research articles Journal of proteome research

A New Computer Program (GlycoX) To Determine Simultaneously the Glycosylation Sites and Oligosaccharide Heterogeneity of Glycoproteins

Hyun Joo An,[†] John S. Tillinghast,[‡] David L. Woodruff,[§] David M. Rocke,[‡] and Carlito B. Lebrilla*[, †, ||]

Department of Chemistry, Division of Biostatistics, Graduate School of Management, Biochemistry and Molecular Medicine, University of California, Davis, California 95616

Received June 16, 2006

A new computer program, GlycoX, was developed to aid in the determination of the glycosylation sites and oligosaccharide heterogeneity in glycoproteins. After digestion with the nonspecific protease, each glycan at a specific glycosylation site contains a small peptide tag that identifies the location of the glycan. GlycoX was developed in MATLAB requiring the entry of the exact masses of the glycopeptide and the glycan spectra in the form of a mass-intensity table and taking advantage of the accurate mass capability of the mass analyzer, in this case a Fourier transform ion cyclotron resonance (FT ICR) mass spectrometer. This program computes not only the glycosylation site but also the composition of the glycans at each site. Several glycoproteins were used to determine the efficacy of GlycoX. These glycoproteins range from the simple, with one site of glycosylation, to the more complex, with multiple (three) sites of glycosylation. The results obtained using the computer program were the same as those determined manually. Model glycoproteins yielded the correct results, and new glycoproteins with unknown glycosylation were examined with the site of glycosylation and the corresponding glycans determined. Furthermore, other functions in GlycoX, including an auto-isotope filter to identify monoisotopic peaks and an oligosaccharide calculator to obtain the oligosaccharide composition, are demonstrated.

Keywords: glycosylation site • oligosaccharide heterogeneity • software • oligosaccharide calculator • isotope filter • mass spectrometry

Introduction

Glycosylation is one of the most common forms of post-translational modification in eukaryotic proteins and is involved in many cell communication and signaling events.[1] Glycans play key roles in protein folding, cell−cell recognition, and the immune system.[2,3] Defective or altered glycosylation can have profound biological implications and has been associated with a number of human diseases, for example, the congenital disorders of glycosylation[4,5] and cancer.[6]

It has been estimated that at least 50% of human proteins are glycosylated.[7] There are two major types, N- and O-glycosylation. The N-linked glycans are linked via an amide bond to an asparagine with a consensus sequence Asn-X-Ser (Thr), where X can be any amino acid except proline. N-linked oligosaccharides have a single core consisting of two *N*-acetylglucosamine (GlcNAc) and three mannoses (Man). The O-linked glycans are bound to either a serine or threonine with no single common core or consensus sequence at the attached peptide.

The determination of glycosylation sites and oligosaccharide heterogeneity is key toward understanding the specific biological roles of glycoproteins. Traditionally, site-specific glycosylation analysis has been extremely challenging due to the high complexity of glycoproteins. There have been several recent reports on glycosylation site analysis.[8−16] Typical approaches to this task are based upon some combination of specific enzymatic proteolysis (usually with trypsin), fractionation of glycopeptides (most often by liquid chromatography or affinity chromatography), and glycopeptide analysis by mass spectrometry (MS).[12−16] In some cases, deglycosylation of glycopeptides is concurrently performed with the information regarding the glycan discarded.[12,16] Unfortunately, many glycoproteins are resistant to trypsin.[17,18] Furthermore, the glycopeptides that are formed by specific enzymatic proteolysis are often too large for MS analysis because glycosylation may produce missed cleavages.[14] For these reasons, the information regarding glycosylation is very often incomplete.

Recently, a new strategy for the rapid determination of N-glycosylation sites and site heterogeneity was introduced by

* Corresponding author. Tel, +1-530-752-6364; fax, +1-530-752-8995; e-mail, cblebrilla@ucdavis.edu.

[†] Department of Chemistry, University of California.

[‡] Division of Biostatistics, University of California.

[§] Graduate School of Management, University of California.

[||] Biochemistry and Molecular Medicine, University of California.

10.1021/pr0602949 CCC: $33.50

Published on Web 09/08/2006

our laboratory.[9] In this approach, a glycoprotein is subjected to nonspecific proteolysis with a highly active mixture of proteases known as Pronase. Nonglycosylated regions of the glycoprotein are digested to amino acids and dipeptides, while glycosylated regions are protected from protease activities by the oligosaccharides, which block the digestion of the associated peptide segment by steric interactions. The resulting mixture of glycopeptides is easily separated from the amino acids and salts using solid phase extraction (SPE) with porous graphitic carbon (PGC). In parallel, the glycoprotein is treated with PNGase F, an enzyme that specifically releases N-linked glycans from glycoproteins. However, this procedure is necessary only for large glycoproteins with several glycosylation sites. Smaller glycoproteins do not typically require the determination of the oligosaccharide constituents. In any case, the released glycans are also purified by SPE with PGC. The purified glycopeptides and glycans are analyzed by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry (MALDI−FTICR−MS), allowing the assignment of site-specific glycosylation and elucidation of glycan heterogeneity at individual sites.

To aid in the interpretation of this datum, we developed a unique and powerful software platform GlycoX that makes use of accurate masses belonging to the glycopeptides and glycans with the known protein sequence for assigning of N-linked and O-linked glycosylation with the additional determination of microheterogeneity. Although several tools such as GlycoMod,[19] FindMod,[20] and Glypeps[21] are available for the determination of glycosylation sites from mass spectra, glycopeptides derived from specific proteases are required to predict the glycosylation sites. Recently, other groups have also utilized the approach using nonspecific or reduced-specificity proteolysis for glycoprotein analysis;[22−26] however, there are no computer programs specifically developed to interpret automatic mass spectra obtained from nonspecific proteases such as Pronase E and protease K.

In this paper, we describe a new computer program (GlycoX) for determining site-specific glycosylation with microheterogeneity. The software described was developed in MATLAB and requires the entry of mass spectra as ASCII files. It takes advantage of the accurate mass capability of FTICR−MS. In addition, this program can be used to determine the glycosylation sites without the glycan information. GlycoX also includes an oligosaccharide calculator that accounts for multiple alkali adduction to acidic oligosaccharides. To validate the program, several glycoproteins were characterized. These glycoproteins range from the simple, with one site of glycosylation, to the more complex, with at least three sites of glycosylation. It is also shown that GlycoX can be used to determine oligosaccharide composition of released glycans. An isotope filter is included to determine the monoisotopic quasimolecular ions in mixtures of glycopeptides or glycans readily.

Experimental Section

General Experimental. 1. Pronase Digestion. Pronase digestion has been described in great detail in an earlier publication.[9] Briefly, Pronase E (~10 U) was added to the glycoprotein (~10 nmol) and incubated at 37 °C for 36−48 h. The reaction mixture was boiled to deactivate the enzyme. The digested glycoprotein was desalted and concentrated on a porous graphitic carbon (PGC) cartridge.

2. Release of N-Linked Glycans by PNGase F. The glycoprotein (~10 nmol) was dissolved in ammonium bicarbonate buffer solution (100 mM, pH 7.6). PNGase F (1 μL, ~10 U) was added, and the solution was incubated at 37 °C in a water bath for 17 h. The digestion mixture was boiled for 3 min to deactivate the enzyme. The digestion mixture was applied to PGC−SPE to purify the released glycans.

3. Release of O-Linked Glycans by Reductive β-Elimination. The glycoprotein was subjected to the reductive β-elimination. Alkaline borohydride solution (200 μL, mixture of 1.0 M sodium borohydride and 0.1 M sodium hydroxide) was added to 200 μg of glycoprotein. The mixture was incubated at 42 °C for 24 h. After the reaction period, a 1.0 M hydrochloric acid solution was slowly added with the reaction mixture in an ice bath to destroy excess sodium borohydride. The released glycans were purified by SPE employing PGC.

4. Mass Spectrometry. Mass spectra were recorded on an external source HiResMALDI (IonSpec corporation, Irvine, CA) equipped with a 7.0 T magnet and a pulsed YAG laser (355 nm) for ionization. 2,5-Dihydroxy benzoic acid (DHB) was used as a matrix (5 mg/100 μL in ethanol). A saturated solution of NaCl (or in some case KCl) in methanol was used as a dopant for the identification of the quasimolecular ion for either glycopeptides or oligosaccharides. The glycopeptide/oligosaccharide solution (1 μL) was applied to the MALDI probe followed by matrix solution (1 μL). The sample was dried under a stream of air prior to mass spectrometric analysis.

5. Operating System. The GlycoX program was written in MATLAB. We have used MATLAB 6 on several Windows platforms including 98, 2000, and XP.

