SUPPLEMENTARY MATERIAL FOR THE ARTICLE:

Profiling of the Glycoforms of the Intact α Subunit of Recombinant Human Chorionic Gonadotropin by High Resolution CE/MS

D. Thakur[1], T. Rejtar[1], B. L. Karger[1,*]

N.Washburn[2], C.J. Bosques[2], N.S.Gunay[2], Z.Shriver[2], G.Venkataraman[2,*]

[1]*Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115*

[2]*Momenta Pharmaceuticals, Cambridge, Massachusetts 02142*

Supplementary Materials and Methods:

Chemicals

Glacial acetic acid was obtained from Acros Organics (Morris Plains, NJ). HPLC grade acetonitrile and water were purchased from Thermo Scientific (Fairlawn, NJ). Trypsin (sequencing grade, modified) was from Promega (Madison, WI), and urea, ammonium bicarbonate, ammonium acetate, dithiothreitol and iodoacetamide were from Sigma-Aldrich. Peptide N-glycosidase F (PNGase F) was obtained from New England Biolabs and β(1-3,4,6)-galactosidase, sialidase A, β(1-2, 3, 4,6)-N-acetylhexosaminidase and α(1-3,4,6)-galactosidase were from Prozyme (San Leandro, CA).

Deglycosylation and Analysis of Released Glycans

For glycan structural analysis, r-αhCG was deglycosylated using PNGase F. The released N-glycans were purified from the digestion buffer and protein using a porous graphitized carbon (PCG) solid phase extraction cartridge (Thermo Scientific). The glycan pool was then labeled with 2-aminobenzamide (2-AB) using the Glyko Signal™ labeling kit (Prozyme), and the fluorescently-labeled glycans were purified from the reaction mixture using a GlycoClean G cartridge (Prozyme). The glycans were loaded and washed with 96% acetonitrile and subsequently eluted from the cartridge using LC-MS grade water. The 2-AB labeled glycans were then analyzed by LC-MS/MS using normal phase LC (Amide 80 column, 2 x 250 mm, Tosoh Biosciences) with fluorescence detection (excitation and emission wavelengths of 330 nm, and 420 nm). The HPLC was coupled to an LTQ ion trap mass spectrometer (Thermo Scientific), and the glycans were fragmented by low energy CID using a normalized collision energy of 35. The LTQ mass spectrometer was operated in the "triple play" mode with one full MS scan followed by 5 ultra zoom scans and 5 MS/MS scans. Relative abundances of the different glycan species were estimated from the peak area of the fluorescent chromatogram. In the cases where more than one glycan species was observed under one fluorescence peak, the relative abundance of each species under the peak was estimated from the area of the respective MS ions. It should be noted that, due to potential differences in glycan ionization efficiencies, the relative abundances for these glycans represent an approximation to the real abundances of the glycans in solution.

The glycan composition was determined based on the intact mass from the full MS scan. Glycan structures were then further interrogated using low energy collision-induced dissociation (CID). Fragments were assigned using the nomenclature proposed by Domon and Costello[1], and glycan fragments were identified through GlycoWorkBench[2]. Since the utility of exoglycosidase

enzyme arrays for N-glycan structural elucidation is well documented[3], the proposed structures were further confirmed using an exoglycosidase enzyme array designed to determine the nature of the non-reducing end structures. In short, the glycan pool was split into two aliquots and incubated overnight with an array of exoglycosidase enzymes; both aliquots were treated with β(1-3,4,6)-galactosidase (Prozyme), sialidase A (Prozyme) and β(1-2, 3, 4, 6)-N-acetylhexosaminidase (Prozyme). To determine the nature of the non-reducing end structures, one aliquot was incubated with α(1-3,4,6)-galactosidase (Prozyme) which is specific for terminal α-galactose capping structures. The two exoglycosidase treated pools were analyzed *via* HILIC LC-MS/MS, as discussed above. The species eluted from the Amide column were monitored using fluorescence detection (ex 330 nm; em 420 nm) and online mass spectrometry, as described above.

Trypsin Digestion of r-αhCG Expressed in a Murine Cell Line

r-αhCG obtained from Sigma-Aldrich (20 μg aliquot) was denatured in 100 μL of 8 M urea and 50 mM ammonium bicarbonate for 1 h at 37°C. The protein was reduced by 20 mM dithiothreitol for 3 h at 37°C. Subsequently, the protein was alkylated by 40 mM iodoacetamide for 90 minutes at room temperature. The solution was next transferred to a Microcon Ultracel YM-3 centrifugal filter device (Millipore) to remove urea. Trypsin (Promega, Madison, WI) was added to the protein solution (protein to enzyme ratio 20:1), and digestion was allowed to proceed for 12 hrs at 37°C before stopping the reaction by addition of 0.1% formic acid.

