

## Abstract

**Motivation:** Glycosylation is one of the most heterogenous and complex post-translational modifications, but.  
**Results:** These are the results for this article.

# Application of Network Smoothing to Glycan LC-MS Profiling

Joshua Klein

July 12, 2017

## 1 Introduction

Glycosylation is one of the most pervasive forms of post-translational modification.

Table 1: Glycan Composition Rule Table

Monosaccharide	Lower Limit	Upper Limit	Constraints
<b>HexNAc</b>	2	9	
<b>Hex</b>	3	10	
<b>Fuc</b>	0	4	<b>HexNAc &gt; Fuc</b>
<b>NeuAc</b>	0	5	<b>(HexNAc – 1) &gt; NeuAc</b>

Table 2: Samples Used

Sample Name	Instrument	Derivatization	Adduction	Source
20150930-06-AGP	QTOF	Native	Formate (1)	Khatri <i>et al.</i> (2016a)
20141031-07-Phil-82	QTOF	Native	Formate(3)	Khatri <i>et al.</i> (2016a)
20141101-04-Phil-BS	QTOF	Native	Formate(3)	Khatri <i>et al.</i> (2016a)
20151002-02-IGG	QTOF	Native	Formate (2)	Khatri <i>et al.</i> (2016b)
20141128-11-Phil-82	QTOF	Deutero-reduced and Permethylated	Ammonium (3)	Khatri <i>et al.</i> (2016a)
AGP-DR-Perm-glycans-1	FTMS	Deutero-reduced and Permethylated	Ammonium (3)	Khatri <i>et al.</i> (2016a)
AGP-permethylated-2ul-inj-55-SLens	FTMS	Reduced and Permethylated	Ammonium (3)	Khatri <i>et al.</i> (2016a)
Perm-BS-070111-04-Human-Serum	FTMS	Reduced and Permethylated	Ammonium (3)	Yu <i>et al.</i> (2013)

## 2 Methods

### 2.1 Glycan Hypothesis Generation

In eukaryotes, *N*-glycans start with a common, conserved core of **HexNAc2 Hex3**, building up to **HexNAc2 Hex9** (Stanley *et al.* (2009)). This structure is refined by sequentially removing monosaccharides and replacing them with more complex structures through a series of glycosidase and glycosyltransferase reactions, the enumeration of which as shown in Akune *et al.* (2016) yields over a million of possible *N*-glycan topologies and epitopes. These topologies define the geometry of the glycan, affecting the glycan’s binding affinities and how the glycan may influence protein folding and accessibility, the glycan’s functional aspects. The medium through which we observe *N*-glycan does not capture the full tree or graph structure of an *N*-glycan , so we reduce the topology to a count of each type of residue.

Starting with the core motif, we generate all combinations of monosaccharides ranging between the limits in Table 1. We created a copy of this database for native, reduced and permethylated, and deuteroreduced and permethylated. Let  $n = 1240$  be the number of glycan compositions  $\mathbf{g}$  in the database.

### 2.2 LC-MS Data Preprocessing

We used samples from several sources including both QTOF and FTMS instruments as shown in Table 2. For details on sample preparation and data acquisition, please see their source citation. All data were converted into mzML format (Martens *et al.* (2011)) prior to analysis with Proteowizard (Kessner *et al.* (2008)) without any data transforming filters. We applied a background reduction method based upon (Kaur and O’Connor (2006)), using a window length of 2 m/z. Next, we picked peaks using a simple gaussian model. Scans were then subjected to iterative charge state deconvolution and deisotoping using an averagine (Senko *et al.* (1995)) formula appropriate to the molecule under study. For native glycans, the formula was **H 1.690 C 1.0 O 0.738 N 0.071**, for permethylated glycans, the formula was **H 1.819 C 1.0 O 0.431 N 0.042**. We used an iterative approach which combines aspects of the dependence graph method (Liu *et al.* (2010)) and with subtraction. All samples were processed using a minimum isotopic fit score of 35 with an isotopic strictness penalty of 2.

Table 3: Chromatogram Feature Definitions

$\mathcal{M}_i$	The neutral mass of the $i$ th chromatogram
$\mathcal{I}_i$	The total intensity array assigned to the $i$ th chromatogram
$\mathcal{I}_{i,j}$	The sum of all peak intensities for peaks observed in the $j$ th scan for the $i$ th chromatogram
$\mathcal{I}_{i,j,k}$	The intensity assigned to the $k$ th peak at the $j$ th scan for the $i$ th chromatogram
$\mathbf{c}_i$	The set of charge states observed for the $i$ th chromatogram
$\mathcal{I}_{i,c=j}$	The total intensity assigned to the $i$ th chromatogram with charge state $j$
$\mathbf{t}_{i,j}$	The time of the $j$ th scan of the $i$ th chromatogram
$\mathbf{env}_{i,j,k}$	The normalized experimental isotopic envelope composing the $k$ th peak of the $j$ th scan of the $i$ th chromatogram, whose members sum to 1
$\mathbf{a}_i$	The set of adduction states observed for the $i$ th chromatogram
$\mathcal{I}_{i,a=j}$	The total intensity assigned to the $i$ th chromatogram with adduct $j$
$\hat{g}_i$	The glycan composition assigned to the $i$ th chromatogram, or $\emptyset$ if there was no matched glycan composition

### 2.3 Chromatogram Aggregation

We clustered peaks whose neutral masses were within 15 parts-per-million error (PPM) of each other. When there were multiple candidate clusters for a single peak, we used the cluster with the lowest mass error. After all peaks were clustered, we sorted each cluster by time, creating a list of aggregated chromatograms. To account for small mass differences, we found all chromatograms

### 2.4 Glycan Composition Matching

For each chromatogram, we queried the target glycan database for compositions whose masses were within  $\delta_{mass} = 10$  PPM mass error for QTOF data, 5 PPM mass error for FTMS data. We merged all features matching the same composition. Then, for each adduct combination, we searched the target glycan database for compositions whose neutral mass were within  $\delta_{mass}$  of the observed neutral mass - adduct combination mass, followed by another round of merging chromatograms with the same assigned composition. We reduced the data by splitting each feature where the time between sequential observation was greater than  $\delta_{rt} = 0.25$  minutes and removed features with fewer than  $k = 5$  data points. We term the remaining assigned and unassigned chromatograms *candidate features*.

### 2.5 Feature Evaluation

For each candidate feature, we computed several statistics to estimate how distinguishable the observed signal was from random noise. We use the following quantities from each LC-MS feature:

#### 2.5.1 Chromatographic Peak Shape

An LC-MS elution profile should be composed of one or more peak-like components, each following a bi-Gaussian peak shape model (Yu and Peng (2010)) or in less ideal chromatographic circumstances, a skewed Gaussian peak shape model. We fit these models using non-linear least squares (NLS). As measures of goodness of fit are not generally available for NLS, we use

the following criterion:

$$\begin{aligned}
\hat{y}_i &= NLS(\mathcal{I}_i, \mathbf{t}_i) \\
e_{i,NLS} &= \mathcal{I}_i - \hat{y}_i \\
\bar{y}_i &= \mathbf{t}_i \left( (\mathbf{t}_i^t \mathbf{t}_i)^{-1} \mathbf{t}_i \mathcal{I}_i \right) \\
e_{i,null} &= \mathcal{I}_i - \bar{y}_i \\
\mathcal{L}_i &= 1 - \frac{\sum e_{i,NLS}^2}{\sum e_{i,null}^2}
\end{aligned} \tag{1}$$

where line score describes how much the peak shape fit improves on a straight line fit null model.

We apply two competitive peak fitting strategies to address distorted, overlapping, or multimodal elution profiles. The first works iteratively by finding a best-matching peak shape using non-linear least squares, subtracting the fitted signal and checks if there is another peak with at least half as tall as the removed peak, if so repeating the process until no peak can be found, saving each peak model so constructed. The second approach starts by locating local minima between putative peaks, and partitioning the chromatogram into sub-groups which would be fit independently. This method generates a candidate list of minima, and selects the case which has the greatest difference between the minimum and its pair of maxima to split the feature at. The strategy which produces the maximum  $\mathcal{L}_i$  is chosen.

### 2.5.2 Composition Dependent Charge State Distribution

As the number of monosaccharides composing a glycan increases, the number of possible sites for charge localization increases. Under normal conditions, we would expect to observe the same molecule in multiple charge states (Maxwell *et al.* (2012)). Which charge states are expected would depend upon the size of the molecule and it's constituent units' electronegativity. In it's native state, **NeuAc**'s acidic group causes glycans with one or more **NeuAc** to have a propensity for higher negative charge states (Varki and Schauer (2009)). To capture this relationship, we modeled the probability of observing a glycan composition for sialylated and unsialylated compositions separately. For permethylated glycans, charge is carried by protons or metallic cation adducts like sodium, the relationship between acidic monosaccharides and charge state propensities is weaker.

$$\begin{aligned}
m_i &= (\lfloor (\mathcal{M}_i/w)/10 \rfloor + 1) * 10 \\
\mathcal{H}_{i,j} &= \frac{\mathcal{I}_{i,c=j}}{\mathcal{I}_i} \\
P(c, m) &= |m| \sum_{m_i \in m} \mathcal{H}_{i,j} \\
\mathcal{C}_i &= \sum_{c_{i,j} \in \mathbf{c}_i} P(c_{i,j}, m_i)
\end{aligned} \tag{2}$$

where  $w$  is the width of the mass bin divided by 10 and  $P(c, m)$  is defined as part of the model estimation procedure.

### 2.5.3 Adduction Rate

For the samples *AGP-permethylated-2ul-inj-55-SLens* and *Perm-BS-070111-04-Human-Serum* we also include an Adduction Frequency model score  $\mathcal{A}_i$ , following the same pattern as the charge state distribution, with the same extension of justification from Maxwell *et al.* (2012). We use one mass scaling model for all glycan compositions as ammonium adduction is not expected to be composition dependent.

$$\begin{aligned}
\mathcal{H}_{i,j} &= \frac{\mathcal{I}_{i,a=j}}{\mathcal{I}_i} \\
P(a, m) &= |m| \sum_{m_i \in m} \mathcal{H}_{i,j} \\
\mathcal{A}_i &= \sum_{a_{i,j} \in \mathbf{a}_i} P(a_{i,j}, m_i)
\end{aligned} \tag{3}$$

We fit the adduction rate model on *AGP-permethylated-2ul-inj-55-SLens* in order to make our comparison to third-party data less biased given limited sample data.

