Polylactosaminoglycan Glycomics: Enhancing the Detection of High-molecular-weight N-glycans in Matrix-assisted Laser Desorption Ionization Time-of-flight Profiles by Matched Filtering

Marshall Bern‡§, Alejandro E. Brito‡, Poh-Choo Pang¶, Angad Rekhi¶, Anne Dell§¶, and Stuart M. Haslam§¶

For over 30 years, protocols based on the mass spectrometry (MS) of permethylated derivatives, complemented by enzymatic degradations, have underpinned glycomic experiments aimed at defining the structures of individual glycans present in the complex mixtures that are characteristic of biological samples. Both MS instrumentation and sample handling have improved markedly in recent years, enabling greater sensitivity and better signal-to-noise ratios, thereby facilitating the detection of glycans at much higher masses than could be achieved in the past. The latter is especially important for the characterization of the biologically important class of N-glycans that carry polylactosaminoglycan chains. Such advances in data acquisition heighten the need for informatics tools to assist in glycan structure assignment. Here, utilizing mouse lung tissue as a model system, we present evidence of polylactosaminoglycan-containing N-glycans with permethylated molecular weights exceeding 13 kDa. We show that antennae branching patterns and lengths can be successfully determined at these high masses via MS/MS experiments, even when MS ion counts are very low. We also describe the development and application of a matched filtering algorithm for assisting high-molecular-weight glycan detection and structure assignment. Molecular & Cellular Proteomics 12: 10.1074/mcp.0112.026377, 996–1004, 2013.

Glycosylation is one of the most common and important post-translational modifications of proteins, yet it is also one of the most difficult to study because of its great complexity. Much of what we know about glycosylation, especially tissue- and organism-specific variation, has been gathered via mass spectrometric profiling of detached glycans (1–4). International, multi-laboratory comparisons of MS1 and chromatographic glycomic data derived from standardized glycoproteins have concluded that the MS analysis of permethylated derivatives of released N- and O-linked glycans is the most robust glycomics strategy with respect to both sensitivity and quantitative reliability (5, 6). MS/permethylation strategies were first implemented in the Imperial laboratory in the early 1980s, when fast atom bombardment MS was revolutionizing the analysis of glycoconjugates (7). Notably, with the exception of MALDI superseding fast atom bombardment as the preferred method of MS ionization in the past decade, permethylation-based glycomic protocols remain virtually unchanged from those employed 30 years ago. Where there has been a dramatic change, however, is in the quality and quantity of data emerging from glycomics investigations. This is exemplified by the success of the Analytical Glycomics Core of the Consortium for Functional Glycomics, which has built a public database of MALDI-TOF profiles from human and mouse tissues and cell lines as a resource for researchers worldwide.

Characterizing complex mixtures of most classes of N-and/or O-glycans is now a routine task for glycomics experts when working with a few micrograms of a glycoprotein sample or, for cell glycomics, about a million cells. However, this is not the case for an important class of N-glycans, namely, those that carry polylactosaminoglycan antennae composed of [3Galβ1–4GlcNAcβ1], or its branched counterpart, [3(Galβ1–4GlcNAcβ1–6)Galβ1–4GlcNAcβ1]. These glycans are challenging because of their large size and because they are often minor constituents of the total glycan population. Nevertheless, they merit detailed investigation because they are implicated in many biological processes (8–12). Much of our current understanding of polylactosaminoglycan

From the ‡Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304; ¶Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, United Kingdom
Received December 4, 2012, and in revised form, January 14, 2013
Published, MCP Papers in Press, January 16, 2013, DOI 10.1074/mcp.O112.026377

1 The abbreviations used are: MALDI-TOF-MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; MS/MS, tandem mass spectrometry.

© 2013 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

This is an Open Access article under the CC BY license.
structure arises from fast atom bombardment experiments on hematopoietic glycoproteins (13–18). These studies indicate that the longest antennae are most abundant on the 6-arm of bi-antennary glycans and that tri- and tetra-antennary glycans tend to carry multiple polyolactosaminoglycan antennae. MALDI glycomics of rabbit erythrocytes (19), Chinese hamster ovary cells (20), human T cells (21) and human umbilical vein endothelial cells (22) have endorsed these early findings. This research has provided a wealth of data on the compositions of large polyolactosaminoglycan-containing glycans, albeit with very little information concerning antennae lengths or arrangements on the core glycans.

