Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.
This journal is © The Royal Society of Chemistry 2014

Supplementary Information

A New Base-Modified Deoxycytidine Capable of Base Pairing With Guanine Using Four Hydrogen Bonds

Ken Yamada, Yoshiaki Masaki, Hirosuke Tsunoda, Akihiro Ohkubo, Kohji Seio, and Mitsuo Sekine*

*Department of Life Science, Tokyo Institute of Technology, 4259 Nagatsuta, Yokohama 226-8501, Japan

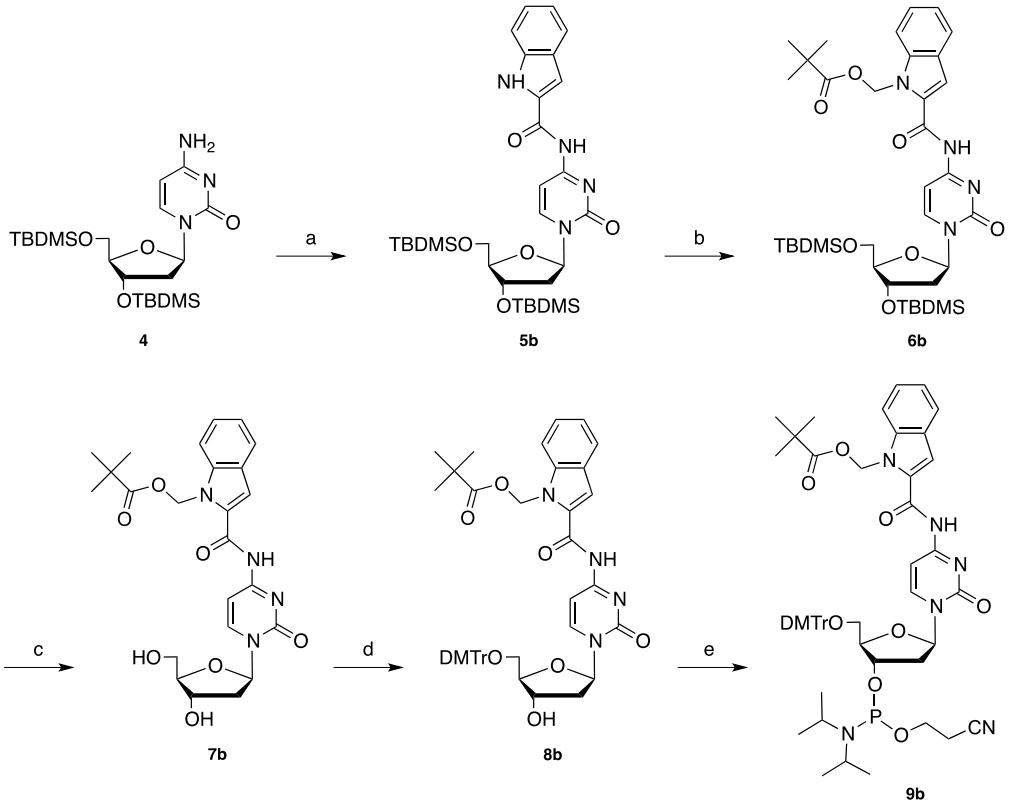
Email address: msekine@bio.titech.ac.jp

Content

Synthetic scheme of phosphoramidite **9b** ---------- S3
Figure S1 ---------- S4
Figure S2 ---------- S4
Figure S3 ---------- S5
Synthesis of DNA oligonucleotides containing dC^{Pyc} (ODNs**1**-**4**) ---------- S6
Table S1 ---------- S6
Synthesis of DNA oligonucleotides containing dC^{Inc} (ODN**5** and ODN**6**) ---------- S7
Table S2 ---------- S7
Figure S4 ---------- S7
Synthesis of Dickerson Drew sequenced-ODN**7** incorporating dC^{Pyc} ---------- S8
Figure S5 ---------- S8
^{1}H-NMR of **1** ----------S9
^{13}C-NMR of **1** ----------S10
^{1}H-NMR of **2** ----------S11
^{13}C-NMR of **2** ----------S12
^{1}H-NMR of **5a** ----------S13
^{13}C-NMR of **5a** ----------S14
^{1}H-NMR of **5b** ----------S15
^{13}C-NMR of **5b** ----------S16
^{1}H-NMR of **5c** ----------S17
^{13}C-NMR of **5c** ----------S18
^{1}H-NMR of **6a** ----------S19
^{13}C-NMR of **6a** ----------S20
^{1}H-NMR of **6b** ----------S21

^{13}C-NMR of **6b** ----------S22
^{1}H-NMR of **7a** ----------S23
^{13}C-NMR of **7a** ----------S24
^{1}H-NMR of **7b** ----------S25
^{13}C-NMR of **7b** ----------S26
^{1}H-NMR of **8a** ---------- S27
^{13}C-NMR of **8a** ----------S28
^{1}H-NMR of **8b** ---------- S29
^{13}C-NMR of **8b** ----------S30
^{1}H-NMR of **9a** ----------S31
^{13}C-NMR of **9a** ----------S32
^{31}P-NMR of **9a** ----------S33
^{1}H-NMR of **9b** ----------S34
^{13}C-NMR of **9b** ----------S35
^{31}P-NMR of **9b** ----------S36
Computational method ----------S37
Figure S6 ---------- S38
Figure S7 ---------- S38
Figure S8 ---------- S39
Figure S9 ---------- S40
Figure S10 ---------- S40

Synthetic scheme of phosphoramidite 9b.



Scheme S1. *Reagents and conditions*: (a) 1-*H*-indole-2-carbonyl chloride (1.2 eq), Pyridine-CH_3CN (1:1, v/v), rt, 2 h, 90% (b) POM-Cl (1.5 eq), NaH (3.0 eq), 99% (c) TBAF (2.4 eq), THF, rt, 2 h, 59%, (d) DMTr-Cl (1.2 eq), pyridine, rt, 2 h, 99% (e) CEO-P[N-(i-Pr_2)$_2$] (1.5 eq), 1*H*-tetrazole (0.6 eq), i-Pr_2NH (0.6 eq), CH_3CN, rt, 2 h, 80%.

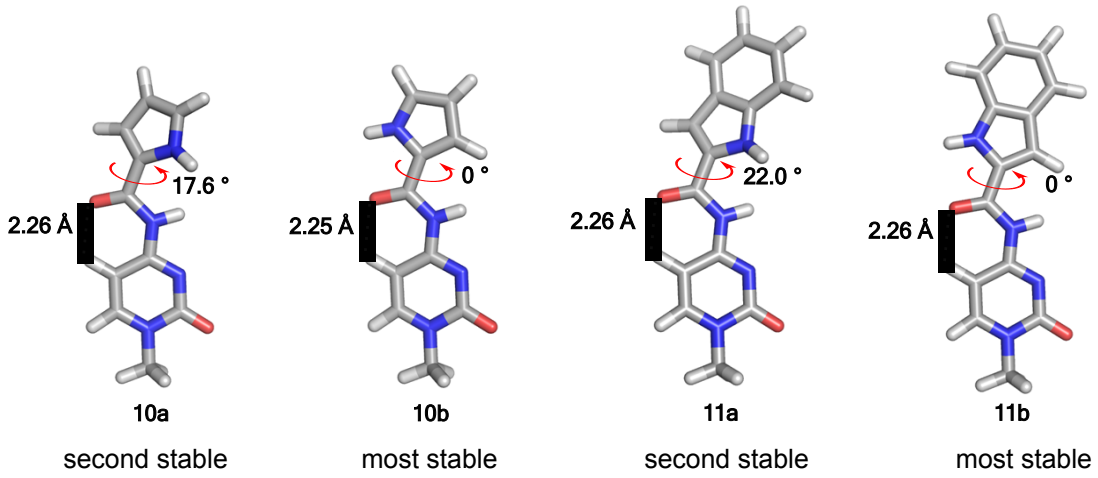
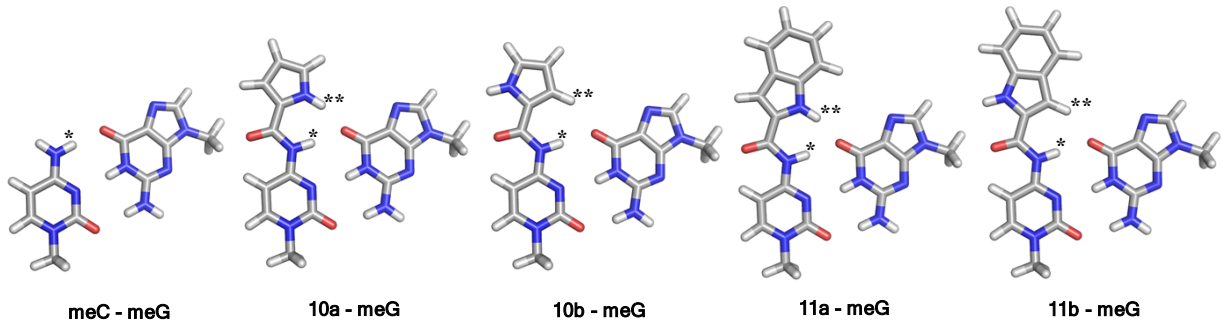


Figure S1. Geometry-optimized structures of rotational isomers of 4-*N*-(1*H*-pyrrol-2-ylcarbonyl)-1-methylcytosine (**10a**, **10b**) and 4-*N*-(1*H*-indole-2-ylcarbonyl)-1-methylcytosine (**11a**, **11b**) determined by *ab initio* MO calculation at the HF/6-31G* level.



distance	meC-meG	**10a**-meG	**10b**-meG	**11a**-meG	**11**-meG
NH* --- O (Å)	1.92	1.99	1.96	1.99	1.95
NH**(or CH**) --- O (Å)	-	1.87	2.18	1.87	2.15
Mulliken charge					
H*	0.51	0.52	0.51	0.53	0.51
H**	-	049	0.32	0.49	0.33
Hydrogen bond energy (kcal/mol)	-25.5	-27.7	-25.0	-27.7	-25.5

Figure S2. Most stable geometry-optimized structures of modified base pairs formed between 4-*N*-acyl-1-methyl-cytosines (**10a**, **10b**, **11a**, **11b**) and 9-methylguanine determined by *ab initio* MO calculation at the HF/6-31G* level; distances between the guanine 6-oxigen atom and one of the aromatic protons of the acyl group capable of hydrogen bonding, and the Mulliken charges of the protons (H* and H**) marked by asterisks.

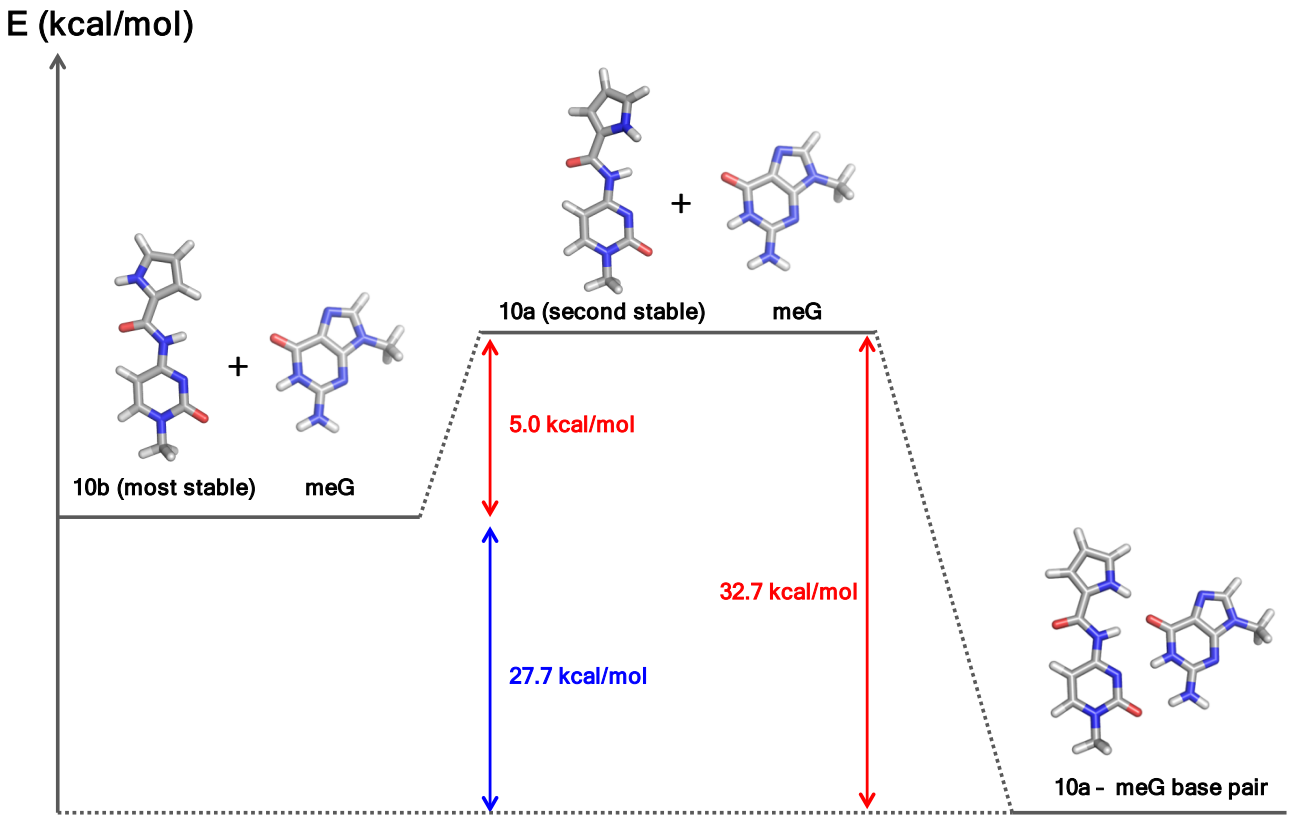


Figure S3. Energy profile in formation of a base pair between **10a** and meG.

Synthesis of DNA oligonucleotides containing dC^{Pyc} (ODNs1-4). The synthesis of DNA oligonucleotides (ODNs**1-4**) having dC^{Pyc} was performed in an ABI 392 DNA synthesizer by the standard 1.0 μmol scale DNA phosphoramidite approach, which consists of detritylation, coupling (using Activator 42), capping and iodine oxidation steps. Releasing of oligonucleotides from CPG resin and the cleavage of protecting groups were carried out by the treatment with 1.5 M NaOMe/MeOH solution at room temperature for 30 min. Quenching of this treatment was underwent using Dowex 50W X8 (H^+ form). The polymer supports were removed by filtration and washed with 0.1 M ammonium acetate buffer (1 mL × 3). The filtrates were purified by C18 Sep-Pak cartridge and reverse-phased HPLC to give oligonulcleotides.

Table S1. Isolated yield and MALDI-TOF Mass data of oligonucleotides containing dC^{Pyc}

			MALDI-TOF mass	
DNA	oligonucleotide sequences[a]	yields (%)	found	calcd
ODN**1**	5′-d(T T C T T C **C^{Pyc}** C T T C T T)-3′	61	3910.5	3909.7
ODN**2**	5′-d(T T **C^{Pyc}** T T **C** **C^{pyc}** **C** T T **C^{Pyc}** T T)-3′	21	4096.6	4096.7
ODN**3**	5′-d(T T C T T **C^{Pyc}** **C^{Pyc}** **C^{Pyc}** T T C T T)-3′	38	4097.2	4096.7
ODN**4**	5′-d(T T **C^{Pyc}** T T **C^{Pyc}** **C^{Pyc}** **C^{Pyc}** T T **C^{Pyc}** T T)-3′	17	4282.6	4282.8

[a]C^{Pyc}; 4-*N*-(1*H*-pyrrol-2-ylcarbonyl)deoxycytidine.

Synthesis of DNA oligonucleotides containing dC^{Inc} (ODN5 and **ODN6).** The synthesis of DNA oligonucleotides (ODN**5** and ODN**6**) having dC^{Inc} was performed in an ABI 392 DNA synthesizer by the standard 1.0 μmol scale DNA phosphoramidite approach, which consists of detritylation, coupling, capping (using Activator 42) and iodine oxidation steps. Releasing of oligonucleotides from CPG resin and the cleavage of protecting groups were carried out by the treatment with 1.5 M NaOMe/MeOH solution, at room temperature for 1 h. Quenching of this treatment was underwent using Dowex 50W X8 (H^+ form). The polymer supports were removed by filtration and washed with 0.1 M ammonium acetate buffer (1 ml × 3). The filtrates were purified by C18 Sep-Pak cartridge and reverse-phased HPLC to give oligonulcleotide.

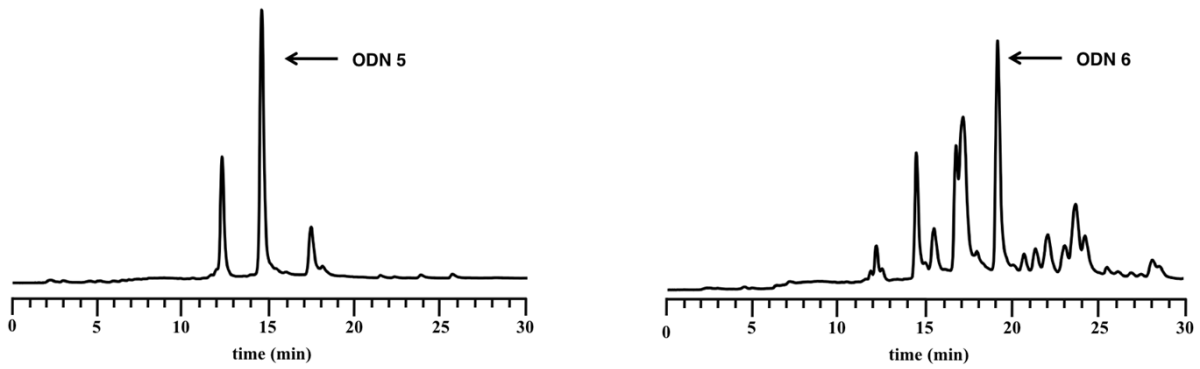


Figure S4. Reversed phase HPLC analysis of (a) ODN**5** and (b) ODN**6** after the purification by C_{18} Sep-Pak cartridge.

Table S2. Isolated yield and MALDI-TOF Mass data of oligonucleotides containing dC^{Inc}

DNA	oligonucleotide sequences[a]	yields (%)	MALDI-TOF mass found	MALDI-TOF mass calcd
ODN**5**	5′-d(T T C T T C **C^{Inc}** C T T C T T)-3′	20	3960.1	3960.7
ODN**6**	5′-d(T T **C^{Inc}** T T C **C^{Inc}** C T T **C^{Inc}** T T)-3′	2	4245.0	4245.8

[a]C^{Inc}; 4-*N*-(1*H*-indol-2-ylcarbonyl)deoxycytidine.