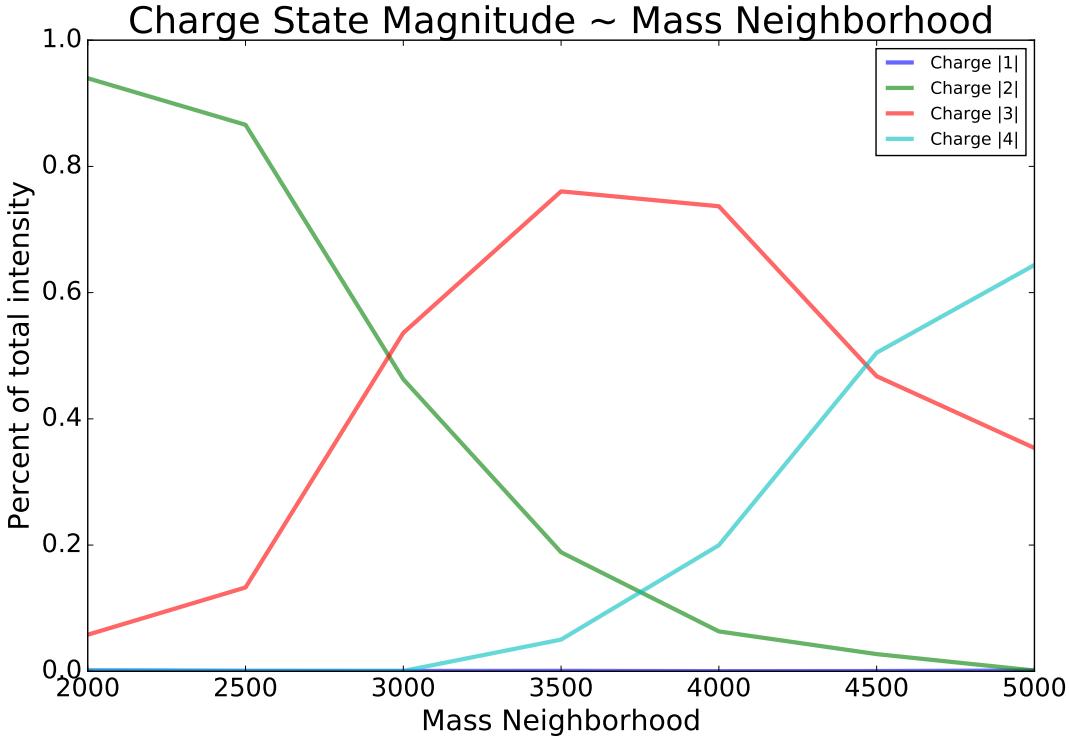


Figure 1: The trend of charge state relative abundance for acidic glycans

#### 2.5.4 Isotopic Pattern Consistency

Our ahead-of-time deconvolution procedure uses an averagine isotopic model and does not capture the consistency of the isotopic pattern that was fit with the isotopic pattern of the glycan composition that matched that peak. The criterion

$$\mathcal{I}_i = 1 - 2\mathcal{I}_i^{-t} \mathbf{I}_i \sum_j^K \sum_k^J \mathcal{I}_{i,j,k} \text{env}_{i,j,k}^t (\ln \text{env}_{i,j,k} - \ln \text{tid}_i) \quad (4)$$

where  $\text{tid}$  is the theoretical isotopic pattern derived from either  $\hat{g}_i$  or an averagine interpolated for  $\mathcal{M}_i$  if  $\hat{g}_i = \emptyset$ . This computes a per-peak intensity weighted mean G-test comparing the goodness of fit between the experimental envelope and the theoretical isotopic pattern.

#### 2.5.5 Observation Spacing Score

The less time between observations of a glycan composition the less likely the chromatogram is to contain peaks missing or caused by isotopic pattern interference or missing information.

$$\mathcal{T}_i = 1 - 2\mathcal{I}_i^{-t} \mathbf{I}_i \sum_{j=1}^J \mathcal{I}_{i,j} (\mathbf{t}_{i,j} - \mathbf{t}_{i,j-1}) \quad (5)$$

#### 2.5.6 Summarization Score

Each scoring feature  $\in [\mathcal{L}_i, \mathcal{C}_i, \mathcal{I}_i, \mathcal{T}_i]$  is penalized by  $\epsilon = 1e-6$  bounded in the range  $[0, 1)$ , with values below 0 set to  $\epsilon$ .

$$s_i = \sum_{f_{i,j} \in \text{features}_i} \ln \frac{f_{i,j}}{1 - f_{i,j}} \quad (6)$$

producing a value between  $(-\infty, \infty)$ .  $s_i < 8$  reflects multiple poor feature scores and is unexpected to be real, while  $s_i > 15$  is consistent with model expectations.

## 2.6 Glycan Composition Network Smoothing

Evidence for individual glycan compositions can often be enough to claim that composition had been detected. Lower abundance may score poorly in one or more features, leading to the glycan composition being discarded. Other methods

Name	Bounds
High Mannose	$\text{HexNAc} = 2 \wedge \text{Hex} \in [3, 10] \wedge \text{NeuAc} = 0$
Hybrid	$\text{HexNAc} \in [2, 4] \wedge \text{Hex} \in [2, 6] \wedge \text{NeuAc} \in [0, 2]$
Bi-Antennary	$\text{HexNAc} \in [3, 5] \wedge \text{Hex} \in [3, 6] \wedge \text{NeuAc} \in [1, 3]$
Asialo-Bi-Antennary	$\text{HexNAc} \in [3, 5] \wedge \text{Hex} \in [3, 6] \wedge \text{NeuAc} \in [0, 1]$
Tri-Antennary	$\text{HexNAc} \in [4, 6] \wedge \text{Hex} \in [4, 7] \wedge \text{NeuAc} \in [1, 4]$
Asialo-Tri-Antennary	$\text{HexNAc} \in [4, 6] \wedge \text{Hex} \in [4, 7] \wedge \text{NeuAc} \in [0, 0]$
Tetra-Antennary	$\text{HexNAc} \in [5, 7] \wedge \text{Hex} \in [5, 8] \wedge \text{NeuAc} \in [1, 5]$
Asialo-Tetra-Antennary	$\text{HexNAc} \in [5, 7] \wedge \text{Hex} \in [5, 8] \wedge \text{NeuAc} \in [0, 0]$
Penta-Antennary	$\text{HexNAc} \in [6, 8] \wedge \text{Hex} \in [6, 9] \wedge \text{NeuAc} \in [1, 5]$
Asialo-Penta-Antennary	$\text{HexNAc} \in [6, 8] \wedge \text{Hex} \in [6, 9] \wedge \text{NeuAc} \in [0, 0]$

Table 4: N-Glycan Neighborhoods

have demonstrated it is advantageous to use relationships between glycans based on biosynthetic or structural rules to adjust the score of a single glycan assignment (Goldberg *et al.* (2009); Kronewitter *et al.* (2014)). This idea has been explored more generically under the name "Manifold Regularization" (Belkin *et al.* (2006)) and specifically "Laplacian Regularization" when the Laplacian matrix of a graph is used to influence the parameter scaling. We apply this idea to weighted networks of related glycans with arbitrarily defined and overlapping sub-populations.

### 2.6.1 Glycan Composition Graph

For each database of theoretical glycan compositions we create, we define each composition to be a coordinate vector in a  $\mathcal{Z}^{+4}$  space, and represented by a node in an undirected glycan composition graph  $\mathcal{G}$ . Under this interpretation, we can compute the  $L_1$ -distance between two glycan compositions. For any two glycan compositions  $g_u, g_v$ , if  $L_1(g_u, g_v) = 1$  we add an edge connecting  $g_u$  and  $g_v$  to  $\mathcal{G}$  with weight  $w = 1$ .

### 2.6.2 Neighborhood Definition

Our definition of distance connects glycan compositions which differ by a single monosaccharide, but we can assert larger collections of glycan compositions are related. We define the following neighborhoods for  $N$ -glycans:

Glycan compositions may belong to zero or more neighborhoods, as there are unusual glycan compositions which do not satisfy any neighborhood's rules, and several neighborhoods intentionally overlap to express broad relationships between groups. We define a matrix  $\mathbf{A}$  as an  $n \times k$  matrix where  $A_{i,k}$  to be the degree to which  $g_i$  belongs  $k$ th neighborhood:

$$A_{i,k} = \frac{1}{|\text{neighborhood}_k|} \sum_{g^* \in \text{neighborhood}_k} L_1(g_i, g^*) \quad (7)$$

To reduce the impact of neighborhood size on the elements of  $\mathbf{A}$ , the columns of  $\mathbf{A}$  are first normalized to sum to 1, and then the rows of  $\mathbf{A}$  are normalized to sum to 1.

We assume that members of the same neighborhood will share a central tendency,  $\tau$ .

### 2.6.3 Laplacian Regularization

We combine the observed score  $\mathbf{s}$  and the structure of  $\mathcal{G}$  to estimate a smoothed score  $\phi$  that combines the evidence for the  $i$ th glycan composition as well as its relatives. As  $\mathbf{s}$  is the size of the set of observed glycan composition  $p$  while  $\phi$  is of size  $n$ , we partition  $\phi$  into a block vector  $\begin{bmatrix} \phi_o \\ \phi_m \end{bmatrix}$  with dimensions  $\begin{bmatrix} p \\ n-p \end{bmatrix}$ .

Let  $\mathbf{L}$  be the weighted Laplacian matrix of  $\mathcal{G}$ , which is an  $n \times n$  matrix. To ensure  $\mathbf{L}$  is invertible, we add  $\mathbf{I}_n$  to  $\mathbf{L}$ . We partition  $\mathbf{L}$  into blocks  $\begin{bmatrix} \mathbf{L}_{oo} & \mathbf{L}_{om} \\ \mathbf{L}_{mo} & \mathbf{L}_{mm} \end{bmatrix}$ . We also partition  $\mathbf{A}$  into  $\begin{bmatrix} \mathbf{A}_o \\ \mathbf{A}_m \end{bmatrix}$  and  $\tau_o = \mathbf{A}_o\tau$ ,  $\tau_m = \mathbf{A}_m\tau$ .

We find the  $\phi$  that minimizes the expression

$$\ell = (\mathbf{s} - \phi_o)^t(\mathbf{s} - \phi_o) + \lambda [\phi_o - \tau_o, \phi_m - \tau_m]^t \begin{bmatrix} \mathbf{L}_{oo} & \mathbf{L}_{om} \\ \mathbf{L}_{mo} & \mathbf{L}_{mm} \end{bmatrix} \begin{bmatrix} \phi_o - \tau_o \\ \phi_m - \tau_m \end{bmatrix} \quad (8)$$

where  $\lambda$  controls how much weight is placed on the network structure and  $\tau$ .

To obtain the optimal  $\phi$ , we take the partial derivative of  $\ell$  w.r.t  $\phi_m$

$$\begin{aligned}
0 &= \frac{\partial \ell}{\partial \phi_m} \left( (\mathbf{s} - \phi_{\mathbf{o}})^t (\mathbf{s} - \phi_{\mathbf{o}}) + \lambda [\phi_o - \tau_o, \phi_m - \tau_m] \begin{bmatrix} \mathbf{L}_{\mathbf{oo}} & \mathbf{L}_{\mathbf{om}} \\ \mathbf{L}_{\mathbf{mo}} & \mathbf{L}_{\mathbf{mm}} \end{bmatrix} \begin{bmatrix} \phi_o - \tau_o \\ \phi_m - \tau_m \end{bmatrix} \right) \\
&= \lambda (\phi_o - \tau_o)^t \mathbf{L}_{\mathbf{om}} + \lambda \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) + \lambda (\phi_m - \tau_m)^t (\mathbf{L}_{\mathbf{mm}}^t + \mathbf{L}_{\mathbf{mm}}) \\
&= 2\lambda \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) + 2\lambda \mathbf{L}_{\mathbf{mm}} (\phi_m - \tau_m)
\end{aligned} \tag{9}$$

$$\begin{aligned}
-\mathbf{L}_{\mathbf{mm}} (\phi_m - \tau_m) &= \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) \\
(\phi_m - \tau_m) &= -\mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) \\
\hat{\phi}_m &= -\mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) + \tau_m
\end{aligned} \tag{10}$$

