Article

A method for in-depth structural annotation of human serum glycans yields the biological variations

Ting Song, Danielle L. Aldredge, and Carlito B. Lebrilla

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.5b01340 • Publication Date (Web): 18 Jun 2015

A method for in-depth structural annotation of human serum glycans yields the biological variations

Ting Song, Danielle Aldredge, Carlito B. Lebrilla*

Department of Chemistry, University of California Davis, Davis, CA, 95616, USA

Abstract

Glycosylation is an important post-translational modification of proteins present in the vast majority of human proteins. For this reason, they are potentially new sources of biomarkers and active targets of therapeutics and vaccines. However, the absence of a biosynthetic template as in the genome and the general complexity of the structures have limited the development of methods for comprehensive structural analysis. Even now, the exact structures of many abundant N-glycans in serum are not known. Structural elucidation of oligosaccharides remains difficult and time consuming. Here, we introduce a means of rapidly identifying released N-glycan structures using their accurate masses and retention times based on a glycan library. This serum glycan library, significantly expanded from a previous one covering glycans released from a handful of serum glycoproteins, has over 170 complete and partial structures and constructed instead from whole serum. The method employs primarily nanoflow liquid chromatography and accurate mass spectrometry. The method allows us to readily profile N-glycans in biological fluids with deep structural analysis. This approach is used to determine the relative abundances and variations in the N-glycans from several individuals providing detailed variations in the abundances of the important N-glycans in blood.

Introduction

Alterations of glycoconjugates are found to correlate with pathological states making them new potential targets for therapeutic drugs and as biomarkers for diseases.[1-10] The human serum is highly glycosylated.[11-14] In general, glycoconjugates are ubiquitous in biological systems and are highly distributed on cell surfaces and in secreted proteins. As a post-translational modification, glycosylation may be more responsive and sensitive to changes in biological states.[2-10,15,16] However, the utility of glycoconjugates and the general understanding of their functions are severely limited by the lack of rapid methods for structural analysis. Indeed, the large variety of structures, the micro-heterogeneity, and the unavailability of a biological template have conspired to severely limit our efforts to translate the glycomic code.[17]

Glycans on proteins are differentiated into at least two major classes: N-glycan and O-glycan, according to the connecting atom on the protein backbone.[18] N-Glycans are significantly more abundant in serum. They are further differentiated into three subtypes depending on their composition as governed by the biosynthetic pathway. They include complex, high mannose, and hybrid types. N-Glycans are produced with a single core structure composed of Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAc. Additional monosaccharides such as mannose (Man), N-acetyl-glucosamine (GlcNAc), galactose (Gal), fucose (Fuc), and N-acetylneuraminic acid (NeuAc) are added to elongate the structure.[18] By adding different numbers and types of monosaccharides through various linkages, branching and multiple antennae are produced resulting in large diversity of structures.[17] Because the biosynthesis of N-glycans involves the competition of a number of glycosyltransferases, there are no templates as in genomics and proteomics to predict N-glycan structures.[18]

Even with the ubiquity of N-glycosylation, many of the exact structures of abundant glycans in serum are still unknown. However, these glycans could provide important targets either as diagnostic markers or for pharmaceutical drugs. They also provide important targets for synthesis. Analytical methods for structural characterization of glycans have progressed significantly.[19-23] For example, accurate mass readily yields glycan compositions, while tandem MS provides sequence and even linkage information.[19] Tandem MS has also been shown to be effective for structural identification of glycans.[19,24] Profiling the human serum N-glycome is performed mostly on the compositional level.[25][Put in more references.]

There are methods that can provide structural profile of glycans. A chromatographic method for profiling 2AB labeled glycans in serum with structural analyses and isomer differentiation was developed by Rudd and co-workers using HILIC.[26,27] The method employs a glycan relational database GlycoBase, which contains 300 partial and complete structures. There are limitations in distinguishing positional isomers of compounds containing more than one sialic acid. Furthermore, because HILIC does not provide extensive isomer separation and detection relies primarily on spectrophotometric techniques, it requires exoglycosidase digestion in every analysis.

N-Glycans and their isomers are readily separated and profiled on a LC-chip with a column packed with porous graphitized carbon (PGC).[28] In our hands, PGC provides the best isomer separation of glycans compared to any other stationary phase. PGC separates both neutrals and highly sialylated glycans effectively with high retention time repeatability.[28] Neutral and sialylated species are separated on the isomer level. Even highly sialylated species are eluted as shown recently by Kronewitter et al with glycans containing as many as

11 sialic acids.[29]

In this study, we developed an in-depth glycomic tool for profiling N-glycans with comprehensive structural analysis including isomer differentiation based on an N-glycan library produced directly from serum. This library resolves most crucial linkages (*e.g.* linkages of the sialic acids of most of the sialylated N-glycans) thereby allowing the identification of N-glycans on an isomeric level. Currently, the serum N-glycan library contains ~50 complete structures, ~100 partial structures, and 177 entries in total. Moreover, because serum glycans are generally representative of human glycosylation, it can be used for other tissues. The method for glycan identification relies on employing accurate masses and nanoflow liquid chromatography (nanoLC) retention times for the vast majority of the compounds. However, tandem MS can also be used where necessary. The high retention time reproducibility, high mass accuracy of nanoLC-CHIP-Q-TOF, and the excellent repeatability of sample preparation that are well tested ensure the effectiveness of this method.[7,15,17,28,30,31]

EXPERIMENTAL SECTION

Chemicals and reagents. Human pooled serum and standard serum proteins including immunoglobulin A, immunoglobulin M, anti-trypsin, and transferrin were purchased from Sigma-Aldrich (St. Louis, MO). Nine individual human sera were collected from healthy patients by the UCDavis Medical Center following a protocol approved by the UCDavis Medical Center IRB. Peptide: N-Glycosidase F (PNGase F) and exoglycosidases including α(2-3) neuraminidase (sialidase), α(1-2,3) mannosidase, α(1-6) mannosidase, and β N-acetyl glucosaminidase (GlcNAcase) were purchased from New England Biolabs (Ipswich,

MA). β(1-4) Galactosidase and α(1-3,4) fucosidase were purchased from Prozyme (Hayward, CA). Sodium borohydride was purchased from Sigma-Aldrich (St. Louis, MO). All reagents were of analytical or HPLC grade.