Subroutines in GlycoX. The ***ReadSpectrum (specfile)*** subroutine reads spectral data from the file *specfile*. The user is asked to put quotes around the name of the file. The format of the specfile is ASCII, and the file contains two columns with masses and the intensities, for example,

```
1500.0010, 72.1513
1500.0084, 66.4353
1500.0158, 60.8504
1500.0233, 55.7041
1500.0307, 51.2539.
```

The output is a spectrum in GlycoX format—a matrix with two columns and many rows. The first column contains masses that are evenly spaced with separation delta (default value 0.01 Da). The second column has the intensities for the evenly spaced masses based on spline interpolation from the input data.

The ***MonoGly(spectrum)*** and ***MonoPep(spectrum)*** subroutines perform the isotope filter by taking the raw spectra and returning the estimate monoisotopic spectra in GlycoX format. The isotopic abundances of glycans and glycopeptides are slightly different and are accounted for in the respective subroutines (see next section).

The ***SolveGlySpec(spectrum, modename, PercentMax, errtol)*** performs the preprocessing by converting the raw spectrum and produces the resulting spectrum in the GlycoX format. ***SolveGlySpec*** runs ***MonoGly*** internally and should not be used on a monoisotopic spectrum. The input *modename* = ‘n+’, ‘o+’, ‘n-’, or ‘o-’ specifies the types of oligosaccharides with ‘n’ and ‘o’ representing N-linked and O-linked oligosaccharides, respectively, and ‘+’ and ‘-’ representing the positive and negative modes in MS. *PercentMax* is a threshold for identifying peaks. No peaks will be identified with height less than this percentage of the highest peak. *errtol* is the error allowed in parts per million of the mass accuracy (ppm). This routine also runs ***GetPeakInfo*** and ***SolveGlyPeaks*** (see below) and displays predicted glycans (oligosaccharide composition).

The ***GetPeakInfo(xy, PercentMax)*** subroutine shows the peak information for any of the GlycoX format spectra. The *xy* input is a spectrum in GlycoX format (isotope-filtered or unfiltered). The input *PercentMax* sets the peak threshold. The subroutine returns *PeakInfo*, a matrix with two columns for each mass peak with mass and corresponding intensity.

The subroutine ***SolveGlyPeaks (PeakInfo, modename, errtol)*** determines the glycan composition based on the mass from exact masses of monosaccharide residues. *PeakInfo* is the file returned by ***GetPeakInfo***. The files *modename* and *errtol* are the same as those used in ***SolveGlySpec***. The subroutine prints an output table of possible solutions for the relevant peaks.

The subroutine ***SolveGlyFile (specfile, modename, PercentMax, errtol)*** solves the glycan composition as ***SolveGlyPeaks***; however, the input can be the raw ASCII data (unfiltered). The output is the same as ***SolveGlyPeaks***.

The subroutine ***SolveGP (GPInfo, GPModeName, GlyInfo, GlyModeName, SeqFile, errtol, MaxPepLen, glylist, GlyMat)*** determines the glycopeptides and the oligosaccharide heterogeneity. The subroutine prints output for masses and intensities given by *GPInfo*, a matrix just like *PeakInfo* in ***GetPeakInfo*** or ***SolveGlyPeaks***, but for the glycopeptide mass values and intensities. *GPModeName* is 'n+', 'o+', 'n-', or 'o-', depending on which mode was used for the glycopeptide measurements. *GlyInfo* is the matrix of glycan mass values and intensities. *GlyModeName* is also 'n+', 'o+', 'n-', or 'o-' for the glycan measurements. *SeqFile* is a protein sequence file in FASTA format. *errtol* is again the error allowed, in parts per million. *MaxPepLen* sets the longest peptide length that GlycoX will consider. *glylist* is a list of the names of the glycans. The default is "glycans" from the routine StartGlycoX. *GlyMat* is a matrix with three rows: glycan masses on top, minimum counts below, and maximum counts below that. In StartGlycoX, we define NglyMat (N-linked glycan) and OGlyMat (O-linked glycan), but others can be created by the user.

Results and Discussion

The GlycoX program has three main functions, a 'glycosylation site search' for the determination of glycosylation sites with oligosaccharide heterogeneity in glycoproteins; an 'oligosaccharide calculator' for assigning oligosaccharide compositions for N-linked, O-linked, and chemically modified oligosaccharides; and an 'auto-isotope filter' for selecting the monoisotopic peaks from the mass spectra of glycopeptides, peptides, and glycans.

1. Isotope Filter. An isotope filter was developed to simplify the mass spectra for analysis. Simplification is needed because isotopes will produce several peaks in the mass spectrum at masses m, $m + 1$, $m + 2$, and so forth. If all peaks are used, there can be interference and confusion between a higher isotopic peak of one molecule and the lower peaks of another. To have an unambiguous peak associated with each compound, we would prefer to work with a *monoisotopic spectrum*. For a given compound, the monoisotopic mass is the mass of the isotopic peak whose elemental composition is composed of the most abundant isotopes of those elements. Monoisotopic masses are used more often than average masses in mass spectrometry to relate peaks with their compositions. The GlycoX method uses the monoisotopic masses only and would give incorrect results if applied to the larger isotopic peaks. There is a need for a robust software to simplify the MS data because complex isotopic patterns make peak interpretation difficult.[27] Recently, a new algorithm, *Iosconv*, for deisotoping electrospray and MALDI mass spectra was developed by Hoffmann and co-workers.[28,29] This algorithm was successfully used for peptides. However, there is currently no readily accessible software to filter the monoisotopic masses of other biomolecules such as glycans and glycopeptides. Different molecule classes need slightly different parameters for more efficient deisotoping algorithms (see below). Thus, *isotope filter* is a unique and efficient software for simplifying the MALDI–MS spectra of oligosaccharides, glycopeptides, and peptides.

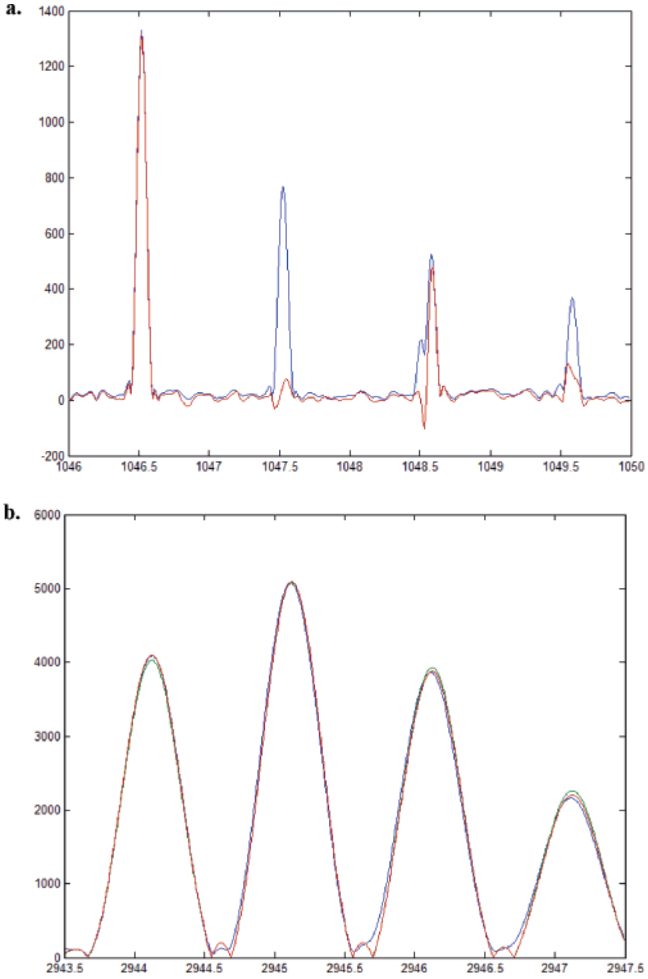


Figure 1. (a) Isotope filter helps to disentangle overlapping peaks in angiotensin fragment (1046) and Pro9-Arg (1048) (b) Estimation of isotope ratios for methylated trisialylated triantennary glycan. Blue line shows the MS spectrum; green line is synthesized using ratios based on $a = 1.27$, $b = 0.180$ (estimates for mass = 2944.05 amu); red line is synthesized using $a = 1.27$, $b = 0.176$ (true values for composition C112–H184–O80–N8).

An isotope filter can be used to deconvolute two overlapping quasimolecular ion signals. For example, shown in Figure 1a are two ionic species that differ by two mass units. The isotopic peaks of the lower mass interfere with the less abundant monoisotopic peaks of the larger mass. Performing the filter eliminates the contribution of the lower mass to the higher, although with some distortion, showing two distinct peaks in the region.