and w.r.t.  $\phi_o$

$$\begin{aligned}
0 &= \frac{\partial \ell}{\partial \phi_o} \left( (\mathbf{s} - \phi_{\mathbf{o}})^t (\mathbf{s} - \phi_{\mathbf{o}}) + \lambda [\phi_o - \tau_o, \phi_m - \tau_m] \begin{bmatrix} \mathbf{L}_{\mathbf{oo}} & \mathbf{L}_{\mathbf{om}} \\ \mathbf{L}_{\mathbf{mo}} & \mathbf{L}_{\mathbf{mm}} \end{bmatrix} \begin{bmatrix} \phi_o - \tau_o \\ \phi_m - \tau_m \end{bmatrix} \right) \\
&= -2\mathbf{s} + 2\phi_o + \lambda (\mathbf{L}_{\mathbf{oo}} + \mathbf{L}_{\mathbf{oo}}^t) (\phi_o - \tau_o) + \lambda \mathbf{L}_{\mathbf{om}} (\phi_m - \tau_m) + \lambda \mathbf{L}_{\mathbf{mo}}^t (\phi_m - \tau_m) \\
&= -2\mathbf{s} + 2\phi_o + 2\lambda \mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) + 2\lambda \mathbf{L}_{\mathbf{om}} (\phi_m - \tau_m) \\
\mathbf{s} &= \phi_o + \lambda (\mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) + \mathbf{L}_{\mathbf{om}} (\phi_m - \tau_m)) \\
&= \phi_o + \lambda (\mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) + \mathbf{L}_{\mathbf{om}} (-\mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o) + \tau_m - \tau_m)) \\
&= \phi_o + \lambda (\mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o)) \\
\mathbf{s} - \tau_o &= \phi_o - \tau_o + \lambda (\mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o)) \\
&= \mathbf{I} (\phi_o - \tau_o) + \lambda (\mathbf{L}_{\mathbf{oo}} (\phi_o - \tau_o) - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}} (\phi_o - \tau_o)) \\
&= [\mathbf{I} + \lambda (\mathbf{L}_{\mathbf{oo}} - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}})] (\phi_o - \tau_o) \\
(\phi_o - \tau_o) &= [\mathbf{I} + \lambda (\mathbf{L}_{\mathbf{oo}} - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}})]^{-1} (\mathbf{s} - \tau_o) \\
\hat{\phi}_o &= [\mathbf{I} + \lambda (\mathbf{L}_{\mathbf{oo}} - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}})]^{-1} (\mathbf{s} - \tau_o) + \tau_o
\end{aligned} \tag{12}$$

To use this method, we must provide values for  $\lambda$  and  $\tau$ . While these values could be chosen based on the expectations of the user for a given experiment, we provide an algorithm for selecting their values. These methods use the topology of the glycan composition graph and the distribution of observed scores, and cannot fully capture boundary cases or related but disconnected parts of the graph.

#### 2.6.4 Parameter Estimation

We model the relationship between  $\mathbf{s}$ ,  $\phi_{\mathbf{o}}$ , and  $\tau$  as a set of gaussian distribution.

$$(\mathbf{s} | \phi_{\mathbf{o}}, \tau) \sim \mathcal{N}(\phi_{\mathbf{o}}, \Sigma) \tag{13}$$

$$\Sigma = \rho \mathbf{I} \tag{14}$$

$$\left( \begin{bmatrix} \phi_{\mathbf{o}} \\ \phi_{\mathbf{m}} \end{bmatrix} \middle| \tau \right) \sim \mathcal{N}(\mathbf{A}\tau, \lambda^{-1} \mathbf{L}^-) \tag{15}$$

$$(\phi_{\mathbf{o}} | \tau) \sim \mathcal{N}(\mathbf{A}_{\mathbf{o}} \tau, \Sigma_{\phi_{\mathbf{o}}}) \tag{16}$$

$$\Sigma_{\phi_{\mathbf{o}}} = \lambda^{-1} (\mathbf{L}_{\mathbf{oo}} - \mathbf{L}_{\mathbf{om}} \mathbf{L}_{\mathbf{mm}}^{-1} \mathbf{L}_{\mathbf{mo}})^{-1} \tag{17}$$

$$\tau \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \tag{18}$$

Fully expanded, this becomes

$$\begin{bmatrix} \mathbf{s} \\ \phi_{\mathbf{o}} \\ \tau \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma + \Sigma_{\phi_{\mathbf{o}}} + \sigma^2 \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t & \Sigma_{\phi_{\mathbf{o}}} + \sigma^2 \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t & \sigma^2 \mathbf{A}_{\mathbf{o}} \\ \Sigma_{\phi_{\mathbf{o}}} + \sigma^2 \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t & \Sigma_{\phi_{\mathbf{o}}} + \sigma^2 \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t & \sigma^2 \mathbf{A}_{\mathbf{o}} \\ \sigma^2 \mathbf{A}_{\mathbf{o}}^t & \sigma^2 \mathbf{A}_{\mathbf{o}}^t & \sigma^2 \mathbf{I} \end{bmatrix} \right) \tag{19}$$

We can form the conditional distribution  $\tau | \mathbf{s}$  which has a mean

$$\mu_{\tau | \mathbf{s}} = 0 + (\sigma^2 \mathbf{A}_{\mathbf{o}}^t) (\Sigma + \Sigma_{\phi_{\mathbf{o}}} + \sigma^2 \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t)^{-1} \mathbf{s} \tag{20}$$

$$\begin{aligned}
&= \mathbf{A}_{\mathbf{o}}^t \left( \frac{\rho}{\sigma^2} \mathbf{I} + \frac{1}{\lambda \sigma^2} \mathbf{L}_{\mathbf{oo}}^- + \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t \right)^{-1} \mathbf{s} \\
&= \mathbf{A}_{\mathbf{o}}^t \left( \tilde{\rho} \mathbf{I} + \frac{1}{\tilde{\lambda}} \mathbf{L}_{\mathbf{oo}}^- + \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t \right)^{-1} \mathbf{s}
\end{aligned} \tag{21}$$

We assume that  $\sigma^2 \gg 1$ , and treat  $\lambda$  and  $\rho$  as relative to  $\sigma^2$ , as  $\tilde{\rho}$  and  $\tilde{\lambda}$ . This model gives us an estimate for  $\tau$  given a value for  $\rho$  and  $\lambda$ . As  $\rho$  has no direct role in the central tendency of  $\phi$  or  $\mathbf{s}$ , we choose to fix the value of  $\tilde{\rho} = 0.1$ , which leaves only  $\tilde{\lambda}$ . We estimate the optimal  $\tilde{\lambda}$  by grid search, minimizing the predicted residual error sum of squares (PRESS) statistic.

$$\arg \min_{\tilde{\lambda}} \frac{\mathbf{s} - \hat{\phi}_{\mathbf{o}}}{\left(1 - \left(\mathbf{I} + \tilde{\lambda} \mathbf{L}\right)^{-1}\right)^2} \quad (22)$$

$$\begin{aligned} \arg \min_{\tilde{\lambda}} & \frac{\mathbf{s} - \left(\left[\mathbf{I} + \tilde{\lambda} (\mathbf{L}_{\mathbf{o}\mathbf{o}} - \mathbf{L}_{\mathbf{o}\mathbf{m}} \mathbf{L}_{\mathbf{m}\mathbf{m}}^{-1} \mathbf{L}_{\mathbf{m}\mathbf{o}})\right]^{-1} (\mathbf{s} - \tau_o) + \tau_o\right)}{\left(1 - \left(\mathbf{I} + \tilde{\lambda} \mathbf{L}\right)^{-1}\right)^2} \\ \arg \min_{\tilde{\lambda}} & \frac{\mathbf{s} - \left(\left[\mathbf{I} + \tilde{\lambda} (\mathbf{L}_{\mathbf{o}\mathbf{o}} - \mathbf{L}_{\mathbf{o}\mathbf{m}} \mathbf{L}_{\mathbf{m}\mathbf{m}}^{-1} \mathbf{L}_{\mathbf{m}\mathbf{o}})\right]^{-1} \left(\mathbf{s} - \mathbf{A}_{\mathbf{o}}^t \left(\tilde{\rho} \mathbf{I} + \frac{1}{\tilde{\lambda}} \mathbf{L}_{\mathbf{o}\mathbf{o}}^{-1} + \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t\right)^{-1} \mathbf{s}\right) + \mathbf{A}_{\mathbf{o}}^t \left(\tilde{\rho} \mathbf{I} + \frac{1}{\tilde{\lambda}} \mathbf{L}_{\mathbf{o}\mathbf{o}}^{-1} + \mathbf{A}_{\mathbf{o}} \mathbf{A}_{\mathbf{o}}^t\right)^{-1} \mathbf{s}\right)}{\left(1 - \left(\mathbf{I} + \tilde{\lambda} \mathbf{L}\right)^{-1}\right)^2} \end{aligned} \quad (23)$$

This formulation depends upon the value of  $\mathbf{s}$  and is sensitive to low scoring matches, which can lead to incorrect estimates of  $\tau$  and PRESS. We therefore perform a grid search over both  $\tilde{\lambda}$  and a minimum threshold for  $\mathbf{s}$ ,  $\gamma$ . As we increase  $\gamma$  we remodel the graph  $\mathcal{G}$ , removing nodes whose score is below  $\gamma$ . For each pair of neighbors of removed node  $g_m$ ,  $(g_u, g_v)$ , if  $L_1(g_u, g_v) > L_1(g_u, g_m) + L_1(g_m, g_v)$ , we add an edge from  $g_u$  to  $g_v$  with weight  $\frac{1}{L_1(g_u, g_m) + L_1(g_m, g_v)}$ , up to a limit of  $L_1(g_k, g_m) < 5$ . We give the result of this grid search the name  $\mathbf{r}$ . At each point, on the grid, we save the value of  $\tau$  in  $r_{\lambda_i, \gamma_j, \tau}$  and the PRESS in  $r_{\lambda_i, \gamma_j, \text{PRESS}}$ .

To select the optimal parameters, we traverse the grid along  $\gamma$ , computing  $\tau_{\gamma}$ :

$$\tau_{\gamma_j} = \sum_{\lambda_i} |r_{\lambda_i, \gamma_j, \tau}| * \left(\frac{\gamma_j}{b} + (1 - \frac{1}{b})\right) \quad (24)$$

where  $b$  is a bias factor defining how much weight to give to higher values of  $\gamma$  which correspond to networks made up of higher confidence assignments. We chose  $b = 4$ . We define  $\bar{\tau}_{\gamma} = \max \tau_{\gamma}$  and define the vector  $\bar{\gamma} = [\gamma_j \leftarrow \tau_{\gamma_j} \geq \bar{\tau}_{\gamma} * 0.9]$ . This favors values of  $\gamma$  where large values of  $\tau$  are selected, meaning that the neighborhoods are well populated, while also giving an estimate for  $\tilde{\lambda}$  that is non-zero. We term the values of  $\gamma$  in  $\bar{\gamma}$  the *target thresholds* of  $\mathbf{s}$ .

To estimate  $\tilde{\lambda}$  and  $\tau$  from these results, we select the columns of the grid  $\mathbf{r}$  at each  $\gamma_j \in \bar{\gamma}$  and set  $\bar{\lambda}_j = \arg \min_{\lambda_i} r_{\lambda_i, \gamma_j, \text{PRESS}}$

$$\bar{\tau}_{\gamma} = \max \tau_{\gamma} \quad (25)$$

$$\bar{\gamma} = [\gamma_j \leftarrow \tau_{\gamma_j} \geq \bar{\tau}_{\gamma} * 0.9] \quad (26)$$

$$\bar{\lambda}_j = \arg \min_{\lambda_i} r_{\lambda_i, \gamma_j, \text{PRESS}} \leftarrow \gamma_j \in \bar{\gamma} \quad (27)$$

$$\mathbf{s}_{\gamma_j} = [s_i \leftarrow s_i > \gamma_j] \quad (28)$$

$$\bar{\tau}_{\mathbf{j}} = \mu_{\tau | \mathbf{s}_{\gamma_j}, \bar{\lambda}_j} \quad (29)$$

$$\hat{\lambda} = \frac{1}{|\bar{\lambda}|} \sum_j \bar{\lambda}_j \quad (30)$$

$$\hat{\tau} = \frac{1}{|\bar{\tau}|} \sum_j \bar{\tau}_{\mathbf{j}} \quad (31)$$

$$\hat{\gamma} = \frac{1}{|\bar{\gamma}|} \sum_j \bar{\gamma}_j \quad (32)$$

where  $\mathbf{s}_{\gamma_j}$  is the set of observed scores which are greater than  $\gamma_j$ , but where the estimation of is carried out with the complete Laplacian  $\mathbf{L}$ , not the reduced network used to compute  $\mathbf{r}$ . This set of averaged estimates of  $\hat{\lambda}$  and  $\hat{\tau}$  are then used to estimate  $\hat{\phi}_o$  by (12).