LC-MS Analysis of r-αhCG Tryptic Digest

The tryptic digest of r-αhCG was separated using a 75 μm i.d. 20 cm long column packed with 3 μm Magic C18AQ (Michrom Bioresources, Auburn, CA) with mobile phase A as 0.1% (v/v) formic acid in water and mobile phase B as 0.1% (v/v) formic acid, 90% (v/v) acetonitrile in water. A linear gradient of mobile phase B from 5% to 80% in 90 min was used to elute the peptides. Both MS and MS/MS data were acquired on an LTQ MS (Thermo Scientific) using data dependent acquisition with every MS scan followed by MS/MS spectra (normalized collision energy of 28) for up to 8 precursors with dynamic exclusion set to 30 sec . Database searching was performed using Sequest within the Bioworks Browser ver. 3.3.1 (Thermo Scientific) with the SwissProt mouse protein database (ver. 55.1) appended with the sequence of r-αhCG. Tryptic digests of both native and deglycosylated r-αhCG were analyzed to facilitate identification of the chromatographic peaks of glycopeptides. MS/MS spectra of glycopeptides were interpreted manually.

Supplementary Results

Analysis of the Released Glycan: Characterization of Gal-α1-3Gal containing species and sulfated and pentasialylated glycans

The analysis procedure for the released glycans is detailed in the above Supplementary Material and methods Section. Initial examination of the different glycans revealed a number of species with a neutral mass, suggesting the presence of an additional hexose residue (i.e. $HexNAc_N Hex_{N+2}$). One of the most abundant of these glycans had an observed m/z of 1108.2 $[M+2H]^{2+}$ and composition of $HexNAc_4Hex_6NeuAc$. Low energy CID of this species suggested

the presence of a Gal-α-Gal disaccharide as a terminal non-reducing capping structure (**Figure S3B**). Secondly, the presence of significant levels of this and other N-glycans containing Gal-α-Gal was confirmed through the use of a targeted exoglycosidase enzyme array. Enzymatic digestion with α-galactosidase, in conjunction with sialidase A, β-galactosidase, and β-N-acetylhexosaminodase, resulted in all species collapsing down to the trimannosyl-chitobiose core (**Figure S4A**), whereas in the absence of the α-galactosidase enzyme, treatment with the other three enzymes resulted in resistant species, which contained the non-reducing end Gal-α-Gal (**Figure S4B**). Susceptibility to α-galactosidase has been extensively used by others to confirm the presence of α-galactose moieties at the non-reducing end of N-glycans[4]. However, since α(1-3,4,6)-galactosidase can cleave terminal galactose residues linked in either an α-1-3, 1-4 or 1-6 linkage, the exact linkage within the Gal-α-Gal structure could not be unequivocally assigned. Nevertheless, previous studies have identified the immunogenic Gal-α1-3-Gal epitope in mouse cell surface glycans,[5, 6] and the enzymatic activity necessary to form these structures has also been well documented in many murine cell types[7]. The combined enzymatic and mass spectrometric approach strongly suggest the presence of significant levels of glycans containing Gal-α1-3-Gal in r-αhCG in the murine cell line. Combining the results from HPLC and MS analyses, we estimate that as much as 60% of the glycans in r-αhCG from the murine cell line contained this important immunogenic epitope.

The neutral mass of several species suggested the presence of low levels (<1%) of sulfated N-glycans, for example, the species with m/z 1212.7 $[M+2H]^{2+}$ (neutral mass of 2423.4 Da) corresponding to a composition of $HexNAc_4Hex_5NeuAc_2$+sulfate. The presence of sulfated structures was confirmed by fragmentation analysis (**Figure S3A**). In addition, these

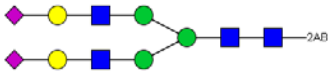
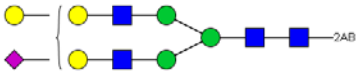
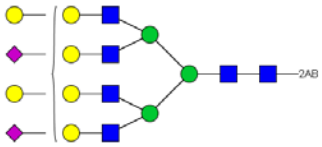
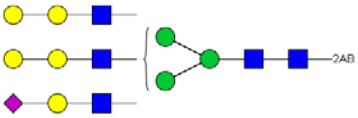
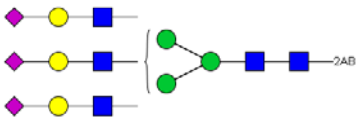
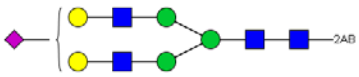
species were found to be resistant to digestion with β-galactosidase on the sulfated antennae, further confirming their presence (data not shown).