Synthesis of Dickerson Drew sequenced-ODN7 incorporating dC^{Pyc}. The synthesis of Dickerson Drew sequenced ODN**7** having dC^{Pyc} [5′-d(CGCGAATTC^{Pyc}GCG)-3′] was performed in an ABI 392 DNA synthesizer by the standard 1.0 μmol scale DNA phosphoramidite approach, which consists of detritylation, coupling, capping (using Activator 42) and iodine oxidation steps. Releasing of oligonucleotides from CPG resin and the cleavage of protecting groups were carried out by the treatment with 1.5 M NaOMe/MeOH solution at room temperature for 30 min. Quenching of this treatment was underwent using Dowex 50W X8 (H^+ form). The polymer supports were removed by filtration and washed with 0.1 M ammonium acetate buffer (1 ml × 3). After freezing dry treatment, the ODN was treated with 28% NH_3 aqueous solution at room temperature for 2 h to deprotect remaining Pac and *i*-Pr-Pac groups of adenine bases and guanine bases. The filtrates were purified by C18 Sep-Pak cartridge and reverse-phased HPLC to give the oligonulcleotide. ODN**7** was characterized by MALDI-TOFF mass analysis (found 3739.3, calcd 3739.5 [M+H]).

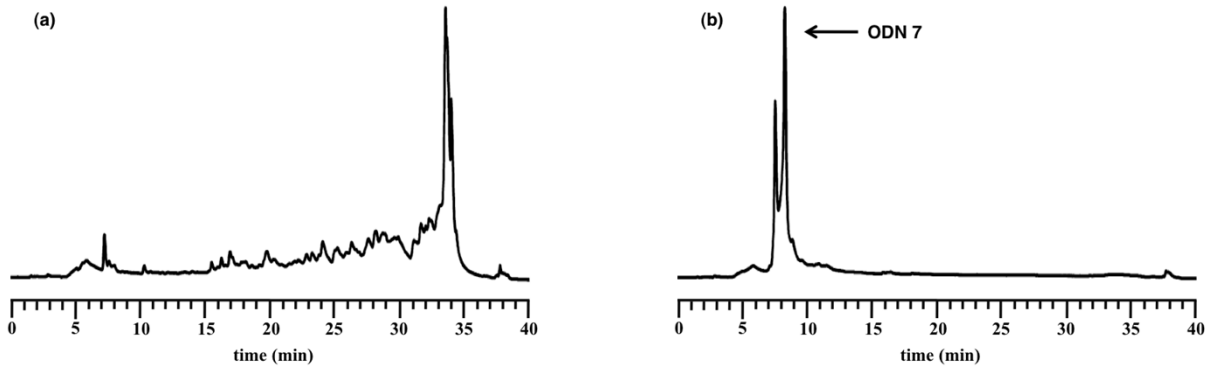
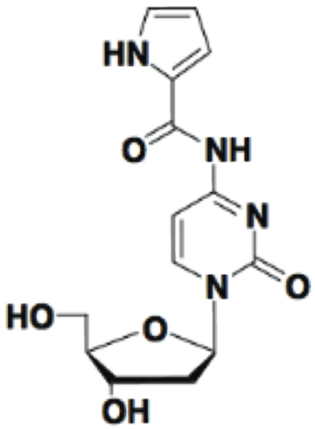
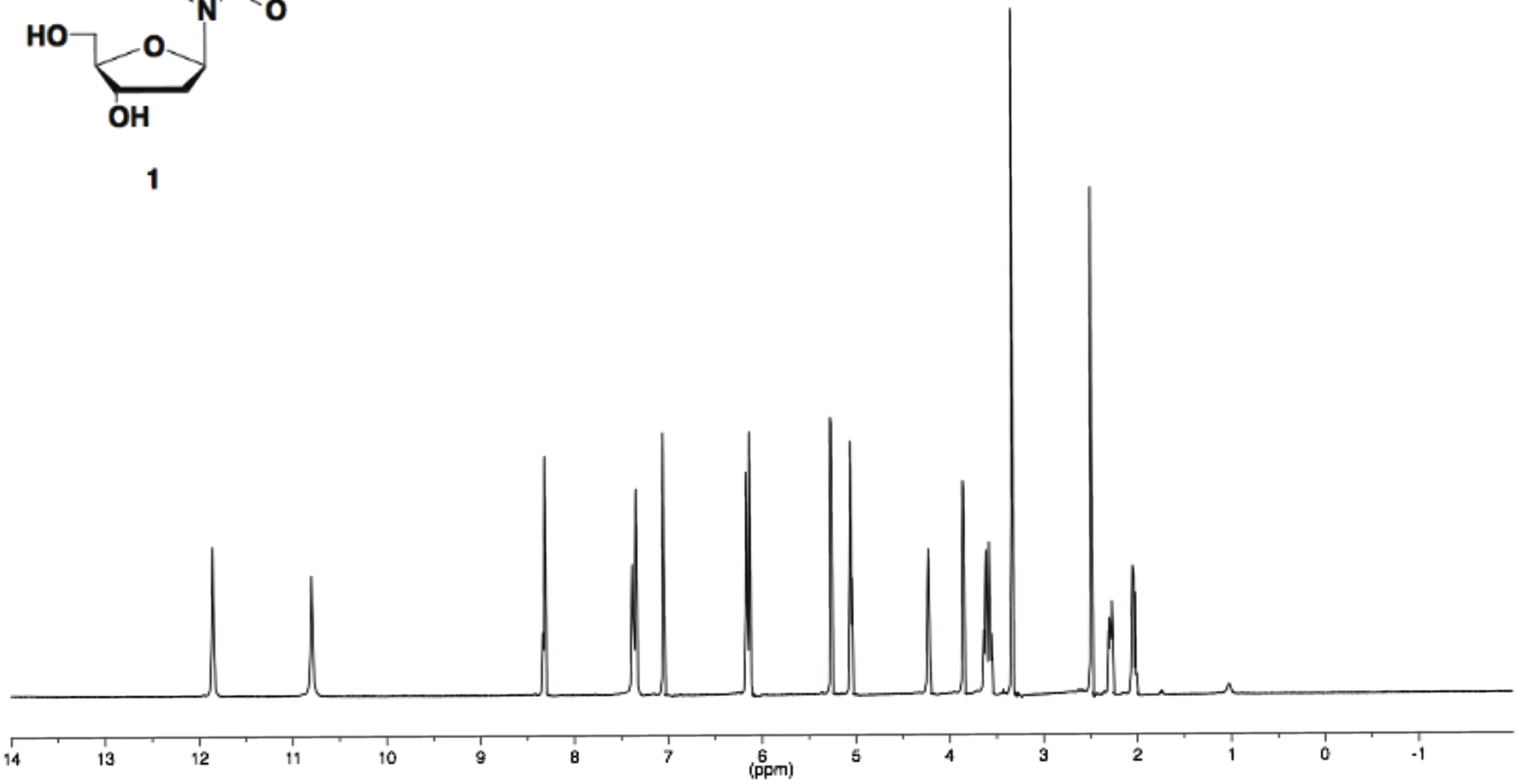
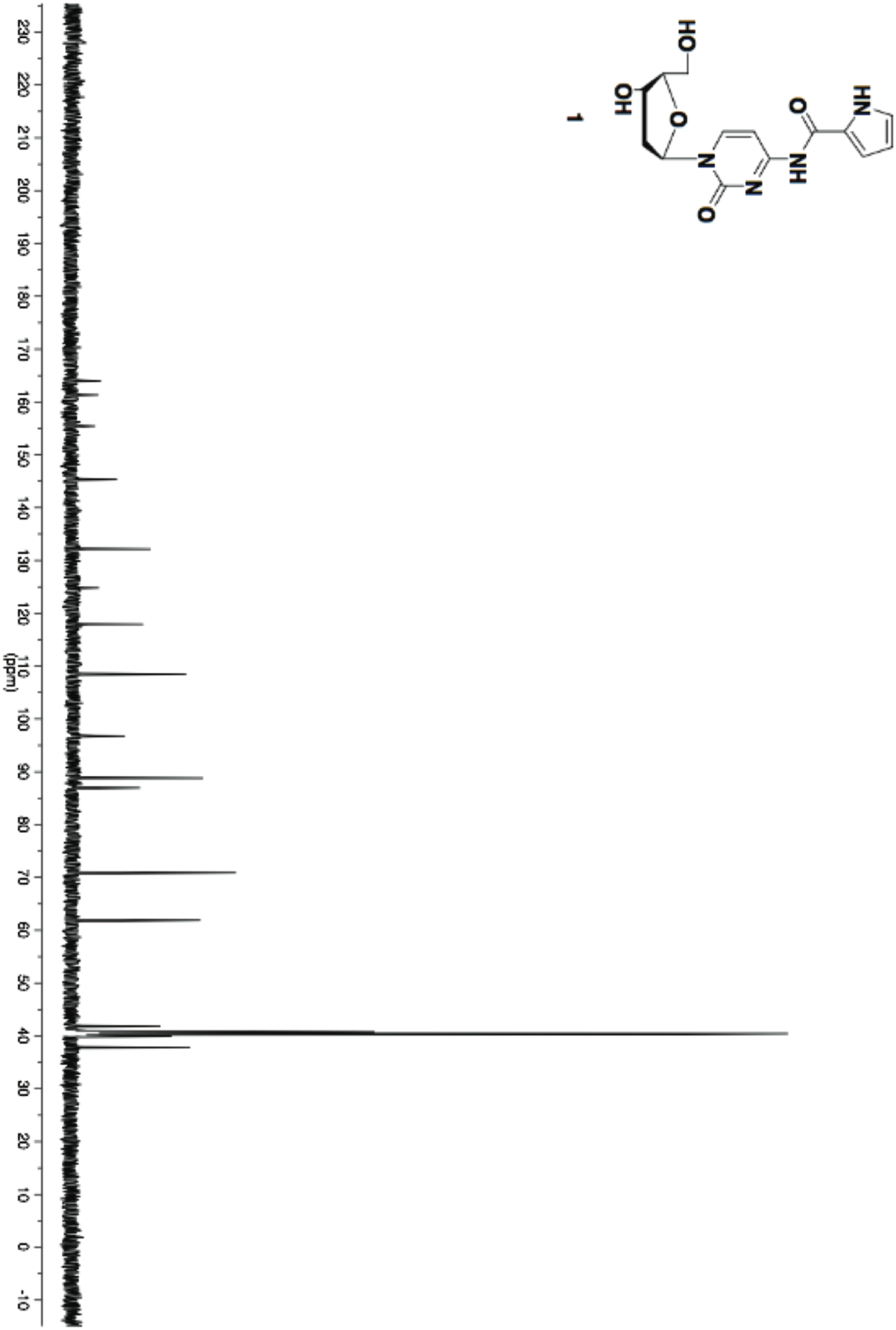


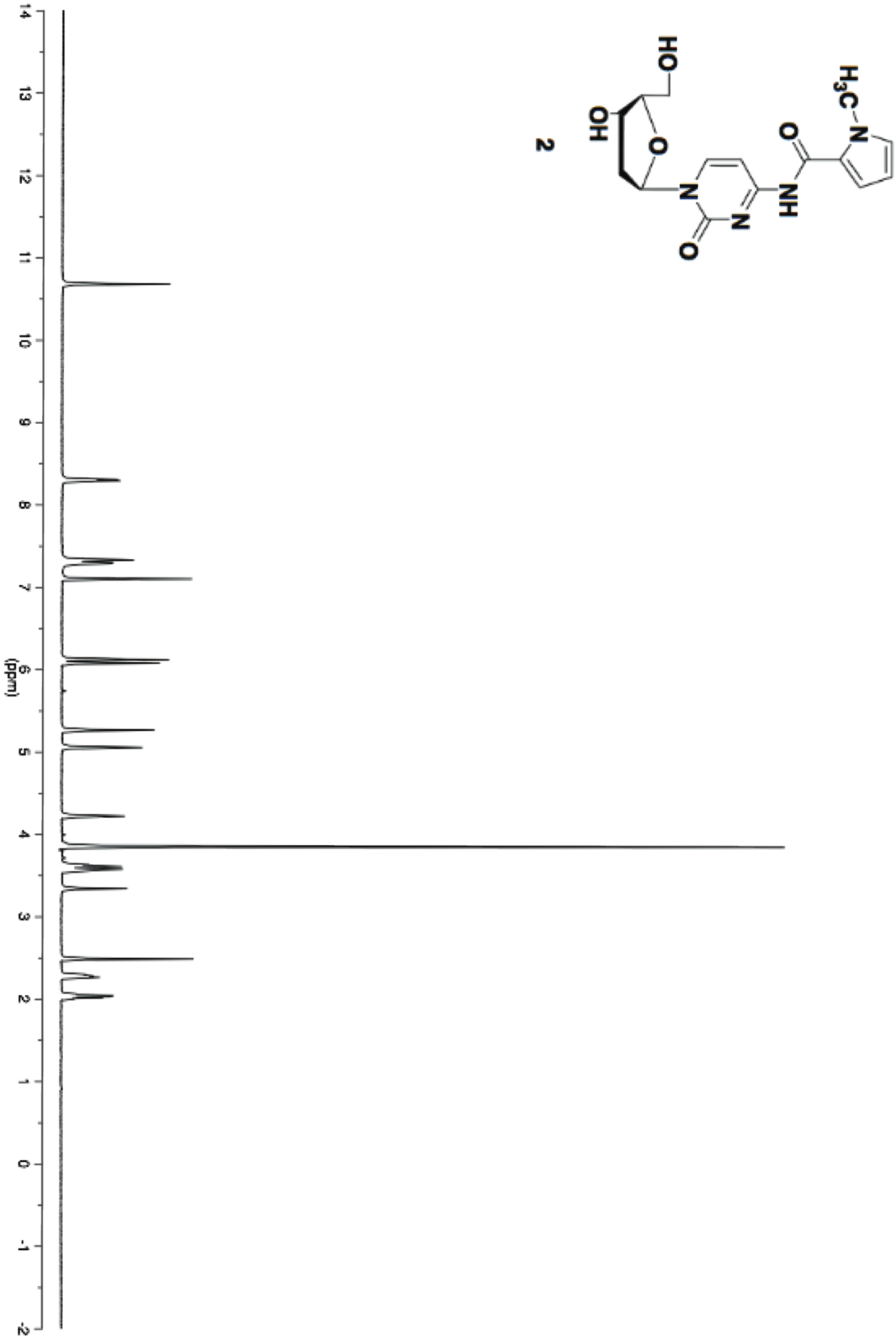
Figure S5. Reverse-phase HPLC analysis of ODN**7** after the treatment of (a) 1.5 M NaOMe/MeOH and (b) 28% NH_3 aq.