## 2.7 Performance Comparison

We compare the performance of the described algorithm with and without network smoothing. State of the art glycan LC-MS profiling software has been designed around Thermo-Fisher Scientific instrumentation, with support for their binary format

(Kronewitter *et al.* (2014), Yu *et al.* (2013)) but not open community formats. MultiGlycan-ESI, though publicly available, was unable to be applied to the majority of our datasets because they were not acquired on that vendor's instruments, and their mzXML alternative did not produce matches consistent with their previously published results on *Perm-BS-070111-04-Human-Serum*, and when ran on *AGP-permethylated-2ul-inj-55-SLens* it ran out of memory. GlyQ-IQ was made available for testing by its authors, but required Thermo-Fisher binary format, but, as it assumed that glycans were in native form, and it did not make its test data publicly available.

(a)	$\tau$	Glycan Compostion	(b)	Unregularized Score	Regularized Score
Group					
high-mannose	0.00	{Fuc:1; Hex:3; HexNAc:4}	15.51	14.77	
hybrid	14.58	{Fuc:1; Hex:4; HexNAc:4}	18.95	14.66	
bi-antennary	8.67	{Fuc:1; Hex:4; HexNAc:4; Neu5Ac:1}	13.23	12.23	
asialo-bi-antennary	14.47	{Fuc:1; Hex:5; HexNAc:4}	18.23	14.52	
tri-antennary	5.30	{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:1}	14.11	12.29	
asialo-tri-antennary	11.61	{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:2}	11.78	10.95	
tetra-antennary	3.09	{Fuc:1; Hex:3; HexNAc:5}	11.41	14.15	
asialo-tetra-antennary	0.00	{Fuc:1; Hex:4; HexNAc:5}	15.34	13.57	
penta-antennary	0.00	{Fuc:1; Hex:5; HexNAc:5; Neu5Ac:1}	10.30	8.80	
asialo-penta-antennary	0.00	{Fuc:1; Hex:5; HexNAc:5; Neu5Ac:2}	14.07	7.54	

Table 5: Search Results for *20151002-02-IGG*. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 13.295275$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

### 3 Results

### 3.1 Native N-Glycans

The results for *20151002-02-IGG-Glycomics* are shown in Table 5 and visualized in Figure 2. The results for *20141031-07-Phil-82* are shown in Table 6 and visualized in Figure 3. The results for *20150930-06-AGP* are shown in Table 7 and visualized in Figure 4. The results for *20141101-04-Phil-BS* are shown in Table 8 and visualized in Figure 5.

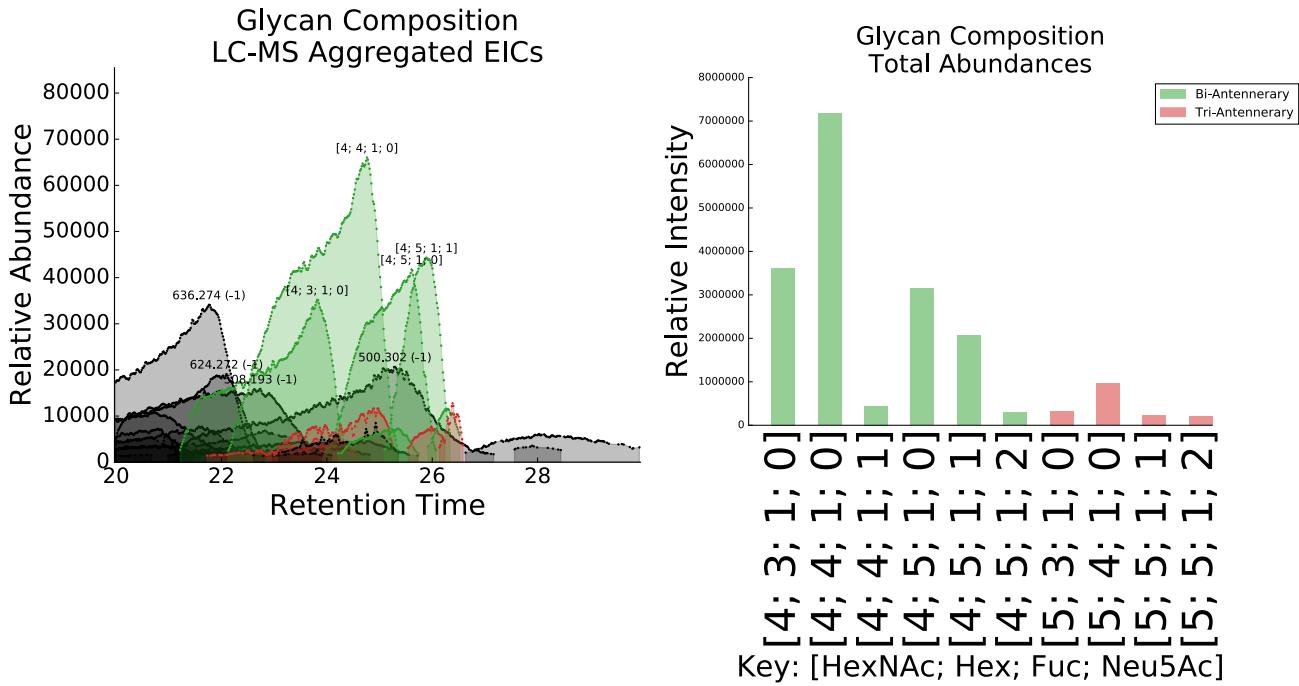


Figure 2: 20151002-02-IGG Glycan Relative Abundances

### 3.2 Deuteroreduced and Permethylated N-Glycans

The results for *20141128-11-Phil-82* are shown in Table 9 and visualized in Figure 6. The results for *AGP-DR-Perm-glycans-1* are shown in Table 10 and visualized in Figure 7.

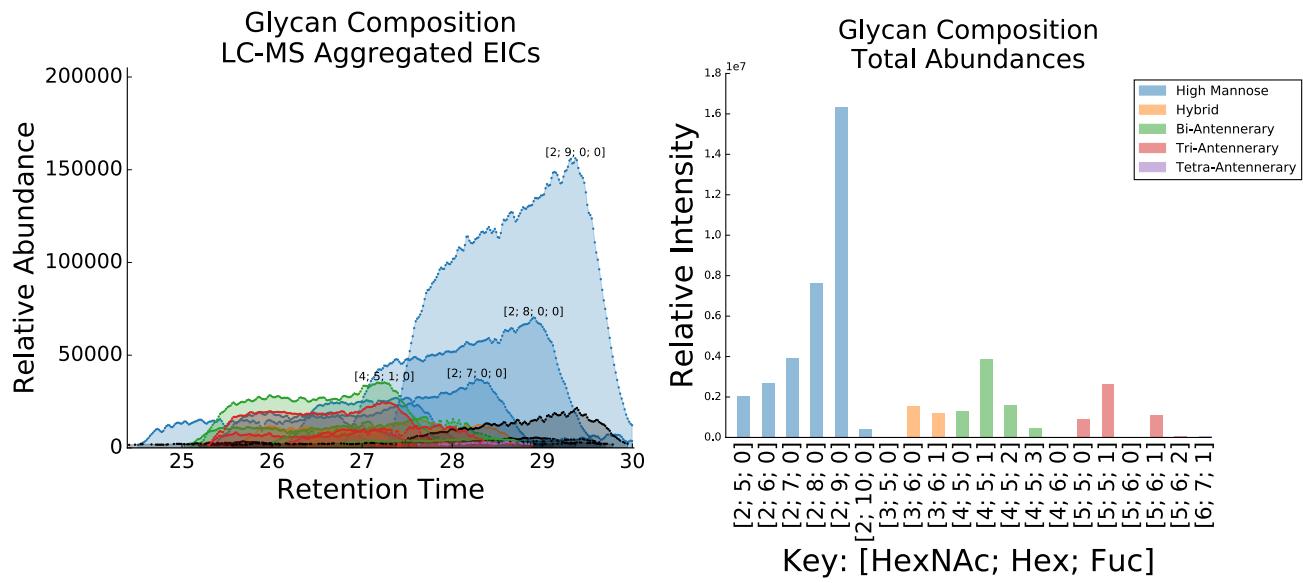


Figure 3: 20141031-07-Phil-82 Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	(b)	
			Unregularized Score	Regularized Score
high-mannose	15.88	{Hex:5; HexNAc:2}	14.89	15.23
hybrid	12.12	{Hex:6; HexNAc:2}	16.81	15.77
bi-antennary	0.00	{Hex:7; HexNAc:2}	17.27	16.33
asialo-bi-antennary	18.57	{Hex:8; HexNAc:2}	18.63	16.52
tri-antennary	0.00	{Hex:9; HexNAc:2}	19.10	16.20
asialo-tri-antennary	11.61	{Hex:10; HexNAc:2}	6.80	13.94
tetra-antennary	0.00	{Hex:5; HexNAc:3}	10.01	14.31
asialo-tetra-antennary	6.18	{Hex:6; HexNAc:3}	17.03	15.17
penta-antennary	0.00	{Fuc:1; Hex:6; HexNAc:3}	16.76	15.20
asialo-penta-antennary	0.00	{Hex:5; HexNAc:4}	16.29	14.46
		{Fuc:1; Hex:5; HexNAc:4}	17.07	14.60
		{Fuc:2; Hex:5; HexNAc:4}	14.69	14.15
		{Fuc:3; Hex:5; HexNAc:4}	7.75	13.12
		{Hex:6; HexNAc:4}	12.66	13.88
		{Hex:5; HexNAc:5}	16.43	13.35
		{Fuc:1; Hex:5; HexNAc:5}	16.51	13.36
		{Hex:6; HexNAc:5}	6.74	11.68
		{Fuc:1; Hex:6; HexNAc:5}	10.94	12.15
		{Fuc:2; Hex:6; HexNAc:5}	7.94	11.79
		{Hex:7; HexNAc:6; Neu5Ac:4}	6.39	1.06
		{Fuc:1; Hex:7; HexNAc:6}	4.62	5.74
		{Fuc:1; Hex:8; HexNAc:7}	8.16	3.49

Table 6: Search Results for 20141031-07-Phil-82. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 15.372934$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

