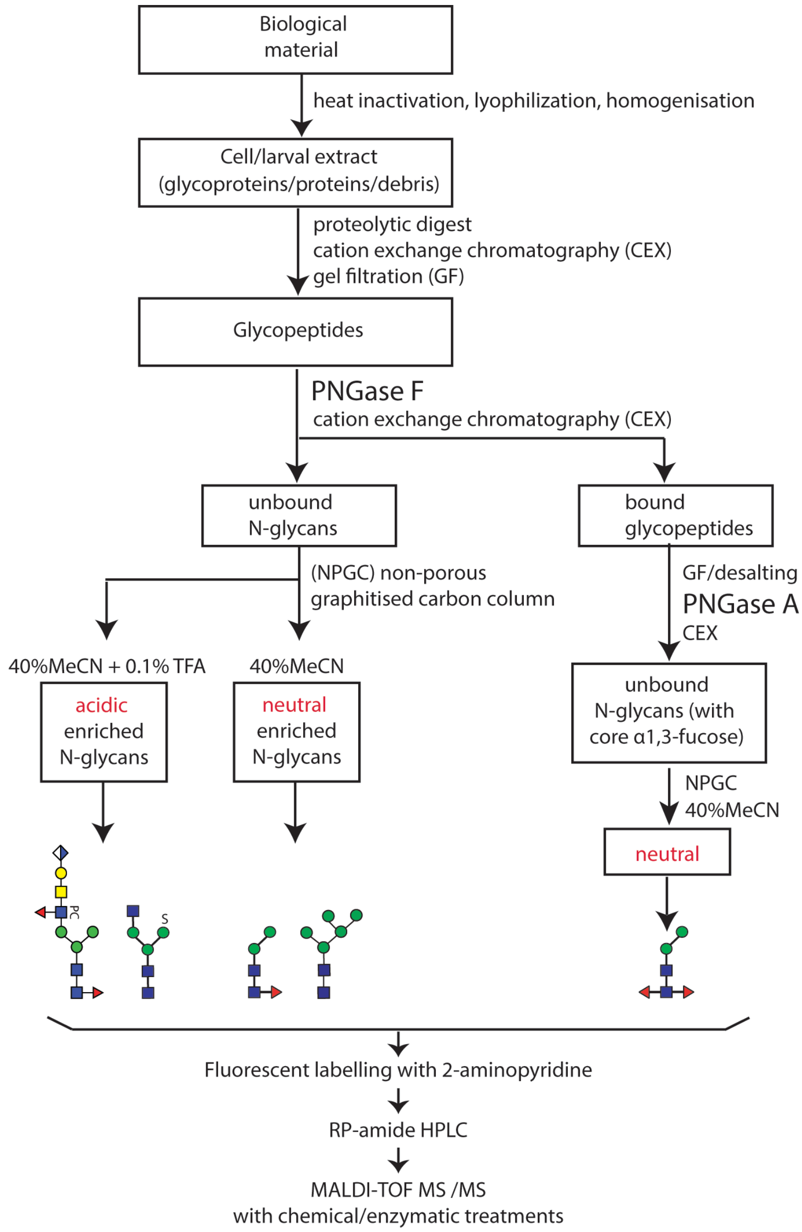
The underestimated N-glycomes of lepidopteran species

Rhiannon Stanton, Alba Hykollari, Barbara Eckmair, Daniel Malzl, Martin Dragosits, Dieter Palmberger, Ping Wang, Iain B. H. Wilson and Katharina Paschinger

Supplementary Data

Scheme: Glycomic workflow employed in this study. Summary of the Experimental Procedures indicating serial digestion with PNGase F then PNGase A followed by solid-phase extraction and labelling steps. Example glycans in the three different pools are shown; MALDI-TOF MS screening of PNGase A released pools resulted in detection of anionic glycans only in the High Five sample.



Further information regarding the glycomic analyses

Definition of the level of the glycan structural analysis

The goal was the in-depth analysis of the N-glycomes of two lepidopteran species and one lepidopteran cell line. Thus individual glycan-containing HPLC fractions were subject to MALDI-TOF MS and MS/MS, a range of chemical and enzymatic treatments and (if appropriate) re-chromatography.

Search parameters and acceptance criteria

a. **Peak lists:** As stated in the methods section: typically 1000 shots were summed for MS and 3000 for MS/MS. Spectra were processed with the manufacturer's software (Bruker Flexanalysis 3.3.80) using the SNAP algorithm with a signal/noise threshold of 6 for MS (unsmoothed) and 3 for MS/MS (four-times smoothed).
b. **Search engine, database and fixed modifications:** All glycan data were manually interpreted and no search engine or database was employed; the fixed modification is the pyridylamine label at the reducing end.
c. **Exclusion of known contaminants and threshold:** All glycan data were manually interpreted; only peaks with an MS/MS consistent with a pyridylaminated core chitobiose were included – the 'threshold' for inclusion was an interpretable MS/MS spectrum.
d. **Enzyme specificity:** A description of the release methods (PNGase F followed by PNGase A) is given in the methods section. Enzymes used during the analysis (glycosyl hydrolases) are defined in the methods by species name and supplier. Citations for in-house purified recombinant enzymes are also given. As previous experience with normalizing glycosidase amounts based on units of activity towards *p*-nitrophenyl sugars reduced digestion efficiency, aliquots of glycans (equivalent to 5 – 50 mV in terms of fluorescence) were incubated with 0.2 µl of the various enzyme preparations (whether commercial, desalted commercial or in-house produced) overnight (except for three hours in the case of FDL digests). These conditions result in no obvious unspecific removal of residues as defined by shifts in mass, MS/MS or retention times, although steric hindrance in some glycans leads to a requirement for longer incubation times (48 hours). Hydroflouric acid treatment (3µl of 48% HF added to the dried glycan) was 24 or 48 hours on ice in the cold room prior to drying under vacuum; expected release of α1,3-fucose and phosphodiesters, but not of other sugars or of sulphate, was observed under these conditions.
e. **Isobaric/isomeric assignments:** For isomeric species, elution and/or digestion data were used for the assignment (as described in the text).

