

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221819195>

# Annotation of a Serum N-Glycan Library for Rapid Identification of Structures

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · MARCH 2012

Impact Factor: 4.25 · DOI: 10.1021/pr2011439 · Source: PubMed

CITATIONS

52

READS

43

5 AUTHORS, INCLUDING:



**Danielle L Aldredge**

University of California, Davis

6 PUBLICATIONS 192 CITATIONS

SEE PROFILE



**Hyun Joo An**

Chungnam National University

83 PUBLICATIONS 2,222 CITATIONS

SEE PROFILE



**Ning Tang**

Agilent Technologies

22 PUBLICATIONS 1,662 CITATIONS

SEE PROFILE

Published in final edited form as:

*J Proteome Res.* 2012 March 2; 11(3): 1958–1968. doi:10.1021/pr2011439.

## Annotation of a Serum N-glycan Library for Rapid Identification of Structures

Danielle Aldredge<sup>1</sup>, Hyun Joo An<sup>2</sup>, Ning Tang<sup>3</sup>, Keith Waddell<sup>3</sup>, and Carlito Lebrilla<sup>1,4,\*</sup>

<sup>1</sup>Department of Chemistry, University of California, Davis, CA 95616, USA

<sup>2</sup>Graduate School of Analytical Science and Technologies, Chungnam National University, Daejeon 305-764, South Korea

<sup>3</sup>Agilent Technologies, Inc., Santa Clara, CA 95051, USA

<sup>4</sup>Department of Biochemistry and Molecular Medicine, University of California, Davis, CA 95616, USA

### Abstract

Glycosylation is one of the most common post-translational modifications of proteins and has been shown to change with various pathological states including cancer. Global glycan profiling of human serum based on mass spectrometry has already led to several promising markers for diseases. The changes in glycan structure can result in altered monosaccharide composition as well as in the linkages between the monosaccharides. High-throughput glycan structural elucidation is not possible due to the lack of a glycan template to expedite identification. In an effort toward rapid profiling and identification of glycans, we have constructed a library of structures for the serum glycome to aid in the rapid identification of serum glycans. N-Glycans from human serum glycoproteins are used as a standard and compiled into a library with exact structure (composition and linkage), liquid chromatography retention time, and accurate mass. Development of the library relies on highly reproducible nanoLC/MS retention times. Tandem MS and exoglycosidase digestions were used for structural elucidation. The library currently contains over 300 entries with 50 structures completely elucidated and over 60 partially elucidated structures. This database is steadily growing and will be used to rapidly identify glycans in unknown biological samples.

### Keywords

N-linked glycan; Glycan structures; Isomers; nano HPLC-Chip TOF MS; on-line detection; retention time; database/library

## INTRODUCTION

Glycosylation is one of the most common post-translational modifications in proteins. It has been estimated that over 50% of human proteins are glycosylated<sup>1</sup>. The glycans serve important structural roles by participating in protein folding, enhancing protein solubility, attaching proteins to cell membranes, protecting proteins from degradation and mediating many important biological functions<sup>2–5</sup>. Changes in glycosylation have been linked to various pathological states including cancer, atherosclerosis, rheumatoid arthritis, Chron's Disease, and inflammation making them important targets for analysis<sup>2, 6–10</sup>. Alteration of

\*Address correspondence to Carlito B. Lebrilla, cblebrilla@ucdavis.edu, Tel.: 530-752-0504, Fax: 530-752-8995.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>.

the glycan structures can take on several forms including different monosaccharide compositions, changes in connectivity (sequence), and most interestingly, changes in the types of linkages between monosaccharides. The ability to monitor these changes quantitatively and specifically can make monitoring of glycosylation a valuable diagnostic tool by providing greater sensitivity and specificity. Indeed, glycomics, the characterization of the glycan constituents of a biological system, is rapidly emerging as a new paradigm for biomarker discovery<sup>8, 9, 11–13</sup>.

Despite the advancements in glycan analysis, structural elucidation remains a complicated and difficult task. The variation in size, polarity, and linkage configuration can yield a dizzying array of structures. Glycans can vary in size from disaccharide to complex carbohydrates; additionally, each monosaccharide can differ in linkages, potentially forming 10<sup>12</sup> distinct structures from as few as six monosaccharides<sup>8, 14, 15</sup>. Furthermore, there is glycan heterogeneity at the site of covalent attachment to the protein meaning a particular site of glycosylation may have a number of different glycans associated with it (site- or micro-heterogeneity). The result is a collection of 'glycoforms' associated with each glycoprotein further complicating the analysis<sup>13, 16</sup>. Uncovering the glycome has been described as a bigger challenge than the genome or proteome not only because of the many variations among glycans, but also because carbohydrate synthesis is not under direct genetic control resulting in no relationship between their synthesis and the organism's genome<sup>17, 18</sup>.

High throughput structural elucidation is therefore not likely with current technology. However, rapid structural identification is a possibility if the glycome is finite and not as large as previously estimated. Recent studies from this group developed a theoretical N-glycan library for serum<sup>19</sup>. Based on this library, it was estimated that there are about 330 distinct compositions in the serum N-glycome. A further study on isomer separation of N-glycans in serum show that there are, on average, about five isomers for each composition<sup>20</sup>. Given these results, it may be concluded that there could be as few as 1600 distinct structures in the serum N-glycome, and far smaller than the possible chemical combinations.

To this end, several databases and libraries<sup>21–24</sup> have been developed to encourage high-throughput glycan identification. Most of these, however, contain structures from previously published studies and are not suitable for specific glycan identification in unknown samples. Glycosuite database contains most of the published N-linked and O-linked glycans including native and recombinant sources, tissues, and cell lines<sup>21</sup>. Glycan structures are cross-referenced to Swiss-Prot/TrEMBL as well as PubMed when available, unfortunately retention time is not offered. The Consortium for Functional Glycomics provides a database with information containing glycan structures, glycan binding proteins, as well as glycosyl transferases. Their efforts provide an "integrated-systems approach to understanding the structure-function relationship of glycans"<sup>22</sup>. EUROCarbDB offers databases and bioinformatic tools enabling research groups to upload their data to share with others<sup>23</sup>. The most pertinent is GlycoBase by Rudd and co-workers, which boasts a fully automatable technology platform with more than 350 N-glycan structures including 117 present in the human serum glycome. Here, glycans are 2-aminobenzamide labeled separated by HILIC and annotated/measured with glucose units<sup>24, 25</sup> to identify structures.

In this study, we have constructed a serum N-glycome library that allows matching of nanoLC retention times and accurate masses to rapidly identify exact glycan structures. The library facilitates rapid structure identification from unknown samples providing comprehensive analysis of underivatized N-linked human serum glycans. There are currently no databases that focus on fully annotating the human serum N-glycome with retention times and on-line mass spectrometry detection. We chose a porous graphitized

carbon (PGC) stationary phase as it provides superior isomer separation and unique retention times for individual isomers<sup>26, 27</sup>. On-line mass spectrometry detection allows for greater confidence in our glycan assignments. A two-part approach was used to build up the library, consisting of identifying known structures and *de novo* sequencing for unknown compounds. The use of serum glycoprotein standards provided some previously characterized structures. Structures were determined and confirmed through sequential exoglycosidase digestions. Also, knowledge of the N-glycan synthesis and biological pathway allowed for the refinement of some structural assignments. Our library is composed of native glycans found in human serum separated by porous graphitized carbon (PGC) whose structural assignments are supported with MS/MS analysis and exoglycosidase digestions.