An algorithm that we call the *isotope filter* estimates the monoisotopic spectrum that would correspond to the measured spectrum with its isotopic peaks. For a given molecule,

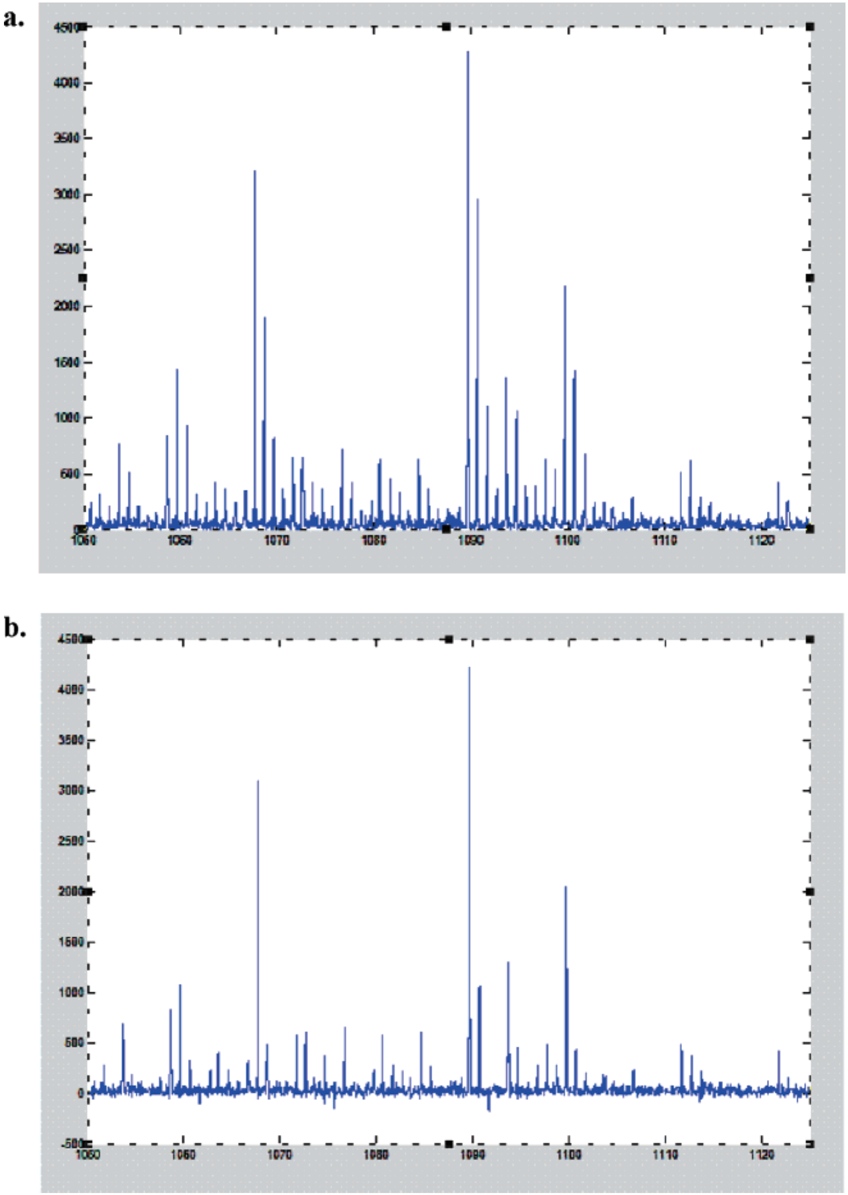


Figure 2. MALDI–MS spectrum of oligosaccharides obtained from human tear (a) before and (b) after the isotope filter. All peaks obtained from the isotope filter correspond to monoisotopic masses.

there are rules for the relative abundances of isotopomer,

$$y_{m+1} = ay_m$$

$$y_{m+2} = \left(\frac{a^2}{2} + b\right)y_m$$

$$y_{m+3} = \left(\frac{a^3}{6} + ab\right)y_m$$

$$y_{m+4} = \left(\frac{a^4}{24} + \frac{a^2}{2}b + \frac{b^2}{2}\right)y_m$$

where,

$$a = 0.00011n_{\mathrm{H}} + 0.011n_{\mathrm{C}} + 0.0036n_{\mathrm{N}}$$

is the expected number of single-isotope atoms per molecule (^{13}C, ^{2}H, and ^{15}N), and

$$b = 0.0022n_{\mathrm{O}} + 0.045n_{\mathrm{S}}$$

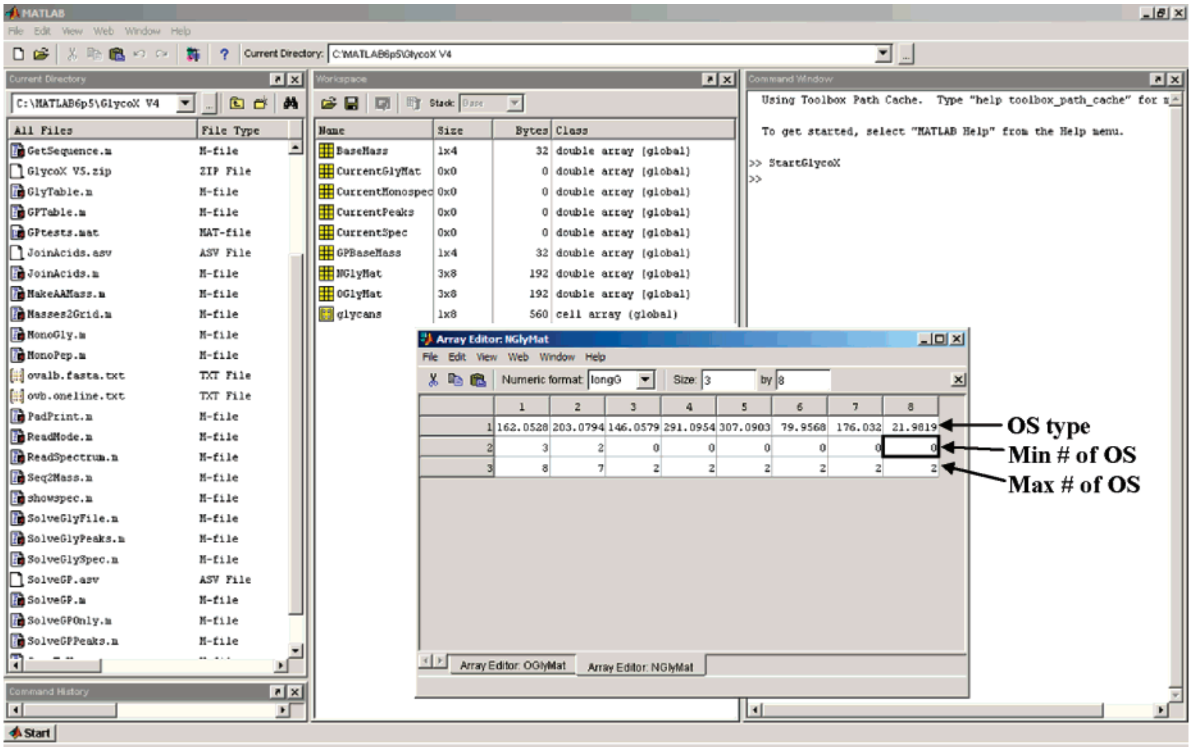
is the expected number of double-isotope atoms per molecule (^{18}O and ^{34}S). Figure 1b shows the spectrum for a glycan with elemental composition $C_{112}H_{184}O_{80}N_8$. Fitted to the spectrum are predictions based on the rules for isotope ratios. The blue trace is the experimental plot. The green trace is based on the average composition of glycans of a mass 2944. The values for *a* and *b* are determined from an "average glycan" residue ("averagose") analogous to "averagine" for peptides.[30] The red trace is the synthesized spectrum with *a* and *b* calculated from the known glycan composition.

