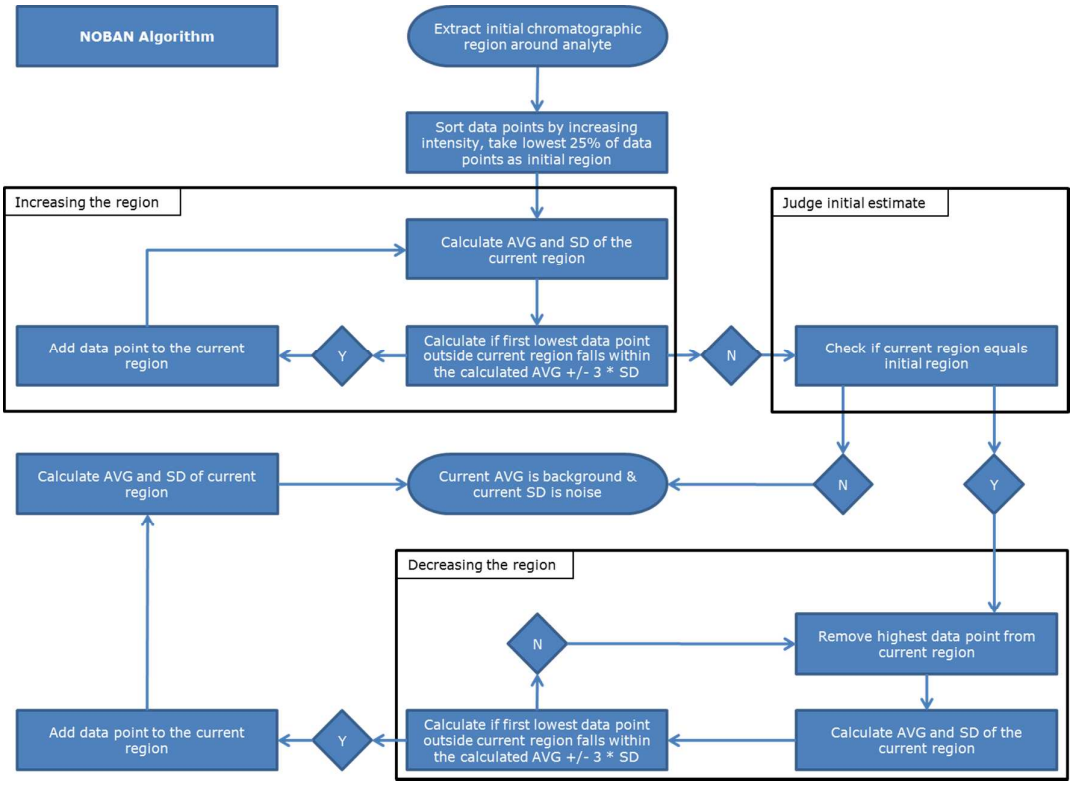
Supporting Information "LaCyTools – a targeted LC-MS data processing package for relative quantitation of glycopeptides"

Bas C. Jansen[1], David Falck[1], Noortje de Haan[1], Agnes L. Hipgrave Ederveen[1], Genadij Razdorov[2], Gordan Lauc[2], Manfred Wuhrer[1]

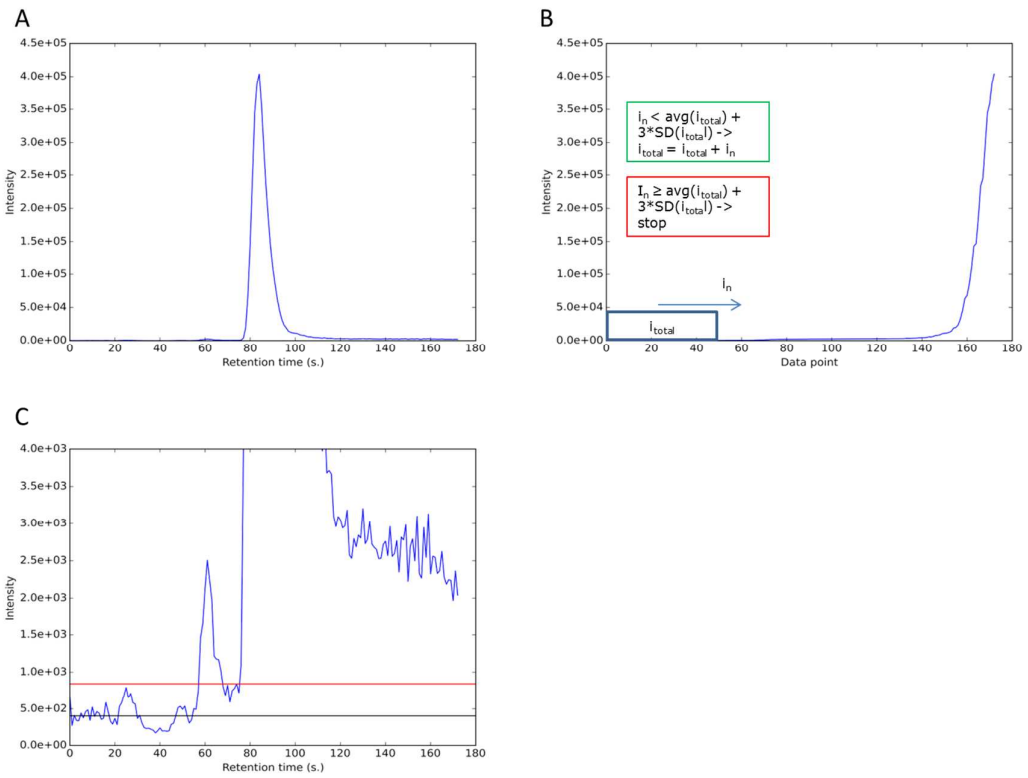
[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300RC Leiden, The Netherlands