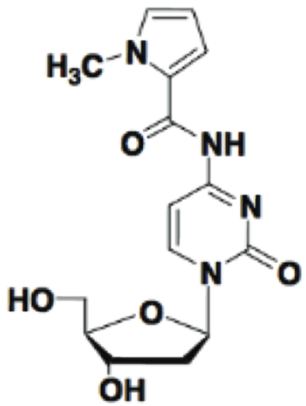


1

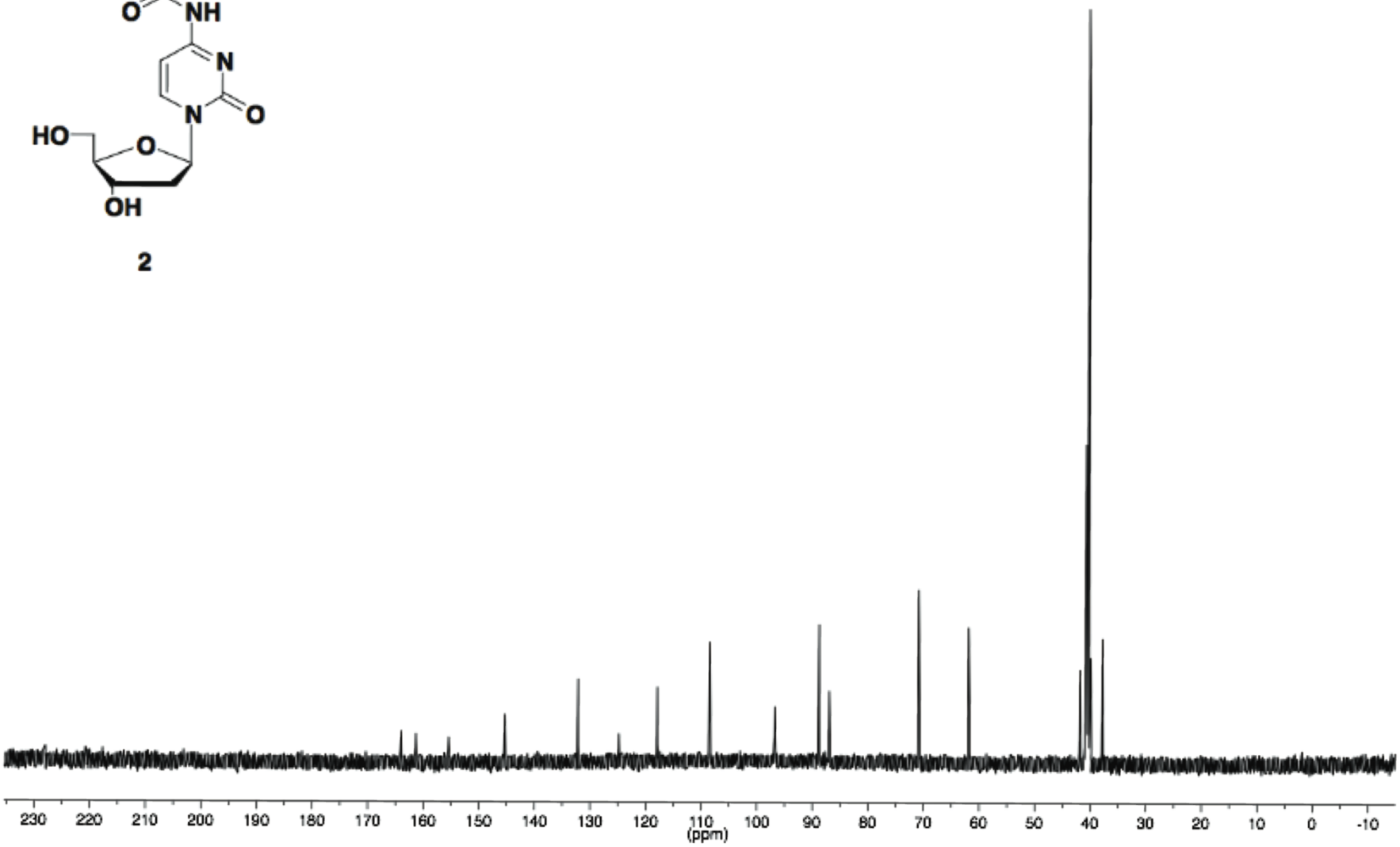


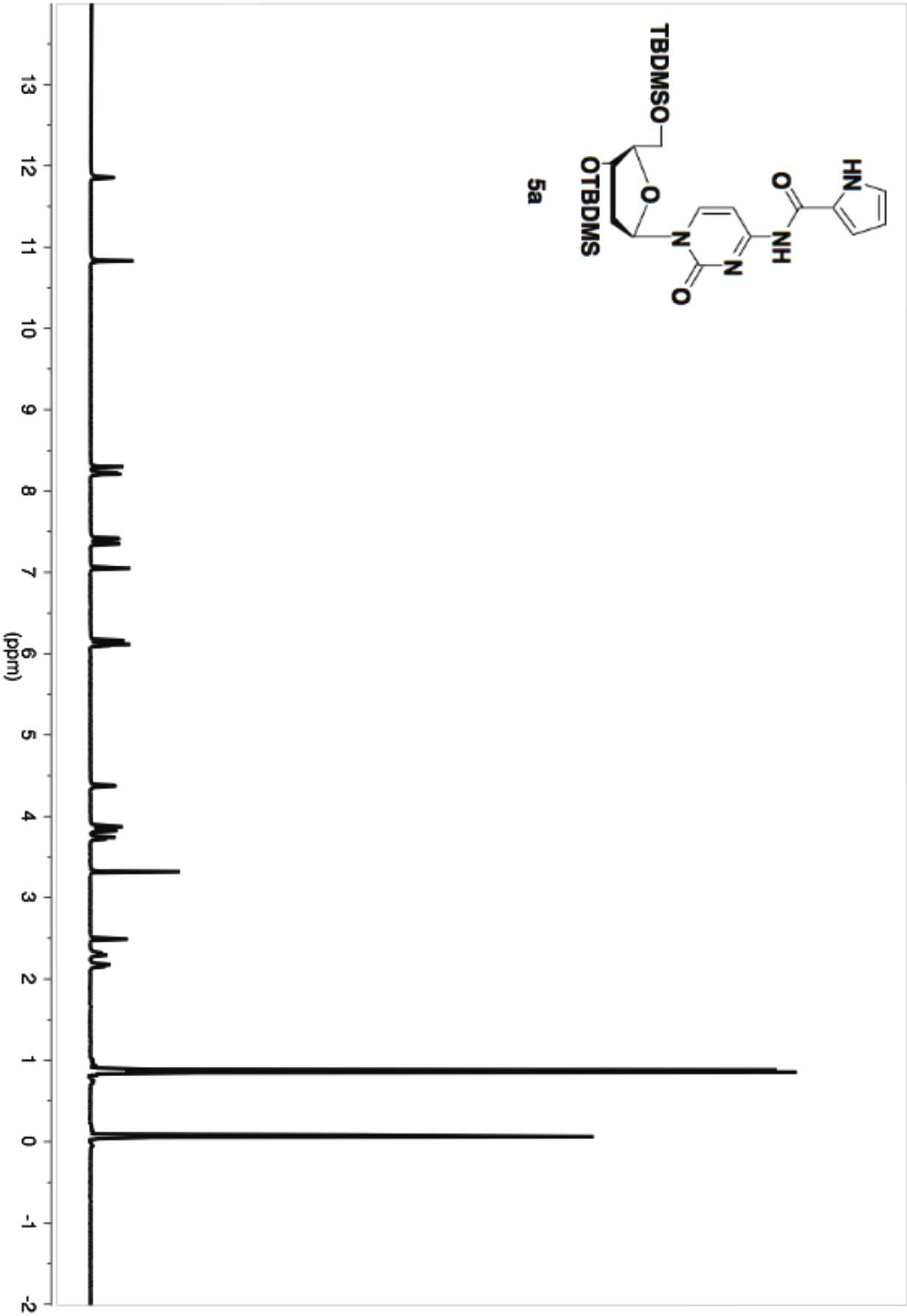

1
HN
O
NH
N
O
N
O
OH
HO
230
220
210
200
190
180
170
160
150
140
130
120
110
(ppm)
100
90
80
70
60
50
40
30
20
10
0
-10

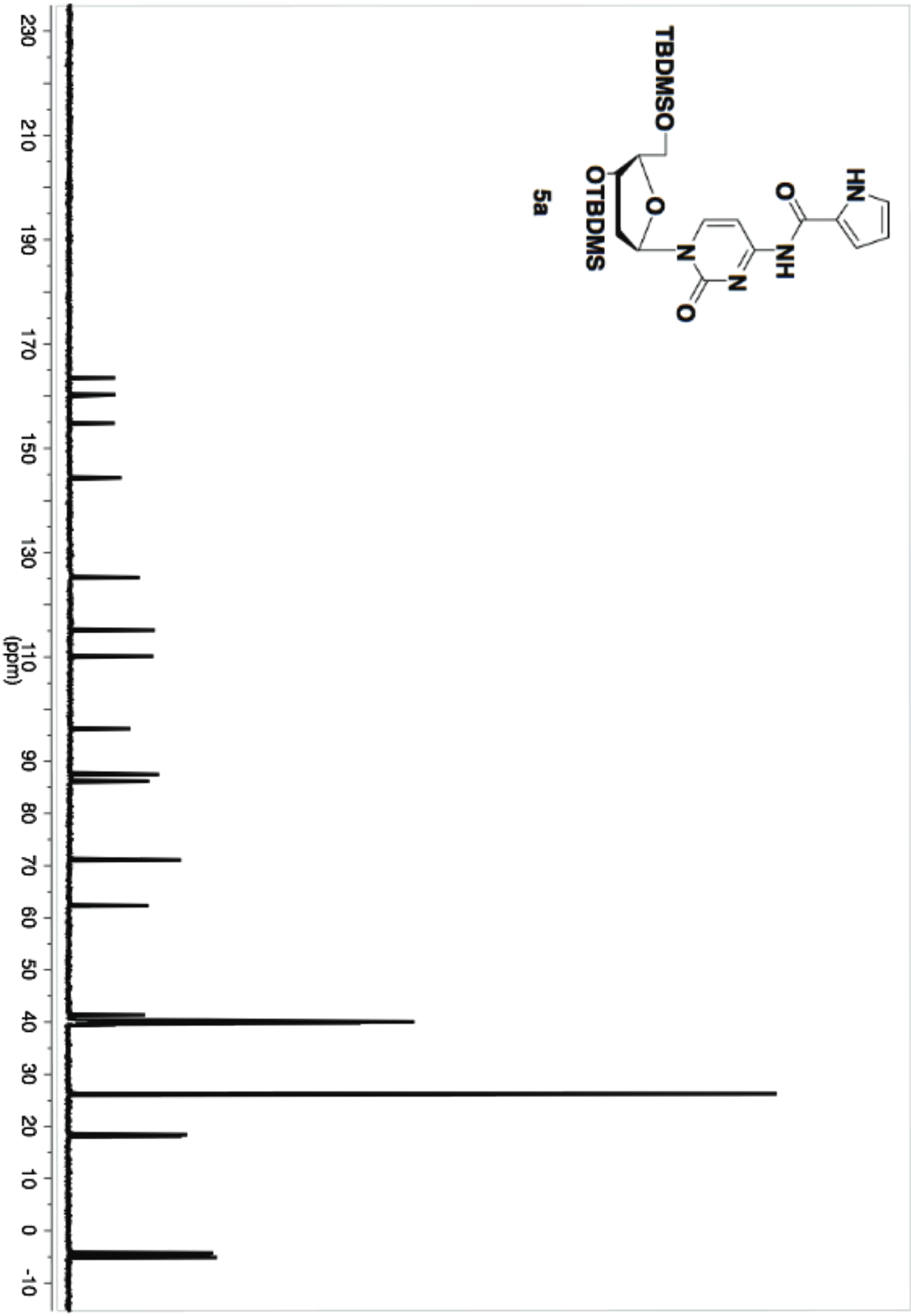

H3C
N
O
NH
N
N
O
O
HO
OH
2
14
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

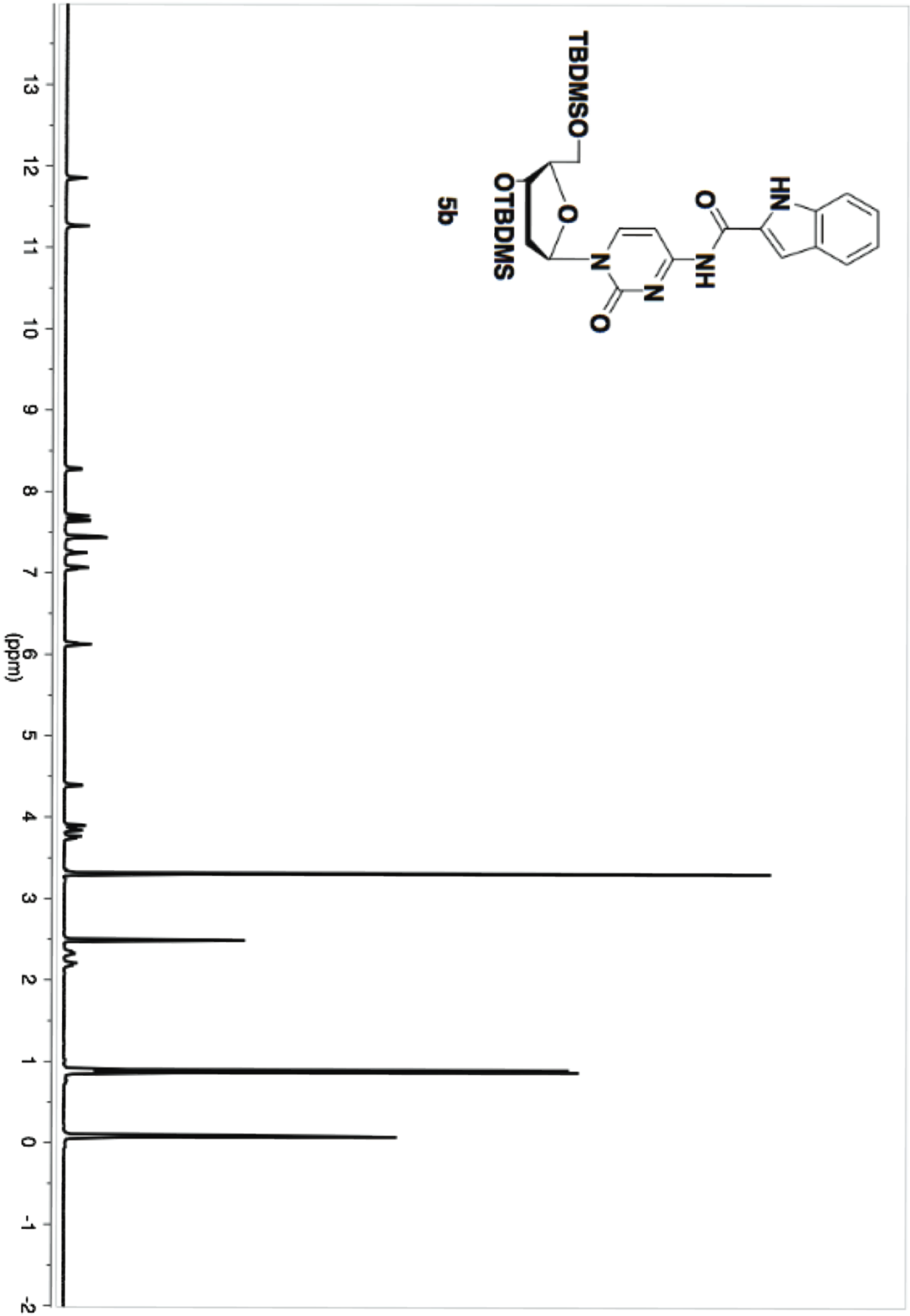

H3C
N
O
NH
N
N
O
HO
O
OH

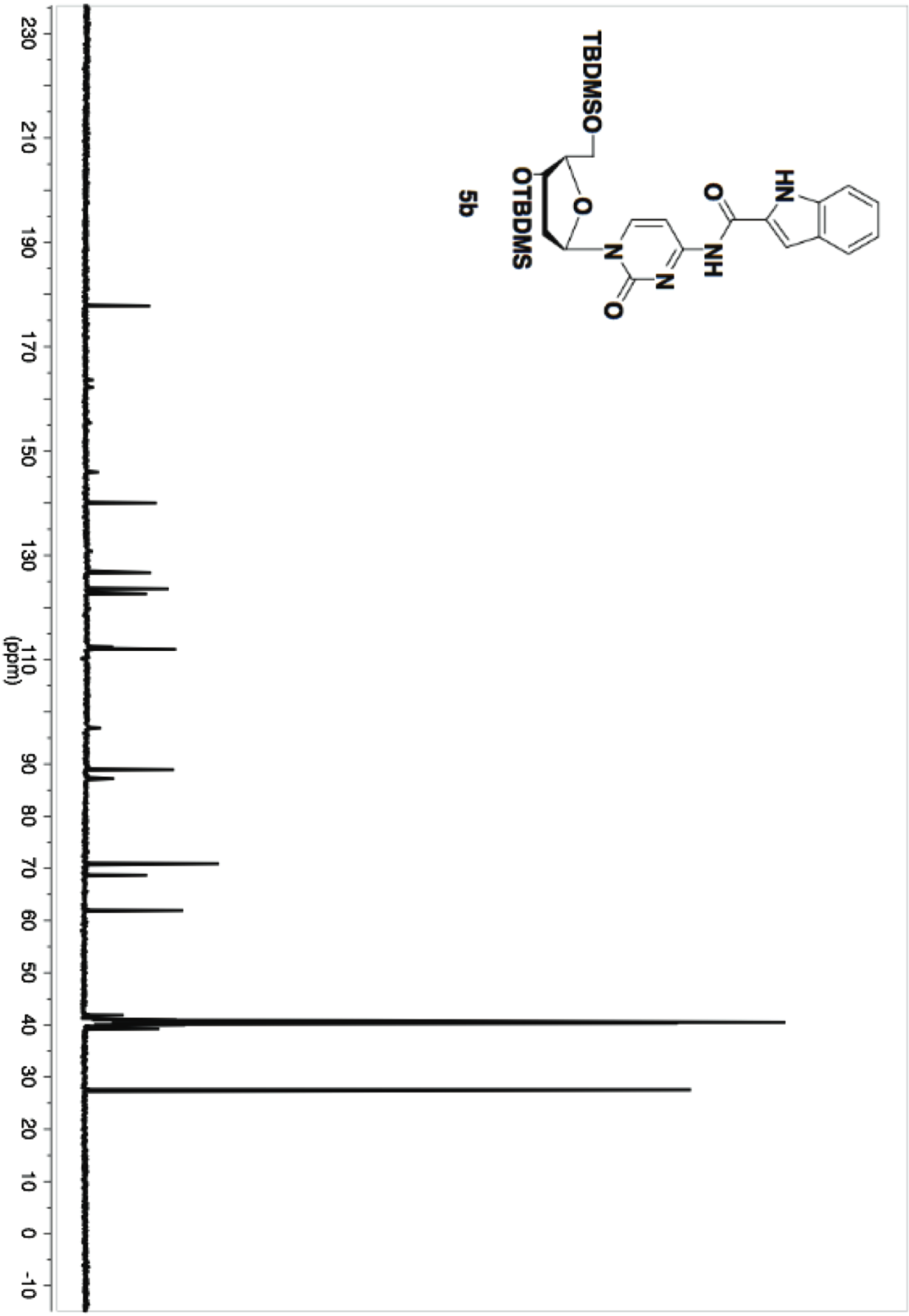
2


230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
(ppm)

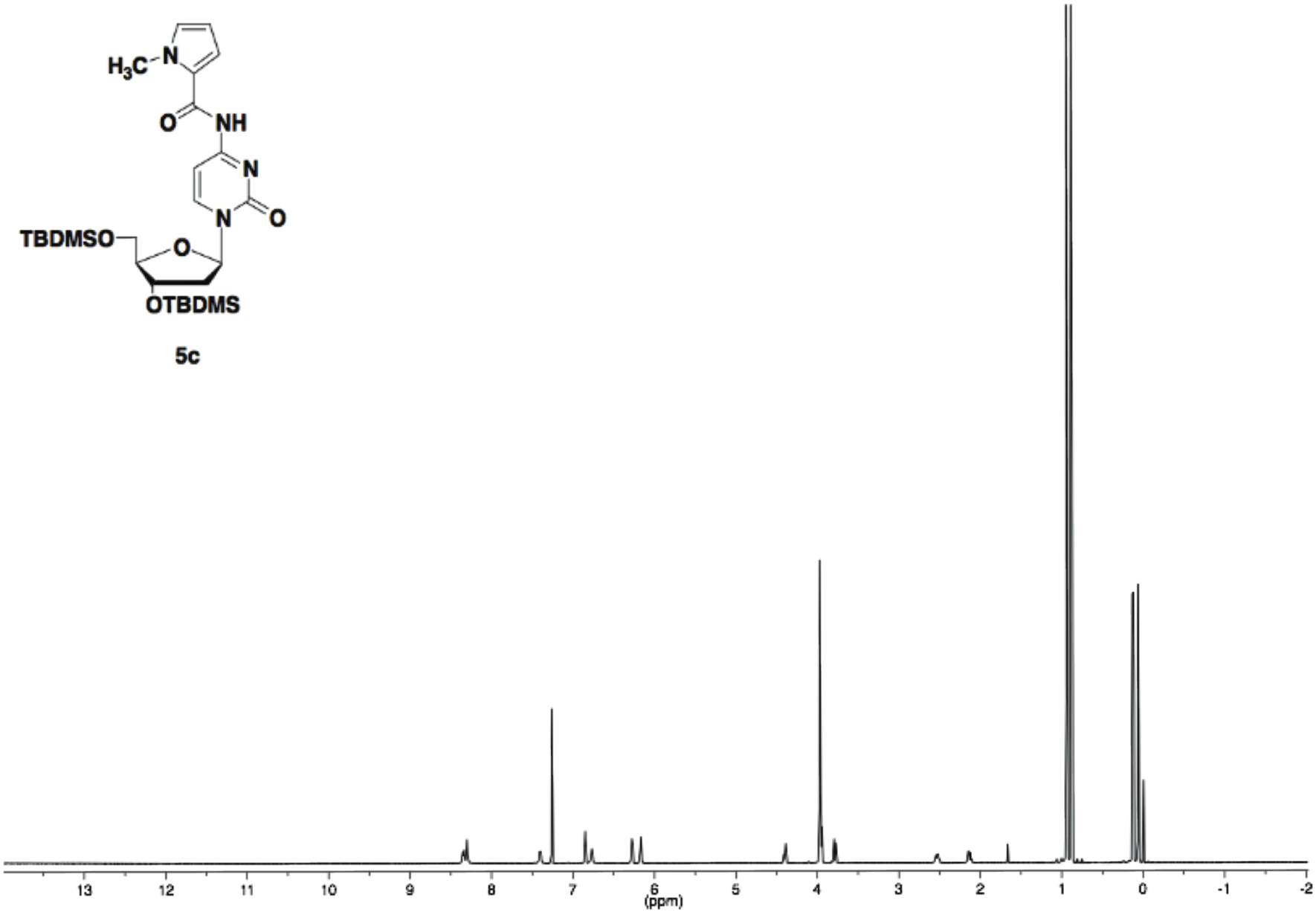

HN
O
NH
N
O
N
O
TBDMSO
OTBDMS
5a
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