This can be written as a matrix relationship,

$$\mathbf{y} = \exp(a\mathbf{S} + b\mathbf{S}^2)\mathbf{y}^{\mathrm{M}}$$

for

$$\mathbf{y} = \begin{bmatrix} y_m \\ y_{m+1} \\ y_{m+2} \\ \vdots \end{bmatrix}, \quad \mathbf{y}^{\mathbf{M}} = \begin{bmatrix} y_m \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$



C:\MATLAB6p5\GlycoX V4

All Files	File Type
GetSequence.m	M-file
GlycoX V5.zip	ZIP File
GlyTable.m	M-file
GPTable.m	M-file
GPtests.mat	MAT-file
JoinAcids.asv	ASV File
JoinAcids.m	M-file
MakeAAMass.m	M-file
Masses2Grid.m	M-file
MonoGly.m	M-file
MonoPep.m	M-file
ovalb.fasta.txt	TXT File
ovb.oneline.txt	TXT File
PadPrint.m	M-file
ReadMode.m	M-file
ReadSpectrum.m	M-file
Seq2Mass.m	M-file
showspec.m	M-file
SolveGlyFile.m	M-file
SolveGlyPeaks.m	M-file
SolveGlySpec.m	M-file
SolveGP.asv	ASV File
SolveGP.m	M-file
SolveGPOnly.m	M-file
SolveGPPeaks.m	M-file

Name	Size	Bytes	Class
BaseMass	1x4	32	double array (global)
CurrentGlyMat	0x0	0	double array (global)
CurrentMonospec	0x0	0	double array (global)
CurrentPeaks	0x0	0	double array (global)
CurrentSpec	0x0	0	double array (global)
GPBaseMass	1x4	32	double array (global)
NGlyMat	3x8	192	double array (global)
OGlyMat	3x8	192	double array (global)
glycans	1x8	560	cell array (global)

```
 Using Toolbox Path Cache.  Type "help toolbox_path_cache" for m

 To get started, select "MATLAB Help" from the Help menu.

>> StartGlycoX
>>
```

Numeric format: longG Size: 3 by 8

	1	2	3	4	5	6	7	8
1	162.0528	203.0794	146.0579	291.0954	307.0903	79.9568	176.032	21.9819
2	3	2	0	0	0	0	0	0
3	8	7	2	2	2	2	2	2

Figure 3. Editor used for oligosaccharide calculator in MATLAB.

and **S** is the shift matrix

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & 0 & \ldots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix}$$

that is, multiplying by **S** just shifts the whole spectrum up by one mass unit. The example shown here is for the simplest case: a spectrum of integer masses. The matrix **S** must always shift the spectrum by one whole mass unit; therefore in a case with precision for 100 data points per amu, the matrix would have ones along the 100th subdiagonal instead of the first.

For large enough masses, we can use an approximation

$$\boldsymbol{a} \approx \alpha \boldsymbol{m}, \quad \text{and } \boldsymbol{b} \approx \beta \boldsymbol{m}$$

where α and β are the average numbers of single and double isotopes per amu. Combining the relations for different molecules leads to

$$\mathbf{y} \approx \exp(\alpha \mathbf{m}\mathbf{S} + \beta \mathbf{m}\mathbf{S}^2)\mathbf{y}^{\mathrm{M}}$$

where **m** is the diagonal matrix that multiplies each intensity by the mass, and **S** is the "shift" matrix that slides the spectrum down by one mass unit.

This means that

$$\mathbf{y}^{\mathrm{M}} \approx \exp(-\alpha \mathbf{m}\mathbf{S} - \beta \mathbf{m}\mathbf{S}^2)\mathbf{y}$$

The GlycoX package has a very efficient way to compute this expression given **y**.

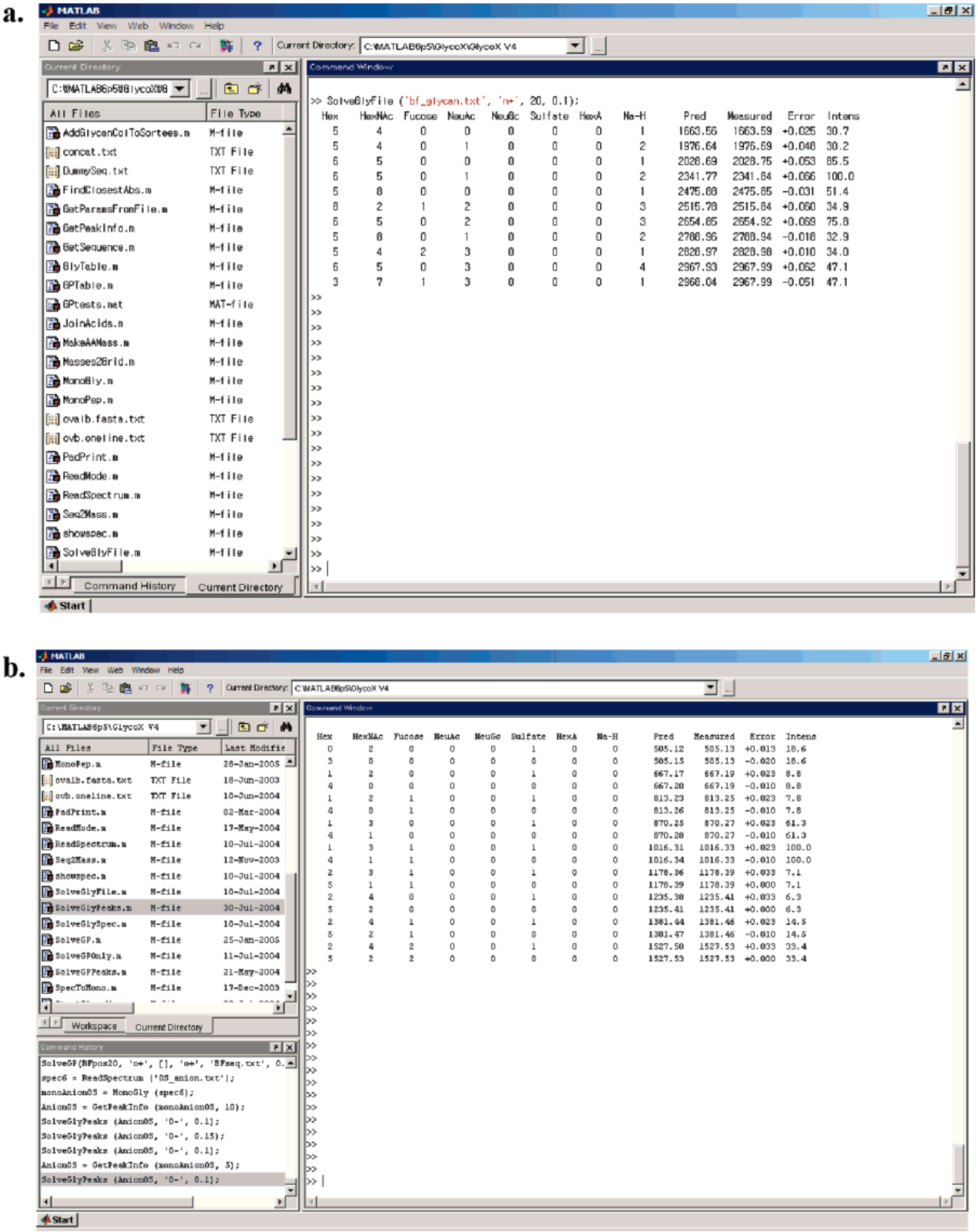
For the molecule classes we have been working with, different parameters for the approximation of the peptides, glycans, and glycopeptides are used to give a more accurate output. We have found that the following coefficients provide a good approximation:

molecule type	*A*	*B*
peptides	5.2×10^{-4}	3.0×10^{-5}
oligosaccharides	4.3×10^{-4}	6.1×10^{-5}
glycopeptides	4.6×10^{-4}	5.0×10^{-5}

Before-and-after pictures show the effect of using the filter. To illustrate the utility of the isotope filter, a complicated mixture of oligosaccharides from human tear was analyzed. Figure 2a shows the MALDI−MS spectrum before the "isotope filter" with several overlapping peaks. After the isotope filter (Figure 2b), the spectrum still contains many peaks, but is considerably reduced in complexity.

2. Oligosaccharide Calculator. GlycoX can be used to calculate the possible glycan compositions from experimentally determined masses. This is a stand-alone feature that can be used on the mass spectra of oligosaccharide mixtures. The user has the choice to define oligosaccharide type (N-linked or O-linked glycan), MS detection mode (positive or negative mode), and the minimum and maximum number of monosaccharides (Figure 3). Furthermore, all glycans types, that is, reduced glycans (alditol), unreduced glycans (aldehyde), and chemically derivatized glycans, can be calculated in GlycoX. A particular advantage of this program is the ability to easily input the sugar's modification and the corresponding mass. To get biologically relevant oligosaccharide compositions, several restrictions were placed on the oligosaccharide calculator. First, the number of fucose residues should be less than or equal to the sum of the number of Hex plus HexNAc residues. Second, the N-linked glycans should have 2HexNAc and 3 Man corresponding to the core structure.