Glycan or glycoconjugate identification

a. **Precursor charge and mass/charge (*m/z*):** All glycans detected were singly-charged and maximally two decimal places used for the *m/z* consistent with the accuracy of MALDI-TOF MS (see *Supplementary Table*); in the figures and due to space limitations, only one decimal place is indicated. For readability reasons, each individual *m/z* measurement in each sample is not listed in the *Supplementary Table* (but refer to the individual Figures for example MS spectra), but previous data indicate an average +0.03 Da (+ 22 ppm) deviation between the measured and the calculated *m/z* values on the instrument used.

b. **All assignments:** A listing of all compositions is given (*Supplementary Table*) for glycans released with either PNGase A or PNGase F. No glycans with core difucosylation (as judged by MS/MS) were detected in the PNGase F digests, consistent with the known specificity of this enzyme. For the glycans present in each pool, see the RP-amide-HPLC chromatograms (*Figures 1-3* and *Supplementary Figures 1-2*) annotated with structures shown according to the Standard Nomenclature for Glycans. Downwardly- and upwardly-drawn core fucose and mannose residues are respectively α1,3- and α1,6-linked.

c. **Modifications observed:** Listed are the *m/z* values for glycans carrying a reducing terminal pyridylamine group (*Supplementary Table*). For the positive mode, the *m/z* values are for protonated forms; in the case of glycans detected in negative mode and carrying two sulphate residues, the *m/z* for sodiated adducts are shown. Depending on the glycan amount or presence of buffers in exoglycosidase preparations, the relative amounts of the H^+, Na^+, K^+ and trace Cu^+ adducts varied.

d. **Number of assigned masses:** No glycan assignments were based on measured mass only; all assignments are based on at least MS/MS (examples are shown in the *Figures 4-9 and Supplementary Figures 2-4*), in most cases corroborated by digest and elution data.

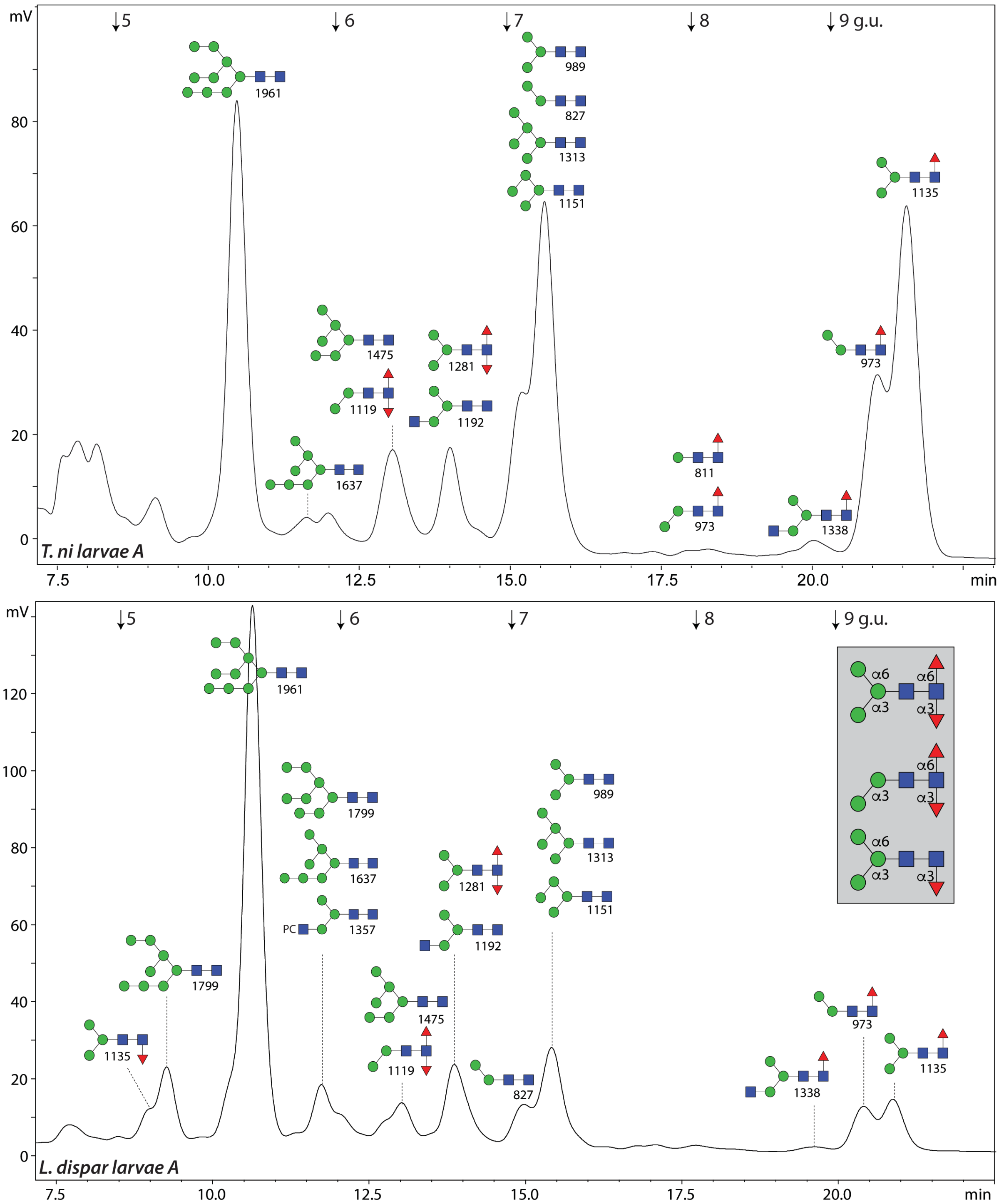
e. **Spectra:** Representative annotated spectra (MS and MS/MS) defining structural elements are given in *Figures 4-9* and *Supplementary Figures 2-5*. In total, MS and/or MS/MS data for some 60 of the approximately 100 glycans are shown. The overall data is based on some 1500 MS and MS/MS spectra. Ten 'complete' MALDI-TOF MS spectra of pyridylaminated glycan pools, as well as a spectrum of an HPLC fraction containing $Man_8GlcNAc_2$, are also submitted in mzXML format, whereby both positive and negative mode spectra are included for the anionic pools.