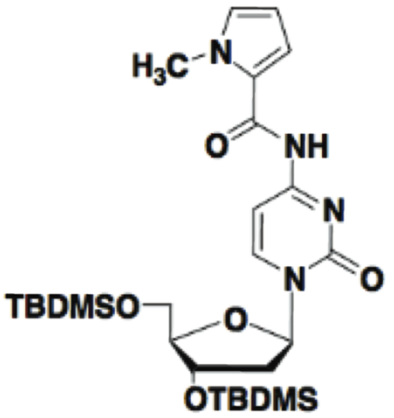

HN
O
NH
N
N
O
O
TBDMSO
OTBDMS
5a
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10


HN
O
NH
N
N
O
O
N
TBDMSO
OTBDMS
5b
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

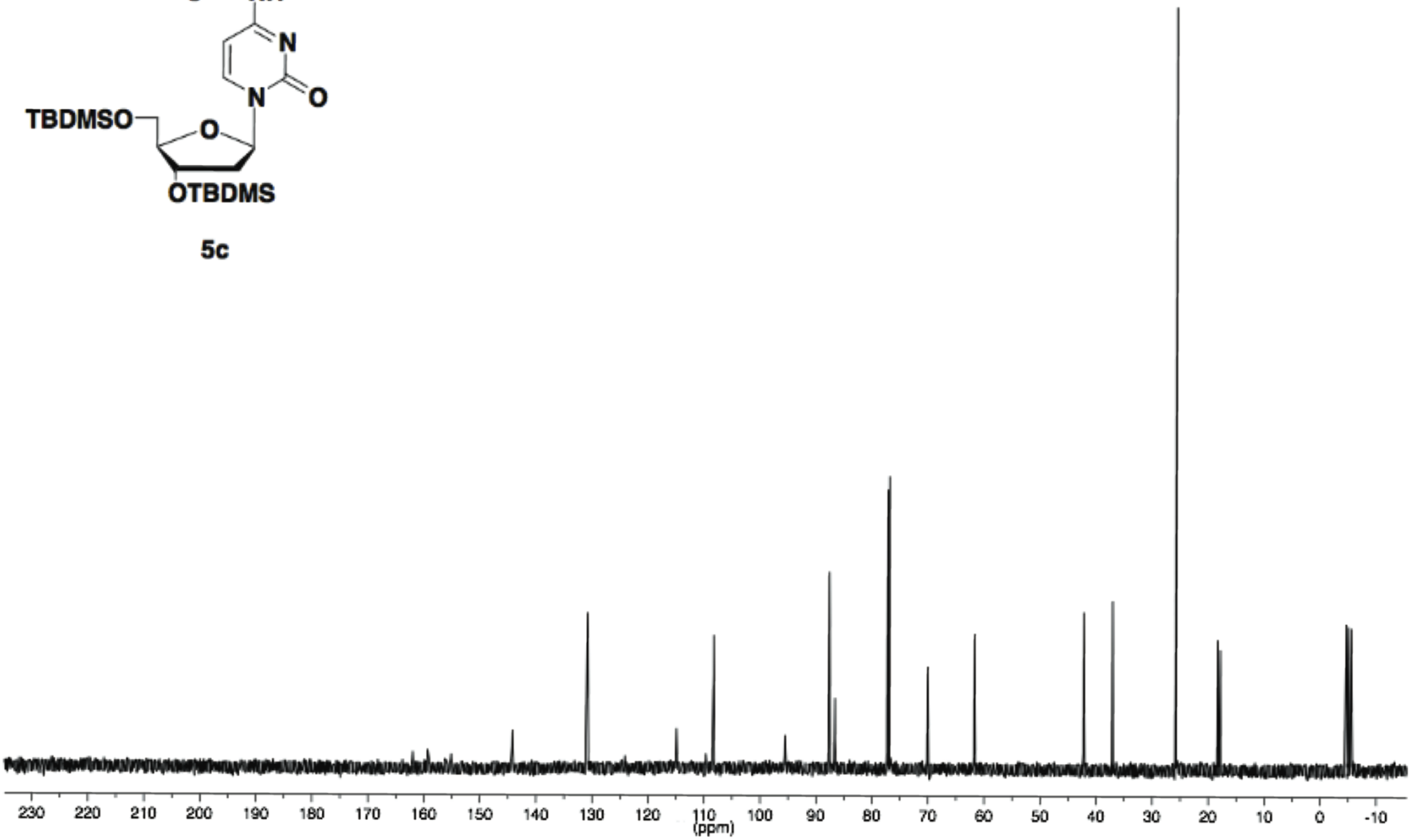

HN
O
NH
N
O
N
O
TBDMSO
O
OTBDMS
5b
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

