Tutorial 2

Aim: Guides the user though the input process when they have determined the mass and charge states themselves. In other words mass is not extracted from the raw data but instead provided by the user. Shows how we can assign all peaks with one button click. Figure 1 shows the monoclonal antibody exoglycosidase array which we will work with.

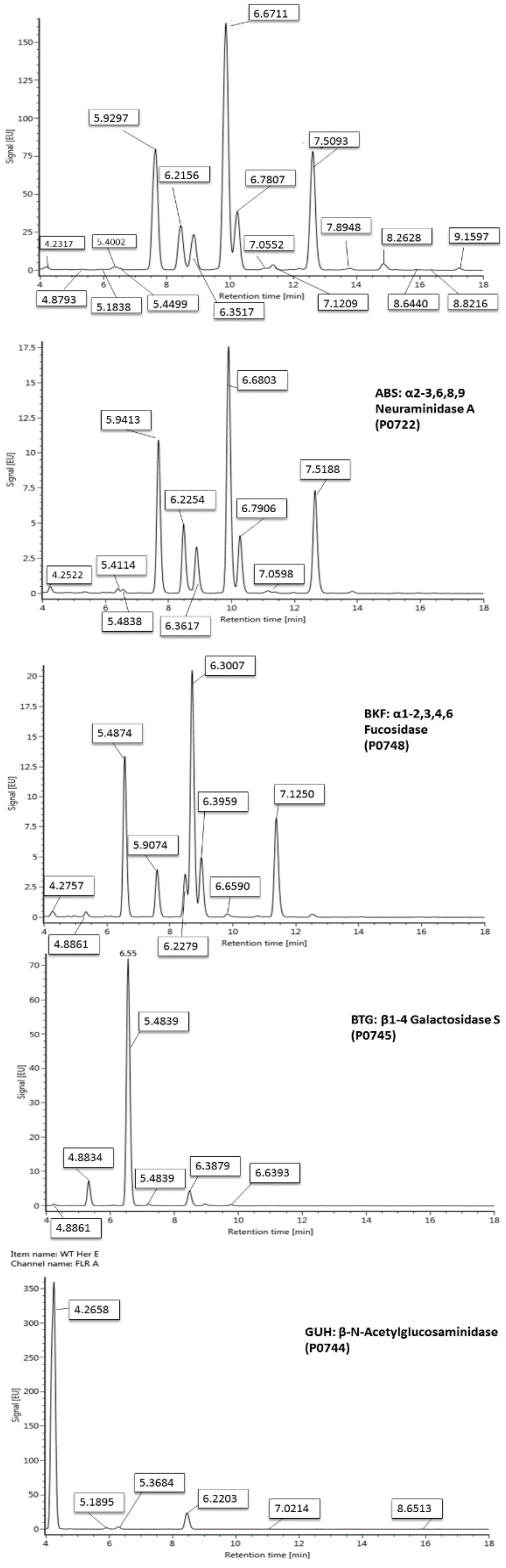


Figure 1. An ABS+BKF+BTG+GUH exoglycosidase array (UPLC). There are three variables: retention time, area of each peak and Glucose Units (boxes). Notice that some peaks have very small size (low abundance) and some are large (high abundance).

Abbreviations:

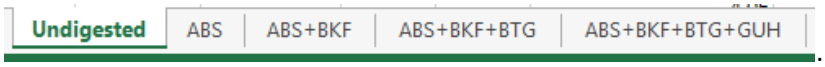
UPLC: Ultra-performance liquid chromatography

LC: Liquid chromatography

MS: Mass spectrometry

GU: Glucose Units

Step 1: An example dataset is located on the GlycanAnalyzer tutorial webpage called Data_Expert_mass.xls. Open the spreadsheet titled Data_Expert_mass.xls and examine the data inside each exoglycosidase pane (Note that it is the same data presented in Figure 1):



Note: if users are supplying their own datasets then they must have it in exactly the same format as Data_Expert_nomass.xls

