Supporting Information: Site-selective glycosylation of hemoglobin on Cys β93

Yalong Zhang[a], Veer S. Bhatt [b], Guoyong Sun[c], Peng G. Wang[a, b] and Andre F. Palmer[c*]

[a]*Department of Chemistry*

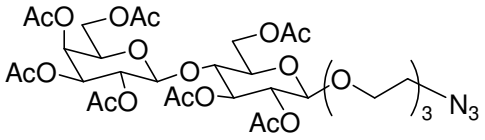
[b]*Biophysics Program and Department of Biochemistry*

[c]*Department of Chemical and Biomolecular Engineering*

The Ohio State University

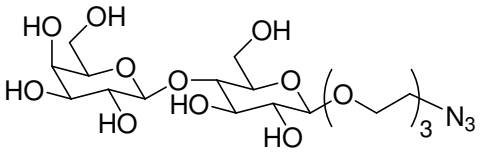
Columbus, OH, 43210

U.S.A.

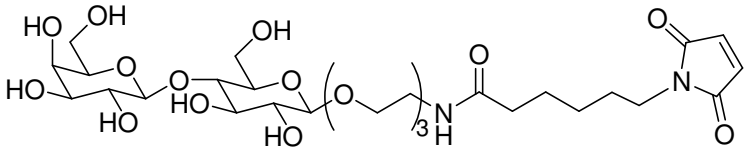


1-azido-3,6-dioxa-9-octyl-*O*-(2,3,6-tri-*O*-acetyl-β-D-glucopyranolsyl)-(1→4)-2,3,4,6-tetra-*O*-acetyl -β-D-galactoside (2). To a solution of peracetyl lactoside **1** (10.0 g, 14.7 mmol) in acetonitrile (150 mL) was added 2-[2-(2-chloroethoxy)ethoxy]ethanol (4.6 mL, 44.2 mmol) and 4 Å MS (Molecular Sieve) (2 g). The mixture was stirred at r.t. for 0.5 h and then cooled to 0 ºC. After that, boron trifluoride ether solution (3.9 mL, 29.4 mmol) was injected into the solution. The solution was slowly warmed to room temperature and stirred overnight. The solution was neutralized with saturated $NaHCO_3$ aqueous solution and extracted with ethyl acetate. The organic phase was dried with Na_2SO_4, filtered and concentrated. The crude product was purified by flash column chromatography with Hexanes/ EtOAc (1:1, v/v). The product (5.0 g, 6.35 mmol) in DMF (50 mL) was reacted with sodium azide (1.24 g, 19.1 mmol) and stirred at 80 ^oC overnight. The DMF was removed under reduced pressure and the crude product was purified by flash column chromatography with Hexanes/ EtOAc (1:1, v/v) to yield **2** (3.2 g) in an overall yield of 30% (two steps) as a colorless oil. ^1H NMR (500 MHz, $CDCl_3$) δ 1.97 (s, 3H), 2.05-2.06 (m, 9H), 2.07 (s, 3H), 2.13 (s, 3H), 2.16 (s, 3H), 3.41 (dd, *J* = 9.5, 4.3 Hz, 2H), 3.63-3.71 (m, 9H), 3.71-3.75 (m, 1H), 3.80 (t, *J* = 9.6 Hz, 1H), 3.86-3.93 (m, 2H), 4.07-4.16 (m, 3H), 4.48-4.51 (m, 2H), 4.58 (d, *J* = 8.0 Hz, 1H), 4.90 (dd, *J* = 9.5, 8.0 Hz, 1H), 4.96 (dd, *J* = 10.4, 3.5 Hz, 1H), 5.11 (dd, *J* = 10.5, 8.0 Hz, 1H), 5.20 (t, *J* = 9.4 Hz, 1H), 5.35-5.36

(m, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 20.49, 20.61, 20.69, 20.80, 20.85, 50.69, 60.81, 62.02, 66.63, 69.05, 69.14, 70.05, 70.40, 71.00, 71.69, 72.65, 72.84, 76.29, 100.64, 101.08, 169.07, 169.65, 169.76, 170.05, 170.14, 170.34, 170.36.



1-azido-3,6-dioxa-9-octyl-*O*-(β-D-glucopyranolsyl)-(1→4)-β-D-galactoside (3). To a solution of compound **2** (3.3 g, 4.16 mmol) in MeOH (20 ml) was added sodium methoxide (22.5 mg, 0.416 mmol). The mixture was stirred at r.t. for 5 h, and neutralized with Dowex 50WX2-100 (H^+) resin. The resultant solution was then concentrated under reduced pressure to yield **3** (1.7 g, 82%) as a white solid. ^{1}H NMR (500 MHz, D_2O) δ 3.23-3.26 (m, 1H), 3.41 (t, *J* = 5.2 Hz, 2H), 3.44-3.46 (m, 1H), 3.48-3.51 (m, 1H), 3.55-3.57 (m, 3H), 3.61-3.63 (m, 7H), 3.65-3.67 (m, 4H), 3.69-3.77 (m, 2H), 3.82 (d, *J* = 3.3 Hz, 1H), 3.87-3.89 (m, 1H), 3.94-3.99 (m, 1H), 4.34 (d, *J* = 7.9 Hz, 1H), 4.42 (d, *J* = 8.0 Hz, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 50.09, 60.03, 60.98, 68.71, 69.17, 69.45, 69.55, 69.65, 70.91, 72.48, 72.78, 74.26, 74.73, 75.31, 78.31, 102.04, 102.90.