In recent years, as a result of improved MS performance and optimization of sample handling, exceptionally large N-glycans have become detectable in glycomics mass spectra, with the record to date being a glycan of composition \( \text{Fuc}_{10}\text{Hex}_{25}\text{HexNAc}_{28} \) (m/z 13,024.3) present in the N-glycome of Chinese hamster ovary cells (20). Such high-molecular-weight glycans are especially challenging to detect because of their low relative abundance, signals spread over many peaks, chemical “noise,” and dynamic range limitations of MS instruments. Nevertheless, because of their potential biological significance, we believe it is important that attempts be made to address the challenges that large, low-abundance glycans pose for detailed characterization and to develop tools and methodologies that will facilitate their detection and the elucidation of structural features such as antennae length and branching patterns.

Utilizing mouse lung tissue as a model system, we have now detected polyolactosaminoglycan-containing N-glycans with permethylated molecular weights exceeding 13 kDa. We show that antennae branching patterns and lengths can be successfully determined at these high masses via MS/MS experiments, even when MS ion counts are very low. We also describe the development and application of an especially sensitive matched filtering algorithm for assisting high-molecular-weight glycan detection and structure assignment.

**EXPERIMENTAL PROCEDURES**

**Samples**—Lung tissues (~0.3 g) from wild-type C57BL/6 mice were produced by the Mouse Phenotype Core at the Consortium for Functional Glycomics.

**Sample Preparation, MS Data Acquisition, and Manual Data Analysis**—Sample processing for N-glycomic profiling of the mouse tissues was carried out as detailed previously (3, 4). The lung tissue preparations were subjected to homogenization, reduction, carboxymethylation, and tryptic digestion. Peptide N-glycosidase F digestion of the purified tryptic glycopeptides was carried out in 50 mM ammonium bicarbonate, pH 8.5, for 24 h at 37 °C (Roche Applied Science, Burgess Hill, UK). The released N-glycans were purified using a Sep-Pak C18 cartridge (Waters Corp., UK). The purified native N-glycans were subsequently permethylated using the sodium hydroxide procedure and purified using a Sep-Pak C18 cartridge. The permethylated N-glycans were then dissolved in methanol before an aliquot was mixed at a 1:1 ratio (v/v) with 10 mg/ml 3,4-diaminobenzophenone in 75% acetonitrile. The glycans were subsequently permethylated in order to enhance the sensitivity of the mass spectrometric analysis and to direct the fragmentation within tandem MS analyses (Fig. 1A).

**Glycomic Strategy**—The mouse lung was prepared for MS via detergent extraction of glycoproteins, reduction, carboxymethylation, and digestion with trypsin to generate peptides/glycopeptides. N-glycans were released from extracted glycopeptides by means of peptide N-glycosidase F digestion, and O-glycans were chemically released via reductive elimination from the glycopeptides remaining after the release of N-glycans. The glycans were permethylated in order to reduce chemical noise and enhance the sensitivity of the mass spectrometric analysis. The glycans were permethylated in order to enhance the sensitivity of the mass spectrometric analysis and to direct the fragmentation within tandem MS analyses (Fig. 1A).

**Glycomic Profiling of Mouse Lung N-glycans**—The MALDI-TOF MS-MS data were obtained using a 4800 MALDI-TOF/TOF mass spectrometer (AB Sciex UK Limited) in the positive ion mode. For MS/MS, the collision energy was set at 1 kV, and argon was used as the collision gas. Typically, MS data were obtained from 0.3% of the total sample, and MS/MS data from 6% of the total sample. The obtained MS and MS/MS data were viewed and processed using Data Explorer 4.9 (AB Sciex UK Limited). The manual assignment of the glycan sequence was done on the basis of knowledge of mammalian biosynthetic pathways and MS/MS of selected molecular ions, and with the aid of a glycomics tool, GlycoWorkBench (23).

**GC/MS Linkage Analysis**—GC/MS linkage analysis of partially permethylated alditol acetates was carried out on a PerkinElmer Life Sciences Clarus 500 instrument fitted with an RTX-5 fused capillary column (Restec Corp.) as described elsewhere (3, 4).