## MATERIALS and METHODS

### Materials and Chemicals

Glycoprotein standards IgG, IgM, Transferrin, alpha-1-acid glycoprotein, alpha-2-macroglobulin, and human serum were obtained from Sigma-Aldrich (St. Louis, MO); alpha-1-antitrypsin, and IgA were obtained from EMD Chemicals (Gibbstown, NJ); alpha-1-antichymotrypsin was obtained from Athens Research and Technology (Athens, GA); and complement component 3 was obtained from Genway Biotech, Inc. (San Diego, CA). Exoglycosidase alpha 2–3 neuraminidase, alpha 1–6 mannosidase, alpha 1–2,3 mannosidase, and beta-N-Acetyl glucosaminidase were obtained from New England Biolabs (Ipswich, MA). All other reagents were of analytical or HPLC grade.

### Reduction and Purification

N-linked glycans were released from protein standards using PNGase F obtained from New England Biolabs (Ipswich, MA), with glycoprotein starting material ranging from 100–500µg. The released N-glycans were reduced using sodium borohydride from Sigma-Aldrich (St. Louis, MO) according to the protocol previously reported by our laboratory<sup>20</sup>. N-glycans were cleaned up via a two-step solid phase extraction (SPE) procedure using C8 cartridges from Suppelco (Bellefonte, PA) and graphitized carbon cartridges from Alltech Associated (Deerfield, IL). C8 cartridges were conditioned with 6mL acetonitrile (ACN) and 6mL nanopure water. The glycan digest was loaded onto the cartridge and the flow-through was collected and loaded on pre-conditioned graphitized carbon cartridges then washed with 6 mL of nanopure water. N-Glycans were eluted in 20% ACN in water (v/v) and 0.05% Trifluoroacetic acid (TFA) in 40% ACN in water (v/v), dried down, then reconstituted in 20 L of nanopure water.

### MALDI FT-ICR MS and HPLC-Chip/TOF MS Analysis

Reduced and purified glycans were first analyzed using an IonSpec HiRes MALDI FT-ICR MS (IonSpec, Lake Forest, CA) equipped with a 355nm pulsed Nd:YAG laser and 7.0 Tesla Magnet. The sample was spotted on a stainless steel MALDI plate with equal volume of DHB matrix made up of 0.05mg/mL DHB in 50% ACN in nanopure water as previously described<sup>28</sup>. Glycans were identified within a 5 ppm accurate mass criterion. The N-glycan samples were analyzed in positive and negative ion modes. Tandem MS via collision-induced dissociation (CID) was then performed for selected peaks.

NanoLC MS analyses were performed using the Agilent 6200 Series HPLC-Chip/TOF MS equipped with a microwell-plate auto sampler, capillary sample loading pump, nano pump, HPLC-Chip Cube interface, and the Agilent 6210 TOF LC/MS. Oligosaccharides were separated using the Agilent HPLC-Chip II comprised of a 40nL enrichment column and a 75 µm × 43mm separation column both packed with graphitized carbon stationary phase. The

system consists of a capillary and nanoflow pump both operating with a binary solvent system consisting of A, 0.1% formic acid in 3% ACN in water and B, 0.1% formic acid in 90% ACN in water. Samples were loaded onto the HPLC-Chip at concentrations ranging 0.6–25 g with the capillary pump delivering 0.1% formic acid in 3% ACN in water at 4 $\mu$ L/min with a 2 $\mu$ L sample injection. The nano pump delivered the 45-minute gradient used to separate oligosaccharides with a flow rate of 0.3 $\mu$ L/min. The samples were eluted with the following gradient: 0%B, 0–2.5 minutes; 0 to 16%B, 2.5–20 minutes; 16 to 44%B, 20–30 minutes; 44 to 100%B, 30–35 minutes; and 100%B, 35–45 minutes. This procedure was followed by equilibration at 0%B, 45.01–65 minutes. Data was acquired in positive ion mode and calibrated with internal calibrant ions covering a wide mass range (ESI-TOF Tuning Mix, Agilent Technologies, Santa Clara, CA). Glycans were identified within a 20 ppm accurate mass criteria using the Molecular Feature Extractor algorithm included in the Mass Hunter Qualitative Analysis Software (Version B.03.01, Agilent Technologies). The software identified individual compounds by considering expected charge states based on a predicted isotopic distribution. The deconvoluted glycan mass, retention time (at the maximum compound height), and abundance (in summed ion counts) were extracted.

### Separation of N-Glycans with off-line HPLC

N-Glycans were separated off-line with an Agilent Hewlett-Packard Series 1100 HPLC system using a Hypercarb PGC column (Thermoquest, Hypersil Division) with the dimensions 100mm  $\times$  2.1 mm. The oligosaccharides were eluted with a solvent system consisting of nanopure water, A, and ACN, B, with a flow rate of 0.250 mL/min. 80 fractions were collected in intervals of 0.5 minutes over the 60-minute gradient, with the last 10 minutes reserved for column reconditioning. The collected fractions were dried and reconstituted in 10 $\mu$ L of nanopure water then analyzed by MALDI FT-ICR.

### Exoglycosidase Digestion

To determine the linkage between monosaccharides, exoglycosidase digestions were performed following an established protocol outlined by Zhang, et al<sup>29</sup>. The buffer was prepared by adding glacial acetic acid to a 0.1M ammonium acetate solution until the pH reached 6.5. Specific enzymes require different pH values and care was taken to ensure that the proper pH values were obtained. Exoglycosidase digestion reactions were optimized according to the enzyme activity and amounts of glycan used. When developing the method, enzyme reactions were tested over various time points to confirm the linkage specificity of the enzyme and to ensure complete digestion. The digestion was carried out in a 37 C water bath with 3 $\mu$ L of buffer, 1 $\mu$ L of sample and 0.5 $\mu$ L of enzyme. Digestion times vary depending on the enzyme as outlined by previous studies<sup>29</sup>. After each digestion the sample was cleaned using Millipore ZipTip C18 pipette tips (Millipore Corporation, Billerica, MA) to remove the enzyme. Each was then dried down and reconstituted in nanopure water for analysis. For sequential exoglycosidase digestions, after the first digestion, the reaction was checked using MALDI FT-ICR and the next enzyme was added to the solution and the digestion was continued.

## RESULTS AND DISCUSSION

The glycan library was constructed from the native N-glycans of the most abundant glycoproteins in human serum. The proteins were chosen because they possess predominantly N-glycans that provide good coverage of all N-glycan types including high mannose, complex, and hybrid. Furthermore, many of these proteins have been extensively studied in literature with several glycan structures already elucidated<sup>30–32</sup>. The known glycan structures were identified and characterized and provided the initial core structures of the library. Glycans were first profiled by MALDI MS and then separated by nanoLC MS.