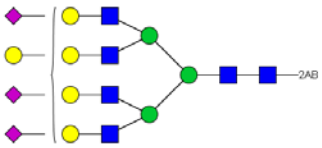
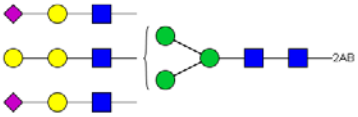
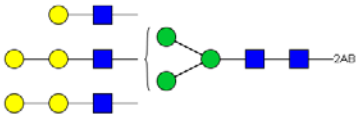
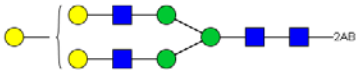
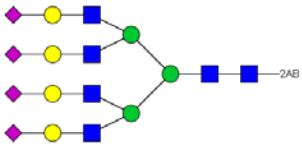
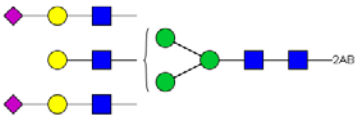
Additional examination of the glycans revealed several complex species with masses that suggested compositions with more than four non-reducing end capping structures. These species contain from two to five N-acetylneuraminic acid moieties and from zero to three α-galactose capping structures. One of these species was found to produce a weak signal at m/z of 1559.6 $[M+3H]^{3+}$ (neutral mass of 4676.5 Da), corresponding to the possible presence of a pentasialylated N-glycan with composition of $HexNAc_8Hex_9NeuAc_5$ (see **Table S1**). Another example of this unusual structure can be seen from the glycan with a composition of $HexNAc_8Hex_{10}NeuAc_4$, based on an observed m/z of 1517.1 $[M+3H]^{3+}$ and a neutral mass of 4547.6 Da. Despite the low intensity of the parent ions, the fragment spectra for these two species suggest the presence of a disialylated antenna (**Figure S5A**). CID fragmentation was able to identify the proposed structure as a likely candidate, and this type of glycosylation has been observed previously in murine transferrin[8] and in other murine urinary proteins[9]. Pentasialylated glycans have also been previously observed in the CE-MS analysis of EPO derived from CHO cells[10]. It is also possible that these structures could contain a disialic acid in the same antennae as has been observed in recombinant murine ICAM-1 expressed in CHO cells [11] and in murine NCAM [12]. Finally, CID fragmentation for multiple species also revealed the presence of N-acetyl lactosamine extended antennae, as determined by the characteristic B-ions (**Figure S5B**).

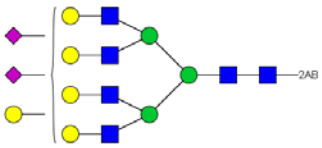
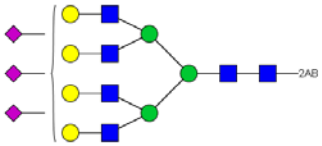
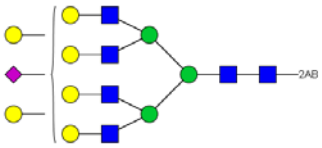
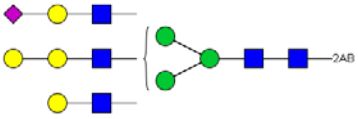
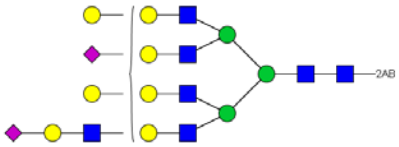
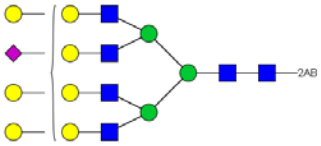
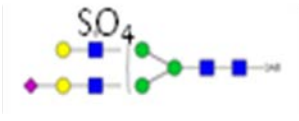
Overall, biantennary species constitute roughly 46% of the total glycan pool, triantennary species ~25 % and tetraantennary glycans ~29%. The tetraantennary species contain a small subset of glycans which possess more than 4 non-reducing end capping structures. (It needs to be emphasized that these latter glycans constitute less than 2% of the total

