Glycoproteomic measurement of site-specific polysialylation

Ruby Pelingon[1¥], Cassandra L. Pegg[2¥], Lucia F. Zacchi[1¥], Toan K. Phung[2], Christopher B. Howard[1], Ping Xu[3], Matthew P. Hardy[3], Catherine M. Owczarek[3], Benjamin L. Schulz[1,2*]

[1]ARC Training Centre for Biopharmaceutical Innovation, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia. Ph: +61 7 334 63880.

[2] School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, 4072, Australia;. Ph: +61 7 336 54875.

[3] Research and Development, CSL Limited; Bio21 Molecular Science and Biotechnology Institute, Parkville Victoria, 3052, Australia. Ph: +61 3 9389 2770

Running title: Site-specific polysialylation glycoproteomics

[*]To whom correspondence should be addressed: Benjamin L. Schulz, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, 4072, Australia. Tel.: +61-7-336-54875; E-mail: b.schulz@uq.edu.au

[¥]These authors contributed equally to this work

Keywords: polysialylation, sialic acid, *N*-linked glycosylation, mass spectrometry, glycoproteomics

Abstract

Polysialylation is the enzymatic addition of a highly negatively charged sialic acid polymer to the non-reducing termini of glycans. Polysialylation plays an important role in development, and is involved in neurological diseases, neural tissue regeneration, and cancer. Polysialic acid (PSA) is also a biodegradable and non-immunogenic conjugate to therapeutic drugs to improve their pharmacokinetics. PSA chains vary in length, composition, and linkages, while the specific sites of polysialylation are important determinants of protein function. However, PSA is difficult to analyse by mass spectrometry (MS) due to its high negative charge and size. Most analytical approaches for analysis of PSA measure its degree of polymerization and monosaccharide composition, but do not address the key questions of site specificity and occupancy. Here, we developed a high-throughput LC-ESI-MS/MS glycoproteomics method to measure site-specific polysialylation of glycoproteins. This method measures site-specific PSA modification by using endosialidase and mild acid hydrolysis to eliminate PSA and sialic acids while leaving the glycan backbone intact, together with protease digestion followed by LC-ESI-MS/MS glycopeptide detection. PSA-modified glycopeptides are not detectable by LC-ESI-MS/MS, but become detectable after desialylation, allowing measurement of site-specific PSA occupancy. This method is an efficient analytical workflow for the study of glycoprotein polysialylation in biological and therapeutic settings.

Introduction

Protein glycosylation, the enzymatic addition of carbohydrates to a protein, is a common post-translational modification that plays a variety of physiological roles.[1] Protein glycosylation is critical in protein folding and transport through the secretory pathway, and is also important for protein solubility and stability, cell-cell interactions, and ligand-receptor binding.[1-4] *N*-linked glycosylation occurs at selected asparagine (N) residues within the consensus "sequon" N-X-S/T (where X ≠ proline).[5] *N*-linked glycans are pre-assembled on the membrane of the endoplasmic reticulum (ER) as a dolichol-linked Glc(3)Man(9)GlcNAc(2) structure (Glc, glucose; Man, mannose), and are transferred *en bloc* by the enzyme oligosaccharyltransferase to the polypeptide in the lumen of the ER.[3] Glycans are modified as proteins traverse the secretory pathway, leading to heterogeneity in the glycan structures in mature glycoproteins.[1, 6, 7] The presence and structures of glycans at particular sites have a major influence on protein conformation, stability, and function.[8]

Mammalian glycans are commonly terminated by sialic acids, 9-carbon carboxylated monosaccharides.[9, 10] Although there are many sialic acid variants, N-acetylneuraminic acid (NeuAc) is the most common in human glycans. Glycans are typically monosialylated, but sialic acids can also be found as short oligosialyl chains (2-7 monosaccharides) and as polysialic acid (PSA) (≥ 8 monosaccharides).[9, 11] PSA is structurally heterogenous due to variability in the sialic acid type, modifications, glycosidic linkages, and degree of polymerization.[10] Only few polysialylated proteins have been identified, including neural cell adhesion molecule (NCAM),[12] synCAM-1,[13] neuropilin-2,[14] voltage-sensitive sodium channel,[15, 16] CCR7

chemokine,[17] E-selecting ligand-1,[18] CD36 in milk,[19] and the polysialyltransferases themselves.[20, 21] NCAM is the best-studied polysialylated protein, and carries α2-8-linked NeuAc PSA in two of its *N*-glycans.[11, 22-24] Bulky and negatively charged PSA-modified NCAM inhibits cell-cell interactions, thereby playing important roles in embryonic development,[11] neurological diseases,[11, 25, 26] tumour metastasis,[27] and neural tissue regeneration.[28] PSA also has important biotechnological applications.[11, 29] PSA is used as a protein conjugate (chemically/enzymatically attached to proteins) or as a delivery system (PSA micelles) to improve pharmacokinetics of therapeutic proteins.[29-34] Thus, polysialylation is a post-translational modification with high physiological, pathological, and biotechnological relevance.

Analysis of PSA-modified glycoproteins is challenging due to the large size, high negative charge, and structural heterogeneity of PSA. Immuno-detection can detect protein polysialylation,[10, 35, 36] and detection with and without endosialidase treatment allows inference of the presence of PSA-modified proteins.[37-39] However, anti-PSA immunodetection provides no information about the sites of polysialylation, PSA structure and composition, the degree of polymerization, or the underlying glycan structures. Several approaches overcome some of these limitations.[40, 41] PSA degree of polymerization, sialic acid composition, and glycosidic linkages can be measured using: mild acid hydrolysis, derivatization, and HPLC;[42-45] fluorometric C7/C9 analysis;[40, 45-47] high performance (high pH) anion-exchange chromatography pulsed electrochemical detection (HPAEC-PED);[48, 49] nuclear magnetic resonance (NMR);[50-53] or mass spectrometry (MS) glycomics [54-57], including matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-TOF MS),[56, 58, 59] gas chromatography-MS (GC-MS),[60] and

electrospray ionization MS (ESI-MS).[57, 61] Therefore, a wide range of chemical and analytical tools are available to characterize released PSA chains. Three additional features of polysialylated glycoproteins are critical: site-specific PSA localisation, PSA occupancy, and the underlying structure of the polysialylated glycan. Sites of PSA modification can be identified by site-directed mutagenesis of glycosylation sites with anti-PSA Western blotting,[22, 62] or immunoenrichment of tryptic PSA-glycopeptides, followed by deglycosylation and MS analysis.[13, 51, 63, 64] However, these methods do not measure site-specific PSA occupancy.

Here, we developed a high-throughput glycoproteomic workflow to measure site-specific polysialylation of glycoproteins (Figure 1). This workflow combines endosialidase and mild acid hydrolysis to remove PSA from glycans, followed by LC-ESI-MS/MS for detection and site-specific relative quantification of previously polysialylated asialoglycopeptides (Figure 1). By comparing glycoform abundance in the desialylated sample with that from the untreated control, site specific polysialylation can be inferred, its occupancy measured, and the underlying polysialylated glycan structure can be elucidated.

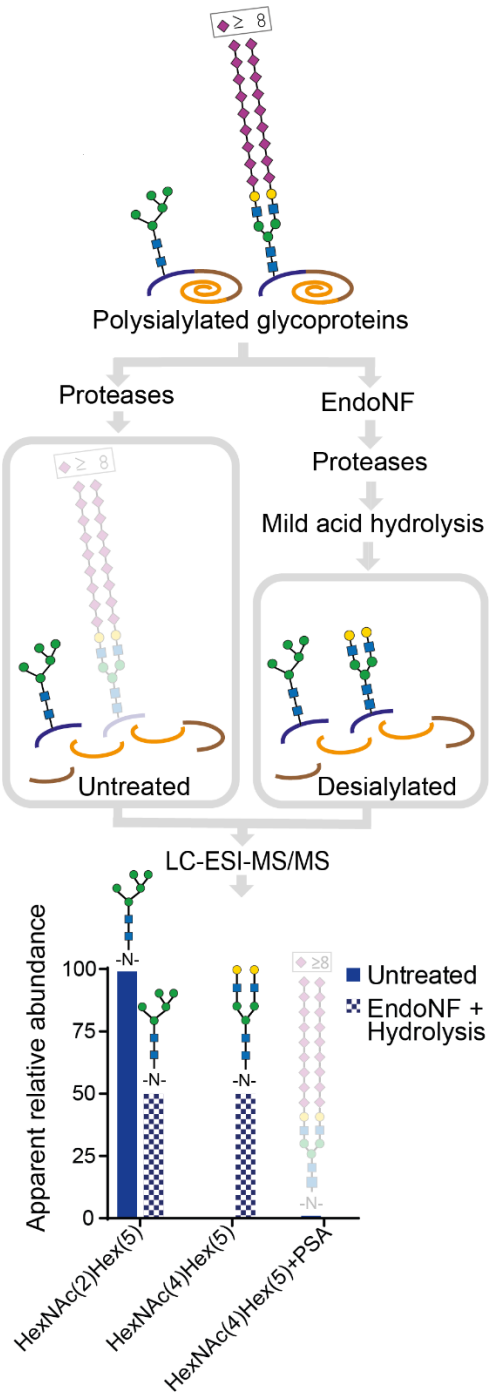


Figure 1. Schematic representation of our glycoproteomic method to detect and measure glycoprotein polysialylation. Polysialylated glycoproteins are protease digested to peptides and analysed by LC-ESI-MS/MS with and without desialylation with Endosialidase NF (EndoNF) and mild acid hydrolysis. Comparison of apparent site-specific glycoform abundance after desialylation and in untreated controls allows inference of site-specific glycan polysialylation and occupancy.