[2]Department of Biochemistry and Molecular Biology, Faculty of Pharmacy and Biochemistry, University of Zagreb, A. Kovačića 1, HR10000, Zagreb, Croatia

Table of Contents

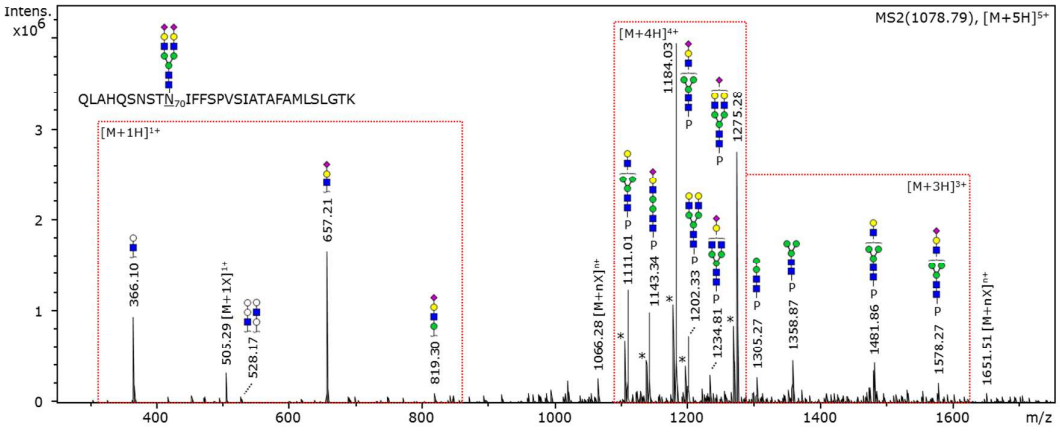


Supporting Information, Figure S-1. NOBAN algorithm decision workflow. The normal distribution based background and noise determination (NOBAN) algorithm calculates the initial estimate for the background and noise based on the *m/z* region around an analyte. The initial estimate is determined using the lowest 25% of the data points, subsequently the algorithm will determine if the next lowest data point falls within the current average (AVG) + 3 * standard deviation (SD). This data point will be added to the current region if it does, after which the process is repeated until this is no longer true. Alternatively, if the next lowest data point fell outside the AVG + 3 * SD of the initial estimate, the highest data point is removed from the current region. This process is repeated until the removed data point falls within AVG + 3 * SD, at which point the removed data point is added again. The calculated AVG and SD are taken as the background and noise values.