Sample preparation. Human serum N-glycans were released, reduced and purified using a well-established protocol in our laboratory.[15] Briefly, N-glycans were enzymatically released by PNGase F from 50 μL of serum in a CEM microwave reactor (CEM Corporation, North Carolina). They were then purified and enriched by solid phase extraction (SPE) employing a graphitized carbon cartridge (GCC) (Alltech Associated, Deerfield, IL) on an automated Gilson GX-274 ASPEC liquid handler. The released N-glycans were chemically reduced with 1 M $NaBH_4$ in 65°C water bath for 1.5 hours and desalted by automated-SPE-GCC. Reduced N-glycans were eluted in 20% acetonitrile (ACN) in water (v/v) and 0.05% trifluoroacetic acid (TFA) in 40% ACN in water (v/v). The sample was dried down and reconstituted with 50μL of nanopure water.

Fractionation by HPLC. Human serum N-glycans were fractioned off-line with an Agilent Hewlett-Packard Series 1100 HPLC system using a Hypercarb PGC column (Thermo Scientific) (100 mm × 0.5 mm I.D., 5 μm particle size) for exoglycosidase sequencing. N-glycans were eluted with a binary solvent system comprised of 0.1% formic acid (FA) in 3% ACN in water (v/v) as solvent A, and 0.1% FA in 90% ACN in water (v/v) as solvent B, with a flow rate of 0.30 mL/min. Eighty fractions were collected from 20 min to 60 min with an interval of 0.5 min over the 70-min gradient. The collected fractions were dried and reconstituted in 10 uL of nanopure water.

Nano-LC-Chip-Q-TOF-MS analysis. Global N-glycan profiling of human serum

was performed using an Agilent nanoLC-Chip-Q-TOF MS. MS spectra were acquired in the positive mode. The instrument is composed of a micro-well plate auto-sampler, a 1200 series nano-LC system equipped with a capillary sample loading pump and a nano-pump, a chip-cube interface, and a 6520 Q-TOF MS. The nano-LC system employs a binary solvent consisting of A (0.1% FA in 3% ACN in water (v/v)) and B (0.1% FA in 90% ACN in water (v/v)). Human serum N-glycans were enriched and separated on the Agilent HPLC-Chip II comprised of a 40 nL enrichment column and a 75 μm x 43 mm analytical column both packed with PGC in 5 μm particle size. The sample was delivered by the capillary pump to the enrichment column at a flow rate of 4 μL/min and separated on the analytical column by the nano-pump at a flow rate of 0.3 μL/min with a gradient that was previously optimized for N-glycans: 0% B, 0-2.5 min; 0-16% B, 2.5-20 min; 16-44% B, 20-30 min; 44-100% B, 30-35 min; and 100% B, 35-45 min followed by pure A for 20 min of equilibration. The instrument was mass-calibrated with internal calibrant ions covering a wide m/z range (622.029, 922.010, 1221.991, 1521.972, 1821.952, 2121.933, 2421.914) to yield mass accuracies <5 ppm for MS and <20 ppm for MS/MS. Tandem MS spectra were acquired via collision-induced dissociation (CID).[24]

In-source fragmentation produces losses of residues that coincide with smaller compounds. This yields an apparent compound that elutes at the same location as the larger parent compound. To determine if an extracted chromatographic peak is due to the in-source fragmentation of a larger compound, the peak is matched to larger co-eluting homologs with nearly identical retention times. For example, a N-glycan loosing a fucose, a labile residue, can produce an apparent smaller compound due to in-source fragmentation with peaks that

coincide but with differing abundances. Generally, the abundances of the fragments can be as much as 10% of the parent.

Exoglycosidase digestion. Several exoglycosidases were used to elucidate the complete N-glycan structures using conditions previously described.[32] Namely, the reaction buffer was prepared by adding glacial acidic acid to 0.1 M ammonium acetate to reach the desired PH for the specific enzyme. In a typical reaction, 5 μL buffer, 3 μL fractioned N-glycan solution, and 0.5 μL enzyme were mixed and the reaction was carried out in a Thermo Precision incubator (Thermo scientific) at 37 °C. The digestion time depends on the activity of the specific enzyme and the amount of N-glycan used. The optimal digestion time for each enzyme was obtained by monitoring the reaction over several time points to ensure the specificity of the exoglycosidase and the completeness of the reaction. The reaction conditions for all of the exoglycosidases used in this study are tabulated in a previous report.[32] The oligosaccharides before and after enzyme digestion were monitored by nano-LC-Chip-Q-TOF MS.

RESULTS AND DISCUSSION

LC retention times and accurate masses are used as distinguishing features to compare unknown compounds against a library thereby achieving rapid in-depth N-glycan structural identification without time consuming exoglycosidase digestions. The serum N-glycome library was constructed to facilitate rapid identification. The library incorporated a more limited one created previously from nine commercial glycoproteins corresponding to the most abundant species in serum.[33] While that previous effort helped guide the current one, it

had severe limitations that is corrected in the present one. Namely, the commercial glycoproteins contained glycans that are not generally found in serum due to the processes involved in commercial preparation. Many of the serum glycans are not represented in the small number of proteins. And, there were technical limitations that were not easily identifiable such as the prevalence of in-source fragmentation in nanoLC MS. This database attempts to correct these limitations by obtaining serum glycans more comprehensively from all serum proteins with no protein enrichment. We further identified in-source fragments as described in the Experimental Section. The resulting library is more comprehensive, focused on serum, with the entries verified as unique compounds.

Extended comprehensive library. The overall workflow for the construction of the library is summarized in **Figure S1.** N-Glycans are first released from serum and reduced to eliminate the interference from anomers. HPLC is performed and the eluent collected into 80 fractions. Each fraction is further analyzed by nanoLC-CHIP-Q-TOF to obtain the number of compounds (including isomers) and their corresponding compositions. The compositions yield putative structures that are used to guide the exoglycosidase reactions. The enzymes are selected based on the putative structures. A number of enzyme reactions are used to determine the structure with the products monitored by nanoLC-CHIP-Q-TOF. HPLC separation can result in the fractionation of isomeric mixtures. To observe the extent of the enzymatic reactions both the loss of the precursor molecule and the rise of the digestion products are observed. The presence of isomers and other compounds aid in the analysis as not all the isomers will react with the same enzyme, and the relative loss of some isomers are indicative of the reaction. The presence of endogenous compounds corresponding to potential

digestion products similarly does not obscure the analysis as the method yields highly reproducible relative abundances. A decrease in the precursor ion corresponds to a proportional increase in the product ion.