Experimental Section

Cloning, expression, and purification of polysialylated recombinant human NCAM

cDNAs encoding human Neural Cell Adhesion Molecule 1 (NCAM) (GenBank Accession number NP_000054), human β1,4-galactosyltransferase 1 (B4GALT1) (NP_001497), human CMP-*N*-acetylneuraminate-β-galactosamide-α-2,3-sialyltransferase 4 (ST3GAL4) (NP_001241686), and human CMP-*N*-acetylneuraminate-poly-α-2,8-sialyltransferase (ST8SIA4) (NP_005659)were codon-optimized for human expression and synthesized (Geneart, Thermo Fisher Scientific). A cDNA encoding a soluble version of human NCAM, with a stop codon introduced after amino acid 718, was generated using standard PCR-based techniques, with a C-terminal 8His tag and a FLAG (DYKDDDDK) tag after the signal peptide cleavage site (rHuNCAM, Figure 2 and Supplementary Figure 2). cDNAs had a Kozak consensus sequence (GCCACC) placed upstream of the initiating methionine. PCR products were digested with NheI and XhoI, and ligated into pcDNA3.1 (Invitrogen, Thermo Fisher Scientific) to produce pcDNA-HuNCAM(1-19)-FLAG-HuNCAM(20-718)-8His. Plasmid DNA was prepared using QIAGEN Plasmid Giga Kits according to manufacturer's instructions.

Expi293F cells (Invitrogen, Thermo Fisher Scientific) were cultured in Expi293 Expression Medium (Invitrogen, Thermo Fisher Scientific) supplemented with Antibiotic-Antimycotic (GIBCO, Thermo Fisher Scientific) and cells were maintained in an incubator at 37 °C and 150 rpm with an atmosphere of 8 % CO_2. Transient transfections of expression plasmids used the Expi293 Expression system and were performed following manufacturer’s recommendations. To produce polysialylated rHuNCAM, expression plasmids pcDNA-HuNCAM(1-19)-FLAG-

HuNCAM(20-718)-8His, pcDNA-Hu-ST8SIA4, pcDNA-Hu-B4GAL1, and pcDNA-Hu-ST3GAL4 with DNA ratios of 0.85:0.05:0.05:0.05 were co-transfected into Expi293 cells, with 0.5 % (w/v) LucraTone Lupin (Celliance) added 18-20 h post-transfection.[65, 66] After 6 days of incubation cells were harvested as described.[67] Expression of secreted polysialylated rHuNCAM was verified by SDS-PAGE and Coomassie Blue staining, and Western blot analysis using anti-His (A01620, Genscript) and anti-PSA-NCAM (MAB5324, Millipore) antibodies. rHuNCAM was purified from the culture supernatant of transiently transfected Expi293F cells using nickel sepharose resin and size-exclusion chromatography as previously described.[68, 69]

Expression, purification, and activity of recombinant Endosialidase NF

The pCW(-lacZ) vector containing the MalE-thrombin-Endosialidase NF construct was a kind gift from Warren Wakarchuk.[37] Expression and purification of MalE-thrombin-Endosialidase NF (EndoNF) used a published method with minor modifications.[70] EndoNF retained in the amylose-agarose resin was eluted with working buffer supplemented with 10 mM maltose. The purity and quality of EndoNF were determined by SDS-PAGE and Coomassie Blue staining (Supplementary Figure 1).

To enzymatically de-polysialylate rHuNCAM, 5 µg of rHuNCAM was diluted in 50 mM sodium phosphate buffer pH 7.5, treated or not with 0.25 µg EndoNF (20:1 sample:EndoNF ratio), and incubated at 37 °C and 1500 rpm for 2 h in a MS100 thermoshaker incubator (LabGene Scientific). Samples were precipitated with 4 volumes of 1:1 acetone:methanol, incubated at -20 °C for 16 h, and centrifuged at 20,000 rcf for 20 min at 4 °C. Pellets were reconstituted in 1X SDS-PAGE loading buffer, incubated at 65 °C for 15 min, and separated by SDS-PAGE using a 6%

polyacrylamide gel. Separated proteins were either transferred onto nitrocellulose membrane (Bio-Rad) using an XCell II Blot Module (Invitrogen, Thermo Fisher Scientific) or stained with Coomassie Blue. Western blots using anti-PSA antibody (mAb735, Abcam) were developed using SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific), and images were captured using a GE Amersham Imager 600.

Mass Spectrometry sample preparation

Four treatments were tested: EndoNF alone, EndoNF followed by mild acid hydrolysis, EndoNF followed by PNGase F (New England Biolabs), and untreated control. For each treatment, 5 µg of purified rHuNCAM or serum purified human IgG (IgG) (I4506, Sigma-Aldrich) was prepared in triplicate. Samples were digested with 0.25 µg EndoNF (20:1 sample:EndoNF ratio) in 50 mM sodium phosphate buffer pH 7.5 at 37 °C and 1500 rpm for 2 h. Proteins were then denatured, reduced, alkylated, and precipitated as described.[71, 72] The precipitated protein pellets were resuspended in 20 µL of 50 mM ammonium bicarbonate and protease digested. For rHuNCAM all samples were sequentially digested with trypsin (30:1 sample:protease ratio), Asp-N (80:1 sample:protease ratio), and Glu-C (20:1 sample:protease ratio) by incubation in a thermoshaker at 37 °C and 1500 rpm for 16 h. Each protease was deactivated before the addition of the next enzyme by heating at 95 °C for 10 min for trypsin [73, 74], and 5 min for Asp-N and Glu-C. For IgG, all samples were digested with trypsin (30:1 sample:protease ratio) by incubation at 37 °C and 1500 rpm for 16 h. Trypsin activity was deactivated by heating at 95 °C for 10 min. Mild acid hydrolysis was performed by acidifying samples with 2% formic acid and heating at 95 °C for 1 h. For PNGase F treatment, samples were digested with 500 units of PNGase F at 37 °C and 1500 rpm for 16 h in a thermoshaker. All samples for MS analysis were desalted with C18 ZipTips

(Millipore, USA), dried in a Genevac miVac centrifugal vacuum concentrator, and reconstituted in 0.1% formic acid at a concentration of 50 ng/µL.

Mass Spectrometry

Samples were analysed with an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) coupled to an UltiMate 3000 UHPLC system (Thermo Fisher Scientific). 6 µL (300 ng) of sample were injected onto a Vydac Everest C18 column (75 µm x 75 mm, 2 µm, 300 Å, Hichrom, UK) with an Acclaim PepMap C18 trap column (0.3 mm x 5 mm, 5 µm, 100 Å, Thermo Fisher Scientific). The mobile phases were A: 0.1% formic acid, 1% acetonitrile, and B: 0.1 % formic acid, 80% acetonitrile. Loaded peptides were initially washed for 3 min, and then were eluted at a flow rate of 0.3 µL/min in a mobile phase gradient of 3-8% solution B for 5 min, followed by 8-50% solution B for 37 min. Two separate LC-ESI-MS/MS runs were completed with MS1 spectra acquired at *m/*z 300-1800 and *m/*z 700-1800 for rHuNCAM samples or *m/z* 300-1800 and *m/z* 600-1800 for IgG samples. Full width at half-maximum (FWHM) resolution was 120K at *m/z* 400 with an automatic gain control (AGC) target of 1,000,000 and a maximum injection time of 200 ms. The 10 most intense precursors with signal ≥1000 and charge state >1 were selected for MS2 fragmentation using higher-energy collision-induced dissociation (HCD) with an isolation width of 2 Da. MS2 scans were acquired in the Orbitrap mass analyzer with normalized collision energy of 35%, resolution setting of 30K with an AGC target of 100,000, a maximum injection time of 200 ms, and fixed first mass of *m/z* 140. The MS1 *m/z* range for glycopeptides was optimized based on *in silico* digestions and empirical data acquired using MS1 *m/z* 300-1800. The MS1 *m/z* 300-1800 setting was used to calculate occupancy, while the MS1 *m/z* 700-1800 and *m/z* 600-1800 setting was used to calculate the abundance of glycoforms of rHuNCAM and IgG, respectively.

Data processing

Peptides and glycopeptides were identified using Proteome Discoverer (v2.0.0.802, Thermo Fisher Scientific) with Byonic (v2.13.17, Protein Metrics) as a node to assign peptide-spectrum matches (PSMs). The Byonic search parameters included a maximum of 2 missed cleavages, a precursor mass tolerance of 10 ppm and fragment mass tolerance of 15 ppm. For rHuNCAM, cleavage was set as semi-specific at arginine (R), lysine (K), aspartic acid (D), and glutamic acid (E). For IgG samples cleavage was set as semi-specific at R and K. The peptide output option was set to show all *N*-glycopeptides regardless of score or FDR cutoff as recommended by the manufacturer for simple samples, and decoys were included. Two common modifications and one rare modification were allowed per peptide. Deamidation at asparagine (N) and glutamine (Q), formation of *N*-pyroglutamyl peptide at Q or E, and oxidation of methionine (M) were common modifications; and *N*-glycosylation was a rare modification. Cysteinyl-S-β-propionamide was set as a fixed modification. The glycan database was from Byonic (*N*-glycan 309 mammalian no sodium) with glycans containing NeuGc removed and HexNAc(3)Hex(5)Fuc(1)NeuAc(1) and 16 PSA glycan structures added (Supplementary Table 1). For final calculations, only *N*-glycans with a minimum structure of HexNAc(2)Hex(3) were included. *N*-glycopeptides containing HexNAc or HexNAc(2) were observed, but were not included in further analysis, as assignment of HexNAc or HexNAc(2) can be ambiguous because these structures can exist as *O*-glycans on serine (S) or threonine (T) residues within peptides that contain an *N*-glycosylation sequon. To calculate area under the curve information for precursors the Precursor Ions Area Detector node was run in Proteome Discoverer.