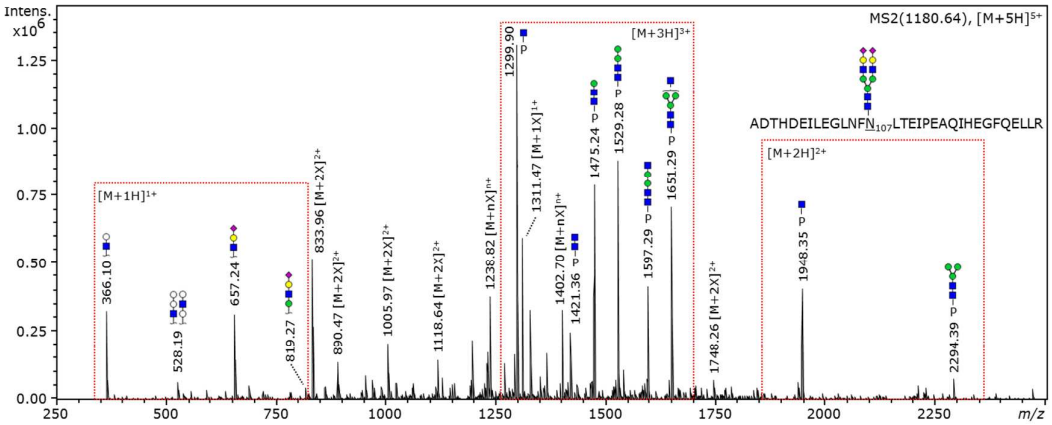


Supporting Information, Figure S-2. Determination of background and noise in the retention time component (LC) of an analysis of AAT. (**A**) Extracted ion chromatogram (EIC) of the main isotope of the N_{107}-H5N4S2 glycoform of alpha-1-antitrypsin (AAT), extracted with a window of ± 0.05 Th. (**B**) EIC sorted by increasing intensity, of the main isotope of N_{107}-H5N4S2. (**C**) EIC of the main isotope of N_{107}-H5N4S2 showing the calculated background and 3 * noise values as the black and red lines, respectively.

