SUPPORTING INFORMATION

Computational and experimental investigation of mono-septanoside binding by Concanavalin A: Correlation of ligand stereochemistry to enthalpies of binding

Michael Duff, W. Shawn Fyvie, Shankar D. Markad, Alexandra Frankel, Challa V. Kumar, José A. Gascón, & Mark W. Peczuh

Computational Approach

Modeling of all protein-ligand complexes was based on the X-ray structure of Concanavalin A (ConA) with **9** (PDB access code: 1GIC, chain A).[1] Since the co-crystal contains one of the ligand studied, we constructed all other six-member ring ligands by *in situ* modification of **9**. For the septanosides we constructed the ligands in the gas phase and aligned common stereocenters to those of **9** inside the protein. After *in situ* modification with the program Maestro[2] and removal of all crystallographic waters except Wat335, each ligand (including **9**) was re-docked with the program Glide,[3] always maintaining the structure of the protein frozen. Wat335 was kept in the system following the studied by Kadirvelraj *et al*[4], in which they show that Wat335 is a highly conserved molecule. Hydrogen atoms were added to this molecule and their orientation optimized to form hydrogen bonds to Asn14, Arg228, and Asp16.

Docking computations were followed by QM/MM energy minimizations using Qsite,[5] where the ligand and Asp208 were treated at the quantum mechanical (QM) level and the rest of the protein at the MM level. In one set of calculations, the MM region was kept frozen in all QM/MM calculations, while in another set the relevant residues Asn14, Leu99, Tyr100, Asp208, Arg228, and ligand were relaxed during the QM/MM minimization. Since the Docking procedure generates a manifold of poses, we considered no just the one with highest score, but others that presented qualitative differences in the conformation of the ligand. The final energies reported in Table 2 are those corresponding to the lowest QM/MM energy found. For the QM/MM calculations, cuts between the QM and MM region were treated with the frozen-orbital method as implemented in Qsite. The MM region was treated with the OPLSAA force field,[6] and the QM region with Density Functional Theory (DFT). The functional B3LYP and basis set 6-31g* were used in all QM calculations. Gas phase calculations were carried out with Jaguar at

the same level of theory specified above.[7] As explained in the main manuscript, the starting geometry for the ligand in the gas phase was that obtained by the highest scored conformation using Glide. However, we considered other conformations with different OH rotamers and ring conformations. Selection of these rotamers was guided by a previous study in which a rigorous conformational analysis of septanosides **5** and **6** was done by Monte Carlo (MC) simulations.[8] In fact, our minimum energy structures of **5** and **6** were the same as those obtained in these previous MC simulations. From the study of DeMatteo *et al*[8] it was concluded that the septanosides studied were rigid enough to prefer one conformation at room temperature. Thus, the sole use of the lowest energy configurations, found in both the Glide/QM/MM and gas phase QM calculations, to correlate binding energies with experimental enthalpies appears to be a reasonable assumption.

Convergence of results with respect to larger QM regions

As explained in the computational methods section, we performed all QM/MM calculations with a minimal QM region (i.e. Asp208 and ligand). This selection was based on several factors: 1) A large number of calculation was required to test different conformers for a same ligand. We typically tested about 5 conformers per ligand. Conformers were selected from the manifold of poses generated by Glide and also by the possible conformers a ligand can adopt in the gas phase. 2) We noticed from the preliminary Docking calculations that Asp208 was the only residue that conserved the number of hydrogen bonds with all ligands studied. 3) Of the two charged residues (208 and 228), Asp208 is the only residue that can substantially polarize the electronic structure of the ligand (Arg228's side chain is actually not in close contact with the ligands and it is not involved in hydrogen bonds).

To further test that this minimal QM region was appropriate, we recomputed the binding energy using a larger QM region: Asn14, Leu99, Tyr100, Asp208, and Arg228. We considered some of the ligands studied as a benchmark. As shown in Table S1, the difference in using the proposed minimal QM region and a larger QM region, spanning the relevant residues inside the cavity, will not alter the interpretation proposed in the manuscript regarding the differential binding among the ligands. Both models agree well with experiment.

Table S1. Comparison between the calculated binding energy with the experimental binding enthalpies considering a minimal QM region (Asp208 + ligand) and a larger QM region (Asn14, Leu99, Tyr100, Asp208, Arg228 + ligand).

Ligand	ΔE_b (minimal QM region (kcal/mol)	ΔE_b (minimal QM region (kcal/mol)	$\Delta\Delta H$ (exp.) (kcal/mol)
6	5.2	5.3	4.67/2.8
7	0.0	0.0	0.0
8	16.9	21.2	no binding
9	1.7	1.8	2.0

Analysis of Methyl 2-*O*-methyl β-septanoside (15)

In the main text of the manuscript we gave a rational on the reason of the difference in the enthalpy of binding between **7** and **6** (both septanosides). We showed that the main difference between them comes from the reorientation of the ring hydroxyl groups in going from the free to the bound state (Figure 8 of the main text). One way to describe these changes was to count the number of electrostatic OH···O interactions before and after binding. In methyl "manno" β-septanoside **7**, this number changes from three to two (see arrows in Figure 8), while in methyl

"gluco" β-septanoside **6** it changes from five to three, accounting for most of the difference in binding energy between these two ligands. A similar analysis can be invoked for ligand **15** (Fig. S1), which has the same stereochemistry as **6**, but contains a methyl ether rather than a hydroxyl group at C2. We can also see that for **15** the number of electrostatic interactions changes from four to two (i.e, changes by two as in **6**), as shown in Fig. S2, which accounts for the similar magnitudes in $\Delta\Delta H$ in **15** than in **6**.