f. **Structural assignments:** As noted in the results section, the typical oligomannosidic structures are assigned based on elution time and fragmentation pattern; it is otherwise assumed that the glycans contain a trimannosyl core consistent with typical eukaryotic N-glycan biosynthesis. MALDI-TOF MS of oligomannosidic glycans shows no evidence for significant in-source fragmentation under the employed analysis conditions (see *Supplementary Figure 3* and $Man_8GlcNAc_2$ mzXML file).

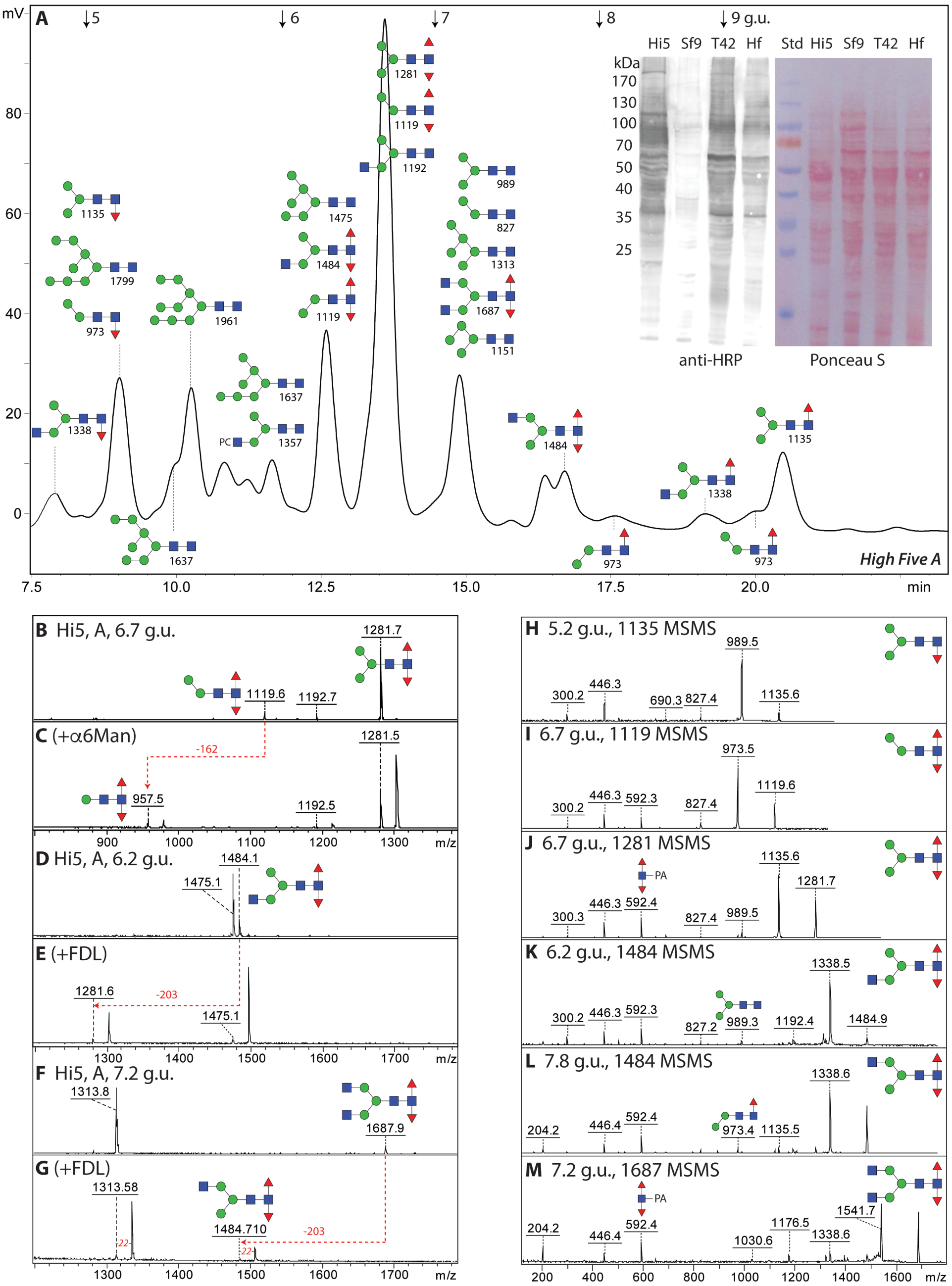
   The assignments of antennal and core fucose residues are based on RP-HPLC retention time, fragmentation pattern and/or susceptibility to digestions. Other antennal modifications (e.g., fucose, galactose, glucuronic acid and *N*-acetylgalactosamine; including type of glycosidic linkage) are defined based on digestions and fragmentation patterns with rechromatography after digestion in some cases. There is no evidence of in-source fragmentation of either neutral terminal monosaccharides (including Lewis-type fucosylation), phosphorylcholine or glucuronic acid (as evidenced by this and previous publications).