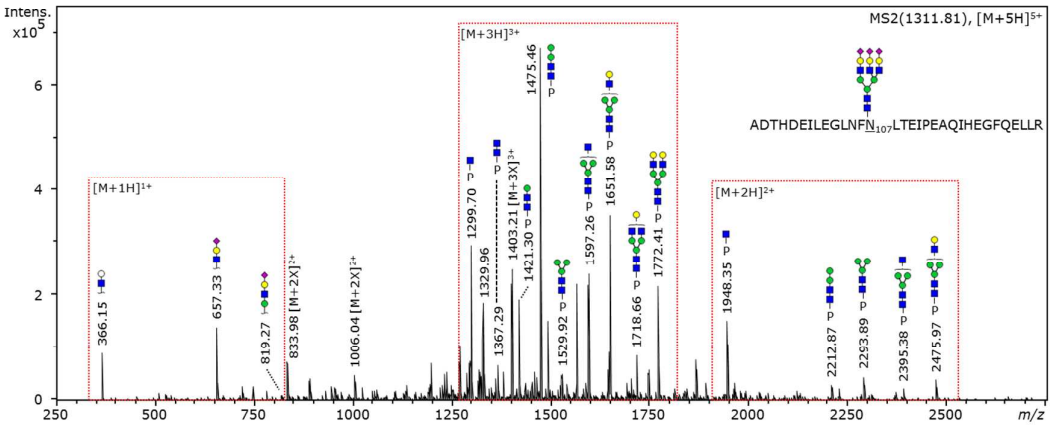
A


MS2(1078.79), [M+5H]5+
QLAHQSNSTN70IFFSPVSIATAFAMLSLGTK
[M+1H]1+
[M+4H]4+
[M+3H]3+
Intens.
x10^6
m/z

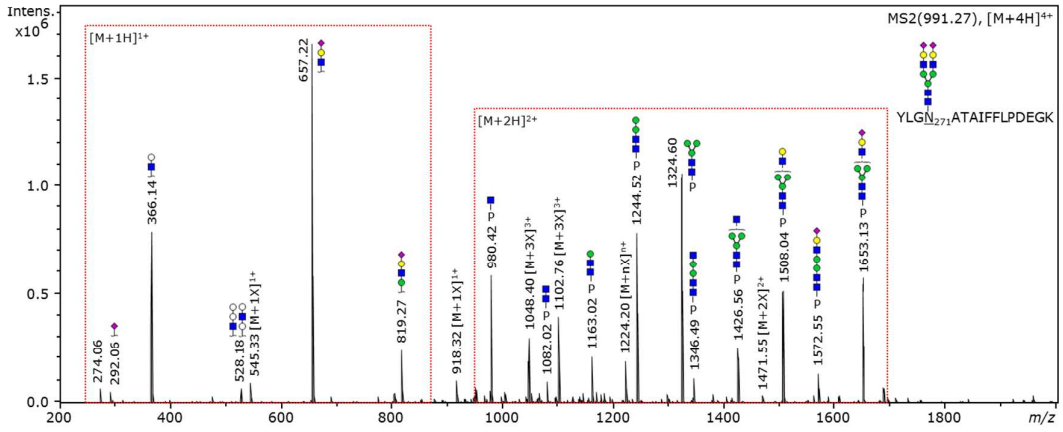
B


MS2(1180.64), [M+5H]5+
ADTHDEILEGLNFN107LTEIPEAQIHEGFQELLR
[M+1H]1+
[M+3H]3+
[M+2H]2+
Intens.
x10^6
m/z

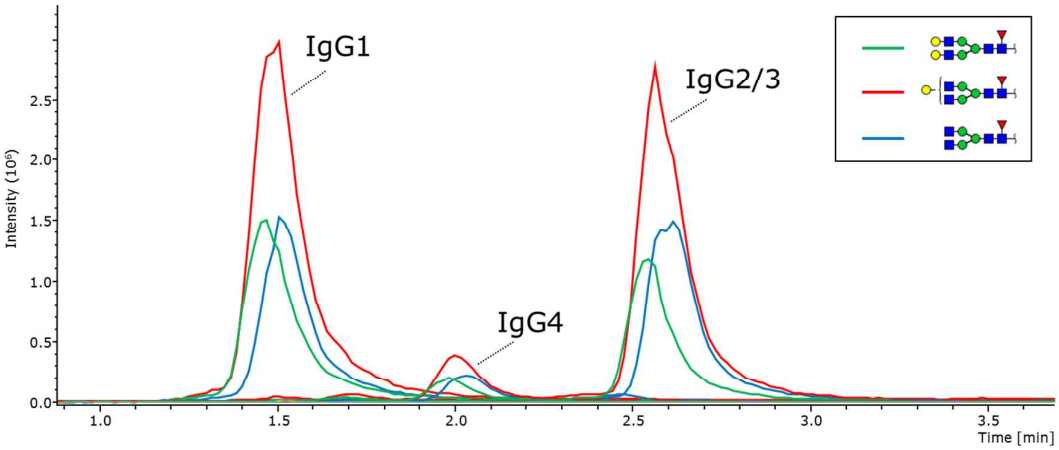
C


MS2(1311.81), [M+5H]5+
ADTHDEILEGLNFN107LTEIPEAQIHEGFQELLR
[M+1H]1+
[M+3H]3+
[M+2H]2+
Intens.
x10^5
m/z

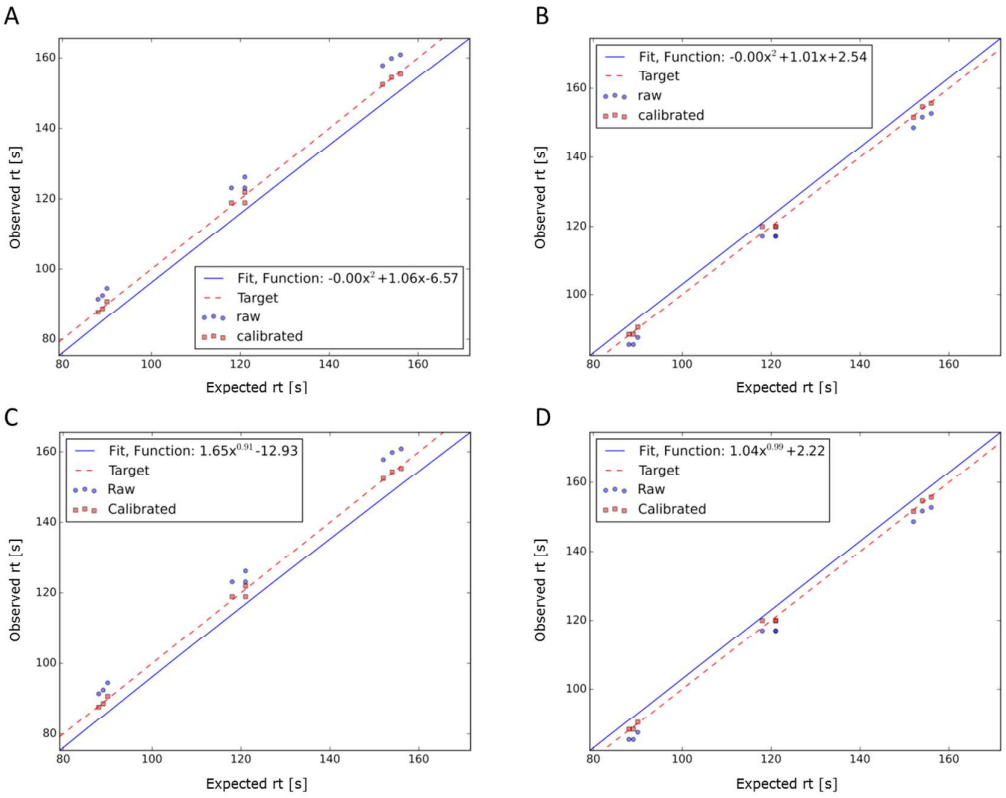
D



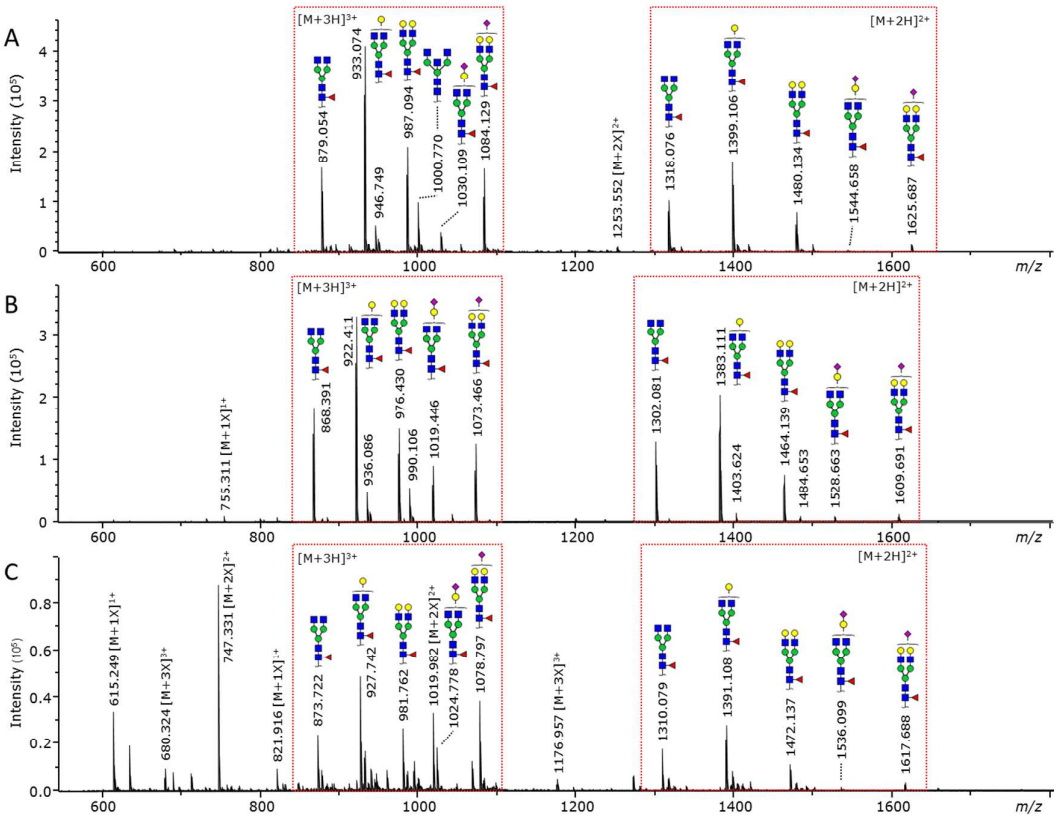
Supporting Information, Figure S-3. AAT glycopeptide fragmentation. (**A**) MS^2 spectrum of the N_{70}-H5N4S2 glycopeptide. (**B**) MS^2 spectrum of the N_{107}-H5N4S2 glycopeptide. (**C**) MS^2 spectrum of the N_{107}-H6N5S3 glycopeptide. (**D**) MS^2 spectrum of the N_{271}-H5N4S2 glycopeptide. Peaks marked with a * indicate a water loss, as seen in panel A. The full