**Software Data Analysis**—The developed detection algorithms were implemented in MATLAB (MathWorks, Natick, MA).

**RESULTS**

**Glycomic Strategy**—The mouse lung was prepared for MS via detergent extraction of glycoproteins, reduction, carboxymethylation, and digestion with trypsin to generate peptides/glycopeptides. N-glycans were released from extracted glycopeptides by means of peptide N-glycosidase F digestion, and O-glycans were chemically released via reductive elimination from the glycopeptides remaining after the release of N-glycans. The glycans were permethylated in order to enhance the sensitivity of the mass spectrometric analysis and to direct the fragmentation within tandem MS analyses (Fig. 1A).

**Glycomic Profiling of Mouse Lung N-glycans**—The MALDI-TOF glycomic profiling of permethylated mouse lung N-glycans produced spectra rich in molecular ion signals ranging from m/z 1579.5–13,386.7. The low-mass spectrum was dominated by complex type N-glycans with compositions consistent with bi-, tri-, and tetra-antennary structures with N-acetyllactosamine tandem repeats (Fig. 1B, supplemental Table S1, supplemental Fig. S1). The major non-reducing end capping groups were sialic acid in both NeuAc and NeuGc forms, with the latter being significantly more abundant at lower molecular weights. Of note are glycans at m/z 3244, 4084, and 4924 whose compositions indicate the presence of more NeuGc residues than potential antennae. Glycans with the Galα1–3Gal sequence (for example, m/z 2652.8 and 2838.8) also were present. The complex glycans had a maximum of one fucose. A striking feature of the glycomic data is the large size of the N-glycans that could be detected, up to a proposed composition of NeuAc₁HexNAC₂₈Hex₂₉Fuc₁, m/z 13,386 (Fig. 2, supplemental Table S1).

**Detailed Structural Assignment via MS/MS**—In order to provide a higher level of N-glycan structural assignment, selected molecular ions up to m/z 13,025 were subjected to MALDI-TOF-TOF MS/MS analysis. Fig. 3A shows the MS/MS profile of a molecular ion of m/z 4084 (NeuGc₂Hex₉HexNAc₂₈) that is predicted to be a tri-anentary structure with either two NeuGc residues attached in a linear fashion or a structure in which the second NeuGc residue is attached to the 6-position of Chinese hamster ovary cells (20), human T cells (21) and human umbilical vein endothelial cells (22) have endorsed these early findings. This research has provided a wealth of data on the compositions of large polyolactosaminoglycan-containing glycans, albeit with very little information concerning antennae lengths or arrangements on the core glycans. In recent years, as a result of improved MS performance and optimization of sample handling, exceptionally large N-glycans have become detectable in glycomics mass spectra, with the record to date being a glycan of composition \( \text{Fuc}_{10}\text{Hex}_{25}\text{HexNAc}_{28} \) (m/z 13,024.3) present in the N-glycome of Chinese hamster ovary cells (20). Such high-molecular-weight glycans are especially challenging to detect because of their low relative abundance, signals spread over many peaks, chemical “noise,” and dynamic range limitations of MS instruments. Nevertheless, because of their potential biological significance, we believe it is important that attempts be made to address the challenges that large, low-abundance glycans pose for detailed characterization and to develop tools and methodologies that will facilitate their detection and the elucidation of structural features such as antennae length and branching patterns.

Utilizing mouse lung tissue as a model system, we have now detected polyolactosaminoglycan-containing N-glycans with permethylated molecular weights exceeding 13 kDa. We show that antennae branching patterns and lengths can be successfully determined at these high masses via MS/MS experiments, even when MS ion counts are very low. We also describe the development and application of an especially sensitive matched filtering algorithm for assisting high-molecular-weight glycan detection and structure assignment.
of the GlcNAc. The B-ion at \( m/z \) 1268 and the Y-ion at \( m/z \) 2838 confirm the presence of a di-NeuGc antenna. The presence of the BZ-ion at \( m/z \) 641 (indicating a NeuGc linked to GlcNAc) and the lack of a B-ion at \( m/z \) 819 (which would have been present if the two NeuGc residues were linked together) and a Y-ion at \( m/z \) 3288 confirm that the second NeuGc is attached to the 6-position of the GlcNAc.