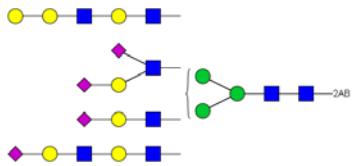
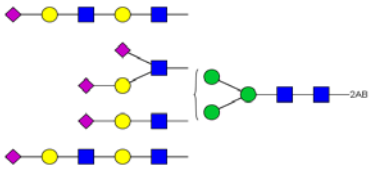
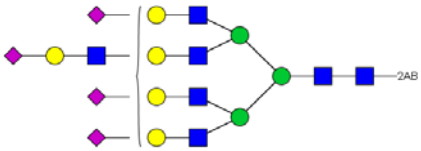
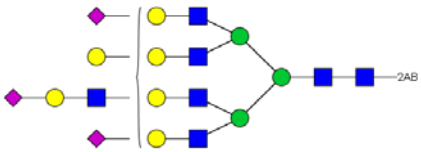
N-glycan pool.) The pool contains primarily monosialylated, disialylated and trisialylated glycans which constitute approximately 33%, 38% and 14% of the total , respectively. Neutral glycans represent approximately 8% of the total pool, and sulfated, tetrasialylated and pentasialylated glycans constitute 3%, 4% and less than 1% , respectively. Interestingly, core fucosylation was not observed in any of the released N-glycans.

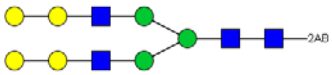
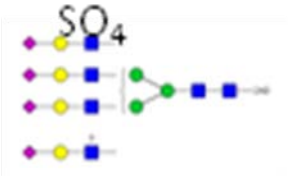
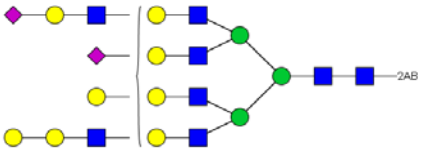
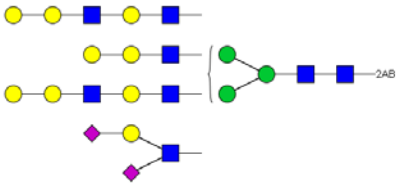
Table S1. Summary of N-linked glycans detected in r-αhCG. Relative abundance, composition and most probable structure are given for each N-glycan.

No.	Relative abundance (%)	m/z / Neutral mass	Composition with most probable structure
1	18	1172.5/ 2342.9	 $HexNAc_4Hex_5NeuAc_2$
2	16	1108.2/ 2213.8	 $HexNAc_4Hex_6NeuAc_1$
3	9	1133.3/ 3397.2	 $HexNAc_6Hex_9NeuAc_2$
4	7	914.3/ 2741.0	 $HexNAc_5Hex_8NeuAc_1$
5	6	1000.2/ 2999.1	 $HexNAc_5Hex_6NeuAc_3$
6	6	1027.1/ 2051.8	 $HexNAc_4Hex_5NeuAc_1$

7	5	1176.8/ 3526.3	 $HexNAc_6Hex_8NeuAc_3$
8	5	958.0/ 2870.0	 $HexNAc_5Hex_7NeuAc_2$
9	4	1225.8/ 2449.9	 $HexNAc_5Hex_8$
10	3	962.3/ 1922.7	 $HexNAc_4Hex_6$
11	3	1219.8/ 3655.8	 $HexNAc_6Hex_7NeuAc_4$
12	2	904.0/ 2708.0	 $HexNAc_5Hex_6NeuAc_2$

13	2	1079.3/ 3235.2	 $HexNAc_6Hex_8NeuAc_2$
14	1	1122.7/ 3364.2	 $HexNAc_6Hex_7NeuAc_3$
15	1	1036.8/ 3106.1	 $HexNAc_6Hex_9NeuAc_1$
16	1	1290.2/ 2578.9	 $HexNAc_5Hex_7NeuAc_1$
17	1	1255.5/ 3762.4	 $HexNAc_7Hex_{10}NeuAc_2$
18	1	1090.7/ 3268.2	 $HexNAc_6Hex_{10}NeuAc_1$
19	<1	1067.3/ 2131.7	 $HexNAc_4Hex_5NeuAc_1$**+SO_4**

20	<1	1517.1/ 4547.6	 $HexNAc_8Hex_{10}NeuAc_4$
21	<1	1559.1/ 4676.1	 $HexNAc_8Hex_9NeuAc_5$
22	<1	1027.3/ 3079.1	 $HexNAc_5Hex_6NeuAc_3$**+SO_4**
23	<1	881.4/ 1760.7	 $HexNAc_4Hex_5$
24	<1	1341.2/ 4020.4	 $HexNAc_7Hex_8NeuAc_4$
25	<1	1298.2/ 3891.4	 $HexNAc_7Hex_9NeuAc_3$