glycopeptide structure is shown inside each fragmentation spectrum. For the diantennary glycoform on site N_{107}, we did not observe all fragments. However the observed fragments together with the observed oxonium ions validate the suggested composition.



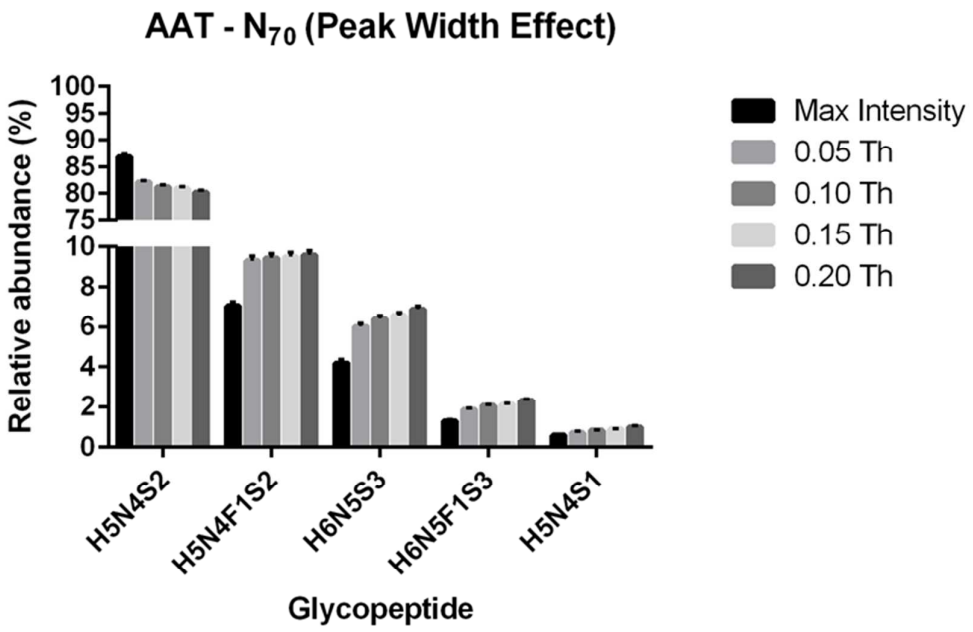
Supporting Information, Figure S-4. Extracted ion chromatograms of main IgG glycoforms. The EICs of the major glycoforms are shown for each subclass of immunoglobulin G (IgG). The displayed chromatograms are extracted with a width of 0.05 Thomson (Th), using the three main isotopes of the doubly and triply charged ions. These results show that the dominant glycoform in all peptide subclasses is the monogalactosylated diantennary fucosylated glycoform, while the agalactosylated and fully galactosylated species are the second and third most abundant glycoforms. These results are in agreement with literature.[1]