Fig. 3B shows the MS/MS profile of a molecular ion of \( m/z \) 10,330 (Hex23HexNAc22Fuc). The spectrum is dominated by both B- and Y-fragment ions, which indicates that after an initial HexHexNAc loss, subsequently Hex2HexNAc2 fragments are lost. This is consistent with the presence of branched polylacNAc antennae containing up to 11 lacNAc units. The Y-ion at \( m/z \) 474 and the lack of any fucosylated antennal ions confirm that the single fucose residue is located exclusively on the N-glycan core.

Fig. 3C shows the MS/MS profile of a molecular ion of \( m/z \) 7458 (NeuAc2Hex15HexNAc14Fuc). The spectrum is dominated by two series of B- and Y-fragment ions, one originating for a sialylated antenna and the other from a non-sialylated antenna. However, unlike in the previous example, the fragment ions are spaced by single HexHexNAc units, which is consistent with the presence of linear polylacNAc antennae containing up to nine lacNAc units. Again, the Y-ion at \( m/z \) 474 and the lack of any fucosylated antennal ions confirm that the single fucose residue is located exclusively on the N-glycan core.

GC/MS Linkage Analysis of Mouse Lung N-Glycans—Important additional structural conclusions can be drawn from the linkage data (supplemental Table S2). High levels of 3,6-linked mannose and 4-linked GlcNAc were in accordance with these residues’ being constituents of the core of all N-gly-
Fucosylated cores in the N-glycan pool were confirmed by the presence of 4,6-linked GlcNAc. The high abundance of 2-linked Man and the very low abundance of 2,4- and 2,6-linked Man indicate that the majority of complex structures are bi-antennary, and that despite the extremely large size of the N-glycans detected, tri- and tetra-antennary complex N-glycans are minor species. The presence of high levels of 3-linked Gal (which can also be produced from other structural features; see below) and 4-linked GlcNAc supports LacNAc extensions, whereas the presence of 3,6-linked Gal confirms that LacNAc extensions can be branched. The presence of both 3- and 6-linked Gal indicates the presence of both α2–3- and α2–6-sialylated glycans, although the latter is of minor abundance. Finally, terminal Gal and 3-linked Gal are consistent with Galα1–3Gal capping.

**Glycomic Profiling of Mouse Lung O-glycans**—The MALDI-TOF glycomic profiling of permethylated mouse lung O-glycans produced a series of molecular ion signals consistent with siaylated core 1 and core 2 structures (supplemental Fig. S2).

**Completeness of Permethylation**—One particular concern with the data acquisition strategy described above is the completeness of the permethylation derivatization at high molecular weight. Even a relatively small N-glycan such as Man9GlcNAc2, which as an [M+Na]+ permethylated molecular ion would have a mass-to-charge ratio of 1579.8, has 23 methylation sites, so that permethylation must be close to 100% complete, or else the molecular ion signal will be split over a number of m/z values, which will reduce the sensitivity of the analysis. The detection of high-molecular-weight glycans therefore requires both a high degree of permethylation completeness and the ability to discriminate between chemical or electronic noise and genuine glycan signals. Firm evidence that our permethylation protocols are suitable for high-mass glycomics is provided by the data in Fig 4A. These peak series show that permethylation is sufficiently complete, and that even in the 10–14 kDa range, the glycan signal will be concentrated at the mass corresponding to complete permethylation.

**Development of Match Fitting Algorithms**—We refer to a series of peaks representing different isotopic forms of the same molecule as peak envelopes. We took a matched filter approach to the detection of peak envelopes in the MALDI-TOF profiles. A matched filter is a finite-impulse-response filter with the same shape as the signal to be detected; it is the optimal linear filter for the detection of a transient signal in the...
presence of additive white Gaussian noise (here “linear” means that the response of the detector scales linearly with the amplitude of the signal; this is a reasonable assumption for signals of unknown amplitude) (24). Peak envelopes change shape with increasing mass, because peaks broaden with decreasing mass resolution, and higher isotope peaks, corresponding to increasing numbers of $^{13}$C and other heavy isotopes, grow in intensity relative to the monoisotopic peak. Thus matched filters must change shape as well. Fig. 4B shows theoretical isotope envelopes used in our peak detection algorithm. Most other peak detection methods proposed for mass spectra (see Ref. 25 for a survey) treat the full mass range uniformly and thus do not account for peak broadening and isotope peaks. An exception is a matched filter approach.
used for peptide mass fingerprinting (26). The following step-wise approach was undertaken:

1. Resampling of mass spectra. The spacing of \( m/z \) bins in a MALDI-TOF profile is nonuniform, with the spacing growing with the square root of \( m/z \) because of the quadratic dependence of TOF on \( m/z \). We resampled the mass spectrum to uniform \( m/z \) bins with spacing of 0.02 Da, which is approximately the finest spacing in a mass spectrum with an \( m/z \) range of 1–14 kDa. To resample, we proportionately interpolated using the original \( m/z \) bins on either side of the new \( m/z \) bin.

2. Matched filters. We precomputed a matched filter for each 164-Da mass range to appear in step 3. Matched filters were 1000 points (20 Da) long, normalized to Euclidean length one, and started at -1.0 Da from the monoisotopic mass to allow for the spread of the monoisotopic peak (Fig. 4B shows only the first 11 Da of the 20 Da filters). We computed the theoretical isotope envelope of a permethylated N-glycan via random sampling of natural isotope abundances of carbon, oxygen, hydrogen, and nitrogen. To set the number of atoms of each type, we used an “average-ose” method and assumed that N-glycans contained equal numbers of hexose and HexNAc.

3. Convolution. We used fast Fourier transforms on successive blocks of 8192 \( m/z \) bins (~164 Da) to implement convolution of the mass spectrum with a reversed filter, which is mathematically equivalent to taking the dot product of the original filter with the mass spectrum at each possible position. For simplicity, we did not window or overlap the blocks, and thus a peak series split between two blocks gives a weakened response. Fig. 4C shows the 13,000–13,100 Da range of the mass spectrum, and Fig. 4D shows the detector response. We subtracted the mean from each \( m/z \) block before convolution to correct for varying baseline levels.

**Fig. 4. MALDI-TOF profiles.** A, The peaks at 3012–3016 Da in A correspond to incompletely permethylated species of the glycan at 3026–3030 Da (HexNAc4, Hex3, Fuc1, NeuGc2). The signals at 2996–3000 are due to a second glycan species (HexNAc4, Hex3, Fuc1, NeuAc1, NeuGc1). The fact that the completely permethylated glycan accounts for ~95% of the total intensity at 3000 Da implies that permethylation will be about (0.95)^4 ~ 80% complete for 12,000 Da glycans. B, Matched filters used in the detection algorithm. C, 13,000–13,100 Da range. There is a barely discernible increase in intensity from 13,020 to 13,040 (inset), with no sign of 1-Da peak spacing. D, Matched filter detection. The detection algorithm locates a peak at monoisotopic mass 13,025.36, an almost perfect match to a singly fucoyslated glycan with 26 lactosamines at 13,025.55 Da.
across the spectrum; thus, the detector response is centered at zero.

4. Low-pass filtering. As seen in Fig. 4D, the detector response, shown in blue, contains side lobes and high-frequency noise, so we next smoothed the response using a finite-impulse-response low-pass filter and obtained the signal shown in red. We used a filter computed by MATLAB’s FIR1 function with nominal frequency cutoff of 0.02 Nyquist, so that with sampling at 50 bins per Dalton, jitters in the detector response smaller than about 5 Da were filtered out.

