Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

German Edition: DOI:

Posttranslational Modifications of Intact Proteins Detected by NMR Spectroscopy: Application to Glycosylation**

Mario Schubert, Michal J. Walczak,* Markus Aebi, and Gerhard Wider*

anie_201502093_sm_miscellaneous_information.pdf

Supporting Information

Supplementary Figures

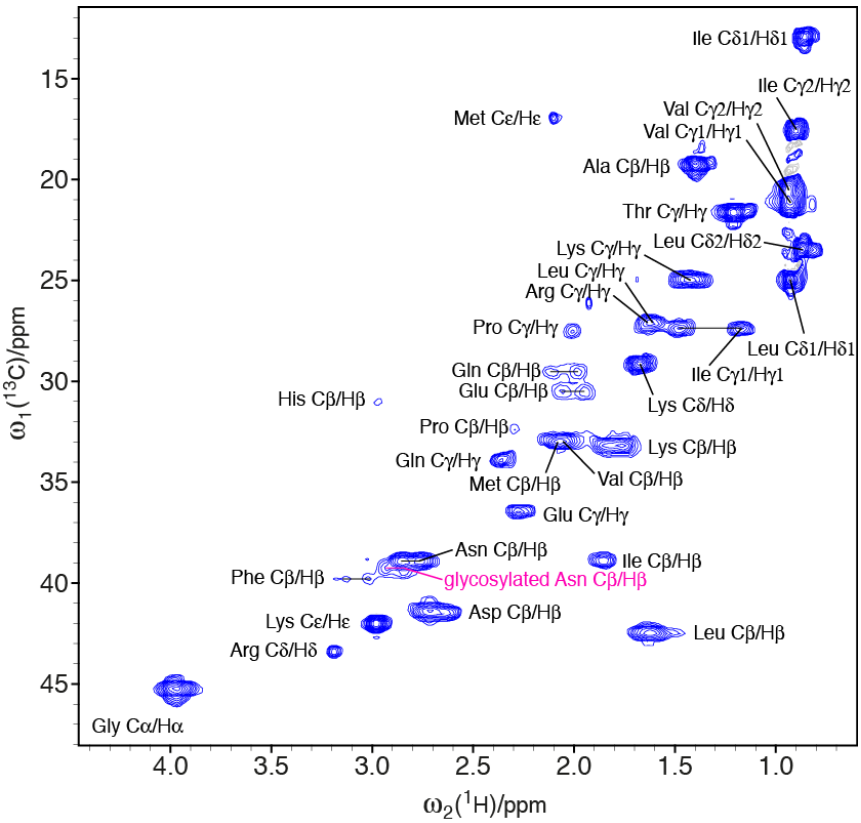

Figure S1. 500 MHz [^{1}H, ^{13}C]-HSQC spectrum of denatured, glycosylated bacterial protein AtaC measured at 298 K displays solely random coil chemical shifts for the protein resonances except for the glycosylated asparagine signal highlighted in magenta. The latter was assigned based on chemical shifts found in the spectra of small peptides with the same modification[1].