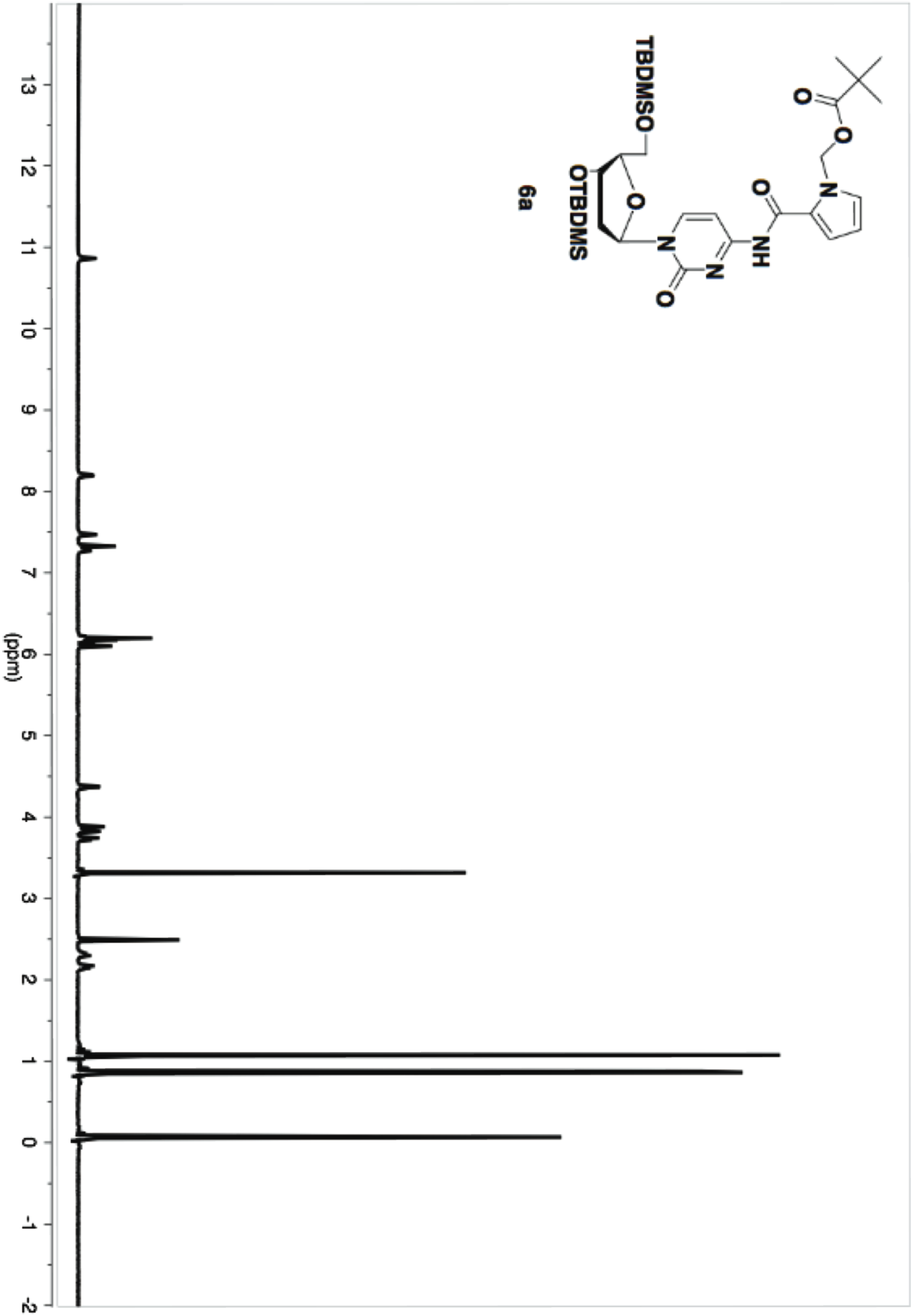
5c

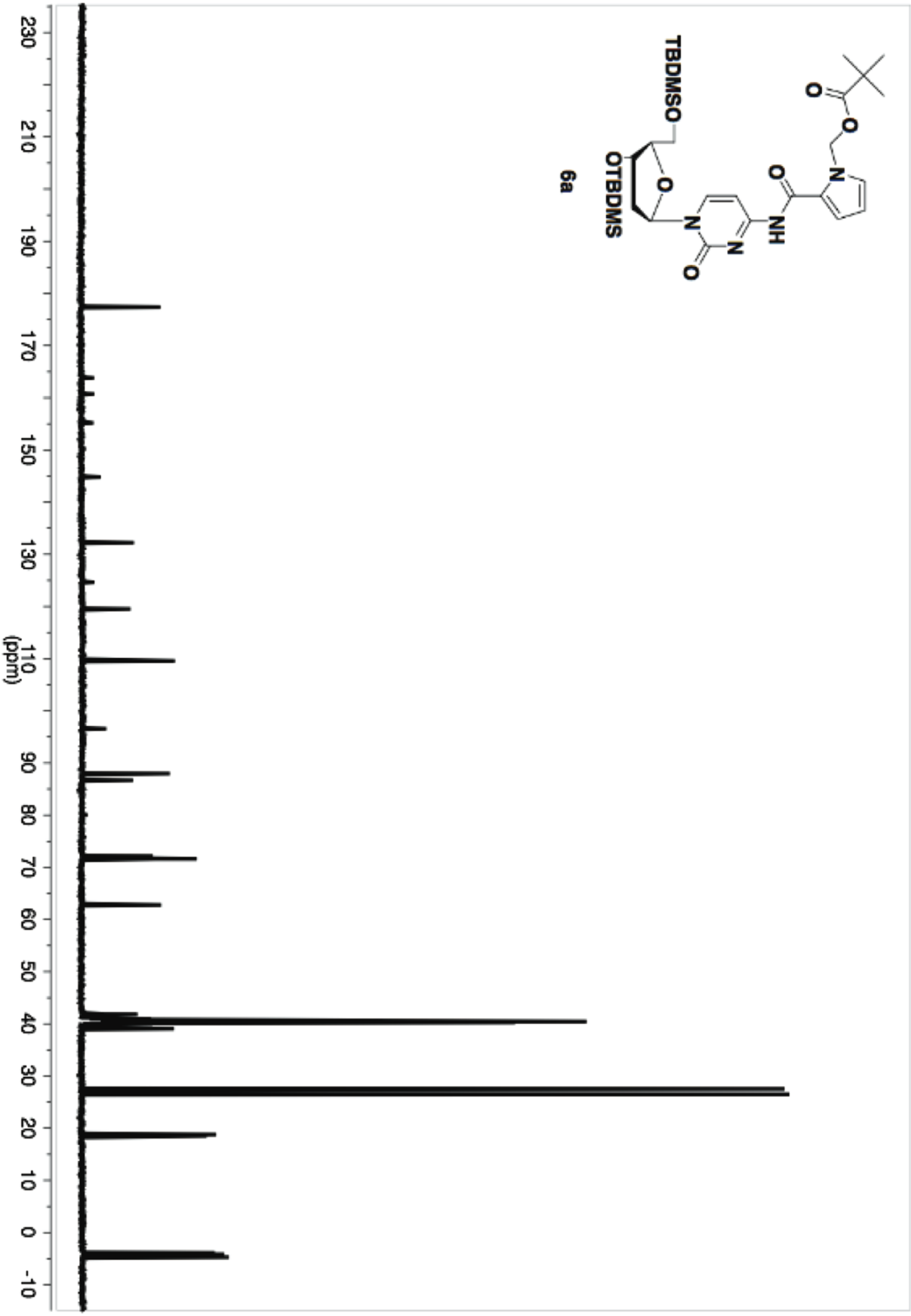


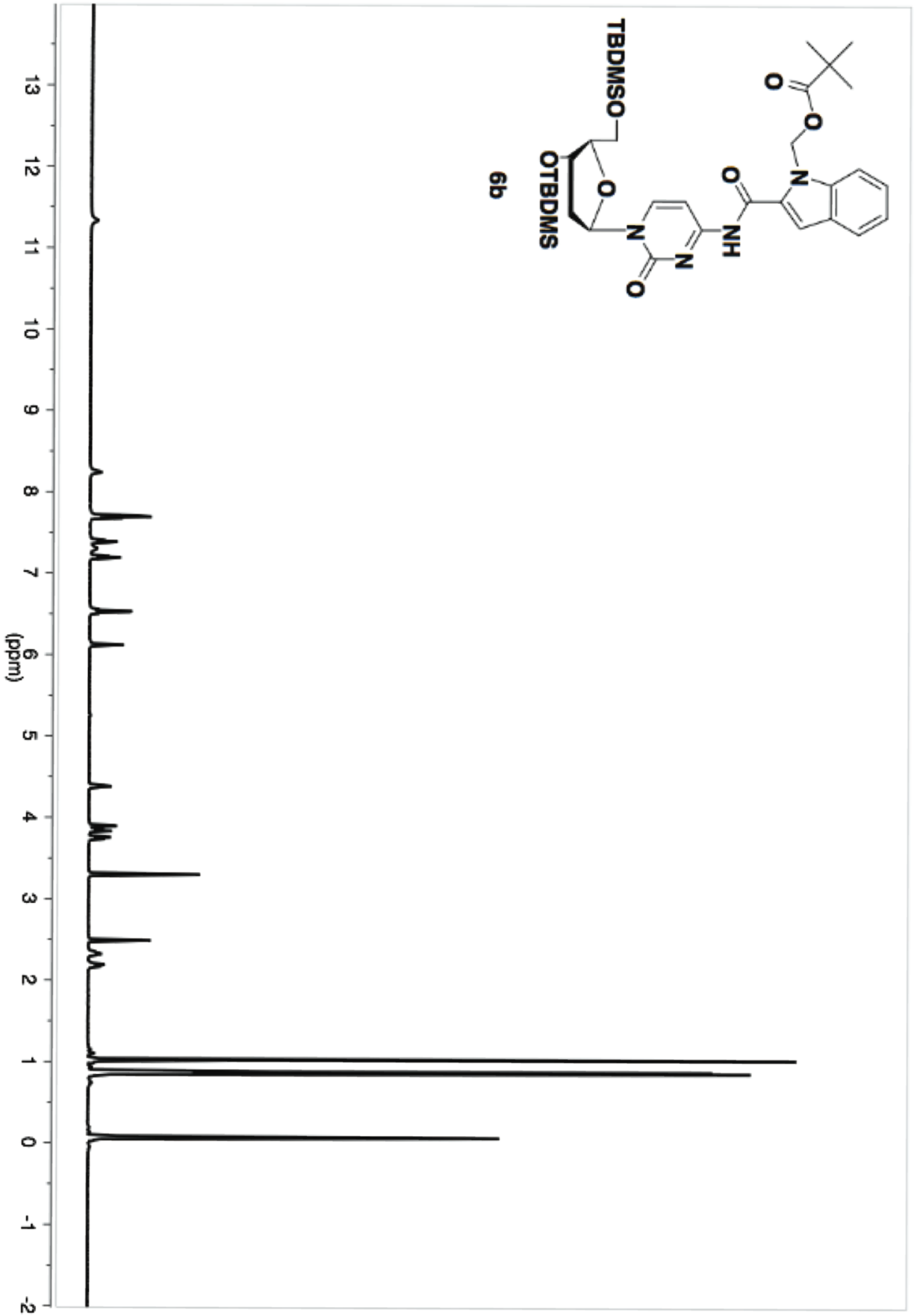

H3C
N
O
NH
N
N
O
TBDMSO
O
OTBDMS

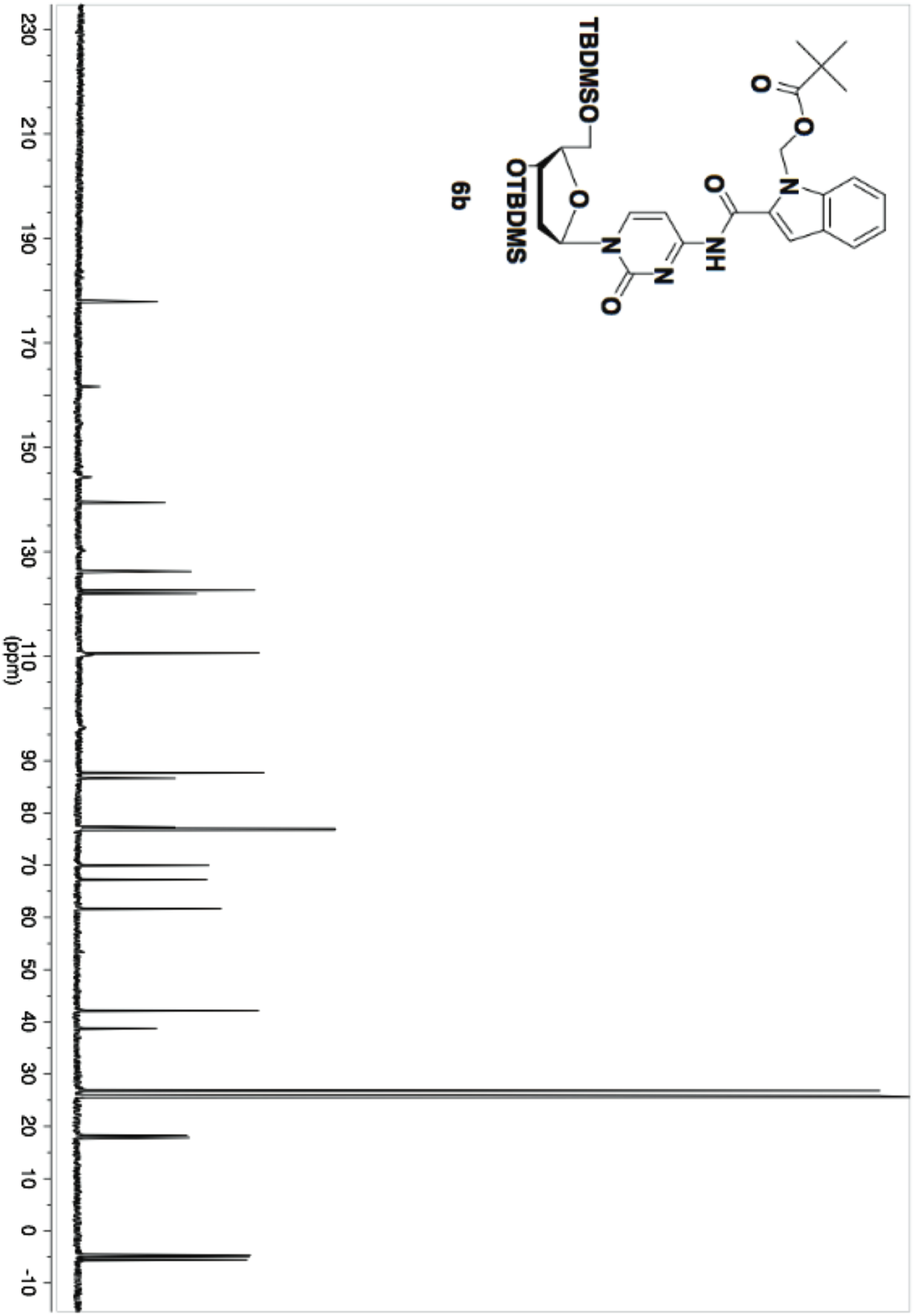
5c

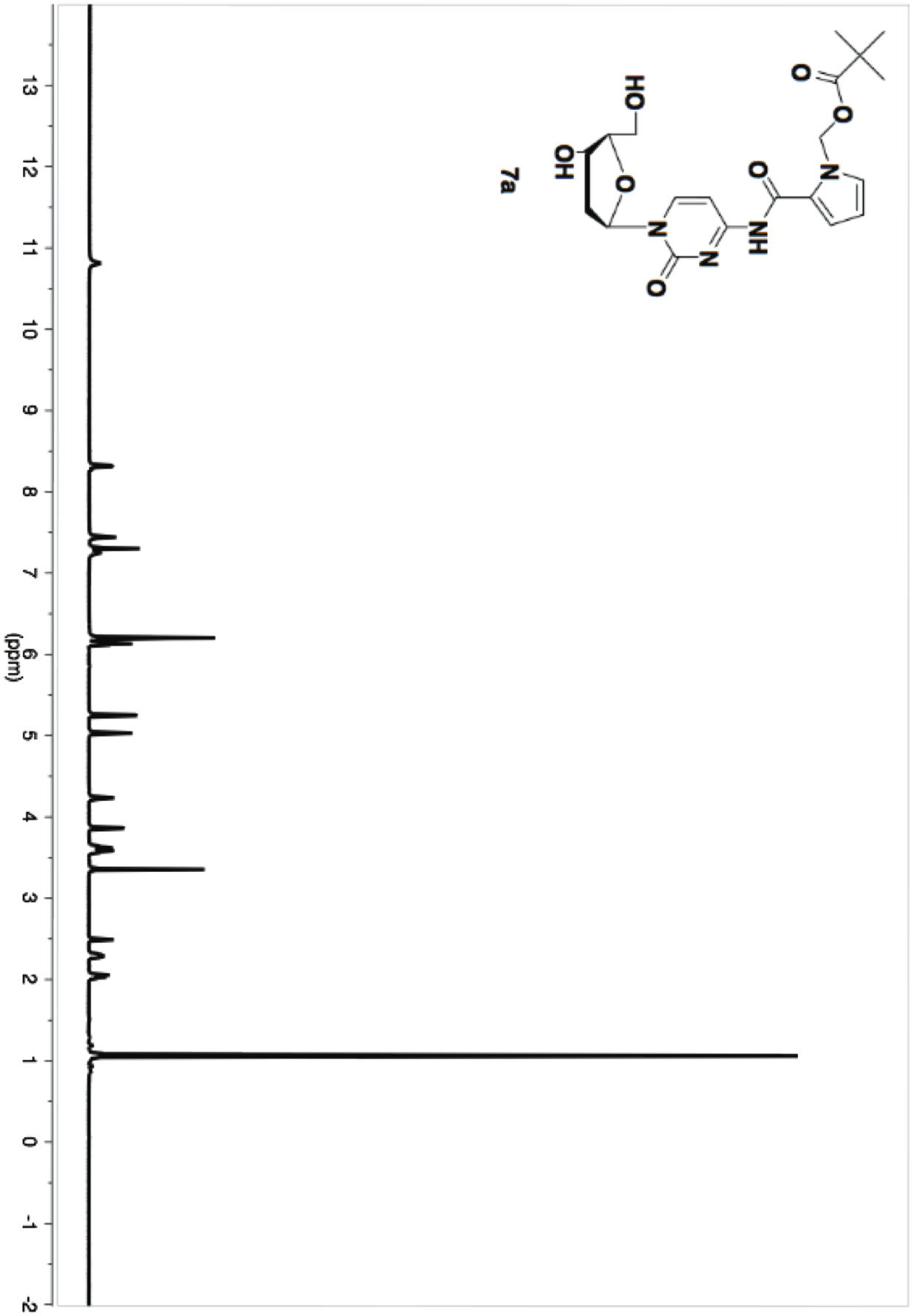

230
220
210
200
190
180
170
160
150
140
130
120
110
(ppm)
100
90
80
70
60
50
40
30
20
10
0
-10

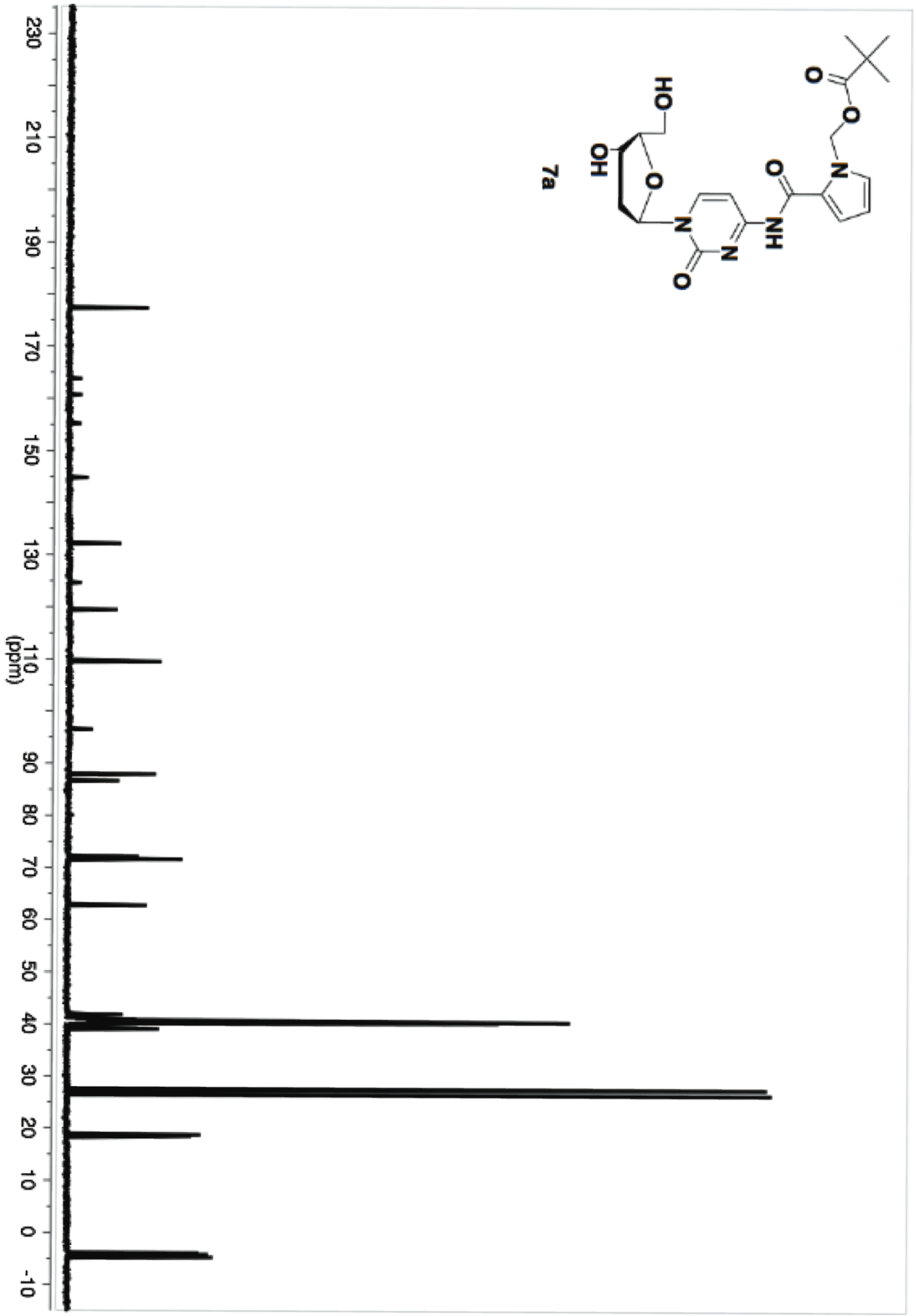

TBDMSO
OTBDMS
6a
(ppm)

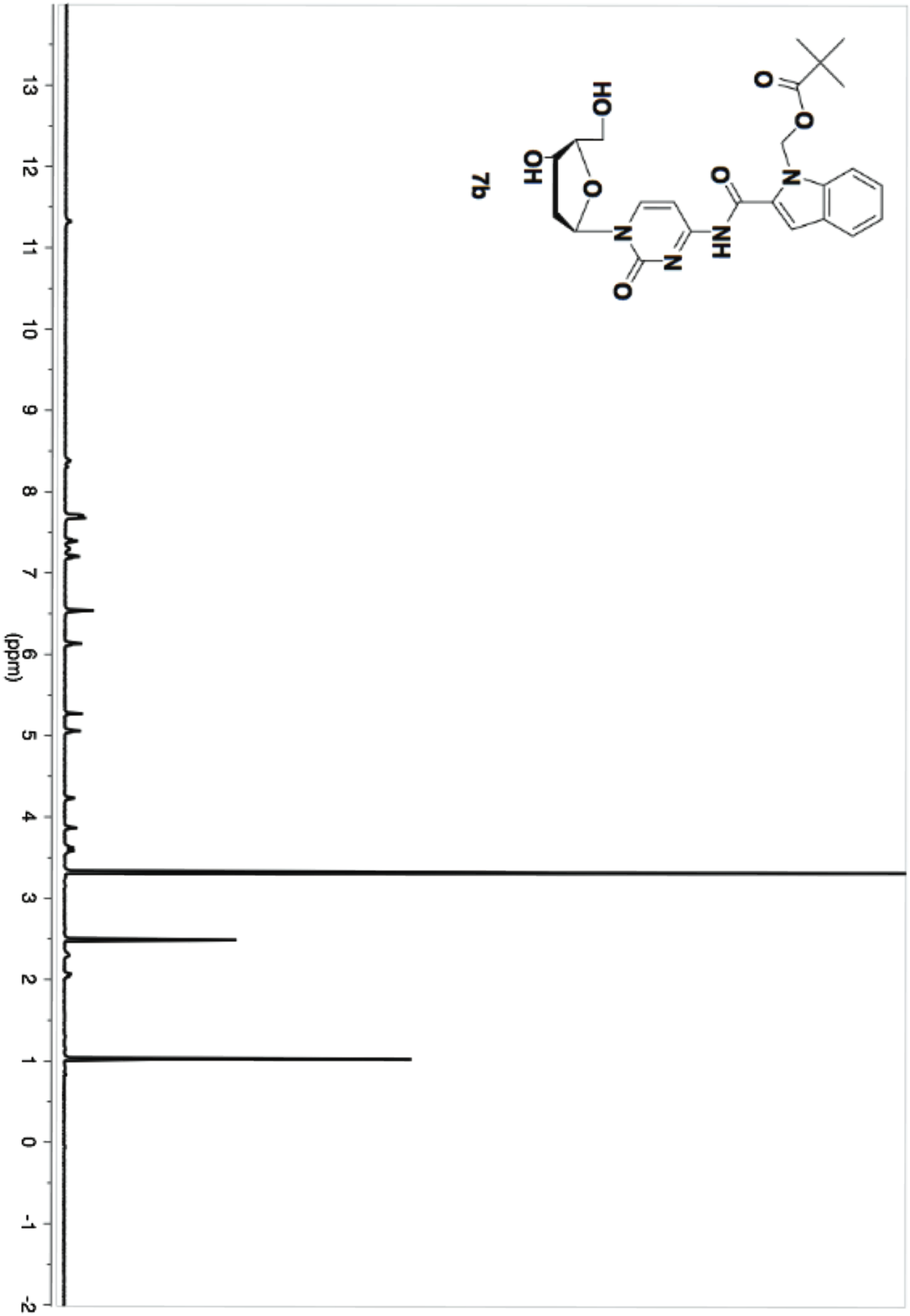

TBDMSO
OTBDMS
6a
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

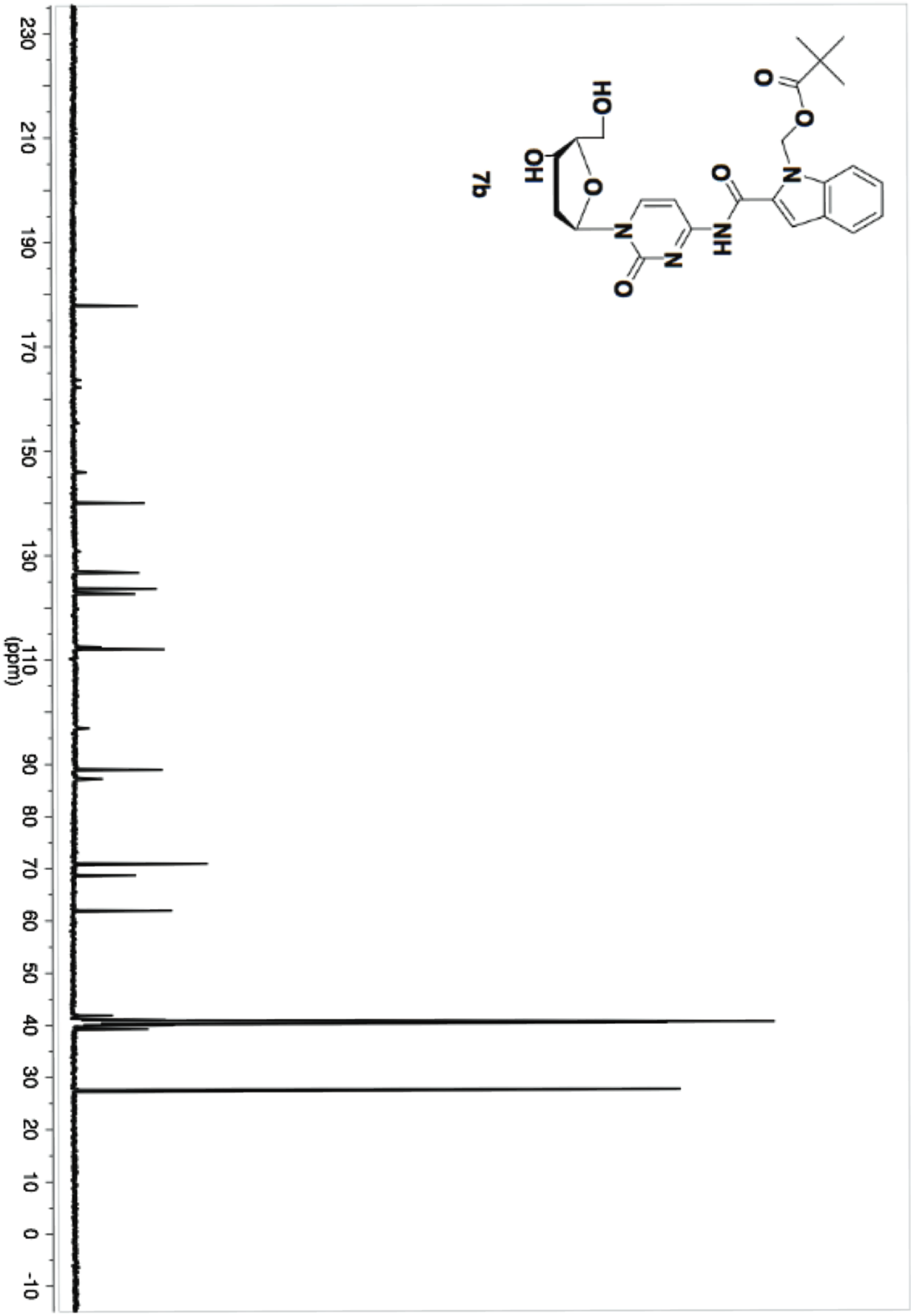

O
O
N
O
NH
N
N
O
O
TBDMSO
OTBDMS
6b
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

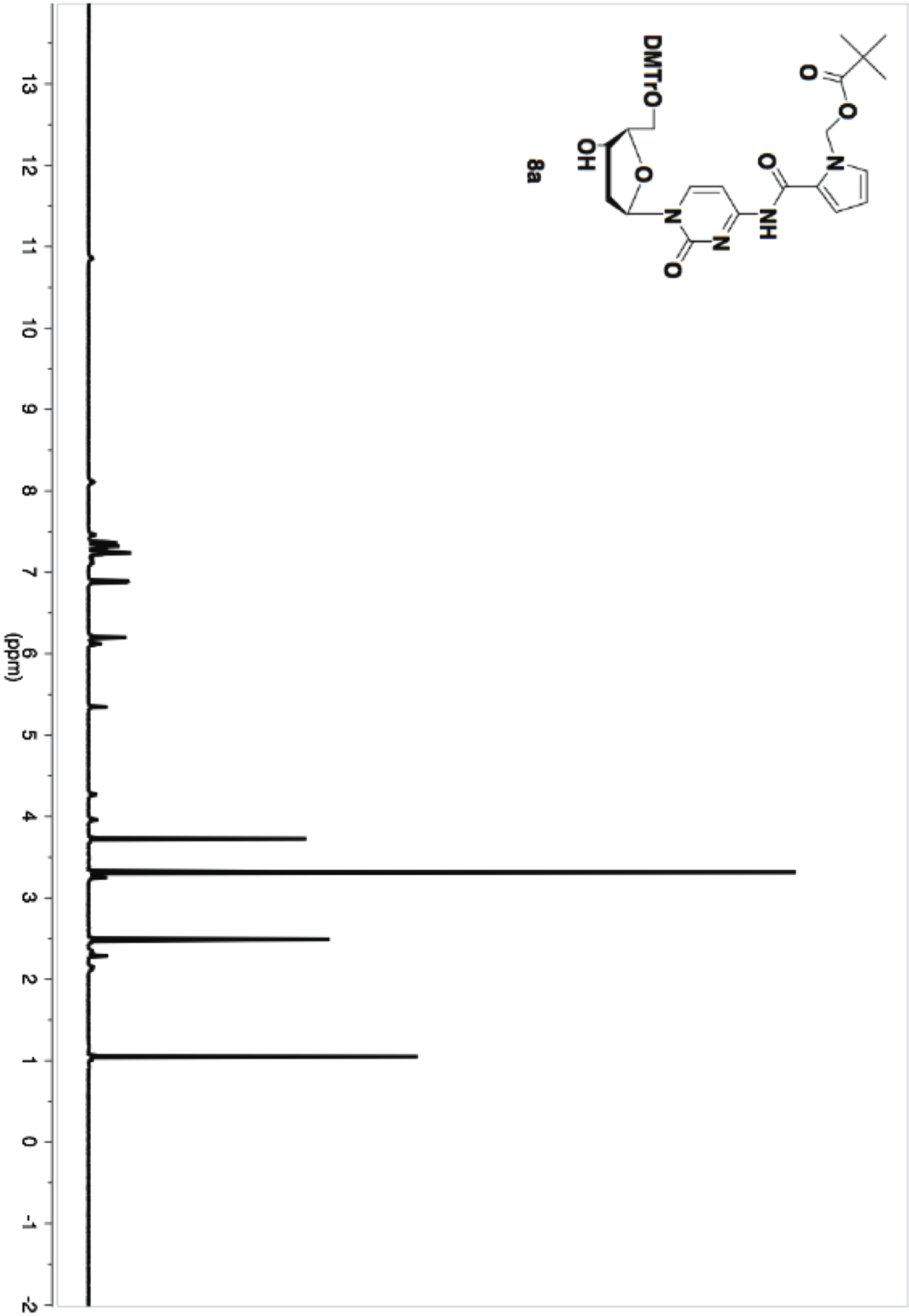

TBDMSO
OTBDMS
6b
(ppm)