26	<1	1212.8/ 2422.6	 $HexNAc_4Hex_5NeuAc_2$**+SO_4**
27	<1	1160.8/ 3477.8	 $HexNAc_6Hex_9NeuAc_2$**+SO_4**
28	<1	1043.7/ 2084.8	 $HexNAc_4Hex_7$
29	<1	1246.3/ 3735.2	 $HexNAc_6Hex_7NeuAc_3$**+SO_4**
30	<1	1377.1/ 4128.4	 $HexNAc_8Hex_{11}NeuAc_2$
31	<1	1430.7/ 4289.5	 $HexNAc_8Hex_{12}NeuAc_2$

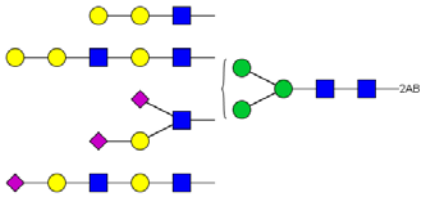
32	<1	1473.7/ 4418.2	 $HexNAc_8Hex_{11}NeuAc_3$

Table S2. Summary of glycopeptides for site 1 and 2 observed in LC-MS analysis of tryptic digest of r-αhCG produced on murine cell line.

Glycan[a]	**Glycan abundance**[b]	**Glycan Composition**	**Peak area for glycosylation site**[c] **N76 (%)**	**Peak area at glycosylation site**[c] **N102 (%)**
1	18%	$HexNAc_4Hex_5NeuAc_2$	7.4	40.7
2	16%	$HexNAc_4Hex_6NeuAc$	4.5	22.0
3	9%	$HexNAc_6Hex_9NeuAc_2$	9.8	1.1
4	7%	$HexNAc_5Hex_8NeuAc$	30.0	1.7
5	6%	$HexNAc_5Hex_6NeuAc_3$	3.7	11.0
6	6%	$HexNAc_4Hex_5NeuAc$	1.0	0.2
7	5%	$HexNAc_5Hex_7NeuAc_2$	17.4	7.9
8	5%	$HexNAc_6Hex_8NeuAc_3$	4.4	1.7
9	4%	$HexNAc_5Hex_8$	0.4	0.1
10	3%	$HexNAc_4Hex_6$	ND	1.6
11	3%	$HexNAc_6Hex_7NeuAc_4$	1.0	1.9
12	2%	$HexNAc_6Hex_8NeuAc_2$	3.7	1.1
13	2%	$HexNAc_5Hex_6NeuAc_2$	2.1	3.7
14	1%	$HexNAc_5Hex_7NeuAc$	5.6	1.4
15	1%	$HexNAc_6Hex_9NeuAc$	4.1	0.6
16	1%	$HexNAc_6Hex_7NeuAc_3$	1.0	2.1
17	1%	$HexNAc_7Hex_{10}NeuAc_2$	0.8	0.4
18	1%	$HexNAc_6Hex_{10}NeuAc$	3.2	0.1

[a] assignment of glycan in **Table 1**. Only glycans with >1% abundance were considered.

[b] abundance of the glycan from **Table 1**.

[c] Peak area was normalized to 100% at each glycosylation site.

Table S3 is in Excel file.

Figure S1:

Diagram of CE-MS system for analysis of intact glycoproteins.

a) BGE reservoir; **b)** separation capillary (20 cm, 50 μm i.d., PVA coated fused silica); **c)** liquid junction; **d)** ESI interface with metal tip, 50 μm i.d.; **e)** pressure valve and manometer; **f)** syringe for replacement of ESI solution.

Figure S2:

(A) HILIC separation with fluorescence detection for 2-AB labeled N-glycans derived from r-αhCG. Over 20 chromatographic peaks were resolved during the 110 minute run. N-glycan compositions were assigned based on the neutral mass of each species; the structures of the N-glycans were determined based on MS/MS fragmentation. **(B)** MS/MS fragmentation of a common N-glycan species as a representative example. N-glycan fragmentation is characterized by cleavage between the antennary N-acetylglucosamine and the trimannosyl chitobiose core. Complementary B/Y and B'/Y' ions ($Y_{4\beta}/B_{3\beta}$ and $Y_{4\alpha}/B_{3\alpha}$ at m/z 366.1/1687.5 and 657.2/1396.4) are the most prominent ions formed, allowing for structural characterization of this species. Importantly, additional assignments exist for many of the fragments observed; fragments were assigned the most likely structure based on the minimum number of bond cleavages.