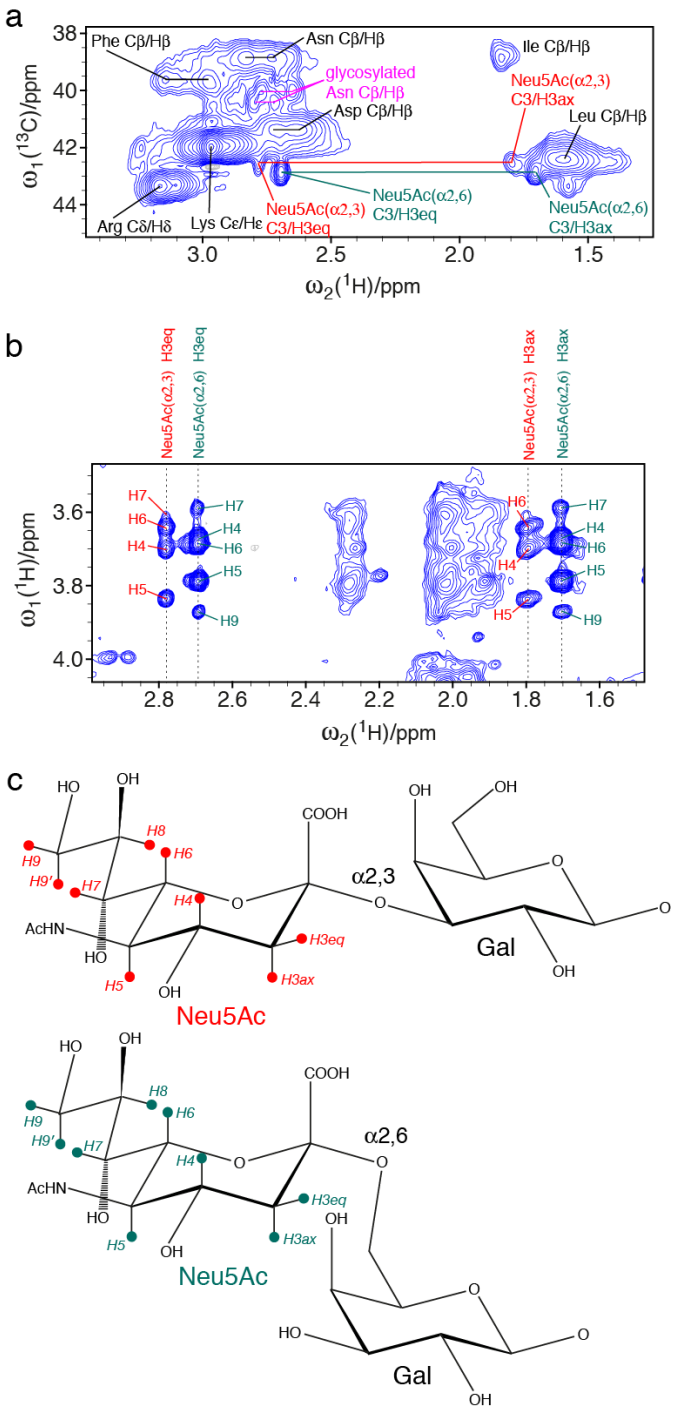

Figure S2. Spectra of denatured human albumin display signals of α2,3-linked sialic acid in addition to the expected strong signals of α2,6-linked sialic acid. a) Region of a [^{1}H, ^{13}C]-HSQC with the methylene signals of sialic acids. b) Region of a [^{1}H, ^{1}H]-TOCSY spectrum (120 ms mixing time) with the methylene signals of sialic acids. The chemical shifts of α2,6-linked sialic acid are identical to those reported for the expected complex biantennary structure with two sialic acid residues[2] whereas the resonances of α2,3-linked sialic acid correspond to values reported for amphibian glycans[3]. c) Schematic illustration of the two populations of terminal sialic acids in albumin.

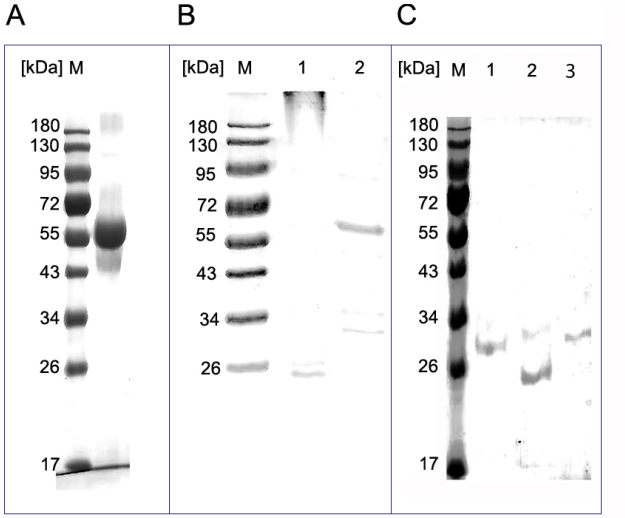

Figure S3. SDS-PAGE gels of the glycoproteins used in this study detected with Coomassie Brilliant Blue. The first lane (labelled M) in all three gels shows a molecular weight protein marker. a) Human albumin (expected molecular weight 62 kDa). b) Baker's yeast invertase; lane 1: untreated invertase with an expected molecular weight of about 270 kDa; lane 2: invertase after treatment with PNGase F confirming N-glycosylation, de-glycosylated invertase indicates an expected molecular weight of approximately 60 kDa. c) Human Tumor Necrosis Factor; lane 1: TNF digested with PNGase F; lane 2: TNF treated with O-glycosidase; lane 3: untreated TNF; digestion protocols were performed according to New England Biolabs protocols; the gel shows that the molecular weight of the O-glycan reaches approximately 7–9 kDa and N-glycan approximately 1–2 kDa; de-glycosylated TNF-EDB shows an expected molecular weight of 25–26 kDa.