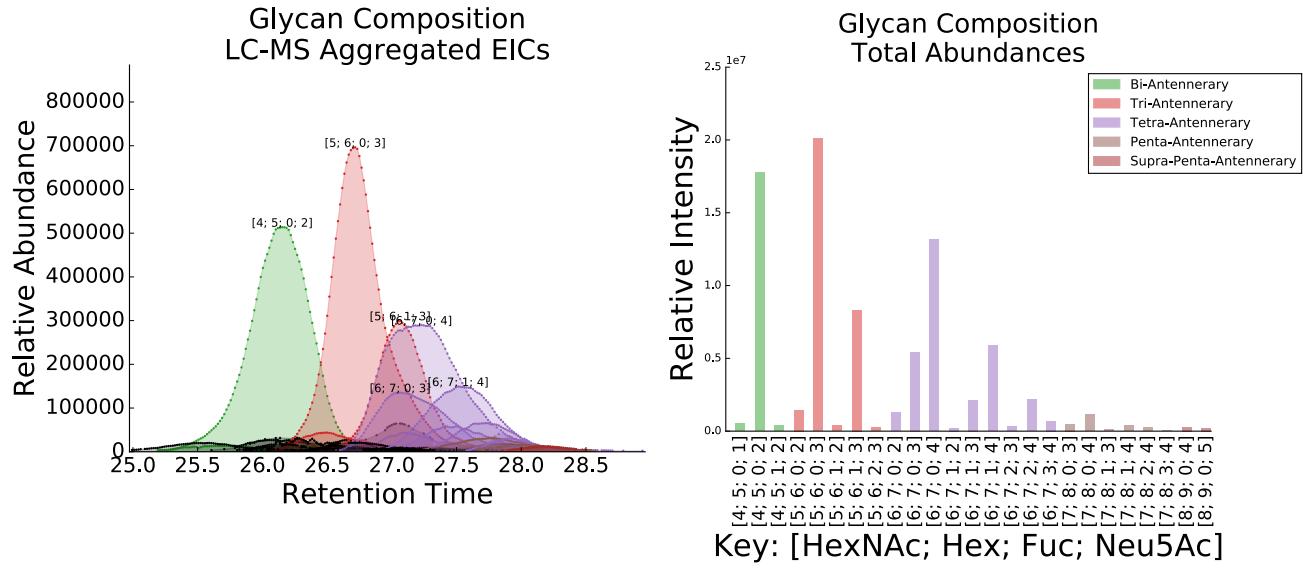


Figure 4: 20150930-06-AGP Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	Unregularized Score	Regularized Score
high-mannose	0.00	{Hex:5; HexNAc:4; Neu5Ac:1}	13.33	10.00
hybrid	11.81	{Hex:5; HexNAc:4; Neu5Ac:2}	23.36	16.47
bi-antennary	15.19	{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:2}	13.92	15.11
asialo-bi-antennary	1.28	{Hex:6; HexNAc:5; Neu5Ac:2}	19.77	17.49
tri-antennary	21.46	{Hex:6; HexNAc:5; Neu5Ac:3}	20.13	17.60
asialo-tri-antennary	0.98	{Fuc:1; Hex:6; HexNAc:5; Neu5Ac:2}	17.68	17.12
tetra-antennary	14.01	{Fuc:1; Hex:6; HexNAc:5; Neu5Ac:3}	17.53	17.08
asialo-tetra-antennary	0.00	{Fuc:2; Hex:6; HexNAc:5; Neu5Ac:3}	12.97	16.40
penta-antennary	9.61	{Hex:7; HexNAc:6; Neu5Ac:2}	17.96	16.21
asialo-penta-antennary	0.00	{Hex:7; HexNAc:6; Neu5Ac:3}	17.35	16.18
		{Hex:7; HexNAc:6; Neu5Ac:4}	20.81	16.55
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:2}	12.55	15.59
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:3}	17.32	15.91
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:4}	17.17	15.95
		{Fuc:2; Hex:7; HexNAc:6; Neu5Ac:3}	8.66	14.98
		{Fuc:2; Hex:7; HexNAc:6; Neu5Ac:4}	16.66	15.61
		{Fuc:3; Hex:7; HexNAc:6; Neu5Ac:4}	13.77	15.28
		{Hex:8; HexNAc:7; Neu5Ac:3}	13.27	12.00
		{Hex:8; HexNAc:7; Neu5Ac:4}	12.81	11.97
		{Fuc:1; Hex:8; HexNAc:7; Neu5Ac:3}	8.78	11.48
		{Fuc:1; Hex:8; HexNAc:7; Neu5Ac:4}	10.66	11.62
		{Fuc:2; Hex:8; HexNAc:7; Neu5Ac:4}	9.66	11.45
		{Fuc:3; Hex:8; HexNAc:7; Neu5Ac:4}	7.59	11.25
		{Hex:9; HexNAc:8; Neu5Ac:4}	11.12	9.80
		{Hex:9; HexNAc:8; Neu5Ac:5}	9.15	9.58

Table 7: Search Results for 20150930-06-AGP. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 14.837758$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

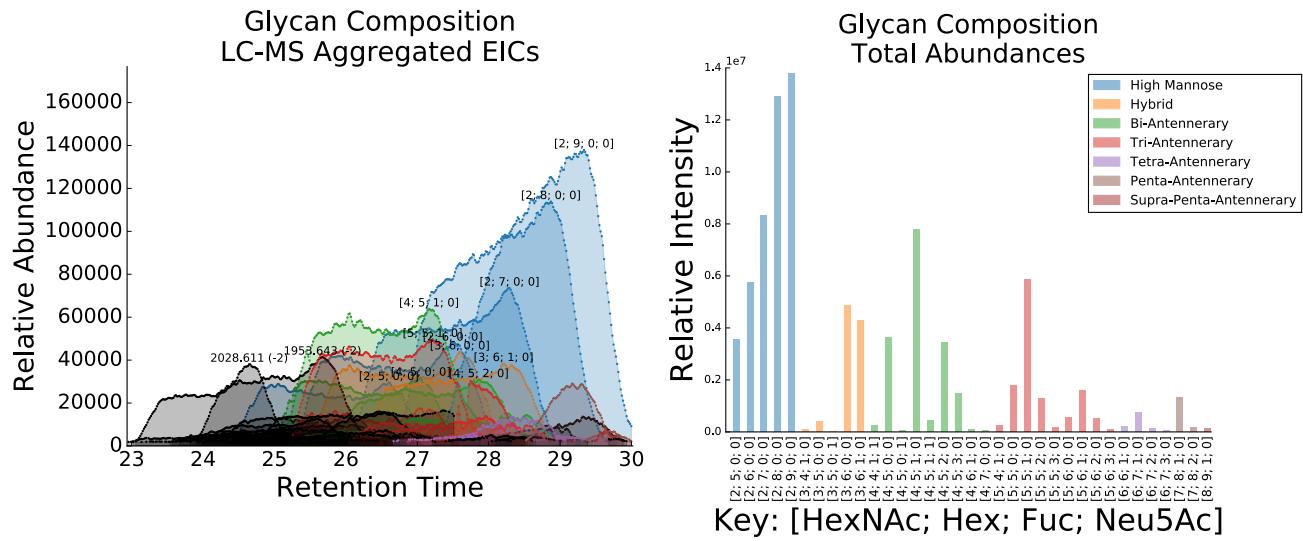


Figure 5: 20141101-04-Phil-BS Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	Unregularized Score	Regularized Score
high-mannose	1.86	{Hex:5; HexNAc:2}	18.44	18.03
hybrid	1.50	{Hex:6; HexNAc:2}	17.74	17.53
bi-antennary	0.00	{Hex:7; HexNAc:2}	17.94	17.64
asialo-bi-antennary	1.70	{Hex:8; HexNAc:2}	19.02	18.64
tri-antennary	0.00	{Hex:9; HexNAc:2}	19.91	19.37
asialo-tri-antennary	0.94	{Fuc:1; Hex:4; HexNAc:3}	10.74	10.40
tetra-antennary	0.00	{Hex:5; HexNAc:3}	11.95	11.90
asialo-tetra-antennary	0.89	{Hex:5; HexNAc:3; Neu5Ac:1}	7.71	7.68
penta-antennary	0.00	{Hex:6; HexNAc:3}	19.15	18.62
asialo-penta-antennary	0.61	{Fuc:1; Hex:6; HexNAc:3}	17.20	16.62
		{Hex:7; HexNAc:3; Neu5Ac:1}	8.19	7.98
		{Fuc:1; Hex:4; HexNAc:4; Neu5Ac:1}	9.02	8.66
		{Hex:5; HexNAc:4}	15.62	15.31
		{Hex:5; HexNAc:4; Neu5Ac:1}	11.41	11.05
		{Fuc:1; Hex:5; HexNAc:4}	17.64	17.15
		{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:1}	7.95	7.87
		{Fuc:2; Hex:5; HexNAc:4}	16.33	15.72
		{Fuc:3; Hex:5; HexNAc:4}	11.08	10.83
		{Fuc:1; Hex:6; HexNAc:4}	12.31	12.11
		{Hex:7; HexNAc:4}	11.97	11.58
		{Hex:8; HexNAc:4; Neu5Ac:2}	6.75	6.50
		{Fuc:1; Hex:4; HexNAc:5}	12.16	11.74
		{Hex:5; HexNAc:5}	17.18	16.63
		{Fuc:1; Hex:5; HexNAc:5}	17.37	16.80
		{Fuc:2; Hex:5; HexNAc:5}	8.08	8.06
		{Fuc:3; Hex:5; HexNAc:5}	7.51	7.33
		{Hex:6; HexNAc:5}	10.03	9.89
		{Fuc:1; Hex:6; HexNAc:5}	11.89	11.63
		{Fuc:2; Hex:6; HexNAc:5}	8.99	8.78
		{Fuc:3; Hex:6; HexNAc:5}	8.19	7.92
		{Hex:9; HexNAc:5; Neu5Ac:2}	11.40	10.72
		{Fuc:1; Hex:6; HexNAc:6}	7.79	7.59
		{Fuc:1; Hex:7; HexNAc:6}	9.13	8.81
		{Fuc:2; Hex:7; HexNAc:6}	7.39	7.20
		{Fuc:3; Hex:7; HexNAc:6}	7.46	7.15
		{Fuc:1; Hex:8; HexNAc:7}	17.25	16.39
		{Fuc:2; Hex:8; HexNAc:7}	15.20	14.46
		{Hex:3; HexNAc:8; Neu5Ac:4}	10.11	9.56
		{Fuc:1; Hex:9; HexNAc:8}	13.44	12.70

Table 8: Search Results for 20141101-04-Phil-BS. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 15.74699$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.01$ .