MALDI MS provided a rapid profile of the glycan compositions and their relative glycan abundances for the specific proteins. NanoLC MS was then performed on the same samples to separate the components into individual compounds as well as provide retention times and mass spectra.

### Release of glycans from glycoprotein IgM with compositional profile by MALDI MS

The analysis of glycans from a single protein standard is illustrated with IgM, where the N-Glycans were released using the protocol described in the Experimental Section. MALDI FT-ICR MS was used to obtain a glycan composition profile where sodiated ions were produced. This MALDI MS spectrum provided confirmation that the release was successful and yielded the types of glycans (complex, hybrid, high mannose) associated with each glycoprotein. In addition, the MALDI MS profile provides semi-quantitative information regarding the relative abundances of the released glycans, which proved valuable for subsequent analyses. Figure 1A shows the positive mode MALDI FT-ICR MS profile of the released glycans from the IgM glycoprotein in the 20% ACN SPE fractions. In the spectrum all three types of N-glycans are observed supporting previous analyses of this glycoprotein<sup>30</sup>. Tandem MS via CID was performed to confirm the composition of the glycans as well as provide information pertaining to the connectivity of the monosaccharides (sequence) allowing for greater confidence in assignments. An example fragmentation profile via CID for the reduced fucosylated neutral oligosaccharide at  $m/z$  1811.6  $[M+Na]^+$  with the composition 5Hexose (Hex), 4N-acetyl hexosamine (HexNAc), 1Fucose (Fuc) is shown in Supplementary Figure 1. The fragmentation profile is consistent with the proposed composition of the glycan based on mass accuracy.

Negative ion mode glycan mass profiles were also obtained by MALDI FT-ICR MS as shown in Figure 1B, which demonstrates an increased detection of anionic oligosaccharides. The negative ion mode is used to detect sialylated glycans from the 40% ACN fraction as sialic acid residues are readily deprotonated, and are not suppressed, as they would be in the positive mode by neutral species<sup>33, 34</sup>. The abundant peaks are labeled with their putative structures based on compositional assignments employing accurate masses and supported by tandem MS.

### NanoLC MS characterization of glycan isomers in IgM

MALDI MS of a mixture of glycans provides a rapid composition profile but provides little to no information regarding the isomers. For nanoLC/MS analysis, the MALDI samples were diluted by nearly five-fold and analyzed via nanoHPLC-Chip/nanoESI/TOF MS (Chip-TOF). Glycans were separated with a PGC stationary phase incorporated into an HPLC-microchip, as described previously<sup>20</sup>. The chip is found to have high sensitivity, reproducibility, and effective isomer separation<sup>35–38</sup>. A further consequence of the separation is minimized suppression effects during ionization resulting in new potential compositions not observed with MALDI MS<sup>39</sup>. The effects of increased sensitivity and decreased suppression effects are observed when comparing the analyses of N-glycans from IgM between MALDI FT-ICR MS and nanoLC/TOF MS. Figure 2 shows the base peak chromatogram of the reduced N-glycans from IgM with the most abundant peaks labeled with their putative structures based on accurate mass. Approximately 30 distinct glycan compositions were identified in the MALDI MS profile from the 20% and the 40% ACN fraction for IgM. For the Chip/TOF MS analysis, the two fractions were combined and analyzed to yield 47 distinct glycan compositions. The additional 17 compositions identified are a result of minimized suppression effects resulting from the nanoLC separation as well as the increased sensitivity of the HPLC-Chip/nanoESI/TOF MS instrument over the MALDI MS analysis. From the nanoLC analysis, 120 compounds are observed due to the



detection of isomers associated with each glycan composition. The additional compounds detected include hybrid, fucosylated, and sialylated type glycans.

In cases where structures have many isomers, the structural assignment becomes increasingly complicated. Assigning glycan structures is further complicated by the separation of anomers. The reducing end can have the anomeric oxygen in alpha or beta form. Therefore, each compound bears two anomers<sup>40, 41</sup>. Because separation on PGC is so effective, anomers are separated, potentially doubling the number of peaks. A comprehensive structural assignment is significantly complicated by the presence of anomeric peaks. To eliminate this ambiguity, the reducing ends were reduced using sodium borohydride, which converts the aldehyde to an alditol. Taking the simple example of the composition 4Hex, 4HexNAc, and 1Fuc, having two isomers, Supplementary Figure 2A yields an EIC (extracted ion chromatogram) with four peaks, or two sets of doublets. Reduction of the mixture yields two overlapping peaks as shown in Supplementary Figure 2B. Interestingly, the anomers often resolve more easily than the positional and linkage isomers.

Overlaying nano LC/MS data from all the samples results in a global glycan pool. Shown in Figure 3 is the BPC (base peak chromatogram) overlay of seven chromatograms representing seven different glycoproteins. For simplicity, the BPC from each glycoprotein is represented by its own color with the abundant compositions and putative structures labeled. Many peaks have multiple glycans eluting at the same time. Also, note there are smaller peaks that exist between the major peaks throughout the chromatogram representing the lower abundant species. The TOF MS instruments inherently boast 5 orders in-spectrum dynamic range allowing for the detection of low abundant compounds in the presence of major components. This provides another advantage during the nanoLC MS analysis allowing the detection of the most abundant glycans in the presence of the minor components.

Because Figure 3 only labels the most abundant compounds, Supplementary Figures 3A and B zoom in and highlight the lower abundant species. We find that high mannose glycans elute earlier, which is highlighted in Supplementary Figure 3A, followed by complex bi-antennary and hybrid glycans. Finally, acidic glycans and complex type glycans with three or more antennae elute. Supplementary Figure 3B highlights the elution of the larger and anionic glycans, where we observe tri- and tetra-sialylated species. These elution trends are consistent with previous studies in our laboratory<sup>20, 26</sup> and in other laboratories<sup>42</sup>.

Chromatograms can be overlaid to rapidly determine identical structures from different proteins. Due to the high retention time reproducibility as shown previously<sup>26, 27</sup>, compounds that are highly overlapped and have the same accurate mass are the same structures. Retention times under these rigorous conditions can be reproducible to within 2.0 seconds<sup>27</sup>. Given this criteria and a similar one for accurate masses, we can determine the structures that are common to different proteins. We find several isomers that are common to the majority of the glycoproteins, and others that are unique to only one. For example, one of the most abundant glycans for all the proteins corresponded to  $m/z$  822.31  $[M+2H]^+$ , 5Hex and 4HexNAc. This glycan composition was found in several isomers with the most abundant isomer corresponding to a retention time of 19.4 minutes  $\pm$  3.0 seconds. The structure was determined to be the well-known bi-antennary complex type glycan structure and corresponds to entry 48 in Supplementary Table 1<sup>31</sup>.