5. Peak detection. Finally, we let \((M, A)\) be the \(m/z\) and amplitude (\(x\)- and \(y\)-axis values as in Fig. 4D) of a local maximum in the low-pass-filtered detector response. We then determined the significance of the local maximum by comparing it to the original intensity values (\(y\)-axis in Fig. 4D) in the \(m/z\) range extending 40 Da per charge on either side of \(M\). We divided \(A\) by the 95% percentile value of \(M\) (that is, the least intensity greater than 95% of all intensities) of the intensities in \([M-40, M+40]\). If this “signal ratio” was larger than some threshold \(T\), which by default was set to 3, then the software annotated a peak series. The example shown in Figs. 4C and 4D gives a signal ratio of 4.19. The choice of 3.0 as the signal ratio threshold was guided by a control experiment in which two MALDI profiles of blank samples (that is, pure matrix with no glycans) were sent through the peak finding program. On the blank samples, no peaks were detected in the 10–20 kDa range with a signal ratio greater than 2.0, but there were a few peaks with ratios of 2.0 to 2.8 in the 20–24 kDa mass range. Table I shows the 17 detected peak series with the highest \(m/z\) ratios. Remarkably, 15 of the 17 detected peaks match masses corresponding to N-glycans of the form shown in Fig. 2, polylactosamine antennae, some with NeuAc termini, and a single fucose. Most of the mass matches are below 100 ppm; obviously this level of agreement is very unlikely to arise by chance from spectra containing pure noise or other families of glycans. Three of the matched N-glycans, the ones with detected \(m/z\) of 10,604.24, 13,474.00 and 13,925.98, were missed by the human expert. Conversely, six peak series annotated by the expert, with \(m/z\) 10,154.3, 11,140.7, 11,951.5, 12,039.1, 12,575.8, and 12,936.7, were missed by the automatic peak detector with a signal ratio threshold of 3.0. Five of these six peak series, however, were detected at lower thresholds—at 10,151.76 with a ratio of 2.16, 11,143.08 with a ratio of 2.46, 11,949.02 with a ratio of 1.12, 12,039.70 with a ratio of 2.85, and 12,576.20 Da with a ratio of 2.55.

**DISCUSSION**

We have shown through the application of a glycomics strategy based on the analysis of permethylated derivatives that the mouse lung expresses a diverse, complex series of N-glycans, and the methodology has allowed the elucidation of some remarkable structural features. Firstly, we draw attention to the large range of sizes of the N-glycans and the exceptionally high-molecular-weight glycans that have been observed, to our knowledge constituting some of the largest mammalian N-glycans ever reported. However, even more striking is that upon detailed examination of the data, structural features become apparent that implicate a highly ordered and regulated biosynthetic process. For example, as the N-glycans increase in size, there is a transition in the dominant form of sialylation from NeuGc to NeuAc. Also, the level of core fucosylation increases with increasing size. Most striking is the regulation of polylactosaminoglycan chains. MS/MS analysis clearly demonstrates that the predominant species on non-sialylated N-glycans are uniformly branched along the whole antenna. In contrast, the polylactosaminoglycan antennae of sialylated N-glycans are linear, irrespective of whether they are sialylated or not. Finally, the MS/MS and

---

**Table I**

<table>
<thead>
<tr>
<th>Detected (m/z)</th>
<th>Signal ratio</th>
<th>Matching composition</th>
<th>Monoisotopic (m/z)</th>
<th>Average isotope (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,242.18</td>
<td>5.26</td>
<td>Core, 19 Lac, 1 NeuAc, 1 Fuc</td>
<td>10,242.14</td>
<td>10,248.21</td>
</tr>
<tr>
<td>10,329.88</td>
<td>8.48</td>
<td>Core, 20 Lac, 1 Fuc</td>
<td>10,330.19</td>
<td>10,336.31</td>
</tr>
<tr>
<td>10,604.24</td>
<td>3.06</td>
<td>Core, 19 Lac, 2 NeuAc, 1 Fuc</td>
<td>10,603.31</td>
<td>10,609.60</td>
</tr>
<tr>
<td>10,690.96</td>
<td>6.89</td>
<td>Core, 20 Lac, 1 NeuAc, 1 Fuc</td>
<td>10,691.36</td>
<td>10,697.70</td>
</tr>
<tr>
<td>10,778.64</td>
<td>4.49</td>
<td>Core, 21 Lac, 1 Fuc</td>
<td>10,779.42</td>
<td>10,785.81</td>
</tr>
<tr>
<td>11,052.10</td>
<td>4.57</td>
<td>Core, 20 Lac, 2 NeuAc, 1 Fuc</td>
<td>11,052.54</td>
<td>11,059.09</td>
</tr>
<tr>
<td>11,228.12</td>
<td>7.02</td>
<td>Core, 22 Lac, 1 Fuc</td>
<td>11,228.64</td>
<td>11,235.30</td>
</tr>
<tr>
<td>11,589.92</td>
<td>5.56</td>
<td>Core, 22 Lac, 1 NeuAc, 1 Fuc</td>
<td>11,589.82</td>
<td>11,596.69</td>
</tr>
<tr>
<td>11,677.64</td>
<td>4.89</td>
<td>Core, 23 Lac, 1 Fuc</td>
<td>11,677.87</td>
<td>11,684.79</td>
</tr>
<tr>
<td>12,126.62</td>
<td>6.37</td>
<td>Core, 24 Lac, 1 Fuc</td>
<td>12,127.09</td>
<td>12,134.28</td>
</tr>
<tr>
<td>12,487.66</td>
<td>3.85</td>
<td>Core, 24 Lac, 1 NeuAc, 1 Fuc</td>
<td>12,488.27</td>
<td>12,495.67</td>
</tr>
<tr>
<td>13,025.36</td>
<td>4.19</td>
<td>Core, 26 Lac, 1 Fuc</td>
<td>13,025.55</td>
<td>13,033.27</td>
</tr>
<tr>
<td>13,386.38</td>
<td>3.46</td>
<td>Core, 26 Lac, 2 NeuAc, 1 Fuc</td>
<td>13,386.72</td>
<td>13,394.66</td>
</tr>
<tr>
<td>13,474.00</td>
<td>3.02</td>
<td>Core, 27 Lac, 1 Fuc</td>
<td>13,474.77</td>
<td>13,482.76</td>
</tr>
<tr>
<td>13,925.98</td>
<td>3.07</td>
<td>Core, 28 Lac, 1 Fuc</td>
<td>13,924.00</td>
<td>13,932.25</td>
</tr>
</tbody>
</table>