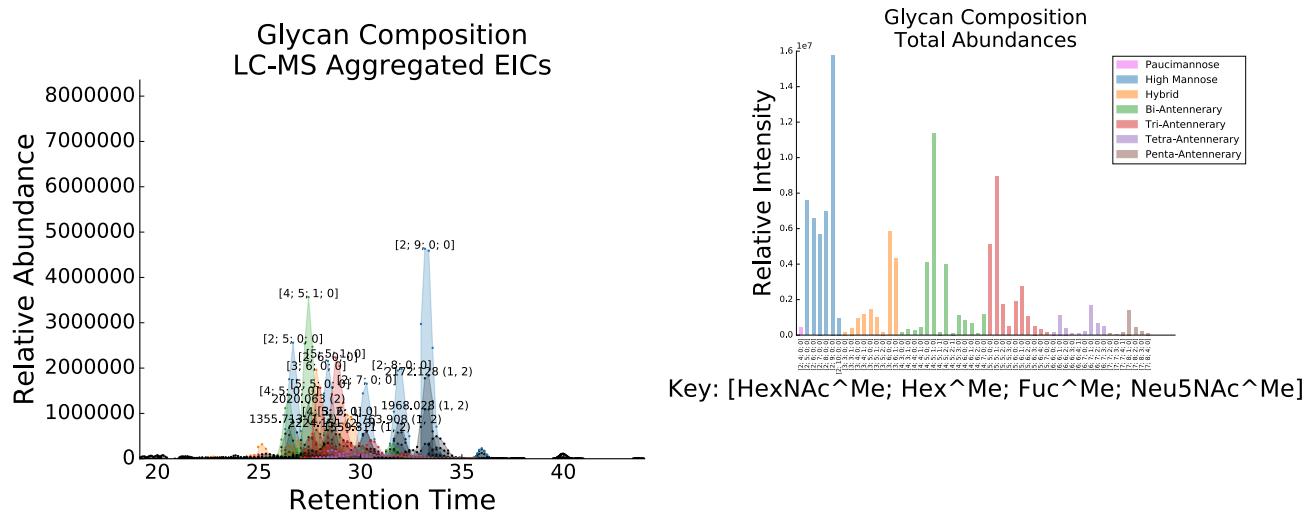


Figure 6: 20141128-11-Phil-82 Glycan Relative Abundances

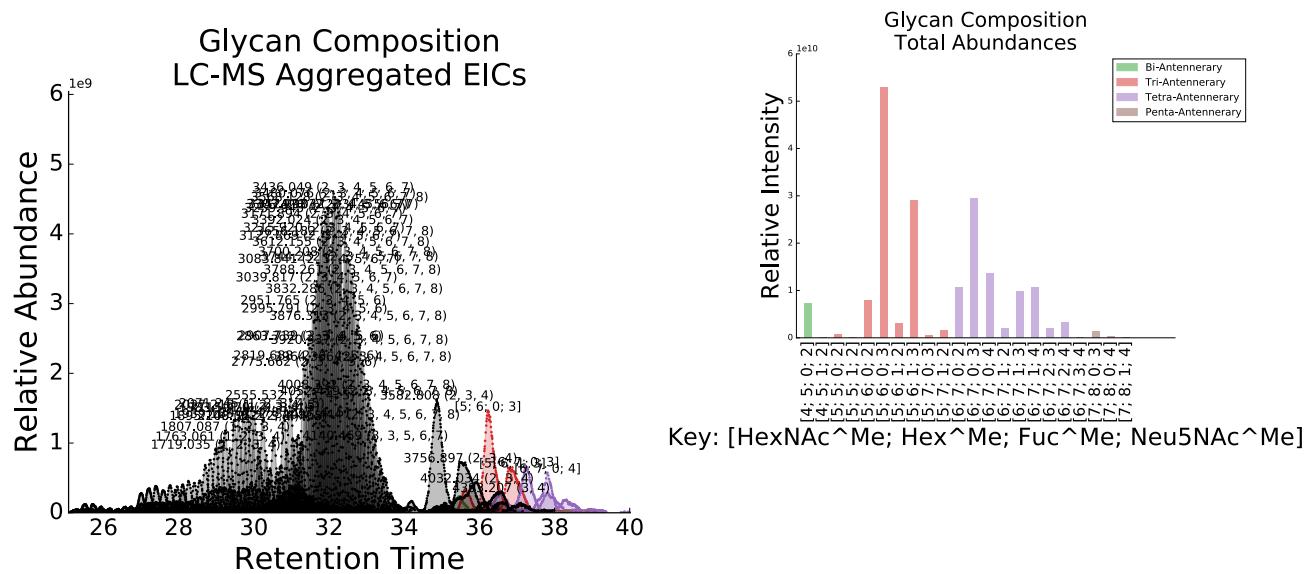


Figure 7: AGP-DR-Perm-glycans-1 Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	(b)	
			Unregularized Score	Regularized Score
high-mannose	18.08	{Hex:4; HexNAc:2}	17.40	17.64
hybrid	15.20	{Hex:5; HexNAc:2}	20.14	18.28
bi-antennary	1.42	{Hex:6; HexNAc:2}	22.02	18.71
asialo-bi-antennary	21.55	{Hex:7; HexNAc:2}	21.78	19.01
tri-antennary	0.74	{Hex:8; HexNAc:2}	15.52	17.70
asialo-tri-antennary	17.12	{Hex:9; HexNAc:2}	16.43	17.43
tetra-antennary	0.00	{Hex:10; HexNAc:2}	12.24	16.62
asialo-tetra-antennary	10.61	{Hex:3; HexNAc:3}	15.09	17.25
penta-antennary	0.00	{Fuc:1; Hex:3; HexNAc:3}	19.38	17.99
asialo-penta-antennary	2.91	{Hex:4; HexNAc:3}	17.48	17.65
		{Fuc:1; Hex:4; HexNAc:3}	20.29	18.07
		{Hex:5; HexNAc:3}	17.18	17.86
		{Fuc:1; Hex:5; HexNAc:3}	20.16	18.15
		{Fuc:2; Hex:5; HexNAc:3}	12.93	17.21
		{Hex:6; HexNAc:3}	15.06	17.76
		{Fuc:1; Hex:6; HexNAc:3}	16.30	17.72
		{Hex:7; HexNAc:3}	6.85	1.02
		{Hex:3; HexNAc:4}	11.87	16.82
		{Fuc:1; Hex:3; HexNAc:4}	17.49	17.69
		{Hex:4; HexNAc:4}	13.19	16.88
		{Fuc:1; Hex:4; HexNAc:4}	9.31	16.76
		{Hex:5; HexNAc:4}	17.27	17.72
		{Fuc:1; Hex:5; HexNAc:4}	21.31	18.21
		{Fuc:1; Hex:5; HexNAc:4; Neu5NAc:1}	16.43	12.93
		{Fuc:2; Hex:5; HexNAc:4}	21.10	18.01
		{Fuc:2; Hex:5; HexNAc:4; Neu5NAc:1}	6.09	11.83
		{Fuc:3; Hex:5; HexNAc:4}	19.42	18.07
		{Hex:6; HexNAc:4}	16.30	17.59
		{Fuc:1; Hex:6; HexNAc:4}	16.56	17.54
		{Fuc:2; Hex:6; HexNAc:4}	7.00	16.46
		{Hex:7; HexNAc:4}	17.23	17.26
		{Hex:5; HexNAc:5}	20.42	17.62
		{Fuc:1; Hex:5; HexNAc:5}	20.33	17.67
		{Fuc:2; Hex:5; HexNAc:5}	20.81	17.55
		{Fuc:3; Hex:5; HexNAc:5}	16.31	16.89
		{Hex:6; HexNAc:5}	16.75	16.90
		{Fuc:1; Hex:6; HexNAc:5}	16.93	16.93
		{Fuc:2; Hex:6; HexNAc:5}	16.08	16.54
		{Fuc:3; Hex:6; HexNAc:5}	15.51	16.21
		{Fuc:4; Hex:6; HexNAc:5}	11.44	15.58
		{Hex:7; HexNAc:5}	14.13	14.08
		{Fuc:2; Hex:4; HexNAc:6; Neu5NAc:3}	6.53	1.47
		{Fuc:3; Hex:5; HexNAc:6; Neu5NAc:2}	6.52	1.20
		{Hex:6; HexNAc:6}	12.92	11.18
		{Fuc:1; Hex:6; HexNAc:6}	17.39	11.71
		{Fuc:2; Hex:6; HexNAc:6}	13.64	11.07
		{Fuc:3; Hex:6; HexNAc:6}	4.15	9.80
		{Fuc:4; Hex:6; HexNAc:6; Neu5NAc:1}	6.51	6.96
		{Hex:7; HexNAc:6}	10.41	10.46
		{Fuc:1; Hex:7; HexNAc:6}	14.80	10.98
		{Fuc:2; Hex:7; HexNAc:6}	12.54	10.55
		{Fuc:3; Hex:7; HexNAc:6}	7.90	9.82
		{Fuc:2; Hex:5; HexNAc:7; Neu5NAc:3}	8.58	1.04
		{Fuc:3; Hex:6; HexNAc:7; Neu5NAc:2}	8.57	1.03
		{Fuc:2; Hex:7; HexNAc:7}	4.18	6.48
		{Fuc:3; Hex:7; HexNAc:7}	5.15	6.51
		{Fuc:4; Hex:7; HexNAc:7; Neu5NAc:1}	8.56	5.14
		{Fuc:1; Hex:8; HexNAc:7}	11.57	7.24
		{Fuc:2; Hex:8; HexNAc:7}	9.62	7.00
		{Fuc:3; Hex:8; HexNAc:7}	10.53	7.11
		{Fuc:4; Hex:8; HexNAc:7}	12.60	7.43
		{Hex:9; HexNAc:7}	10.32	4.05

Table 9: Search Results for 20141128-11-Phil-82. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 17.653258$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

Group	$\tau$	Glycan Compostion	Unregularized Score	Regularized Score
high-mannose	0.00	{Hex:5; HexNAc:4; Neu5NAc:2}	18.73	13.71
hybrid	7.67	{Fuc:1; Hex:5; HexNAc:4; Neu5NAc:2}	4.28	11.87
bi-antennary	15.20	{Hex:5; HexNAc:5; Neu5NAc:2}	9.36	16.09
asialo-bi-antennary	0.00	{Fuc:1; Hex:5; HexNAc:5; Neu5NAc:2}	6.60	15.61
tri-antennary	22.11	{Hex:6; HexNAc:5; Neu5NAc:2}	18.70	17.12
asialo-tri-antennary	0.00	{Hex:6; HexNAc:5; Neu5NAc:3}	20.89	17.53
tetra-antennary	14.57	{Fuc:1; Hex:6; HexNAc:5; Neu5NAc:2}	8.45	15.93
asialo-tetra-antennary	0.00	{Fuc:1; Hex:6; HexNAc:5; Neu5NAc:3}	19.75	17.33
penta-antennary	8.22	{Hex:7; HexNAc:5; Neu5NAc:1}	8.29	17.20
asialo-penta-antennary	0.00	{Hex:7; HexNAc:6; Neu5NAc:1}	8.48	5.81
		{Hex:7; HexNAc:6; Neu5NAc:2}	19.28	16.27
		{Hex:7; HexNAc:6; Neu5NAc:3}	19.70	16.11
		{Hex:7; HexNAc:6; Neu5NAc:4}	18.42	16.03
		{Fuc:1; Hex:7; HexNAc:6; Neu5NAc:2}	9.07	15.04
		{Fuc:1; Hex:7; HexNAc:6; Neu5NAc:3}	14.38	15.45
		{Fuc:1; Hex:7; HexNAc:6; Neu5NAc:4}	15.64	15.52
		{Fuc:2; Hex:7; HexNAc:6; Neu5NAc:3}	10.21	15.06
		{Fuc:2; Hex:7; HexNAc:6; Neu5NAc:4}	14.24	15.12
		{Fuc:3; Hex:7; HexNAc:6; Neu5NAc:4}	7.55	14.38
		{Hex:8; HexNAc:7; Neu5NAc:3}	7.12	10.75
		{Hex:8; HexNAc:7; Neu5NAc:4}	6.38	10.56
		{Fuc:1; Hex:8; HexNAc:7; Neu5NAc:4}	5.68	10.58

Table 10: Search Results for *AGP-DR-Perm-glycans-1*. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 16.783889$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

### 3.3 Reduced and Permethylated N-Glycans

The results for *AGP-permethylated-2ul-inj-55-SLens* are shown in Table 11 and visualized in Figure 8. The results for *Perm-BS-070111-04-Human-Serum* are shown in Table 12 and visualized in Figure 9.