N-Glycans before and after the exoglycosidase reactions were monitored by nanoLC-CHIP-Q-TOF. **Figure 1** provides one example demonstrating the procedure for exoglycosidase sequencing. The specificity of the enzymes was determined and extensively in previous studies.[34-36] **Figure 1a** shows the ECC of the major N-glycan, a mono-sialyated diantennary species (putative structure and MS inset), obtained in a single fraction using offline-HPLC of whole human serum. After 1 hour digestion with an α(2-3) neuraminidase, the peak for this compound diminished completely and a new peak corresponding to a loss of one sialic acid appeared as shown in **Figure 1b,** thereby confirming the linkage of the terminal sialic acid as α(2-3) (structure and MS inset). To confirm further the position of the terminal sialic acid, a cocktail of enzymes was used consisting of a β(1-4) galactosidase, β N-acetylglucosaminidase, and α(1-2, 3) mannosidase over a 24 hour reaction time. Because the exoglycosidases only cleave off the terminal monosaccharide from the non-reducing end, the branch with a sialic acid as the protecting cap will remain while the branch with the Gal exposed will be digested away. If the sialic acid is on the 1-6 arm (two arms differentiated by the linkage of the two branching mannoses), one Gal, one GlcNac and one Man will be cleaved off. Conversely, if the sialic acid is on the 1-3 arm, only one Gal and one GlcNac will be cleaved off while the α(1-6)Man will remain intact even in the presence of α(1-2,3) mannosidase. **Figure 1c** shows the ECC of the enzyme digestion product with the mass spectrum (inset) exhibiting the loss of one Gal, one GlcNac and one Man meaning the

terminal sialic acid is on the 1-6-antenna.

Tandem MS spectra were used to further distinguish isomers and to add an additional level of validation for the structures.[19,24,37] **Figure 2** shows the tandem MS spectra of four isomers corresponding to the most abundant composition in serum, namely the bi-antennary di-sialylated species. The composition yields four isomers: N5402a, N5402b, N5402c and N5402d. The structures are given a systematic name corresponding such as N5402a for N-glycan (N) with Hex_5:$HexNAc_4$:Fuc_0:$NeuAc_2$. The lower case letter corresponds to the isomer designation in order of abundance, with "a" being the most abundant in serum. The most abundant isomer N5402a is the structure with both sialic acid being linked in α(2-3) manner. The compound N5402c is the one with both sialic acid linked in α(2-6), and is significantly of lower in abundance. The compound N5402b and N5402d are mixed linkages with one sialic acid being α(2-3) while the other α(2-6). These structures were confirmed by α(2-3) neuraminidase digestion of the isomeric mixtures. Tandem MS spectra further confirmed the differences between isomers producing unique fragmentation. The four spectra corresponding to the isomers (**Figure 2**) produce the same sialic acid peaks, 273.09 Da (neutral mass) and 291.09 Da corresponding to NeuAc-H_2O and NeuAc, but with significantly differing intensities. Each isomer apparently has its own favored fragmentation pathway. For example, N5402c (**Figure 2c**) dissociates from the reducing end losing the core GlcNac. Conversely, the other three isomers first lose sialic acid from the non-reducing end. Isomer N5402c is the only one that does not have a α(2-3) sialic acid. The other isomers contain at least one α(2-3) sialic acid suggesting that the α(2-3) linkage is more labile than α(2-6). This observation is difficult to confirm due to the lack of standards, but it may be a

useful indicator of this specific structural feature. The CID of N5402d also appears to induce a sialic acid migration producing a loss of an internal Gal residue. There is precedence for monosaccharides migration in the tandem MS of native N-glycans.[38] The biantennary di-sialylated compounds illustrate a limitation of this method. We were able to determine the linkages on the more abundant mixed-linked species (N5402b) by tandem MS and enzymatic digestion, however we could not isolate sufficient amounts of N5402d for similar analysis. Instead, the structure was assigned by deduction and elimination of the three more abundant species.

Table S1 summarizes the comprehensive library containing over 170 distinct compounds with over 50 complete structures and 100 partially elucidated structures. Linkages and positions of terminal sialic acids of most sialylated species are resolved. The compounds containing only partial structures could not be fully elucidated due to their low abundances, or due to the lack of availability of specific enzymes. For example, entry number 14, 21, 60, 97 are four isomers of the tri-antennary di-sialylated N-glycans observed in serum with distinct retention times. Linkages of the sialic acids of number 14 and 21 are resolved, however the positions cannot be resolved due to the lack of an efficient and commercial β(1-4) and β(1-6) N-acetylglucosaminidases. Linkages of sialic acids of isomer 60 and 97 are not resolved because of their extremely low abundances. Furthermore, although 14 and 21 have sialic acids that cannot be localized, the two compounds are distinct and can be defined by their retention times. N-Glycans in this library are sorted approximately in decreasing order of relative abundances in serum. A more systematic study of the abundances is discussed below. Systematic names, LC retention times, accurate neutral masses of reduced form,

symbolic structures with annotated linkages, and relative abundances normalized to the total N-glycan abundance are provided in **Table S1**.

Application of profiling method for structural analysis of glycans from commercial glycoproteins. In order to validate the method for profiling with instantaneous structural identification, we examined N-glycans released from several commercial human serum glycoproteins, including immunoglobulin A (IgA), immunoglobulin M (IgM), anti-trypsin, and transferrin. The glycans were released from the glycoprotein and identified by matching retention times and accurate masses to the library. The predicted structures were then validated by exoglycosidase sequencing. **Figure S2** shows a glycan from immunoglobulin A. The ECC (MS inset) of N5411a shows the compound before digestion and the structure identified by matching with the library (**Figure S2a**). After 1 hour reaction with an α(2-3) neuraminidase, the original peak diminished and a new peak corresponding to N5410 emerged (**Figure S2b**) confirming that the sialic acid linkage is α(2-3). The ECC of the product with a loss of one GlcNac and one Gal after treatment with β(1-4) galactosidase, β N-acetylglucosaminidase, and α(1-2/3) mannosidase over a 24-hour reaction time provides the location of the sialic acid as being on 1-3-antenna (**Figure S2c**). The result is indeed consistent with the library assignment. This method was used to test approximately 20 randomly selected, relatively abundant structures from the above proteins. All of the assignments were correctly predicted as confirmed by enzymatic digestion.