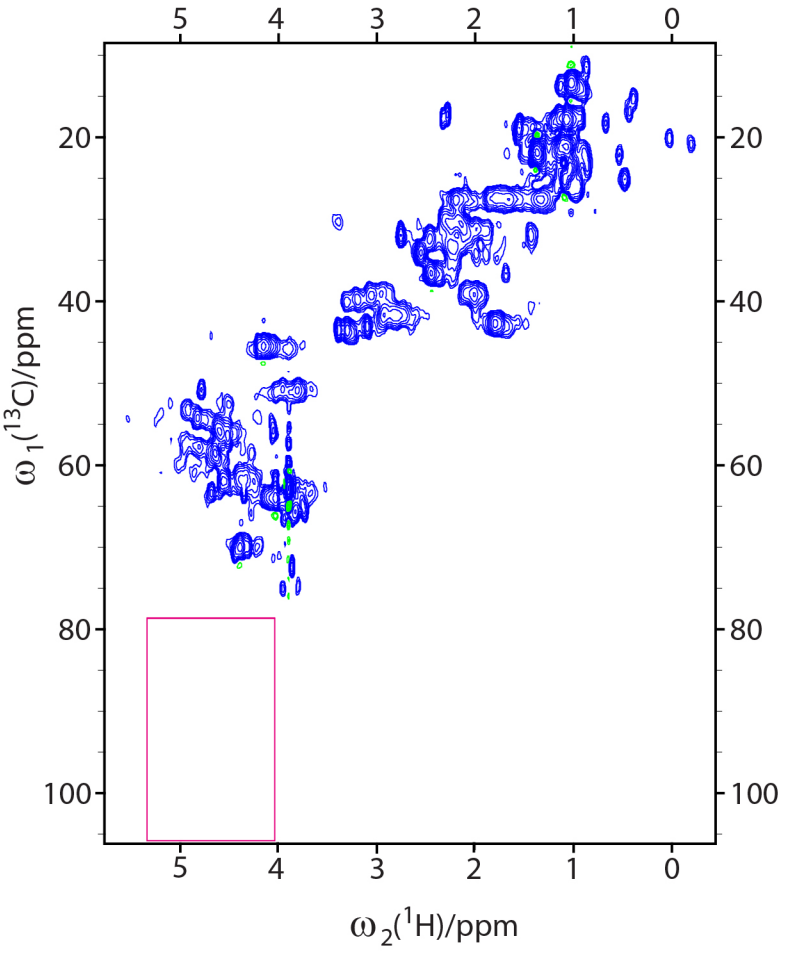

Figure S4. [^{1}H, ^{13}C]-HSQC spectrum of the free EDB domain of fibronectin expressed in HEK 293 cells. The spectrum was recorded at 600 MHZ and 313 K in a urea buffer as the other proteins. The magenta rectangle corresponds to the anomeric region of the TNF-EDB fusion protein shown in Figure 2e; there are no glycan signals in EDB.