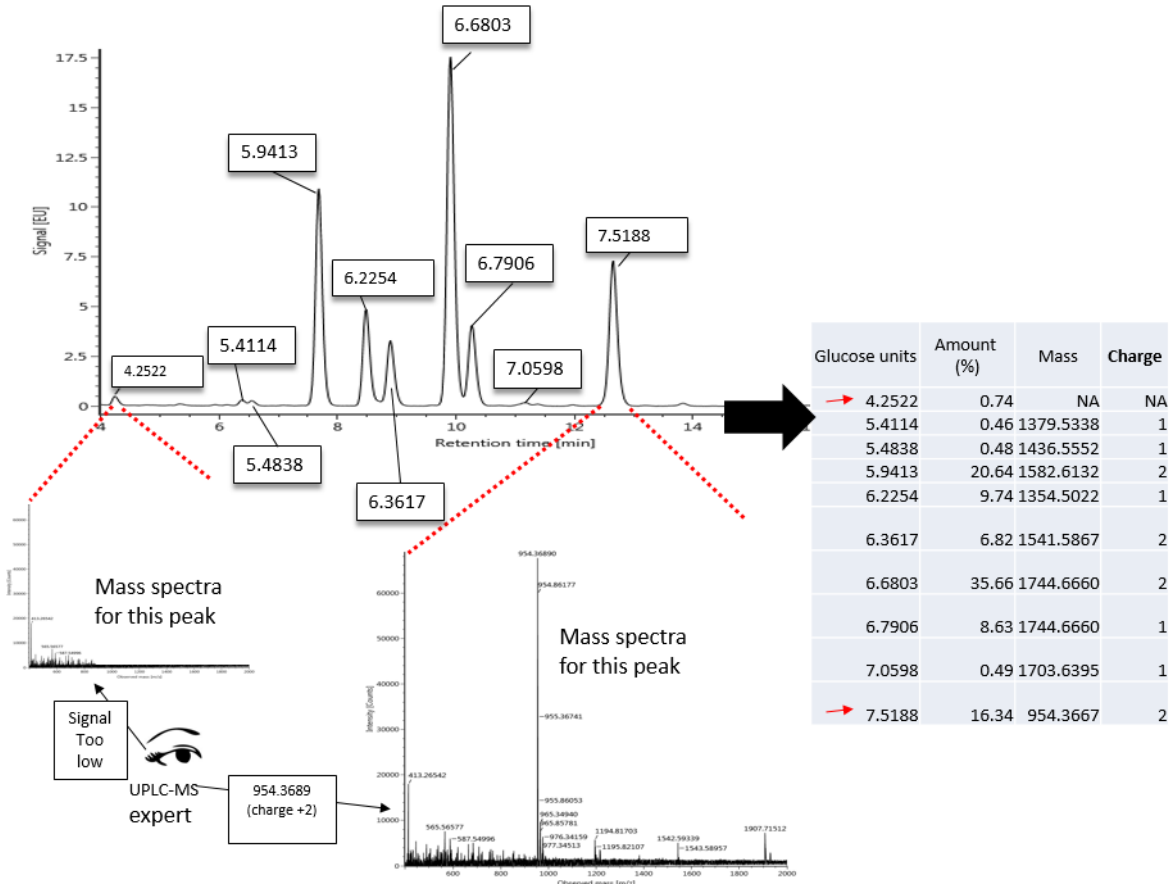
Step 2: Notice that the LC data (Fig 1) is tabulated in the spreadsheet under the columns "GU", "Amount (%)", and the MS is summarized in the tabs "Observed Mass" and "Observed charge":

GU	Amount (%)	Observed Mass	Observed charge
4.2317	0.39	1030.39653	1
4.8793	0.13	1233.4759	1
5.1838	0.15	1192.44935	1
5.4002	0.5	1379.53381	1
5.4499	0.16	NA	NA
5.9297	18.85	1582.61318	2
6.2156	6.67	1354.50218	1
6.3517	5.54	1541.58663	2
6.6711	37.6	1744.66601	2
6.7807	8.92	1744.66601	1
7.0552	0.27	1516.555	1
7.1209	0.81	1703.63946	1
7.5093	18.07	1906.71883	2
7.8948	0.37	1678.60782	1
8.2628	1.1	2197.81425	2
8.644	0.11	1840.66065	1
8.8216	0.07	1840.66065	1
9.1597	0.31	2488.90966	2

Each peak has now been annotated with a mass and a charge. The user has determined these themselves.

Note that in this case the 2AB label is added to the masses. Also note that one peak in the undigested profile with a very low relative abundance could not be determined ('NA' in Observed Mass and Observed charge columns).

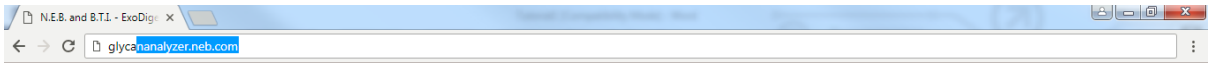
Each peak is therefore defined using four variables as showing in Figure 3.



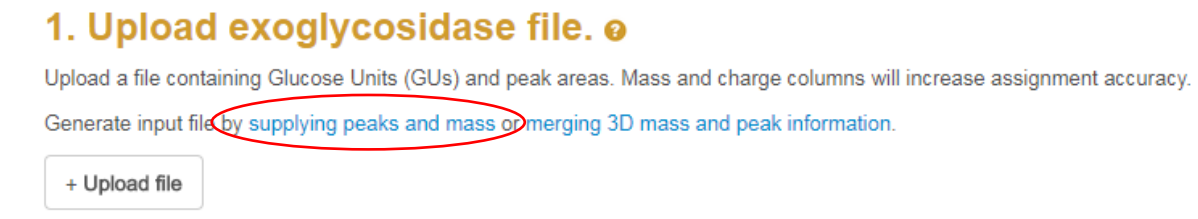
Glucose units	Amount (%)	Mass	Charge
4.2522	0.74	NA	NA
5.4114	0.46	1379.5338	1
5.4838	0.48	1436.5552	1
5.9413	20.64	1582.6132	2
6.2254	9.74	1354.5022	1
6.3617	6.82	1541.5867	2
6.6803	35.66	1744.6660	2
6.7906	8.63	1744.6660	1
7.0598	0.49	1703.6395	1
7.5188	16.34	954.3667	2

Figure 2. Peak input with mass observed by the user.