***N*-3',6'-dioxa-9'-octyl-*O*-(β-D-glucopyranolsyl)-(1→4)-β-D-galactoside-6-maleimido-hexanamide (maleimide-lactoside) (4).** To a solution of compound **3** (0.20 g, 0.40 mmol) in MeOH (10 mL) was added a catalytic amount of Pd/C. The mixture was stirred under hydrogen atmosphere at r.t. for 24 h. Consequently, the Pd/C was removed by filtration and the solution was concentrated under reduced pressure to yield the amine compound. The amine compound (50 mg, 0.106 mmol) in DMF (5.0 mL) was reacted with 6-maleimidohexanoic acid *N*-hydroxysuccinimide ester (34.0 mg, 0.111 mmol) and Et_3N (21 μL, 0.148 mmol). The solution was stirred at 40 ^{o}C overnight. The solvent was removed under reduced pressure and the residue was purified by flash chromatography with EtOAc:IPA:H_2O (6:3:1) to yield the

product as a colorless oil (26.1 mg) with an overall total yield of 37% (two steps). ^{1}H NMR (500 MHz, MeOD) δ 1.30-1.34 (m, 2H), 1.59-1.66 (m, 4H), 2.21 (t, *J* = 7.4 Hz, 2H), 3.29 (t, *J* = 8.3 Hz, 1H), 3.32-3.34 (m, 4H), 3.36-3.38 (m, 2H), 3.43-3.46 (m, 1H), 3.50-3.53 (m, 3H), 3.54-3.57 (m, 4H), 3.59-3.64 (m, 4H), 3.68-3.74 (m, 5H), 3.75-3.82 (m, 2H), 3.84-3.87 (m, 2H), 3.92 (dd, *J* = 12.2, 2.6 Hz, 1H), 4.01-4.05 (m, 1H), 4.38 (d, *J* = 7.6, 5.6 Hz, 2H), 6.83 (s, 2H). ^{13}C NMR (125 MHz, MeOD) δ 25.05, 25.92, 27.88, 35.41, 37.04, 38.92, 60.58, 61.09, 68.30, 68.90, 69.24, 69.80, 70.08, 70.14, 71.16, 73.34, 73.44, 74.90, 75.12, 75.69, 79.31, 102.90, 103.73, 133.98, 171.23, 174.73. HRMS: Calculated for $C_{28}H_{46}N_2O_{16}Na$ (M+Na^+) 689.2745; found 689.2744.

The proteolytic MS-MS data of the βChain of the glycosylated bovine hemoglobin

Trypsin

Measured m/z	Calculated M+H	Theoretical M+H	Protein Sequence	MASCOT Score	# Being detected
821.36^{1+}/411.24^{2+}	821.48	821.4073	^{1}MLTAEEK^{7}	490	19
837.53^{1+}/419.31^{2+}	837.62	837.4022	^{1}M$_{(Ox)}$LTAEEK^{7}	82	3
950.38^{1+}/475.80^{2+}	950.60	950.5094	^{8}AAVTAFWGK^{16}	1415	30
584.98^{3+}	1752.94	1752.8989	^{1}MLTAEEKAAVTAFWGK^{16}	481	12
1328.82^{1+}/664.87^{2+}	1328.74	1328.7169	^{17}VKVDEVGGEALGR^{29}	2542	47
551.30^{2+}	1101.60	1101.5535	^{19}VDEVGGEALGR^{29}	229	6
425.56^{3+}/637.83^{2+}	1274.66	1274.7256	^{30}LLVVYPWTQR^{39}	955	24
697.19^{3+}	2089.57	2089.9535	^{40}FFESFGDLSTADAVMNNPK^{58}	3403	45
702.71^{3+}/1053.70^{2+}	2106.13	2104.9484	^{40}FFESFGDLSTADAVM$_{(Ox)}$NNPK^{58}	881	14
613.32^{2+}	1225.64	1225.6245	^{65}KVLDSFSNGMK^{75}	272	9
549.17^{2+}	1097.34	1097.5296	^{66}VLDSFSNGMK^{75}	602	18
557.34^{2+}	1113.68	1113.5245	^{66}VLDSFSNGM$_{(Ox)}$K^{75}	35	1
724.14^{3+}/1085.65^{2+}	2170.30	2170.0597	^{76}HLDDLKGTFAALSELHC$_{(Cam)}$DK^{94}	287	4
843.66^{3+}	2528.98	2528.2238	^{82}GTFAALSELHC$_{(Cam)}$DKLHVDPENFK^{103}	75	2
782.76^{3+}	2346.28	2345.3704	^{95}LHVDPENFKLLGNVLVVVLAR^{115}	89	2
422.55^{3+}/633.70^{2+}	1265.40	1265.8304	^{104}LLGNVLVVVLAR^{115}	1039	22
935.10^{3+}/623.63^{3+}	1868.89	1868.9541	^{116}NFGKEFTPVLQADFQK^{131}	1416	36
474.80^{3+}/711.88^{2+}	1422.76	1422.7264	^{120}EFTPVLQADFQK^{131}	1029	21
589.36^{2+}	1177.72	1177.6800	^{132}VVAGVANALAHR^{143}	1396	18