In any LC/MS analysis, unwanted fragmentation can cause problems by complicating analysis. This problem is more common with oligosaccharides, which fragment more readily than, for example, peptides. While fragmentation has been minimized to less than 5%, but

more commonly less than 1%, it may still present a problem when identifying the minor components. To ensure that the proper molecular ions are assigned, EICs of compounds are inspected for larger co-eluting compounds that may be fragmenting to yield the smaller species. A simple test is to examine the mass spectrum at a given time point for any smaller related species that are co-eluting. If these two compounds have identical retention times there is a possibility that the smaller one could be a fragment. This examination helps eliminate false structural assignments due to any in-source fragmentation.

### Structure Elucidation with enzymatic reactions

For structure analyses, there are two types of glycans: those with previously characterized structures and those with unknown structures. Glycans with known structures require the identification of a few residues with their linkages to confirm. Unknown structures require *de novo* characterization of entire structures, although, the entire exercise is facilitated by the conserved structure in the core region. Therefore, even *de novo* analyses require only a limited number of glycosidase reactions. Furthermore, each product glycan can be tested to determine whether it matches in retention time and accurate mass to a previously identified structure, so that structure elucidation simply requires only one or two exoglycosidases. Additionally, once a structure is characterized, its unique retention time and mass can be used to identify the same structure from other protein sources.

The first step in either process is to isolate the compound of interest through the use of an off-line HPLC system with a porous graphitized column and a fraction collector. Although complete isolation is desirable, it is not necessary if the structures differ in compositions. Because the PGC stationary phase is better at separating isomers than oligomers, compounds that differ in size by one or two residues may prove more difficult to separate than isomers with different linkages. In this case, the mass spectrometry analysis is ideal, as it complements the isomer-specific data obtained through the chromatographic separation. HPLC fractions are first analyzed by MALDI MS to determine the different compositions present in the fraction. This mixture is then analyzed by Chip TOF MS to determine the number of isomers present in the fraction. The mixture is then reacted with specific exoglycosidases monitored by MALDI MS and Chip TOF MS. Similar methods have been published with greater detail from this laboratory<sup>28, 43</sup>.

The annotation of a glycan structure using these techniques is shown in the example of  $m/z$  814.29  $[M+2H]^+$ , which corresponds to 4Hex, 4HexNAc, and 1Fuc. Based on the nanoLC MS, this composition consists of two major isomers, which are abundant in IgG and shown in Figure 4A. The putative structure of this composition is bi-antennary with a terminal galactose and a terminal N-Acetylglucosamine. The uncertainty is in the linkage of the galactose and its position on the terminus *i.e.* the 1–3 or 1–6 antennae. The schematic, inset in Figure 4A, summarizes the exoglycosidase digestion strategy needed to elucidate the structure. The mixture was separated further by HPLC to isolate one compound. Figure 4B shows an EIC of  $m/z$  814.3  $[M+2H]^+$ , obtained from HPLC separation (fraction 23) and analyzed by Chip TOF MS (retention time 21.6 minutes).

The compound was first treated with a general  $\beta$ -N-acetylglucosaminidase followed by an  $\alpha$  1–3 mannosidase. Figure 4C shows an EIC of  $m/z$  631.7, which correlates to the loss of a HexNAc and Hex, confirming the respective products from these enzymes. The results indicate that the galactose is attached to the  $\alpha$  1–6 antenna and not the  $\alpha$  1–3 antenna. The same compound was also treated with an  $\alpha$  1–6 mannosidase (data not shown). The result showed no mass shift or reaction confirming the results. The corresponding mass spectra are shown. The most abundant ions are those for the doubly charged except for the product in Figure 4C where the singly charged is also abundant. The same experiments were performed



on the other isomer on Figure 4A and it was found to be the isomer with the galactose on the 1–3 antenna.

For additional verification the distinct tandem MS for these two positional isomers is shown in Figure 4D. Although the isomer-specific tandem MS profiles appear to be very similar there are significant differences between them. In particular the earlier eluting isomer has a major ion at 203.1 Da representing a free HexNAc monosaccharide. In the spectrum of the later eluting isomer this peak is hardly expressed. Additionally, each isomer has its own preferred fragmentation pathway with different intensities for their monosaccharide losses. For example, the earlier eluting isomer has 1261.4 with higher intensity where the later eluting isomer has 1257.4. The differences in their fragmentation profiles confirm these two compounds as distinct isomers.

For some of the structures, the enzymes can be applied directly to the glycan mixture to obtain structural information. Figures 5A shows the BPC chromatogram of the glycans released from the protein transferrin. The mixture was treated with  $\alpha 2$ –3 sialidase, and the resulting base peak chromatogram in Figure 5B, shows peaks that decrease after the reaction. The extracted ion chromatogram shows more clearly the affect of the glycosidase. Figure 5C and D, are the EICs for  $m/z$  1113.40  $[M+2H]^{+2}$ , before and after  $\alpha 2$ –3 sialidase treatment, respectively. This mass corresponds to the composition 5Hex, 4HexNAc, and 2NeuAc. Figure 5C shows two isomers for the di-sialylated bi-antennary species. Peak labeled **I** in Figure 5C corresponds to the more abundant species and **II** is the less abundant species. After digestion with the  $\alpha 2$ –3 sialidase, peak **II** disappears while peak **I** remains untouched. The results indicate that both sialylated residues in **I** are linked  $\alpha 2$ –6 while both sialic acid groups in **II** are  $\alpha 2$ –3 linked. We are able to conclude this because there is no change in the mono-sialylated EIC after the sialidase digestion. If one sialic acid were left attached we would expect an increase in the mono-sialylated EIC. The reaction mixture also includes the mono-sialylated homolog. Figure 5E and F, show the EIC for  $m/z$  967.85  $[M+2H]^{+2}$ , before and after  $\alpha 2$ –3 sialidase treatment, respectively. This composition corresponds to 5Hex, 4HexNAc, and 1NeuAc. The EIC shows two isomers that are not fully chromatographically resolved and were identified as bi-antennary mono-sialylated glycans. After treatment with the  $\alpha 2$ –3 sialidase, there is no change in the abundance of either. This result leads to the conclusion that both isomers have an  $\alpha 2$ –6 linked sialic acid. Reaction of the mixture with a non-specific sialidase diminishes these two peaks (data not shown).

### Application of Library to Rapidly Identify Structures in Unknown Samples

The library is a continual work in progress with an established group of abundant structures that can be identified rapidly. The glycans from the selected group of proteins number in excess of 350 structures. The 200 most abundant are listed in Supplementary Table 1. There are approximately 55 complete structures and over 60 nearly complete with only one or two linkages missing.

To identify a structure in an unknown mixture, the accurate mass and retention times for that compound are used. Tandem MS is also available but would require significant programming, which is outside the scope of this manuscript. Future publications will address identification of glycans using isomer-specific tandem MS.