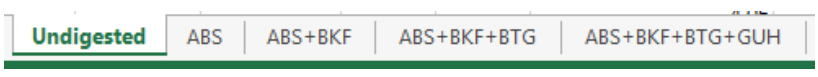
Step 3: Load GlycanAnalyzer web application in your browser. Type http://glycananalyzer.neb.com



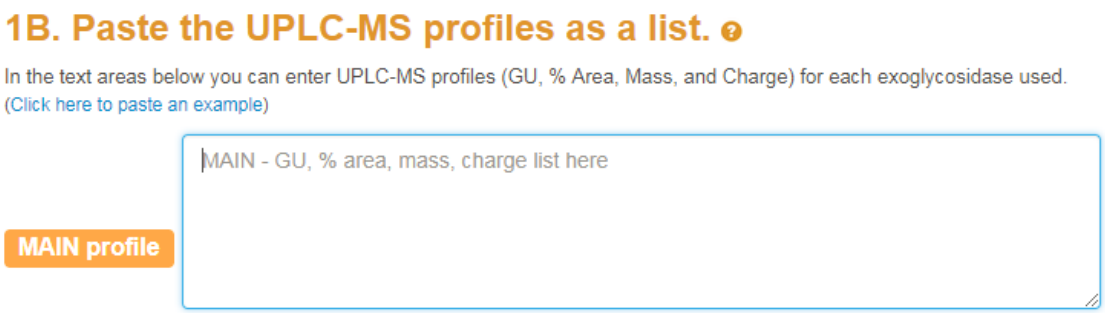
Step 4: We want to generate input for GlycanAnalyzer. Click the link circled on the main page:



Step 5: in the undigested data tab in our spreadsheet:



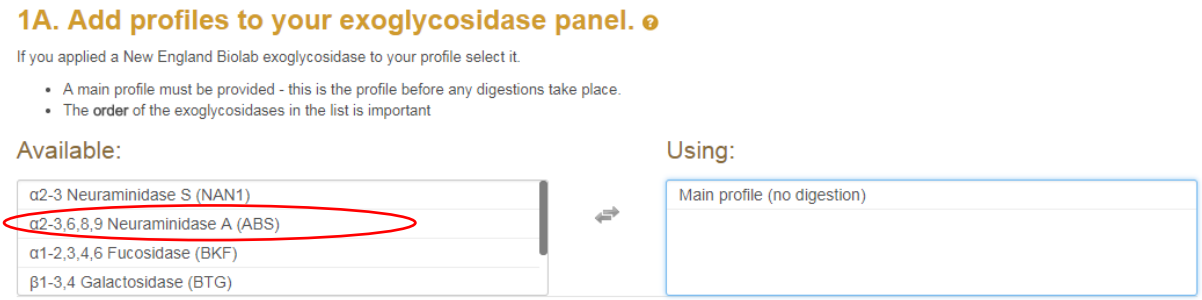
Copy the undigested data into the main profile.:



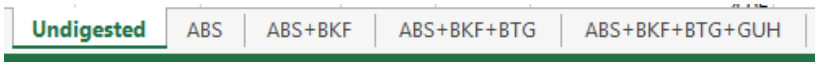
This is the main profile on which exoglycosidase digestions will be applied and glycan annotations will be assigned.

Note that the columns GU, Amount (%), Observed Mass and Observed charge should be separated by tabs. Copy and pasting from spreadsheets automatically separates by tabs.

Step 6: Add data from the spreadsheet by adding exoglycosidases **in order** of application (order ABS, BKF, BTG and GUH – see Figure 1). To add ABS click here:

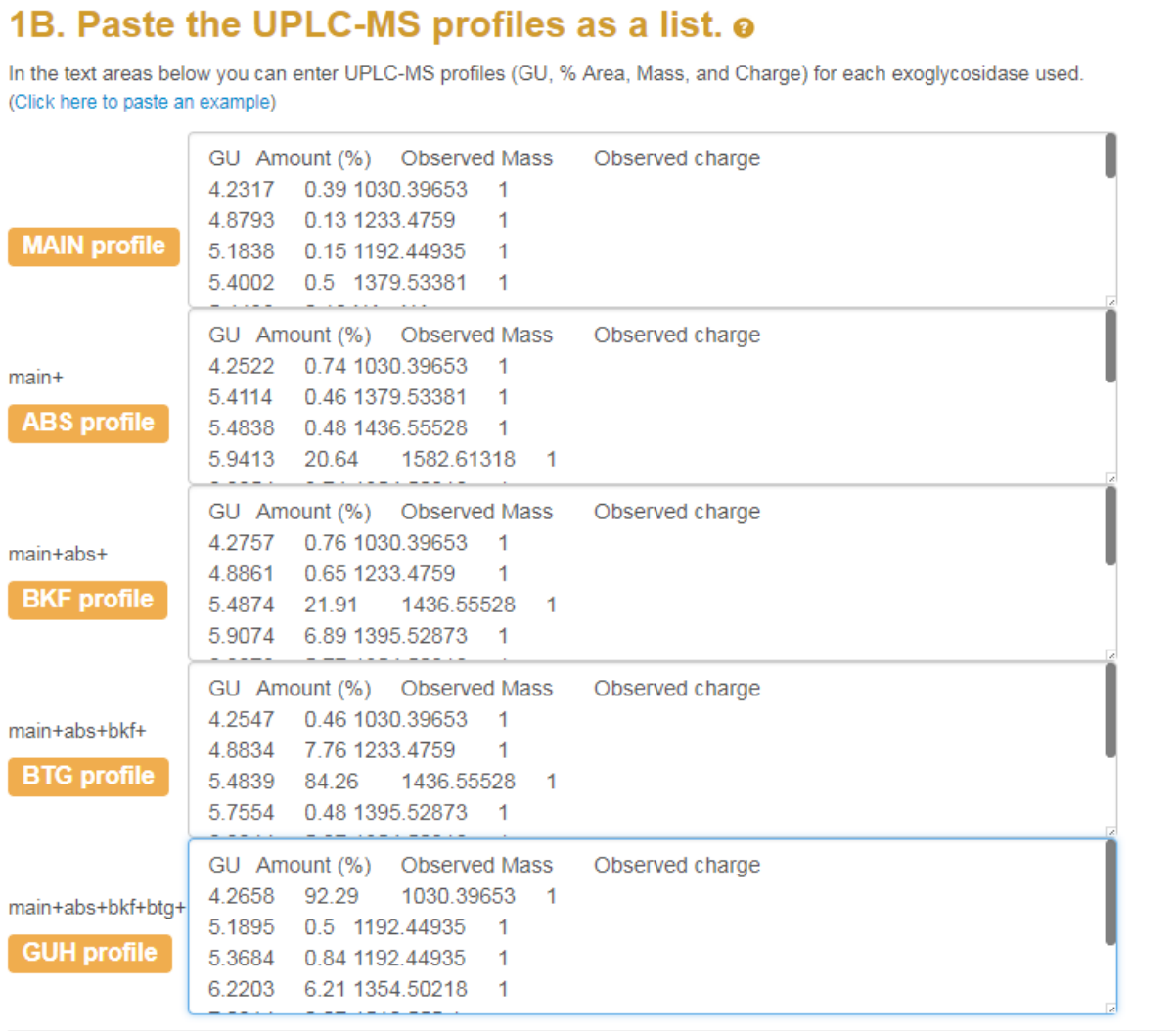


Then add the data in the ABS tab below exactly as in step 5:



and repeat for BKF, BTG and GUH enzymes as ordered in Figure 1.

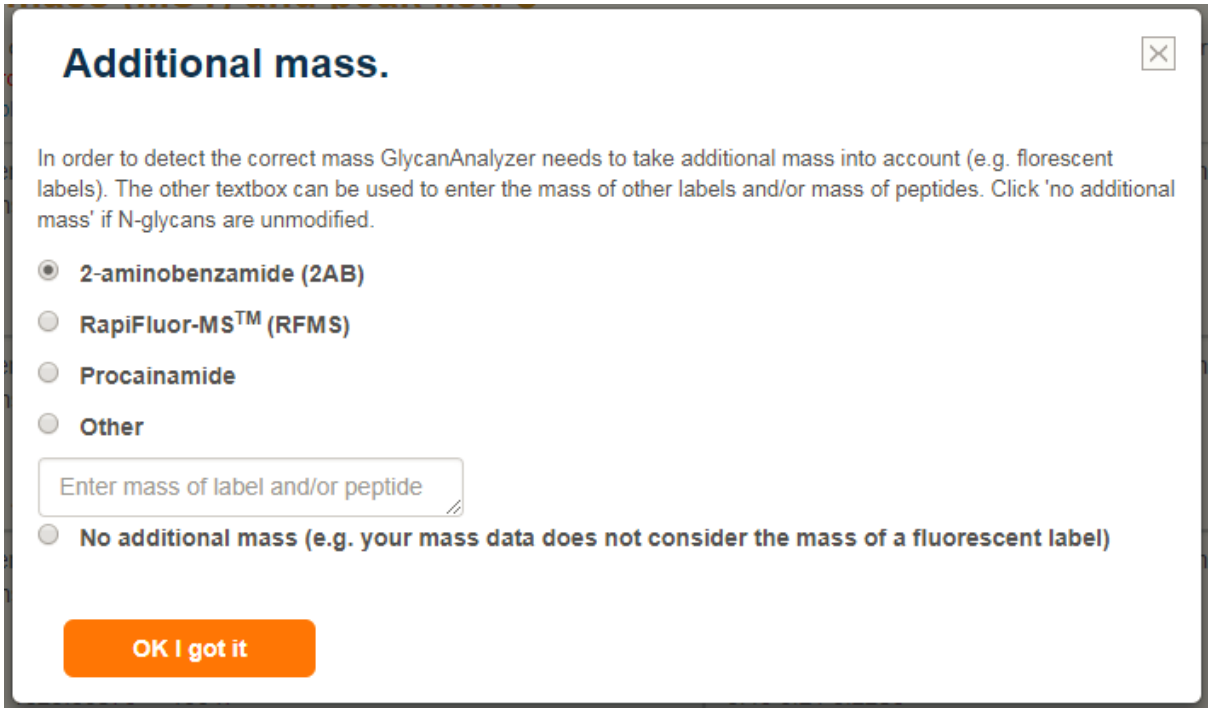
Step 7: The final data page should resemble the following:



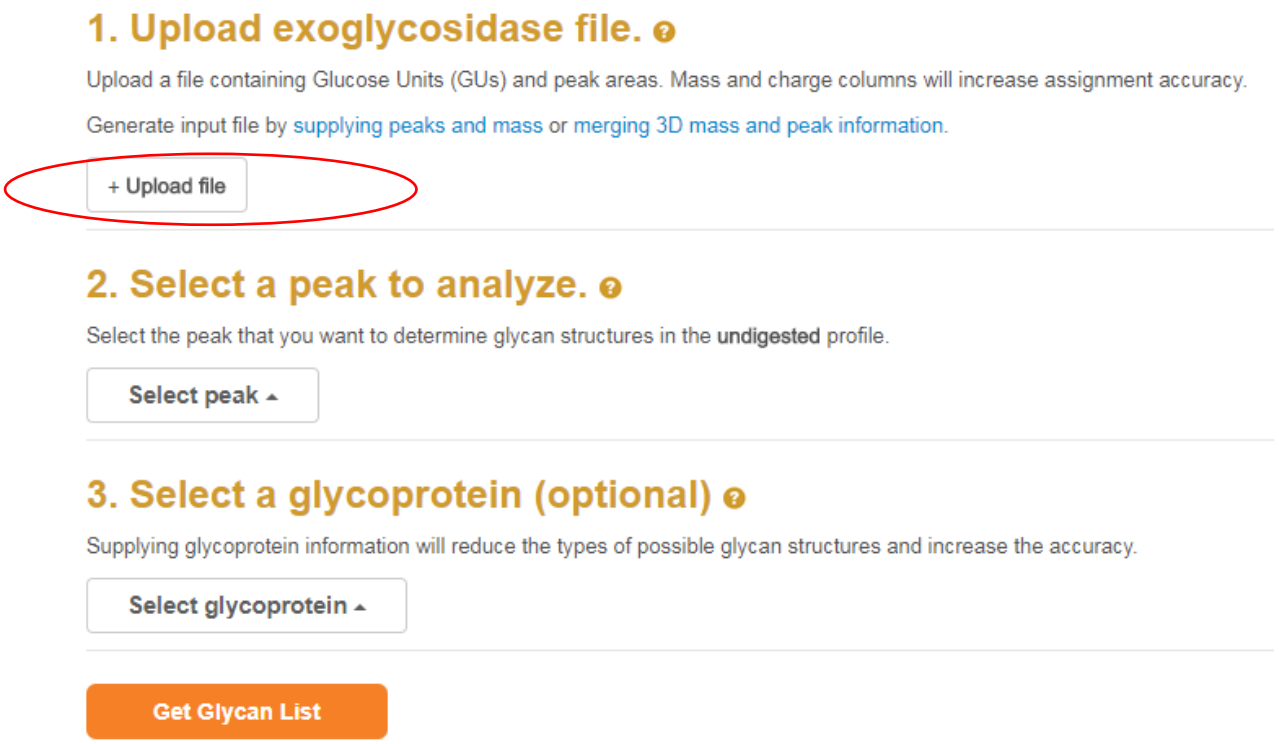
Click the button 

and save the file to your PC.