Chymotrypsin

Measured m/z	Calculated M+H	Theoretical M+H	Protein Sequence	MASCOT Score	# Being detected
691.42^{2+}	1381.39	1381.7032	^{1}MLTAEEKAAVTAF^{13}	143	4
784.42^{2+}	1567.84	1567.7825	^{1}MLTAEEKAAVTAFW^{14}	110	3
570.02^{2+}	1139.08	1137.5786	^{3}TAEEKAAVTAF^{13}	27	1
663.13^{2+}	1325.26	1323.6579	^{3}TAEEKAAVTAFW^{14}	21	1
642.66^{3+}	1925.98	1926.0807	^{14}WGKVKVDEVGGEALGRLL^{31}	113	2
651.11^{2+}	1301.22	1300.7101	^{15}GKVKVDEVGGEAL^{27}	73	2
580.73^{3+}/870.77^{2+}	1740.54	1740.0014	^{15}GKVKVDEVGGEALGRLL^{31}	344	6
663.37^{1+}	663.37	663.3501	^{32}VVYPW^{36}	195	8
531.29^{2+}	1061.58	1061.5051	^{37}TQRFFESF^{44}	22	1
674.00^{2+}	1347.00	1346.6375	^{37}TQRFFESFGDL^{47}	40	1
782.32^{3+}/521.48^{3+}	1562.44	1562.7632	^{71}SNGMKHLDDLKGTF^{84}	243	9
527.05^{3+}	1579.15	1578.7581	^{71}SNGM$_{(ox)}$KHLDDLKGTF^{84}	25	1
603.28^{1+}	603.28	603.3348	^{85}AALSEL^{90}	42	2
418.98^{3+}/628.81^{2+}	1256.62	1256.6303	^{85}AALSELHC$_{(Cam)}$DKL^{95}	64	2
699.52^{3+}/1048.20^{2+}	2095.40	2094.9913	^{85}AALSELHC$_{(Cam)}$DKLHVDPENF^{102}	82	2
756.600^{2+}	1512.20	1510.6743	^{91}HC$_{(Cam)}$DKLHVDPENF^{102}	48	1
614.30^{3+}	1840.90	1839.9388	^{114}ARNFGKEFTPVLQADF^{129}	25	1
676.36^{2+}	1351.72	1351.6892	^{118}GKEFTPVLQADF^{129}	376	9
801.50^{3+}/1202.18^{2+}	2402.50	2402.3078	^{118}GKEFTPVLQADFQKVVAGVANAL^{140}	87	2
971.04^{2+}	1941.08	1941.0804	^{122}TPVLQADFQKVVAGVANAL^{140}	180	4
766.06^{2+}	1531.12	1530.8275	^{126}QADFQKVVAGVANAL^{140}	50	1
535.40^{2+}	1069.80	1069.6364	^{130}QKVVAGVANAL^{140}	95	3

Table 1. Data-collection statistics.

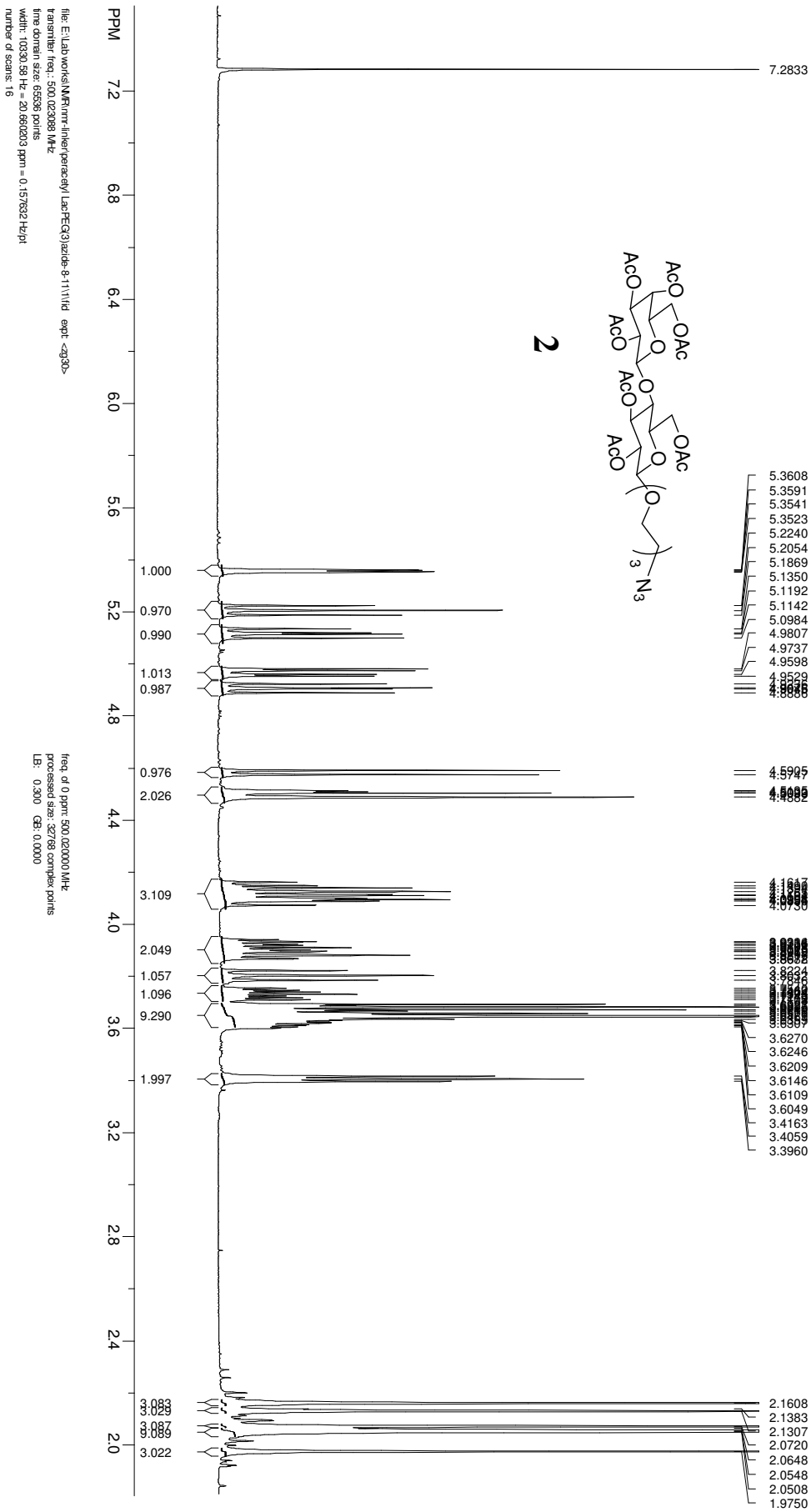
Values in parentheses are for the last shell.