   The definition of sulphate is based on detection in negative ion mode, in-source loss in positive ion mode, resistance to hydrofluoric acid (as compared to isobaric phosphate), non-digestion of the underlying residue and (as appropriate) co-elution with structures from mosquito shown to be partially sensitive to solvolysis with methanolic HCl.

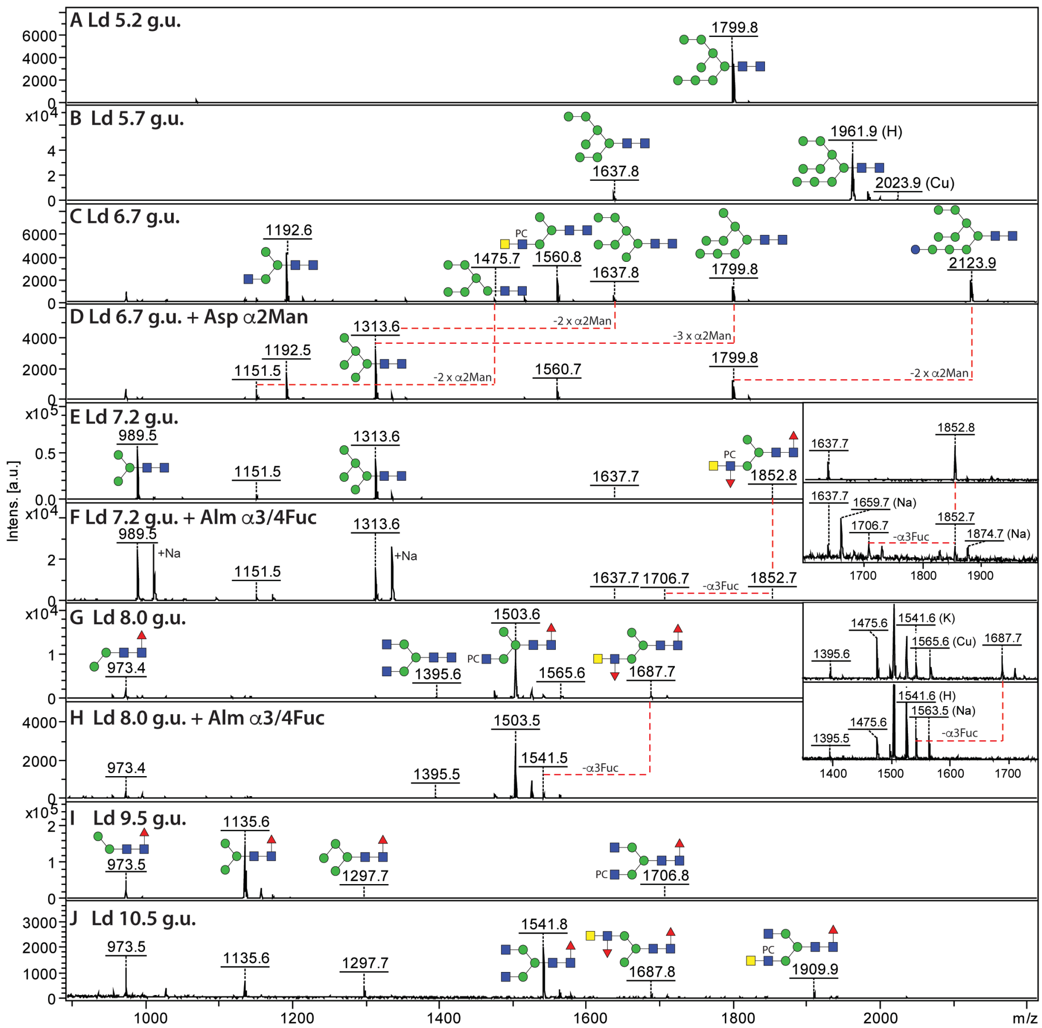
Supplementary Figure 1: PNGase A-digests of *T. ni* and *L. dispar* larvae. RP-amide chromatograms of the pyridylaminated N-glycans of lepidopteran larvae resulting from PNGase A digestion of glycopeptides remaining after a PNGase F digest are annotated with proposed structures and the glucose units. The dominant glycans in these preparations represent residual oligosaccharides which were not released by PNGase F despite the lack of core α1,3-fucose; thereby, core α1,3-fucosylated and α1,3/6-difucosylated glycans are present in low amounts (see inset in lower panel for relevant linkage annotations).



