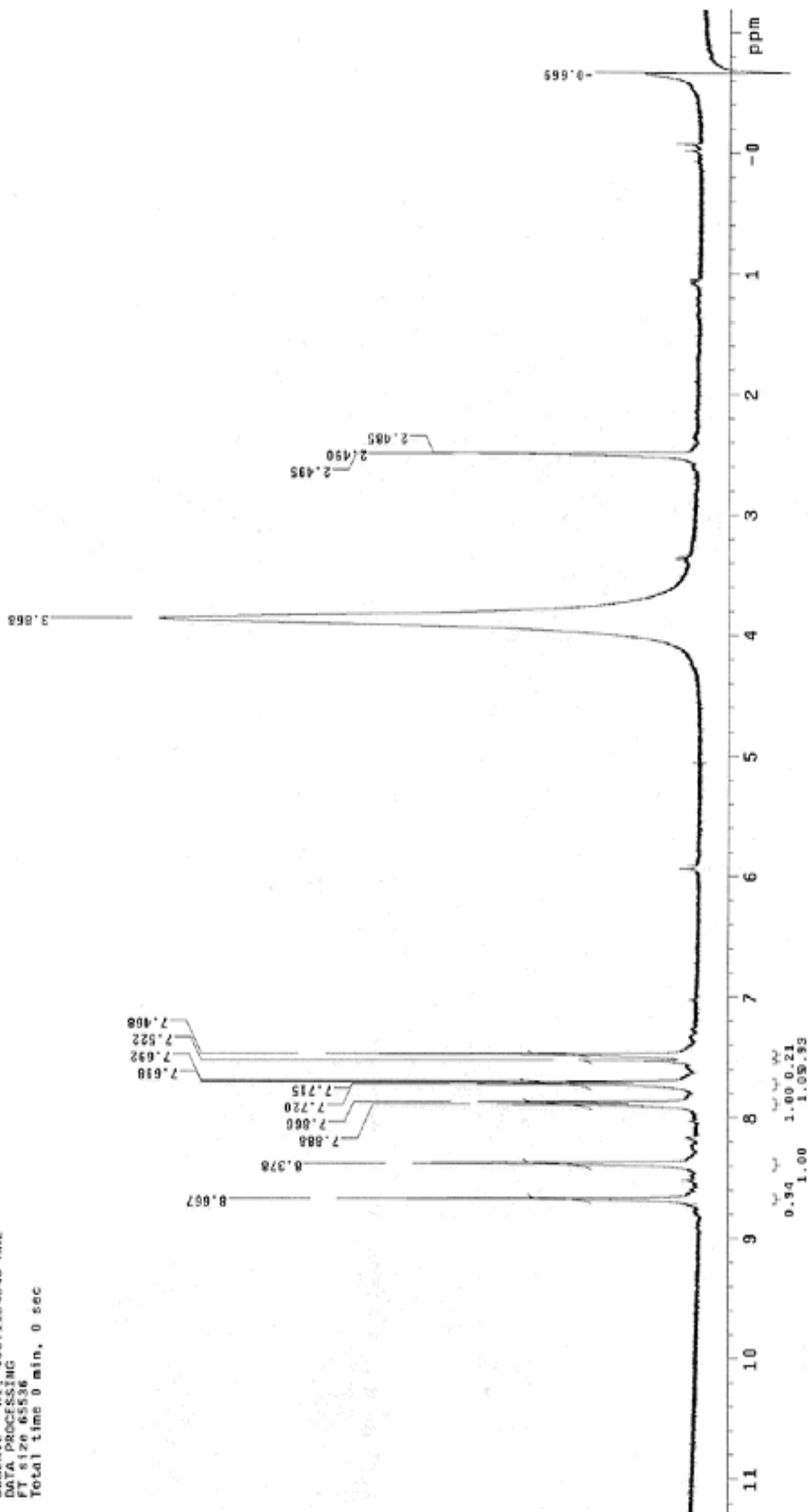


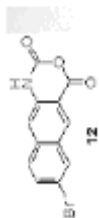




STANDARD 1H OBSERVE

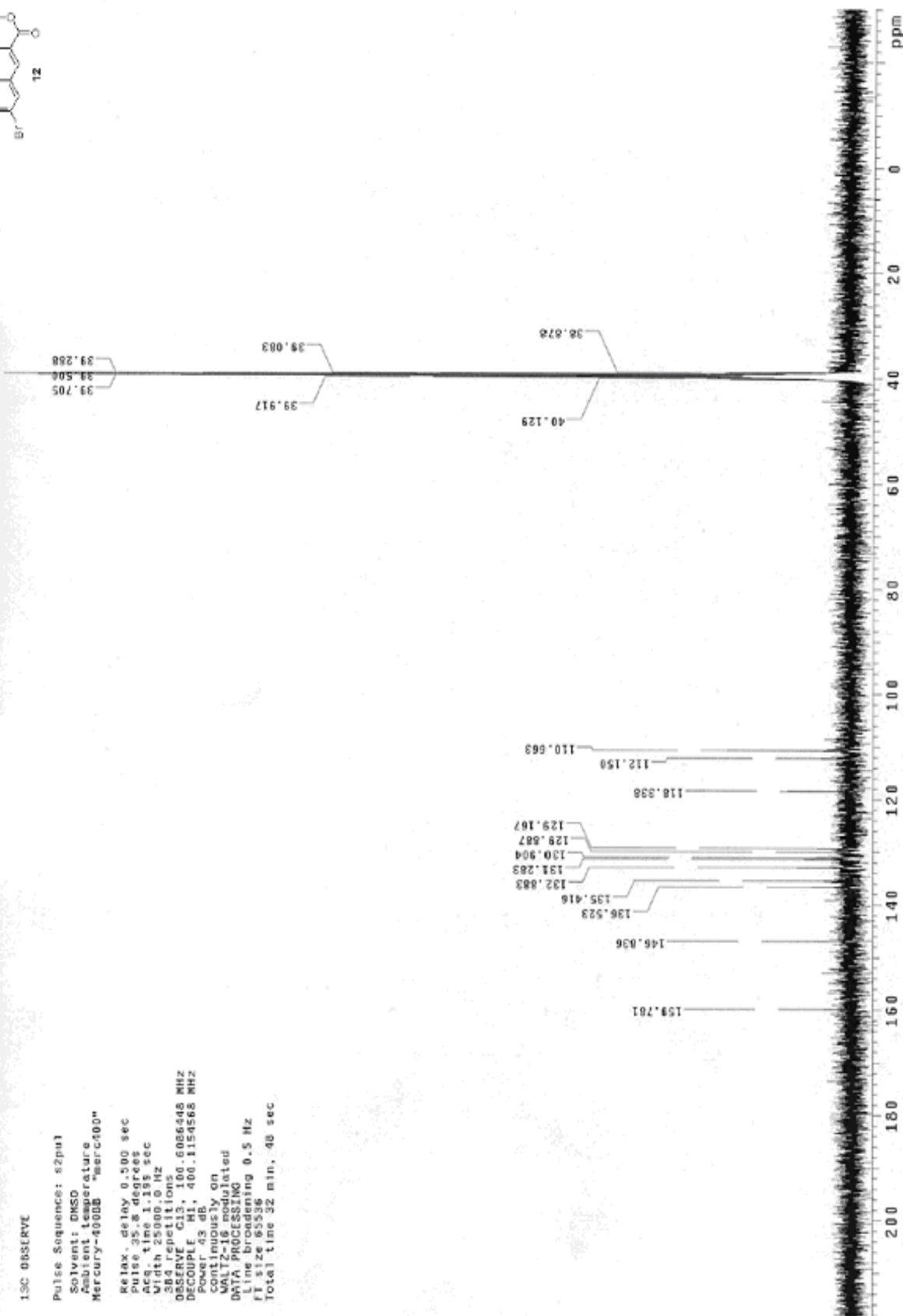
Pulse Sequence: s2pul  
 Solvent: DMSO  
 Ambient Temperature  
 Mercury-400BB "merc400"  
 Relax. delay 0.500 sec  
 Pulse 51.4 degrees  
 Acq. time 4.002 sec  
 Width 4997.5 Hz  
 4 repetitions  
 DATE\_000.1134348.MHZ  
 DATA ACQUISITION  
 FT size 65536  
 Total time 0 min, 0 sec