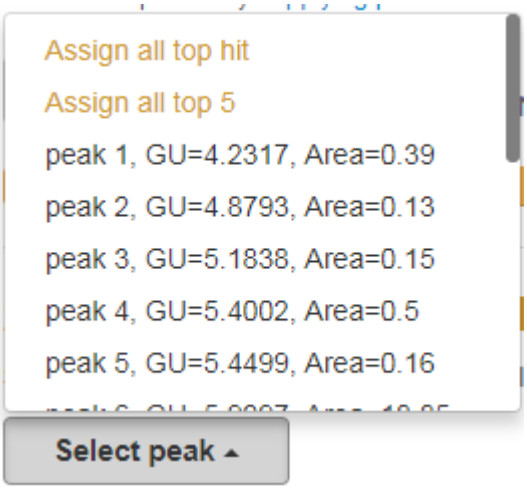
After clicking 'Merge data to input file' the user is asked if their MS data contains a florescent label. Users should select one of three things: (i) select a label supported by GlyanAnalyzer (2AB, RFMS, procanaminde), (ii) select 'other' and enter the labels mass explicitly or (iii) select label free. In this tutorial our dataset was generated using the 2AB label and we therefore should make sure this button is clicked:



Step 8: Once the input is generated you will be directed back to the main input page. On the main input page upload the file generated in step 7 by clicking the button circled:



Step 9: In this tutorial we will assign all peaks to the chromatogram. Select 'Assign all top 5' in the select peak drop up menu seen below.



Step 10: Click the button 

and wait for the assignment to complete. This computation is quite demanding and it may take 10-20minutes to return assignments. However, your PC is not being used the calculations are taking place on our servers at new England Biolabs. Note that we did not 'Selct glycoprotein' on the input page.

Step 11: Figure 4 shows assigned N-glycans summary page.