Application of the rapid identification method to nine individual sera. N-Glycans in individual serum were identified by matching the LC retention times and accurate masses from the library to profile and rapidly identify structures. Mass accuracy is consistent with

little drifts over several runs because they are optimized with known standards as part of the procedure for their operation. For separation, there is no internal optimization to provide identical retention times. Furthermore, there are no well-characterized mixtures of standard oligosaccharides. Nonetheless, the retention times are suitably reproducible over a two-three day period. Variations in retention times over delayed periods of several days and months were corrected by using the glycans in commercial serum as the standard oligosaccharide mixture as described below.

The N-glycans released from serum are highly reproducible in the composition and relative abundances. We have previously shown that there are generally small variations in the abundances of the major components of N-glycans over hundreds of samples.[28,39] Indeed, serum N-glycans are so reproducible that they make a suitable standard mixture. In this method, we use serum glycans to align chromatograms allowing this method to be potentially transportable to other instruments and other laboratories.

Alignment of the chromatogram is performed systematically and stepwise using groups of isomers. We first extract the chromatogram of a specific mass, for example that corresponding to the composition N5402 shown in **Figure 3a**. The extracted chromatogram (for 2224.80 Da - reduced) shows four isomers with a distinctive abundance pattern. **Figure 3b** is the chromatogram from a commercial serum sample. **Figure 3c** is from a single individual. Based on this comparison, it is relatively simple to identify each component and assign the structure with the peak. Using nanoLC/MS in this way, we can comprehensively assign the N-glycans associated with serum.

The reproducibility of the technical replicates and the overall procedure was

examined by comparing retention times and abundances (**Figure S3**). There is high reproducibility in the total ion chromatogram as shown for both multiple injections (5x) and multiple sample preparations (5x). The compound extraction chromatogram in **Figure S4** shows the excellent overlap for five injections of the disialylated, biantennary structure.

Shown in **Figure 4** is a comprehensively annotated chromatogram using the top 100 N-glycans. The numbers correspond to the entries with structures provided in **Table S1**. The profiling method was employed to determine the variations in the released N-glycans from nine individual sera. The variations in terms of glycan types are illustrated for nine individuals in **Figure 5,** with the abundances of the total N-glycan, sialylated (C/H-S), fucosylated (C/H-F), high mannose (HM), and non-fucosylated non-sialylated hybrid and complex species (C/H). Coefficient of variation (CV) was determined for each category corresponding to be 13.2%, 15.5%, 14.4%, 11.4% and 14.0%, respectively, indicating the biological diversity between individuals.

To illustrate the biological variations in abundances between individuals, the 10 most abundant N-glycans in serum were selected and identified (**Figure S5**) using direct ion abundances. While there are accepted differences in ionization and detection efficiencies, these effects are generally minimized when the compounds are separated as they are in liquid chromatography.[40] The structures correspond to N5402a, N5401a, N5511a, N5401b, N5410a, N5400a, N3410a, N4410a, N5411a, and N5412a. Included in the figure are the potential protein sources for each glycan based on the earlier study.[33] The most abundant glycoprotein in serum corresponds to IgG. However, while the glycans from IgG are among the top 10 structures, the most abundant N-glycan N5402a is from several other proteins including

α-2-macroglobulin, transferrin and anti-trypsin.[33]

As expected the CV in the individual structures is greater than those for the entire glycan types. The CV for each structure is an indication of the biological diversity. The structure with the highest CV corresponds to N3410a (31.5 %). For comparison the method replicate yields a CV of 2.8% (**Figure S6**). This compound N3410a is the most abundant glycan in IgG and plays an important role in the effector function of the antibody.[32] The second highest CV is N4410a, which is also derived from IgG. Conversely, the least fluctuating N-glycan is the structure with N5400a with a CV of 8.7%. These results suggest that the greatest variation in glycosylation is perhaps related to IgG glycosylation.

Figure 6 shows the N-glycome of human serum based on nine individual sera with the relative abundances averaged for the nine individual. This "glycan wheel" illustrates the diversity of the human serum N-glycome. Where possible, the complete structures are provided along with the respective abundances. The chart shows that the N-glycome is much more diverse than human plasma proteome, where a small number of proteins account for the vast majority of the abundances. The most abundant N-glycan accounts for 10 % of the total N-glycan abundances. It has long been known that the most abundant composition in serum corresponds to the di-sialylated bi-antennary species. We can now assign this structure as being N5402a, the disialylated compound with both sialic acids linked via α(2,3).The second most abundant structure (N5401a) accounts for about 6 %, the third to the seventh accounting for 3 % individually, the eighth to the fifteenth accounting for 2 % individually, and each of the rest only accounting for 1 % or less. The top 20 N-glycans account for 52 % of the total N-glycome in human serum. The remaining 48 % consists of over a hundred (and perhaps

many more) structures.

CONCLUSION

The identification of N-glycan structures based on a functional library of serum glycans is used to profile structures in biological samples with deep structural analysis. This method provides a way forward for rapid throughput glycomics. To use this library, accurate masses and LC retention times are the most critical parameters to match between study samples and the library. Commercial serum N-glycan pool was used as the standard to correct the retention times relative to the library when needed.

The N-glycan variation in serum was monitored between nine individuals from the most abundant to the least, representing over four orders of magnitude and corresponding to over 170 structures. Interestingly, the most abundant N-glycan is not from the most abundant glycoprotein, IgG, but originated from several glycoproteins. Furthermore, the abundances are distributed throughout many structures that vary slightly between different individuals. An annotated global serum N-glycome will allow us to visualize the complexity and diversity of N-glycans providing useful targets for future biomarker and therapeutic studies.

ACKNOWLEDGEMENTS

Funding provided by the National Institutes of Health (R01GM049077) is gratefully acknowledged.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

ASSOCIATED CONTENT

Supporting Information Available

The supplementary figures as noted in text.