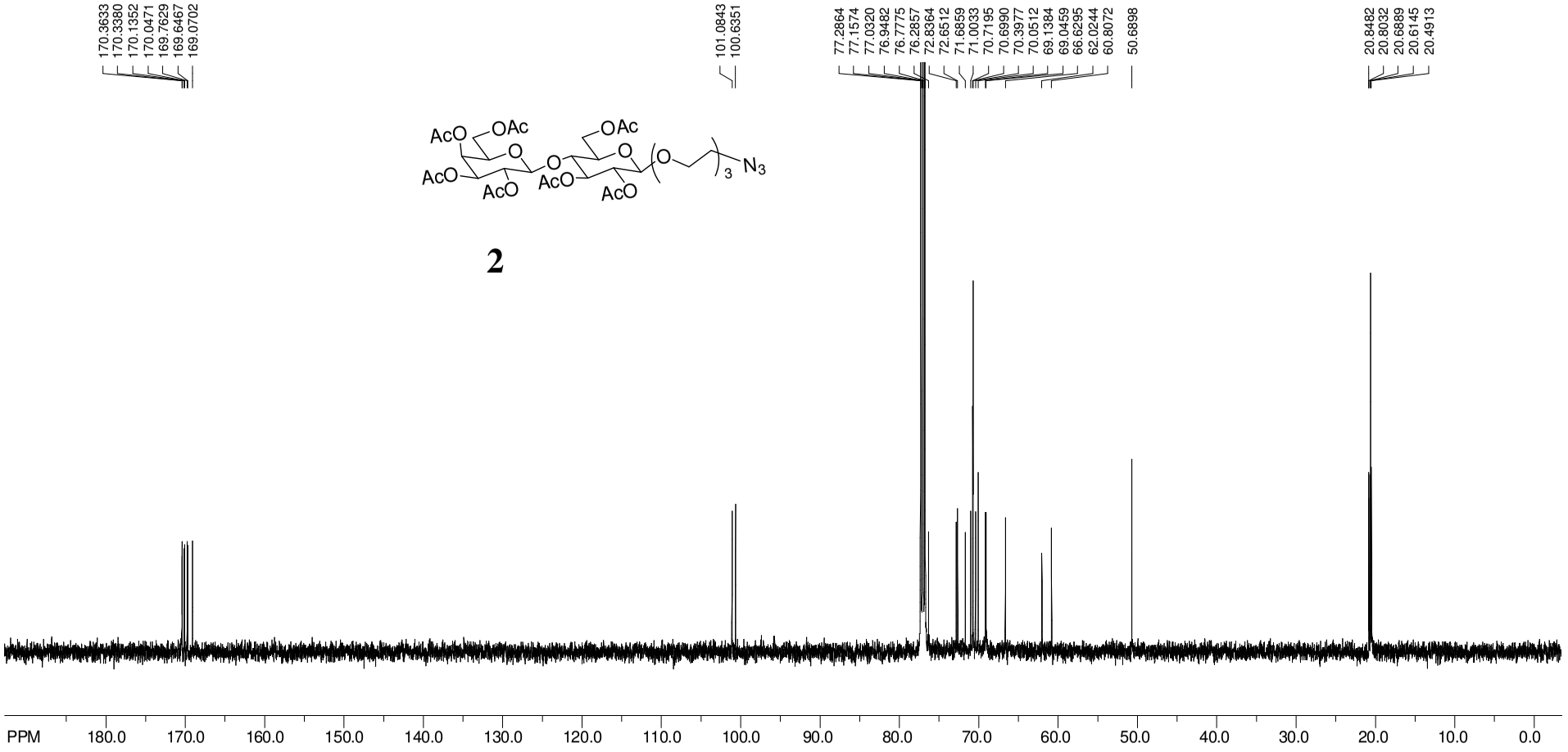
Wavelength (Å)	1.54
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 62.6, b = 74.0, c = 129.8
V_M (Å^3 Da^{-1})	2.35
Solvent content (%)	47.7
Resolution range (Å)	24.5 - 3.5 (3.6 - 3.5)
No. of observed reflections	23286
No. of unique reflections	7999
Completeness (%)	96.4 (92.5)
R_{sym} (%)	13.1 (33.2)
I/I_σ	5.3 (2.4)

Table 2. Summary of crystallographic refinement.