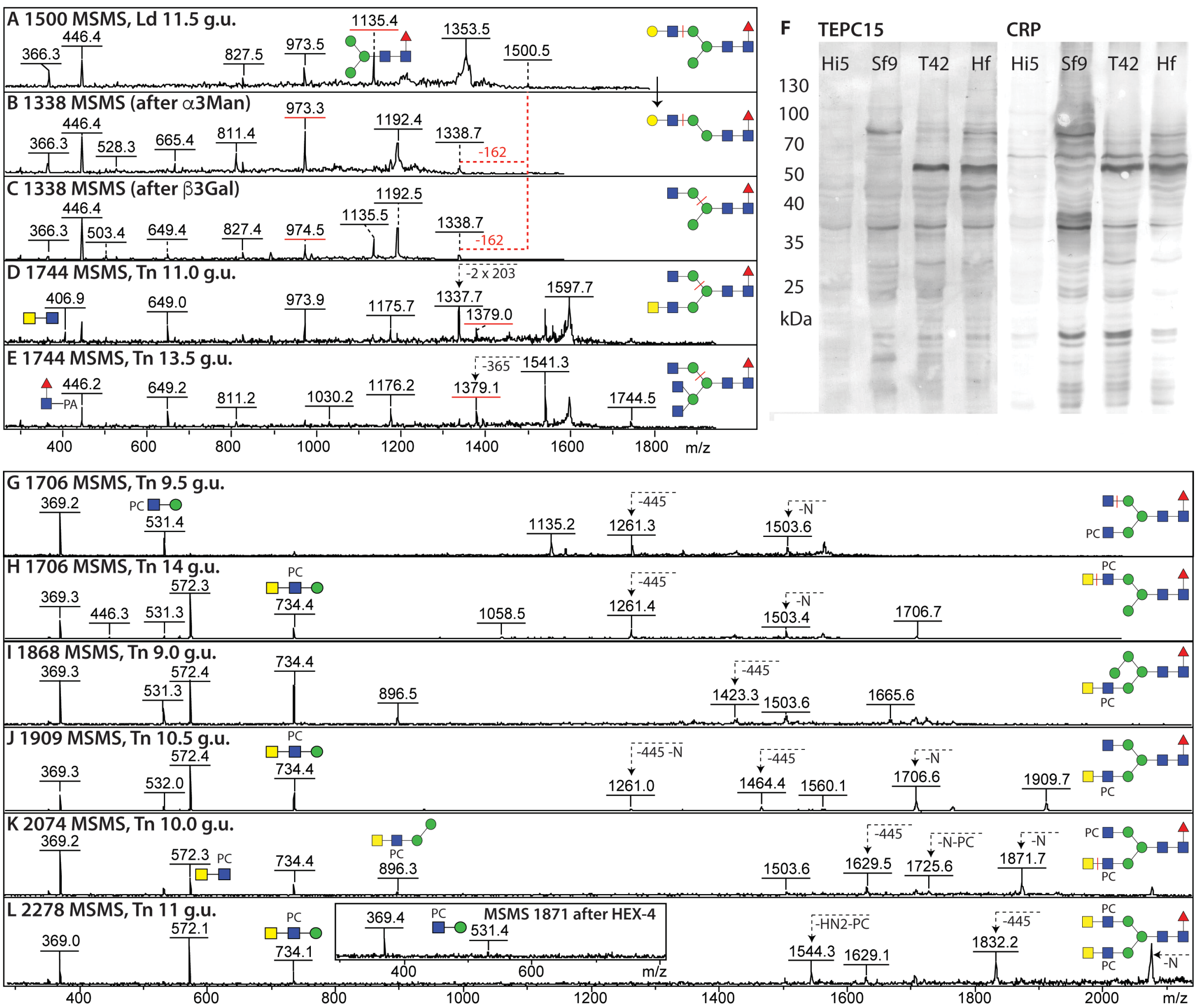
Supplementary Figure 2: Neutral PNGase A-released glycans from High Five cells. (**A**) RP-amide chromatogram of the pyridylaminated N-glycans of *T. ni* High Five cells resulting from PNGase A digestion of glycopeptides (which remained after PNGase F digestion) are annotated with proposed structures and the glucose units (g.u.). (**B-G**) Positive mode MALDI-TOF MS of three fractions before and after diagnostic α1,6-mannosidase or FDL β-hexosaminidase treatment. (**H-M**) Positive mode MALDI-TOF MS/MS of core α1,3-fucosylated and α1,3/6-difucosylated glycans (only protonated forms were fragmented despite the high abundance of sodiated adducts in the FDL digest); characteristic for mono- and difucosylation of pyridylaminated glycans are the *m/z* 446 and 592 Y1 fragment ions. The inset in A shows the result of anti-HRP Western blotting and Ponceau S staining of three *T. ni* cell lines as well as of Sf9 cells; the low anti-HRP staining of Sf9 extracts is in accordance with previous data indicating minimal core α1,3-fucosylation in this cell line.