Oligosaccharides readily coordinate with Na^+. Acidic oligosaccharides exchange acidic protons with Na^+ to yield



```
C:\MATLAB6p5\GlycoX\G
All Files                  File Type
AddGlycanColToSortees.m    M-file
concat.txt                 TXT File
DummySeq.txt               TXT File
FindClosestAbs.m           M-file
GetParamsFromFile.m        M-file
GetPeakInfo.m              M-file
GetSequence.m              M-file
GlyTable.m                 M-file
GPTable.m                  M-file
GPtests.mat                MAT-file
JoinAcids.m                M-file
MakeAAMass.m               M-file
Masses2Grid.m              M-file
MonoGly.m                  M-file
MonoPep.m                  M-file
ovalb.fasta.txt            TXT File
ovb.oneline.txt            TXT File
PadPrint.m                 M-file
ReadMode.m                 M-file
ReadSpectrum.m             M-file
Seq2Mass.m                 M-file
showspec.m                 M-file
SolveGlyFile.m             M-file
```

```
>> SolveGlyFile ('bf_glycan.txt', 'n+', 20, 0.1);
  Hex    HexNAc  Fucose  NeuAc   NeuGc  Sulfate  HexA    Na-H      Pred     Measured   Error   Intens
    5       4       0       0       0       0       0       1     1663.56   1663.59   +0.025   30.7
    5       4       0       1       0       0       0       2     1976.64   1976.69   +0.048   30.2
    6       5       0       0       0       0       0       1     2028.69   2028.75   +0.053   85.5
    6       5       0       1       0       0       0       2     2341.77   2341.84   +0.066   100.0
    5       8       0       0       0       0       0       1     2475.88   2475.85   -0.031   51.4
    8       2       1       2       0       0       0       3     2515.78   2515.84   +0.060   34.9
    6       5       0       2       0       0       0       3     2654.85   2654.92   +0.069   75.8
    5       8       0       1       0       0       0       2     2788.96   2788.94   -0.018   32.9
    5       4       2       3       0       0       0       1     2828.97   2828.98   +0.010   34.0
    6       5       0       3       0       0       0       4     2967.93   2967.99   +0.062   47.1
    3       7       1       3       0       0       0       1     2968.04   2967.99   -0.051   47.1
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>> |
```

```
C:\MATLAB6p5\GlycoX V4
All Files           File Type   Last Modifie
MonoPep.m           M-file      28-Jan-2005
ovalb.fasta.txt     TXT File    18-Jun-2003
ovb.oneline.txt     TXT File    10-Jun-2004
PadPrint.m          M-file      02-Mar-2004
ReadMode.m          M-file      17-May-2004
ReadSpectrum.m      M-file      10-Jul-2004
Seq2Mass.m          M-file      12-Nov-2003
showspec.m          M-file      10-Jul-2004
SolveGlyFile.m      M-file      10-Jul-2004
SolveGlyPeaks.m     M-file      30-Jul-2004
SolveGlySpec.m      M-file      10-Jul-2004
SolveGP.m           M-file      25-Jan-2005
SolveGPOnly.m       M-file      11-Jul-2004
SolveGPPeaks.m      M-file      21-May-2004
SpecToMono.m        M-file      17-Dec-2003
```

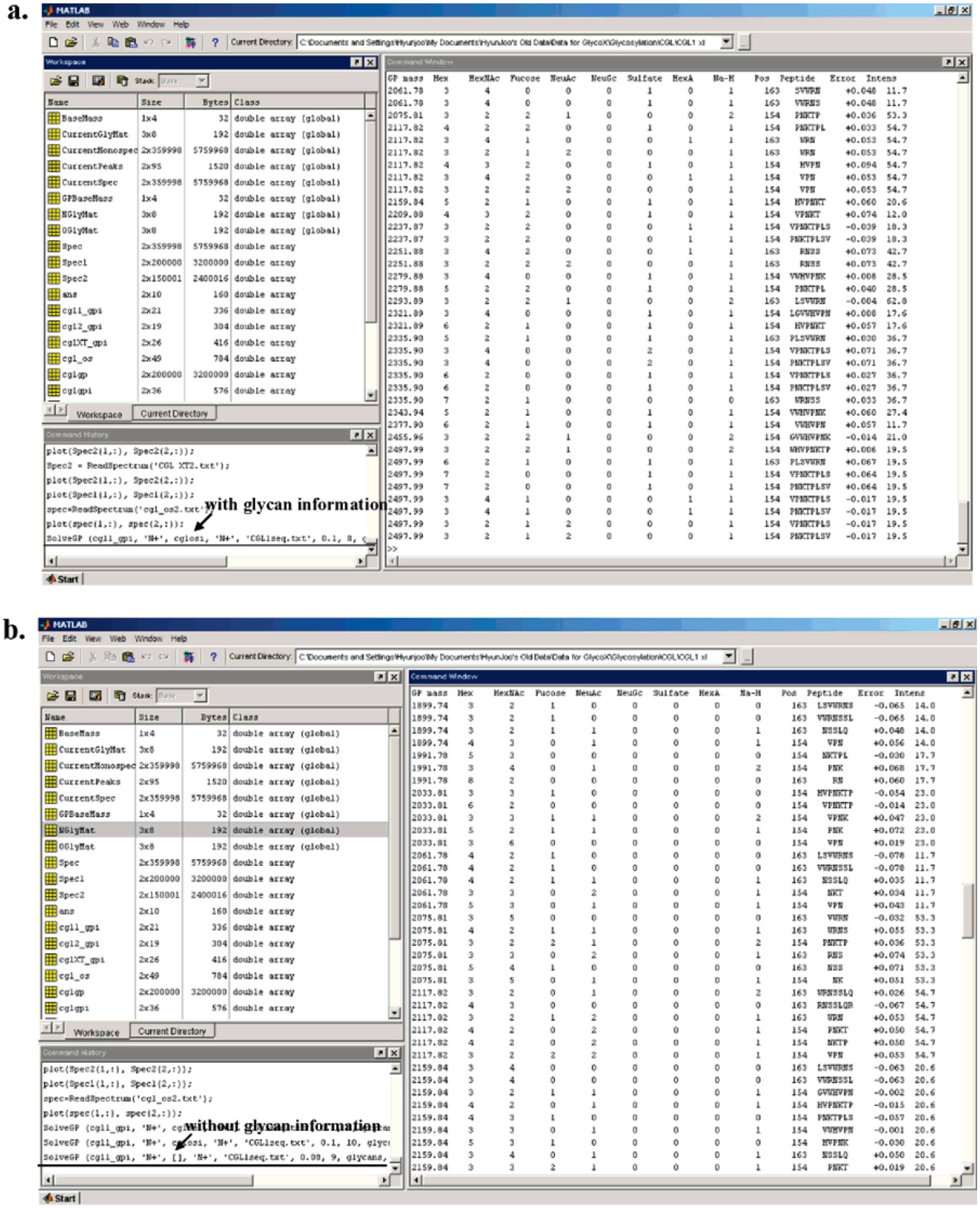
```
SolveGP(BFpos20, 'o+', [], 'o+', 'BFseq.txt', 0.
spec6 = ReadSpectrum ('OS_anion.txt');
monoAnionOS = MonoGly (spec6);
AnionOS = GetPeakInfo (monoAnionOS, 10);
SolveGlyPeaks (AnionOS, 'O-', 0.1);
SolveGlyPeaks (AnionOS, 'O-', 0.15);
SolveGlyPeaks (AnionOS, 'O-', 0.1);
AnionOS = GetPeakInfo (monoAnionOS, 5);
SolveGlyPeaks (AnionOS, 'O-', 0.1);
```

```
  Hex    HexNAc  Fucose  NeuAc   NeuGc  Sulfate  HexA    Na-H      Pred     Measured   Error   Intens
    0       2       0       0       0       1       0       0      505.12    505.13   +0.013   18.6
    3       0       0       0       0       0       0       0      505.15    505.13   -0.020   18.6
    1       2       0       0       0       1       0       0      667.17    667.19   +0.023   8.8
    4       0       0       0       0       0       0       0      667.20    667.19   -0.010   8.8
    1       2       1       0       0       1       0       0      813.23    813.25   +0.023   7.8
    4       0       1       0       0       0       0       0      813.26    813.25   -0.010   7.8
    1       3       0       0       0       1       0       0      870.25    870.27   +0.023   61.3
    4       1       0       0       0       0       0       0      870.28    870.27   -0.010   61.3
    1       3       1       0       0       1       0       0     1016.31   1016.33   +0.023   100.0
    4       1       1       0       0       0       0       0     1016.34   1016.33   -0.010   100.0
    2       3       1       0       0       1       0       0     1178.36   1178.39   +0.033   7.1
    5       1       1       0       0       0       0       0     1178.39   1178.39   +0.000   7.1
    2       4       0       0       0       1       0       0     1235.38   1235.41   +0.033   6.3
    5       2       0       0       0       0       0       0     1235.41   1235.41   +0.000   6.3
    2       4       1       0       0       1       0       0     1381.44   1381.46   +0.023   14.5
    5       2       1       0       0       0       0       0     1381.47   1381.46   -0.010   14.5
    2       4       2       0       0       1       0       0     1527.50   1527.53   +0.033   33.4
    5       2       2       0       0       0       0       0     1527.53   1527.53   +0.000   33.4
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>>
>> |
```

Figure 4. An output page of (a) N-linked glycans of bovine fetuin analyzed in the positive mode and (b) O-linked glycans of *X. tropicalis* eggs analyzed in the negative mode.

quasimolecular ions with compositions [M − *n*H + (*n*+1)Na]$^+$, where *n* = no. of acidic residues. We often observe multisodium adduct peaks in the mass spectra of acidic oligosaccharides. The Na$^+$-exchanged peaks yield multiple quasimolecular ions that often complicate MS spectrum. A few oligosaccharide calculator programs have recently been described, but there are no programs that account for multisodium-adducted peaks. GlycoX is unique because it shows all common possible sodium combinations with the oligosaccharides in order to calculate the acidic glycan composition.