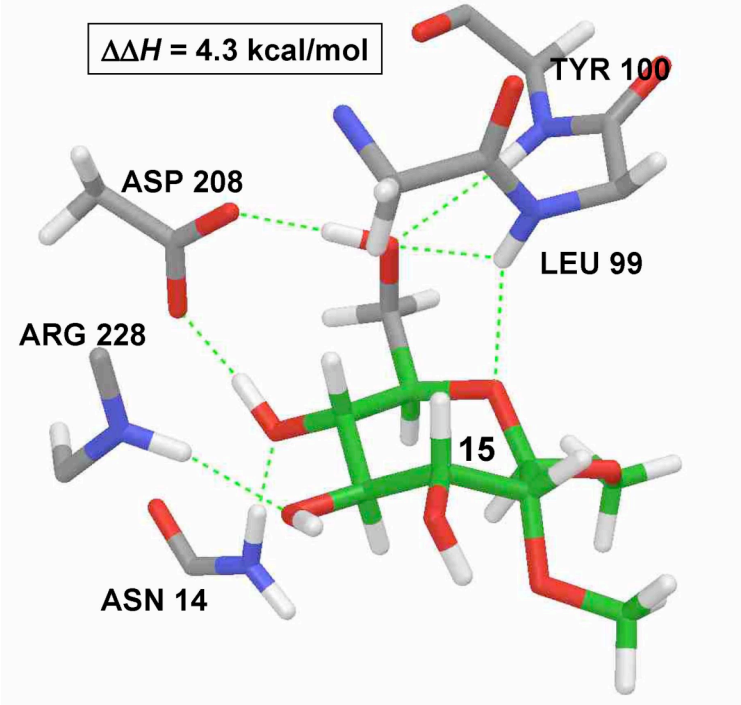


Figure S1

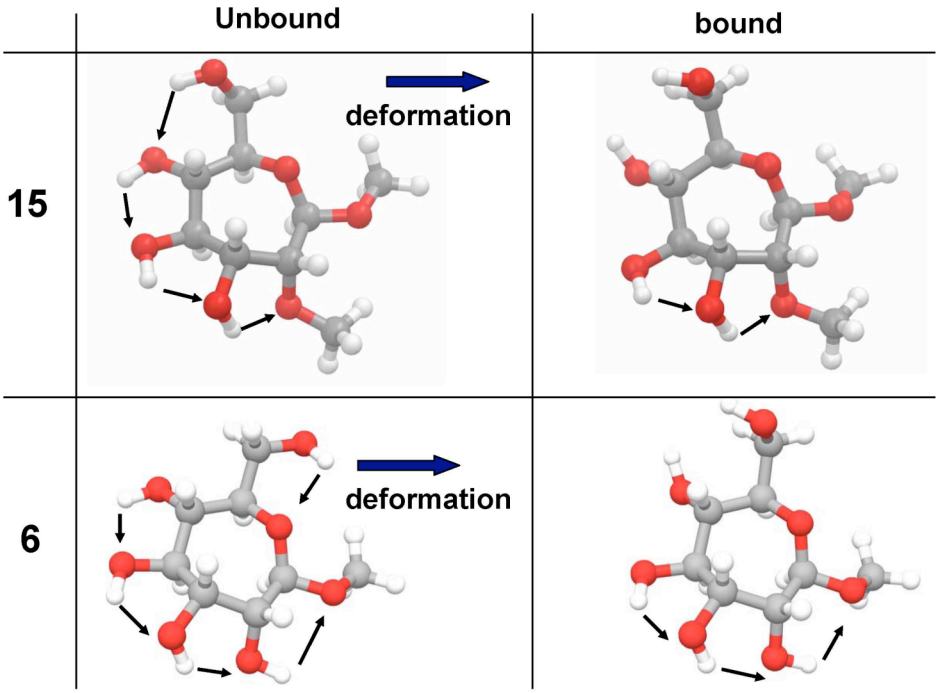


Figure S2

Analysis of Methyl "gluco" α-septanoside (5)

The large enthalpy of binding of the ConA•**5** complex relative to ConA•**7** (8.8 kcal/mol) correlates well with no-binding events seen in ITC. The QM/MM structure of this complex reveals that no hydrogen bond is lost with respect to **7** (Figure S3). Thus all its destabilization with respect to **7** comes from deformation energy upon binding. While in **7** one OH···O contact is lost, in **5** two of these contacts are lost (Figure S4).

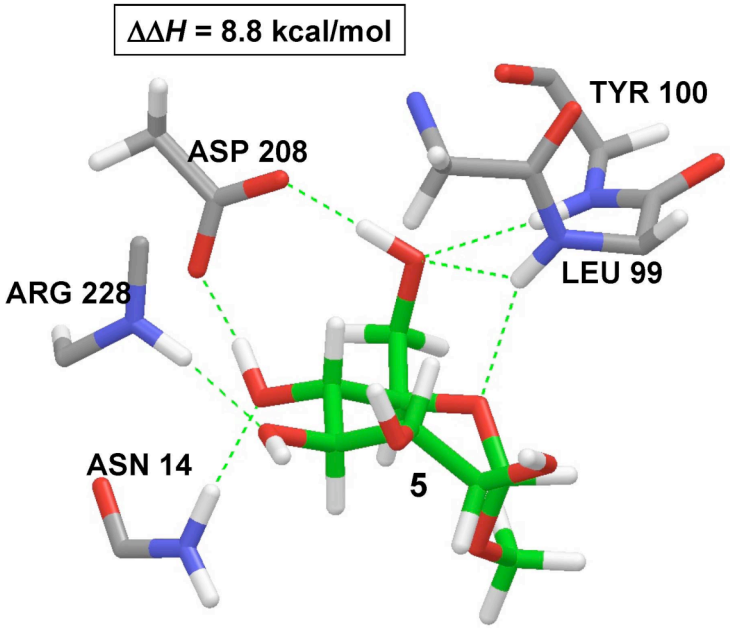


Figure S3

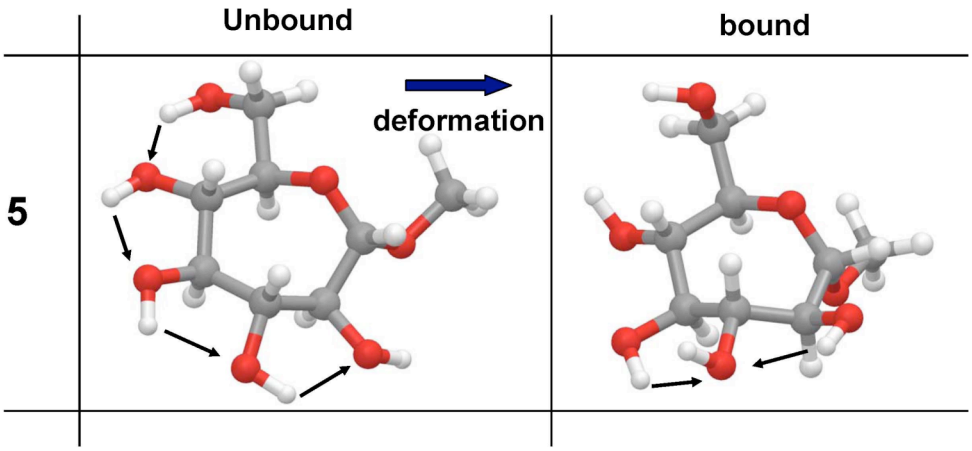


Figure S4

Comparison of Scoring functions with experimental binding affinities

Table S2 presents scoring results of the best pose for all ligands studied. Notice that the ranking of affinities according to the Docking calculations only marginally correlate with the ranking given by the experimental free energy of binding. The reason for this is twofold: 1) Although the scoring function has in it some sort of binding energy incorporated (via Molecular Mechanics terms), the docking algorithm misses the reorientation of the OH bonds that must necessary take

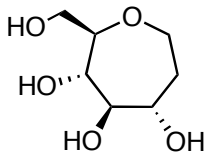
place upon binding. 2) Also, it has been shown before by DeMatteo et al (ref 8) that the relative energies of the various conformers according to a molecular mechanics force field did not coincide with the QM relative energies, not even qualitatively.