PDB code	
Resolution limits (Å)	24.5 - 3.5
R_{cryst} (%)	26.8
R_{free} (%)	33.2
Total number of non-hydrogen atoms	4582
Total number of atoms in succinimide linked lactose	1(44)
Water molecules	8
R.m.s. deviations†(Å)	
Bond lengths	0.006
Bond angles	0.98
Dihedral angles	0.05
Overall G factor§	0.26
Average B factors $(Å)^2$	91.2
Residues in most favored regions (%)	91.1
Residues in additionally allowed regions (%)	8.9

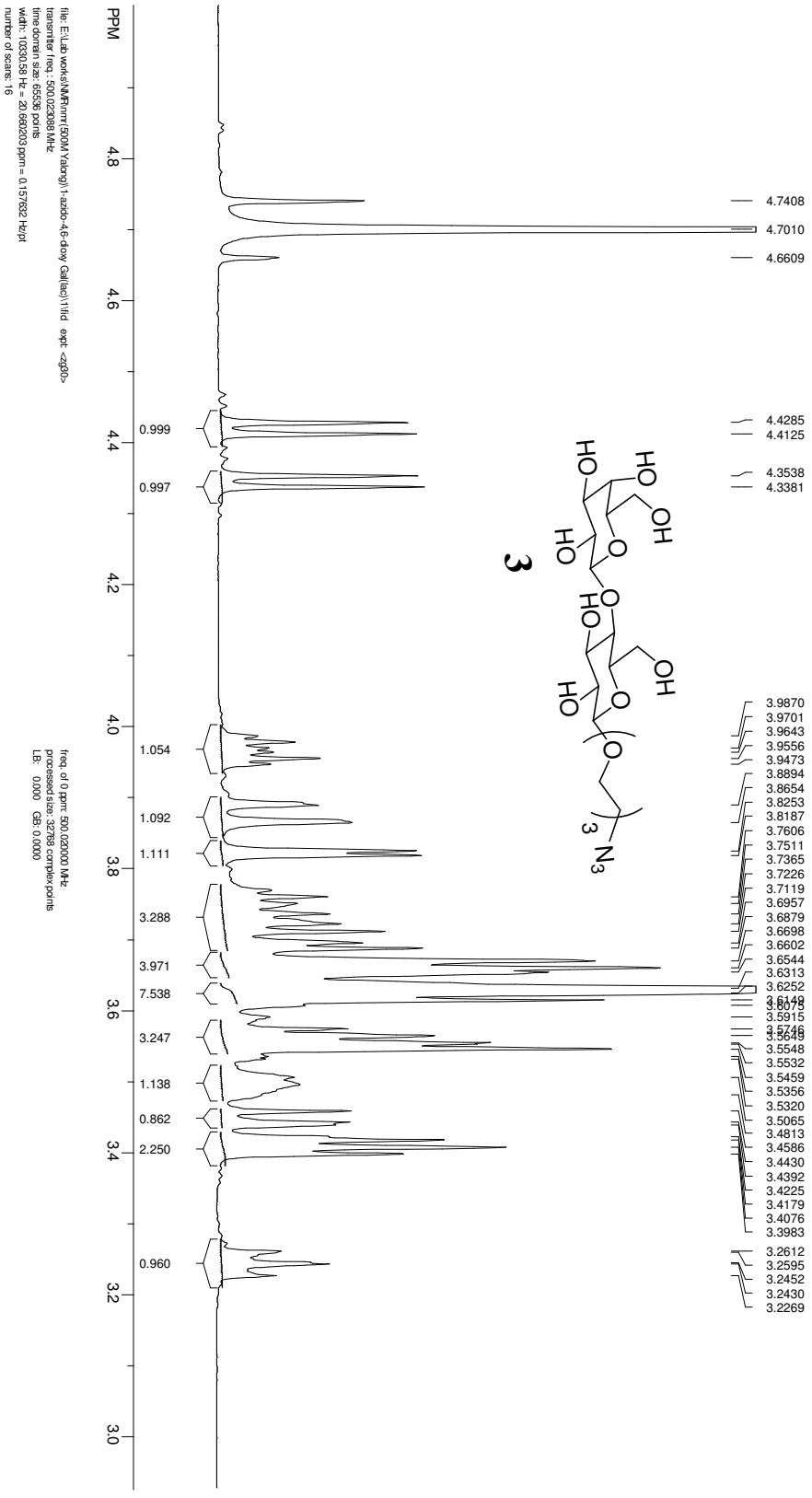
†Target stereochemistry from literature[62] § As reported by PROCHECK[63]