The utility of this library is illustrated with N-glycans released from an unknown serum. Figure 6A shows the EIC of  $m/z$  741.28 ( $z=2$ ), corresponding to 4 Hex, 4 HexNAc, from human serum with seven major isomers matching this composition. Three isomers are highlighted in Figure 6A (**III**, **IV**, and **V**) that can be identified based on accurate mass and retention time from the structure library. Figure 6B–E shows the EICs for  $m/z$  741.28  $[M+2H]^{+2}$ , from different glycoprotein standards, IgG, IgA, Complement C3, and anti-

chymotrypsin, respectively. Based on the accurate mass and the chromatogram retention time, it is observed that **V** (RT 18.5 min) is found as the most abundant isomer in IgA, but it is not found in any of the other proteins. This structure was determined to be a hybrid type and was validated with tandem MS. The other two compounds (**III** and **IV**- RT 20.6 minutes and 21.3 minutes, respectively) are abundant in serum and found in several of the proteins including IgG, Complement C3, and anti-chymotrypsin eluting at 20.6 and 21.3 minutes. These structures were determined via exoglycosidase digestions from previous experiments and can be matched in this unknown sample via their unique retention time and accurate mass. The main isomer from IgA elutes at 18.5 minutes and is not observed in any of the other glycoproteins analyzed. This figure demonstrates the excellent retention time reproducibility across different samples and illustrates how specific glycan structures can be identified over several samples. This example clearly demonstrates the goal of this library, where we use previously characterized glycan structures to accurately identify isomers. This same treatment can be applied on a larger scale to rapidly identify glycan structures based on their accurate mass and unique retention time.

## CONCLUSION

An N-glycan library was developed based on the most abundant proteins in serum. Despite the small number of proteins, a large number of oligosaccharides' structures were obtained. Furthermore, despite the large number of structures there were large structural correlations between the glycans of different proteins. We propose that an extensive library can be readily developed for any biological systems based on a small number of proteins. The library includes glycan structures, accurate masses, unique retention times, and protein sources. The most abundant glycan structures were readily identified based on knowledge of the conserved biological processes and confirmed with tandem MS and exoglycosidase digestions. The lower abundant structures are more difficult to fully elucidate without enrichment. In these cases partial structures are offered in the database. With this library, we are able to quickly identify glycan structures based on their unique retention time and accurate mass. The limitations of full structural elucidation are a function of the difficulty in fully separating and isolating isomers for sequential exoglycosidase mapping and the limited commercial availability of specific exoglycosidases.

Although the nanoLC has been shown to be exceptionally reproducible, to account for any major retention time shifts, a pool of glycan standards can be used to calibrate the system. Because the elution order of isomers will not change, it is possible to simply realign retention times according to their shift. This will enable users to adjust library retention times making this library widely applicable for those with similar LC/MS systems.

This library provides a powerful template for rapid N-glycan structure identification in unknown samples. Our laboratory is employing the library to rapidly identify the most abundant glycan structures in biological samples, including serum, milk, and saliva samples. Any new structures identified are included in the growing database. The library's applicability extends to several fields involving rapid glycan analysis and identification.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

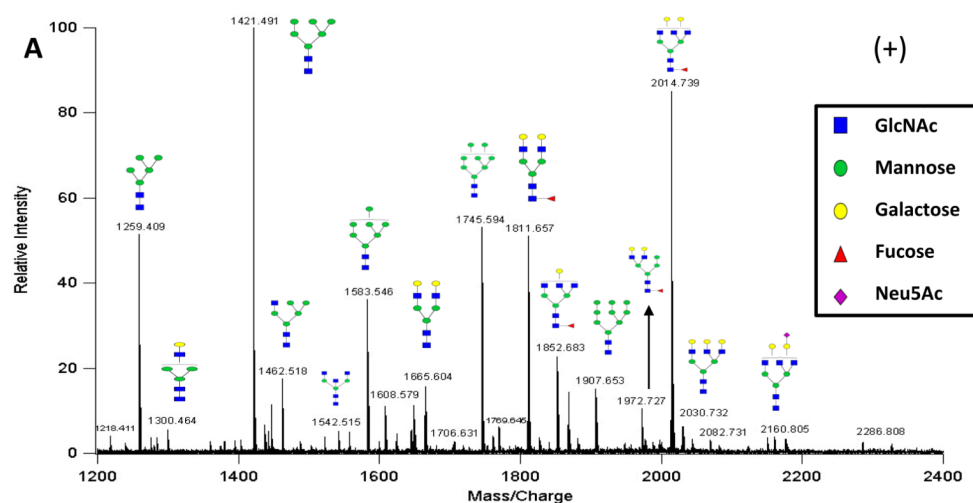
Funding provided by National Institute of Health (R01GM049077 and R01HD059127) is gratefully acknowledged.

## REFERENCES

1. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 1999; 1473(1):4–8.
2. Morelle W, Canis K, Chirat F, Faïd V, Michalski JC. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics*. 2006; 6(14):3993–4015. [PubMed: 16786490]
3. Dwek RA. Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev*. 1996; 96(2): 683–720. [PubMed: 11848770]
4. Varki, A.; Cummings, R.; Esko, J.; Freeze, H., et al. *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor; New York: 1999.
5. Ratner DM, Adams EW, Disney MD, Seeberger PH. Tools for Glycomics: Mapping Interactions of Carbohydrates in Biological Systems. *ChemBioChem*. 2004; 5(10):1375–1383. [PubMed: 15457538]
6. Alavi A, Axford JS. Sweet and sour: the impact of sugars on disease. *Rheumatology*. 2008; 47(6): 760–770. [PubMed: 18375404]
7. Varki A. Sialic acids in human health and disease. *Trends in Molecular Medicine*. 2008; 14(8):351–360. [PubMed: 18606570]
8. Callesen AK, Madsen JS, Vach W, Kruse TA, Mogensen O, Jensen ON. Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool? *Proteomics*. 2009; 9(6): 1428–1441. [PubMed: 19235169]
9. Abbott KL, Aoki K, Lim J-M, Porterfield M, Johnson R, O'Regan RM, Wells L, Tiemeyer M, Pierce M. Targeted Glycoproteomic Identification of Biomarkers for Human Breast Carcinoma. *Journal of Proteome Research*. 2008; 7(4):1470–1480. [PubMed: 18271524]
10. Zhang XL. Roles of Glycans and Glycopeptides in Immune System and Immune-Related Diseases. *Current Medicinal Chemistry*. 2006; 13:1141–1147. [PubMed: 16719775]
11. Srivastava S. Move Over Proteomics, Here Comes Glycomics. *Journal of Proteome Research*. 2008; 7(5):1799–1799. [PubMed: 18509903]
12. Packer NH, von der Lieth C-W, Aoki-Kinoshita KF, Lebrilla CB, Paulson JC, Raman R, Rudd P, Sasisekharan R, Taniguchi N, York WS. Frontiers in glycomics: Bioinformatics and biomarkers in disease. An NIH White Paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006). *Proteomics*. 2008; 8(1):8–20. [PubMed: 18095367]
13. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology*. 1999; 9(8):747–755. [PubMed: 10406840]
14. Willy Morelle J-CM. The Mass Spectrometric Analysis of Glycoproteins and their Glycan Structures. *Current Analytical Chemistry*. 2005; 1:29–57.
15. Laine RA. Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields  $1.05 \sqrt{6}$  1012 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*. 1994; 4(6):759–767. [PubMed: 7734838]
16. Arnold JN, Saldova R, Hamid UMA, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics*. 2008; 8(16):3284–3293. [PubMed: 18646009]
17. Brooks, SA.; Dwek, MV.; Schumacher, U. *Functional and Molecular Glycobiology*. BIOS Scientific Publishers Limited; Oxford, UK: 2002.
18. Lis H, Sharon N. Protein glycosylation. *European Journal of Biochemistry*. 1993; 218(1):1–27. [PubMed: 8243456]
19. Kronewitter SR, An HJ, de Leoz ML, Lebrilla CB, Miyamoto S, Leiserowitz GS. The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. *Proteomics*. 2009; 9(11):2986–2994. [PubMed: 19452454]
20. Chu CS, Ninonuevo MR, Clowers BH, Perkins PD, An HJ, Yin H, Killeen K, Miyamoto S, Grimm R, Lebrilla CB. Profile of native N-linked glycan structures from human serum using high

- performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. *Proteomics*. 2009; 9(7):1939–51. [PubMed: 19288519]
21. Joshi HJ, Harrison MJ, Schulz BL, Cooper CA, Packer NH, Karlsson NG. Development of a mass fingerprinting tool for automated interpretation of oligosaccharide fragmentation data. *Proteomics*. 2004; 4(6):1650–64. [PubMed: 15174134]
  22. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Meth*. 2005; 2(11):817–824.
  23. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008; 7(4):1650–9. [PubMed: 18311910]
  24. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*. 2008; 24(9):1214–6. [PubMed: 18344517]
  25. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem*. 2008; 376(1):1–12. [PubMed: 18194658]
  26. Hua S, An HJ, Ozcan S, Ro GS, Soares S, DeVere-White R, Lebrilla CB. Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst*. 2011; 136(18):3663–3671. [PubMed: 21776491]
  27. Hua S, Lebrilla C, An HJ. Application of nano-LC-based glycomics towards biomarker discovery. *Bioanalysis*. 2011; 3(22):2573–2585. [PubMed: 22122604]
  28. Wu S, Tao N, German JB, Grimm R, Lebrilla CB. Development of an annotated library of neutral human milk oligosaccharides. *J Proteome Res*. 2010; 9(8):4138–51. [PubMed: 20578730]
  29. Zhang J, Lindsay LL, Hedrick JL, Lebrilla CB. Strategy for Profiling and Structure Elucidation of Mucin-Type Oligosaccharides by Mass Spectrometry. *Analytical Chemistry*. 2004; 76(20):5990–6001. [PubMed: 15481946]
  30. Arnold JN, Wormald MR, Suter DM, Radcliffe CM, Harvey DJ, Dwek RA, Rudd PM, Sim RB. Human Serum IgM Glycosylation. *Journal of Biological Chemistry*. 2005; 280(32):29080–29087. [PubMed: 15955802]
  31. Field MC, Amatayakul-Chantler S, Rademacher TW, Rudd PM, Dwek RA. Structural analysis of the N-glycans from human immunoglobulin A1: comparison of normal human serum immunoglobulin A1 with that isolated from patients with rheumatoid arthritis. *Biochem J*. 1994; 299(Pt 1):261–75. [PubMed: 8166649]
  32. Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics*. 2008; 8(14):2858–2871. [PubMed: 18655055]
  33. Karlsson NG, Wilson NL, Wirth H-J, Dawes P, Joshi H, Packer NH. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid Communications in Mass Spectrometry*. 2004; 18(19):2282–2292. [PubMed: 15384149]
  34. Thomsson KA, Bäckström M, Holmér Larsson JM, Hansson GC, Karlsson H. Enhanced Detection of Sialylated and Sulfated Glycans with Negative Ion Mode Nanoliquid Chromatography/Mass Spectrometry at High pH. *Analytical Chemistry*. 2010; 82(4):1470–1477. [PubMed: 20092260]
  35. Ninonuevo MR, Park Y, Yin H, Zhang J, Ward RE, Clowers BH, German JB, Freeman SL, Killeen K, Grimm R, Lebrilla CB. A Strategy for Annotating the Human Milk Glycome. *Journal of Agricultural and Food Chemistry*. 2006; 54(20):7471–7480. [PubMed: 17002410]
  36. Knox, JH., et al. Carbon-based Packing Materials for Liquid Chromatography-Structure, Performance, and Retention Mechanisms. Vol. Vol. 37. Marcel Dekker; New York: 1997.
  37. Ross, P., et al. Carbon-based Packing Materials for Liquid Chromatography: Applications. Vol. Vol. 37. Marcel Dekker; New York: 1997.
  38. Davies M, Smith KD, Harbin A-M, Hounsell EF. High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. *Journal of Chromatography A*. 1992; 609(1–2):125–131.

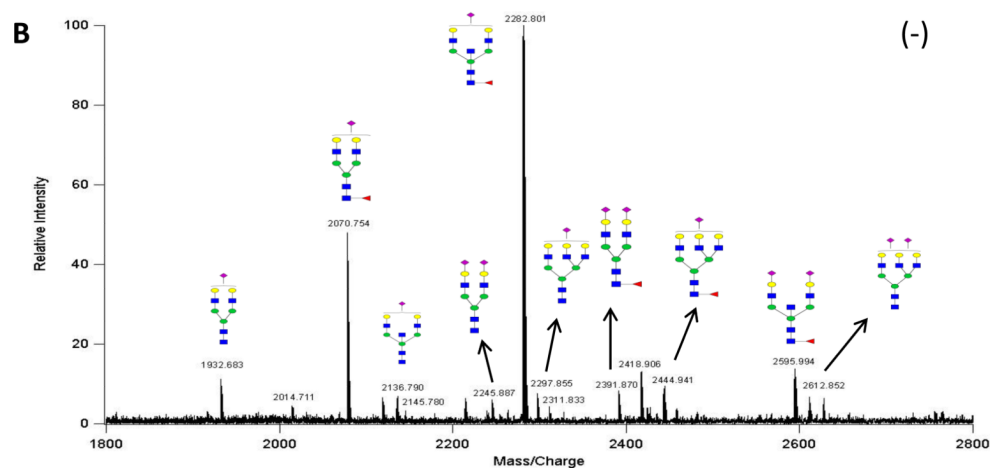
39. Annesley TM. Ion Suppression in Mass Spectrometry. *Clin Chem.* 2003; 49(7):1041–1044. [PubMed: 12816898]
40. Koizumi K. High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *Journal of Chromatography A.* 1996; 720(1–2):119–126. [PubMed: 8601188]
41. Koizumi K, Okada Y, Fukuda M. High-performance liquid chromatography of mono- and oligo-saccharides on a graphitized carbon column. *Carbohydrate Research.* 1991; 215(1):67–80.
42. Itoh S, Kawasaki N, Ohta M, Hyuga M, Hyuga S, Hayakawa T. Simultaneous microanalysis of N-linked oligosaccharides in a glycoprotein using microbore graphitized carbon column liquid chromatography-mass spectrometry. *J Chromatogr A.* 2002; 968(1–2):89–100. [PubMed: 12236519]
43. Li B, Russell SC, Zhang J, Hedrick JL, Lebrilla CB. Structure determination by MALDI-IRMPD mass spectrometry and exoglycosidase digestions of O-linked oligosaccharides from *Xenopus borealis* egg jelly. *Glycobiology.* 2011; 21(7):877–894. [PubMed: 21220250]