Table S2. Comparison between the calculated binding energy with the experimental binding enthalpies. Values in parenthesis correspond to a QM/MM calculation in which residues Asn14, Leu99, Tyr100, Asp208, and Arg228 are relaxed. Also shown are the scoring function according to docking calculations and their comparison with experimental binding affinities. Ranks are presented based on the scoring function and free energy of binding.

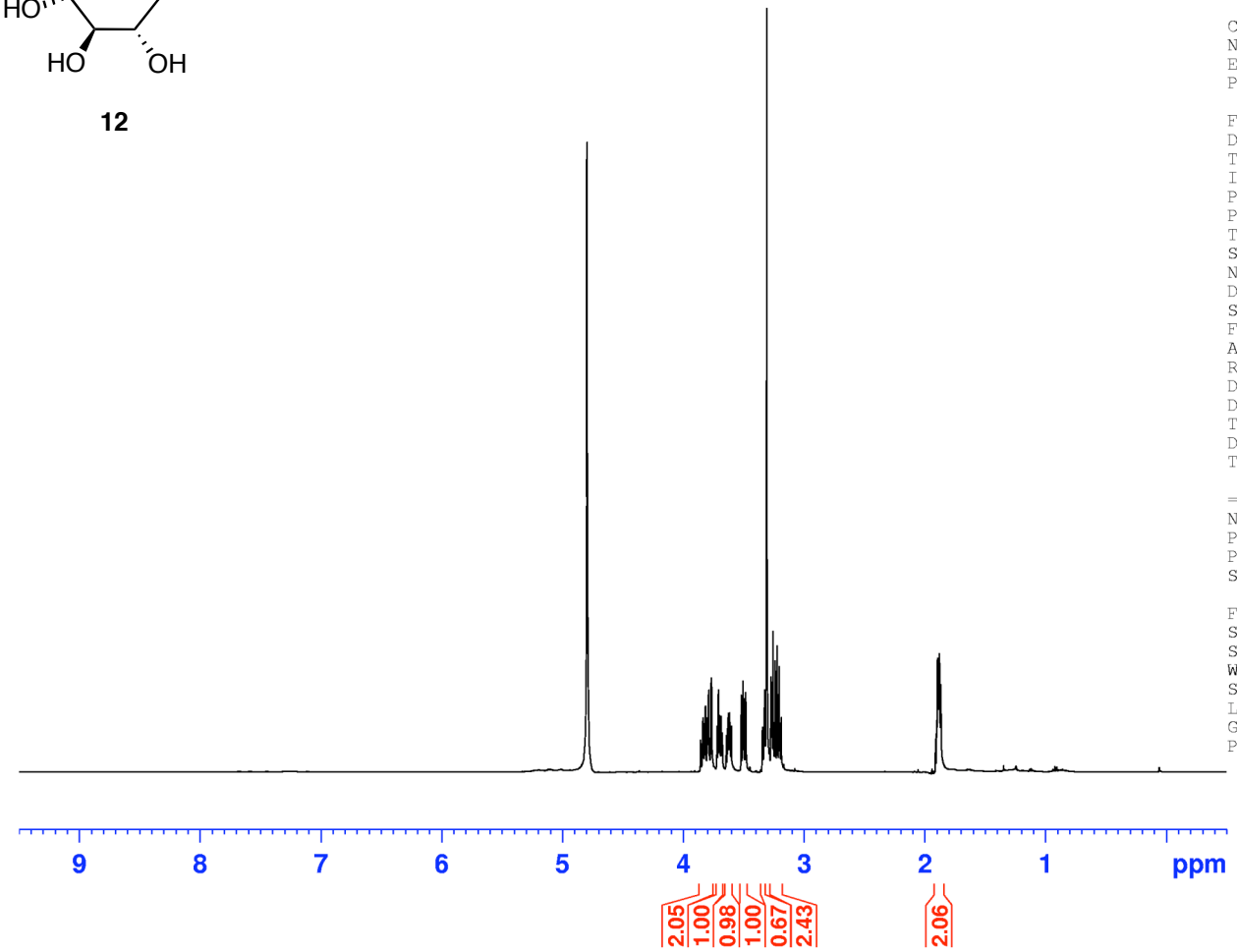
Ligand	ΔE_b (kcal/mol)	$\Delta\Delta H$ (exp.) (kcal/mol)	Scoring Function	Scoring Rank	$\Delta\Delta G$ (exp.) (kcal/mol)	Experimental Rank
5	8.8 (**9.0**)	no binding	-6.58	3	no binding	
6	5.2 (**5.5**)	4.67/2.8	-5.45	7	-3.7/3.6	4
7	0.0 (**0.0**)	0.0	-6.04	4	-4.0	3
8	16.9 (**15.4**)	no binding	-5.84	5	no binding	
9	1.7 (**1.1**)	2.0	-6.72	1	-4.5	2
10	1.0 (**1.0**)	1.0	-6.70	2	-5.5	1
15	4.3 (**4.7**)	5.37/4.1	-5.62	6	-3.6/3.1	5