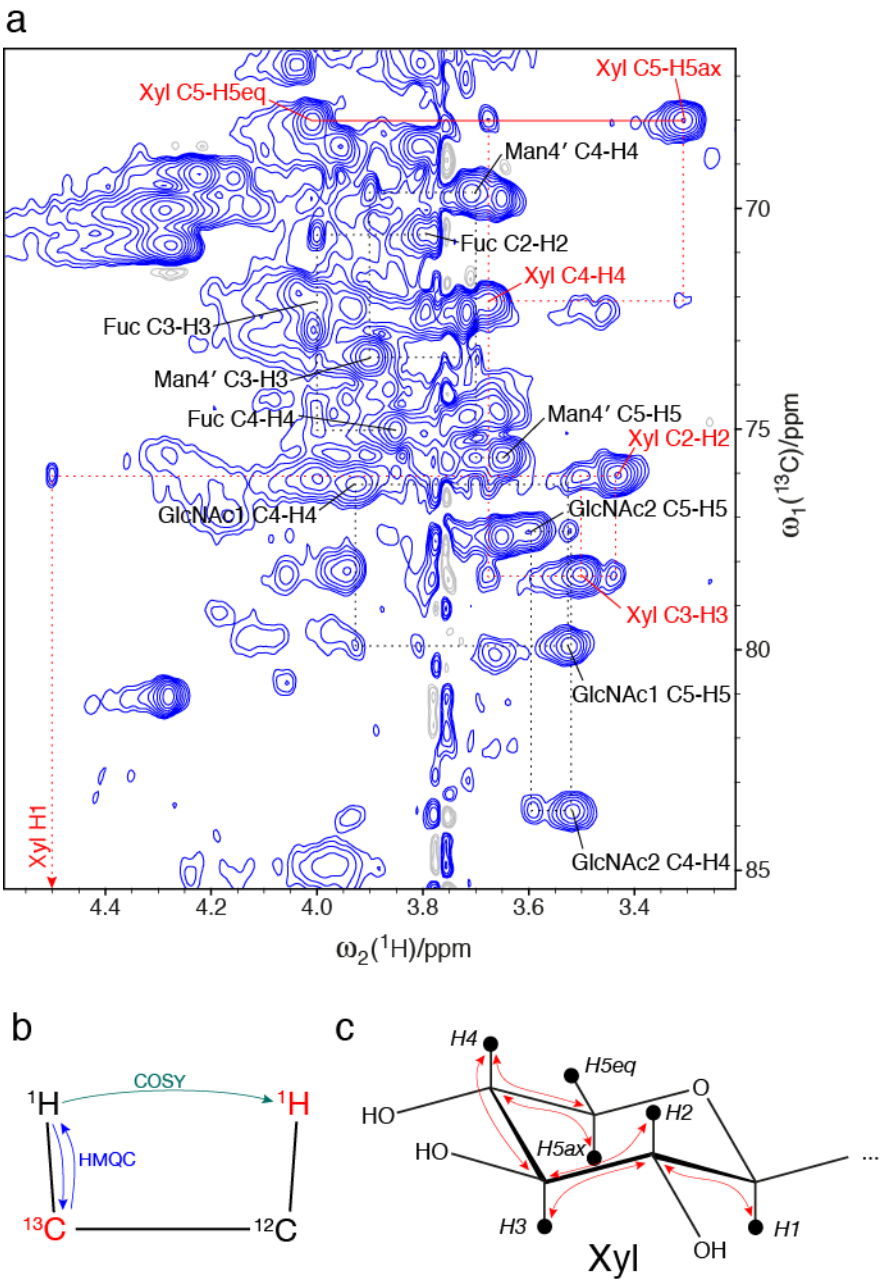

Figure S5. Glycan chemical shift assignment based on [^{1}H, ^{13}C]-HMQC-COSY correlations. a) 2D [^{1}H, ^{13}C]-HMQC-COSY spectrum of denatured bromelain recorded at 500 MHz and 298 K. In addition to resonances of ^{13}C-^{1}H moieties small correlation peaks to the next ^{1}H neighbor based on $^{3}J_{HH}$ are observed, resulting in (H)C-(C)H correlations. The dotted lines connect the latter correlations to the ^{13}C-^{1}H resonances (labelled with their assignment) allowing their assignment. The assignment path for Xylose is highlighted in red (see text). b) Schematic illustration of correlations in an HMQC-COSY experiment where also ^{1}H-^{13}C cross-peaks between the nuclei marked in red occur. c) Observed C-H two-bond correlations of xylose in its pyranose form (red arrows). Together with single-bond ^{1}H-^{13}C correlations an assignment walk through the entire spin system is easily possible (see text).

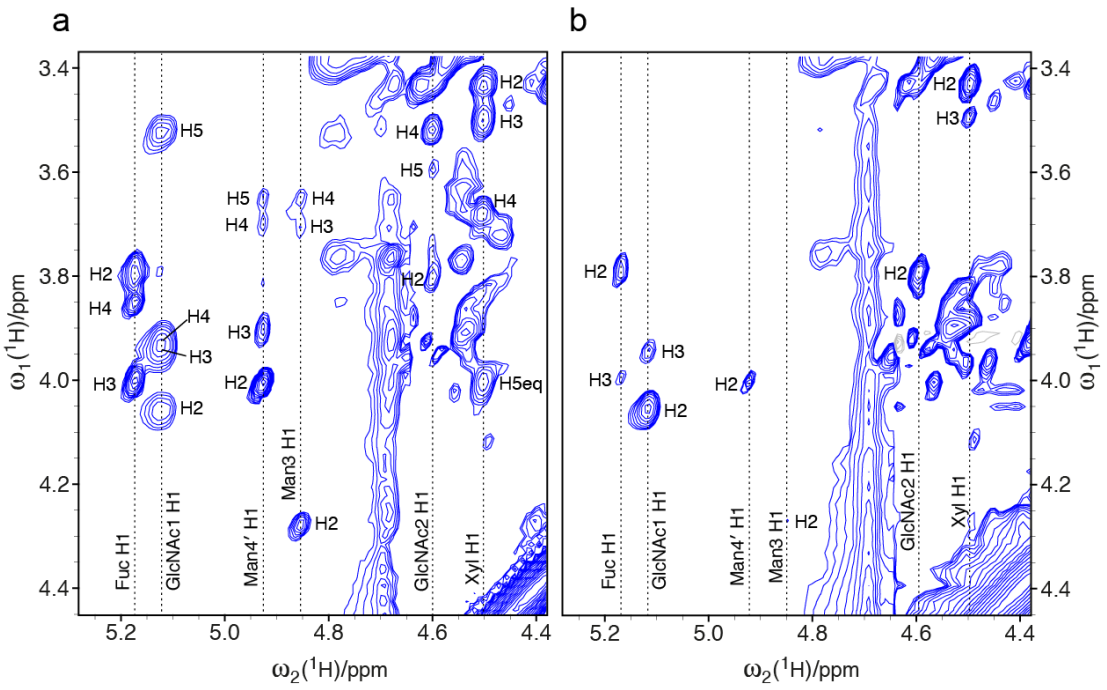

Figure S6. 2D [^{1}H, ^{1}H] TOCSY spectra of bromelain recorded at 500 MHz and 298 K using mixing times of 120 ms (a) or 13 ms (b).