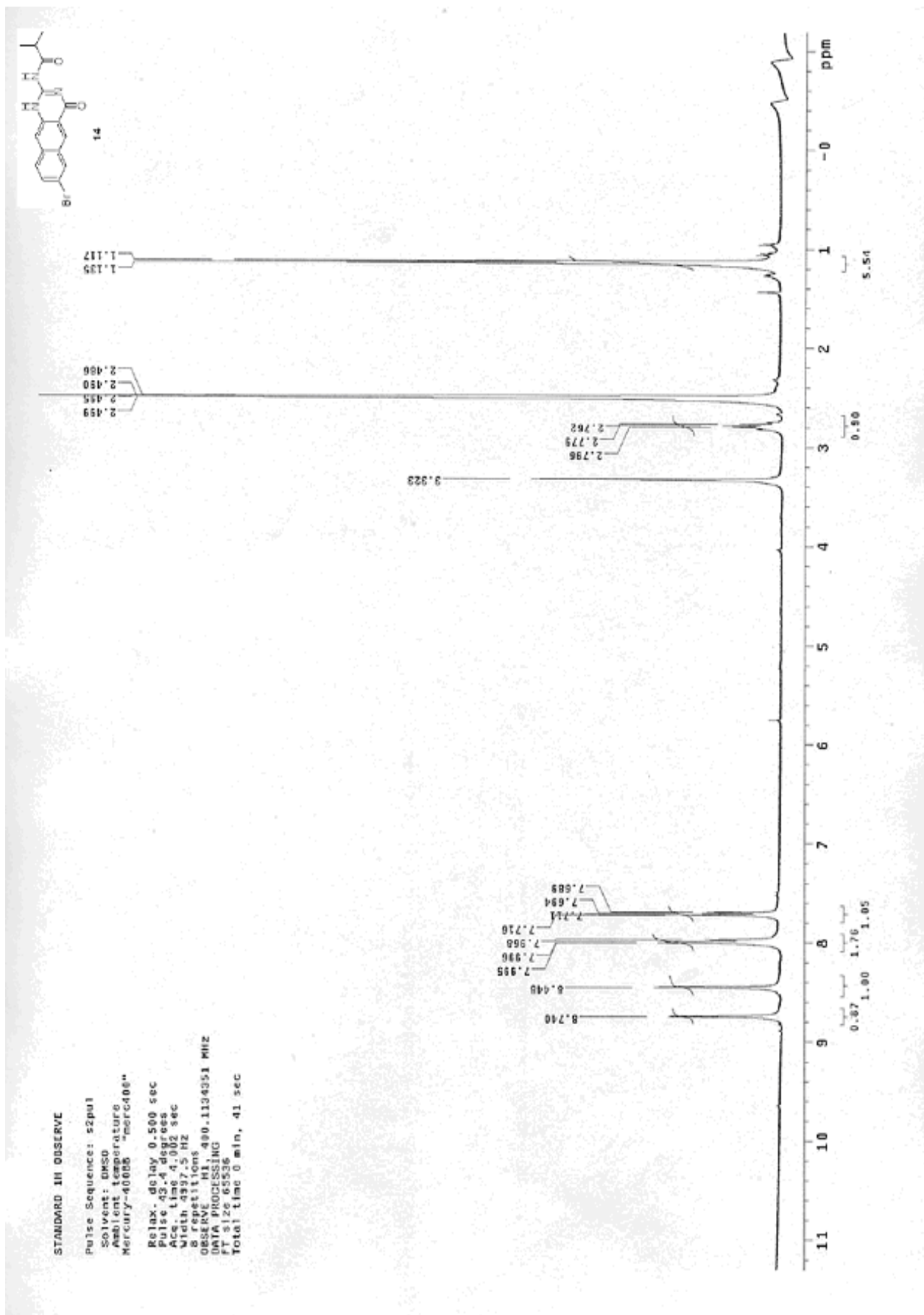


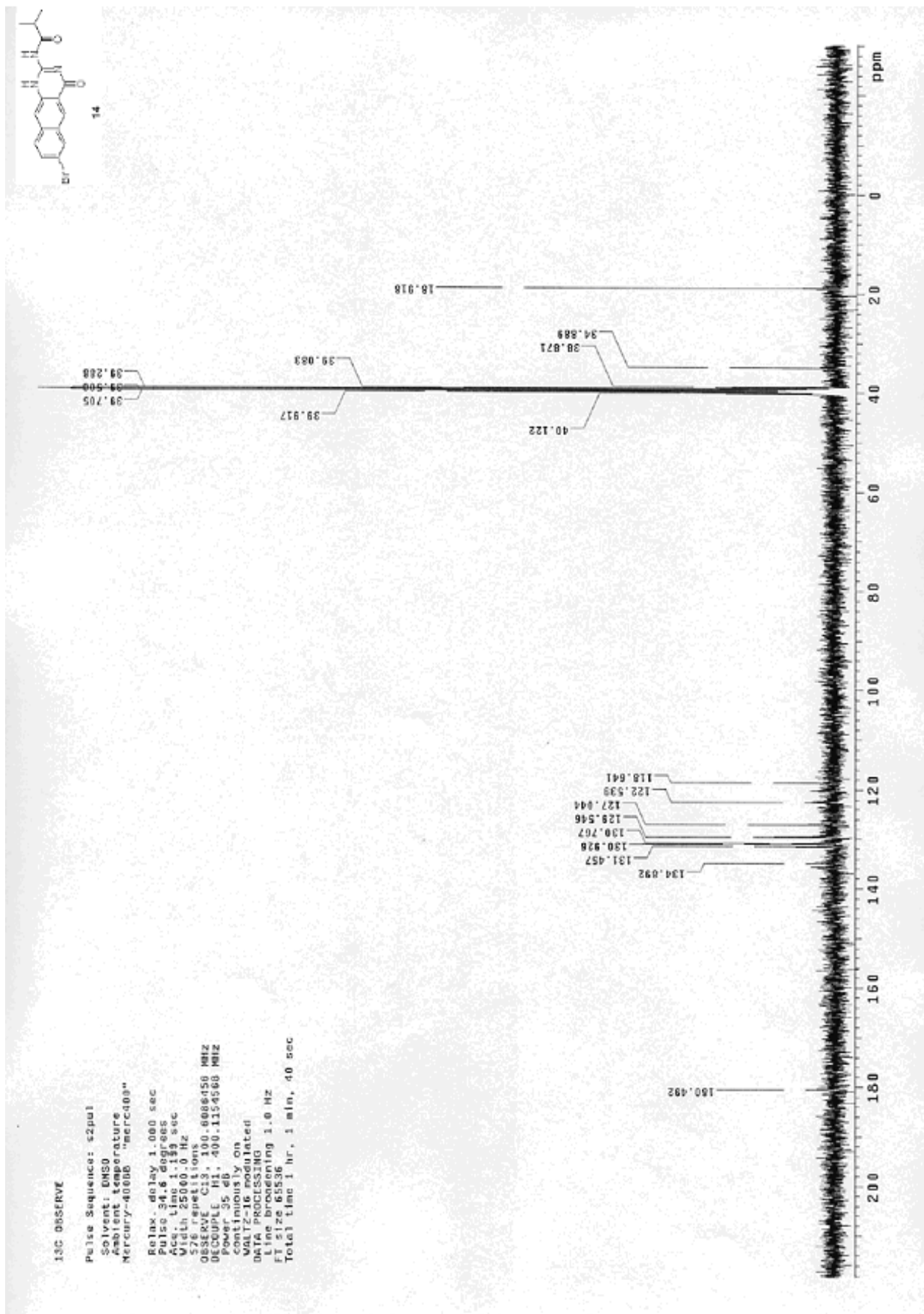


13C OBSERVE

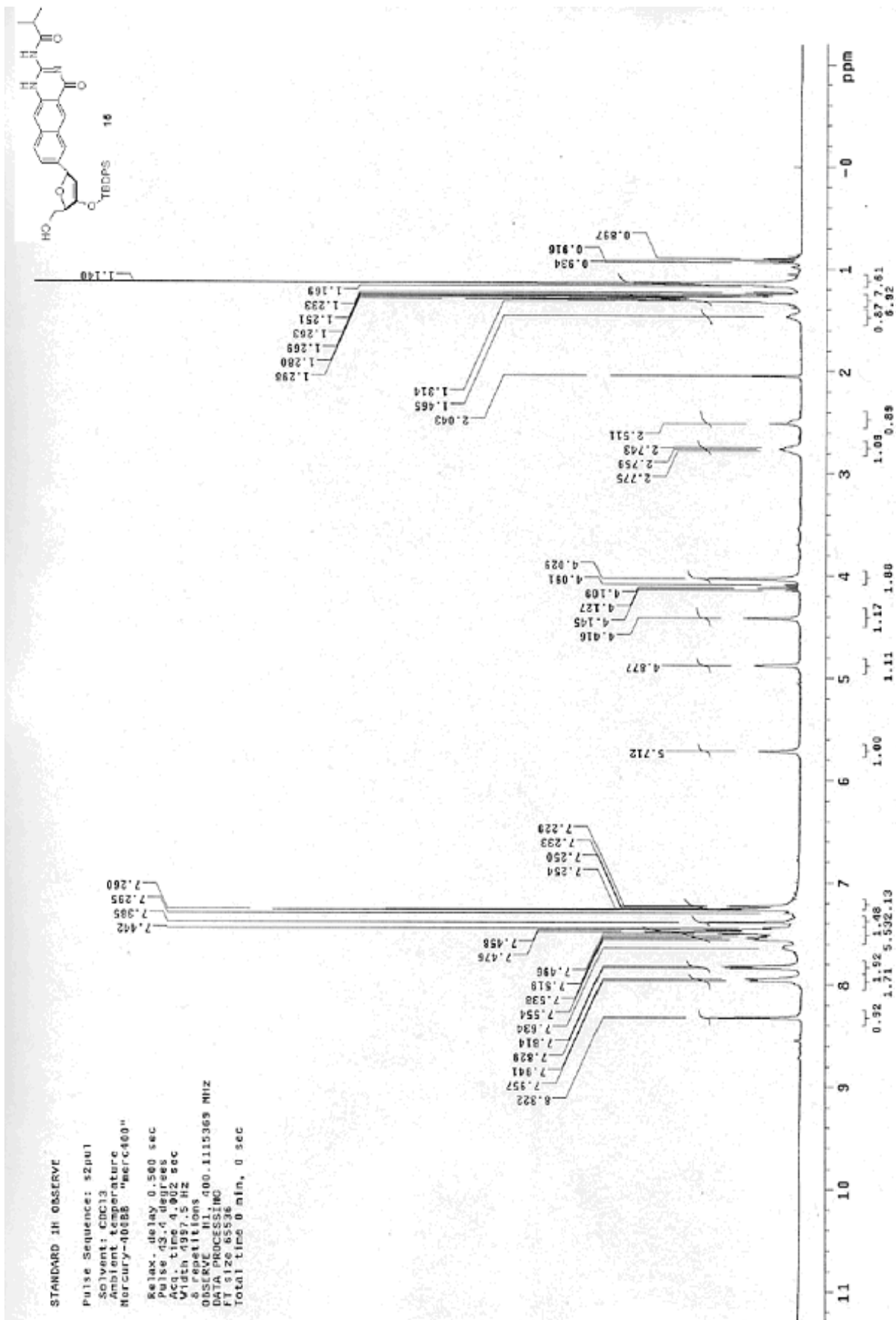
Pulse Sequence: s2pu1  
 Solvent: DMSO  
 Ambient temperature  
 Mercury-600DB "merc400"  
 Relax. delay 0.500 sec  
 Pulse 35.8 degrees  
 Acq. time 1.195 sec  
 Width 25000.0 Hz  
 Spectra 1000  
 DECOUPLE HI, 400.1154568 MHz  
 Power 43 dB, 400.1154568 MHz  
 continuously on  
 WALTZ-16 modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FI size 65536  
 Total time 32 min, 48 sec