References

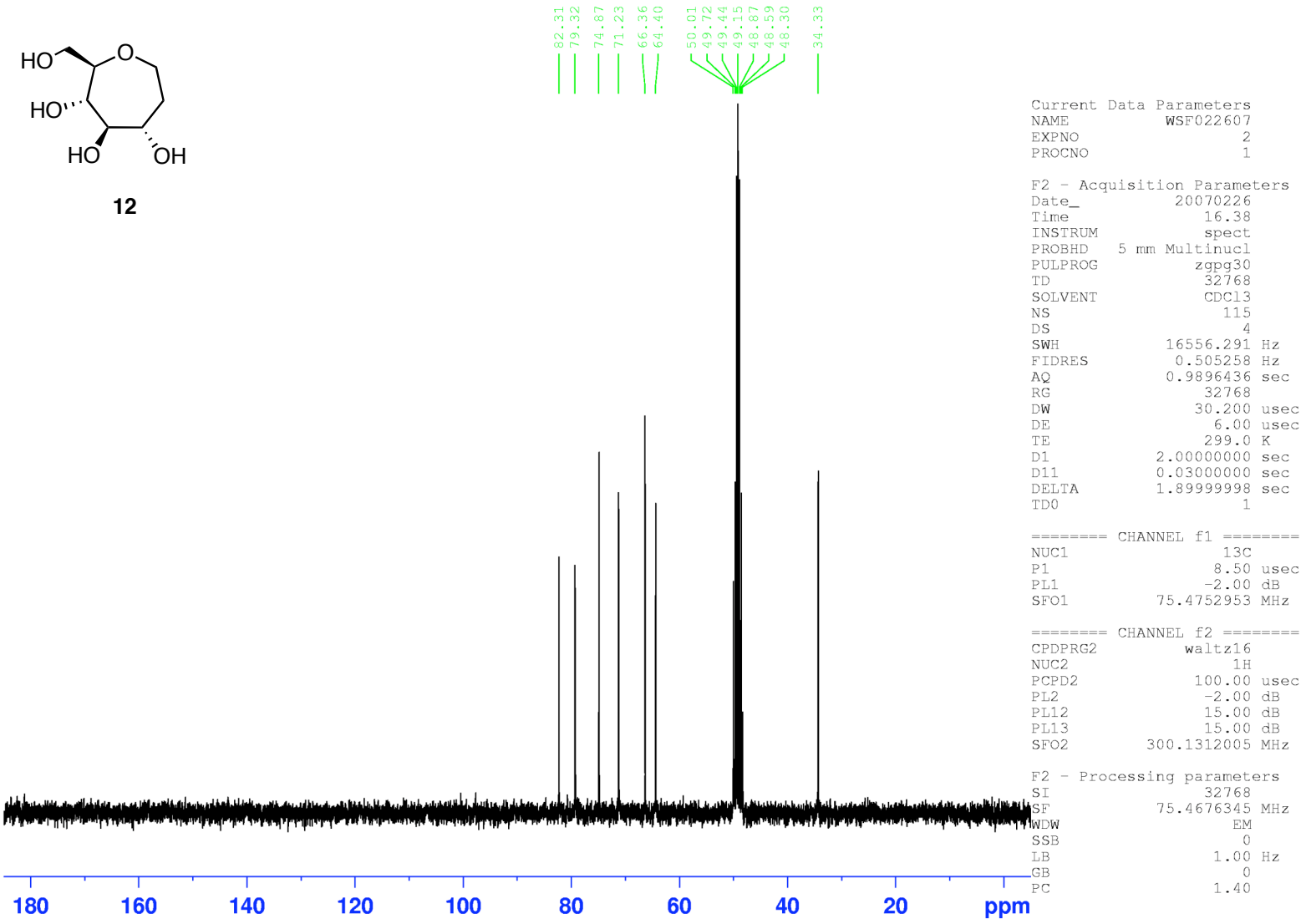
1. S. J. Harrop, J. R. Helliwell, T. C. M. Wan, A. J. Kalb, L. Tong and J. Yariv, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1996, **52**, 143-155.
2. *Maestro, version 9.0*, Schrodinger LLC, New York, NY.
3. *Glide, version 502076*, Schrodinger LLC, New York, NY.
4. R. Kadirvelraj, B. L. Foley, J. D. Dyekjaer and R. J. Woods, *J. Am. Chem. Soc.*, 2008, **130**, 16933-16942.
5. *Qsite, version 7.5*, Schrodinger LLC, New York, NY.
6. W. L. Jorgensen, D. S. Maxwell and J. TiradoRives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236.
7. *Jaguar, version 7.5*, Schrodinger LLC, New York, NY.
8. M. P. DeMatteo, N. L. Snyder, M. Morton, D. M. Baldisseri, C. M. Hadad and M. W. Peczuh, *J. Org. Chem.*, 2005, **70**, 24-38.