This material and the library are available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Lin, S. Q.; Kemmner, W.; Grigull, S.; Schlag, P. M. *Experimental cell research* **2002**, *276*, 101-110.
(2) An, H. J.; Kronewitter, S. R.; de Leoz, M. L. A.; Lebrilla, C. B. *Curr Opin Chem Biol* **2009**, *13*, 601-607.
(3) An, H. J.; Miyamoto, S.; Lancaster, K. S.; Kirmiz, C.; Li, B. S.; Lam, K. S.; Leiserowitz, G. S.; Lebrilla, C. B. *J Proteome Res* **2006**, *5*, 1626-1635.
(4) de Leoz, M. L. A.; An, H. J.; Kronewitter, S.; Kim, J.; Beecroft, S.; Vinall, R.; Miyamoto, S.; White, R. D.; Lam, K. S.; Lebrilla, C. *Dis Markers* **2008**, *25*, 243-258.
(5) de Leoz, M. L. A.; Young, L. J. T.; An, H. J.; Kronewitter, S. R.; Kim, J. H.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B. *Mol Cell Proteomics* **2011**, *10*.
(6) Hua, S.; An, H. J.; Ozcan, S.; Ro, G. S.; Soares, S.; DeVere-White, R.; Lebrilla, C. B. *Analyst* **2011**, *136*, 3663-3671.
(7) Kirmiz, C.; Li, B. S.; An, H. J.; Clowers, B. H.; Chew, H. K.; Lam, K. S.; Ferrige, A.; Alecio, R.; Borowsky, A. D.; Sulaimon, S.; Lebrilla, C. B.; Miyamoto, S. *Mol Cell Proteomics* **2007**, *6*, 43-55.
(8) Lebrilla, C. B.; An, H. J. *Mol Biosyst* **2009**, *5*, 17-20.
(9) Leiserowitz, G. S.; Lebrilla, C.; Miyamoto, S.; An, H. J.; Duong, H.; Kirmiz, C.; Ll, B.; Liu, H.; Lam, K. S. *Int J Gynecol Cancer* **2008**, *18*, 470-475.
(10) Packer, N. H.; von der Lieth, C. W.; Aoki-Kinoshita, K. F.; Lebrilla, C. B.; Paulson, J. C.; Raman, R.; Rudd, P.; Sasisekharan, R.; Taniguchi, N.; York, W. S. *Proteomics* **2008**, *8*, 8-20.
(11) Alavi, A.; Axford, J. S. *Rheumatology* **2008**, *47*, 760-770.
(12) Dube, D. H.; Bertozzi, C. R. *Nat Rev Drug Discov* **2005**, *4*, 477-488.
(13) Robinson, L. N.; Artpradit, C.; Raman, R.; Shriver, Z. H.; Ruchirawat, M.; Sasisekharan, R. *Electrophoresis* **2012**, *33*, 797-814.
(14) Varki, A. *Trends Mol Med* **2008**, *14*, 351-360.
(15) Hua, S.; Lebrilla, C.; An, H. J. *Bioanalysis* **2011**, *3*, 2573-2585.
(16) Taniguchi, N.; Paulson, J. C. *Proteomics* **2007**, *7*, 1360-1363.
(17) Kronewitter, S. R.; An, H. J.; de Leoz, M. L.; Lebrilla, C. B.; Miyamoto, S.; Leiserowitz, G. S. *Proteomics* **2009**, *9*, 2986-2994.
(18) Brooks, S. A.; Dwek, M. V.; Schumacher, U. *Functional and molecular glycobiology*; BIOS Scientific ; Distributed exclusively in the United States of America, its dependedent territories and ... by Springer-Verlag: Oxford New York, 2002, p xvi, 354 p.
(19) An, H. J.; Lebrilla, C. B. *Mass Spectrom Rev* **2011**, *30*, 560-578.

(20) Gupta, G.; Surolia, A.; Sampathkumar, S. G. *Omics* **2010**, *14*, 419-436.
(21) Takegawa, Y.; Deguchi, K.; Nakagawa, H.; Nishimura, S. I. *Anal Chem* **2005**, *77*, 6062-6068.
(22) Singh, S.; Reinhold, V. *Glycobiology* **2003**, *13*, 848-848.
(23) Zhang, H. L.; Singh, S.; Reinhold, V. *Glycobiology* **2004**, *14*, 1192-1192.
(24) Wu, S.; Salcedo, J.; Tang, N.; Waddell, K.; Grimm, R.; German, J. B.; Lebrilla, C. B. *Anal Chem* **2012**, *84*, 7456-7462.
(25) Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B. *Proteomics* **2009**, *9*, 1939-1951.
(26) Saldova, R.; Asadi Shehni, A.; Haakensen, V. D.; Steinfeld, I.; Hilliard, M.; Kifer, I.; Helland, A.; Yakhini, Z.; Borresen-Dale, A. L.; Rudd, P. M. *J Proteome Res* **2014**, *13*, 2314-2327.
(27) Campbell, M. P.; Royle, L.; Radcliffe, C. M.; Dwek, R. A.; Rudd, P. M. *Bioinformatics* **2008**, *24*, 1214-1216.
(28) Ruhaak, L. R.; Taylor, S. L.; Miyamoto, S.; Kelly, K.; Leiserowitz, G. S.; Gandara, D.; Lebrilla, C. B.; Kim, K. *Anal Bioanal Chem* **2013**, *405*, 4953-4958.
(29) Kronewitter, S. R.; Marginean, I.; Cox, J. T.; Zhao, R.; Hagler, C. D.; Shukla, A. K.; Carlson, T. S.; Adkins, J. N.; Camp, D. G., 2nd; Moore, R. J.; Rodland, K. D.; Smith, R. D. *Anal Chem* **2014**, *86*, 8700-8710.
(30) Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H. F.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B. *Proteomics* **2009**, *9*, 1939-1951.
(31) Kronewitter, S. R.; de Leoz, M. L. A.; Peacock, K. S.; McBride, K. R.; An, H. J.; Miyamoto, S.; Leiserowitz, G. S.; Lebrilla, C. B. *J Proteome Res* **2010**, *9*, 4952-4959.
(32) Song, T.; Ozcan, S.; Becker, A.; Lebrilla, C. B. *Anal Chem* **2014**, *86*, 5661-5666.
(33) Aldredge, D.; An, H. J.; Tang, N.; Waddell, K.; Lebrilla, C. B. *J Proteome Res* **2012**, *11*, 1958-1968.
(34) Zhang, J. H.; Lindsay, L. L.; Hedrick, J. L.; Lebrilla, C. B. *Anal Chem* **2004**, *76*, 5990-6001.
(35) Wu, S. A.; Grimm, R.; German, J. B.; Lebrilla, C. B. *J Proteome Res* **2011**, *10*, 856-868.
(36) Wu, S. A.; Tao, N. N.; German, J. B.; Grimm, R.; Lebrilla, C. B. *J Proteome Res* **2010**, *9*, 4138-4151.
(37) Penn, S. G.; Cancilla, M. T.; Lebrilla, C. B. *Anal Chem* **1996**, *68*, 2331-2339.
(38) Franz, A. H.; Lebrilla, C. B. *J Am Soc Mass Spectr* **2002**, *13*, 325-337.
(39) Kronewitter, S. R.; de Leoz, M. L.; Peacock, K. S.; McBride, K. R.; An, H. J.; Miyamoto, S.; Leiserowitz, G. S.; Lebrilla, C. B. *J Proteome Res* **2010**, *9*, 4952-4959.
(40) Lebrilla, C. B.; An, H. J. *Mol Biosyst* **2009**, *5*, 17-20.