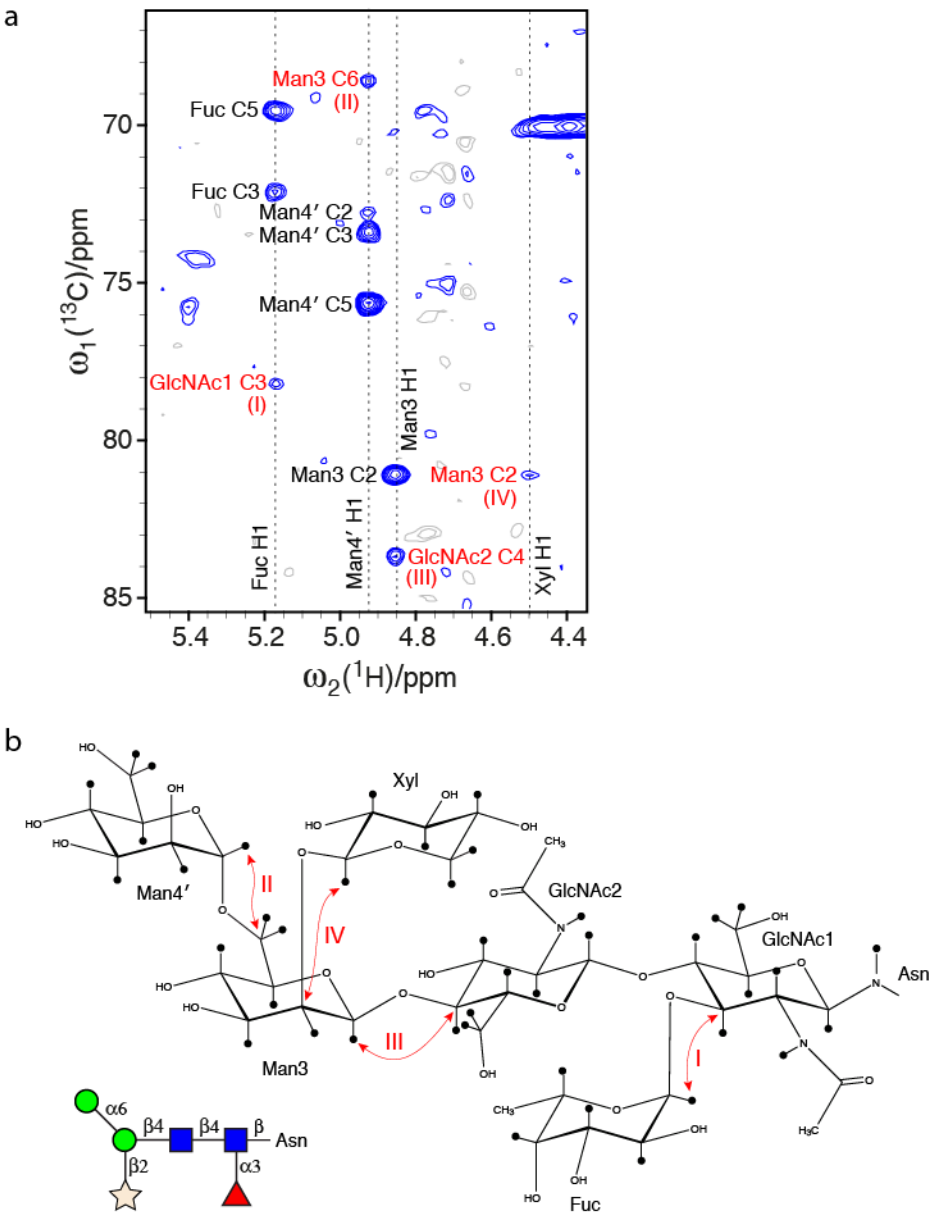

Figure S7. Establishing the glycan identity by connecting the saccharides using correlations across the glycosidic linkages. a) 2D [^{1}H, ^{13}C]-HMBC spectrum of bromelain recorded at 500 MHz and 298 K. A standard [^{1}H, ^{13}C]-HSQC pulse sequence with an INEPT period of 26 ms was used. In addition to intra-saccharide long range $^{n}J_{CH}$ correlations valuable correlations across glycosidic linkages (red labels with roman numbers (see b)) are observed. b) Illustration of the trans-glycosidic linkage correlations that establish the connections in bromelain; the roman numbers identify the connectivities indicated by the cross peaks with red labels in a).

Supplementary Tables

Table S1. Summary of the NMR experiments recorded for the assignments of glycans of glycoproteins[1]

Protein (MW)	conc. [mM]	Experiment	NS	TD	Spectrometer Frequency [MHz]	Remarks
AtaC (58.2 kDa)	0.5	[^{1}H, ^{13}C] - HSQC	196	172	500	Δ = 1.7ms
		[^{1}H, ^{13}C] - HMBC	128	312	500	Δ = 13ms
		[^{1}H, ^{13}C] - HMQC-COSY	96	310	500	Δ = 13ms
		[^{1}H, ^{1}H] - TOCSY	4	424	500	τ_m = 13ms
		[^{1}H, ^{1}H] - TOCSY	4	480	500	τ_m =120ms
Albumin (66.5 kDa)	2.5	[^{1}H, ^{13}C] - HSQC	112	840	900	Δ =1.7ms
		[^{1}H, ^{13}C] - HMBC	124	700	600	Δ = 13ms
		[^{1}H, ^{13}C] - HMQC-COSY	184	700	600	Δ = 13ms
		[^{1}H, ^{1}H] - TOCSY	16	1536	900	τ_m = 13ms
		[^{1}H, ^{1}H] - TOCSY	16	1536	900	τ_m =120ms
Bromelain (23 kDa)	3.5	[^{1}H, ^{13}C] - HSQC	96	700	500	Δ = 1.7ms
		[^{1}H, ^{13}C] - HMBC	104	700	500	Δ = 13ms
		[^{1}H, ^{13}C] - HMQC-COSY	240	700	500	Δ = 13ms
		[^{1}H, ^{1}H] - TOCSY	28	1024	500	τ_m = 13ms
		[^{1}H, ^{1}H] - TOCSY	28	1024	500	τ_m =120ms
Invertase (270 kDa)	0.9	[^{1}H, ^{13}C] - HSQC	88	980	700	Δ = 1.7ms
		[^{1}H, ^{13}C] - HMBC	96	980	700	Δ = 13ms
		[^{1}H, ^{13}C] - HMQC-COSY	172	980	600	Δ = 13ms
		[^{1}H, ^{1}H] - TOCSY	24	896	700	τ_m = 13ms
		[^{1}H, ^{1}H] - TOCSY	24	896	700	τ_m =120ms
TNF-EDB (27 kDa)	0.6	[^{1}H, ^{13}C] - HSQC	96	512	900	Δ = 1.7ms
		[^{1}H, ^{13}C] - HMBC	144	512	900	Δ = 13ms
		[^{1}H, ^{13}C] - HMQC-COSY	272	512	900	Δ = 13ms
		[^{1}H, ^{1}H] - TOCSY	24	512	900	τ_m = 13ms
		[^{1}H, ^{1}H] - TOCSY	24	512	900	τ_m =120ms