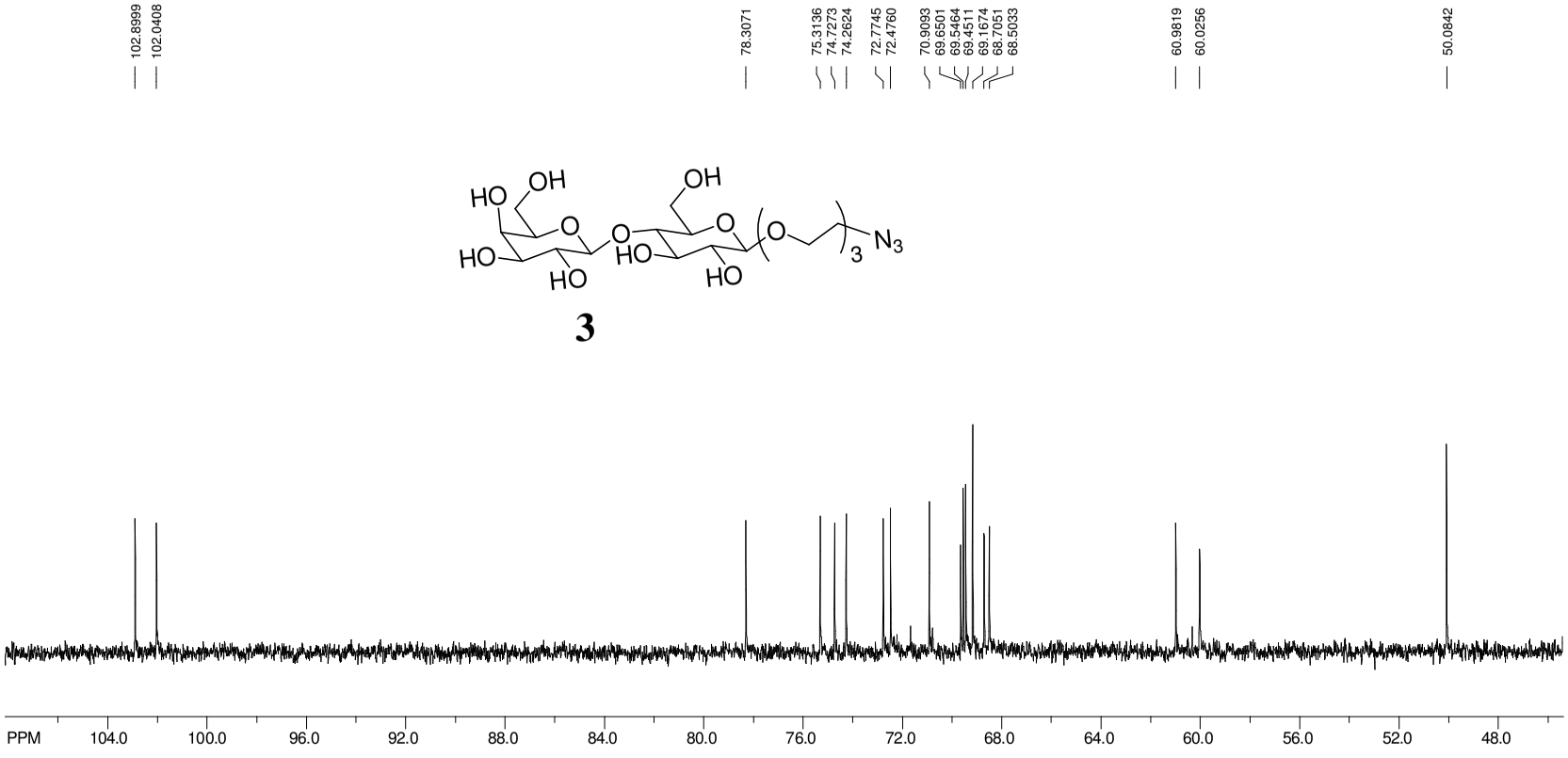

2
AcO
AcO
AcO
AcO
OAc
OAc
3
N3
7.2833
5.3608
5.3591
5.3541
5.3523
5.2240
5.2054
5.1869
5.1350
5.1192
5.1142
5.0984
4.9807
4.9737
4.9598
4.9529
4.5905
4.5747
4.4882
3.6270
3.6246
3.6209
3.6146
3.6109
3.6049
3.4163
3.4059
3.3960
2.1608
2.1383
2.1307
2.0720
2.0648
2.0548
2.0508
1.9750
1.000
0.970
0.990
1.013
0.987
0.976
2.026
3.109
2.049
1.057
1.096
9.290
1.997
3.083
3.029
3.087
3.089
3.022
PPM
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
file: E:\Lab works\NMR\nmr-linker\peracetyl LacPEG(3)azide-8-11\1\fid expt: <zg30>
transmitter freq.: 500.023088 MHz
time domain size: 65536 points
width: 10330.58 Hz = 20.660203 ppm = 0.157632 Hz/pt
number of scans: 16
freq. of 0 ppm: 500.020000 MHz
processed size: 32768 complex points
LB: 0.300 GB: 0.0000



file: E:\Lab works\NMR\nmr-linker\peracetyl LacPEG(3)azide-8-11\2\fid expt: <zgpg30>
transmitter freq.: 125.742702 MHz
time domain size: 65536 points
width: 30030.03 Hz = 238.821256 ppm = 0.458222 Hz/pt
number of scans: 897

freq. of 0 ppm: 125.730129 MHz
processed size: 32768 complex points
LB: 1.000 GB: 0.0000