FIGURE CAPTION

Figure 1. The extracted compound chromatograms (ECCs) of a mono-sialylated diantennary species (putative structure inset) before and after exoglycosidase digestion yields the linkage and position of the sialic acid. (a) The chromatogram of N5401 from an HPLC fraction before enzyme digestion. (Neutral mass of 1933.71 Da was used to produce the ECC). (b) The chromatogram of the N-glycan after one hour digestion with α(2-3) sialidase at 37 ℃. Neutral mass of 1642.61 Da is used to produce the ECC. (c) The chromatogram of the N-glycan residue after 24 hours digestion with a cocktail of exoglycosidases including β (1-4) galactosidase, β N-acetylglucosaminidase and α(1-2,3) mannosidase at 37 ℃ of this fraction. (Neutral mass of 1406.53 Da is used to produce the ECC).

Figure 2. Tandem mass spectra of the four isomers of N5402 (symbolic structures inset, and the complete structures as elucidated by exoglycosidases).

Figure 3. (a) The ECC of mass 2224.80 Da from the library. (b) The ECC of mass 2224.80 Da from the commercial serum N-glycans run together with study samples. (c) The ECC of mass 2224.80 Da from an individual serum of the study sample.

Figure 4. The total compound chromatogram of the reference library with peaks annotated for structures. Listed are the 100 most abundant components. Less abundant components are not labeled for clarity. The annotation numbers correspond to structures provided in **Table S1**.

Figure 5. The variations in relative abundances of the N-glycan subclasses for the nine individual human serum samples.

Figure 6. A human serum "glycan wheel" based on the relative abundances averaged for the nine individual sera.

Figure 1.

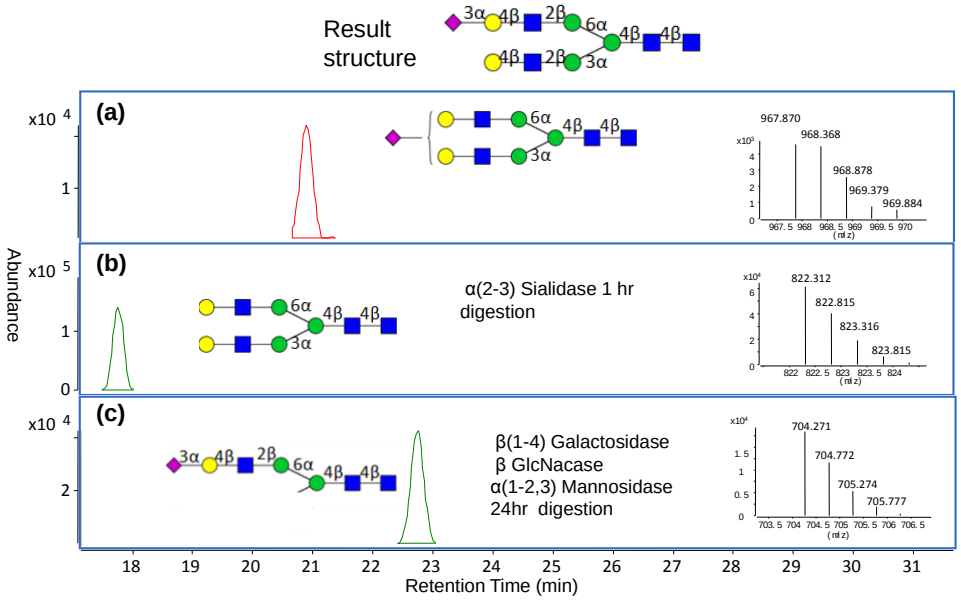

Result structure
(a)
(b)
(c)
α(2-3) Sialidase 1 hr digestion
β(1-4) Galactosidase
β GlcNacase
α(1-2,3) Mannosidase
24hr digestion
Abundance
Retention Time (min)

Figure 2.

