NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation

Supporting Information

Vadim Slynko, Mario Schubert, Shin Numao, Michael Kowarik, Markus Aebi, Frédéric H.-T. Allain

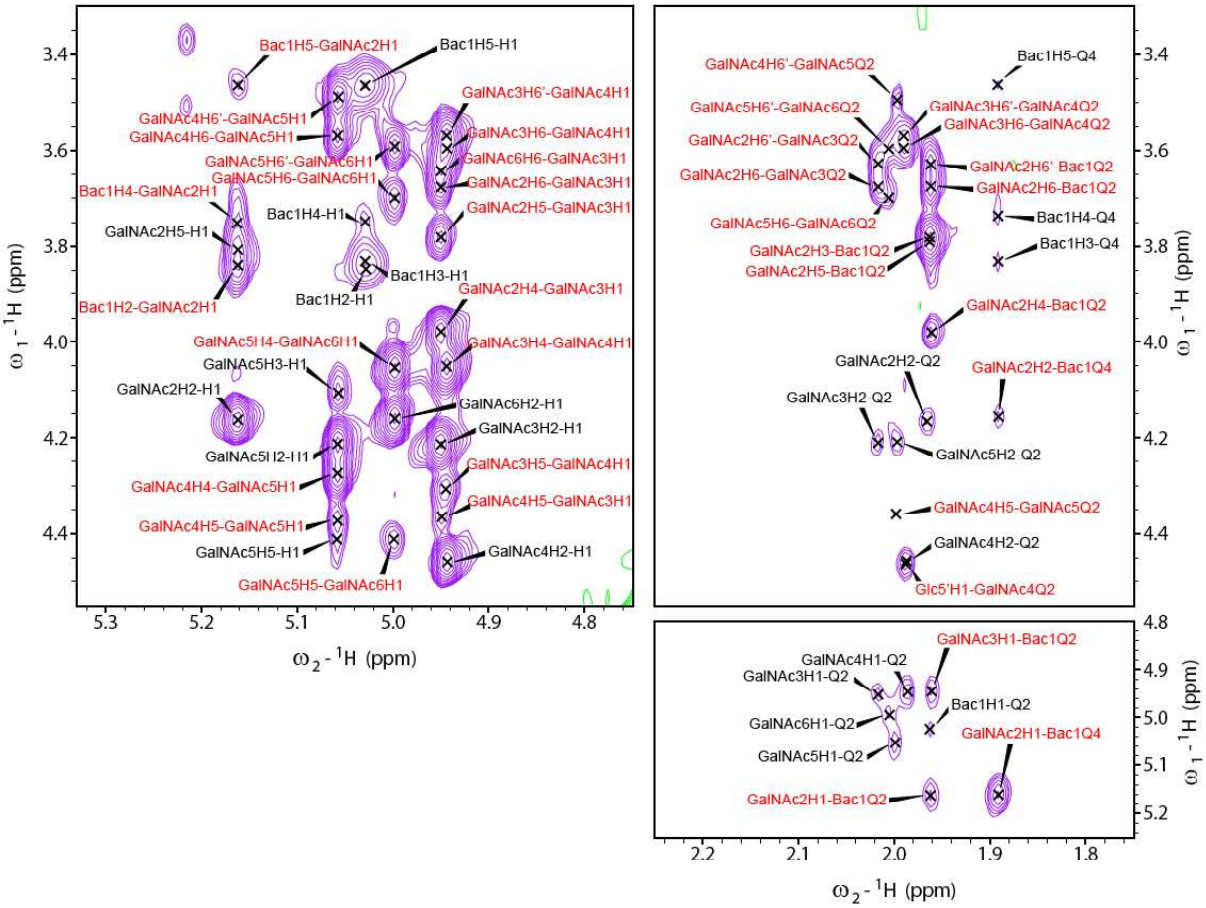

Figure S1: Detailed views of the ^{13}C-filtered-filtered 2D NOESY spectrum of glycosylated AcrA$^{61-210\Delta\Delta}$ (^{13}C/^{15}N labeled protein and unlabeled glycan). Sequential NOEs are labeled in red and intra-residue NOEs in black. The chemical shifts of the acetamido methyl groups are labeled as Q2 and Q4.