This table lists all the detected peaks in the 10–14 kDa range in Fig. 1 with signal ratios greater than 3.0. All but two of the detections correspond to masses of N-glycans of the form shown in Fig. 1: lactosamine antennae with NeuAc termini and a single fucose. The peaks at 10,125.76 and 11,786.62 Da could be either noise or N-glycans of other forms.
GC/MS linkage data indicate that despite the exceptionally large size of the N-glycans, bi-antennary structures are much more abundant than tri- and tetra-antennary complex N-glycans.

The assignment of the high-molecular-weight glycans was greatly assisted by the development and utilization of a matched filter approach to peak detection. This algorithm can find N-glycan peak envelopes in MALDI-TOF profiles with accuracy comparable to that of a human expert, without any prior knowledge of likely N-glycan masses or biosynthetic pathways. In particular, the human expert relied on prior knowledge that long chains of lactosamine are biosynthetically likely. The success of the matched filter algorithm is not altogether surprising, because N-glycan peak envelopes are highly predictable from the natural abundances of isotopes, affording an accurate model of the signal to be detected, so that on this particular pattern recognition problem a computer algorithm enjoys some advantage over human vision. N-glycans are especially well suited to this matched filter approach, because N-glycans are almost identical in their elemental composition, which for permethylated glycans is about 32.43% carbon, 2.7% oxygen, 63.52% hydrogen, and 1.35% nitrogen, along with one sodium, which has only one stable isotope. By comparison, peptides vary more widely in their elemental compositions, and methionine and cysteine contents can have a noticeable effect on isotopic ratios, because $^{34}\text{S}$ has a natural abundance over 4%. Finally, the detailed data analysis afforded by the matched filtering algorithms clearly demonstrates the efficiency of the permutation de- derivation reaction, which underpins the entire glycomic analytical strategy.

* M.B. and A.E.B. were supported in part by NIH Grant No. R01GM085718. A.D. and S.M.H. were supported by Analytical Glycotechnology Core (Core C) of the Consortium for Functional Glycomics Grant No. GM 62116 and by Biotechnology and Biological Sciences Research Council Grant Nos. BBF0083091 and B19088.

Text: This article contains supplemental material.

§ To whom correspondence should be addressed: Marshall Bern, Ph.D., Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304, Tel.: (650) 812-4443, Fax: (650) 812-4471, E-mail: bern@parc.com; Professor Anne Dell, Division of Molecular Biosciences, Imperial College London, London, SW7 2AZ, UK, Tel.: +44 (0)20 75945219, E-mail: a.dell@imperial.ac.uk; Dr. Stuart Haslam, Division of Molecular Biosciences, Imperial College London, London, SW7 2AZ, UK, Tel.: +44 20 75945222, Fax: +44 20 7225 0458, E-mail: s.haslam@imperial.ac.uk.

REFERENCES