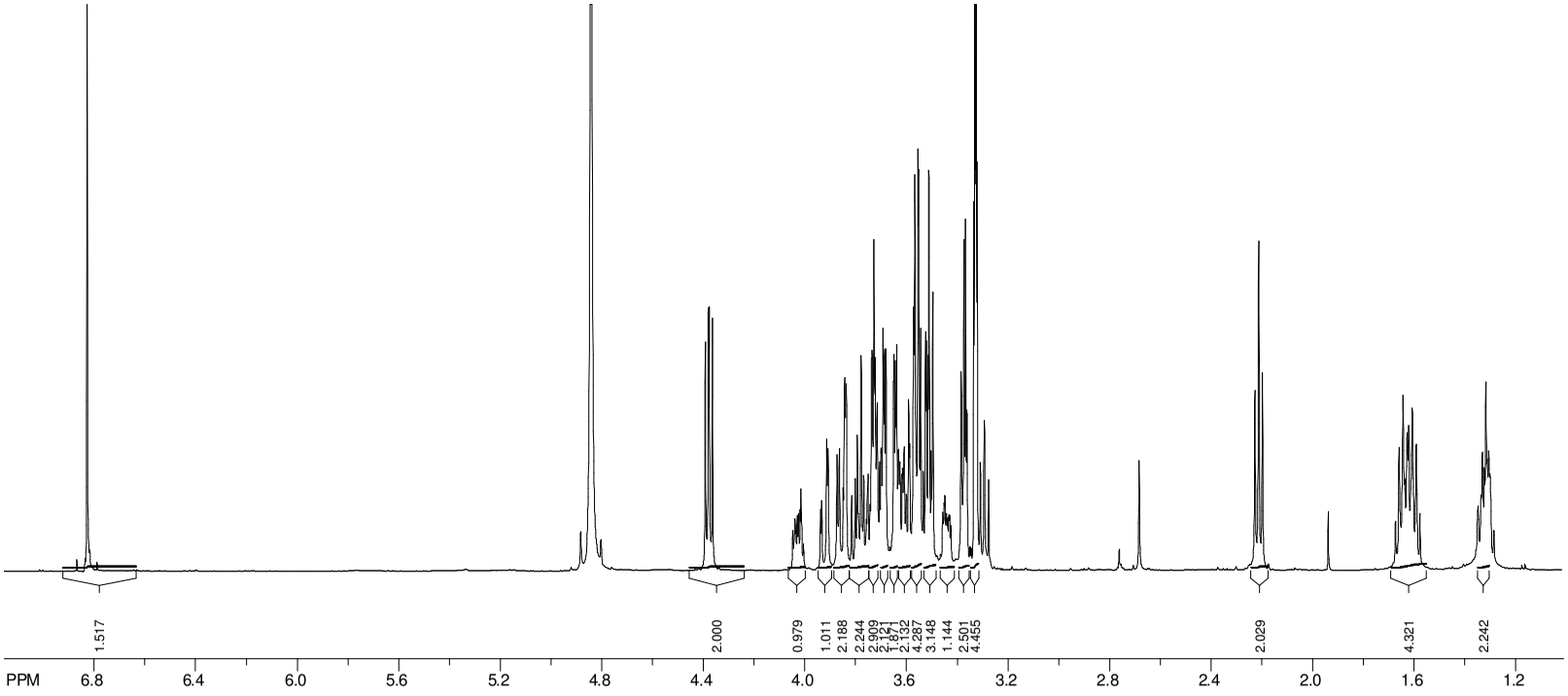

4.7408
4.7010
4.6609
4.4285
4.4125
4.3538
4.3381
3.9870
3.9701
3.9643
3.9556
3.9473
3.8894
3.8654
3.8253
3.8187
3.7606
3.7511
3.7365
3.7226
3.7119
3.6957
3.6879
3.6698
3.6602
3.6544
3.6313
3.6252
3.5915
3.5548
3.5532
3.5459
3.5356
3.5320
3.5065
3.4813
3.4586
3.4430
3.4392
3.4225
3.4179
3.4076
3.3983
3.2612
3.2595
3.2452
3.2430
3.2269
0.999
0.997
1.054
1.092
1.111
3.288
3.971
7.538
3.247
1.138
0.862
2.250
0.960
PPM
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
HO
OH
O
N3
3
file: E:\Lab works\NMR\nmr\500M Yalong\1-azido-4,6-doxy Gal(lac)\1\fid expt: <zg30>
transmitter freq.: 500.023098 MHz
time domain size: 65536 points
width: 10330.58 Hz = 20.662203 ppm = 0.157632 Hz/pt
number of scans: 16
freq. of 0 ppm: 500.020000 MHz
processed size: 32768 complex points
LB: 0.000 GB: 0.0000



3

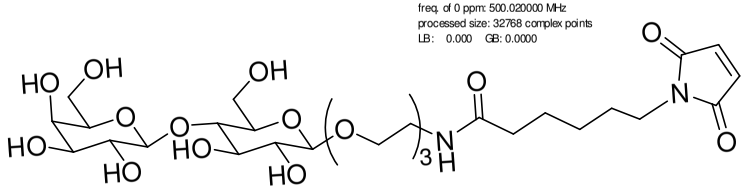
file: E:\Lab works\NMR\nmr(500M Yalong)\1-azido-4,6-dioxy Gal(lac)\2\fid expt: <zgpg30>
transmitter freq.: 125.742702 MHz
time domain size: 65536 points
width: 30030.03 Hz = 238.821256 ppm = 0.458222 Hz/pt
number of scans: 668

freq. of 0 ppm: 125.730129 MHz
processed size: 32768 complex points
LB: 0.000 GB: 0.0000