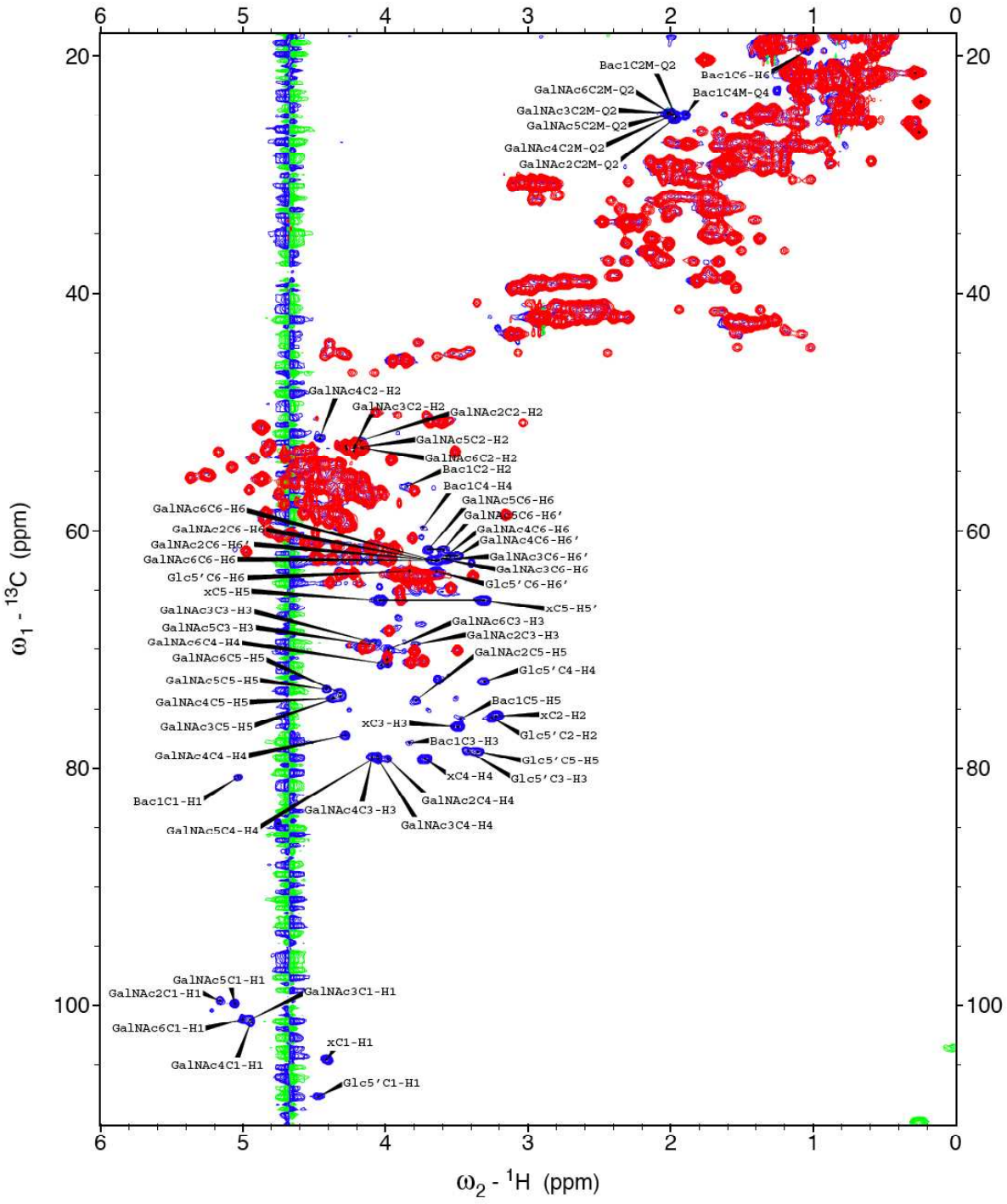

Figure S2: Natural abundance ^{13}C-HSQC spectrum of the unlabeled glycoprotein (blue/green) overlaid by a ^{13}C-HSQC spectrum of the ^{13}C/^{15}N protein-labeled glycoprotein (red). Experimental details are given in **Table S1**. Xylose signals are labeled with x.