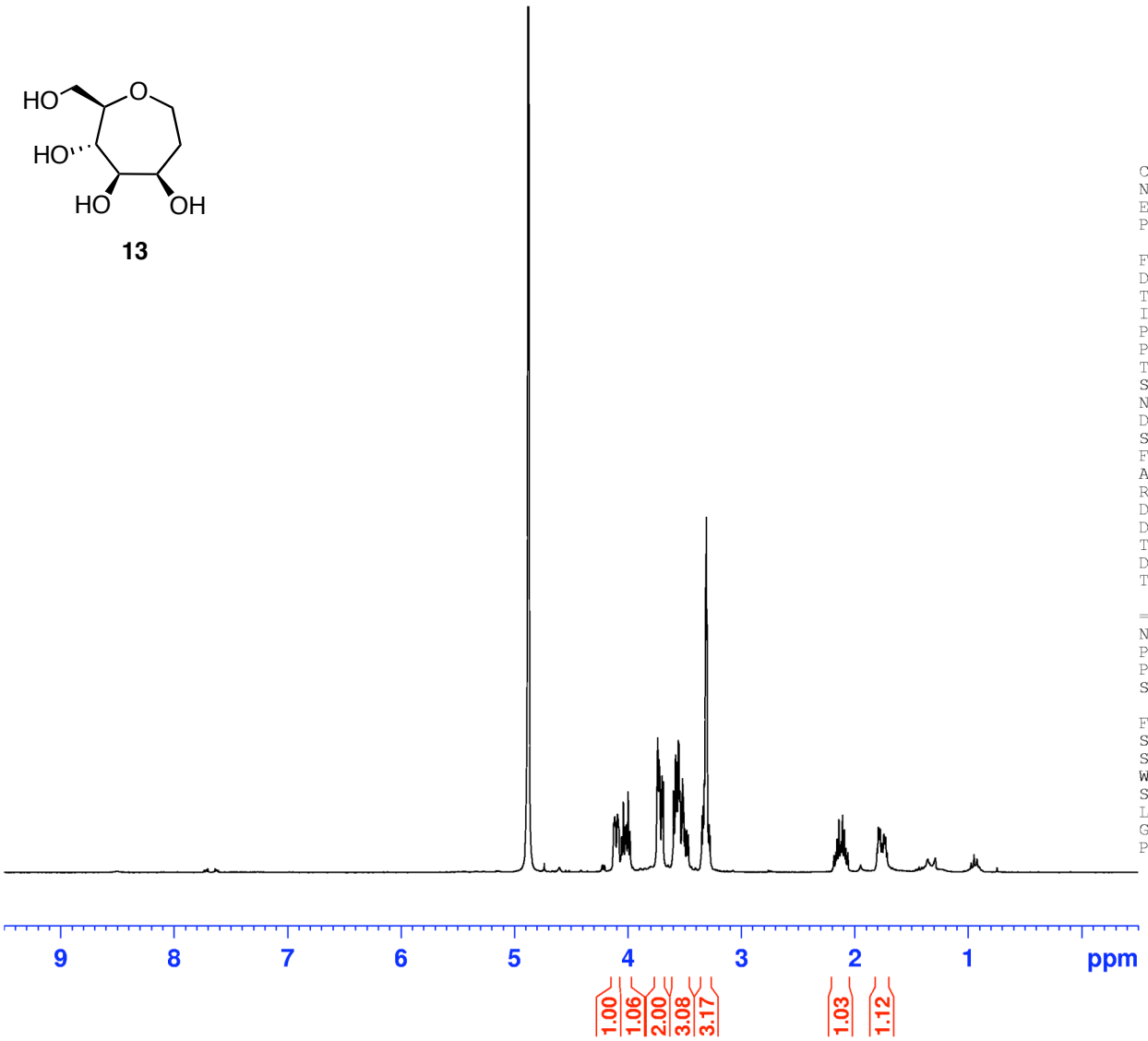


12

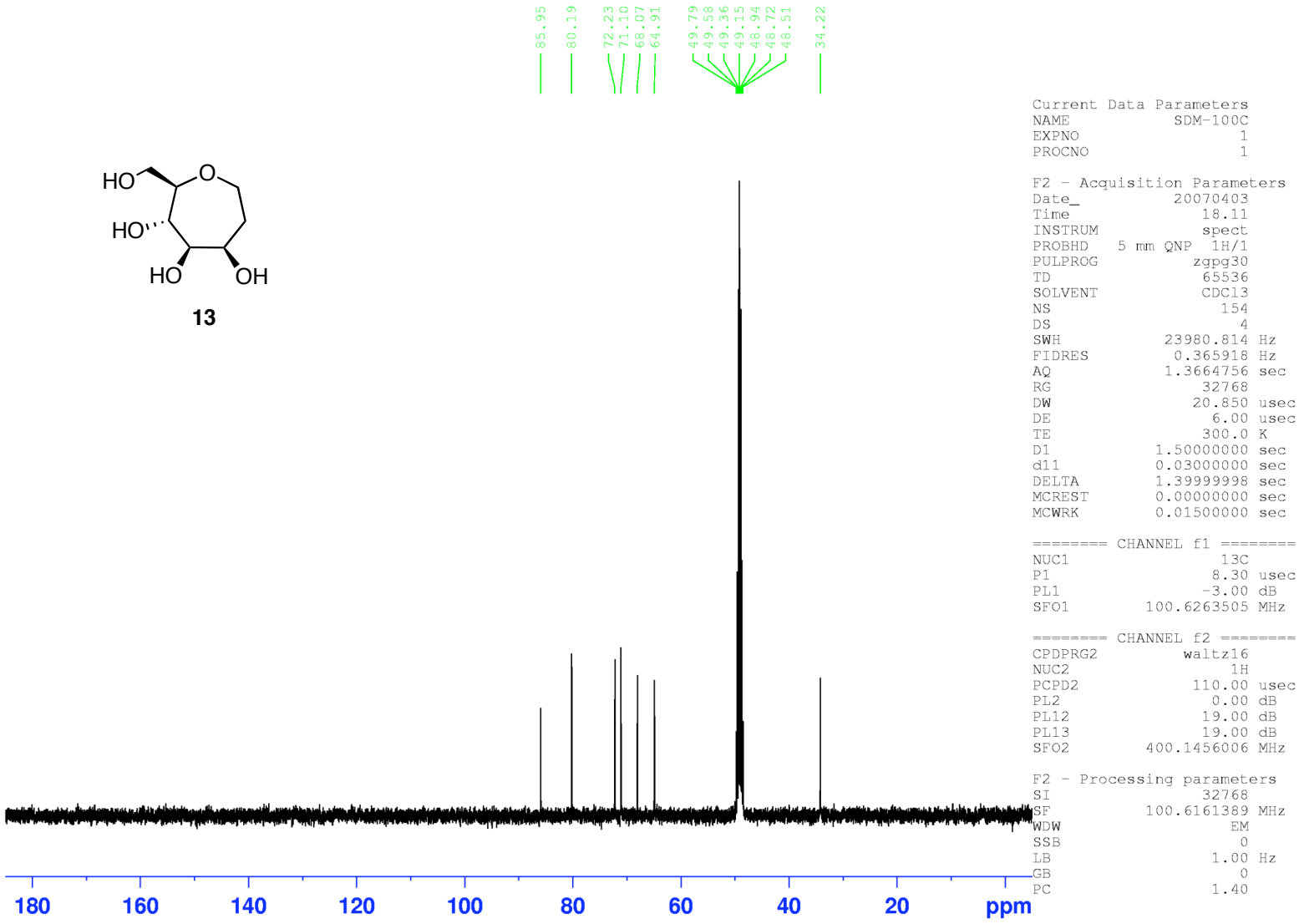


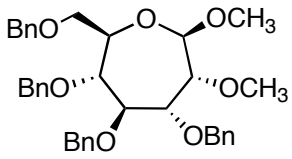

```
Current Data Parameters
NAME          wsf022607b
EXPNO                  1
PROCNO                 1

F2 - Acquisition Parameters
Date_           20070226
Time               11.04
INSTRUM            spect
PROBHD   5 mm TXI 1H-13
PULPROG             zg30
TD                 32768
SOLVENT            CDCl3
NS                    16
DS                     2
SWH           10330.578 Hz
FIDRES         0.315264 Hz
AQ            1.5860696 sec
RG                  25.4
DW                48.400 usec
DE                 6.00 usec
TE                298.2 K
D1           1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
NUC1                  1H
P1                12.00 usec
PL1               -1.00 dB
SFO1        500.1330885 MHz

F2 - Processing parameters
SI                 32768
SF          500.1300300 MHz
WDW                   EM
SSB                    0
LB                0.30 Hz
GB                     0
PC                 1.00
```