The protein databases contained the amino acid sequence of rHuNCAM (Supplementary Figure 2) or the human proteome database from UniProt (UP000005640 downloaded 20 April 2018; 20,303 reviewed proteins), and a custom contaminants database from Proteome Discoverer and UniProt (304 proteins). PSMs for *N*-glycopeptides were included based on the quality of MS/MS fragmentation with the following minimum requirements: presence of Y1 (peptide + HexNAc) or Y0 (peptide) ions, and presence of relevant HexNAc (*m/z* 204.0866), Hex (*m/z* 163.0601), NeuAc (*m/z* 292.1027) and NeuAc-H_2O (*m/z* 274.0921) glycan oxonium ions. The relative abundance of each glycoform at each site was calculated as the summed intensity of all *N*-glycopeptides with a particular glycan divided by the summed intensities of all *N*-glycopeptides containing that site. The summed intensity of each glycoform at a site included different peptide proteolysis forms, variable peptide modifications, and all charge states. *N*-glycan occupancy was calculated in samples that had been treated with PNGase F as the summed intensities of deamidated previously glycosylated peptides at a specific site (including all variations of peptide sequences) divided by the summed intensities of all deamidated and unmodified peptides for that site.

Precursor area under the curve data was obtained through the Precursor Ions Area Detector node of Proteome Discoverer. These values were used for PSMs identified through the Byonic search. A python script was used to facilitate high-throughput parsing of unique PSMs and calculation of relative abundances (Supplementary information, https://github.com/bschulzlab/glypnir). The script allowed the user to select the protein of interest, a minimum area under the value, and the number of *N*-glycosylation sites allowed per peptide. The script selected only PSMs with a search engine rank of 1, indicating high confidence PSMs. The unique PSMs selected were sorted based on peptide sequence, modifications, and charge state. Only the highest area under the curve for

each unique PSM was chosen for quantification. Each sequon was sorted as either unmodified, deamidated, or glycosylated, and annotated onto the original protein sequence. For glycosylated sequons, the script had an option to group the sequons according to glycan composition. To calculate either occupancy or glycan abundance, the summed area of each unique PSM (based on peptide composition, modifications of the peptide, and peptide charge state) containing the target sequon was divided by the summed area of all unique PSMs containing the target sequon. The output displayed by the script was the occupancy at each sequon or the distribution of glycoforms at each sequon for each sample. All unique glycopeptide PSMs containing sequons of interest used in the computations were manually validated (Supplementary information). The relative abundance of each glycoform at each site was compared between untreated samples and samples treated with EndoNF followed by mild acid hydrolysis using two-tailed, unpaired t-test.

Results and Discussion

Generation of polysialylated recombinant HuNCAM

To develop a glycoproteomics workflow capable of measuring site-specific polysialylation, we required a polysialylated glycoprotein substrate. One of the best-studied polysialylated glycoproteins is NCAM,[11, 75] a single-pass type 1 transmembrane protein with an extracellular domain composed of five Immunoglobulin-like (Ig-like) domains and two fibronectin type III (FN) domains (Figure 2). Human NCAM has six *N*-linked glycosylation sites within the Ig-like domains, and one *N*-linked site close to the transmembrane domain (Figure 2 and Supplementary Figure 2) (https://www.uniprot.org/uniprot/P13591). The 5th and 6th *N*-glycans of native human

NCAM are polysialylated (N459 and N488, numbering as per native human protein sequence) (https://www.uniprot.org/uniprot/P13591).[11, 22-24] To obtain polysialylated human NCAM, we expressed a recombinant soluble human NCAM variant (rHuNCAM) in mammalian cells (Figure 2 and Supplementary Figure 2). Polysialylation of rHuNCAM was achieved by co-expression with human B4GALT1, human ST3GAL4, and human ST8SIA4. To verify that the purified rHuNCAM was polysialylated we performed Western blot analysis using anti-PSA and anti-NCAM antibodies with and without pretreatment with EndoNF endosialidase to remove PSA (Figure 3). EndoNF is a tail-spike protein from bacteriophage K1F that cleaves PSA leaving an oligosialic acid chain attached to the glycan.[37] rHuNCAM showed anti-PSA reactivity (Figure 3, lane 4) that was lost after EndoNF treatment (Figure 3, lane 5), while there was anti-NCAM antibody reactivity in both samples (Figure 3, lanes 6 and 7). We also observed a reduction in the apparent molecular weight of rHuNCAM after EndoNF treatment by Coomassie Blue stained SDS-PAGE (Figure 3, lanes 2 and 3). Together, these results indicated that rHuNCAM was polysialylated. rHuNCAM was subsequently used as our polysialylated glycoprotein substrate for MS workflow development.

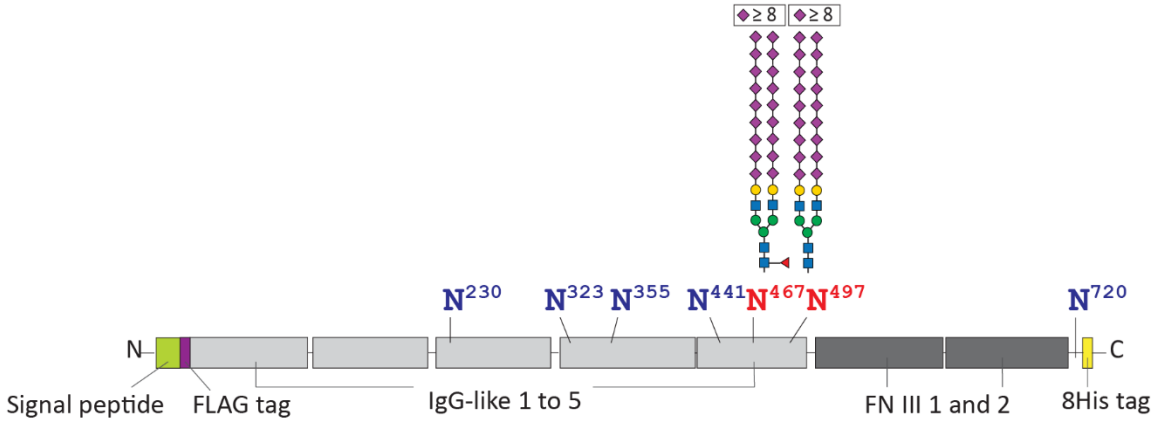


Figure 2. *N*-linked glycosylation sites on rHuNCAM. The protein domain structure of rHuNCAM is shown as per UniProtKB (P13591), with five Ig-like domains (light grey) and two fibronectin type III (FN) domains (dark grey), joined by short linker regions of variable size. The protein is truncated at amino acid 718 to generate a soluble version of HuNCAM; a FLAG tag (violet) is located at the N-terminus after the signal sequence (light green); and an 8His tag is attached to the C-terminal end of rHuNCAM (yellow). There are six *N*-linked glycosylation sites in NCAM's Ig-like domains and one *N*-linked site close to the transmembrane domain. The 5th and 6th *N*-linked sites at N467 and N496 of rHuNCAM are reported to be polysialylated in NCAM.[11, 22-24]

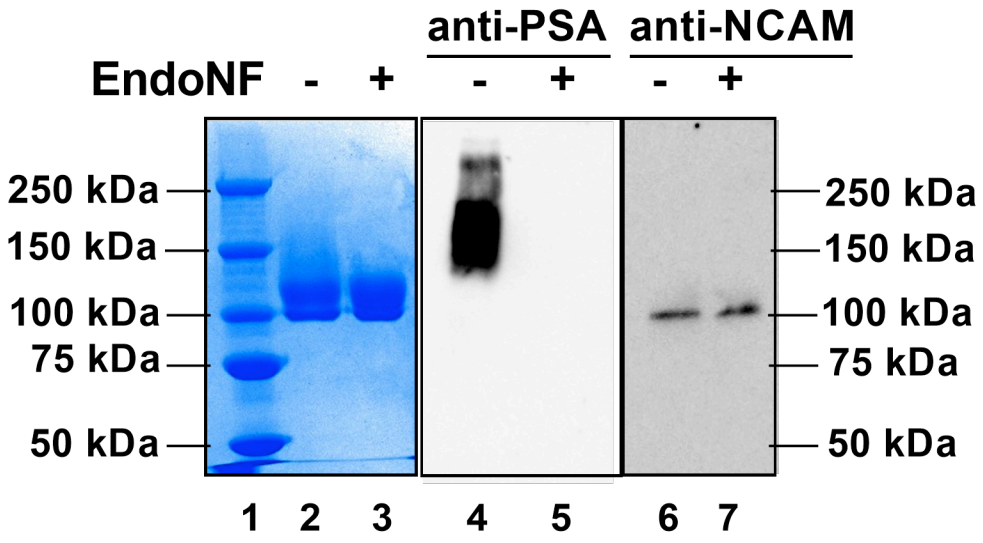


Figure 3. rHuNCAM was partially polysialylated. Purified rHuNCAM was incubated with EndoNF to remove PSA. After incubation, the samples were separated by SDS-PAGE, and either stained with Coomassie Blue (lanes 2 and 3) or transferred to nitrocellulose and used in Western blot analysis (lanes 4-7). PSA and NCAM were detected with anti-PSA (lanes 4-5) and anti-NCAM antibodies (lanes 6-7) respectively.