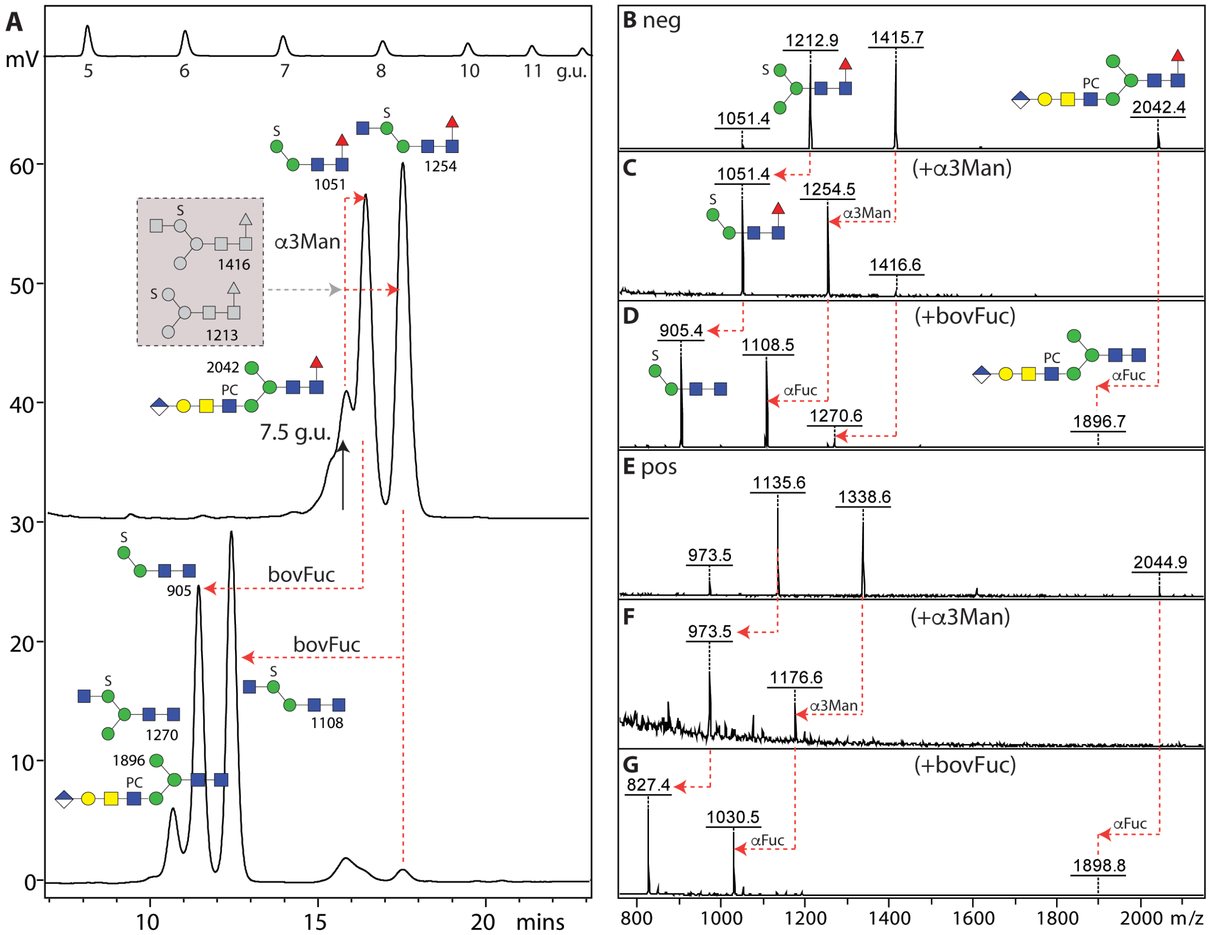
Supplementary Figure 3: Selected mass spectra of *L. dispar* neutral PNGase F-released HPLC-fractionated N-glycans with or without exoglycosidase treatments. The depicted MALDI-TOF MS spectra represent the 'full' data for parts of Figures 4 and 5 from the main text; the major structures are annotated with *m/z* values of the $[M+H]^+$ ions, in some cases also other adducts with sodium (+22), potassium (+38) or copper (+62/+64), as based on retention time and fragmentation data. Panels A and B are spectra of major oligomannosidic fractions and indicate that in source dissociation is absent in the case of $Man_{7,8,9}GlcNAc_2$ (see also mzXML file for $Man_8GlcNAc_2$). Panels C and D are fuller-range depictions of the MS spectra in Figure 4 G and H in the main text (refer to Figure 4 S-X for MS/MS spectra of selection ions before and after *Aspergillus* α1,2-mannosidase digestion). Panels E, F, G and H are fuller-range depictions of the MS spectra in Figure 5 A, B, D and F in the main text (zoomed-in sections of the spectra are also shown to highlight the effects of almond α1,3/4-fucosidase on the glycans containing antennal Lewis-like fucose modifications – see also changes in MS/MS patterns in Figure 5 of the main text); no evidence of unspecific removal of core α1,6-fucose of the co-eluting *m/z* 1503 structure (*m/z* 1541/1565 as $[M+K]^+$/$[M+Cu]^+$) was observed, but some shift towards sodiated adducts occurred. Panels I and J depict MS spectra of the two neighbouring 9.5 and 10.5 g.u. fractions (hence some overlap in the contained structures); panel J is a fuller-range spectrum of that shown in Figure 5 C.