N-linked glycans of bovine fetuin are highly glycosylated. To test the program, the oligosaccharides were released by PNGase F and analyzed by MALDI−MS in the positive ion mode. The resulting mass spectrum was entered into GlycoX to calculate the glycan compositions. Figure 4a shows the output page for the N-linked glycans of bovine fetuin. The results correctly determine the bi- and triantennary glycan compositions. All observed quasimolecular ions have [M − *n*H + (*n*+1)Na]$^+$, where *n* represents the number of sialic acid residues. Peaks corresponding to [M − H + 2Na]$^+$ of monosialylated bianten-


a.
MATLAB
File Edit View Web Window Help
Current Directory: C:\Documents and Settings\Hyunjoo\My Documents\HyunJoo's Old Data\Data for GlycoX\Glycosylation\CGL\CGL1 xl
Workspace
Stack: Base
Workspace
Current Directory
Command History
Command Window
with glycan information
Start
b.
without glycan information

```
Name              Size       Bytes    Class
BaseMass          1x4        32       double array (global)
CurrentGlyMat     3x8        192      double array (global)
CurrentMonospec   2x359998   5759968  double array (global)
CurrentPeaks      2x95       1520     double array (global)
CurrentSpec       2x359998   5759968  double array (global)
GPBaseMass        1x4        32       double array (global)
NGlyMat           3x8        192      double array (global)
OGlyMat           3x8        192      double array (global)
Spec              2x359998   5759968  double array
Spec1             2x200000   3200000  double array
Spec2             2x150001   2400016  double array
ans               2x10       160      double array
cgl1_gpi          2x21       336      double array
cgl2_gpi          2x19       304      double array
cglXT_gpi         2x26       416      double array
cgl_os            2x49       784      double array
cglgp             2x200000   3200000  double array
cglgpi            2x36       576      double array
```

```
plot(Spec2(1,:), Spec2(2,:));
Spec2 = ReadSpectrum('CGL XT2.txt');
plot(Spec2(1,:), Spec2(2,:));
plot(Spec1(1,:), Spec1(2,:));
spec=ReadSpectrum('cgl_os2.txt');
plot(spec(1,:), spec(2,:));
SolveGP (cgl1_gpi, 'N+', cglos1, 'N+', 'CGL1seq.txt', 0.1, 8, c
```

```
GP mass  Hex  HexNAc  Fucose  NeuAc  NeuGc  Sulfate  HexA  Na-H   Pos  Peptide    Error   Intens
2061.78   3    4       0       0      0      1        0     1      163  SVWRN      +0.048  11.7
2061.78   3    4       0       0      0      1        0     1      163  VWRNS      +0.048  11.7
2075.81   3    2       2       1      0      0        0     2      154  PNKTP      +0.036  53.3
2117.82   4    2       2       0      0      1        0     1      154  PNKTPL     +0.033  54.7
2117.82   3    4       1       0      0      0        1     1      163  WRN        +0.053  54.7
2117.82   3    2       1       2      0      0        0     1      163  WRN        +0.053  54.7
2117.82   4    3       2       0      0      1        0     1      154  HVPN       +0.094  54.7
2117.82   3    4       2       0      0      0        1     1      154  VPN        +0.053  54.7
2117.82   3    2       2       2      0      0        0     1      154  VPN        +0.053  54.7
2159.84   5    2       1       0      0      1        0     1      154  HVPNKT     +0.060  20.6
2209.88   4    3       2       0      0      1        0     1      154  VPNKT      +0.074  12.0
2237.87   3    2       2       0      0      0        1     1      154  VPNKTPLS   -0.039  18.3
2237.87   3    2       2       0      0      0        1     1      154  PNKTPLSV   -0.039  18.3
2251.88   3    4       2       0      0      0        1     1      163  RNSS       +0.073  42.7
2251.88   3    2       2       2      0      0        0     1      163  RNSS       +0.073  42.7
2279.88   3    4       0       0      0      1        0     1      154  VWHVPNK    +0.008  28.5
2279.88   5    2       2       0      0      1        0     1      154  PNKTPL     +0.040  28.5
2293.89   3    2       2       1      0      0        0     2      163  LSVWRN     -0.004  62.8
2321.89   3    4       0       0      0      1        0     1      154  LGVWHVPN   +0.008  17.6
2321.89   6    2       1       0      0      1        0     1      154  HVPNKT     +0.057  17.6
2335.90   5    2       1       0      0      1        0     1      163  PLSVWRN    +0.030  36.7
2335.90   3    4       0       0      0      2        0     1      154  VPNKTPLS   +0.071  36.7
2335.90   3    4       0       0      0      2        0     1      154  PNKTPLSV   +0.071  36.7
2335.90   6    2       0       0      0      1        0     1      154  VPNKTPLS   +0.027  36.7
2335.90   6    2       0       0      0      1        0     1      154  PNKTPLSV   +0.027  36.7
2335.90   7    2       1       0      0      0        0     0      163  WRNSS      +0.033  36.7
2343.94   5    2       1       0      0      1        0     1      154  VWHVPNK    +0.060  27.4
2377.90   6    2       1       0      0      1        0     1      154  VWHVPN     +0.057  11.7
2455.96   3    2       2       1      0      0        0     2      154  GVWHVPNK   -0.014  21.0
2497.99   3    2       2       1      0      0        0     2      154  WHVPNKTP   +0.006  19.5
2497.99   6    2       1       0      0      1        0     1      163  PLSVWRN    +0.067  19.5
2497.99   7    2       0       0      0      1        0     1      154  VPNKTPLS   +0.064  19.5
2497.99   7    2       0       0      0      1        0     1      154  PNKTPLSV   +0.064  19.5
2497.99   3    4       1       0      0      0        1     1      154  VPNKTPLS   -0.017  19.5
2497.99   3    4       1       0      0      0        1     1      154  PNKTPLSV   -0.017  19.5
2497.99   3    2       1       2      0      0        0     1      154  VPNKTPLS   -0.017  19.5
2497.99   3    2       1       2      0      0        0     1      154  PNKTPLSV   -0.017  19.5
>>
```

```
Name              Size       Bytes    Class
BaseMass          1x4        32       double array (global)
CurrentGlyMat     3x8        192      double array (global)
CurrentMonospec   2x359998   5759968  double array (global)
CurrentPeaks      2x95       1520     double array (global)
CurrentSpec       2x359998   5759968  double array (global)
GPBaseMass        1x4        32       double array (global)
NGlyMat           3x8        192      double array (global)
OGlyMat           3x8        192      double array (global)
Spec              2x359998   5759968  double array
Spec1             2x200000   3200000  double array
Spec2             2x150001   2400016  double array
ans               2x10       160      double array
cgl1_gpi          2x21       336      double array
cgl2_gpi          2x19       304      double array
cglXT_gpi         2x26       416      double array
cgl_os            2x49       784      double array
cglgp             2x200000   3200000  double array
cglgpi            2x36       576      double array
```

```
plot(Spec2(1,:), Spec2(2,:));
plot(Spec1(1,:), Spec1(2,:));
spec=ReadSpectrum('cgl_os2.txt');
plot(spec(1,:), spec(2,:));
SolveGP (cgl1_gpi, 'N+', cgl
SolveGP (cgl1_gpi, 'N+', cglosi, 'N+', 'CGL1seq.txt', 0.1, 10, glyc
SolveGP (cgl1_gpi, 'N+', [], 'N+', 'CGL1seq.txt', 0.08, 9, glycans,
```

```
GP mass  Hex  HexNAc  Fucose  NeuAc  NeuGc  Sulfate  HexA  Na-H   Pos  Peptide    Error   Intens
1899.74   3    2       1       0      0      0        0     0      163  LSVWRNS    -0.065  14.0
1899.74   3    2       1       0      0      0        0     0      163  VWRNSSL    -0.065  14.0
1899.74   3    2       1       1      0      0        0     1      163  NSSLQ      +0.048  14.0
1899.74   4    3       0       1      0      0        0     1      154  VPN        +0.056  14.0
1991.78   5    3       0       0      0      0        0     0      154  NKTPL      -0.030  17.7
1991.78   3    4       0       1      0      0        0     2      154  PNK        +0.068  17.7
1991.78   8    2       0       0      0      0        0     0      163  RN         +0.060  17.7
2033.81   3    3       1       0      0      0        0     0      154  HVPNKTP    -0.054  23.0
2033.81   6    2       0       0      0      0        0     0      154  VPNKTP     -0.014  23.0
2033.81   3    3       1       1      0      0        0     2      154  VPNK       +0.047  23.0
2033.81   5    2       1       1      0      0        0     1      154  PNK        +0.072  23.0
2033.81   3    6       0       0      0      0        0     0      154  VPN        +0.019  23.0
2061.78   4    2       1       0      0      0        0     0      163  LSVWRNS    -0.078  11.7
2061.78   4    2       1       0      0      0        0     0      163  VWRNSSL    -0.078  11.7
2061.78   4    2       1       1      0      0        0     1      163  NSSLQ      +0.035  11.7
2061.78   3    3       0       2      0      0        0     1      154  NKT        +0.034  11.7
2061.78   5    3       0       1      0      0        0     1      154  VPN        +0.043  11.7
2075.81   3    5       0       0      0      0        0     0      163  VWRN       -0.032  53.3
2075.81   4    2       1       1      0      0        0     1      163  WRNS       +0.055  53.3
2075.81   3    2       2       1      0      0        0     2      154  PNKTP      +0.036  53.3
2075.81   3    3       0       2      0      0        0     1      163  RNS        +0.074  53.3
2075.81   5    4       1       0      0      0        0     0      163  NSS        +0.071  53.3
2075.81   3    5       0       1      0      0        0     1      154  NK         +0.051  53.3
2117.82   3    2       0       1      0      0        0     2      163  WRNSSLQ    +0.026  54.7
2117.82   4    3       0       0      0      0        0     0      163  RNSSLQR    -0.067  54.7
2117.82   3    2       1       2      0      0        0     1      163  WRN        +0.053  54.7
2117.82   4    2       0       2      0      0        0     1      154  PNKT       +0.050  54.7
2117.82   4    2       0       2      0      0        0     1      154  NKTP       +0.050  54.7
2117.82   3    2       2       2      0      0        0     1      154  VPN        +0.053  54.7
2159.84   3    4       0       0      0      0        0     0      163  LSVWRNS    -0.063  20.6
2159.84   3    4       0       0      0      0        0     0      163  VWRNSSL    -0.063  20.6
2159.84   3    2       1       1      0      0        0     1      154  GVWHVPN    -0.002  20.6
2159.84   4    2       0       1      0      0        0     1      154  HVPNKTP    -0.015  20.6
2159.84   4    3       1       0      0      0        0     0      154  PNKTPLS    -0.057  20.6
2159.84   3    3       0       1      0      0        0     1      154  VWHVPN     -0.001  20.6
2159.84   5    3       1       0      0      0        0     0      154  HVPNK      -0.030  20.6
2159.84   3    4       0       1      0      0        0     1      163  NSSLQ      +0.050  20.6
2159.84   3    3       2       1      0      0        0     1      154  PNKT       +0.019  20.6
```