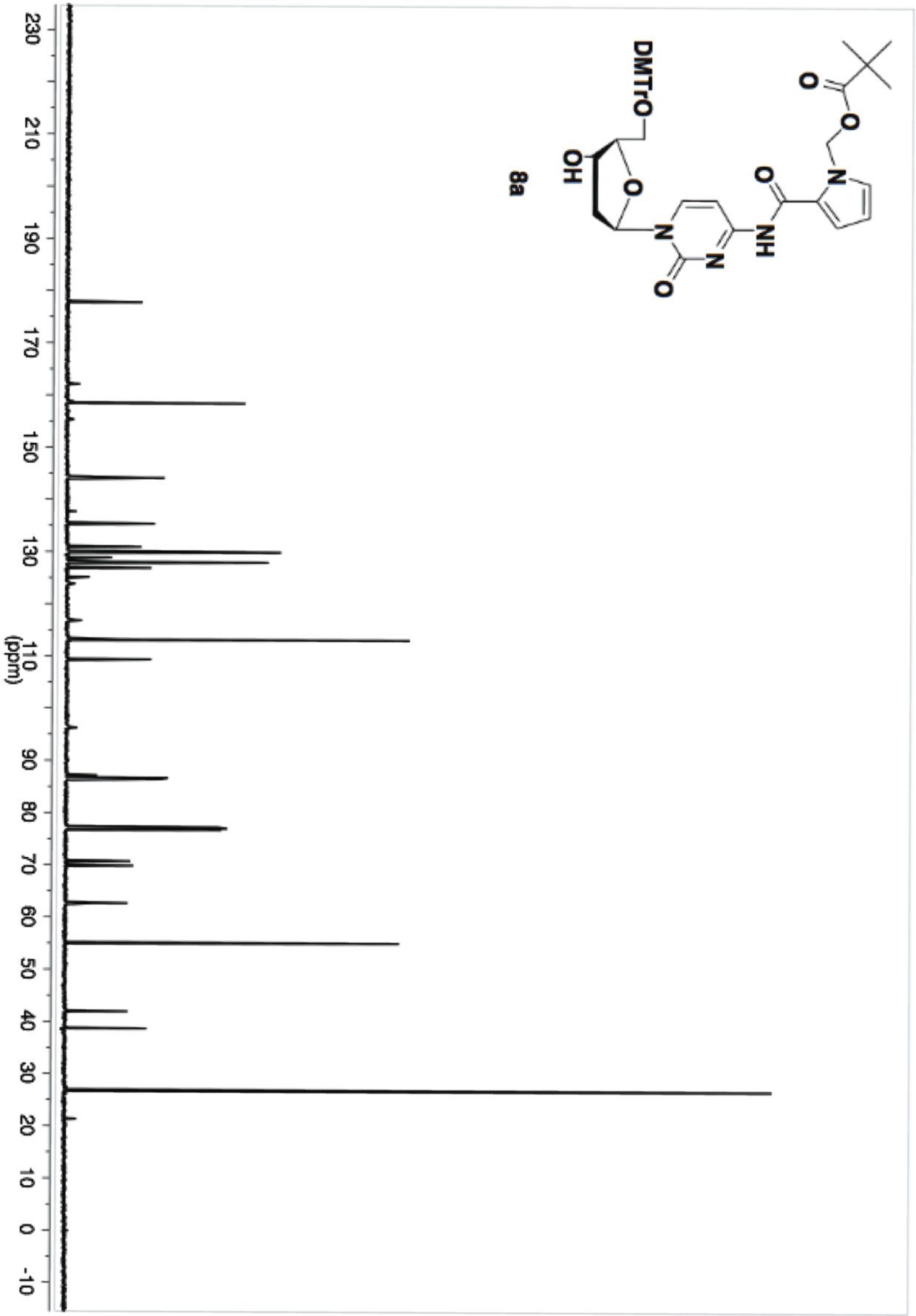

7a
(ppm)

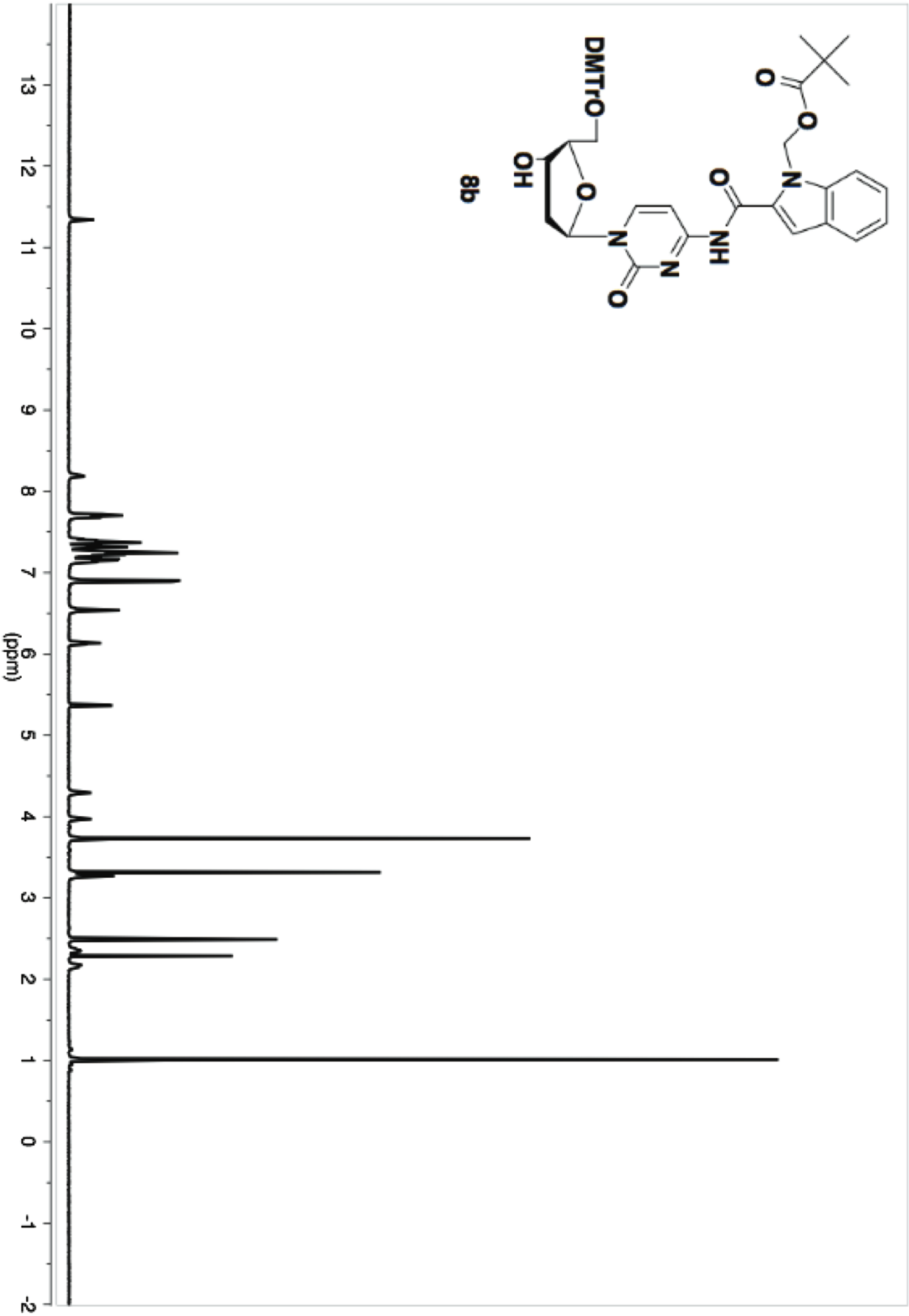

7a
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

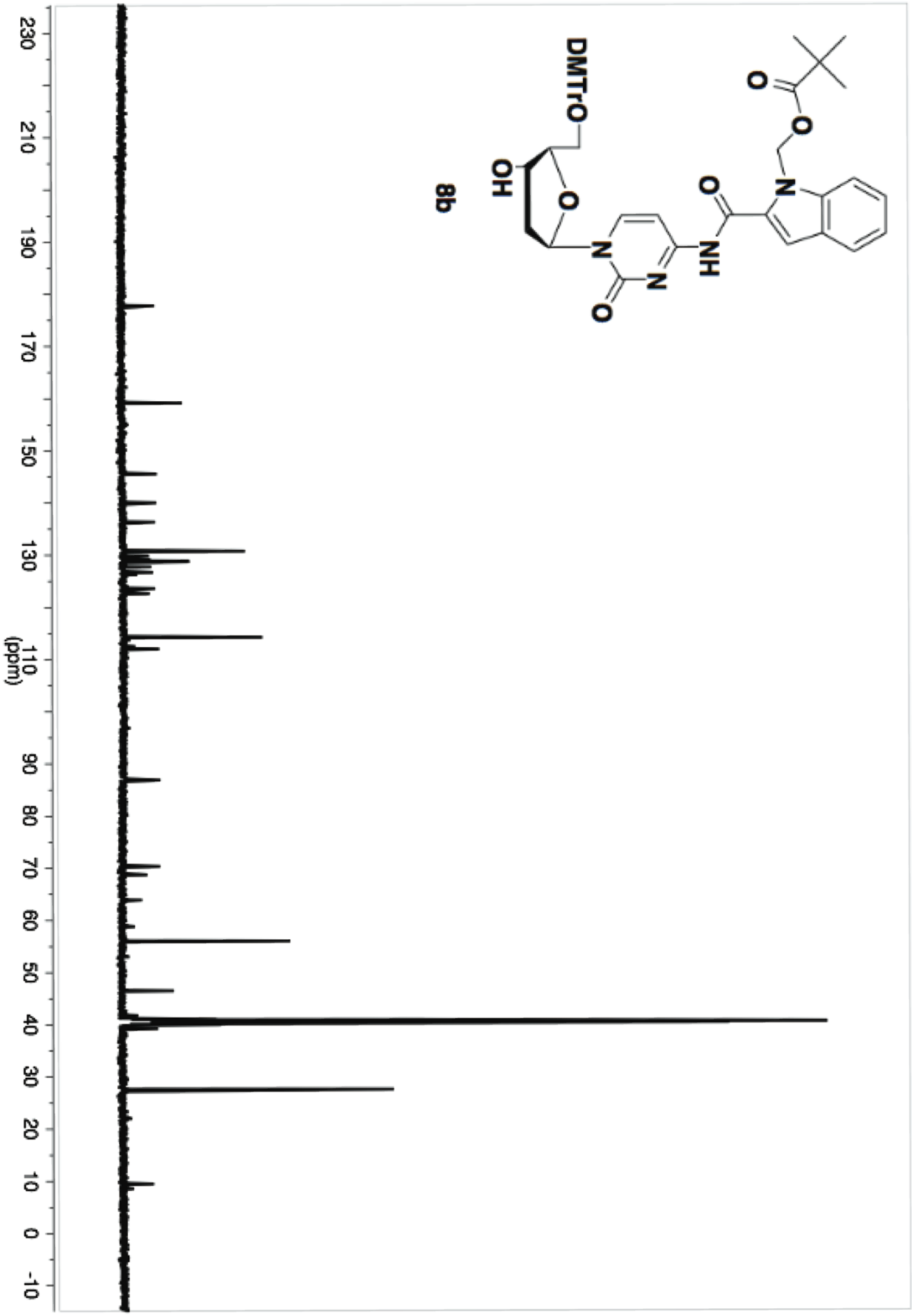

7b
(ppm)

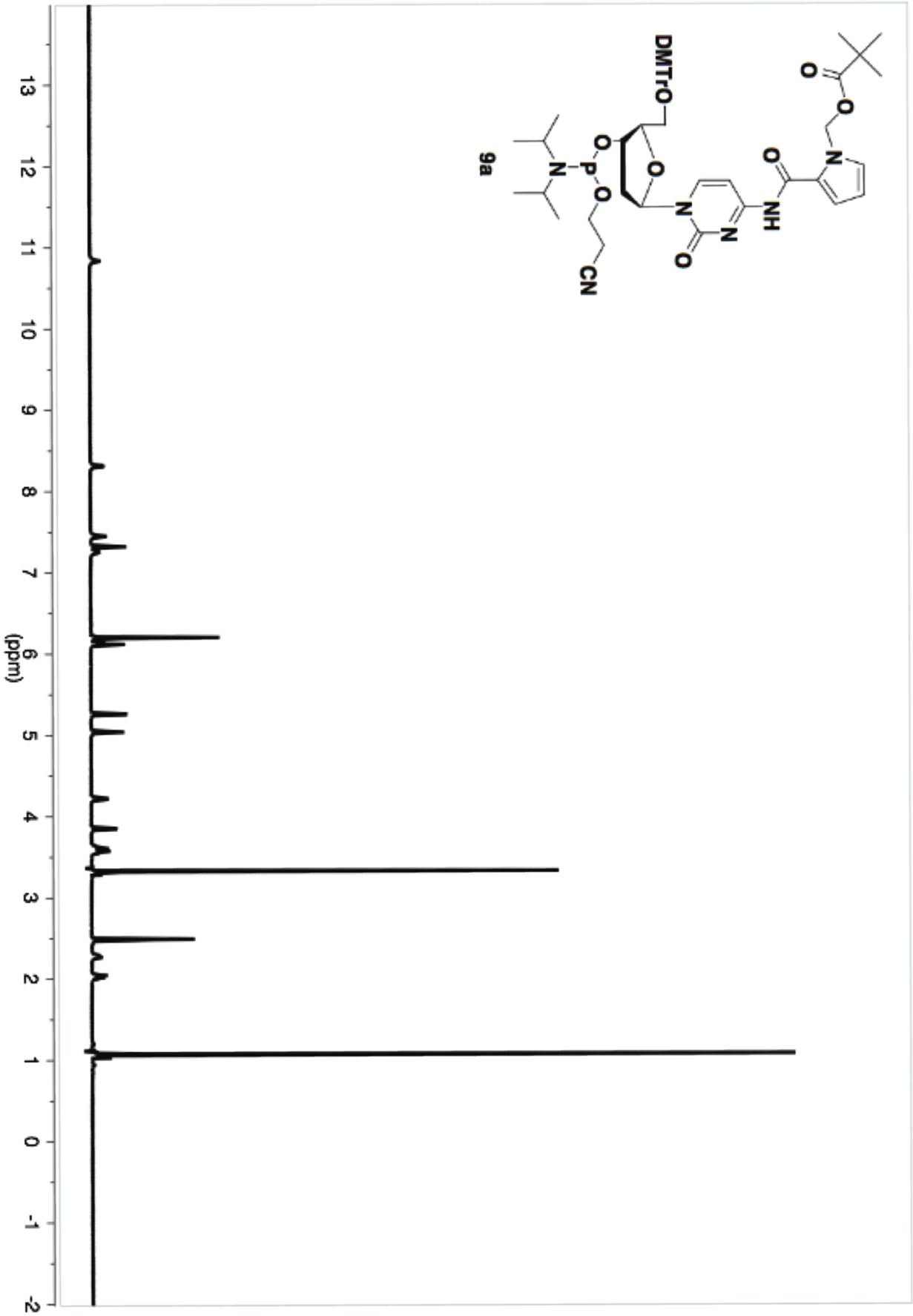

7b
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

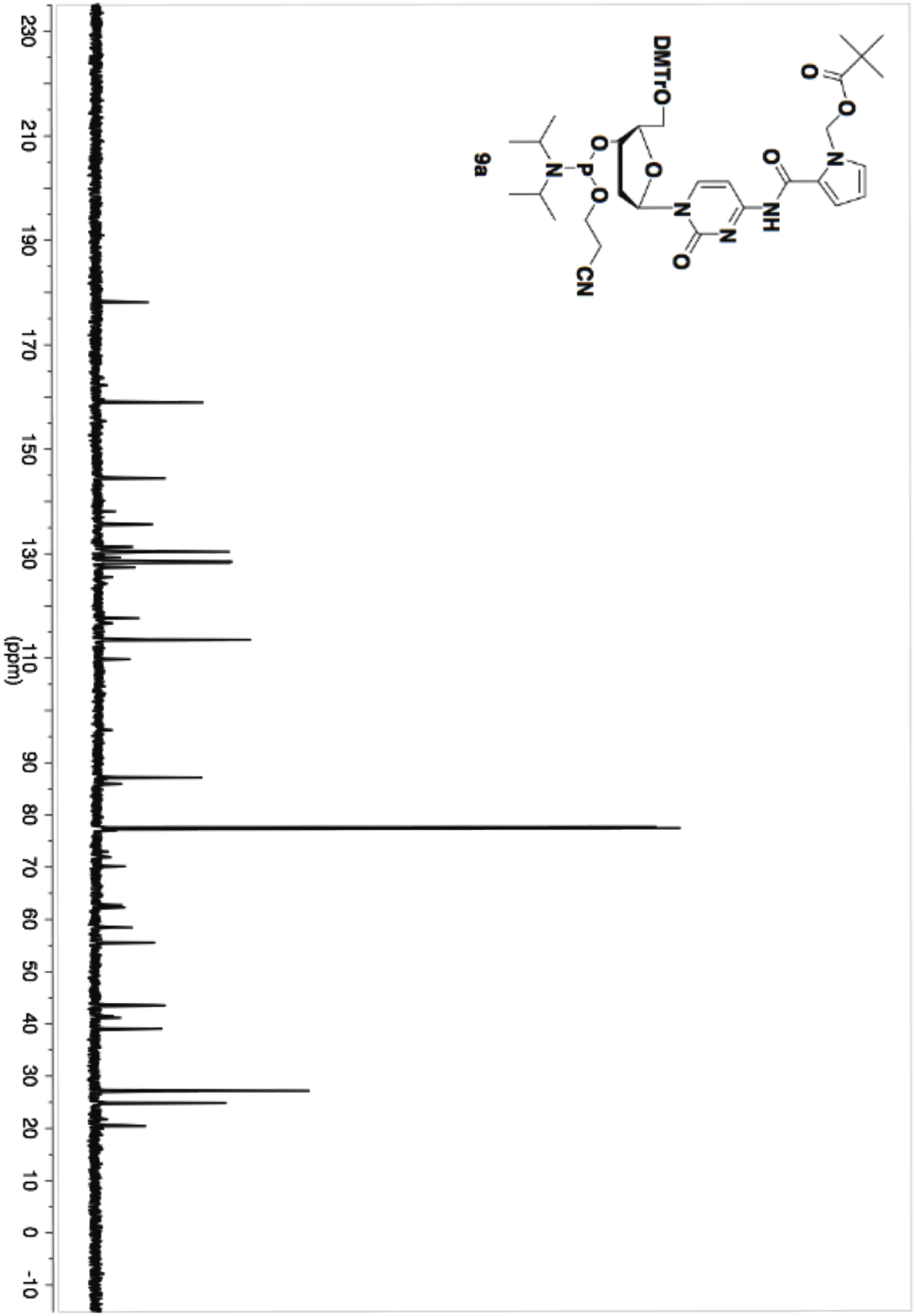

8a
DMTrO
OH
O
N
O
NH
N
O
N
O
O
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