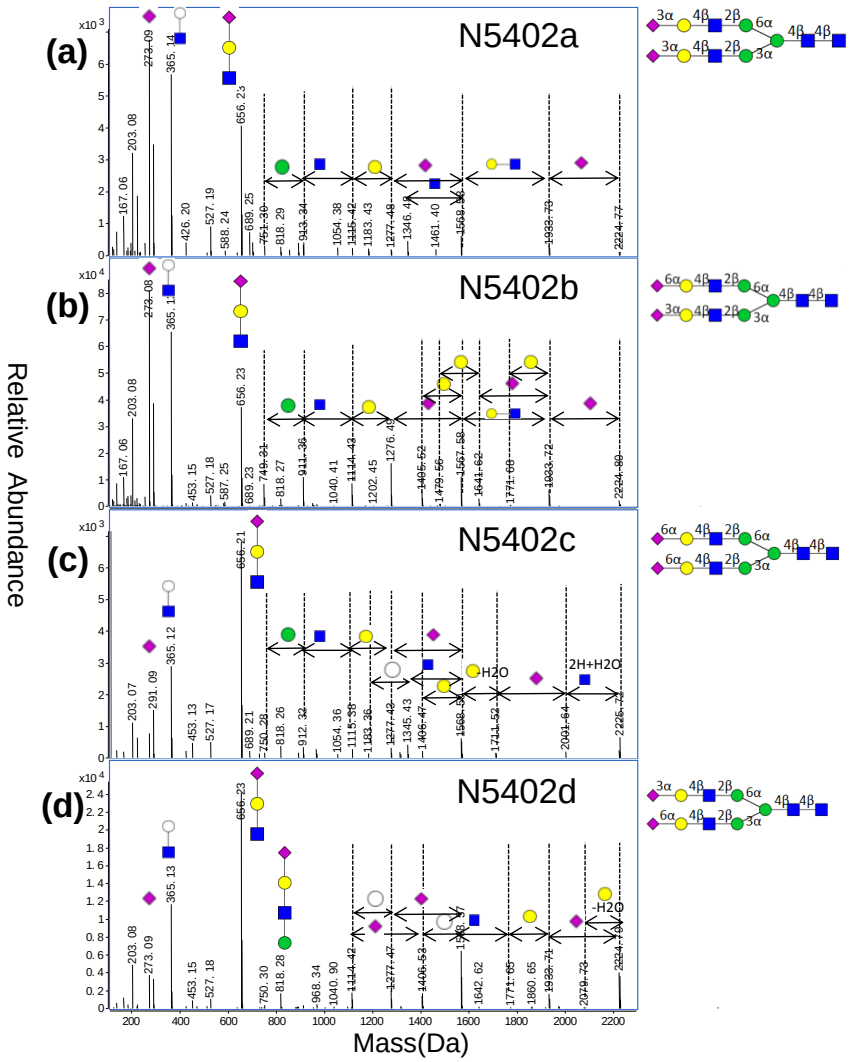

(a)
N5402a
(b)
N5402b
(c)
N5402c
(d)
N5402d
Relative Abundance
Mass(Da)
H2O
2H+H2O
-H2O

Figure 3.

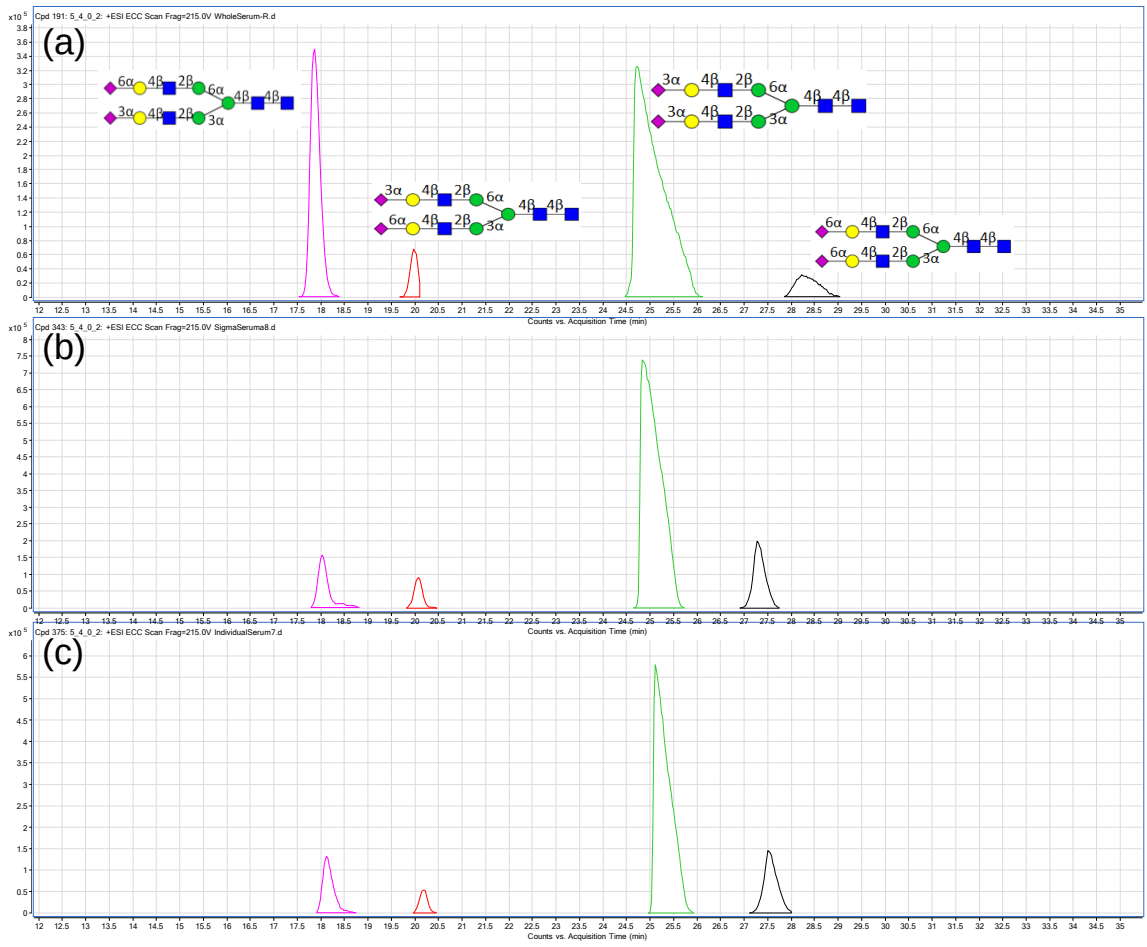


Figure 4.