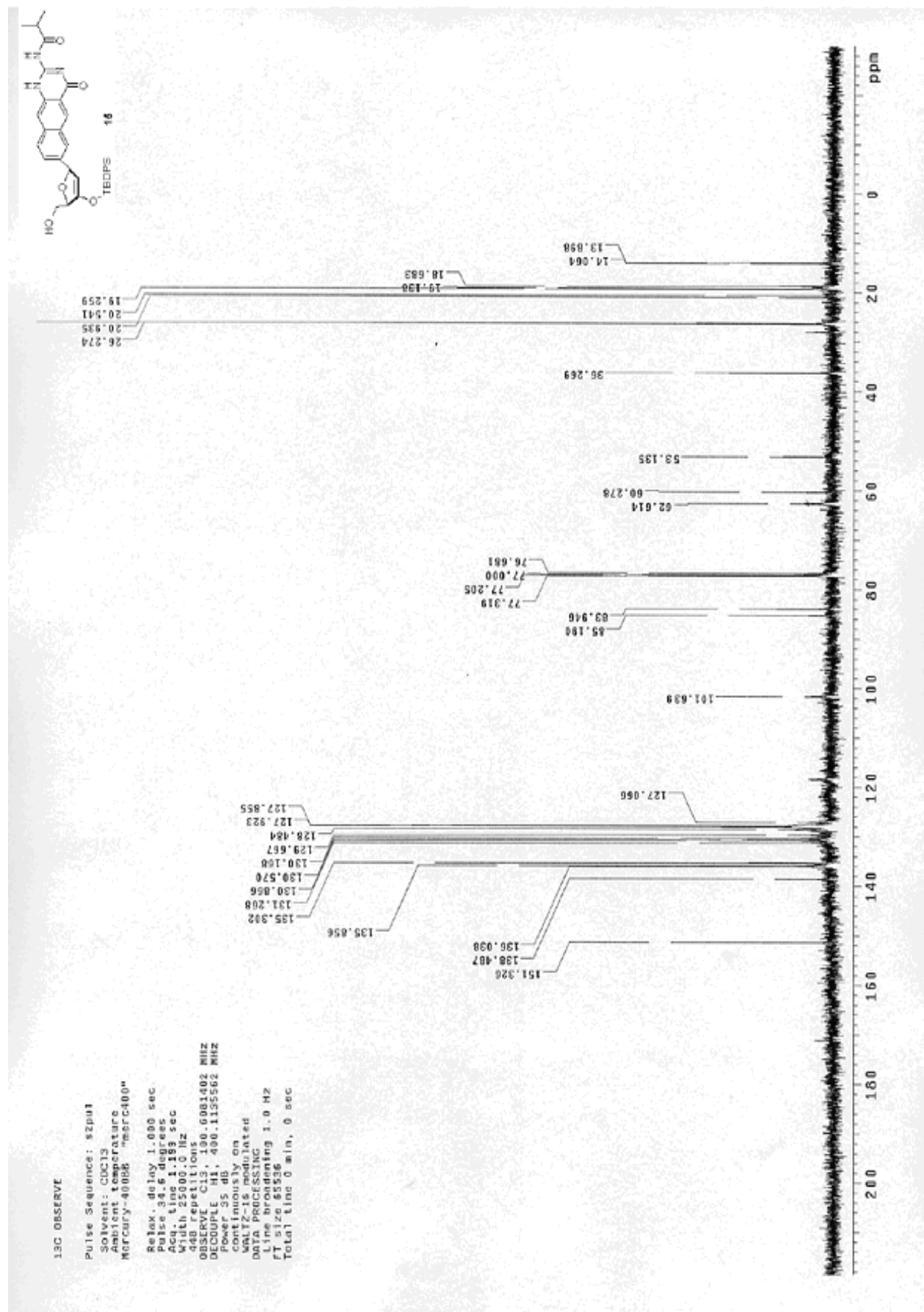


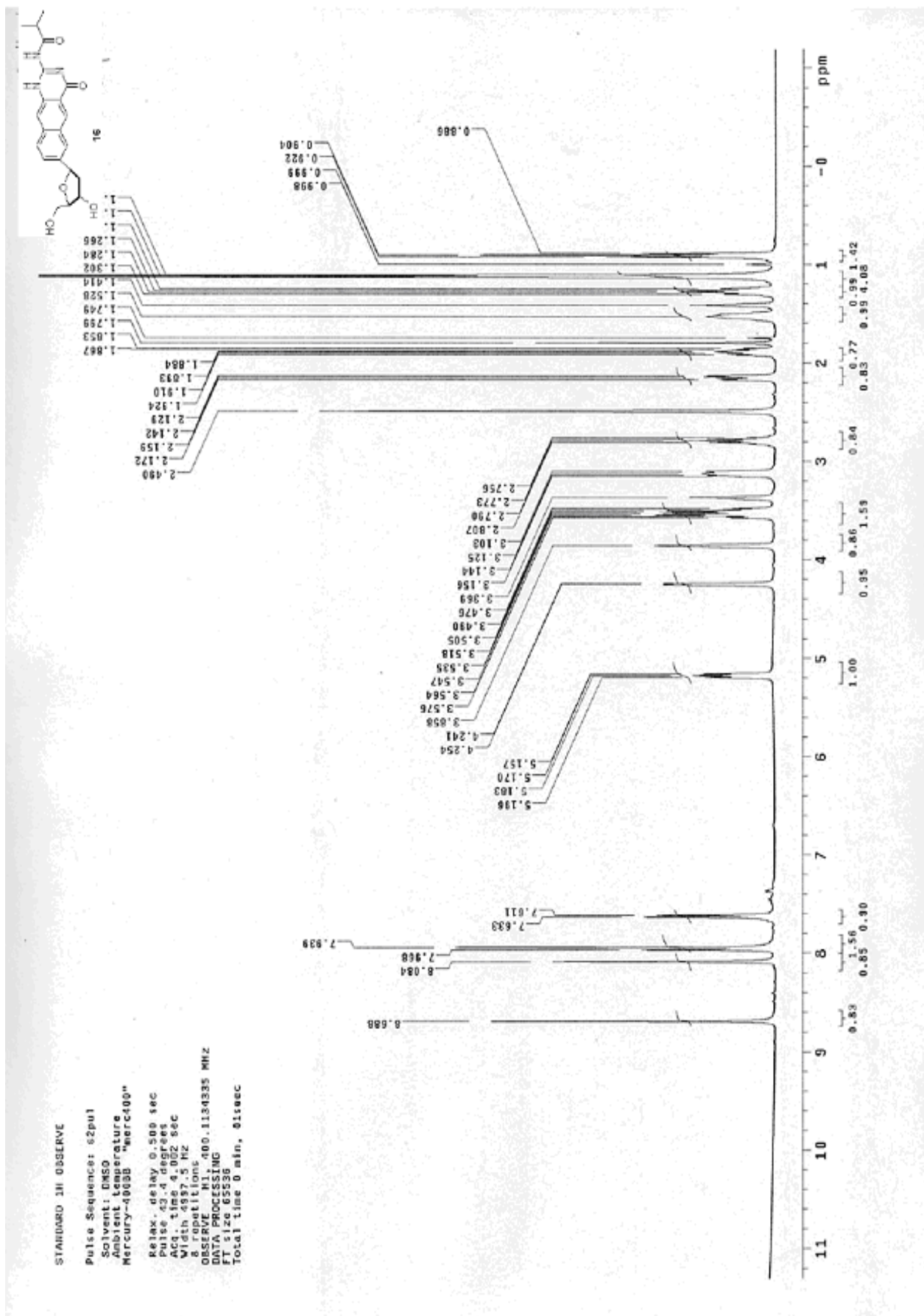


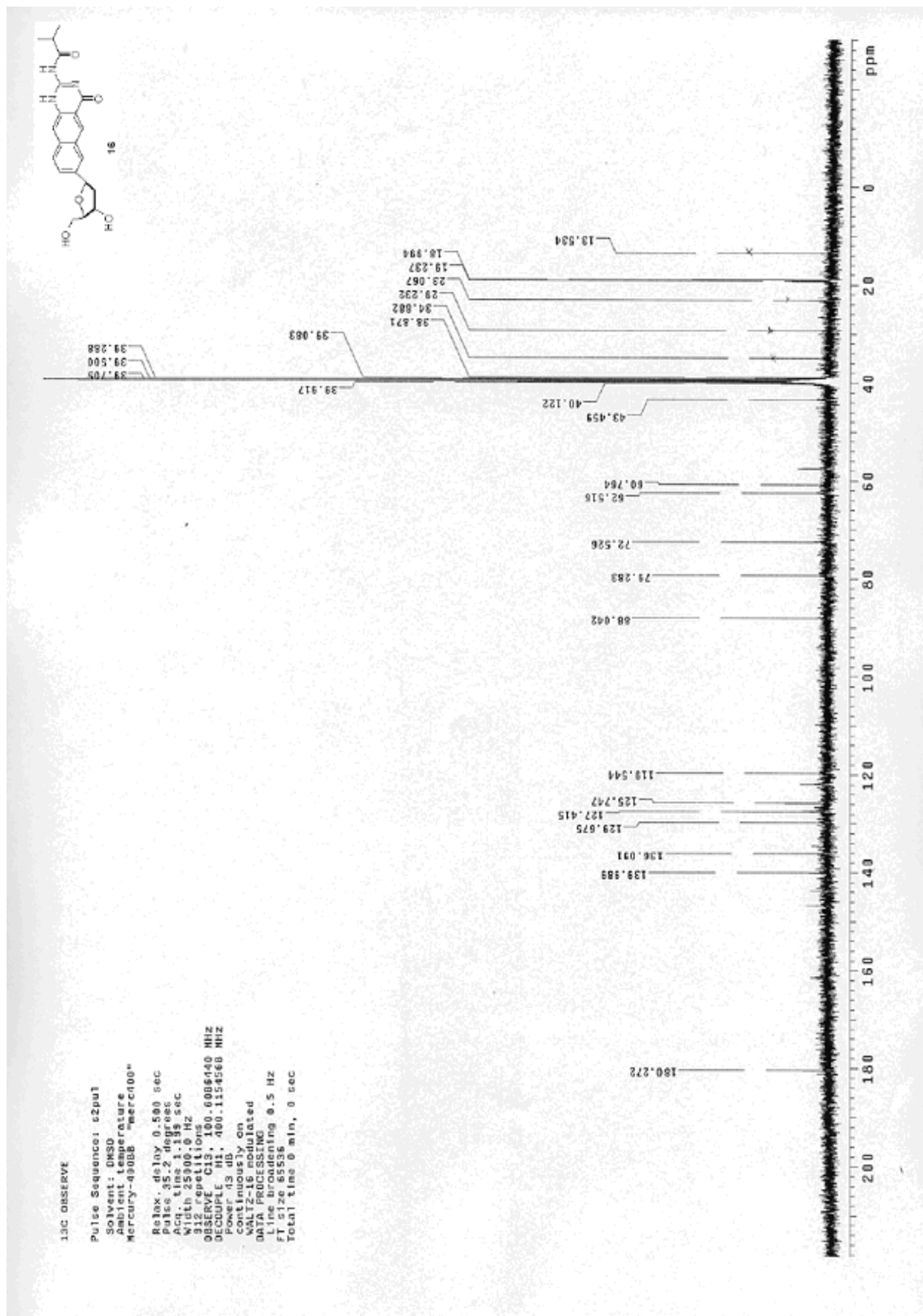


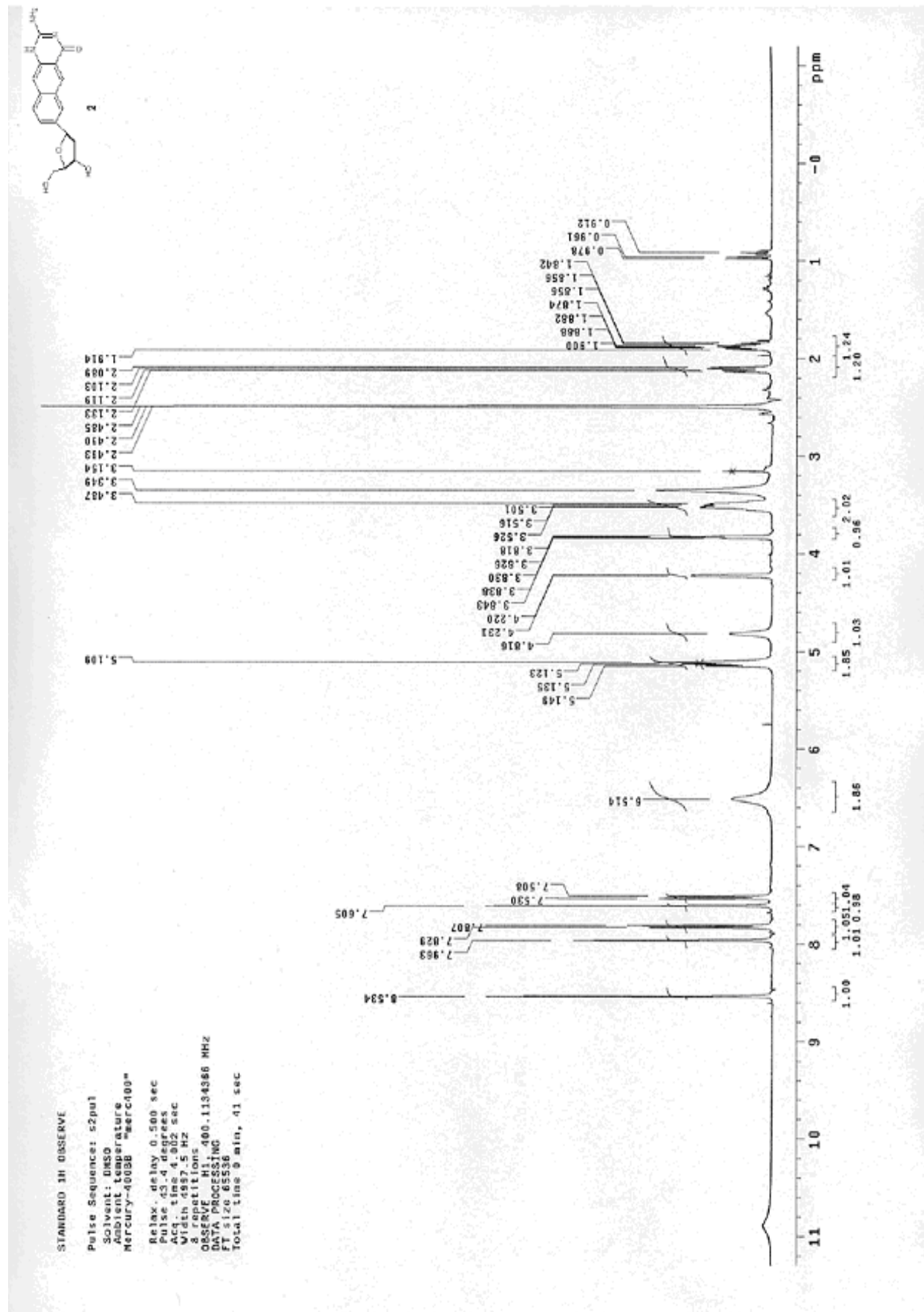


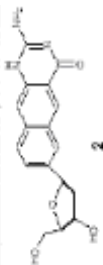