Figure S3:

LC/MS/MS analysis of sulfated and α-galactose containing N-glycans. (**A**) Fragmentation analysis of N-glycan species containing a sulfate with neutral mass of 2423.4 Da. Abundant B–ions seen which help to confirm the antennary structure at m/z 657.27, corresponding to the sialylated lactosamine antenna, and at 737.38, corresponding to the sulfated sialylated

lactosamine antenna. (**B**) Fragmentation analysis of an α-galactose containing N-glycan. Abundant B–ions are observed which help to confirm the antennary structure at m/z 657.27 corresponding to the sialylated lactosamine antenna and, at 528.18, corresponding to the α-galactose capped lactosamine antenna.

Figure S4:

Exoglycosidase characterization of a galactose-α-galactose containing species. (**A**) HILIC fluorescence chromatogram of N-glycan pool treated with sialidase A, β-galactosidase, α-galactosidase and β-N-acetylhexosaminodase. (**B**) Glycan pool treated with sialidase A, β-galactosidase and β-N-acetylhexosaminodase.

Figure S5

LC/MS/MS fragmentation of two N-glycan species having 5 apparent non-reducing end structures. Fragmentation of these N-glycans are characterized by prominent B-ions corresponding to the four proposed antennae. (**A**) B-ions are seen which correspond to the sialylated dilactosamine antenna (m/z 1022.2), the disialylated lactosamine antenna (m/z 948.4), the α1-3Gal capped dilactosamine antenna (m/z 894.2) and the sialylated lactosamine antenna (m/z 657.2). (**B**) B-ions are seen which correspond to the sialylated dilactosamine antenna (m/z 1022.5), the disialylated lactosamine antenna (m/z 947.9), and the sialylated lactosamine antenna (m/z 657.2).

References:

(1) Domon, B.; Costello, C. *Biochemistry* 1988, *27*, 1534-1543.

(2) Ceroni, A.; Dell, A.; Haslam, S. *Source Code Biol Med* 2007, 2:3.

(3) Edge, C. J.; Rademacher, T. W.; Wormald, M. R.; Parekh, R. B.; Butters, T. D.; Wing, D. R.; Dwek, R. A. *Proc Natl Acad Sci USA* 1992, *89*, 6338-6342.

(4) Kim, Y.-G.; kim, S.-Y.; Hur, Y.-M.; Joo, H.-S.; Chung, J.; Lee, D.-S.; Royle, L.; Rudd, P. M.; Dwek, R. A.; Harvey, D. J.; Kim, B.-G. *Proteomics* 2006, *6*, 1133-1142.

(5) Chung, T. W.; Kim, K. S.; Moon, S. K.; Lee, J. W.; Song, E. Y.; Chung, T. H.; Yeom, Y. I.; Kim, C. H. *Moll. Cells* 2003, *16*, 343-353.

(6) Easton, R. L.; Patankar, M. S.; Lattanzio, F. A. L., T.H; Morris, H. R.; Clark, G. F.; Dell, A. *J.Biol.Chem* 2000, *275*, 7731-7742.

(7) Joziasse, D. H.; Shaper, N. L.; Kim, D.; van den Eijnden, D. H.; Shaper, J. H. *J.Biol.Chem* 1992, *267*, 5534-5541.

(8) Coddeville, B.; Regoeczi, E.; Strecker, G.; Plancke, Y.; Spik, G. *Biochimica et Biophysica Acta* 2000, *1475*, 321-328.

(9) Mechref, Y.; Chen, P.; Novotony, M. *Glycobiology* 1999, *3*, 227-234.

(10) Balaguer, E.; Demelbauer, U.; Pelzing, M.; Sanz-Nebot, V.; Barbosa, J.; Neusub, C. *Electrophoresis* 2006, *27*, 2638-2650.

(11) Otto, V. I.; Damoc, E.; Cueni, L. N.; Schurpf, T.; Frei, R.; Ali, S.; Callewaert, N.; Moise, A.; Leary, J. A.; Folkers, G.; Przybylski, M. *Glycobiology* 2006, *16*, 1033-1044.

(12) Kojima, N.; Tachida, Y.; Yoshida, Y.; Tsuji, S. *J.Biol.Chem* 1996, *271*, 19457-19463.