An LC-ESI-MS/MS method to measure site-specific polysialylation

Polysialylated glycopeptides are difficult to detect by standard LC-ESI-MS/MS glycoproteomic workflows due to the high molecular weight and negative charge of PSA glycans.[76] To enable site-specific measurement of polysialylated glycopeptides by LC-ESI MS/MS, PSA must be removed (Figure 1). Indeed, using standard glycoproteomic workflows with only protease digestion and analysis of intact glycopeptides we could not detect PSA-modified glycopeptides from rHuNCAM (Figure 4A and B, Supplementary Tables 2 and 3). We incorporated EndoNF digestion into a standard glycoproteomic workflow as EndoNF can digest PSA to oligosialic acid,[37, 77, 78] and we had observed that it efficiently removed anti-PSA antibody reactivity from rHuNCAM (Figure 3). We then analysed the types and relative abundance of glycans detected at N467 or N496 in rHuNCAM, which correspond to the 5th and 6th *N*-linked sites of NCAM that are reported to be polysialylated on native NCAM.[11, 22-24] However, we observed no difference in the relative abundances of glycoforms at these sites in rHuNCAM samples with or without EndoNF digestion (Figure 4A and B, Supplementary Table 2 and 3). The predominant glycoform observed at N467 and N496 in both the untreated and EndoNF treated samples was the oligomannose HexNAc(2)Hex(5) glycan (Figure 4A and B, Supplementary Table 2 and 3). These results suggested that oligosialylated glycopeptides were not detectable by LC-ESI-MS/MS. The negatively charged oligosialic acid chain remaining after EndoNF treatment may reduce the ionisation efficiency of the glycopeptides in positive ion mode, resulting in diminished MS1 signals and limiting detection. To overcome this problem, we treated the samples with mild acid hydrolysis, to completely remove sialic acids.[16, 73, 74] As expected, after mild acid hydrolysis we could only detect asialylated glycoforms by MS (Figure 4A and B, Supplementary Table 2 and 3). This treatment also allowed for the detection of several new glycoforms at both N467 and N496

(Figure 4A and B, Supplementary Table 2 and 3). At N467, glycopeptides with HexNAc(3)Hex(4)Fuc(1), HexNAc(3)Hex(5)Fuc(1), HexNAc(3)Hex(6), HexNAc(3)Hex(6)Fuc(1), and HexNAc(4)Hex(5)Fuc(1) glycans were previously undetectable, and increased in abundance after acid hydrolysis (Figure 4A, Supplementary Table 2). Similarly, at the N496 site, glycopeptides with HexNAc(3)Hex(4), HexNAc(3)Hex(5), HexNAc(3)Hex(6), and HexNAc(4)Hex(5) glycans were only detectable after mild acid hydrolysis (Figure 4B and 5, Supplementary Table 3). The new glycoforms detected after hydrolysis (hybrid and complex *N*-glycans) were also the glycan scaffolds that have been previously reported to be polysialylated in bovine NCAM.[64] Therefore, complete removal of sialic acids by mild acid hydrolysis allowed for the detection and relative quantification of polysialylated glycopeptides by LC-ESI-MS/MS.

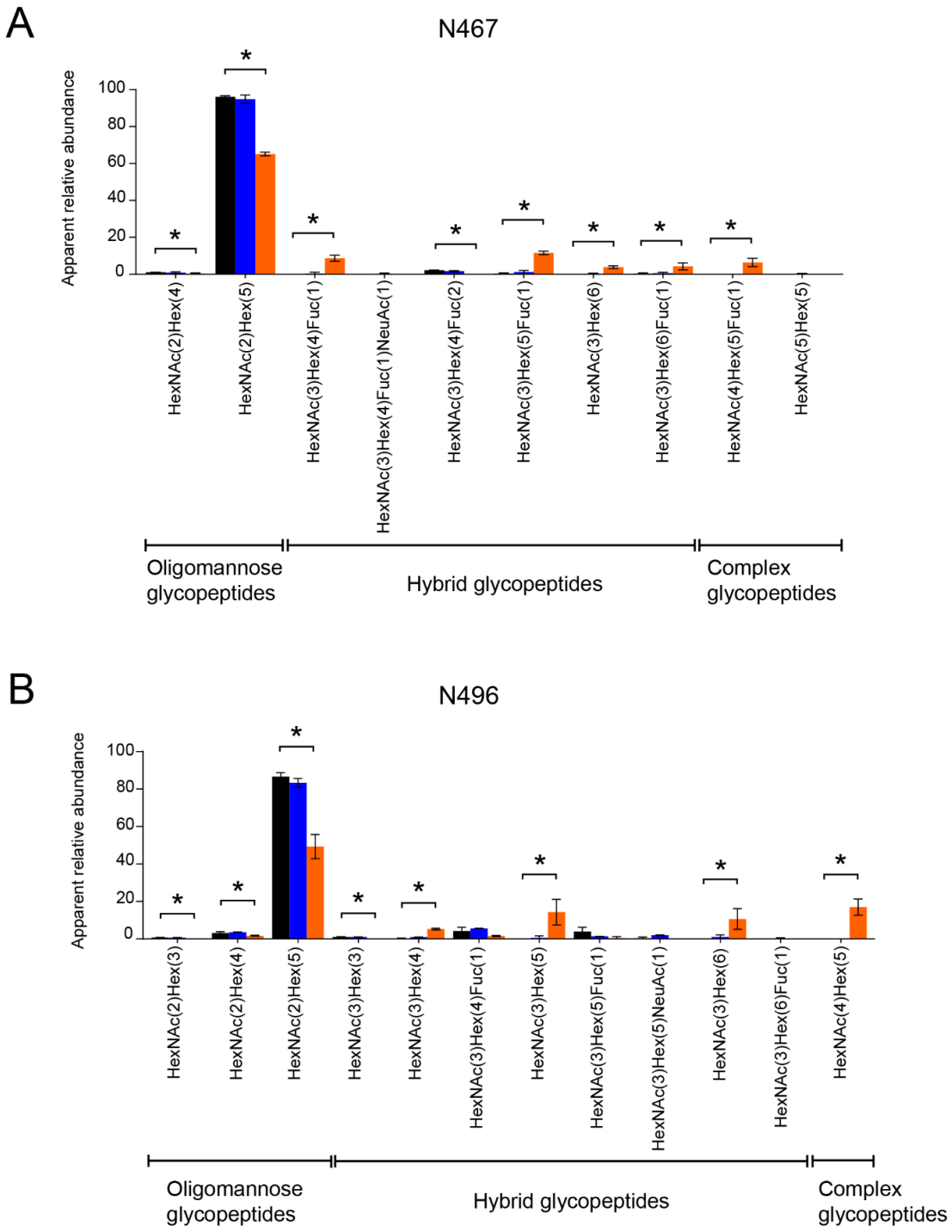


Figure 4. Desialylation allows inference of site specific polysialylation on rHuNCAM. Relative abundance of glycoforms detected by LC-ESI-MS/MS at glycosylation sites **(A)** N467 and **(B)** N496 of rHuNCAM. Black, untreated; blue, EndoNF; orange, EndoNF and mild acid hydrolysis. Values show mean ± SD, n=3. *, $p < 0.05$. Full data is shown in Supplementary Tables 2 and 3.