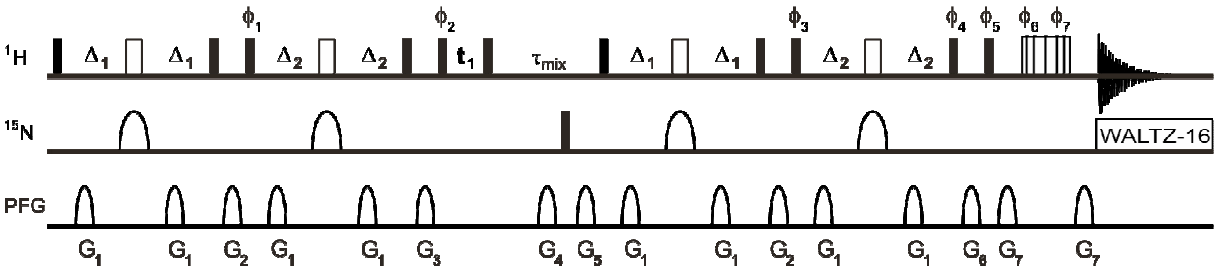

Figure S3: Pulse sequence of the 2D ^{15}N filtered-filtered NOESY. Narrow filled and wide unfilled rectangles correspond to 90° and 180° rectangular pulses, respectively. Magnetic field gradients are represented by narrow sine shapes. The wide sine shapes on the ^{15}N channel stand for a band-selective 180° REBURP pulse with a pulse length of 1250 μs [1]. The lengths of the pulses are given for a 900 MHz spectrometer. Unless indicated otherwise, pulses are applied with phase x. Proton hard pulses and the hard ^{15}N pulse were applied with 21.9 kHz and 5.1 kHz field strength, respectively. Water suppression was obtained using WATERGATE implemented with a 3-9-19 pulse [2]. The carrier frequencies were centered at ^{1}H = 4.703 ppm, ^{15}N = 118 ppm. The following delays were used: τ_{mix} = 150 ms, $\Delta_1 = \Delta_2$ = 2.71 ms corresponding to $1/2J_{NH}$ with J_{NH} =92 Hz. To achieve quadrature detection in the indirect dimension States-TPPI phase cycling [3] was applied to phase ϕ2. The phase cycling was: ϕ_1 = 4 (x), 4 (−x) ; ϕ_2 = 8 (x), 8 (−x) ϕ_3 = 16 (x), 16 (−x); ϕ_4 = x, y, −x, -y; ϕ_5 = y, −x; ϕ_6 = -y, x; ϕ_{rec} = x, y, −x, -y, 2 (−x, -y, x, y), x, y, −x, -y, −x, -y, x, y, 2 (x, y, −x, -y), −x, -y, x, y . The gradients were applied as a sinusoidal function from 0 to π and had the following duration and strength: G_1 = 400 μs (16 G/cm in x), G_2 = 900 μs (16.5 G/cm in z), G_3 = 1.1 ms (18.7 G/cm in z), G_4 = 3 ms (16.5 G/cm in z), G_5 = 300 μs (18.7 G/cm in z), G_6 = 800 μs (19.3 G/cm in z).