[1] First column: name and molecular weight (MW) of protein; second column: molar concentration; third column: type of the NMR experiment, fourth column: number of transients accumulated for a given experiment (NS); fifth column: number of time domain data points in the indirect dimension (TD); sixth column: spectrometer frequency, and seventh column: remarks on the experimental setup; τ_m is the TOCSY mixing time and Δ is half the duration of the heteronuclear transfer. All spectra were recorded with spectral widths of 140 (offset 65 ppm) and 12 ppm (offset 4.7 ppm) in the carbon and proton dimension, respectively; temperature T = 298 K except for TNF (T = 308K).

Table S2. Glycan $^{13}C/^{1}H$ chemical shifts of $AtaC^{1866-2428}$ (referenced to DSS).

Residue with residue number	C1/ H1	C2/ H2	C3/ H3	C4/ H4	C5/ H5	C6/ H6;H6′
Glc 1	82.1/ 4.96	74.6/ 3.41	79.3/ 3.55	72.0/ 3.45	80.4/ 3.47	63.3/ 3.86;3.74

Table S3. Glycan $^{13}C/^{1}H$ chemical shifts of bromelain (referenced to DSS).

Residue with residue number	C1/ H1	C2/ H2	C3/ H3	C4/ H4	C5/ H5ax;H5eq	C6/ H6;H6′
GlcNAc 1	80.9/ 5.12	57.1/ 4.06	78.2/ 3.94	76.2/ 3.92	79.9/ 3.52	
GlcNAc 2	103.2/ 4.60	57.8/ 3.80	75.1/ 3.79	83.6/ 3.52	77.3/ 3.59	
Fuc 3′	101.4/ 5.17	70.6/ 3.80	72.1/ 4.00	75.0/ 3.85	69.6/ 4.78	18.4/ 1.32
Man 3	103.5/ 4.86	81.1/ 4.28	74.7/ 3.71	69.8/ 3.65	77.5/ 3.65	68.6/ 3.94;3.80
Man 4′	102.6/ 4.93	72.8/ 4.01	73.4/ 3.90	69.6/ 3.70	75.6/ 3.65	3.81
Xyl	107.2/ 4.50	76.0/ 3.43	78.3/ 3.50	72.1/ 3.68	68.0/ 3.31;4.01	

Table S4. Glycan $^{13}C/^{1}H$ chemical shifts of albumin (referenced to DSS).

Residue with residue number	C1/ H1	C2/ H2	C3/ H3ax;H3eq	C4/ H4	C5/ H5	C6/ H6;H6′	C7/ H7	C8/ H8	C9/ H9;H9′
GlcNAc 1	81.1/ 5.04	56.4/ 3.86	3.71	83.4/ 3.63	78.9/ 3.50				
Man 3	103.2/ 4.75	73.0/ 4.26	83.5/ 3.78	68.3/ 3.81		68.8/ 3.64;3.64			
Man 4	102.4/ 5.15	79.3/ 4.20	3.76	70.1/ 3.51					
Man 4′	99.7/ 4.92	79.4/ 4.09	72.2/ 3.89	70.1/ 3.51					
GlcNAc 5/5′ [a]	102.3/ 4.57	57.4/ 3.76	74.9/ 3.76	83.3/ 3.64	77.3/ 3.59	63.0/ 3.97;3.84			
Gal 6/6′ [a]	106.4/ 4.43	73.5/ 3.56	75.3/ 3.66	71.3/ 3.91	76.5/ 3.80	66.2/ 3.98;3.59			
Neu5Ac 7/7′ [a]			42.9/ 1.71;2.69	71.0/ 3.68	54.7/ 3.79	75.4/ 3.69	71.1/ 3.59	74.5/ 3.87	65.3/ 3.87;3.67
Gal# 6/6′ [a,b]	105.6/ 4.52	72.1/ 3.59	78.3/ 4.08	70.3/ 3.96	3.78	66.7/ 3.99;3.75			
Neu5Ac# 7/7′ [a,b]			42.5/ 1.80;2.78	71.1/ 3.70	54.6/ 3.84	75.6/ 3.64	70.8/ 3.61		

[a] the chemical shifts of the two branches (Neu5Acα2,6Galβ1,4GlcNAcβ1,2) of the complex glycan are identical

[b] second population containing Neu5Acα2,3Gal

Table S5. Glycan $^{13}C/^{1}H$ chemical shifts of invertase with spin systems ordered according their anomeric ^{1}H chemical shift (referenced to DSS).