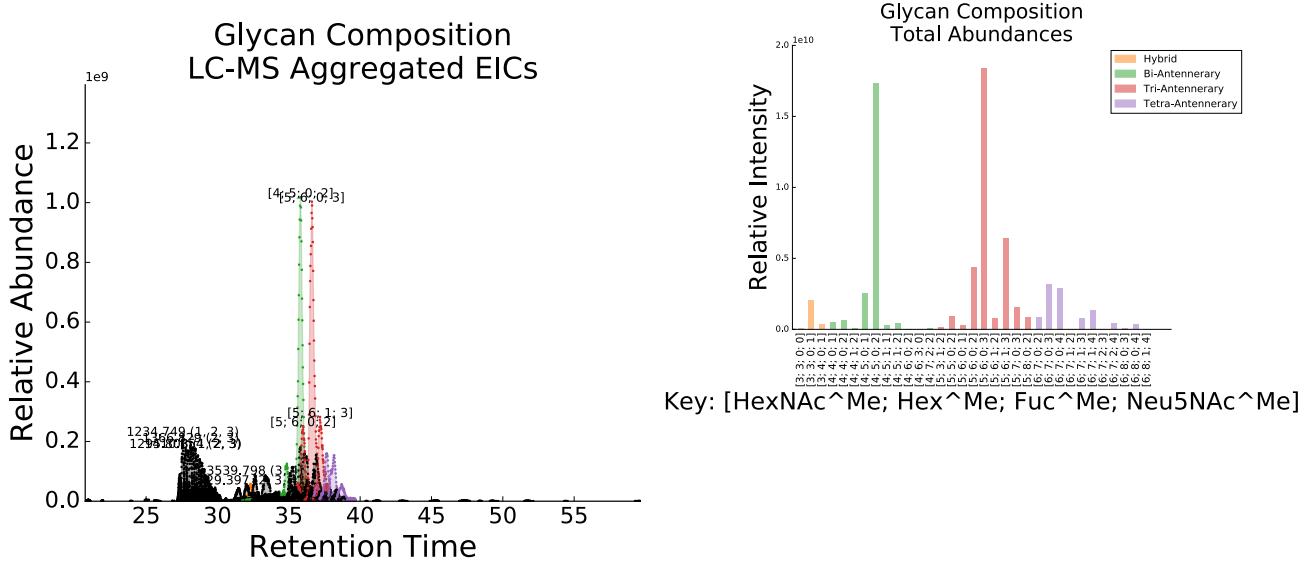


Figure 8: *AGP-permethylated-2ul-inj-55-SLens* Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	Unregularized Score	Regularized Score
high-mannose	0.00	{Hex:3; HexNAc:3}	13.17	16.87
hybrid	20.22	{Hex:3; HexNAc:3; Neu5Ac:1}	20.04	18.03
bi-antennary	18.50	{Hex:4; HexNAc:3; Neu5Ac:1}	17.30	17.75
asialo-bi-antennary	14.15	{Fuc:2; Hex:8; HexNAc:3}	5.85	1.01
tri-antennary	21.33	{Fuc:1; Hex:9; HexNAc:3}	1.37	0.29
asialo-tri-antennary	11.71	{Hex:4; HexNAc:4; Neu5Ac:1}	19.72	17.30
tetra-antennary	16.72	{Hex:4; HexNAc:4; Neu5Ac:2}	16.01	18.70
asialo-tetra-antennary	1.43	{Fuc:1; Hex:4; HexNAc:4; Neu5Ac:2}	3.49	3.49
penta-antennary	8.13	{Hex:5; HexNAc:4; Neu5Ac:1}	22.71	17.63
asialo-penta-antennary	0.00	{Hex:5; HexNAc:4; Neu5Ac:2}	24.88	19.63
		{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:1}	15.95	16.71
		{Fuc:1; Hex:5; HexNAc:4; Neu5Ac:2}	7.21	17.51
		{Hex:6; HexNAc:4; Neu5Ac:2}	3.53	3.53
		{Fuc:3; Hex:6; HexNAc:4}	9.22	14.52
		{Fuc:2; Hex:7; HexNAc:4; Neu5Ac:2}	nan	18.98
		{Fuc:1; Hex:8; HexNAc:4; Neu5Ac:2}	5.15	0.57
		{Fuc:2; Hex:9; HexNAc:4; Neu5Ac:1}	6.79	0.88
		{Fuc:3; Hex:10; HexNAc:4}	8.29	1.65
		{Fuc:1; Hex:3; HexNAc:5; Neu5Ac:2}	7.96	16.94
		{Hex:5; HexNAc:5; Neu5Ac:2}	13.88	18.23
		{Hex:6; HexNAc:5; Neu5Ac:1}	8.41	12.65
		{Hex:6; HexNAc:5; Neu5Ac:2}	21.65	18.71
		{Hex:6; HexNAc:5; Neu5Ac:3}	24.97	19.67
		{Fuc:1; Hex:6; HexNAc:5; Neu5Ac:2}	11.26	17.89
		{Fuc:1; Hex:6; HexNAc:5; Neu5Ac:3}	17.38	18.64
		{Hex:7; HexNAc:5; Neu5Ac:3}	15.78	18.73
		{Hex:7; HexNAc:6; Neu5Ac:2}	10.36	15.41
		{Hex:7; HexNAc:6; Neu5Ac:3}	16.80	16.03
		{Hex:7; HexNAc:6; Neu5Ac:4}	18.28	16.26
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:2}	9.17	15.42
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:3}	14.56	15.85
		{Fuc:1; Hex:7; HexNAc:6; Neu5Ac:4}	19.27	16.17
		{Fuc:2; Hex:7; HexNAc:6; Neu5Ac:3}	9.60	15.34
		{Fuc:2; Hex:7; HexNAc:6; Neu5Ac:4}	10.14	15.08
		{Hex:8; HexNAc:6; Neu5Ac:3}	8.75	12.08
		{Hex:8; HexNAc:6; Neu5Ac:4}	9.87	11.96
		{Fuc:1; Hex:8; HexNAc:6; Neu5Ac:4}	5.50	11.58
		{Hex:9; HexNAc:7}	4.48	nan

Table 11: Search Results for *AGP-permethylated-2ul-inj-55-SLens*. (a) Estimated  $\tau$  for grid with  $\hat{\gamma} = 16.167868$ . (b) Scores For Identified Glycans of grid with  $\hat{\lambda} = 0.99$ .

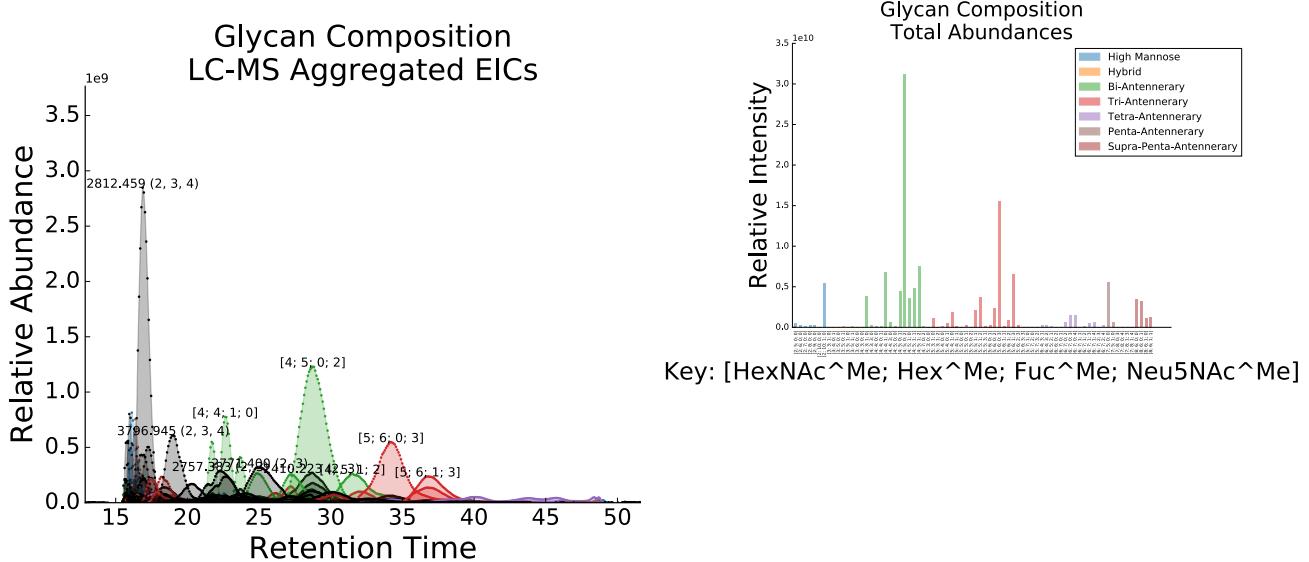


Figure 9: *Perm-BS-070111-04-Human-Serum* Glycan Relative Abundances

Group	$\tau$	Glycan Compostion	(b)	
			Unregularized Score	Regularized Score
high-mannose	20.10	{Hex:5; HexNAc:2}	22.94	20.45
hybrid	19.79	{Hex:6; HexNAc:2}	24.33	20.77
bi-antennary	18.47	{Hex:7; HexNAc:2}	21.66	20.63
asialo-bi-antennary	23.00	{Hex:8; HexNAc:2}	21.96	20.71
tri-antennary	22.98	{Hex:9; HexNAc:2}	22.02	20.31
asialo-tri-antennary	16.47	{Hex:10; HexNAc:2}	13.40	18.48
tetra-antennary	17.52	{Fuc:1; Hex:10; HexNAc:2}	15.92	18.94
asialo-tetra-antennary	5.92	{Fuc:1; Hex:3; HexNAc:3}	12.75	19.71
penta-antennary	10.69	{Hex:4; HexNAc:3; Neu5NAc:1}	6.86	17.61
asialo-penta-antennary	3.56	{Hex:5; HexNAc:3}	19.17	20.27
		{Hex:5; HexNAc:3; Neu5NAc:1}	17.49	19.02
		{Fuc:1; Hex:5; HexNAc:3}	13.17	19.70
		{Fuc:1; Hex:5; HexNAc:3; Neu5NAc:1}	6.76	17.97
		{Hex:6; HexNAc:3}	13.25	19.92
		{Fuc:2; Hex:8; HexNAc:3}	13.62	2.60
		{Fuc:2; Hex:10; HexNAc:3}	8.59	2.07
		{Hex:3; HexNAc:4}	18.15	20.31
		{Fuc:1; Hex:3; HexNAc:4}	25.84	21.54
		{Fuc:3; Hex:3; HexNAc:4}	16.38	20.21
		{Hex:4; HexNAc:4}	23.30	20.00
		{Hex:4; HexNAc:4; Neu5NAc:1}	14.16	18.76
		{Fuc:1; Hex:4; HexNAc:4}	25.87	20.72
		{Fuc:3; Hex:4; HexNAc:4}	18.07	19.41
		{Hex:5; HexNAc:4}	17.49	19.50
		{Hex:5; HexNAc:4; Neu5NAc:1}	22.56	19.81
		{Hex:5; HexNAc:4; Neu5NAc:2}	22.25	19.77
		{Fuc:1; Hex:5; HexNAc:4}	24.08	20.34
		{Fuc:1; Hex:5; HexNAc:4; Neu5NAc:1}	22.73	19.91
		{Fuc:1; Hex:5; HexNAc:4; Neu5NAc:2}	19.52	19.67
		{Hex:7; HexNAc:4; Neu5NAc:2}	nan	19.71
		{Fuc:2; Hex:9; HexNAc:4; Neu5NAc:1}	11.11	1.70
		{Fuc:3; Hex:9; HexNAc:4; Neu5NAc:1}	7.84	1.65
		{Fuc:3; Hex:10; HexNAc:4}	18.74	4.01
		{Hex:3; HexNAc:5}	12.52	20.91
		{Fuc:1; Hex:3; HexNAc:5}	24.17	22.85
		{Fuc:1; Hex:3; HexNAc:5; Neu5NAc:2}	6.88	16.70
		{Hex:4; HexNAc:5}	12.73	18.75
		{Hex:4; HexNAc:5; Neu5NAc:1}	20.78	19.50
		{Fuc:1; Hex:4; HexNAc:5}	23.25	20.26
		{Fuc:1; Hex:4; HexNAc:5; Neu5NAc:1}	10.53	18.77
		{Hex:5; HexNAc:5}	16.81	16.03
		{Hex:5; HexNAc:5; Neu5NAc:1}	16.35	17.07
		{Hex:5; HexNAc:5; Neu5NAc:2}	6.86	17.80
		{Fuc:1; Hex:5; HexNAc:5; Neu5NAc:1}	21.02	17.63
		{Fuc:1; Hex:5; HexNAc:5; Neu5NAc:2}	21.99	19.43
		{Fuc:3; Hex:5; HexNAc:5; Neu5NAc:1}	6.82	15.97
		{Hex:6; HexNAc:5; Neu5NAc:1}	14.77	16.50
		{Hex:6; HexNAc:5; Neu5NAc:2}	18.97	18.82
		{Hex:6; HexNAc:5; Neu5NAc:3}	23.46	20.04
		{Fuc:1; Hex:6; HexNAc:5; Neu5NAc:1}	10.92	16.06
		{Fuc:1; Hex:6; HexNAc:5; Neu5NAc:2}	11.56	18.34
		{Fuc:1; Hex:6; HexNAc:5; Neu5NAc:3}	24.83	19.92
		{Fuc:2; Hex:6; HexNAc:5; Neu5NAc:3}	11.41	18.06
		{Fuc:3; Hex:6; HexNAc:5; Neu5NAc:3}	8.65	17.66
		{Fuc:1; Hex:7; HexNAc:5}	5.29	10.24
		{Fuc:2; Hex:7; HexNAc:5}	6.54	10.34
		{Fuc:3; Hex:3; HexNAc:6}	6.57	0.83
		{Fuc:3; Hex:4; HexNAc:6; Neu5NAc:1}	9.45	17.52
		{Fuc:3; Hex:4; HexNAc:6; Neu5NAc:2}	4.93	20.56
		{Fuc:2; Hex:5; HexNAc:6; Neu5NAc:1}	7.87	13.83
		{Hex:6; HexNAc:6; Neu5NAc:2}	7.92	16.62
		{Fuc:1; Hex:6; HexNAc:6; Neu5NAc:4}	1.95	1.95
		{Hex:7; HexNAc:6; Neu5NAc:1}	4.73	10.47
		{Hex:7; HexNAc:6; Neu5NAc:2}	12.75	16.69
		{Hex:7; HexNAc:6; Neu5NAc:3}	16.51	17.53
		{Hex:7; HexNAc:6; Neu5NAc:4}	27.88	19.15