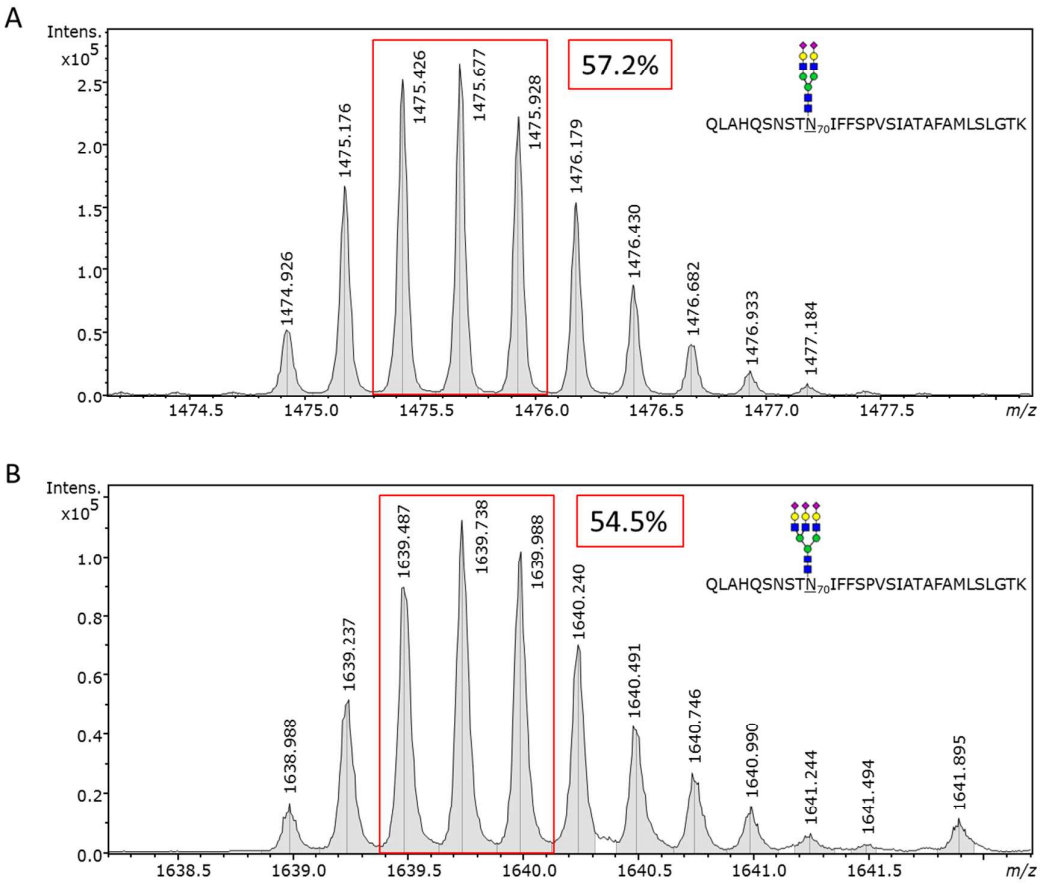
Supporting Information, Figure S-5. Alignment of IgG runs. A total of two IgG runs using differing alignment functions (polynomial in panels (**A**) and (**B**), power law in panels (**C**) and (**D**)) were aligned using LaCyTools. The pictures show the features that were used for alignment before and after alignment (blue and red markers). Furthermore, the figure shows the function that was used for the alignment (blue) as well as the target line (red, striated).