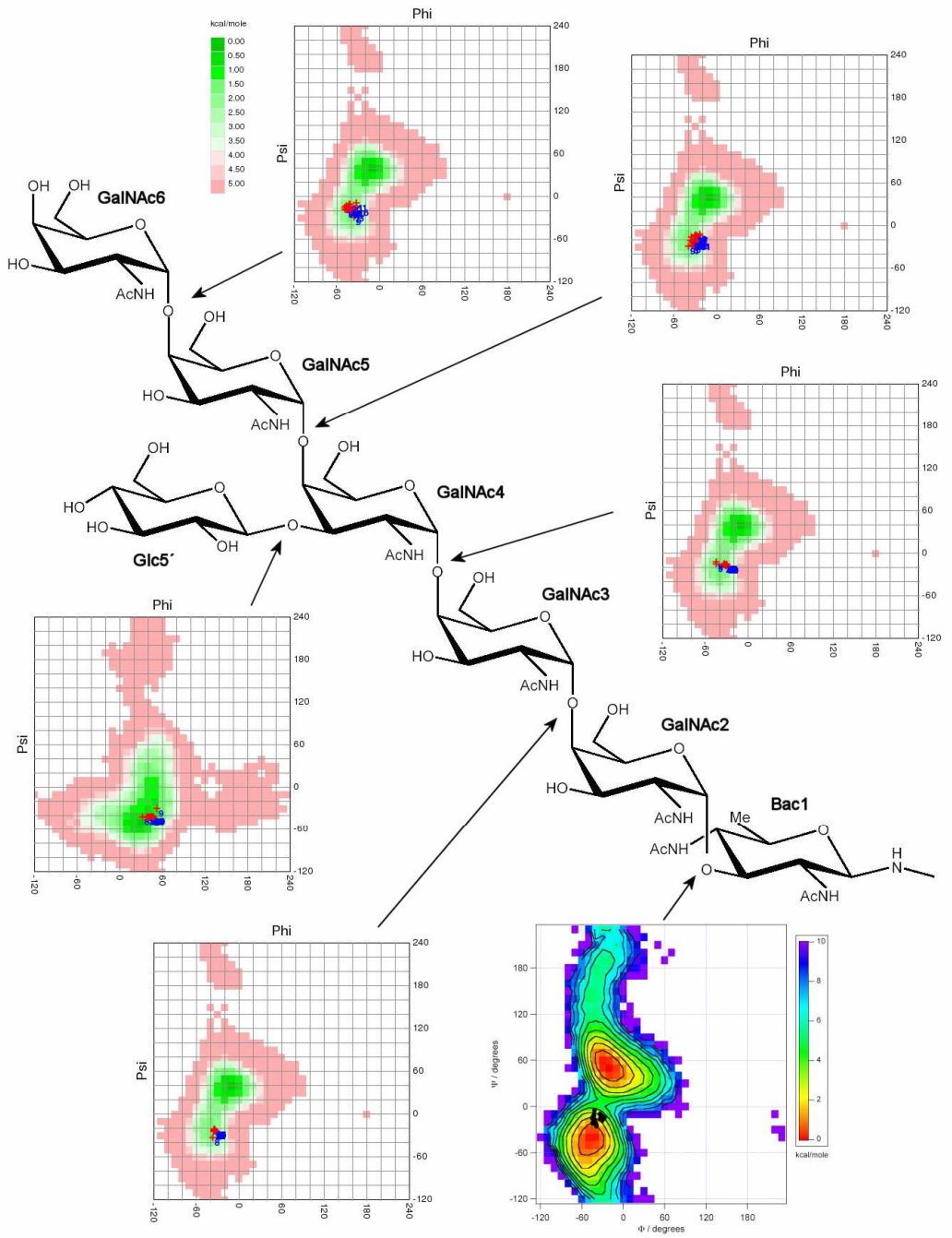

Figure S4: Plot of the glycosidic torsion angles φ and ψ of all glycosidic linkages of the 20 best structures together with the energy landscape. φ and ψ are defined as H1-C1-O1-Cx´ and C1-O1-Cx´-Hx´, respectively. The plots were generated by the software CARP [4] that analyzes coordinates and has the option to use energy landscapes from the GlycoMapsDB database [5] as background. Since CARP reads only one structure, the glycans of the ensemble were spatially separated and saved as one structure. The energy landscape of the GalNAc2-α1,3-Bac1 linkage was kindly generated and provided by Martin Frank.