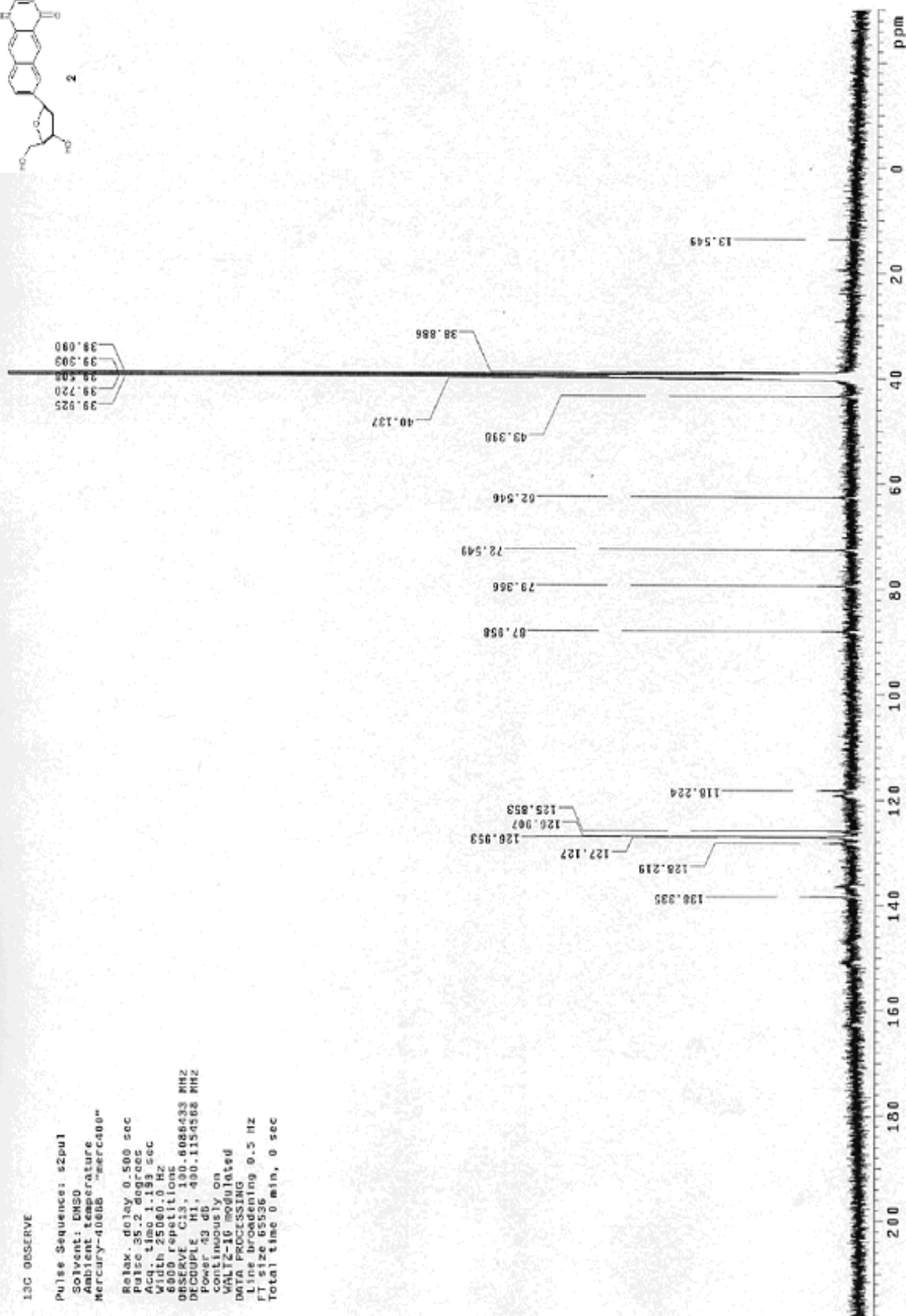


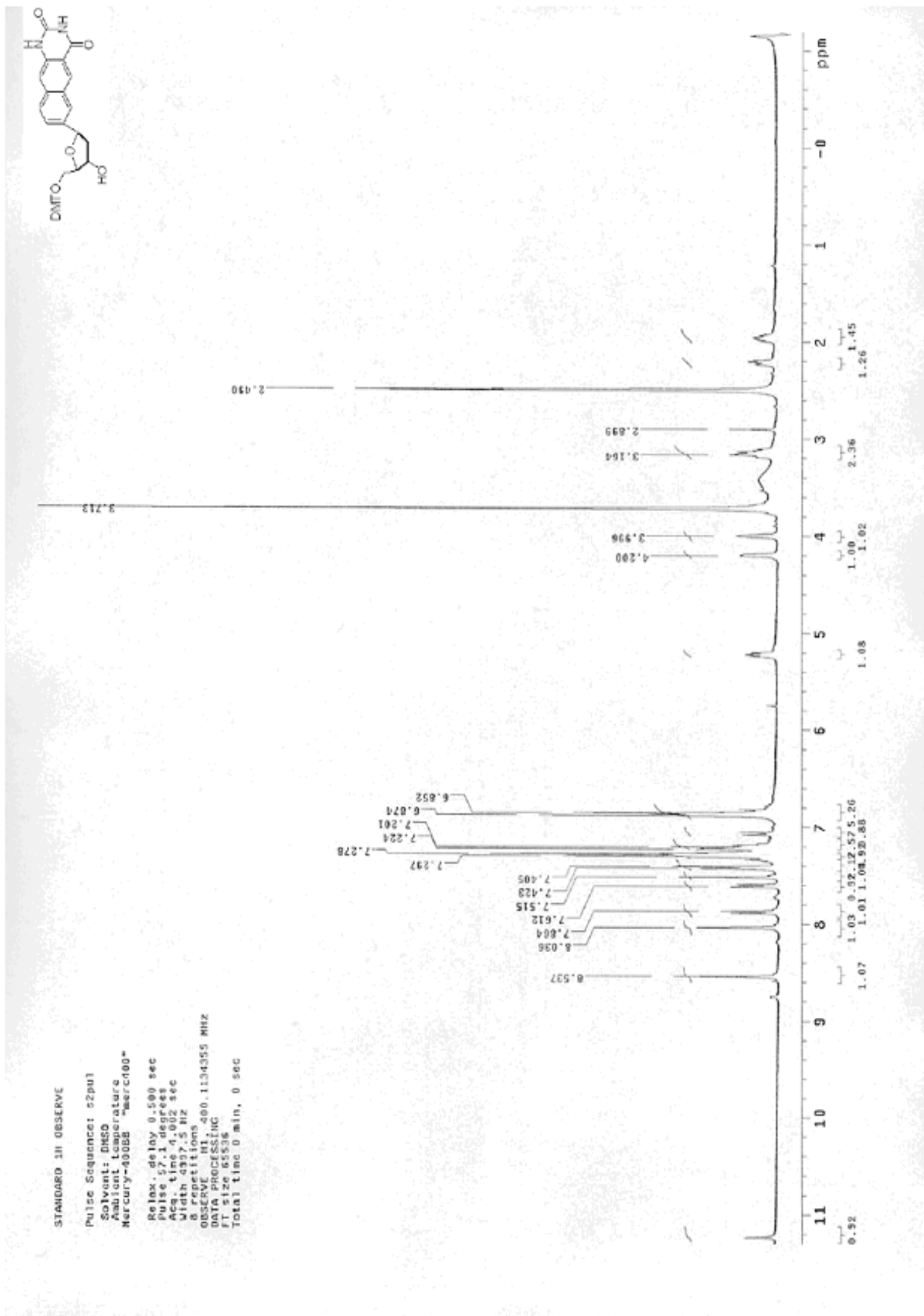


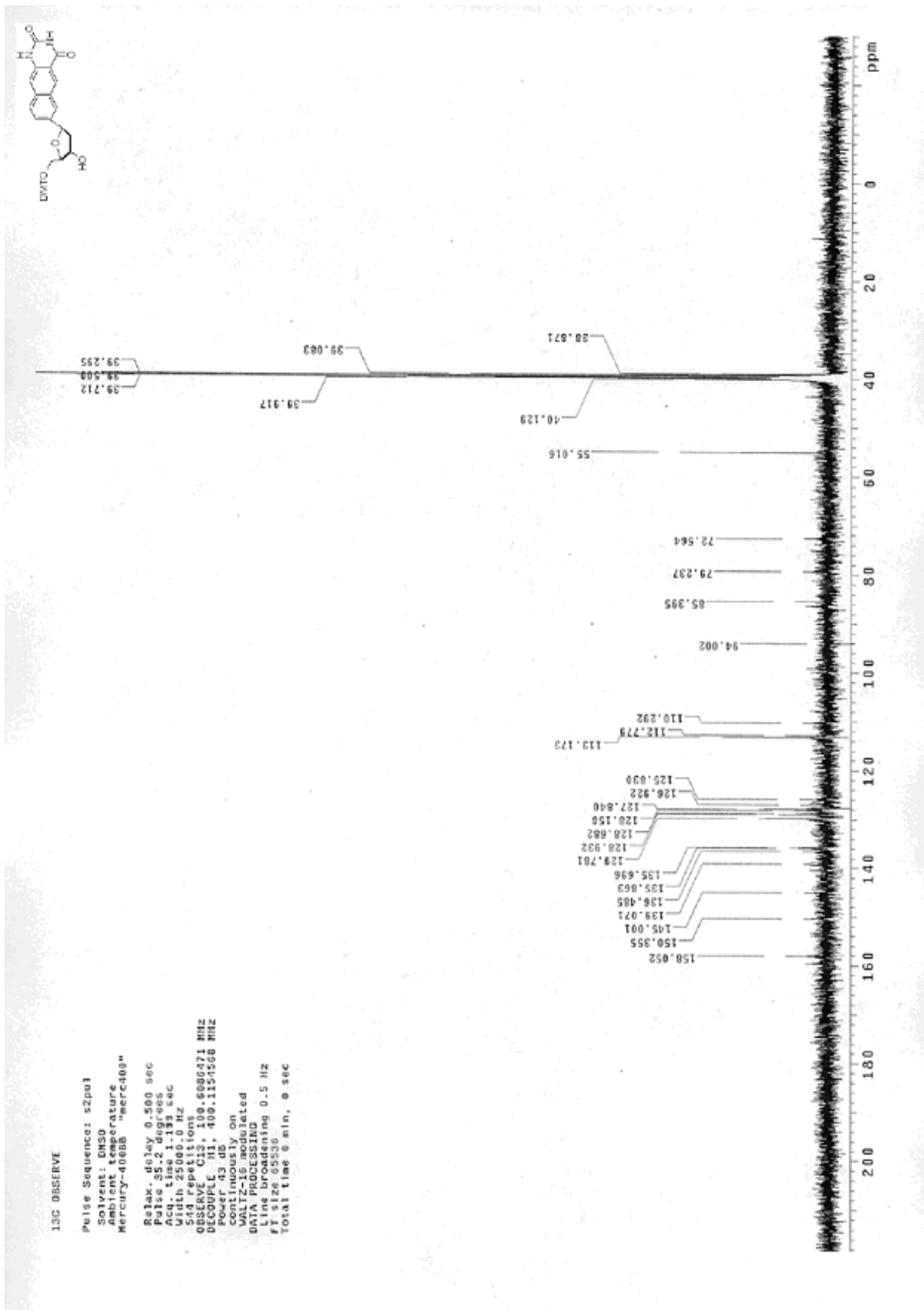


13C OBSERVE

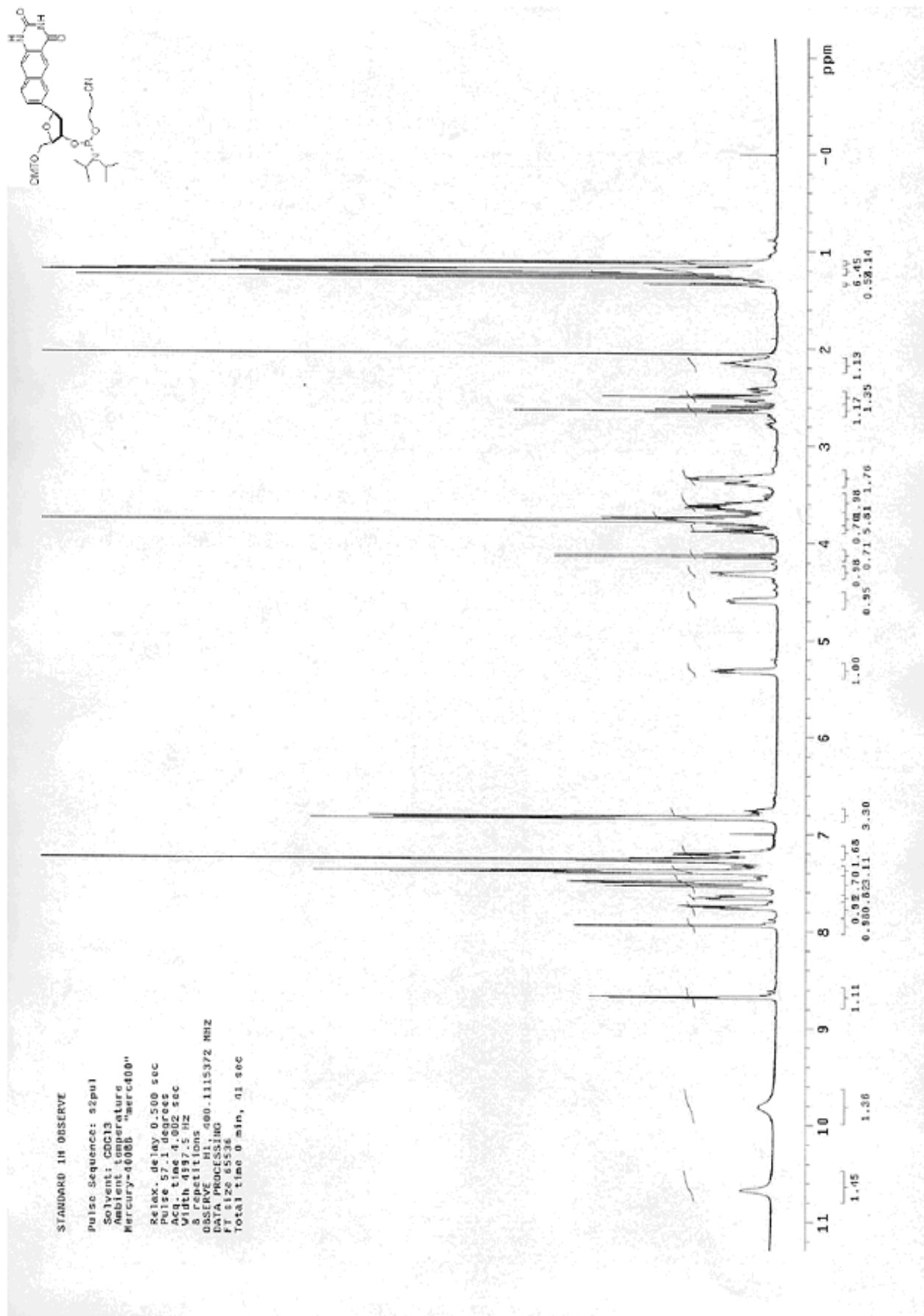
Pulse Sequence: s2pu1  
 Solvent: DMSO  