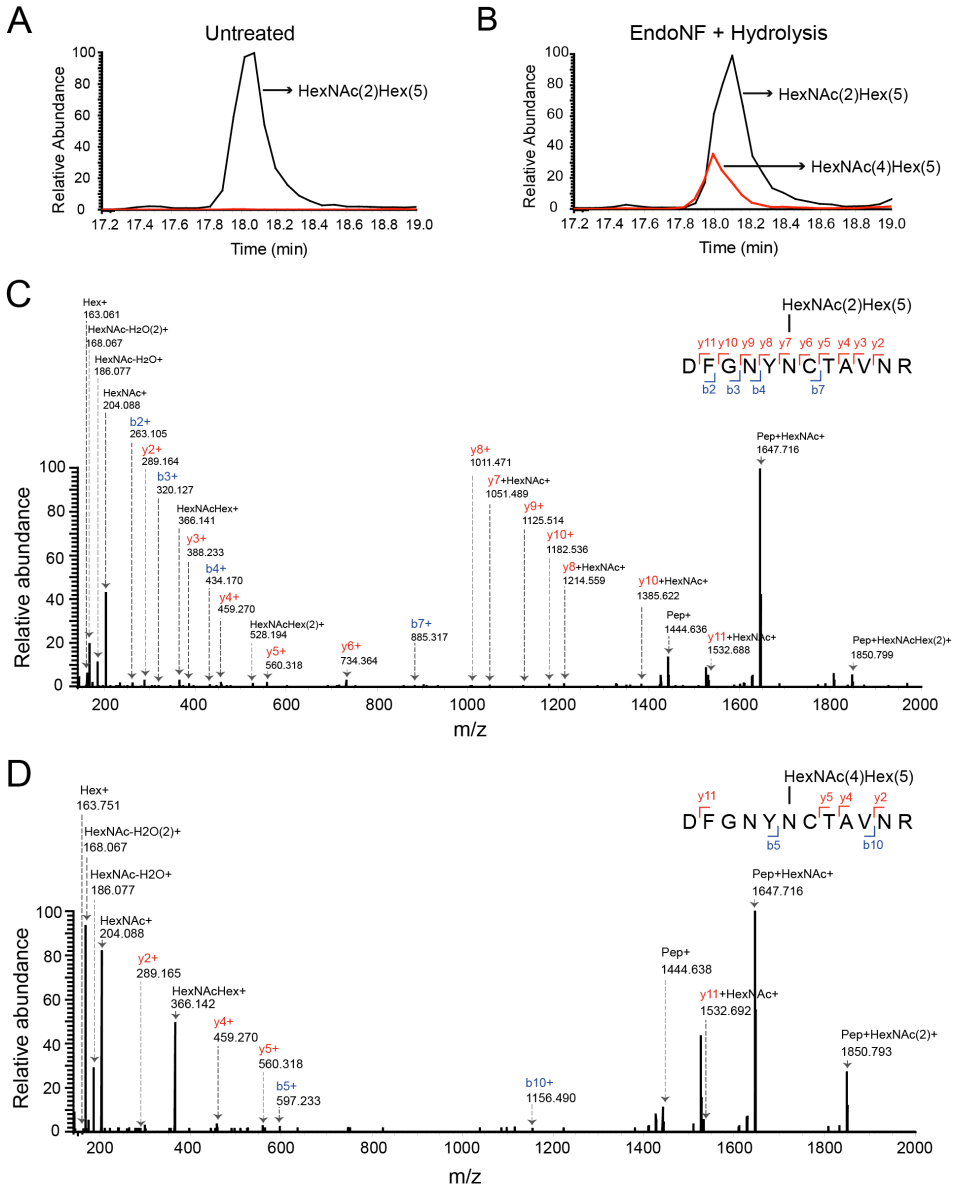


Figure 5. Apparent glycoform abundance of selected DFGNYN^{496}CTAVNR glycopeptides of rHuNCAM with and without desialylation. Extracted ion chromatograms of DFGNYNCTAVNR glycopeptide with oligomannose HexNAc(2)Hex(5) (non-polysialylated glycan) or biantennary HexNAc(4)Hex(5) (de-polysialylated glycan) in (**A**) untreated or (**B**) EndoNF and mild acid hydrolysis treated rHuNCAM samples. MS/MS spectra identifying DFGNYNCTAVNR glycopeptides with (**C**) HexNAc(2)Hex(5) or (**D**) HexNAc(4)Hex(5) glycan. All cysteines (C) are alkylated.

To allow focused analysis of glycopeptides containing N467 and N496, the sites known to be polysialylated in native NCAM,[11, 22-24] we optimized protease digestion and LC-ESI-MS/MS acquisition methods accordingly. However, this generated suboptimal conditions for detection of other *N*-glycopeptides from rHuNCAM. Therefore, to verify that our workflow did not perturb analysis of glycopeptides without PSA, we tested the workflow on serum-purified human immunoglobulin G (IgG), a non-polysialylated glycoprotein (Figure 6, Supplementary Table 4). This analysis focused on IgG2, for which glycopeptides were most robustly detected. IgG2 has one *N*-linked glycosylation site at N176 (https://www.uniprot.org/uniprot/P01859), which is predominantly occupied by bi-antennary glycans.[79] IgG2 glycoforms observed in the untreated sample included HexNAc(4)Hex(3)Fuc(1), HexNAc(4)Hex(4)Fuc(1), HexNAc(4)Hex(5)Fuc(1), HexNAc(4)Hex(4)Fuc(1)NeuAc(1), and HexNAc(4)Hex(5)Fuc(1)NeuAc(1) (Figure 6, Supplementary Table 4). This glycoform distribution is similar to previous reports.[79] After EndoNF treatment, we did not detect any new glycoforms on IgG2, and there was no significant change in the relative abundance of the observed glycoforms compared to the untreated samples (Figure 6, Supplementary Table 4). This result was expected, given that IgG2 is not reported to contain PSA-modified glycans. Importantly, the only changes observed in IgG2 glycoform structure and abundance in samples treated with EndoNF and mild acid hydrolysis compared to untreated samples were the disappearance of previously monosialylated structures, and the increase in abundance of the corresponding unsialylated forms of these structures (Figure 6, Supplementary Table 4). For example, we observed the disappearance of the HexNAc(4)Hex(4)Fuc(1)NeuAc(1) and HexNAc(4)Hex(5)Fuc(1)NeuAc(1) glycoforms in desialylated samples, and an increase in the abundance of the corresponding HexNAc(4)Hex(4)Fuc(1) and HexNAc(4)Hex(5)Fuc(1) structures (Figure 6, Supplementary

Table 4). We did not detect any new glycoform in the desialylated sample. Since the relative abundance of nonsialylated glycoforms increases after desialylation, our workflow is more effective when comparing potentially polysialylated glycoproteins with their non-polysialylated counterparts. Together, these results show that our workflow is a useful method to detect site-specific polysialylation in glycoproteins.

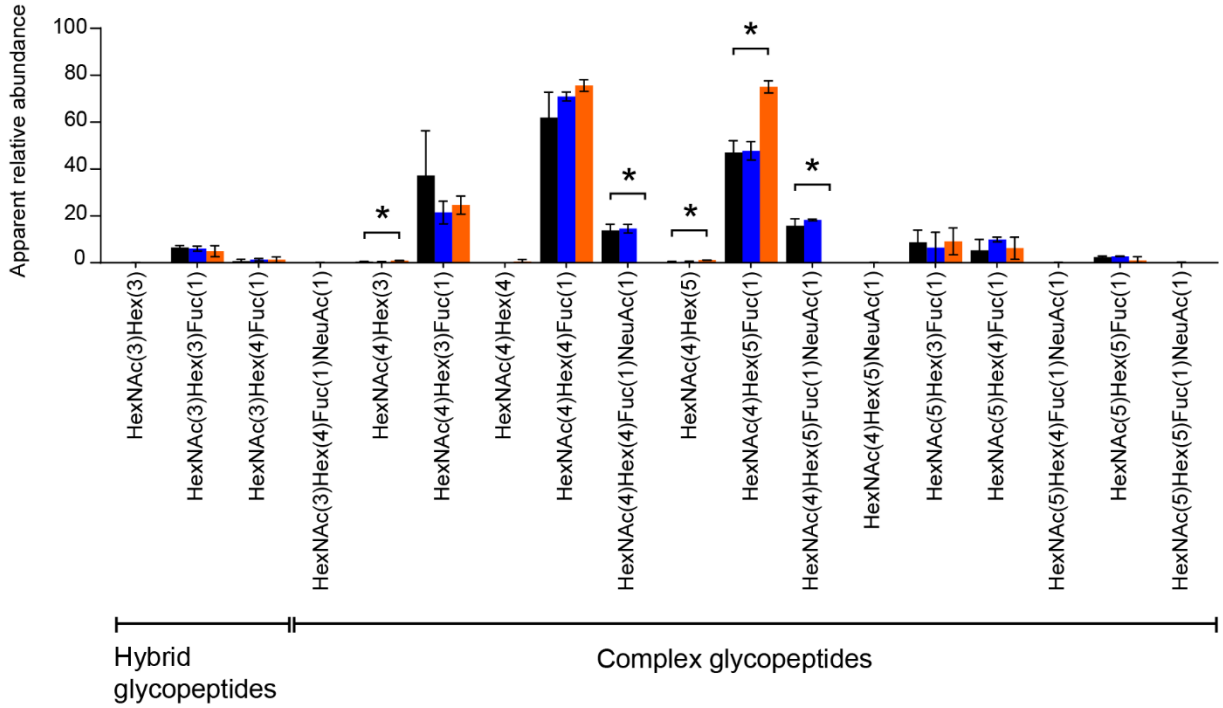


Figure 6. Desialylation does not strongly affect the glycan structures on human serum IgG. Relative abundance of glycoforms detected by LC-ESI-MS/MS at the glycosylation site N176 of IgG2. Black, untreated; blue, EndoNF; orange, EndoNF and mild acid hydrolysis. Values show mean ± SD, n=3. *, $p < 0.05$. Full data is shown in Supplementary Table 4.

Conclusions

PSA has key physiological roles and therapeutic applications. To facilitate the study of polysialylation and the development of new PSA technologies, we have developed an LC-ESI-MS/MS method that allows for the detection and measurement of site-specific polysialylation. Our method combines carefully selected proteases to generate peptides of appropriate length for LC-ESI-MS/MS detection with enzymatic and chemical methods to eliminate sialic acids and newly developed Python scripts for data processing and calculation of glycoform abundance. We have demonstrated that this method distinguishes PSA- and non-PSA modified glycopeptides, can identify the precise peptides modified by PSA, and allows measurement of the site-specific extent of polysialylation.

Acknowledgements

We thank Dr Amanda Nouwens and Peter Josh at The University of Queensland, School of Chemistry and Molecular Biosciences Mass Spectrometry Facility for their assistance and expertise.

Conflict of Interest

PX, CMO and MPH are employees of CSL Ltd

Funding

BLS was funded by an Australian National Health and Medical Research Council RD Wright Biomedical (CDF Level 2) Fellowship APP1087975. This work was funded by an Australian Research Council Discovery Project DP160102766 to BLS and an Australian Research Council Industrial Transformation Training Centre IC160100027 to BLS, CBH, and CMO.