## 4 Discussion

### 4.1 Glycan Assignment Performance

#### 4.1.1 IgG

In *20151002-02-IGG* and *20150930-06-AGP* the known common glycoforms were assigned without ambiguity, both with and without formate adducts, confirmed manually. In these cases, network smoothing is not necessary. The estimates of  $\tau$  in *20151002-02-IGG* are low because the number of glycan compositions observed is low and the overlap in  $\mathbf{A}$  is large.

#### 4.1.2 Alpha-1 Acid Glycoprotein

In *20150930-06-AGP*, there are many more compositions to assign, which in turn leads to larger  $\tau$  estimates, with good support for all observations with the exception of {Hex:5; HexNAc:4; Neu5Ac:1} which depends upon the asialo-bi-antennary neighborhoods, and is also the only observed member of this neighborhoods. The deutero-reduced and permethylated sample *AGP-DR-Perm-glycans-1* shows similar distributions for tri-, tetra, and penta-antennary forms with low abundance of the bi-antennary forms.

The reduced and permethylated sample *AGP-permethylated-2ul-inj-55-SLens* has these groups as well, though it also shows higher ammoniation and supports the presence of hybrid N-glycans not found in the other AGP samples. Several of them are ambiguous assignments of the same signal due to the mass shift imposed by ammonium (17.026 Da, H3 N1) compared to a proton adduct, is nearly the same as the difference between a permethylated **Neu5Ac** and permethylated **FucHex** (17.015 Da, H3 C1 O1 N-1).

Without high quality tandem mass spectra, it is not possible to definitively resolve this ambiguity. As we do not require or interpret for our method, we conservatively assume that if the ambiguous adducted chromatogram's score is less than the score of the score of the chromatogram which uses it as an adduct of, its signal is reported for both the un-adducted and adducted chromatograms, otherwise the adducted chromatogram's signal is deducted from the unadducted chromatogram and is reported independently. Our tables show those cases where either the glycan composition is unambiguous, or the highest scoring member of an ambiguous chromatogram group.

#### 4.1.3 Influenza Strains

In *20141031-07-Phil-82*, the unregularized case contains some ambiguous matches where the biological context implies they should not be possible, {Hex:7; HexNAc:6; Neu5Ac:4}, or where the biological context and neighboring observations support the presence of a glycan composition but the evidence does not satisfy the scoring function, {Hex:10; HexNAc:2}, {Hex:6; HexNAc:5}, and {Fuc:1; Hex:7; HexNAc:6}. By applying the smoothing procedure with the parameters automatically estimated with grid search (Table 6), {Hex:7; HexNAc:6; Neu5Ac:4} was eliminated, while {Hex:10; HexNAc:2} and {Hex:6; HexNAc:5} were boosted into a higher confidence score range. The change to {Fuc:1; Hex:7; HexNAc:6} was insufficient to reach a high confidence score range, and the score for {Fuc:1; Hex:8; HexNAc:7} was dropped from a plausible score range to a low confidence range. Both these larger asialo-N-glycans have been previously assigned in Khatri *et al.* (2016a), but the automated procedure drops these compositions while estimating  $\hat{\gamma}$  leading to empty neighborhoods when estimating  $\tau$ . A user-generated  $\tau$  would contain a value greater than 5 in those neighborhoods.

This would be another table entry in `tbl:phil82_score_table`, already computed but uncertain how to integrate into the flow yet

We are justified in {Hex:7; HexNAc:6; Neu5Ac:4}'s removal based on its lack of supporting intermediary glycoforms and relatives above the selected  $\hat{\gamma}$ . If its own score exceeded  $\hat{\gamma}$ , it would itself result in a non-zero value for its related neighborhoods, providing itself with a non-zero minimum value and adjusting its rate of decay with  $\lambda$ . This is not to say that the LC-MS evidence that was observed is not real signal, merely that the assignment of that signal the glycan composition {Hex:7; HexNAc:6; Neu5Ac:4} is unlikely given the context.

The related *20141128-11-Phil-82* sample we see a similar pattern of glycoforms though with a wider range of fucosylation. In this case, we also have chemical noise from permethylation and a low degree of ammonium adduction, which can lead to more spurious matches. We match several multiply sialylated glycan compositions in this sample which do not satisfy any of the neighborhood conditions in Table 4, which results in their score decaying rapidly as  $\lambda \rightarrow 1$ . If these compositions were viable under the user's glycome model, then a different set of neighborhood rules would need to be specified which covered this group. In other cases, monosialylated compositions are either not eliminated or receive a larger score after smoothing because they are connected to the asialo neighborhoods, which are strongly supported in this glycome. We manually confirmed that the signal assigned to {Fuc:1; Hex:5; HexNAc:4; Neu5NAc:1} is not a deconvolution artefact, but the signal for {Fuc:2; Hex:5; HexNAc:4; Neu5NAc:1} is partially overlapped and difficult to manually separate. While this connection between monosialo and asialo forms may make sense in some cases, the link may not be appropriate for the biological context of this sample where a viral protein Neuraminadase removes **Neu5Ac**, and the user could redefine the neighborhoods to omit

that overlap. Similar commentary can be made for *20141101-04-Phil-BS* , which while not permethylated and ammoniated, shows the presence of sialylated compositions due to variable adduction.

## 4.2 Utility of Network Smoothing

## References

- Akune, Y., Lin, C.-H., Abrahams, J. L., Zhang, J., Packer, N. H., Aoki-Kinoshita, K. F., and Campbell, M. P. (2016). Comprehensive analysis of the N-glycan biosynthetic pathway using bioinformatics to generate UniCorn: A theoretical N-glycan structure database. *Carbohydrate Research*, **431**, 56–63.
- Belkin, M., Niyogi, P., and Sindhwan, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, **7**(2006), 2399–2434.
- Goldberg, D., Bern, M., North, S. J., Haslam, S. M., and Dell, A. (2009). Glycan family analysis for deducing N-glycan topology from single MS. *Bioinformatics*, **25**(3), 365–371.
- Kaur, P. and O'Connor, P. B. (2006). Algorithms for automatic interpretation of high resolution mass spectra. *Journal of the American Society for Mass Spectrometry*, **17**(3), 459–468.
- Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*, **24**(21), 2534–2536.
- Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Lemarie, N., Woods, R. J., Hartshorn, K. L., Zaia, J., Leymarie, N., Woods, R. J., Hartshorn, K. L., and Zaia, J. (2016a). Integrated omics and computational glycobiology reveal structural basis for Influenza A virus glycan microheterogeneity and host interactions. *Molecular & cellular proteomics : MCP*, **13975**(615).
- Khatri, K., Klein, J. A., and Zaia, J. (2016b). Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Analytical and Bioanalytical Chemistry*.
- Kronewitter, S. R., Slysz, G. W., Marginean, I., Hagler, C. D., LaMarche, B. L., Zhao, R., Harris, M. Y., Monroe, M. E., Polyukh, C. A., Crowell, K. L., Fillmore, T. L., Carlson, T. S., Camp, D. G., Moore, R. J., Payne, S. H., Anderson, G. a., and Smith, R. D. (2014). GlyQ-IQ: Glycomics quintavariate-informed quantification with high-performance computing and glycagrid 4D visualization. *Analytical Chemistry*, **86**(13), 6268–6276.
- Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C., Edwards, N., Souda, P., Whitelegge, J. P., Bafna, V., and Pevzner, P. A. (2010). Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Molecular & cellular proteomics : MCP*, **9**(12), 2772–2782.
- Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W. H., Römpf, A., Neumann, S., Pizarro, A. D., Montecchi-Palazzi, L., Tasman, N., Coleman, M., Reisinger, F., Souda, P., Hermjakob, H., Binz, P.-a., and Deutsch, E. W. (2011). mzML—a community standard for mass spectrometry data. *Molecular & cellular proteomics : MCP*, **10**(1), R110.000133.
- Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., Conley, S., Staples, G. O., Slysz, G. W., Smith, R. D., and Zaia, J. (2012). GlycReSoft: a software package for automated recognition of glycans from LC/MS data. *PloS one*, **7**(9), e45474.
- Senko, M. W., Beu, S. C., and McLafferty, F. W. (1995). Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry*, **6**(4), 229–233.
- Stanley, P., Schachter, H., and Taniguchi, N. (2009). *N-Glycans*. Cold Spring Harbor Laboratory Press.
- Varki, A. and Schauer, R. (2009). *Sialic Acids*. Cold Spring Harbor Laboratory Press.
- Yu, C.-Y. C.-Y., Mayampurath, A., Hu, Y., Zhou, S., Mechref, Y., and Tang, H. (2013). Automated annotation and quantification of glycans using liquid chromatography-mass spectrometry. *Bioinformatics*, **29**(13), 1706–1707.
- Yu, T. and Peng, H. (2010). Quantification and deconvolution of asymmetric LC-MS peaks using the bi-Gaussian mixture model and statistical model selection. *BMC bioinformatics*, **11**(1), 559.