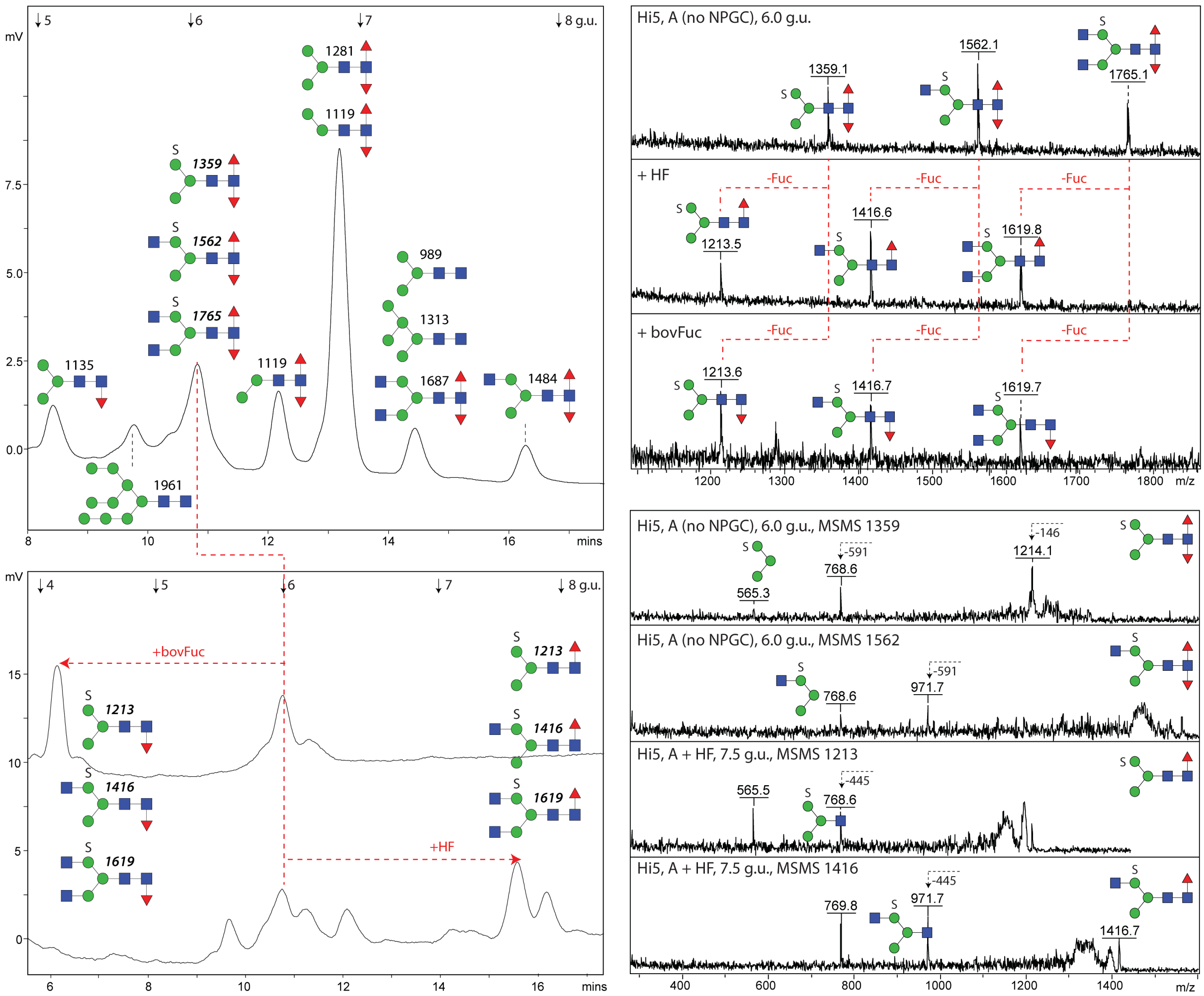
Supplementary Figure 4: Characterisation of neutral and zwitterionic antennal modifications of *L. dispar* and *T. ni* glycans. (**A-C**) MS/MS spectra of $Gal_1Man_3GlcNAc_3Fuc_1$ (*m/z* 1500) before and after specific *Xanthomonas* α1,2/3-mannosidase or *Xanthomonas* β1,3-galactosidase treatment; this isomer eluting at 11.5 g.u is distinct from the $Man_4GlcNAc_3Fuc_1$ glycan eluting at 8.4 g.u. (see Figure 4 D and Q). (**D and E**) MS/MS of two isomers of $Hex_3HexNAc_5Fuc_1$ (*m/z* 1744) of different elution times isolated from *T. ni* larvae, whereby the B ion of *m/z* 407 is diagnostic for a HexdiNAc motif; the elution time of the triantennary glycan corresponds to that of one from mosquito run on the same column (13.5 g.u.), but contrasts with an isomer with a putative upper arm β1,6-GlcNAc as found in nematodes (elution at 8.2 g.u.). (**F**) Western blotting of three *T. ni* cell lines as well as Sf9 cells with the murine TEPC15 antibody and human C-reactive protein, which both recognise phosphorylcholine; protein loading was similar as judged by Ponceau S staining (see inset in Supplementary Figure 2A). (**G and H**) MS/MS of two further isomers of $Hex_3HexNAc_4Fuc_1PC_1$ (*m/z* 1706) with a biantennary or 'upper arm' pseudohybrid structure; (**I-K**) MS/MS of $Hex_{3-4}HexNAc_{4-5}Fuc_1PC_{1-2}$ glycans; (**L**) MS/MS of $Hex_3HexNAc_6Fuc_1PC_2$ (*m/z* 2278) before and after (see inset) *C. elegans* HEX-4 β-*N*-acetylgalactosaminidase treatment, thereby verifying the presence of two unsubstituted terminal GalNAc residues. Key B-fragments are annotated with structures and selected losses yielding Y-fragments are also indicated.



Supplementary Figure 5: RP-amide HPLC and MALDI-TOF-MS analysis of *L. dispar* sulphated glycans before and after glycosidase digestion. (**A**) Digestion of the 7.5 g.u. anionic fraction (solid arrow) with *Xanthomonas* α1,2/3-mannosidase resulted in shifts to higher retention time (only ~90% due to steric effect in case of *m/z* 1416); subsequent bovine α-fucosidase then resulted in a lower retention time reflecting shifts in *m/z* as judged by negative (**B-D**) and positive (**E-G**) MALDI-TOF MS. The original sulphated glycans are indicated on the chromatogram in grey and the digestion products as well as the mannosidase-resistant *m/z* 2042 glycan (see also **Figure 8 M and Figure 9 F and G** in the main text) in colour.



Supplementary Figure 6: Anionic High Five PNGase A-released glycans. An aliquot of PNGase A-released N-glycans was pyridylaminated without prior solid phase extraction. This revealed at peak eluting at 6 g.u. containing three glycans with core difucosylation and hexose-linked sulphate. This fraction was subject to treatment with either bovine α-fucosidase (removal of α1,6-fucose) or hydrofluoric acid (removal of α1,3-fucose) leading to shifts in retention time (note that different chromatogram windows are shown) as well as *m/z* (see panels on upper right); original and product glycans are annotated together with the relevant positive or negative mode *m/z* value (bold in the case of negative ions). The relevant negative mode MS/MS of the original *m/z* 1359 and 1562 glycans as well as the *m/z* 1213 and 1416 HF products demonstrate sulphation of the core trimannosyl region (panels on lower right).