 Ambient Temperature  
 Mercury-4000B "merca00"  
 Relax\_dely 0.500 sec  
 Pulse 35.2 degrees  
 Acq. 1.000 1.193 sec  
 Cq. 1.000 1.193 sec  
 6000 Spot11cmz  
 OBSERVE C13: 130.6086433 MHz  
 DECOUPLE H1: 400.1154568 MHz  
 Power 43 db  
 continuously on  
 WALTZ-16 modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FT size 65536  
 Total time 0 min, 0 sec

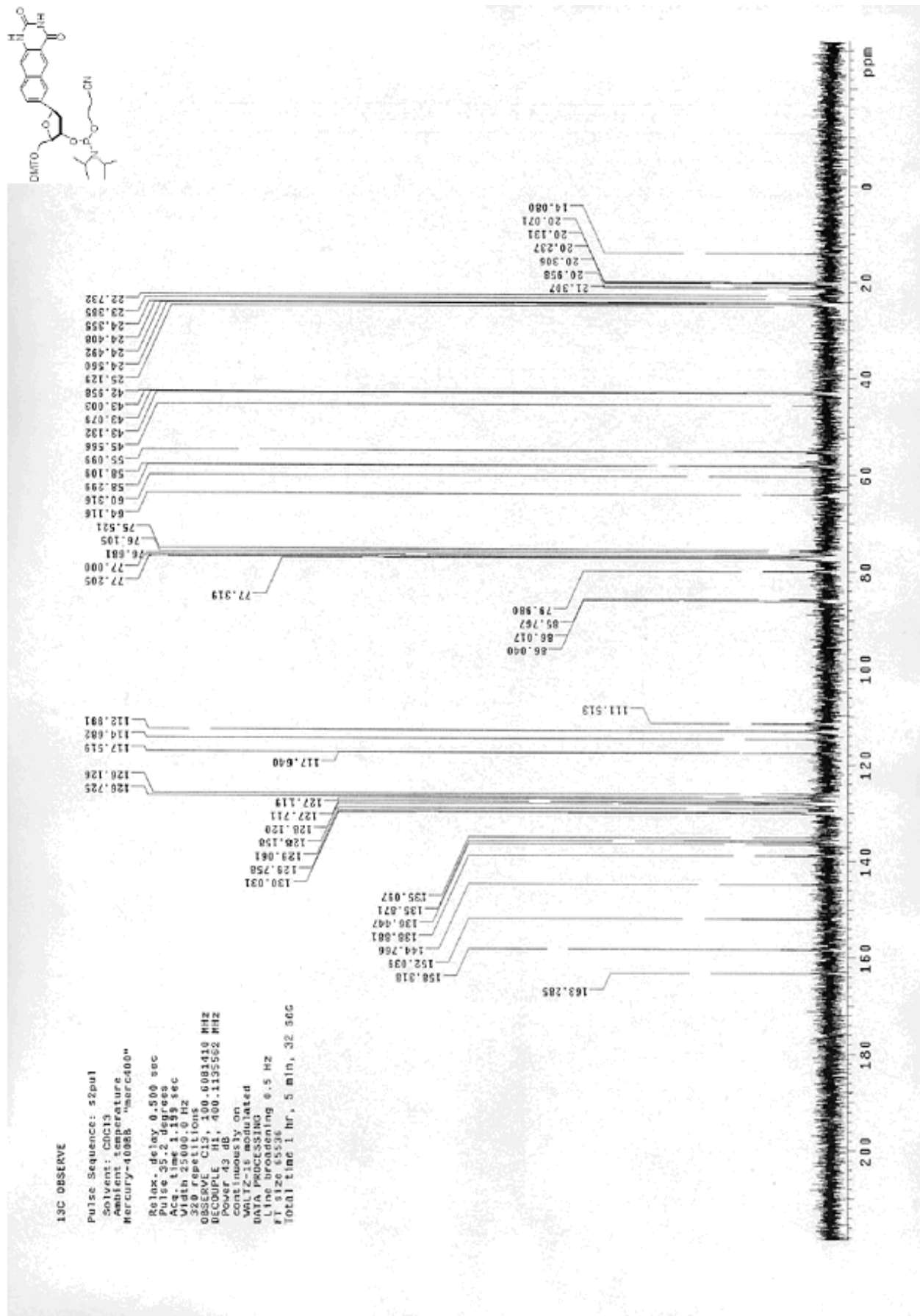
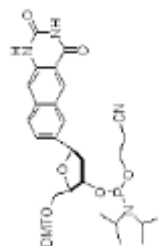