HO
O
HO
HO
OH
12
82.31
79.32
74.87
71.23
66.36
64.40
50.01
49.72
49.44
49.15
48.87
48.59
48.30
34.33
180
160
140
120
100
80
60
40
20
ppm
Current Data Parameters
NAME WSF022607
EXPNO 2
PROCNO 1
F2 - Acquisition Parameters
Date_ 20070226
Time 16.38
INSTRUM spect
PROBHD 5 mm Multinucl
PULPROG zgpg30
TD 32768
SOLVENT CDCl3
NS 115
DS 4
SWH 16556.291 Hz
FIDRES 0.505258 Hz
AQ 0.9896436 sec
RG 32768
DW 30.200 usec
DE 6.00 usec
TE 299.0 K
D1 2.00000000 sec
D11 0.03000000 sec
DELTA 1.89999998 sec
TD0 1
======== CHANNEL f1 ========
NUC1 13C
P1 8.50 usec
PL1 -2.00 dB
SFO1 75.4752953 MHz
======== CHANNEL f2 ========
CPDPRG2 waltz16
NUC2 1H
PCPD2 100.00 usec
PL2 -2.00 dB
PL12 15.00 dB
PL13 15.00 dB
SFO2 300.1312005 MHz
F2 - Processing parameters
SI 32768
SF 75.4676345 MHz
WDW EM
SSB 0
LB 1.00 Hz
GB 0
PC 1.40

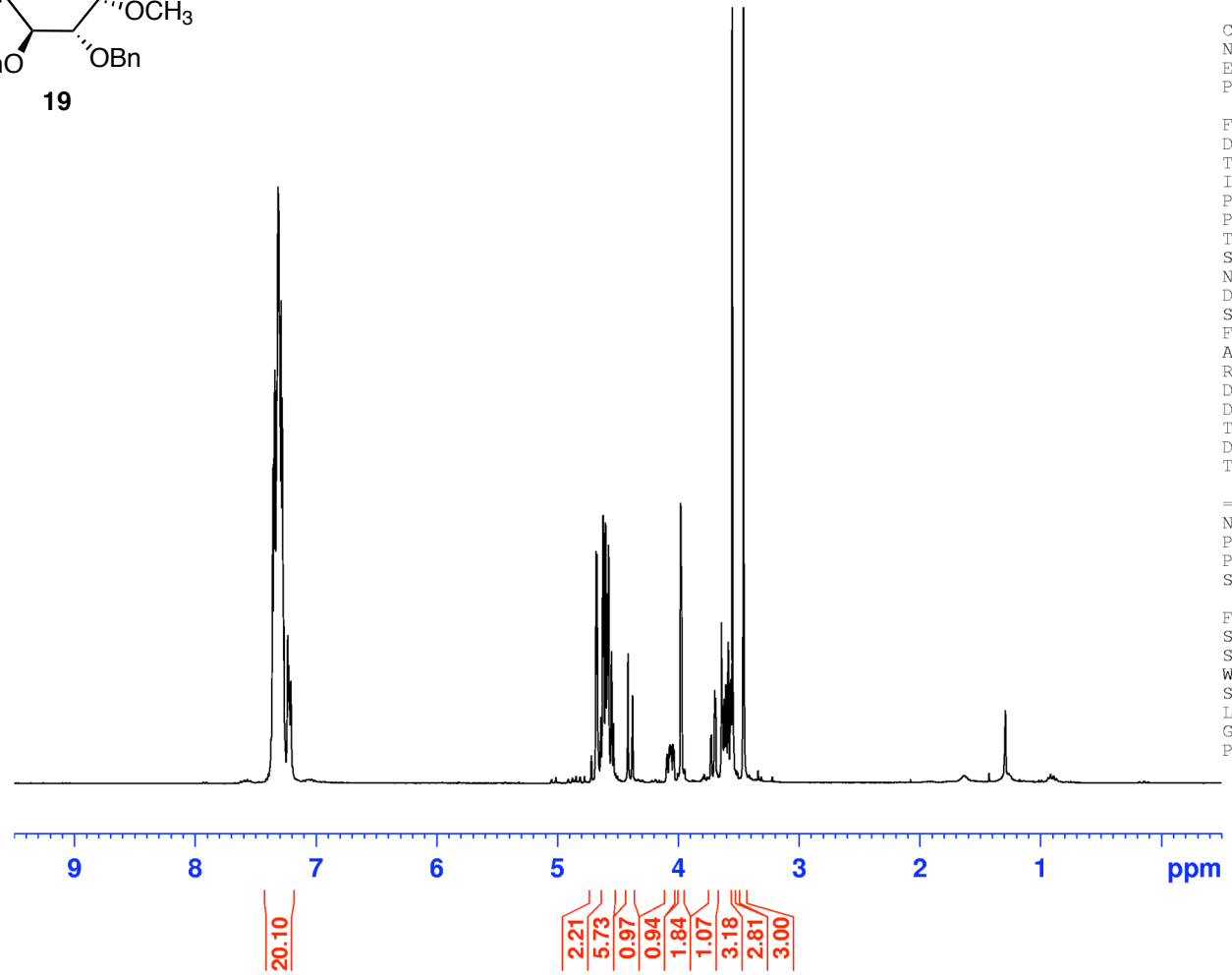



```
Current Data Parameters
NAME                  BS-01-89
EXPNO                        1
PROCNO                       1

F2 - Acquisition Parameters
Date_                 20071025
Time                     18.16
INSTRUM                  spect
PROBHD    5 mm Multinucl
PULPROG                  zg30
TD                      32768
SOLVENT                  CDCl3
NS                          32
DS                           2
SWH                   4194.631 Hz
FIDRES                0.128010 Hz
AQ                    3.9059956 sec
RG                         256
DW                     119.200 usec
DE                       6.00 usec
TE                      296.5 K
D1                 1.00000000 sec
TD0                          1

======== CHANNEL f1 ========
NUC1                        1H
P1                       8.20 usec
PL1                      0.00 dB
SFO1               300.1315007 MHz

F2 - Processing parameters
SI                      32768
SF                 300.1300048 MHz
WDW                        EM
SSB                         0
LB                       0.30 Hz
GB                          0
PC                       1.00
```