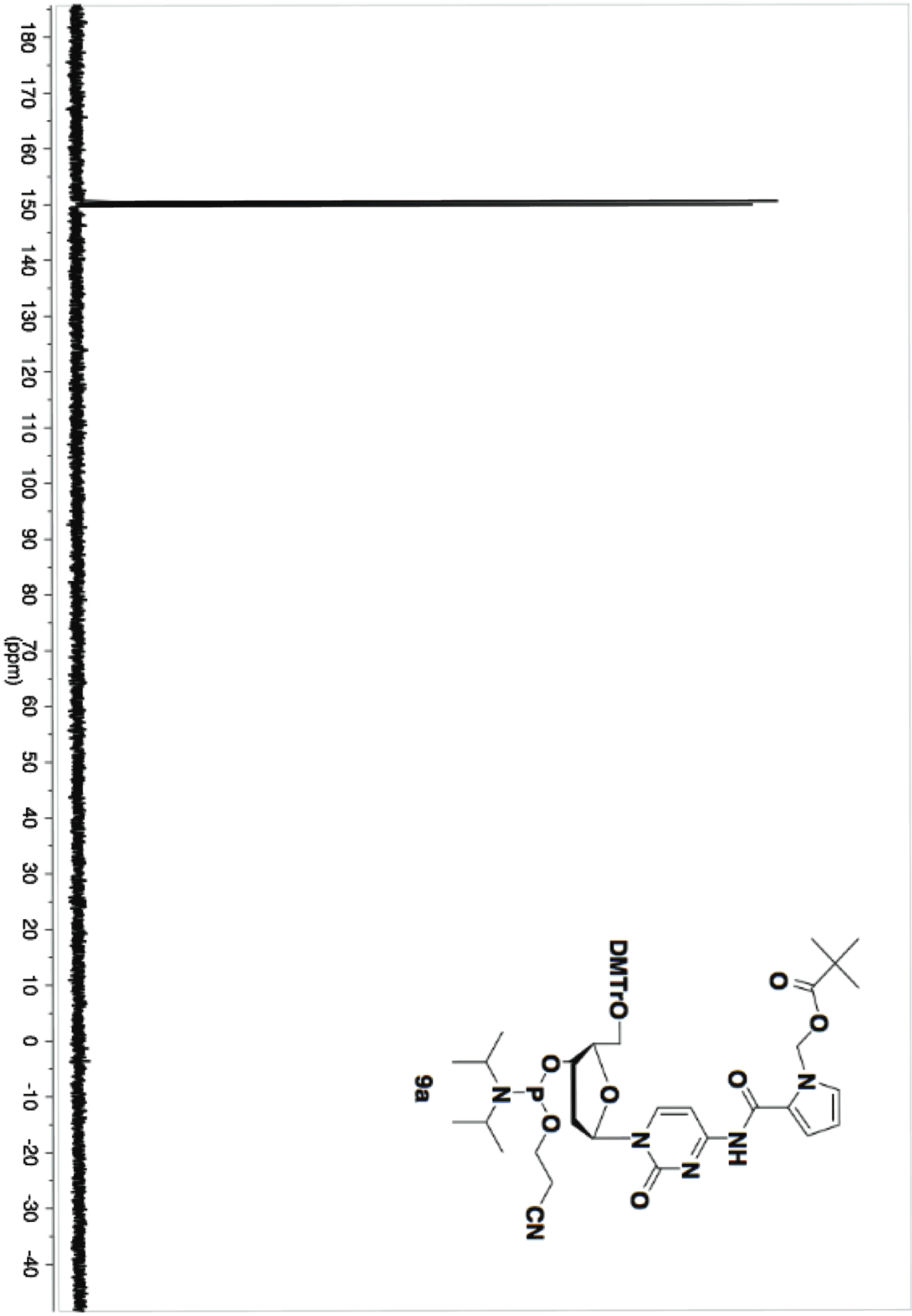

DMTrO
OH
8a
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

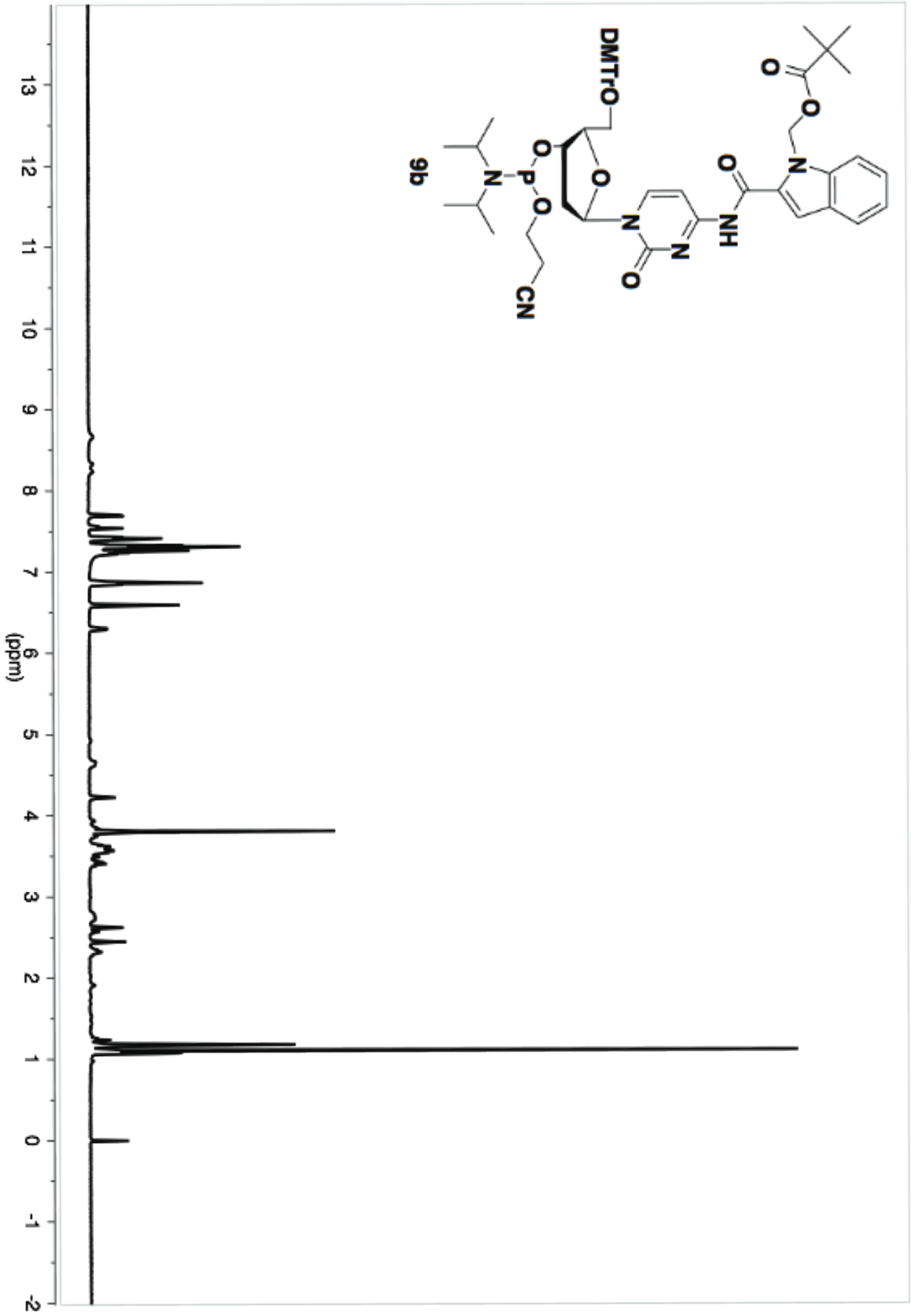

8b
DMTrO
OH
O
N
O
N
NH
O
O
O
N
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2

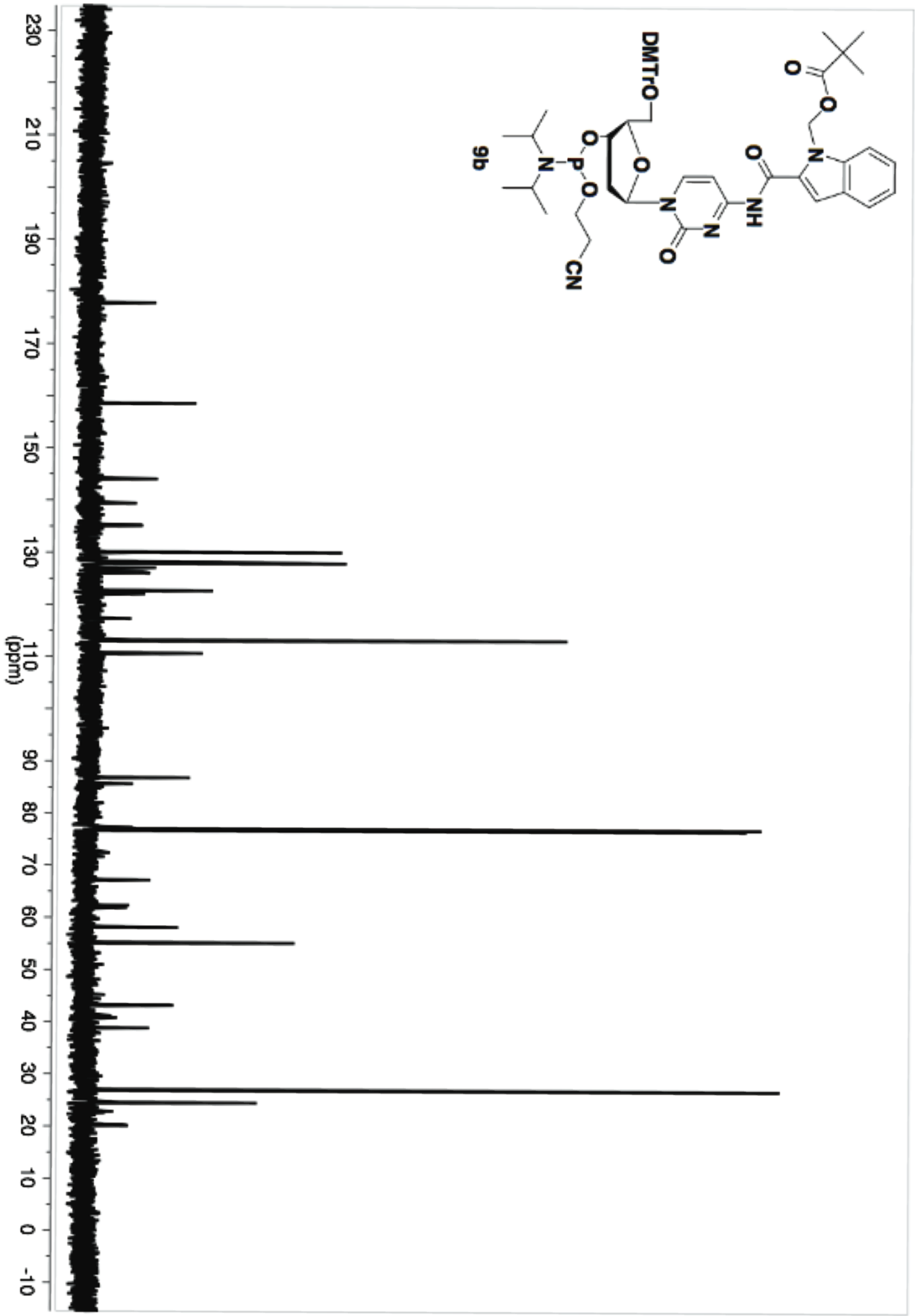

DMTrO
OH
8b
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10

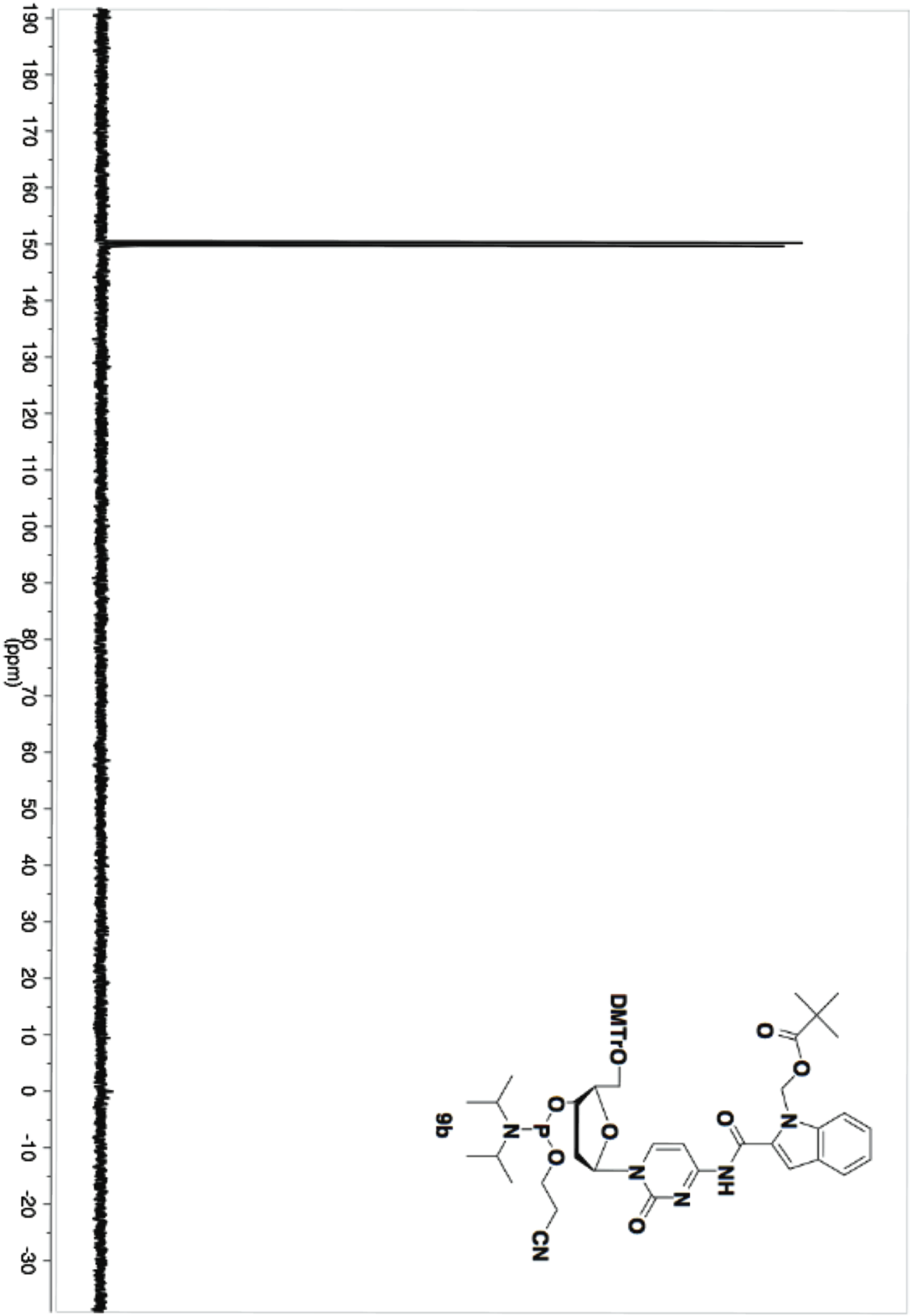

DMTrO
O
N
NH
O
N
O
O
P
O
N
CN
O
O
N
9a
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2


DMTrO
O
NH
O
N
N
O
O
O
P
O
N
CN
9a
O
O
N
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10


DMTrO
O
O
N
O
NH
O
N
N
O
O
P
O
N
CN
O
O
N
9a
180
170
160
150
140
130
120
110
100
90
80
70
(ppm)
60
50
40
30
20
10
0
-10
-20
-30
-40


9b
DMTrO
O
NH
N
CN
13
12
11
10
9
8
7
6
(ppm)
5
4
3
2
1
0
-1
-2


DMTrO
O
P
O
N
CN
O
N
O
NH
O
N
O
O
9b
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0
-10


DMTrO
O
O
N
O
NH
O
N
N
O
O
P
O
N
CN
9b
O
O
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
(ppm)

Computational methods. The simulations were carried out using the AMBER 12.0 program package.[1] The initial structures were built with the canonical B-form parameters frame with the NAB module. The non-canonical modified residues were parameterized by the RESP/6-31G(d) charges fitting.[2] Parm99 AMBER force fields[3] with the latest force field improvements for nucleic acids, parmbsc0[4], were used in the MD simulations. The additional force field parameters were taken from GAFF force field and fitted by using paramfit module embedded in AmberTools 13.[5] The RNA duplexes were solvated in a periodic box with a 10 Å buffer of water molecules, explicitly described by the TIP3P water model, and neutralized, resulting in a concentration of added NaCl of approximately 0.1 M by using ions08.lib.[6] The minimization and equilibration processes were run according to the protocol successfully used in previous works.[7, 8] The unrestrained extended simulations were performed for 10 ns with the Berendsen algorithm to maintain the temperature.[9] During the MD calculation, hydrogen vibrations were removed using SHAKE bond constraints, allowing a longer time step of 2 fs.[10] Long range electrostatic interactions were treated using the Particle Mesh Ewald approach[11] and a 10 Å cutoff. Postprocessing analysis of the trajectories was carried out by using cpptraj module.