Figure 5. An output page of N-linked glycosylation sites of *XL* CGL1 using (a) glycan and glycopeptide masses together and (b) only glycopeptide masses in GlycoX.

nary and triantennary were observed in the mass spectrum at m/z 1976.69 and 2341.84, respectively. A disialylated triantennary composition were also found at m/z 2654.92 ($[M - 2H + 3Na]^+$). A trisialylated triantennary composition was observed at m/z 2967.99 ($[M - 3H + 4Na]^+$).

The composition of O-linked oligosaccharides can also be determined. In addition, the negative mode can also be analyzed as readily as the positive mode. The MS spectrum of an O-linked oligosaccharide was obtained in the negative ion mode and examined by GlycoX. The oligosaccharides were released from the egg jelly coats of *Xenopus tropicalis* by β-elimination. The output is shown in Figure 4b with all the oligosaccharides correctly determined. All of the peaks corresponded to the deprotonated species ($[M - H]^-$).

3. Determination of Glycosylation Sites. The glycoproteins were digested with a nonspecific protease and purified as described previously.[9] A portion of the glycoprotein sample was treated with PNGase F to release N-linked oligosaccharides. The digested glycopeptides and glycans were purified by PGC−SPE. The PGC fractions of the glycopeptides and the N-linked glycans were analyzed by MALDI−FTMS in both the positive and negative mode. The mass intensity (*M/I*) table, saved as

Table 1

GP mass	Hex	HexNAc	Fuc	NeuAc	no. Na	Pos	peptide	error
1991.78	3	2	0	0	1	154	VWHVPNKTP	0.08
2117.82	5	2	0	0	1	154	VWHVPNK	0.04
2159.84	5	2	0	0	1	154	LGVWHVPN	0.03
2167.85	4	2	0	0	1	154	WHVPNKTPL	0.07
2209.88	5	5	0	0	1	154	NKT	0.06
2237.87	4	5	0	1	1	154	NK	0.06
2279.88	6	2	0	0	1	154	VWHVPNK	0.04
2279.88	4	5	0	0	1	154	HVPNK	0.02
2321.89	6	2	0	0	1	154	LGVWHVPN	0.04
2321.89	3	5	1	0	1	154	WHVPN	0.00
2321.89	3	4	0	0	1	154	LGVWHVPN	0.01
2335.90	3	4	0	0	1	154	(V)PNKTPLS(V)	0.07
2335.90	6	2	0	0	1	154	(V)PNKTPLS(V)	0.03
2343.94	9	2	0	0	1	154	VPNK	0.08
2377.90	6	2	1	0	1	154	VWHVPN	0.06
2497.99	7	2	0	0	1	154	(V)PNKTPLS(V)	0.06
2497.99	3	4	1	0	1	154	(V)PNKTPLS(V)	0.02
2520.02	4	4	0	0	1	154	GVWHVPNKT	0.02
1899.74	5	2	0	0	1	163	(S)VWRN(S)	0.02
2061.78	6	2	0	0	1	163	(S)VWRN(S)	0.00
2061.78	4	5	0	0	1	163	RNS	0.02
2061.78	3	4	0	0	1	163	(S)VWRN(S)	0.05
2075.81	3	3	0	0	1	163	PLSVWRNS	0.05
2251.88	4	5	1	0	1	163	NSSL	0.05
2293.89	5	4	0	0	1	163	WRNSS	0.05
2455.96	3	4	1	0	1	163	VWRNSSLQ	0.05

ASCII files, and the corresponding glycoprotein sequences from the Swiss-Prot/TrEMBL database saved in FASTA format (text file) were entered into GlycoX.

To determine the glycosylation sites and the accompanying glycans, the isotope filtered masses were then enumerated for combinations of glycan and peptides. Peptides of variable lengths, up to the user-specified lengths, were examined. The peptide composition was determined from the calculated mass by comparing all possible sequences from the monopeptide to the user-specified maximum peptide sequence. For glycan masses, the program generates possible combinations that can fit in the given mass. Alternatively, masses from the glycan profile spectra can be used. For complicated glycosylation, the measured glycan profile should be used.

Several glycoproteins were used to determine the efficacy of GlycoX. These glycoproteins range from the simple, with one site of glycosylation, to the more complex, with multiple sites of glycosylation. For comparison, the glycans and glycosylation sites were also determined manually using the same data. The results obtained using the GlycoX were identical to those determined manually. In addition, GlycoX found several more glycopeptides that were missed by the manual treatment. Model glycoproteins with known glycosylation sites such as ribonuclease B (with one glycosylation site) and chicken ovalbumin (with two sites and one occupied) were tested and yielded the correct results (data not shown).[31-33]

For examples of more complicated glycoproteins (with unknown glycosylation sites), cortical granule lectins (CGL1 and CGL2) from *Xenopus laevis* (*XL*) eggs and CGL from *X. tropicalis* (*XT*) eggs were examined. An output page of N-linked glycosylation site of *XL* CGL1 by GlycoX is shown in Figure 5a. Glycopeptides were sorted by increasing glycopeptide masses.

GlycoX can be used to predict the glycosylation site without the glycan information. Figure 5b shows the output table of determined glycosylation sites using only glycopeptides masses without the experimentally determined glycan profile.

Figure 5a shows the partial output for of CGL1 from *X. laevis* determined with the experimental glycan profile included in the input. The output lists the glycopeptides by increasing masses. This allows the listing of multiple hits for the same mass to be examined together. Table 1 provides the same information with the glycopeptides sorted according to the glycosylation sites. The compositions are consistent with the known glycans from the glycan profile. Each glycan composition is represented by several glycopeptides. For example, the $GlcNAc_2Man_6$ (oligo-Man6) oligosaccharide is represented by six peptides. Figure 5b shows the output of GlycoX for the same glycoprotein determined without the experimental glycan profile. In the output, there is an entry for a Man12, which is not present in CGL1. However, there is only a single entry for this glycan, which makes it unlikely to be present in the glycoprotein. Single entries are a good indication that the glycopeptide is a false-positive hit and can be excluded. These results illustrate that the glycan profile is useful but not absolutely necessary for glycoproteins with even two sites of glycosylation.