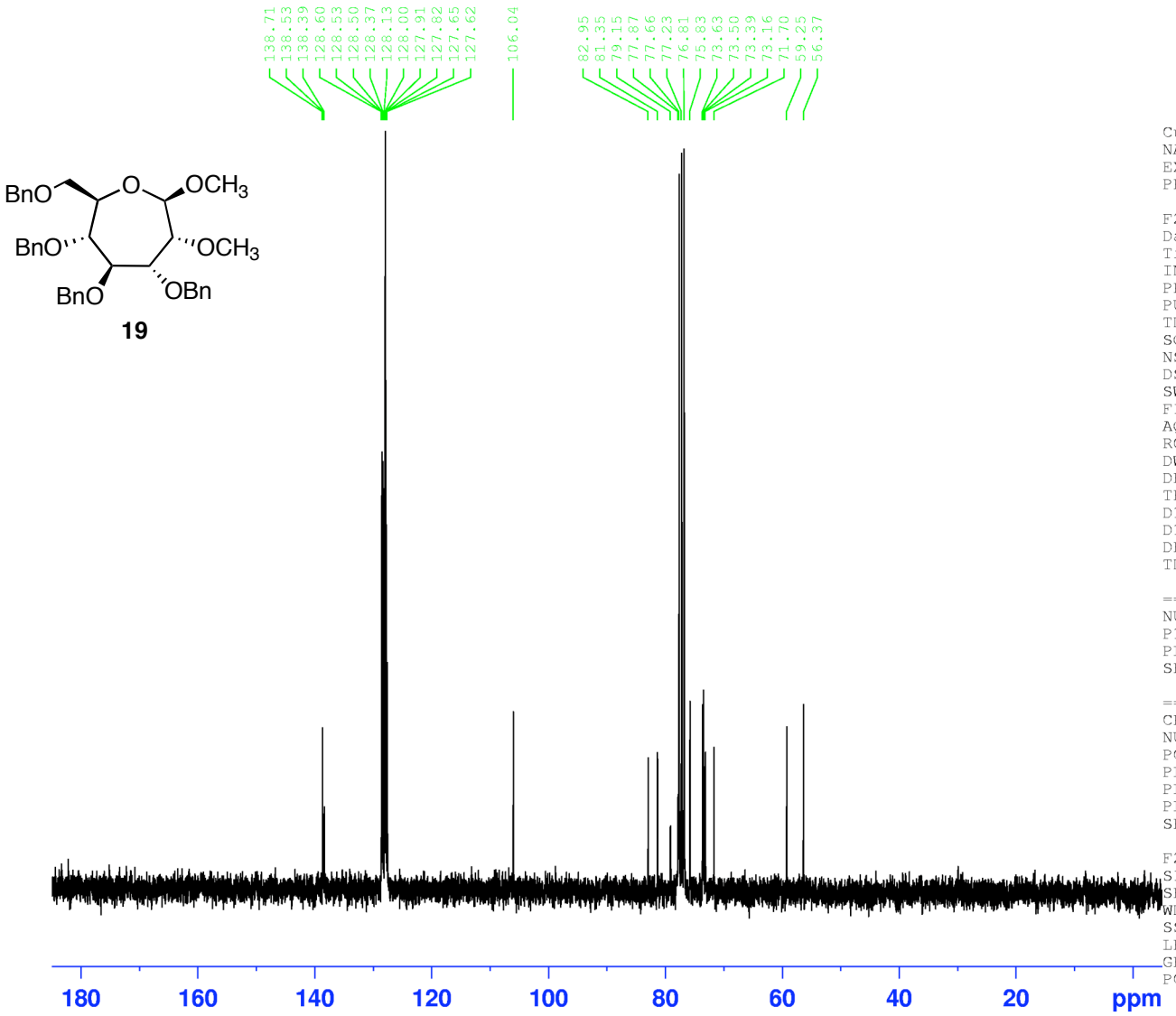

85.95
80.19
72.23
71.10
68.07
64.91
49.79
49.58
49.36
49.15
48.94
48.72
48.51
34.22
HO
HO
HO
O
OH
13
180
160
140
120
100
80
60
40
20
ppm
Current Data Parameters
NAME SDM-100C
EXPNO 1
PROCNO 1
F2 - Acquisition Parameters
Date_ 20070403
Time 18.11
INSTRUM spect
PROBHD 5 mm QNP 1H/1
PULPROG zgpg30
TD 65536
SOLVENT CDCl3
NS 154
DS 4
SWH 23980.814 Hz
FIDRES 0.365918 Hz
AQ 1.3664756 sec
RG 32768
DW 20.850 usec
DE 6.00 usec
TE 300.0 K
D1 1.50000000 sec
d11 0.03000000 sec
DELTA 1.39999998 sec
MCREST 0.00000000 sec
MCWRK 0.01500000 sec
======== CHANNEL f1 ========
NUC1 13C
P1 8.30 usec
PL1 -3.00 dB
SFO1 100.6263505 MHz
======== CHANNEL f2 ========
CPDPRG2 waltz16
NUC2 1H
PCPD2 110.00 usec
PL2 0.00 dB
PL12 19.00 dB
PL13 19.00 dB
SFO2 400.1456006 MHz
F2 - Processing parameters
SI 32768
SF 100.6161389 MHz
WDW EM
SSB 0
LB 1.00 Hz
GB 0
PC 1.40



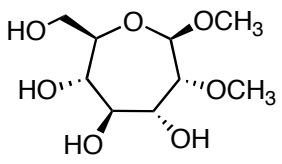
19




```
Current Data Parameters
NAME          WSF120906
EXPNO                 1
PROCNO                1

F2 - Acquisition Parameters
Date_          20061211
Time              14.04
INSTRUM           spect
PROBHD   5 mm Multinucl
PULPROG            zg30
TD                32768
SOLVENT           CDCl3
NS                   16
DS                    2
SWH            4194.631 Hz
FIDRES         0.128010 Hz
AQ             3.9059956 sec
RG                  114
DW              119.200 usec
DE                6.00 usec
TE                297.1 K
D1           1.00000000 sec
TD0                   1

======== CHANNEL f1 ========
NUC1                  1H
P1                12.00 usec
PL1                0.00 dB
SFO1        300.1315007 MHz

F2 - Processing parameters
SI                32768
SF          300.1300000 MHz
WDW                  EM
SSB                   0
LB                 0.30 Hz
GB                    0
PC                 1.00
```