Supplementary Table. Compositions and theoretical *m/z* values for pyridylaminated lepidopteran N-glycans. The ninety different compositions are of the form $H_xN_yF_{1\text{-}2}PC_{0\text{-}2}GlcA_{0\text{-}1}S_{0\text{-}2}$ (i.e., $Hex_xHexNAc_yFuc_{1\text{-}2}PC_{0\text{-}2}GlcA_{0\text{-}1}S_{0\text{-}2}$, shown as, e.g., H4N5FPCGlcAS, whereby PC and S are the abbreviations for phosphorylcholine and sulphate), but do not take account of isomers; single fucose, PC, glucuronic acid or sulphate modifications are given without a number. The presence of a particular glycan mass in a sample is indicated by a tick. Theoretical *m/z* were calculated by Glycoworkbench; generally the observed values (refer to individual spectra) are within 0.2 mass units.

Composition	$[M+H]^+$	$[M-H]^-$	$[M-2H+Na]^-$	*T. ni*	Hi 5	*L. dispar*
H1N2F	811.34			✓	✓	✓
H2N2	827.34			✓	✓	✓
H2N2S		905.28		✓	✓	✓
H2N2F	973.39			✓	✓	✓
H3N2	989.39			✓	✓	✓
H2N2FS		1051.34		✓	✓	✓
H3N2S		1067.33		✓	✓	✓
H2N3S		1108.36		✓	✓	
H2N2F2	1119.46			✓	✓	✓
H3N2F	1135.45			✓	✓	✓
H4N2	1151.45			✓	✓	✓
H3N2S2			1169.27	✓	✓	✓
H2N3F	1176.48			✓	✓	✓
H3N3	1192.47			✓	✓	✓
H3N2FS		1213.39		✓	✓	✓
H2N3FS		1254.42		✓	✓	✓
H3N3S		1270.41		✓	✓	✓
H3N2F2	1281.51			✓	✓	✓
H4N2F	1297.50			✓	✓	✓
H5N2	1313.50			✓	✓	✓
H3N2FS2			1315.33	✓	✓	✓
H3N3F	1338.53			✓	✓	✓
H4N3	1354.52			✓	✓	✓
H3N3PC	1357.53			✓	✓	✓
H3N2F2S		1359.45			✓	
H3N3S2			1372.35	✓	✓	✓
H3N4	1395.55			✓	✓	✓
H3N3FS		1416.47		✓	✓	✓
H4N3S		1432.47		✓	✓	
H3N3PCS		1435.47		✓	✓	
H3N4S		1473.49		✓	✓	✓
H6N2	1475.55			✓	✓	✓
H3N3F2	1484.59				✓	
H5N2F	1459.56			✓	✓	✓
H4N3F	1500.58			✓	✓	✓
H3N3FPC	1503.59			✓	✓	✓
H5N3	1516.58			✓	✓	✓
H3N3FS2			1518.41	✓	✓	✓
H4N3GlcA	1530.56	1528.54		✓		✓
H3N4F	1541.61			✓	✓	✓
H3N4PC	1560.61			✓	✓	✓
H3N3F2S		1562.53			✓	
H3N4S2			1575.43	✓	✓	✓

Composition	$[M+H]^+$	$[M-H]^-$	$[M-2H+Na]^-$	*T. ni*	Hi 5	*L. dispar*
H4N3FS		1578.52		✓	✓	✓
H3N3FPCS		1581.53		✓	✓	
H4N3GlcAS		1608.50		✓		✓
H3N4FS		1619.55		✓	✓	
H6N2F	1621.61			✓	✓	✓
H7N2	1637.60			✓	✓	✓
H3N4PCS		1638.46		✓	✓	
H5N3F	1662.64			✓	✓	
H4N3FPC	1665.64			✓		
H4N3FGlcA	1676.61	1674.60		✓	✓	✓
H3N4F2	1687.67				✓	✓
H3N4FPC	1706.66			✓	✓	✓
H3N4FS2			1721.49	✓	✓	
H4N4PC	1722.66			✓		
H4N4GlcA	1733.64	1731.62		✓		
H3N5F	1744.69			✓	✓	
H4N3FGlcAS		1754.56		✓		✓
H3N5PC	1763.69			✓		
H3N4F2S		1765.61		✓	✓	
H4N4FS		1781.60		✓	✓	
H7N2F	1783.66			✓	✓	
H3N4FPCS		1784.61		✓	✓	✓
H8N2	1799.66			✓	✓	✓
H3N4PC2S		1803.60			✓	
H5N3FPC	1827.69			✓		
H3N5PCS		1841.63		✓	✓	
H3N4F2PC	1852.72			✓	✓	✓
H4N4FPC	1868.72			✓		
H3N4FPC2	1871.72				✓	✓
H4N4FGlcA	1879.69	1877.68		✓		✓
H4N4PCGlcA	1898.69	1896.68		✓	✓	✓
H3N5FPC	1909.74			✓	✓	✓
H3N4F2PCS		1930.66			✓	
H3N4FPC2S		1949.66			✓	
H9N2	1961.71			✓	✓	✓
H3N5FPCS		1987.69		✓	✓	✓
H4N4FPCGlcA	2044.75	2042.73		✓	✓	✓
H3N5FPC2	2074.80			✓		
H4N5FGlcA	2082.77	2080.76				✓
H4N5PCGlcA	2101.77	2099.76			✓	✓
H10N2	2123.76			✓	✓	✓
H4N4F2PCGlcA	2190.81	2188.79				✓
H4N5FPCGlcA	2247.83	2245.81		✓	✓	✓
H3N6FPC2	2277.88			✓	✓	✓
H11N2	2285.81			✓		
H4N5FPCGlcAS		2325.77				✓
H12N2	2447.69			✓		
H4N6FPCGlcA	2450.91	2448.89				✓
H4N6FPC2GlcA	2615.96	2613.95				✓