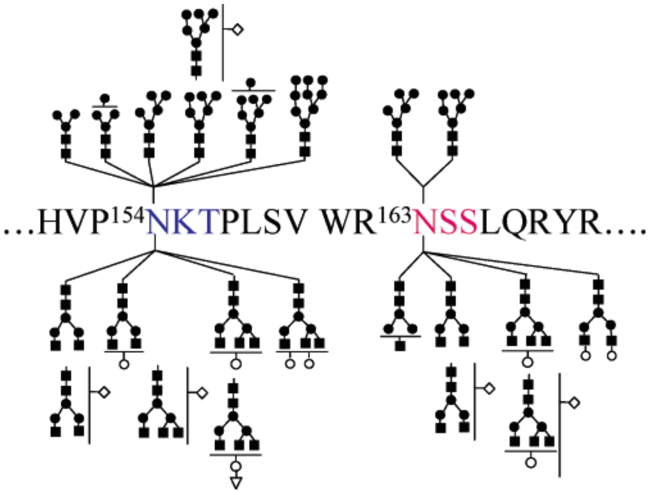


Figure 6. Distribution of the glycans on the two glycosyaltion sites.

The high mass accuracy is another important factor in eliminating false-positive results, especially in the determination of glycosylation sites without the glycan information. To illustrate the effects of mass error tolerance, the glycosylation sites of CGL2 of *X. laevis* using only glycopeptides masses were examined by GlycoX. As expected, one obtains more glycopeptide hits with the lower tolerance. For this example, 1836 hits were obtained with a mass tolerance of 1.0 Da, 502 with 0.1 Da, and 181 with 0.05 Da. The number of hits decreases further to 42 with a mass tolerance of 0.02 Da. Upon inspection, we found that all 42 were correct and all other hits above this number were false positive. However, when glycan information is provided, then the mass accuracy can be relaxed to 0.1 Da. In this case, 37 hits were obtained corresponding to "true" glycopeptides. These results confirm that high mass accuracy (<10 ppm) is needed for the determination of glycosylation sites when there is no glycan information. However, if some glycan information is known, the false-positive hits can be decreased even with less mass accuracy. These experiments can therefore be performed on high mass accuracy instruments such as orbitraps and time-of-flight, and maybe even ion traps or quadrupoles, when glycan information is available.

The CGL1 of *X. laevis* has two potential glycosylation sites. On the basis of this analysis, we found that both sites are occupied (^{154}Asn and ^{163}Asn). The oligosaccharides associated with each sites are shown in cartoon form in Figure 6. The putative glycan structures in Figure 6 were based solely on the glycan masses and were not verified with additional analysis. However, these assignments are consistent with several structures as determined independently by NMR.[34] CGL2 of *X. laevis*, with three potential sites, was also examined and was found to contain only two occupied sites (^{154}Asn and ^{217}Asn). The glycosylation sites of this protein have been determined manually in the previous publication.[9] The GlycoX analysis yields the same exact results.

Conclusions

The program GlycoX aids in the interpretation of MS data from a nonspecific protease treatment of a glycoprotein. The determination includes both the occupation of the site and site heterogeneity. We employed primarily FTICR−MS for the analysis; however, similar analyses can be performed on other high-mass accuracy instruments that are now becoming widely available. Additionally, both ionization methods, MALDI and electrospray ionization (ESI), can be used for this analysis. The method of nonspecific protease digestion with the automated data analysis could find wide applications in the determination of glycosylation sites on glycoproteins. The current method of tryptic digestion has specific limitations, but it provides well-defined peptide chains. For this reason, the two methods may be complementary.

The software will be made available upon requests. To fully automate the analysis, it would be desirable to have a graphical interface that draws the glycans and the associated sites automatically.

Acknowledgment. We gratefully acknowledge the financial support provided to C.B.L. by the National Institutes of Healthy (R01 GM049077) and to D.M.R. by the National Cancer Institute (P30 CA093373-04) and the National Institute of Environmental Health Sciences (P42-ES0469). We also thank Drs. Jerry Hedrick and Thomas Peavy for providing the sample of the *Xenopus cortical granule lectin*.

References

(1) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(2) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364−2369.
(3) Lowe, J. B. *Cell* **2001**, *104*, 809−812.
(4) Butler, M.; Quelhas, D.; Critchley, A. J.; Carchon, H.; Hebestreit, H. F.; Hibbert, R. G.; Vilarinho, L.; Teles, E.; Matthijs, G.; Schollen, E.; Argibay, P.; Harvey, D. J.; Dwek, R. A.; Jaeken, J.; Rudd, P. M. *Glycobiology* **2003**, *13*, 601−622.
(5) Marquardt, T.; Freeze, H. *Biol. Chem.* **2001**, *382*, 161−177.
(6) Brockhausen, I. *Biochim. Biophys. Acta* **1999**, *1473*, 67−95.
(7) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4−8.
(8) Juhasz, P.; Martin, S. A. *Int. J. Mass Spectrom.* **1997**, *169*, 217−230.
(9) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **2003**, *75*, 5628−5637.
(10) Imre, T.; Schlosser, G.; Pocsfalvi, G.; Siciliano, R.; Molnar-Szollosi, E.; Kremmer, T.; Malorni, A.; Vekey, K. *J. Mass Spectrom.* **2005**, *40*, 1472−1483.
(11) Mormann, M.; Paulsen, H.; Peter-Katalinic, J. *Eur. J. Mass Spectrom.* **2005**, *11*, 497−511.
(12) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P. *J. Proteome Res.* **2004**, *3*, 556−566.
(13) Medzihradszky, K. F.; Maltby, D. A.; Hall, S. C.; Settineri, C. A.; Burlingame, A. L. *J. Am. Soc. Mass Spectrosc.* **1994**, *5*, 350−358.
(14) Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, J.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2020−2030.
(15) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(16) Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wisniewski, J. R. *Proteomics* **2004**, *4*, 454−465.
(17) Dezutterdambuyant, C.; Schmitt, D. A.; Dusserre, N.; Hanau, D.; Kolbe, H. V. J.; Kieny, M. P.; Gazzolo, L.; Mace, K.; Pasquali, J. L.; Olivier, R.; Schmitt, D. *Res. Virol.* **1991**, *142*, 129−138.
(18) Bezouska, K.; Sklenar, J.; Novak, P.; Halada, P.; Havlicek, V.; Kraus, M.; Ticha, M.; Jonakova, V. *Protein Sci.* **1999**, *8*, 1551−1556.
(19) Cooper, C. A.; Gasteiger, E.; Packer, N. H. *Proteomics* **2001**, *1*, 340−349.
(20) Wilkins, M. R.; Gasteiger, E.; Gooley, A. A.; Herbert, B. R.; Molloy, M. P.; Binz, P. A.; Ou, K. L.; Sanchez, J. C.; Bairoch, A.; Williams, K. L.; Hochstrasser, D. F. *J. Mol. Biol.* **1999**, *289*, 645−657.
(21) Lehmann, W. D.; Bohne, A.; von der Lieth, C. W. *J. Mass Spectrom.* **2000**, *35*, 1335−1341.
(22) Wuhrer, M.; Koeleman, C. A. M.; Hokke, C. H.; Deelder, A. M. *Anal. Chem.* **2005**, *77*, 886−894.
(23) Wuhrer, M.; Balog, C. I. A.; Koeleman, C. A. M.; Deelder, A. M.; Hokke, C. H. *Biochim. Biophys. Acta* **2005**, *1723*, 229−239.
(24) Jiang, H.; Desaire, H.; Butnev, V. Y.; Bousfield, G. R. *J. Am. Soc. Mass Spectrosc.* **2004**, *15*, 750−758.
(25) Jiang, H.; Irungu, J.; Desaire, H. *J. Am. Soc. Mass Spectrosc.* **2005**, *16*, 340−348.
(26) Larsen, M. R.; Hojrup, P.; Roepstorff, P. *Mol. Cell. Proteomics* **2005**, *4*, 107−119.
(27) Quadroni, M.; Ducret, A.; Stocklin, R. *Proteomics* **2004**, *4*, 2211−2215.
(28) Wehofsky, M.; Hoffmann, R.; Hubert, M.; Spengler, B. *Eur. J. Mass Spectrom.* **2001**, *7*, 39−46.
(29) Wehofsky, M.; Hoffmann, R. *J. Mass Spectrom.* **2002**, *37*, 223−229.
(30) Johnson, K. L.; Muddiman, D. C. *J. Am. Soc. Mass Spectrosc.* **2004**, *15*, 437−445.
(31) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.
(32) Suzuki, T.; Kitajima, K.; Emori, Y.; Inoue, Y.; Inoue, S. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6244−6249.
(33) Fu, D. T.; Chen, L.; Oneill, R. A. *Carbohydr. Res.* **1994**, *261*, 173−186.
(34) Song, Y. W., N. J.; Shimoda, Y.; Hedrick, J. L. unpublished results.

PR0602949