1. D.A. Case TAD, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012) AMBER. University of California, San Francisco,
2. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry 97(40):10269-10280
3. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc 117(19):5179-5197
4. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, 3rd, Laughton CA, Orozco M (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys J 92(11):3817-3829
5. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157-1174
6. Joung IS, Cheatham TE, 3rd (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B 112(30):9020-9041
7. Masaki Y, Miyasaka R, Ohkubo A, Seio K, Sekine M (2010) Linear relationship between deformability and thermal stability of 2'-O-modified RNA hetero duplexes. J Phys Chem B 114(7):2517-2524
8. Masaki Y, Miyasaka R, Hirai K, Tsunoda H, Ohkubo A, Seio K, Sekine M (2012) Prediction of the stability of modified RNA duplexes based on deformability analysis: oligoribonucleotide derivatives modified with 2'-O-cyanoethyl-5-propynyl-2-thiouridine as a promising component. Chemical Communications (Cambridge) 48(58):7313-7315
9. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 81(8):3684-3690
10. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 23(3):327-341
11. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98(12):10089

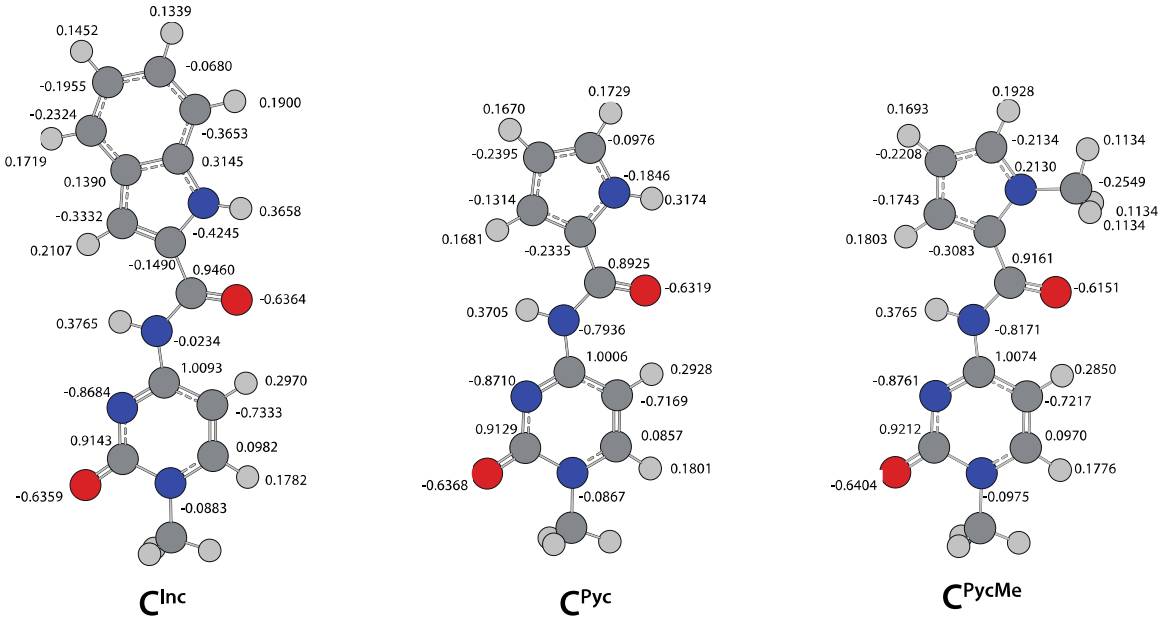


Figure S6 RESP fitted charges

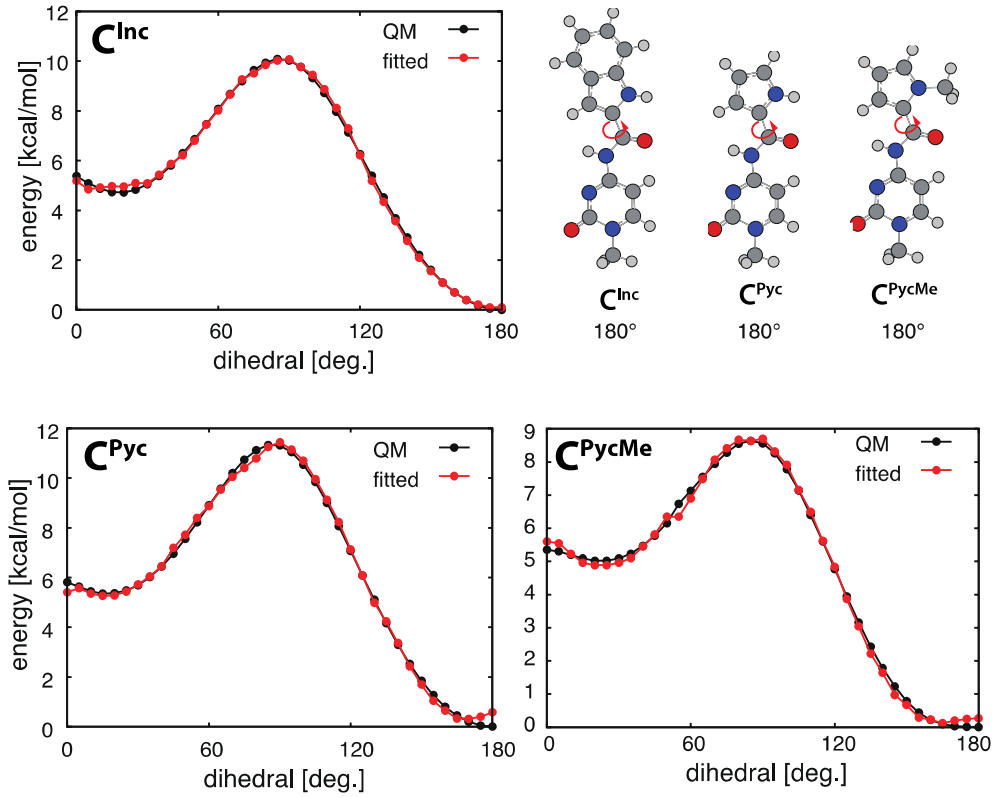


Figure S7 N-C-C-N Torsional angle parameterization using paramfit: red and black curves are rotational profiles produced by fitted parameters and B3LYP/6-31+G(d), respectively.

C^{Pyc}

```
BOND
N2-C   390.50   1.407       same as c2-n
C -CC  449.90   1.406       same as c -c2
CC-C*  418.30   1.429       same as cc-cc
C*-C*  418.30   1.429       same as cc-cc
C*-HA  344.30   1.087       same as c2-hc

ANGLE
CA-N2-C    64.290     123.710   same as c -n -ca
N2-C -O    75.830     122.030   same as n -c - o
N2-C -CC   68.800     123.450   same as c2-c2-n
H -N2-C    47.400     117.860   same as c2-n -hn
C -CC-NA   69.987     114.105   Calculated with empirical approach
C -CC-C*   67.930     120.700   same as c -c2-c2
O -C -CC   72.770     119.120   same as c2-c -o
CC-NA-CW   67.800     110.370   same as c2-na-c2
CC-C*-C*   67.880     110.700   same as cc-cc-cc
CC-C*-HA   50.300     119.700   same as c2-c2-hc
NA-CC-C*   69.830     121.380   same as c2-c2-na
CW-C*-C*   67.880     110.700   same as cc-cc-cc
CW-C*-HA   50.300     119.700   same as c2-c2-hc
C*-C*-HA   50.300     119.700   same as c2-c2-hc

DIHE
CA-N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
CA-N2-C -CC   1    0.650       180.000         -2.000      same as c2-c2-n -c
CA-N2-C -CC   1    1.200       180.000          1.000      same as c2-c2-n -c
O -C -CC-NA   1   -1.7075       95.814           2
O -C -CC-C*   1    1.3991       59.119           3
N2-C -CC-NA   1    4.9000      167.433           2
N2-C -CC-C*   1    1.9080        0.000           2
H -N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
H -N2-C -CC   2    8.300       180.000          2.000      same as X -c -n2-X
C -CC-C*-C*   1    6.650       180.000          2.000      same as X -c2-c2-X
C -CC-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
CC-C*-C*-CW   1    4.000       180.000          2.000      same as X -cc-cc-X
CC-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
NA-CC-C*-C*   1    6.650       180.000          2.000      same as X -c2-c2-X
NA-CC-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
CW-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
HA-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
```

C^{Inc}

```
BOND
N2-C   390.50   1.407       same as c2-n
C -CC  449.90   1.406       same as c -c2
CC-C*  418.30   1.429       same as cc-cc
C*-HA  344.30   1.087       same as c2-hc

ANGLE
CA-N2-C    64.290     123.710   same as c -n -ca
N2-C -O    75.830     122.030   same as n -c - o
N2-C -CC   68.800     123.450   same as c2-c2-n
H -N2-C    47.400     117.860   same as c2-n -hn
C -CC-NA   69.987     114.105   Calculated with empirical approach
C -CC-C*   67.930     120.700   same as c -c2-c2
O -C -CC   72.770     119.120   same as c2-c -o
CC-NA-CN   67.800     110.370   same as c2-na-c2
CC-C*-CB   67.880     110.700   same as cc-cc-cc
CC-C*-HA   50.300     119.700   same as c2-c2-hc
NA-CC-C*   69.830     121.380   same as c2-c2-na
CB-C*-HA   50.300     119.700   same as c2-c2-hc

DIHE
CA-N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
CA-N2-C -CC   1    0.650       180.000         -2.000      same as c2-c2-n -c
CA-N2-C -CC   1    1.200       180.000          1.000      same as c2-c2-n -c
N2-C -CC-C*   1   -1.1798       92.484          2.000
N2-C -CC-NA   1    2.2361       35.998          2.000
H -N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
H -N2-C -CC   2    8.300       180.000          2.000      same as X -c -n2-X
C -CC-C*-CB   1    6.650       180.000          2.000      same as X -c2-c2-X
C -CC-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
O -C -CC-C*   1    0.2494      125.321          3.000
O -C -CC-NA   1    5.3876      179.701          2.000
NA-CC-C*-CB   1    6.650       180.000          2.000      same as X -c2-c2-X
NA-CC-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
```

C^{PycMe}

```
BOND
N2-C   390.50   1.407       same as c2-n
C -C*  449.90   1.406       same as c -c2
C*-N*  411.10   1.391       same as c2-na
C*-C*  418.30   1.429       same as cc-cc
N*-CA  411.10   1.391       same as c2-na
CA-C*  411.70   1.434       same as ca-cc
C*-HA  344.30   1.087       same as c2-hc

ANGLE
CA-N2-C    64.290     123.710   same as c -n -ca
N2-C -O    75.830     122.030   same as n -c - o
N2-C -C*   68.800     123.450   same as c2-c2-n
H -N2-C    47.400     117.860   same as c2-n -hn
C -C*-N*   69.987     114.105   Calculated with empirical approach
C -C*-C*   67.930     120.700   same as c -c2-c2
O -C -C*   72.770     119.120   same as c2-c -o
C*-N*-CA   67.800     110.370   same as c2-na-c2
C*-N*-CT   64.230     117.200   same as c2-na-c3
C*-C*-C*   67.880     110.700   same as cc-cc-cc
C*-C*-HA   50.300     119.700   same as c2-c2-hc
N*-C*-C*   69.830     121.380   same as c2-c2-na
N*-CA-C*   69.830     121.380   same as c2-c2-na
N*-CA-H4   51.180     112.420   same as ha-c2-na
CA-N*-CT   64.230     117.200   same as c2-na-c3
CA-C*-C*   67.660     111.040   same as ca-cc-cc
CA-C*-HA   50.300     119.700   same as c2-c2-hc
C*-CA-H4   50.040     120.940   same as c2-c2-ha

DIHE
CA-N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
CA-N2-C -C*   1    0.650       180.000         -2.000      same as c2-c2-n -c
CA-N2-C -C*   1    1.200       180.000          1.000      same as c2-c2-n -c
N2-C -C*-N*   1   -2.0650       84.962          2.000
N2-C -C*-C*   1    2.7780       31.445          2.000
H -N2-C -O    2    8.300       180.000          2.000      same as X -c -n2-X
H -N2-C -C*   2    8.300       180.000          2.000      same as X -c -n2-X
C -C*-N*-CA   1    0.625       180.000          2.000      same as X -c2-na-X
C -C*-N*-CT   1    0.625       180.000          2.000      same as X -c2-na-X
C -C*-C*-C*   1    6.650       180.000          2.000      same as X -c2-c2-X
C -C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
O -C -C*-N*   1    5.7622      175.670          2.000
O -C -C*-C*   1    0.1089      179.999          3.000
C*-N*-CA-C*   1    0.625       180.000          2.000      same as X -c2-na-X
C*-N*-CA-H4   1    0.625       180.000          2.000      same as X -c2-na-X
C*-C*-C*-CA   1    4.000       180.000          2.000      same as X -cc-cc-X
C*-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
N*-C*-C*-C*   1    6.650       180.000          2.000      same as X -c2-c2-X
N*-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
N*-CA-C*-C*   1    6.650       180.000          2.000      same as X -c2-c2-X
N*-CA-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
CA-N*-C*-C*   1    0.625       180.000          2.000      same as X -c2-na-X
CA-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
C*-CA-N*-CT   1    0.625       180.000          2.000      same as X -c2-na-X
C*-C*-N*-CT   1    0.625       180.000          2.000      same as X -c2-na-X
C*-C*-CA-H4   1    4.000       180.000          2.000      same as X -cc-cc-X
HA-C*-C*-HA   1    6.650       180.000          2.000      same as X -c2-c2-X
HA-C*-CA-H4   1    6.650       180.000          2.000      same as X -c2-c2-X
H4-CA-N*-CT   1    0.625       180.000          2.000      same as X -c2-na-X
```

Figure S8 Additional and fitted force field.

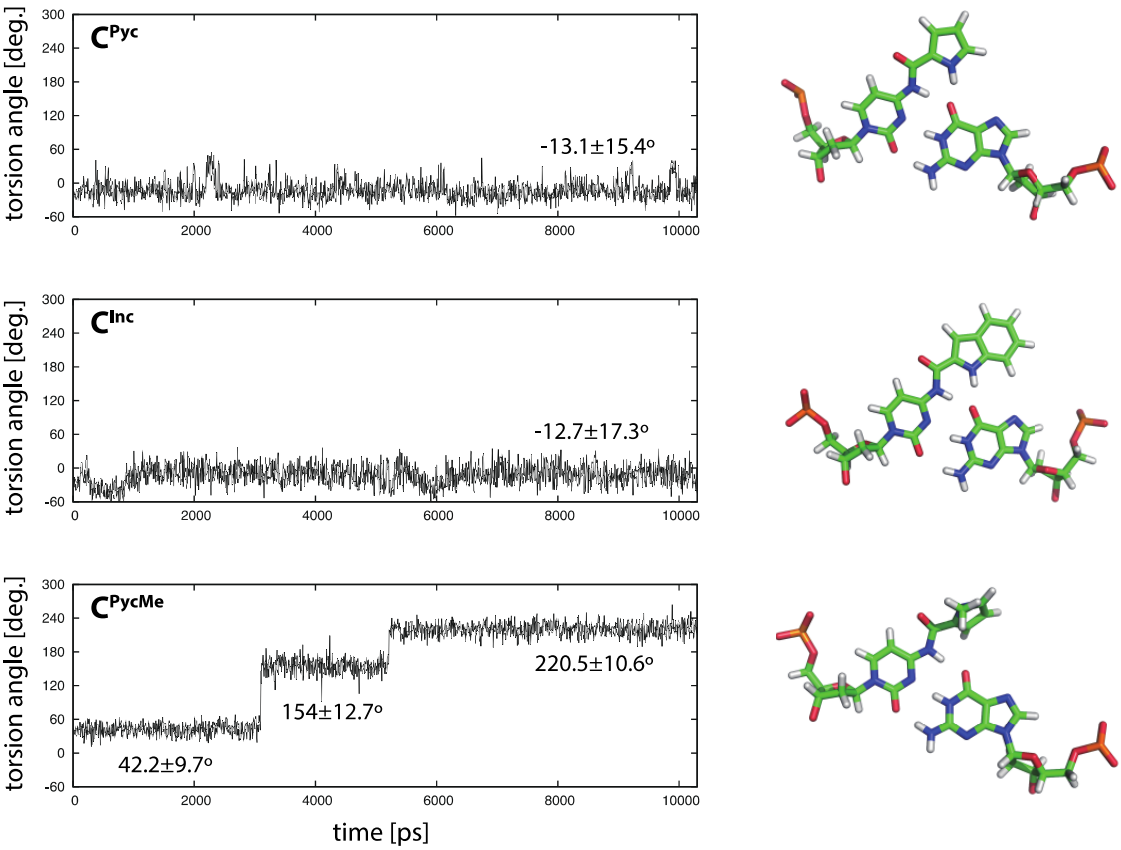


Figure S9 N-C-C-N torsion angle during MD simulation. The mean angles and standard deviations were denoted.

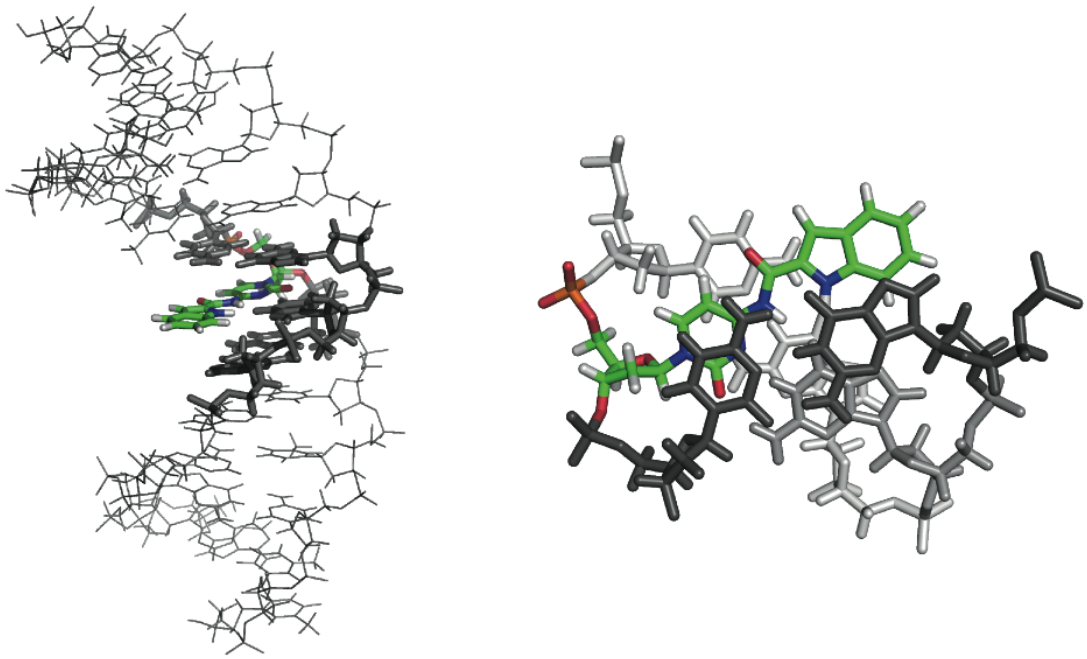


Figure S10 MD simulated structure of C^{Inc}