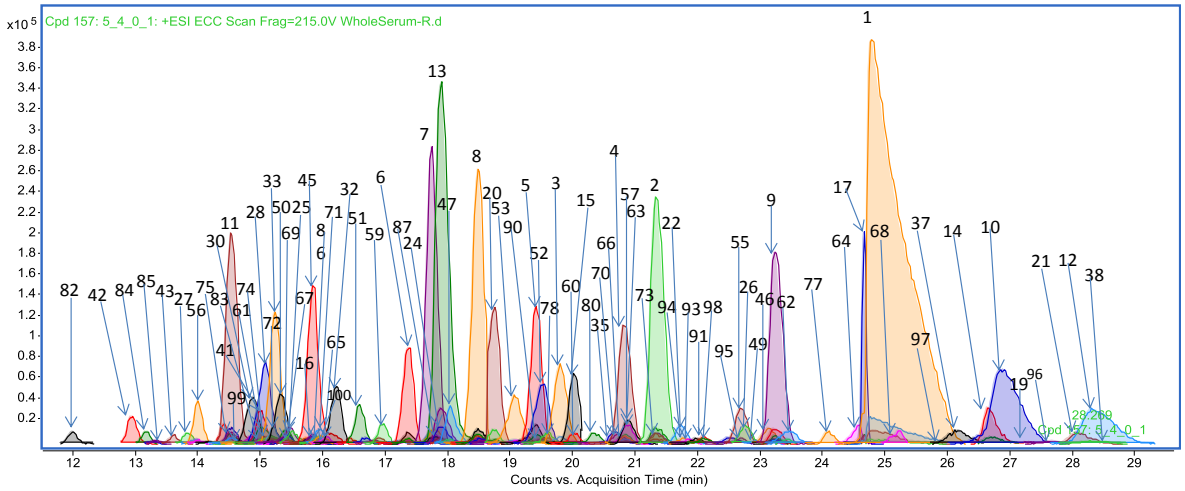

Cpd 157: 5_4_0_1: +ESI ECC Scan Frag=215.0V WholeSerum-R.d
x10 5
Counts vs. Acquisition Time (min)

Figure 5.

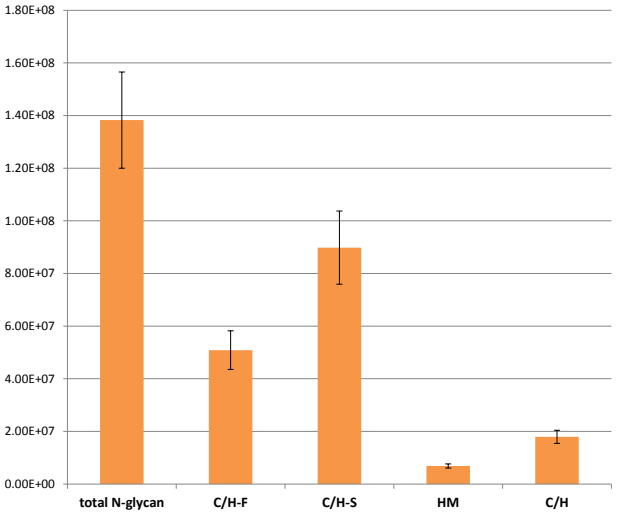
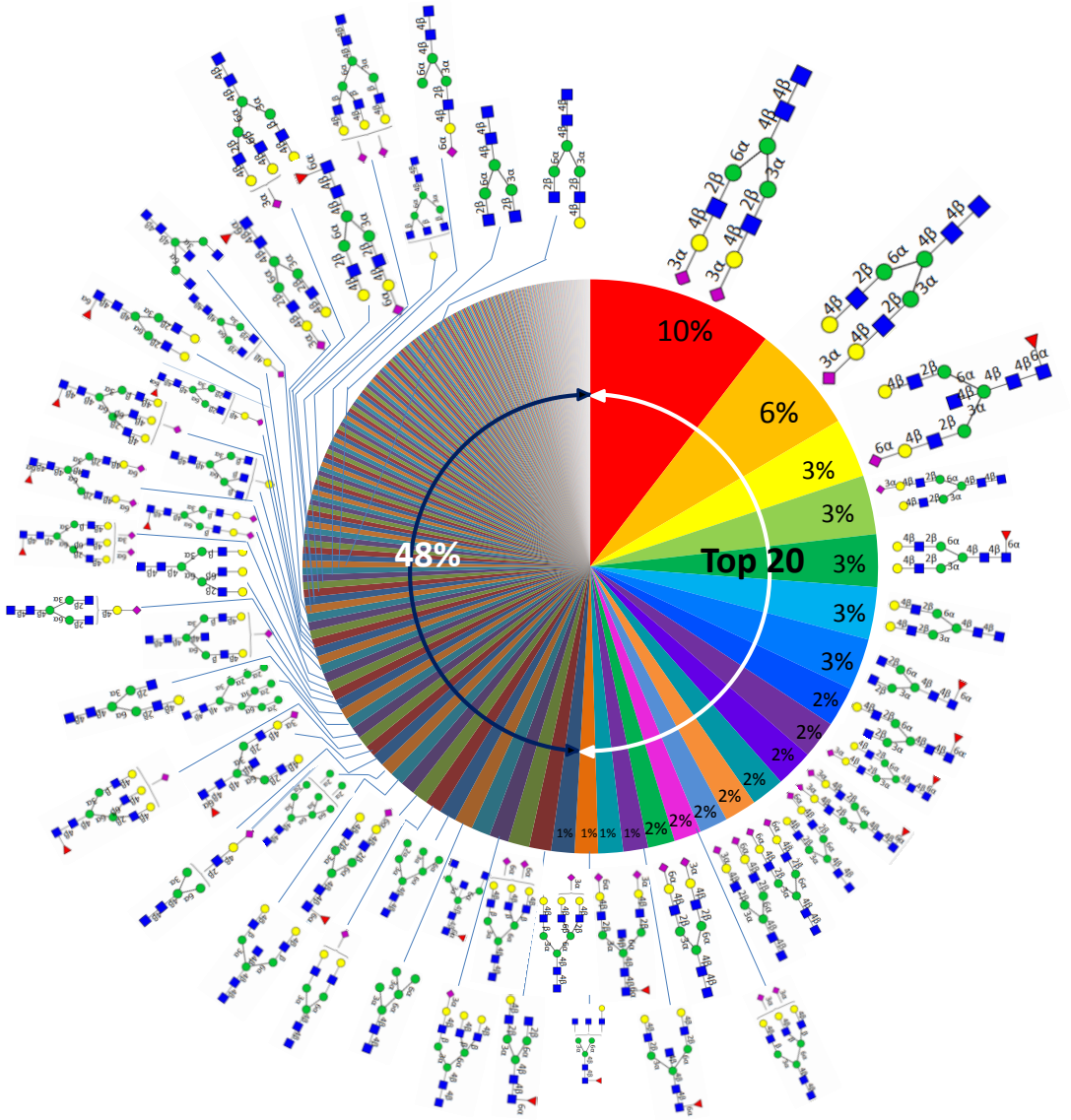

1.80E+08
1.60E+08
1.40E+08
1.20E+08
1.00E+08
8.00E+07
6.00E+07
4.00E+07
2.00E+07
0.00E+00
total N-glycan
C/H-F
C/H-S
HM
C/H
CV%: 13.2% 14.4% 15.5% 11.4% 14.0%

Figure 6.



For table of content