Residue [a]	Residue in environment	C1/ H1	C2/ H2	C3/ H3	C4/ H4	C5/ H5	C6/ H6;H6′
Man(a)	Manβ1,2ManαP	98.9/ 5.46	4.16	80.6/ 4.01	3.89[b]		
Man(a′)	Manα1,2ManαP	99.0/ 5.48	4.03	3.94[c]	3.76	3.84	3.92;3.76
Man(b,c)	Manα1,2Manα1,2Manα1,2... and Manα1,2Manα1,2Manα1,6...	103.3/ 5.30	81.2/ 4.13	3.94	3.75[b]		
Man(d)	Manβ1,2Manα1,2Man ...	102.4/ 5.19					
Man(e)	Manα1,3Manα1,2 ...	105.0/ 5.18	72.9/ 4.10	73.2/ 3.92	3.69	76.2/ 3.81	3.78
Man(f)	Manα1,6[...Manα1,2]Manα1,6Manα ...	101.0/ 5.12	81.4/ 4.04				
Man(g)	Manα1,2Manα1,2 ...	104.9/ 5.08	4.10	3.88	3.68	3.78	
Man(h)	Manα1,3Manα1,2Manα1,2	104.9/ 5.07	4.23	3.96		3.84[b]	
Man(h)	Manα1,2Manα1,6 ...	104.9/ 5.07	4.10				
GlcNAc1	GlcNAcβ1,4GlcNAcβAsn	81.1/ 5.05					
Man(m)	?	4.94	4.33	3.71			
Man(n)	Manα1,6Manα1,6 ... and Manα1,6Manα1,6Manα1,6 ...	102.1/ 4.92	4.02	3.86			

[a] nomenclature according to Jawhara et al.
[b] position in the spin system not clear
[c] overlapped resonance

Supplementary Methods

Sample preparation. The preparation of *in vitro* glycosylated bacterial glycoprotein $AtaC^{1866-2428}$ was reported earlier [4]. The other three proteins were purchased as powder from Sigma-Aldrich: pineapple stem bromelain 65% purity (cat. no. B4882), human serum albumin 96% purity (cat. no. A1653) and *S. cerevisiae* invertase 96% (cat. no. I4504). Bromelain was dissolved in 50 mM Tris-HCl, pH 7.0 and purified with cation exchange chromatography as reported previously [5] and the other two proteins were used without further purification. All proteins were dissolved in D_2O with 50 mM Tris-d_{11} (Sigma-Aldrich, cat. no. 486248) and lyophilized. The proteins were re-suspended and twice lyophilized. For the NMR measurements all lyophilized proteins were dissolved in 7 M Urea-d_4 (98% D, Armar Chemicals, cat. no. 049500.3041), 2 mM DTT (DTT-D_{10} Cambridge Isotope, cat. no. DLM-2622), pD 5.5 (adjusted with acetic acid-D_4, Sigma-Aldrich cat. no. 233315) with a concentration as indicated in Supplementary Table 1.

TNF sample was a generous gift of Prof. Dario Neri (ETH Zurich) and it was expressed in fusion with the EDB domain of fibronectin (EDB-TNF)[6]. The lack of posttranslational modifications of the EDB domain was confirmed by [^{1}H, ^{13}C]-HSQC spectrum, Supplementary Figure 4.

NMR spectroscopy. All spectra were recorded on Burker Avance III or Avance HD spectrometers equipped with triple-resonance cryogenic probes at 298 K (except for TNF, T = 308 K) and ^{1}H frequencies of 500, 600, 700 or 900 MHz. Sample volumes of 500 µl in standard 5 mm NMR tubes were used, except for TNF for which 250 µl in a 5 mm Shigemi® tube were measured. Standard 2D [^{1}H, ^{13}C]-HSQC, [^{1}H, ^{13}C]-HMBC and [^{1}H, ^{1}H] - TOCSY experiments, were recorded together with a 2D [^{1}H, ^{13}C]-HMQC-COSY. For details see Supplementary Table 1. All spectra were processed with Topspin 3.2 (Bruker Biospin), referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) and further analyzed in

Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, USA). ^{1}H and ^{13}C chemical shifts are found in Supplementary Tables 2-5.

Chemical shift assignments. Resonance assignments and the identification of saccharides were achieved by several parallel strategies that are mostly well described[7]. The simplest approach was based on the rarely used 2D ^{13}C-^{1}H HMQC-COSY experiment[8] that results in addition to directly linked ^{13}C-^{1}H correlations to relay correlations based on $^{3}J_{HH}$ scalar couplings (Supplementary Fig. 5). Considering a ^{13}C-^{1}H group, the HMQC-COSY correlates the ^{1}H with the ^{13}C of the neighboring CH-moiety and the ^{13}C with the ^{1}H of the neighboring CH-moiety. This is illustrated for the xylose spin system of bromelain in Supplementary Figure 5, with the anomeric proton H1 showing a cross peak with the neighboring carbon C2 also identifying H2. H2 correlates to C3 and H3 to C2 and to C4 (missing in Fig. S2a). With both correlations (e.g. H2-C3 and C3-H2) present, identification of the neighboring CH-moiety was straightforward (dotted lines in Fig. S2). However, the $^{3}J_{HH}$ scalar couplings were not always large enough to result in detectable signals and a reliable determination of the spin systems required additional data. To this end 2D [^{1}H, ^{1}H]-TOCSY spectra were used (Fig. 2 and Supplementary Fig. 6). Most, sometimes all, ^{1}H resonances of a spin system showed correlations to the anomeric H1 resonance in a TOCSY spectrum with a long mixing time (120 ms), whereas only one strong correlation to the neighboring H2 and sometimes a weaker correlation to H3 was observed with a very short mixing time (Supplementary Fig. 6b). An alternative or supplemental approach to complete the assignment of a spin system involved a HMBC experiment that used long-range $^{n}J_{CH}$ scalar couplings. Typically H1 resonances showed correlations to C3 and C5 (Supplementary Fig. 7). The latter correlations were very useful as $^{3}J_{H4H5}$ couplings were often too weak to establish connections between H5 to the rest of the spin system in a ^{1}H-^{1}H TOCSY spectrum. With this procedure, the resonances of the entire spin system could often be assigned to the position in the spin system, whereas in other cases the position of some resonances remained unclear.