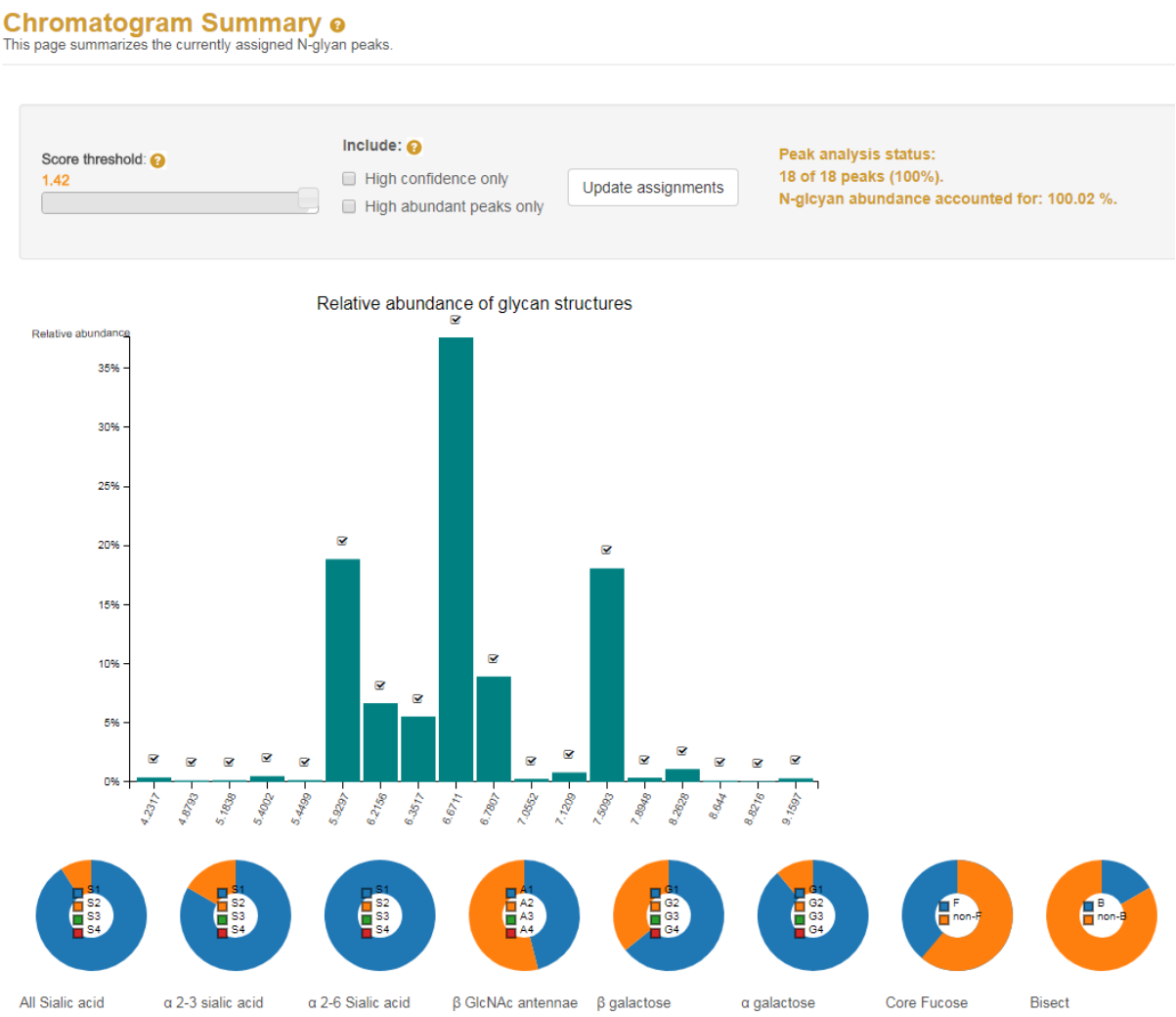


Figure 4. . The summary output page for the monoclonal antibody. This view is returned when all peaks are computed. The bar chart shows each peaks relative abundance where a tick indicates user acceptance of the peak. The pie charts give the distribution of sialic acids, GlcNAc antennae, galactose and other monosaccharides. The tables present the N-glycans annotated in the accepted peaks. If we take as an example Peak 13 and 14. Peak 13 contains the best possible sources of information: matching mass, matching GU and mass and GU shifts. Peak 14 has only two pieces of evidence: matching mass and matching GU, there were no shifting peaks found due to its small size. Peak 13's N-glycan assignment can be considered to have strong support while peak 14's has medium support. The weakest level of supporting evidence is GU similarity alone. The Mass, Shifts and Glucose unit buttons can be clicked to visualize the evidence.

Step 12: Each peak now has an assignment and a corresponding table of ranked glycans. In fact, there should be at most 5 assignments all ranked by the score in each peak. However, not all 5 should be accepted by the user – see step 13.

Step 13: The glycans ranked in the bottom 2 of peak 15 in (Figure 5) should definitely be rejected as the mass was not matched very well. Once they are rejected by clicking the 'x' button all the

statistics are updated. Since there is no mass spectra uploaded no mass spectra images can be generated the best we can do is match the mass supplied by the user. This can be seen in the 'Details' column (Figure 5). Note that the masses are not exactly matched and may differ slightly because the mass in our dataset is the isotopic average mass. Secondly, the M7 mannose structure ranked 1st, 2nd and 3rd, in Figure 5, are all isomers and perhaps one can accept that it is either M7 or M7 D1 since the GU values are very close to the ones observed in the database. We could reject M7 D3.

Peak 14 at 7.8948 GUs

Reject	Score	Oxford notation	Diagram	Details	More information	Evidence
☒	0	M7	GlycoCT encoding	**Mass (Observed / Expected):** 1678.61 / 1679.46 **GU (Observed / Expected):** 7.8948 / 7.8983 ΔGU :0.0035	1 2 3 4 5 GlycoStore 1	Mass Shifts Glucose units
☒	0.07	M7 D1	GlycoCT encoding	**Mass (Observed / Expected):** 1678.61 / 1679.46 **GU (Observed / Expected):** 7.8948 / 7.9685 ΔGU :0.0737	1 2 3 4 5 GlycoStore 1	Mass Shifts Glucose units
☒	0.13	M7 D3	GlycoCT encoding	**Mass (Observed / Expected):** 1678.61 / 1679.46 **GU (Observed / Expected):** 7.8948 / 7.763 ΔGU :0.1318	1 2 3 4 5 GlycoStore 1	Mass Shifts Glucose units
☒	1.02	M4A1G(4)1S(6)1	GlycoCT encoding	**Mass (Observed / Expected):** 1678.61 / 1849.63 **GU (Observed / Expected):** 7.8948 / 7.916 ΔGU :0.0212	1 2 3 GlycoStore 1	Mass Shifts Glucose units
☒	1.02	A2G(4)2Ga(3)1	GlycoCT encoding	**Mass (Observed / Expected):** 1678.61 / 1923.71 **GU (Observed / Expected):** 7.8948 / 7.9122 ΔGU :0.0174	1 2 GlycoStore 1	Mass Shifts Glucose units

Figure 5. Assignments for peak 14 in our data example. All the glycan masses in peak 14 match except for the bottom ranked two.

Step 14: The following information is available.

1. If a user clicks on a bar (in Figure 4) they will be presented with a peak shift graph. This presents a directed graph showing the estimated peak movement (Figure 6). The size of a circle is proportional to the area of the peak. Clicking the circles reveals the glycan structure assigned to the peak in each digestion profile.

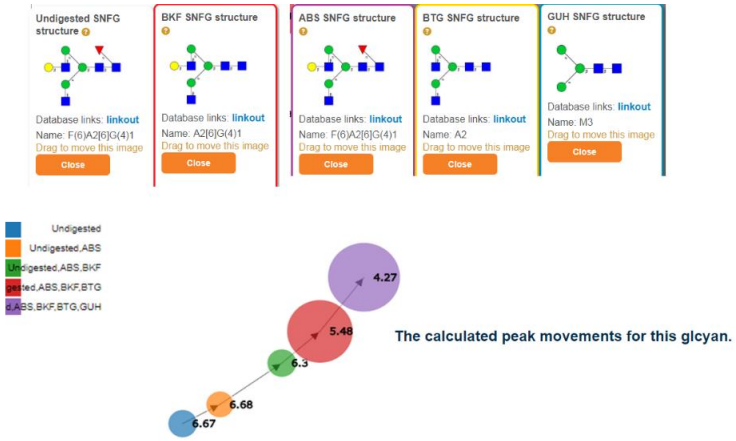


Figure 6. The directed graph representation of the peak movements.