References

1. Colley, K.J., A. Varki, and T. Kinoshita, *Cellular Organization of Glycosylation*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015: Cold Spring Harbor (NY). p. 41-49.
2. Zacchi, L.F., et al., *Endoplasmic Reticulum-Associated Degradation and Protein Quality Control*, in *Encyclopedia of Cell Biology*, R.A.B.a.P.D. Stahl, Editor. 2016, Academic Press. p. 596-611.
3. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annu Rev Biochem, 2004. **73**: p. 1019--1049.
4. Caramelo, J.J. and A.J. Parodi, *A sweet code for glycoprotein folding.* FEBS Lett, 2015. **589**(22): p. 3379-87.
5. Stanley, P., N. Taniguchi, and M. Aebi, *N-Glycans*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015: Cold Spring Harbor (NY). p. 99-111.
6. Rini, J., J. Esko, and A. Varki, *Chapter 5 Glycosyltransferases and Glycan-processing Enzymes*, in *Essentials of Glycobiology [Internet]*, V. A, et al., Editors. 2015-2017, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
7. Rudd, P.M. and R.A. Dwek, *Glycosylation: heterogeneity and the 3D structure of proteins.* Crit Rev Biochem Mol Biol, 1997. **32**(1): p. 1-100.
8. Zacchi, L.F. and B.L. Schulz, *N-glycoprotein macroheterogeneity: biological implications and proteomic characterization.* Glycoconj J, 2016. **33**(3): p. 359-76.
9. Varki, A., R.L. Schnaar, and R. Schauer, *Sialic Acids and Other Nonulosonic Acids*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015: Cold Spring Harbor (NY). p. 179-195.
10. Sato, C. and K. Kitajima, *Disialic, oligosialic and polysialic acids: distribution, functions and related disease.* The Journal of Biochemistry, 2013. **154**(2): p. 115-136.
11. Colley, K.J., K. Kitajima, and C. Sato, *Polysialic acid: biosynthesis, novel functions and applications.* Crit Rev Biochem Mol Biol, 2014. **49**(6): p. 498-532.

12. Finne, J., et al., *Occurrence of alpha 2-8 linked polysialosyl units in a neural cell adhesion molecule.* Biochem Biophys Res Commun, 1983. **112**(2): p. 482-7.
13. Sebastian, P.G., et al., *Synaptic cell adhesion molecule SynCAM 1 is a target for polysialylation in postnatal mouse brain.* Proceedings of the National Academy of Sciences, 2010. **107**(22): p. 10250.
14. Curreli, S., et al., *Polysialylated Neuropilin-2 Is Expressed on the Surface of Human Dendritic Cells and Modulates Dendritic Cell-T Lymphocyte Interactions.* Polysialylated Neuropilin-2 Is Expressed on the Surface of Human Dendritic Cells and Modulates Dendritic Cell-T Lymphocyte Interactions, 2007. **282**(42): p. 30346-30356.
15. Zuber, C., et al., *Polysialic acid is associated with sodium channels and the neural cell adhesion molecule N-CAM in adult rat brain.* The Journal of biological chemistry, 1992. **267**(14): p. 9965-9971.
16. James, W.M. and W.S. Agnew, *Multiple oligosaccharide chains in the voltage-sensitive Na channel from electrophorus electricus: evidence for alpha-2,8-linked polysialic acid.* Biochem Biophys Res Commun, 1987. **148**(2): p. 817-26.
17. Kiermaier, E., et al., *Polysialylation controls dendritic cell trafficking by regulating chemokine recognition.* Science, 2016. **351**(6269): p. 186-90.
18. Werneburg, S., et al., *Polysialylation and lipopolysaccharide-induced shedding of E-selectin ligand-1 and neuropilin-2 by microglia and THP-1 macrophages.* Glia, 2016. **64**(8): p. 1314-30.
19. Yabe, U., et al., *Polysialic acid in human milk. CD36 is a new member of mammalian polysialic acid-containing glycoprotein.* The Journal of biological chemistry, 2003. **278**(16): p. 13875.
20. Muhlenhoff, M., et al., *Autocatalytic polysialylation of polysialyltransferase-1.* Embo j, 1996. **15**(24): p. 6943-50.
21. Close, B.E. and K.J. Colley, *In vivo autopolysialylation and localization of the polysialyltransferases PST and STX.* J Biol Chem, 1998. **273**(51): p. 34586-93.
22. Nelson, R.W., P.A. Bates, and U. Rutishauser, *Protein determinants for specific polysialylation of the neural cell adhesion molecule.* The Journal of biological chemistry, 1995. **270**(29): p. 17171.
23. Close, B.E., et al., *The minimal structural domains required for neural cell adhesion molecule polysialylation by PST/ST8Sia IV and STX/ST8Sia II.* The Journal of biological chemistry, 2003. **278**(33): p. 30796.
24. Foley, D.A., et al., *Structure and Mutagenesis of Neural Cell Adhesion Molecule Domains Evidence for Flexibility in the Placement of Polysialic Acid Attachment Sites.* J. Biol. Chem., 2010. **285**(35).
25. Sato, C. and K. Kitajima, *Impact of structural aberrancy of polysialic acid and its synthetic enzyme ST8SIA2 in schizophrenia.* Front. Cell. Neurosci., 2013. **7**.
26. Barbeau, D., et al., *Decreased Expression of the Embryonic Form of the Neural Cell Adhesion Molecule in Schizophrenic Brains.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(7): p. 2785-2789.
27. Sara, M.E., et al., *Polysialic acid sustains cancer cell survival and migratory capacity in a hypoxic environment.* Scientific Reports, 2016. **6**(1).
28. El Maarouf, A. and U. Rutishauser, *Use of PSA-NCAM in repair of the central nervous system.* Advances in Experimental Medicine and Biology, 2010. **663**: p. 137-147.

29. Bader, R.A. and P.R. Wardwell, *Polysialic acid: overcoming the hurdles of drug delivery.* Ther Deliv, 2014. **5**(3): p. 235-7.
30. Wang, X.-J., et al., *Endogenous Polysialic Acid Based Micelles for Calmodulin Antagonist Delivery against Vascular Dementia.* ACS applied materials & interfaces, 2016. **8**(51): p. 35045-35058.
31. Zhang, R., et al., *Development and testing of solid dose formulations containing polysialic acid insulin conjugate: next generation of long-acting insulin.* Journal of diabetes science and technology, 2010. **4**(3): p. 532.
32. Kontermann, R.E., *Strategies for extended serum half-life of protein therapeutics.* Current Opinion in Biotechnology, 2011. **22**(6): p. 868-876.
33. Fernandes, A.I. and G. Gregoriadis, *The effect of polysialylation on the immunogenicity and antigenicity of asparaginase: implication in its pharmacokinetics.* International Journal of Pharmaceutics, 2001. **217**(1): p. 215-224.
34. Lindhout, T., et al., *Site-specific enzymatic polysialylation of therapeutic proteins using bacterial enzymes.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7397-402.
35. Frosch, M., et al., *NZB mouse system for production of monoclonal antibodies to weak bacterial antigens: isolation of an IgG antibody to the polysaccharide capsules of Escherichia coli K1 and group B meningococci.* Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(4): p. 1194.
36. Sato, C., et al., *Characterization of the antigenic specificity of four different anti- (α2→8-linked polysialic acid) antibodies using lipid-conjugated oligo/polysialic acids.* Journal of Biological Chemistry, 1995. **270**(32): p. 18923-18928.
37. Morley, T.J., et al., *New Sialidase Mechanism: BACTERIOPHAGE K1F ENDO-SIALIDASE IS AN INVERTING GLYCOSIDASE.* New Sialidase Mechanism: BACTERIOPHAGE K1F ENDO-SIALIDASE IS AN INVERTING GLYCOSIDASE, 2009. **284**(26): p. 17404-17410.
38. Katharina, S., et al., *Crystal structure of the polysialic acid–degrading endosialidase of bacteriophage K1F.* Nature Structural & Molecular Biology, 2004. **12**(1): p. 90.
39. Aalto, J., et al., *Mutant bacteriophage with non-catalytic endosialidase binds to both bacterial and eukaryotic polysialic acid and can be used as probe for its detection.* Official Journal of the International Glycoconjugate Organization, 2001. **18**(10): p. 751-758.
40. Kitajima, K., N. Varki, and C. Sato, *Advanced Technologies in Sialic Acid and Sialoglycoconjugate Analysis.* Top Curr Chem, 2015. **367**: p. 75-103.
41. Galuska, S.P., *Chapter 13: Advances in Sialic Acid and Polysialic Acid Detection Methodologies*, in *Sialobiology: Structure, Biosynthesis and Function*, J.T.a.I. Martinez-Duncker, Editor. 2013, Bentham Science Publishers. p. 448-475.
42. Inoue, S. and Y. Inoue, *A challenge to the ultrasensitive chemical method for the analysis of oligo- and polysialic acids at a nanogram level of colominic acid and a milligram level of brain tissues.* Biochimie, 2001. **83**(7): p. 605-613.
43. Inoue, et al., *An ultrasensitive chemical method for polysialic acid analysis.* Glycobiology, 2001. **11**(9): p. 759-767.
44. Inoue, S. and Y. Inoue, *Ultrasensitive analysis of sialic acids and oligo/polysialic acids by fluorometric high-performance liquid chromatography.* Methods Enzymol, 2003. **362**: p. 543-60.