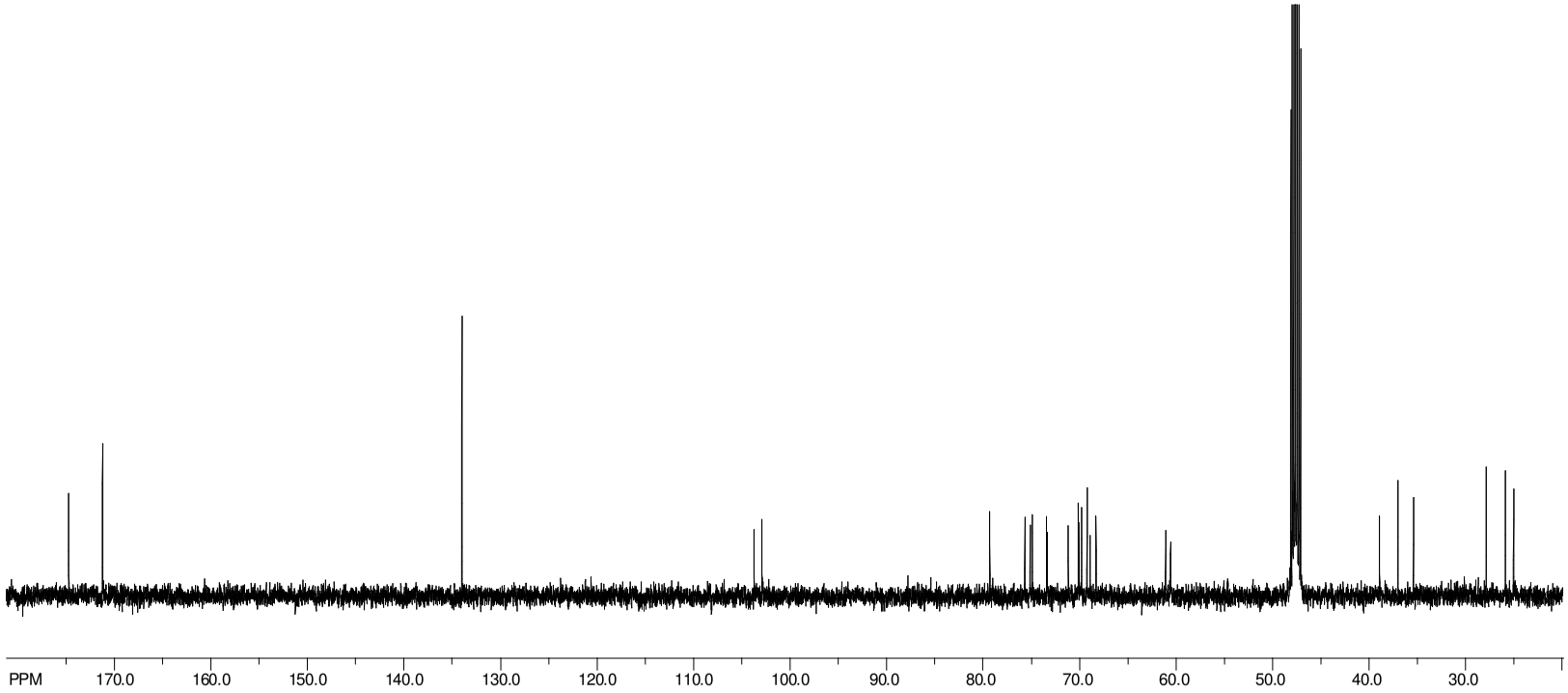


file: E:\Lab works\NMR\nmr(500M Yalong)\LacPEG(3)C(6)suc\1\fid expt: <zg30>
transmitter freq.: 500.023088 MHz
time domain size: 65536 points
width: 10330.58 Hz = 20.660203 ppm = 0.157632 Hz/pt
number of scans: 16

freq. of 0 ppm: 500.020000 MHz
processed size: 32768 complex points
LB: 0.000 GB: 0.0000

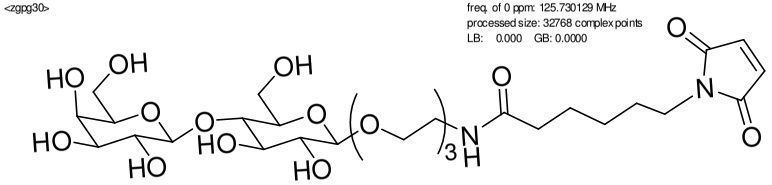


4



file: E:\Lab works\NMR\nmr(500M Yalong)\LacPEG(3)C(6)suc\2\fid expt: <zgpg30>
transmitter freq.: 125.742702 MHz
time domain size: 65536 points
width: 30030.03 Hz = 238.821256 ppm = 0.458222 Hz/pt
number of scans: 574

freq. of 0 ppm: 125.730129 MHz
processed size: 32768 complex points
LB: 0.000 GB: 0.0000



4