2. 'Score calculation' (Figure 6): This is perhaps the most important information revealed by GlycanAnalyzer. It reveals the peak shift in the H/UPLC chromatograms. However, because the user supplied their own observed masses then there will be no images generated for the m/z shifts. Also, the score does not take mass into account.

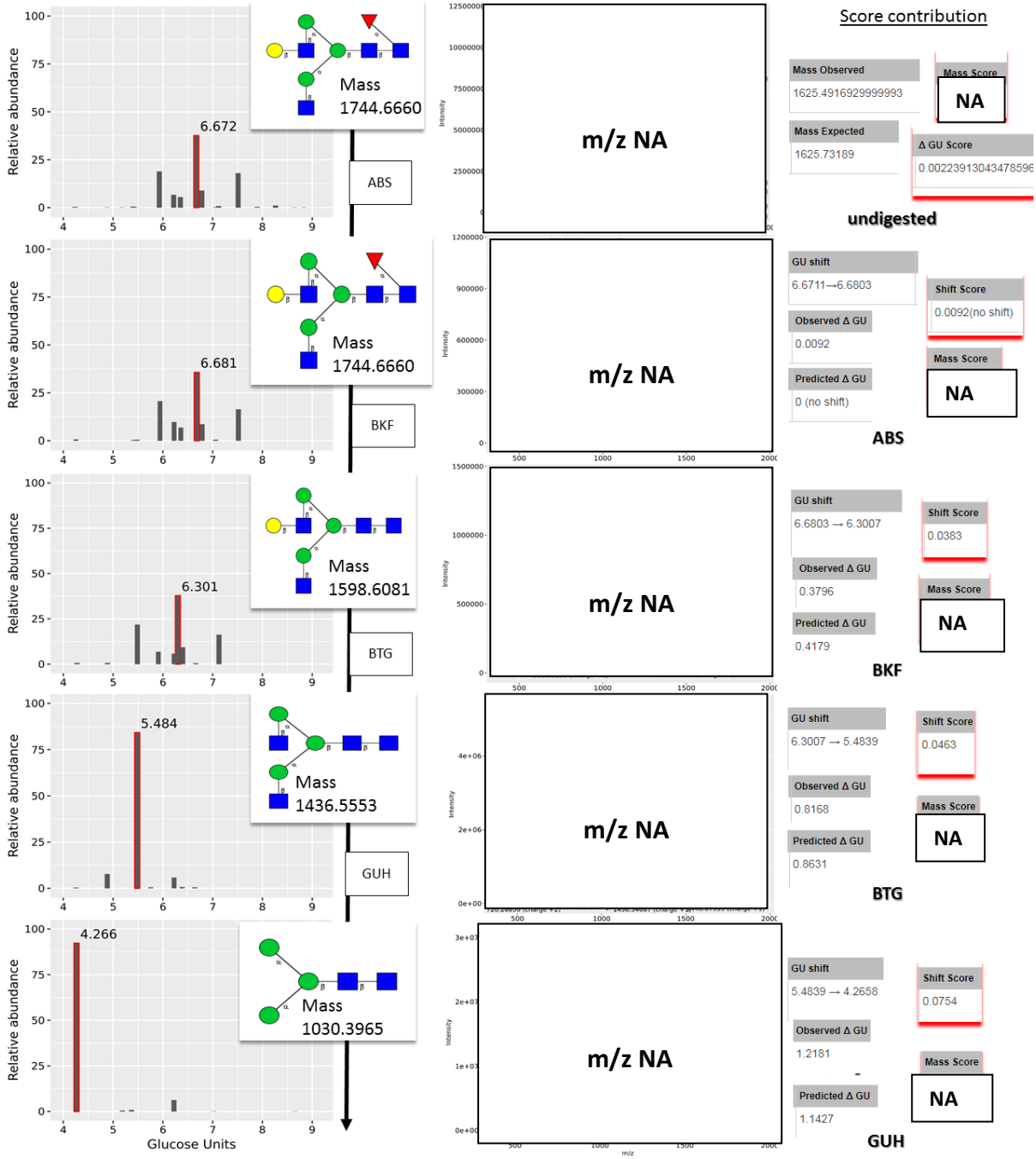


Figure 6. Score calculation for peak 9 of the monoclonal antibody. The total score is 0.1714 coming from the following contributions: ΔGU = 0.0022, Mass score = NA, Shift score= 0.0092 + 0.0383 + 0.0463 + 0.0754. Total score is 0.0022 + 0 + 0.0092 + 0.0383 + 0.0463 + 0.0754.