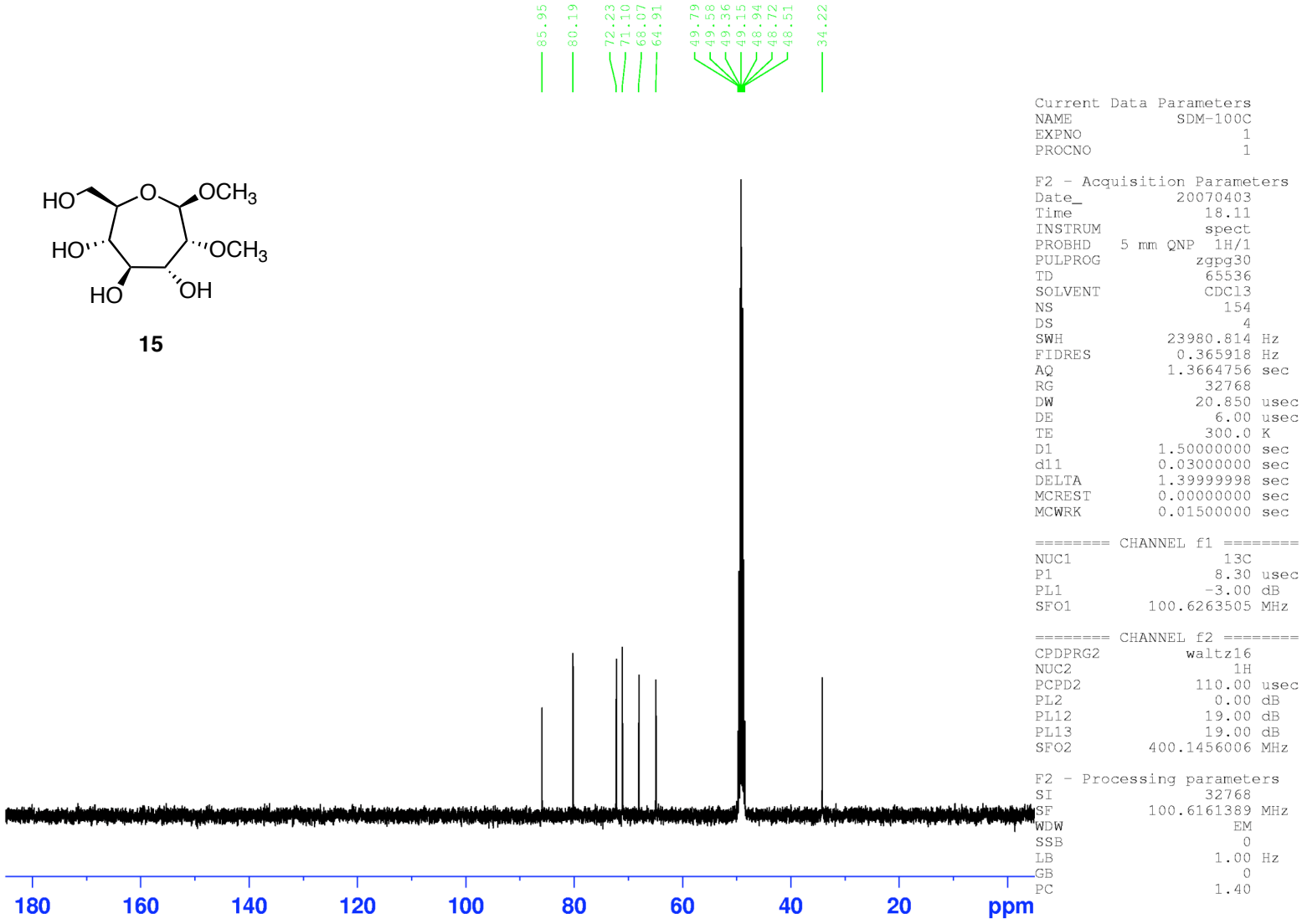



```
Current Data Parameters
NAME                 WSF120906
EXPNO                        2
PROCNO                       1

F2 - Acquisition Parameters
Date_                 20061211
Time                     14.09
INSTRUM                  spect
PROBHD     5 mm Multinucl
PULPROG                 zgpg30
TD                       32768
SOLVENT                  CDCl3
NS                         264
DS                           4
SWH                  16556.291 Hz
FIDRES               0.505258 Hz
AQ                   0.9896436 sec
RG                       32768
DW                      30.200 usec
DE                       6.00 usec
TE                      298.9 K
D1                  2.00000000 sec
D11                 0.03000000 sec
DELTA               1.89999998 sec
TD0                          1

======== CHANNEL f1 ========
NUC1                       13C
P1                       8.50 usec
PL1                     -2.00 dB
SFO1                75.4752953 MHz

======== CHANNEL f2 ========
CPDPRG2                waltz16
NUC2                       1H
PCPD2                  100.00 usec
PL2                     -2.00 dB
PL12                    15.00 dB
PL13                    15.00 dB
SFO2               300.1312005 MHz

F2 - Processing parameters
SI                       32768
SF                  75.4677339 MHz
WDW                         EM
SSB                          0
LB                        1.00 Hz
GB                           0
PC                        1.40
```



15


```
Current Data Parameters
NAME          WSF121206
EXPNO                 1
PROCNO                1

F2 - Acquisition Parameters
Date_          20061212
Time              17.09
INSTRUM           spect
PROBHD   5 mm Multinucl
PULPROG            zg30
TD                32768
SOLVENT           CDCl3
NS                   16
DS                    2
SWH            4194.631 Hz
FIDRES         0.128010 Hz
AQ             3.9059956 sec
RG                71.8
DW              119.200 usec
DE                6.00 usec
TE               297.2 K
D1           1.00000000 sec
TD0                   1

======== CHANNEL f1 ========
NUC1                  1H
P1                12.00 usec
PL1                0.00 dB
SFO1        300.1315007 MHz

F2 - Processing parameters
SI                32768
SF          300.1300047 MHz
WDW                  EM
SSB                   0
LB                 0.30 Hz
GB                    0
PC                 1.00
```


85.95
80.19
72.23
71.10
68.07
64.91
49.79
49.58
49.36
49.15
48.94
48.72
48.51
34.22
HO
O
OCH3
HO
OCH3
HO
OH
15
180
160
140
120
100
80
60
40
20
ppm
Current Data Parameters
NAME SDM-100C
EXPNO 1
PROCNO 1
F2 - Acquisition Parameters
Date_ 20070403
Time 18.11
INSTRUM spect
PROBHD 5 mm QNP 1H/1
PULPROG zgpg30
TD 65536
SOLVENT CDCl3
NS 154
DS 4
SWH 23980.814 Hz
FIDRES 0.365918 Hz
AQ 1.3664756 sec
RG 32768
DW 20.850 usec
DE 6.00 usec
TE 300.0 K
D1 1.50000000 sec
d11 0.03000000 sec
DELTA 1.39999998 sec
MCREST 0.00000000 sec
MCWRK 0.01500000 sec
======== CHANNEL f1 ========
NUC1 13C
P1 8.30 usec
PL1 -3.00 dB
SFO1 100.6263505 MHz
======== CHANNEL f2 ========
CPDPRG2 waltz16
NUC2 1H
PCPD2 110.00 usec
PL2 0.00 dB
PL12 19.00 dB
PL13 19.00 dB
SFO2 400.1456006 MHz
F2 - Processing parameters
SI 32768
SF 100.6161389 MHz
WDW EM
SSB 0
LB 1.00 Hz
GB 0
PC 1.40