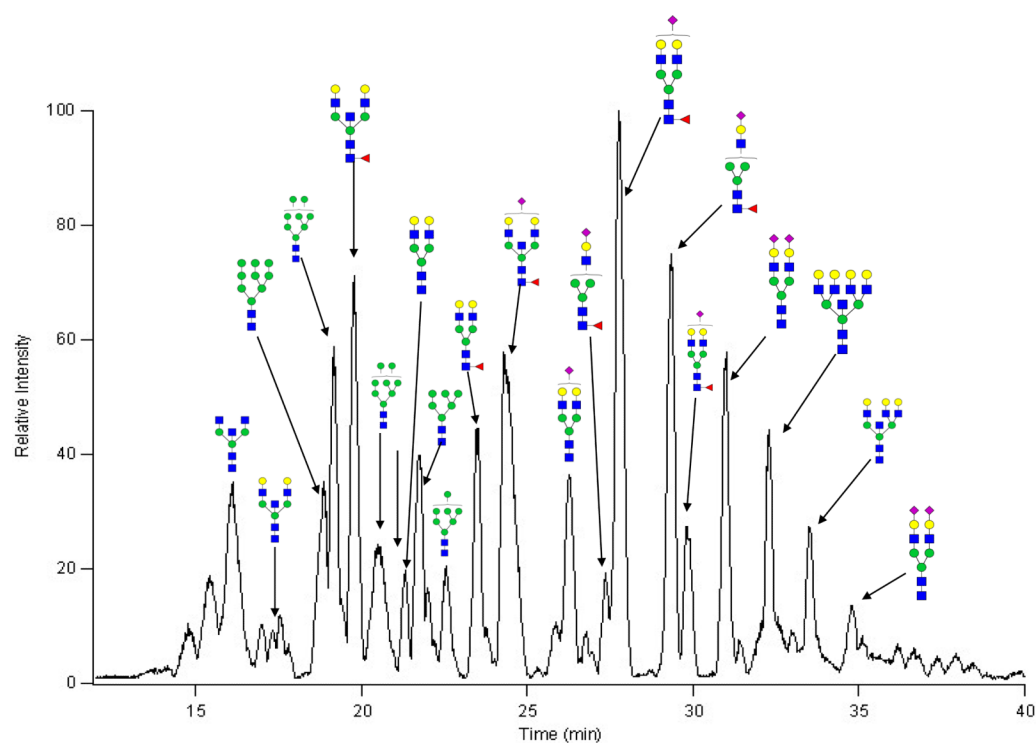
**Figure 1A.**

MALDI MS N-Glycan Profile of IgM in Positive Ion Mode. This spectrum shows the released N-glycans from the glycoprotein IgM with representation from all three glycan types: high mannose, complex, and hybrid. Sodiated ions are predominately produced. Putative structures are presented based on glycan composition. Blue squares represent N-Acetyl glucosamine, green circles represent mannose, yellow circles represent galactose, purple diamonds represent sialic acid, and red triangles represent fucose.



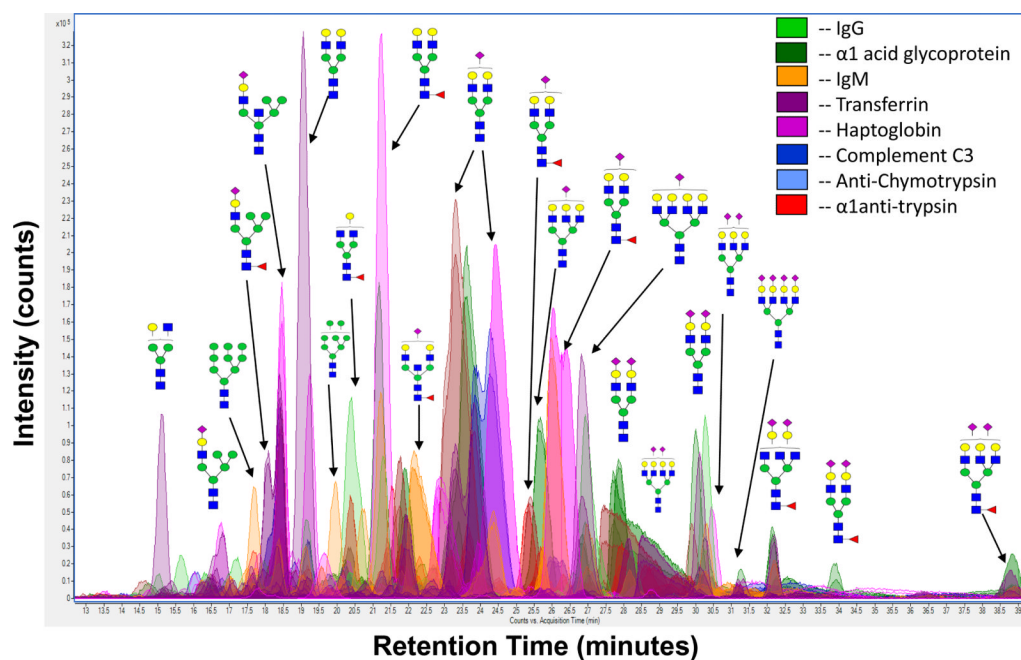


**Figure 1B.** MALDI MS N-Glycan Profile of IgM in Negative Ion Mode. This spectrum shows the increased detection of sialylated N-glycans from IgM. Deprotonated ions are predominately produced.