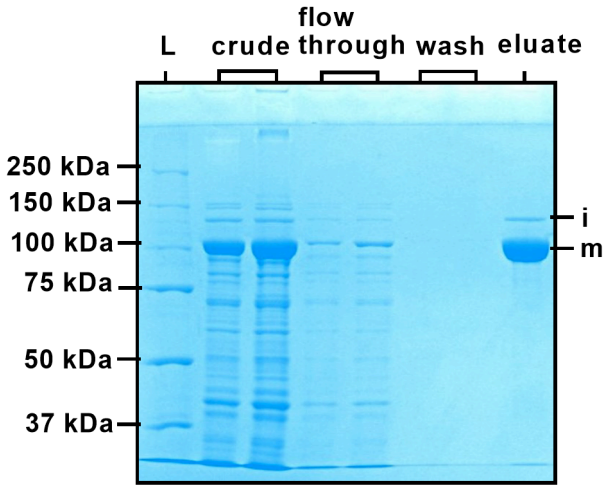
45. Sato, C., et al., *Fluorescent-Assisted Detection of Oligosialyl Units in Glycoconjugates.* Analytical Biochemistry, 1999. **266**(1): p. 102-109.
46. Sato, C., et al., *Development of a highly sensitive chemical method for detecting alpha2-->8-linked oligo/polysialic acid residues in glycoproteins blotted on the membrane.* Anal Biochem, 1998. **261**(2): p. 191-7.
47. Rohr, T.E. and F.A. Troy, *Structure and biosynthesis of surface polymers containing polysialic acid in Escherichia coli.* J Biol Chem, 1980. **255**(6): p. 2332-42.
48. Zhang, Y., et al., *Separation of Oligo/Polymers of 5-N-Acetylneuraminic Acid, 5-N-Glycolylneuraminic Acid, and 2-Keto-3-deoxy-d-glycero-d-galacto-nononic Acid by High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detector.* Analytical Biochemistry, 1997. **250**(2): p. 245-251.
49. Shu-Ling, L., Y. Inoue, and S. Inoue, *Evaluation of high-performance anion-exchange chromatography with pulsed electrochemical and fluorometric detection for extensive application to the analysisof homologous series of oligo- and polysialic acids in bioactive molecules.* Glycobiology, 1999. **9**(8): p. 807.
50. Michon, F., J.R. Brisson, and H.J. Jennings, *Conformational Differences between Linear α(2→8)-Linked Homosialooligosaccharides and the Epitope of the Group B Meningococcal Polysaccharide.* Biochemistry, 1987. **26**(25): p. 8399-8405.
51. Inoue, S. and M. Iwasaki, *Isolation of a novel glycoprotein from the eggs of rainbow trout: Occurrence of disialosyl groups on all carbohydrate chains.* Biochemical and Biophysical Research Communications, 1978. **83**(3): p. 1018-1023.
52. Ijuin, T., et al., *Isolation and identification of novel sulfated and nonsulfated oligosialyl glycosphingolipids from sea urchin sperm.* Official Journal of the International Glycoconjugate Organization, 1996. **13**(3): p. 401-413.
53. Kitazume, S., et al., *Identification of polysialic acid-containing glycoprotein in the jelly coat of sea urchin eggs. Occurrence of a novel type of polysialic acid structure.* J Biol Chem, 1994. **269**(36): p. 22712-8.
54. Thaysen-Andersen, M., et al., *Structural analysis of glycoprotein sialylation – Part I: pre-LC-MS analytical strategies* RSC Advances, 2013. **3**(45): p. 22683-22705.
55. Palmisano, G., et al., *Structural analysis of glycoprotein sialylation – part II: LC-MS based detection.* RSC Advances, 2013. **3**(45): p. 22706-22726.
56. Galuska, C.E., K. Maass, and S.P. Galuska, *Mass Spectrometric Analysis of Oligo- and Polysialic Acids.* Methods Mol Biol, 2015. **1321**: p. 417-26.
57. Galuska, S.P., et al., *Glycomic strategy for efficient linkage analysis of di-, oligo- and polysialic acids.* J Proteomics, 2012. **75**(17): p. 5266-78.
58. Galuska, S., et al., *Characterization of oligo- and polysialic acids by MALDI-TOF-MIS.* Anal. Chem., 2007. **79**(18): p. 7161-7169.
59. Galuska, S.P., et al., *Mass spectrometric fragmentation analysis of oligosialic and polysialic Acids.(Author abstract).* Analytical Chemistry, 2010. **82**(5): p. 2059.
60. Miyata, S., et al., *A major flagellum sialoglycoprotein in sea urchin sperm contains a novel polysialic acid, an 2,9-linked poly- N -acetylneuraminic acid chain, capped by an 8- O -sulfated sialic acid residue.* Glycobiology, 2004. **14**(9): p. 827-840.
61. Kronewitter, S.R., et al., *Polysialylated N-glycans identified in human serum through combined developments in sample preparation, separations, and electrospray ionization-mass spectrometry.* Anal Chem, 2014. **86**(17): p. 8700-10.

62. Rollenhagen, M., et al., *Polysialic acid on neuropilin-2 is exclusively synthesized by the polysialyltransferase ST8SiaIV and attached to mucin-type o-glycans located between the b2 and c domain.* J Biol Chem, 2013. **288**(32): p. 22880-92.
63. Liedtke, S., et al., *Characterization of N-glycans from mouse brain neural cell adhesion molecule.* Glycobiology, 2001. **11**(5): p. 373-384.
64. Von Der Ohe, M., et al., *Localization and characterization of polysialic acid-containing N-linked glycans from bovine NCAM.* Glycobiology, 2002. **12**(1): p. 47-47.
65. Chen, C.-G., et al., *One-step zero-background IgG reformatting of phage-displayed antibody fragments enabling rapid and high-throughput lead identification.* Nucleic acids research, 2014. **42**(4): p. e26-e26.
66. Pham, P.L., et al., *Transient gene expression in HEK293 cells: Peptone addition posttransfection improves recombinant protein synthesis.* Biotechnology and Bioengineering, 2005. **90**(3): p. 332-344.
67. Schmidt, P.M., et al., *A robust robotic high-throughput antibody purification platform.* Journal of Chromatography A, 2016. **1455**: p. 9-19.
68. Panousis, C., et al., *CSL311, a novel, potent, therapeutic monoclonal antibody for the treatment of diseases mediated by the common β chain of the IL-3, GM-CSF and IL-5 receptors.* mAbs, 2016. **8**(3).
69. Schaer, C.A., et al., *Phenotype-specific recombinant haptoglobin polymers co-expressed with C1r-like protein as optimized hemoglobin-binding therapeutics.* BMC Biotechnology, 2018. **18**(1): p. <xocs:firstpage xmlns:xocs=""/>.
70. Jamaluddin, M.F.B., et al., *Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p.* Protein Science, 2011. **20**(5): p. 849-855.
71. Nguyen, L.T., et al., *Adipose tissue proteomic analyses to study puberty in Brahman heifers.* Journal of Animal Science, 2018. **96**(6): p. 2392-2398.
72. Bailey, U.-M., et al., *Analysis of the extreme diversity of salivary alpha-amylase isoforms generated by physiological proteolysis using liquid chromatography–tandem mass spectrometry.* Journal of Chromatography B, 2012. **911**: p. 21-26.
73. Varki, A. and S. Diaz, *The release and purification of sialic acids from glycoconjugates: Methods to minimize the loss and migration of O-acetyl groups.* Analytical Biochemistry, 1984. **137**(1): p. 236-247.
74. Manzi, A., *Acid hydrolysis for release of monosaccharides.* Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], 2001. **17**: p. Unit17.16.
75. Berezin, V., *Structure and Function of the Neural Cell Adhesion Molecule NCAM*. 2010: New York, NY : Springer New York.
76. Nwosu, C.C., et al., *Enhanced detection and identification of glycopeptides in negative ion mode mass spectrometry.* Anal Chem, 2010. **82**(23): p. 9654-62.
77. Jakobsson, E., et al., *Endosialidases: Versatile Tools for the Study of Polysialic Acid.* Top Curr Chem, 2015. **367**: p. 29-73.
78. Schulz, E.C., et al., *Structural Basis for the Recognition and Cleavage of Polysialic Acid by the Bacteriophage K1F Tailspike Protein EndoNF.* Journal of Molecular Biology, 2010. **397**(1): p. 341-351.
79. Wuhrer, M., et al., *Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum.* Proteomics, 2007. **7**(22): p. 4070-81.

80. Gerardy-Schahn, R., P. Delannoy, and M. von Itzstein, *SialoGlyco Chemistry and Biology II Tools and Techniques to Identify and Capture Sialoglycans*. 1st ed. 2015.. ed. 2015: Cham : Springer International Publishing : Imprint: Springer.
81. Schwarzer, D., et al., *Proteolytic release of the intramolecular chaperone domain confers processivity to endosialidase F.* Journal of Biological Chemistry, 2009. **284**(14): p. 9465-9474.
82. Finne, J., *Occurrence of unique polysialosyl carbohydrate units in glycoproteins of developing brain.* Journal of Biological Chemistry, 1982. **257**(20): p. 11966-11970.

Supplementary Information

Supplementary Figures



Supplementary Figure 1. Expression and purification of EndoNF. EndoNF expressed in an *E. coli* Rosetta strain was purified using amylose-agarose resin. Aliquots from the crude, flow through, wash, and eluate samples were separated by SDS-PAGE and the gel was stained with Coomassie Blue. EndoNF appeared as two bands in the eluate. The most prominent band was ~ 118 kDa (marked 'm'), corresponding to the mature EndoNF enzyme; while the top band of ~ 136 kDa corresponded to the immature EndoNF (marked 'i') that still contains the C-terminal chaperone domain (verified by in gel digestion and mass spectrometry analysis (data not shown)) which becomes self-cleaved during maturation [80, 81].