The chemical shifts of the entire spin systems were clearly different from each other, depending on the saccharide type, configuration and linkages. A search in chemical shift databases like www.glycosciences.de[9] revealed the saccharide type and the linkages fitting to the observed spin system. Typically the combination of chemical shifts was unique and resulted in a single assignment solution. For some spin systems, unique characteristic spectroscopic features were diagnostic of the type of saccharide, e.g. a methyl group at C6 indicated L-fucose, which is the only methyl-containing saccharide among the most abundant hexoses. A methylene group at C5 was a signature of D-xylose. If required, the distinction between α and β configuration can be obtained from characteristic chemical shift regions of H1 (axial, typically α: 4.8-5.8 ppm and equatorial, typically β: 4.3-4.8 ppm) and more importantly the size of the scalar coupling constant $^{3}J_{H1H2}$ (7-8 Hz if both H1 and H2 are axial, approx. 4 Hz for equatorial-axial, approx. 2 Hz for equatorial-equatorial or axial-equatorial)[7].

An alternative approach for assigning glycan resonances is suggesting a variety of glycan structures typical for certain organisms, cell types or similar proteins and to either extract the chemical shifts from databases or from literature reports, or to predict them using powerful prediction algorithms like CASPER[10]. It is straightforward to match the chemical shifts of all suggested glycans to the observed ones. In the case of the N-glycan hexasaccharide of bromelain, 2 spin systems could be completely assigned and matched to terminal α-L-fucose and β-D-Xylose and nearly complete assignments were obtained and matched to β-D-GlcNAc1, β-D-Man3 and β-D-Man4′. The [^{1}H, ^{13}C]-HMBC established the connections Xylα1,2Man, Fucα1,3GlcNAc1, Man4′α1,6Man3 and Man3β1,4GlcNAc2 (Supplementary Fig. 7). Since eukaryotic N-glycans always have a GlcNAcβ1,4GlcNAc core, the remaining spin system was likely GlcNAc2 and indeed the observed chemical shifts matched to values for eukaryotic N-glycans. For bromelain the entire N-glycan had already been analyzed by NMR as part of a glycopeptide obtained from bromelain by digestion[5]. All observed chemical shifts matched the literature values. However, we observed some smaller

additional signals that could originate either from another glycosylation site (e.g. O-glycan) or of a minor population of a further extended or modified N-glycan (Fig. 2).

Supplementary References:

[1] F. Schwarz, Y.-Y. Fan, M. Schubert, M. Aebi, *J. Biol. Chem.* **2011**, *286*, 35267-35274.

[2] J.-M. Wieruszeski, J.-C. Michalski, J. Montreuil, G. Strecker, *Glycoconjugate J.* **1989**, *6*, 183-194.

[3] D. Machytka, R. A. Klein, H. Egge, *Carbohyd. Res.* **1994**, *254*, 289-294.

[4] A. Naegeli, C. Neupert, Y.-Y. Fan, C.-W. Lin, K. Poljak, A. M. Papini, F. Schwarz, M. Aebi, *J. Biol. Chem.* **2014**, *289*, 2170-2179.

[5] J. B. Bouwstra, E. C. Spoelstra, P. Dewaard, B. R. Leeflang, J. P. Kamerling, J. F. G. Vliegenthart, *Eur. J. Biochem.* **1990**, *190*, 113-122.

[6] F. Buller, Y. Zhang, J. Scheuermann, J. Schaefer, P. Buehlmann, D. Neri, *Chem. Biol.* **2009**, *16*, 1075-1086.

[7] J. Ø. Duus, C. H. Gotfredsen, K. Bock, *Chem. Rev.* **2000**, *100*, 4589-4614.

[8] S. W. Fesik, R. T. Gampe Jr, E. R. Zuiderweg, *J. Am. Chem. Soc.* **1989**, *111*, 770-772.

[9] T. Lütteke, A. Bohne-Lang, A. Loss, T. Goetz, M. Frank, C.-W. von der Lieth, *Glycobiology* **2006**, *16*, 71R-81R.

[10] M. Lundborg, G. r. Widmalm, *Anal. Chem.* **2011**, *83*, 1514-1517.