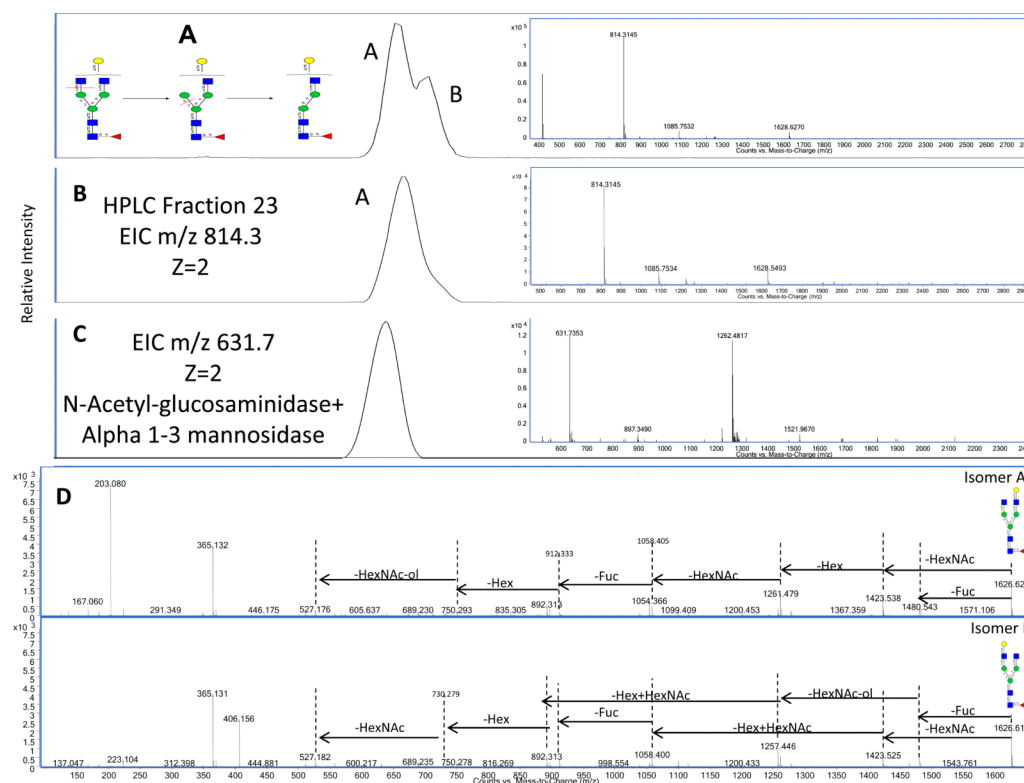


**Figure 2.**

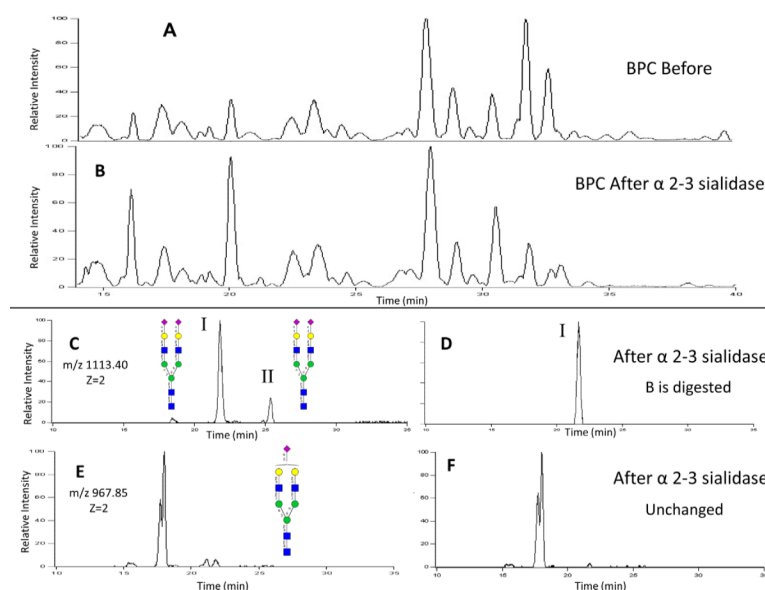
Base Peak Chromatogram of Glycans from IgM. NanoLC/MS Chromatogram of N-Glycans from IgM with glycan compositions labeled for the most abundant peaks. Isomer separation allows for the detection of isomers associated with each compound as well as the detection of more compounds due to decreased competition during ionization.



**Figure 3.** Global Glycan Pool. Chromatogram overlay of N-glycans from seven glycoproteins where each color represents a different glycoprotein. Abundant compositions from some of the major peaks have been labeled.

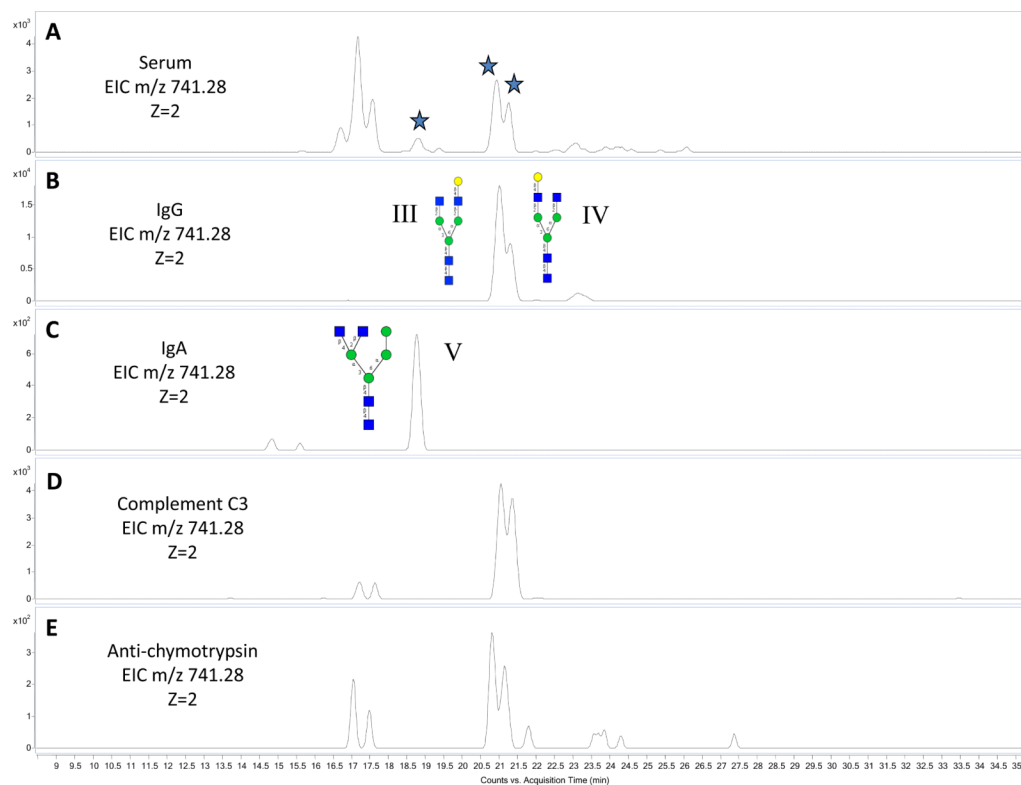
**Figure 4.**

A: Extracted Ion Chromatogram for m/z 814.29  $[M+2H]^+2$  showing two isomers associated with that composition. B: EIC of m/z 814.29  $[M+2H]^+2$  after off-line HPLC fractionation to isolate one isomer. C: EIC of m/z 631.74  $[M+2H]^+2$  after sequential exoglycosidase digestion. Mass spectrum for each chromatogram is inset.



**Figure 5.**

A: Base Peak Chromatogram (BPC) of Transferrin N-Glycans before enzyme digestion. B: BPC of Transferrin N-Glycans after alpha 2–3 sialidase digestion. C: EIC of  $m/z$  1113.40  $[M+2H]^{+2}$  before digestion. D: EIC of  $m/z$  1113.40  $[M+2H]^{+2}$  after digestion. E: EIC of  $m/z$  967.85  $[M+2H]^{+2}$  before digestion. F: EIC of  $m/z$  967.36  $[M+2H]^{+2}$  after digestion.

**Figure 6.**

A–E: EICs of  $m/z$  741.28  $[M+2H]^{+2}$  from different protein sources, serum, IgG, IgA, complement C3, and anti-chymotrypsin.