MLQTKDLIWTLFFLGTAVSDYKDDDDKLQVDIVPSQGEISVGESKFFLCQVAGDA
KDKDISWFSPNGEKLTPNQQRISVVWNDDSSSTLTIYNANIDDAGIYKCVVTGED
GSESEATVNVKIFQKLMFKNAPTPQEFREGEDAVIVCDVVSSLPPTIIWKHKGRD
VILKKDVRFIVLSNNYLQIRGIKKTDEGTYRCEGRILARGEINFKDIQVIVNVPP
TIQARQNIV**N**^{230}ATANLGQSVTLVCDAEGFPEPTMSWTKDGEQIEQEEDDEKYIF
SDDSSQLTIKKVDKNDEAEYICIAENKAGEQDATIHLKVFAKPKITYVE**N**^{323}QTA
MELEEQVTLTCEASGDPIPSITWRTSTR**N**^{355}ISSEEKASWTRPEKQETLDGHMVV
RSHARVSSLTLKSIQYTDAGEYICTASNTIGQDSQSMYLEVQYAPKLQGPVAVYT
WEGNQV**N**^{441}ITCEVFAYPSATISWFRDGQLLPSS**N**^{467}YSNIKIYNTPSASYLEVT
PDSENDFGNY**N**^{496}CTAVNRIGQESLEFILVQADTPSSPSIDQVEPYSSTAQVQFD
EPEATGGVPILKYKAEWRAVGEEVWHSKWYDAKEASMEGIVTIVGLKPETTYAVR
LAALNGKGLGEISAASEFKTQPVQGEPSAPKLEGQMGEDGNSIKVNLIKQDDGGS
PIRHYLVRYRALSSEWKPEIRLPSGSDHVMLKSLDWNAEYEVYVVAENQQGKSKA
AHFVFRTSAQPTAIPA**N**^{720}GSPTSGHHHHHHHH

Supplementary Figure 2. Amino acid sequence of the HuNCAM(1-19)-FLAG-HuNCAM(20-718)-8His used in this study. The amino acid sequence of rHuNCAM consists of an endogenous signal peptide (green), followed by a FLAG tag (purple), then the NCAM open reading frame truncated at amino acid 718 to remove the transmembrane domain, and finally an 8His tag attached to the C-terminal end of rHuNCAM (yellow). The asparagine residues within the *N*-linked glycosylation sites are highlighted in blue (no polysialylation) or red (polysialylated) based on previous reports [22, 23, 82]

Supplementary Tables

Supplementary Table 1. Additional polysialylated *N*-glycan structures added to the Byonic database
HexNAc(4)Hex(5)Fuc(1)NeuAc(3)
HexNAc(4)Hex(5)Fuc(1)NeuAc(4)
HexNAc(4)Hex(5)Fuc(1)NeuAc(5)
HexNAc(4)Hex(5)Fuc(1)NeuAc(6)
HexNAc(4)Hex(5)Fuc(1)NeuAc(7)
HexNAc(4)Hex(5)Fuc(1)NeuAc(8)
HexNAc(4)Hex(5)Fuc(1)NeuAc(9)
HexNAc(4)Hex(5)Fuc(1)NeuAc(10)
HexNAc(4)Hex(5)NeuAc(3)
HexNAc(4)Hex(5)NeuAc(4)
HexNAc(4)Hex(5)NeuAc(5)
HexNAc(4)Hex(5)NeuAc(6)
HexNAc(4)Hex(5)NeuAc(7)
HexNAc(4)Hex(5)NeuAc(8)
HexNAc(4)Hex(5)NeuAc(9)
HexNAc(4)Hex(5)NeuAc(10)

Supplementary Table 2. Glycoform abundance at N467 in rHuNCAM in untreated, EndoNF only treated, and EndoNF and mild acid hydrolysis treated samples. *, $p < 0.05$ between untreated versus EndoNF and mild acid hydrolysis treated.

	Untreated	**EndoNF**	**EndoNF and mild acid hydrolysis**
Glycoforms	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**
HexNAc(2)Hex(4)	1.1 ± 0.1	0.9 ± 0.4	0.2 ± 0.4*
HexNAc(2)Hex(5)	96.1 ± 0.5	94.8 ± 2.2	65.1 ± 1.0*
HexNAc(3)Hex(4)Fuc(1)		0.4 ± 0.7	8.7 ± 1.7*
HexNAc(3)Hex(4)Fuc(1)NeuAc(1)		0.2 ± 0.4	
HexNAc(3)Hex(4)Fuc(2)	2.1 ± 0.2	1.6 ± 0.3	*
HexNAc(3)Hex(5)Fuc(1)	0.4 ± 0.3	1.2 ± 0.9	11.6 ± 1.0*
HexNAc(3)Hex(6)		0.2 ± 0.4	3.8 ± 0.7*
HexNAc(3)Hex(6)Fuc(1)	0.4 ± 0.3	0.6 ± 0.5	4.2 ± 1.9*
HexNAc(4)Hex(5)Fuc(1)			6.4 ± 2.2*
HexNAc(5)Hex(5)		0.2 ± 0.3	

Supplementary Table 3. Glycoform abundance at N496 in rHuNCAM in untreated, EndoNF only treated, and EndoNF and mild acid hydrolysis treated samples. *, $p < 0.05$ between untreated versus EndoNF and mild acid hydrolysis treated.

	Untreated	EndoNF	EndoNF and mild acid hydrolysis
Glycoforms	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**
HexNAc(2)Hex(3)	0.7 ± 0.1	0.6 ± 0.1	*
HexNAc(2)Hex(4)	3.1 ± 0.7	3.6 ± 0.1	1.7 ± 0.2*
HexNAc(2)Hex(5)	86.5 ± 2.2	83.3 ± 2.3	49.3 ± 6.4*
HexNAc(3)Hex(3)	1.0 ± 0.1	1.0 ± 0.1	*
HexNAc(3)Hex(4)	0.2 ± 0.3	0.9 ± 0.1	5.2 ± 0.4*
HexNAc(3)Hex(4)Fuc(1)	4.2 ± 2.0	5.6 ± 0.1	1.5 ± 0.2
HexNAc(3)Hex(5)		0.6 ± 1.0	14.3 ± 6.8*
HexNAc(3)Hex(5)Fuc(1)	3.9 ± 2.3	1.4 ± 0.1	0.4 ± 0.7
HexNAc(3)Hex(5)NeuAc(1)	0.4 ± 0.6	2.0 ± 0.2	
HexNAc(3)Hex(6)		1.0 ± 1.2	10.6 ± 5.5*
HexNAc(3)Hex(6)Fuc(1)		0.2 ± 0.4	
HexNAc(4)Hex(5)			17.0 ± 4.3*

Supplementary Table 4. Glycoform abundance at N176 in IgG2 in untreated, EndoNF only treated, and EndoNF and mild acid hydrolysis treated samples. *, $p < 0.05$ between untreated versus EndoNF and mild acid hydrolysis treated.

	Untreated	EndoNF	EndoNF and mild acid hydrolysis
Glycoforms	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**
HexNAc(3)Hex(3)		0.02 ± 0.03	
HexNAc(3)Hex(3)Fuc(1)	3.2 ± 0.4	3.0 ± 0.5	2.5 ± 1.2
HexNAc(3)Hex(4)Fuc(1)	0.3 ± 0.4	0.6 ± 0.3	0.6 ± 0.6
HexNAc(3)Hex(4)Fuc(1)NeuAc(1)		0.03 ± 0.1	
HexNAc(4)Hex(3)	0.2 ± 0.04	0.1 ± 0.1	0.5 ± 0.02*
HexNAc(4)Hex(3)Fuc(1)	18.6 ± 9.5	10.7 ± 2.4	12.3 ± 1.9
HexNAc(4)Hex(4)			0.3 ± 0.5
HexNAc(4)Hex(4)Fuc(1)	31.0 ± 5.4	35.4 ± 0.9	37.8 ± 1.2
HexNAc(4)Hex(4)Fuc(1)NeuAc(1)	6.9 ± 1.3	7.3 ± 0.9	*
HexNAc(4)Hex(5)	0.1 ± 0.1	0.2 ± 0.2	0.5 ± 0.03*
HexNAc(4)Hex(5)Fuc(1)	23.5 ± 2.5	23.8 ± 2.0	37.5 ± 1.3*
HexNAc(4)Hex(5)Fuc(1)NeuAc(1)	7.9 ± 1.5	9.1 ± 0.2	*
HexNAc(4)Hex(5)NeuAc(1)		0.02 ± 0.04	
HexNAc(5)Hex(3)Fuc(1)	4.4 ± 2.6	3.2 ± 3.3	4.5 ± 2.9
HexNAc(5)Hex(4)Fuc(1)	2.6 ± 2.3	5.0 ± 0.5	3.1 ± 2.4
HexNAc(5)Hex(4)Fuc(1)NeuAc(1)		0.04 ± 0.1	
HexNAc(5)Hex(5)Fuc(1)	1.2 ± 0.2	1.4 ± 0.04	0.5 ± 0.8
HexNAc(5)Hex(5)Fuc(1)NeuAc(1)		0.1 ± 0.1	