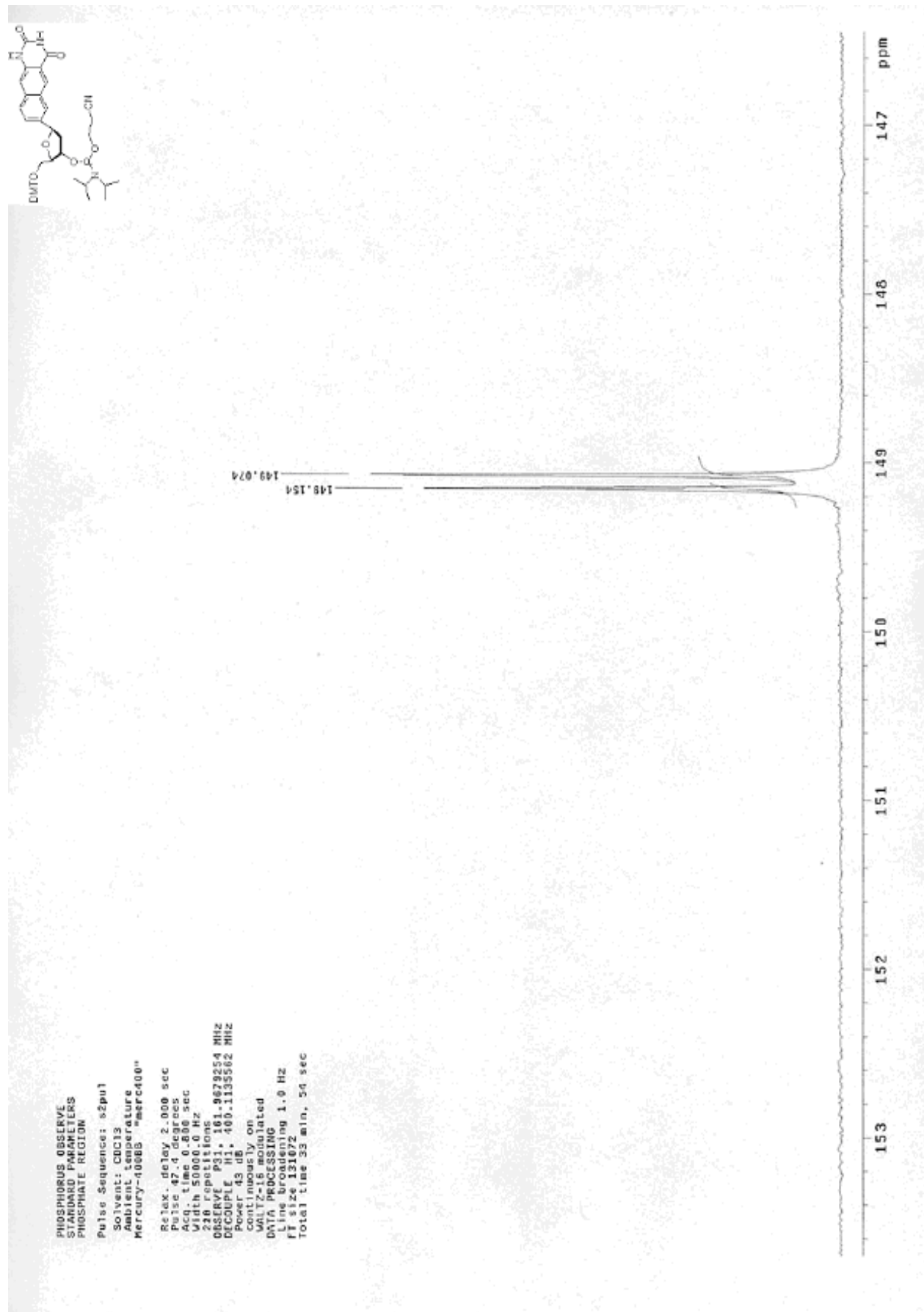


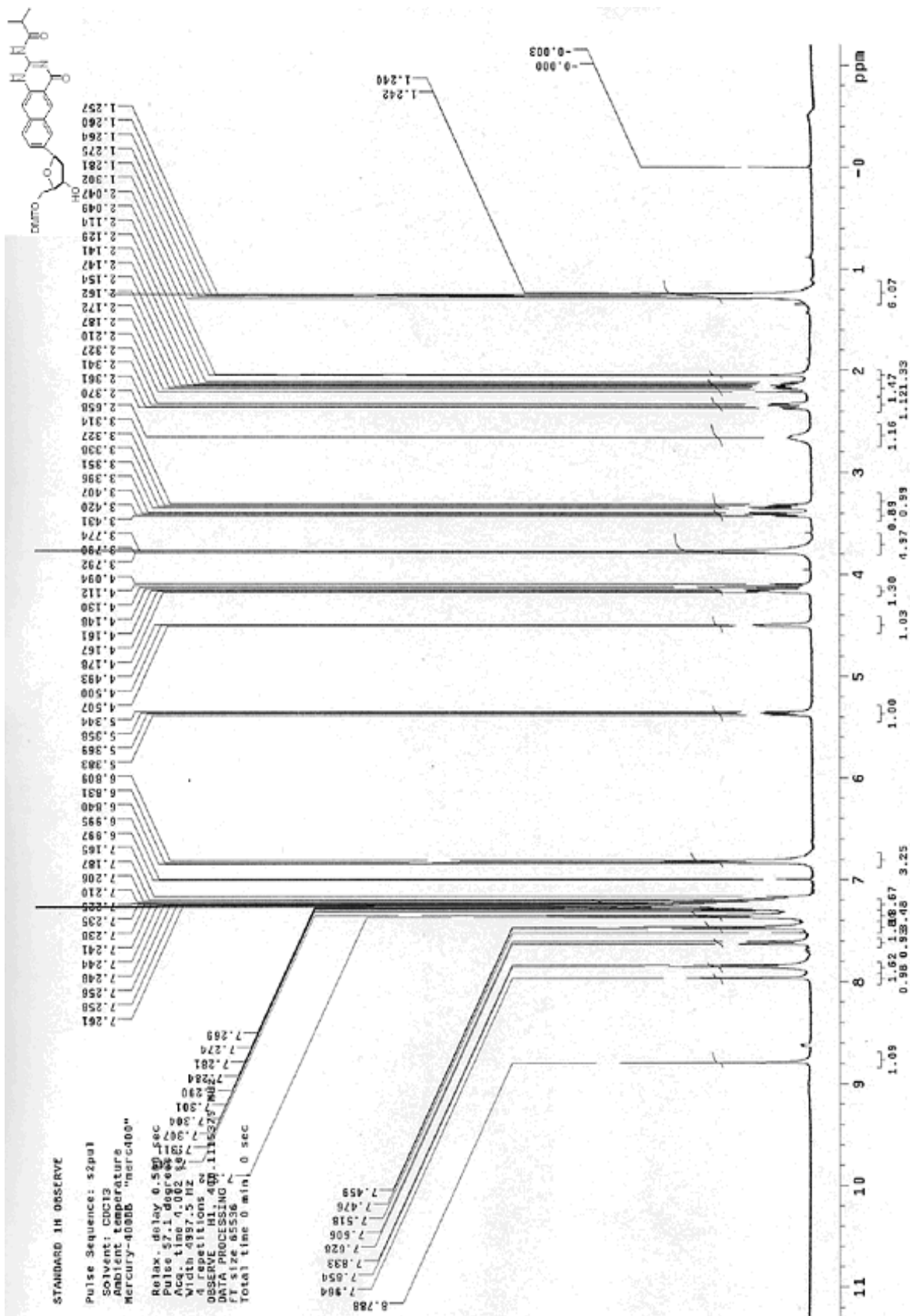


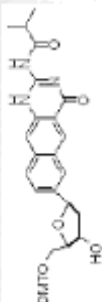












13C OBSERVE

Pulse Sequence: s2pu1  
 Solvent: CDC13  
 Ambient temperature  
 Mercury-4000B "mercs00"  
 Relax. delay 0.500 sec  
 Pulse 35.2 degrees  
 Acq. time 1.199 sec  
 Width 25000.0 Hz  
 S28 repetitions  
 OBSERVE C13, 100.6081639 MHz  
 DECOUPLE H1, 400.1135562 MHz  
 Powering up by on  
 WALTZ-16 modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FT size 65536  
 Total time 2 hr, 11 min, 5 sec