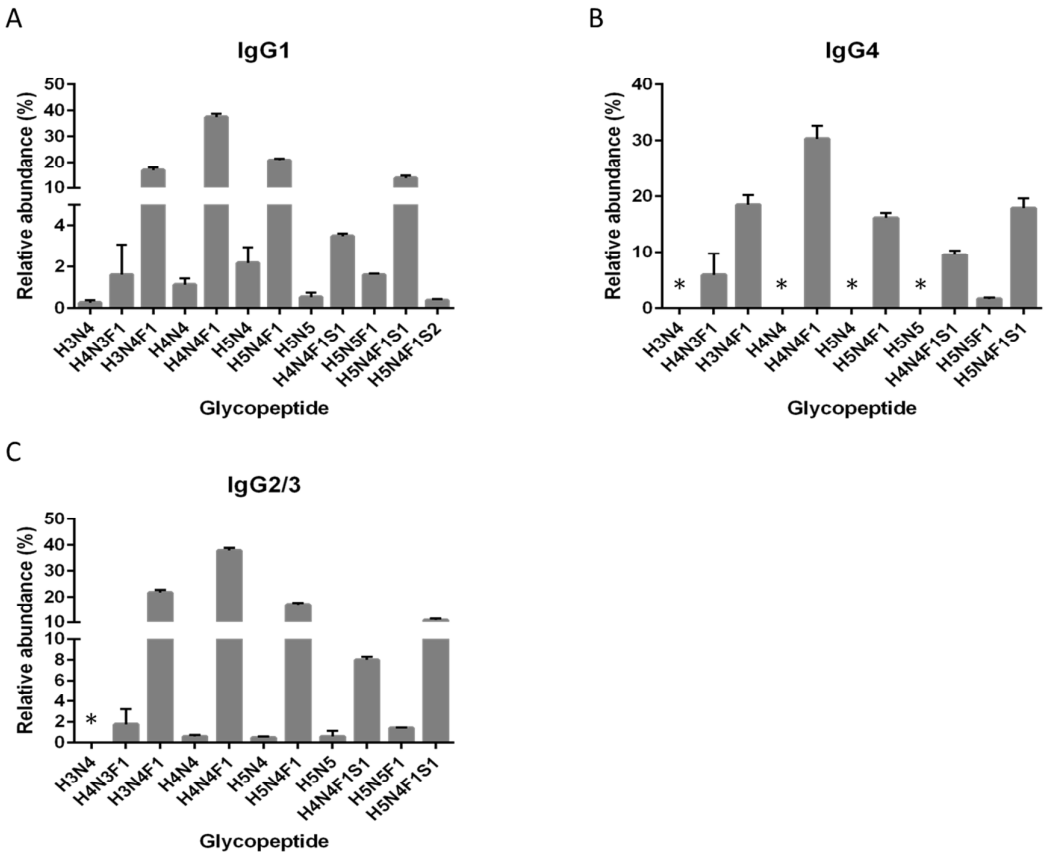
Supporting Information, Figure S-6. Sum spectra of IgG. (**A**) Sum spectrum of the IgG1 EEQYNSTYR peptide cluster. (**B**) Sum spectrum of the IgG2/3 EEQFNSTFR peptide cluster. (**C**) Sum spectrum of the IgG4 EEQFNSTYR peptide cluster. Displayed glycopeptide structures are based on their accurate mass and literature information.[1]



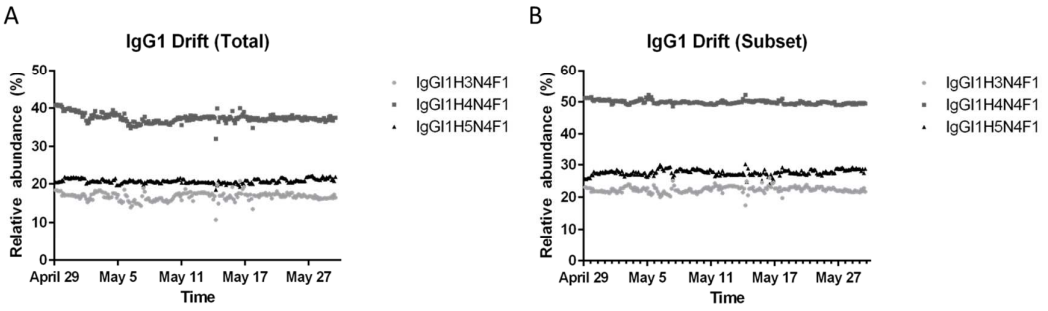
Supporting Information, Figure S-7. AAT quantitation using varying integration regions. Comparison between several quantitations of the AAT glycoforms on N_{70}. The relative area of the main glycopeptide (H5N4S2) increases with a narrower integration region, proving that the bias that is observed between MS peak intensity quantitation and MS peak area integration can in part be attributed to the peak width.



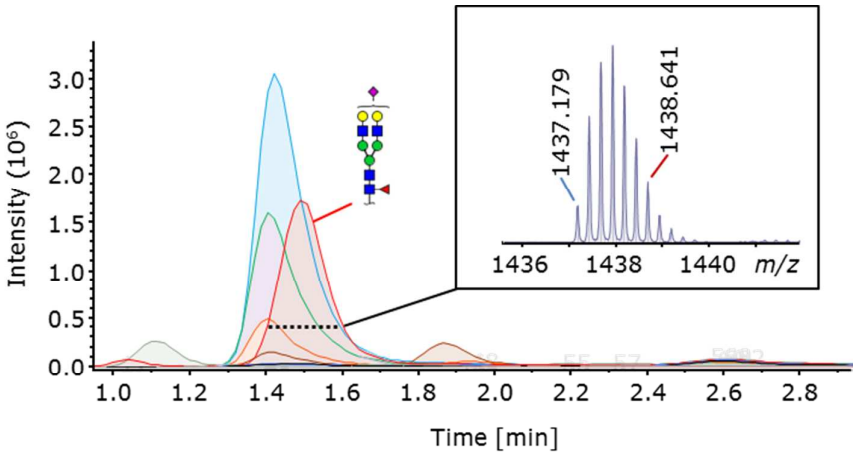
Supporting Information, Figure S-8. Isotopic pattern bias. (**A**) Isotopic pattern of the N_{70}-H5N4S2 glycopeptide. (**B**) Isotopic pattern of N_{70}-H6N5S3 glycopeptide. The percentage of the total theoretical isotopic pattern is displayed next to the three main isotopes, clearly showing a 2.7% bias towards the smaller glycopeptide. This bias can be reduced by extracting a certain, preferably high percentage of the isotopic pattern, as LaCyTools does.



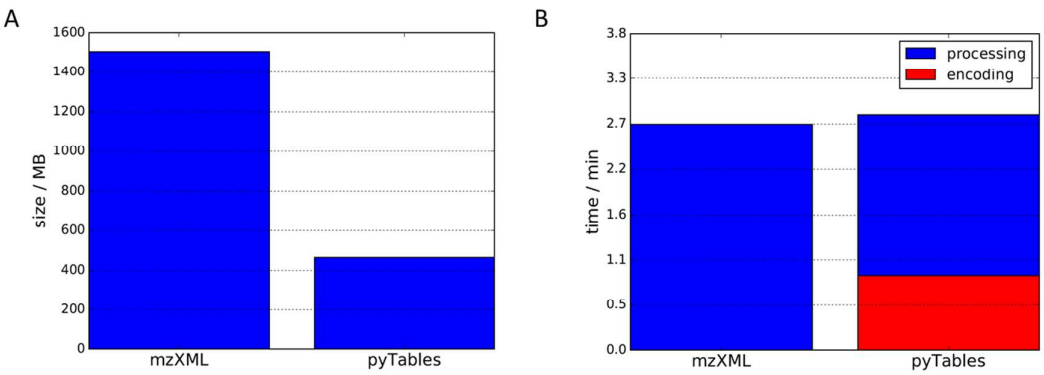
Supporting Information, Figure S-9. Integration of IgG glycoforms. A total of 16 glycoforms that were previously identified, were integrated using LaCyTools[5]. However, the afucosylated glycoforms were automatically discarded for IgG4, as they overlap with the tailing IgG1 glycoforms. Additionally, the glycopeptides that were not detected are indicated by a star. The results show a robust integration with very low variation, as demonstrated by a standard deviation of 1.2% for H4N4F1 on IgG1. Furthermore, the results are fully in agreement with previously published literature.[1]



Supporting Information, Figure S-10. Observed relative abundances of IgG glycoforms over a longer time period. (**A**) The relative abundance of the three major glycoforms of IgG1 as calculated using the total set of 16 glycoforms. (**B**) The relative abundances using only the three major glycoforms of IgG1. Both panels are plotted against the measurement time, demonstrating a decrease in relative abundance of the IgG1-H4N4F1 glycopeptide over time.



Supporting Information, Figure S-11. AAT analyte curation. The displayed chromatograms of the N_{107} peptide cluster show the glycoforms H5N4S2 as the blue trace, H6N5S3 as the purple trace and H5N4F1S1 as the red trace. The curation with LaCyTools showed that N_{107}-H5N4F1S1 glycopeptide had a deviation from the expected isotopic pattern of more than 10%. A detailed view of the mass spectrum acquired by summing the region marked by the striated line, demonstrated that there was an interference at *m/z* 1437.179 (blue line in insert) that overlapped with the expected glycopeptide at *m/z* 1438.641 (red line in insert).



Supporting Information, Figure S-12. Size and performance comparison between mzXML and pyTables options of LaCyTools. AAT raw data has been converted to the mzXML format using ProteoWizard.[2] (**A**) Size comparison in megabytes (MB) between four AAT measurements stored in the mzXML format and four AAT measurements stored in the pyTables format, converted using LaCyTools. (**B**) Performance comparison between running LaCyTools from the mzXML file format versus using the pyTables format. Four AAT measurements were processed in triplicate, with the colored bars indicating the minimum time, in minutes, that was required per step.

References

1. Selman, M. H. J.; Derks, R. J. E.; Bondt, A.; Palmblad, M.; Schoenmaker, B.; Koeleman, C. A. M.; van de Geijn, F. E.; Dolhain, R. J. E. M.; Deelder, A. M.; Wuhrer, M. Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface. *J. Proteomics* **2012,** 75, 1318-1329.
2. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008,** 24, 2534-2536.