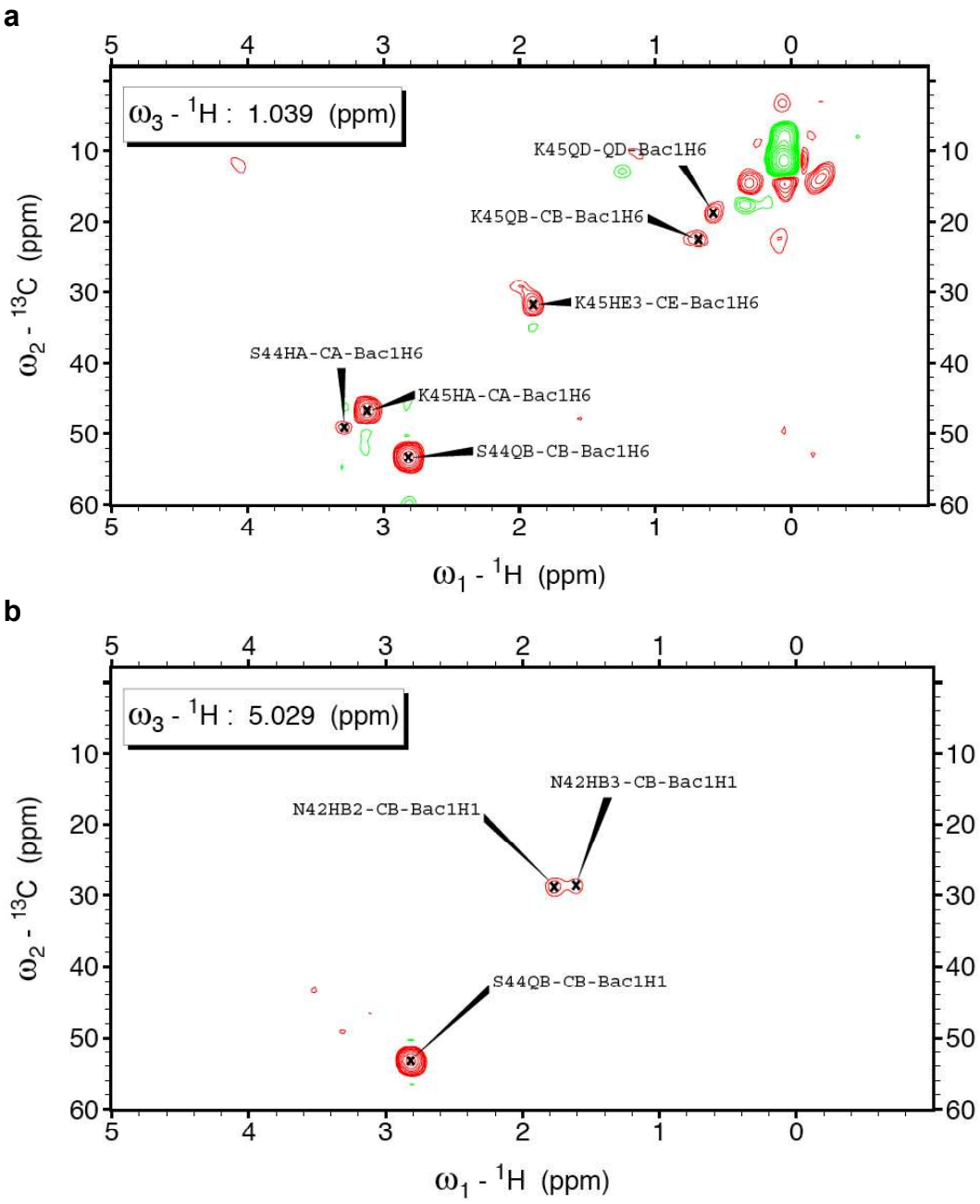

Figure S5: NOE signals between the carbohydrate and the protein. Planes of the 3D ^{13}C-edited-filtered NOESY spectrum recorded at 303 K on a 900 MHz spectrometer with a mixing time of 150 ms showing NOEs to (**a**) Bac1 H6 and (**b**) Bac1 H1. Experimental details are given in **Table S1**.

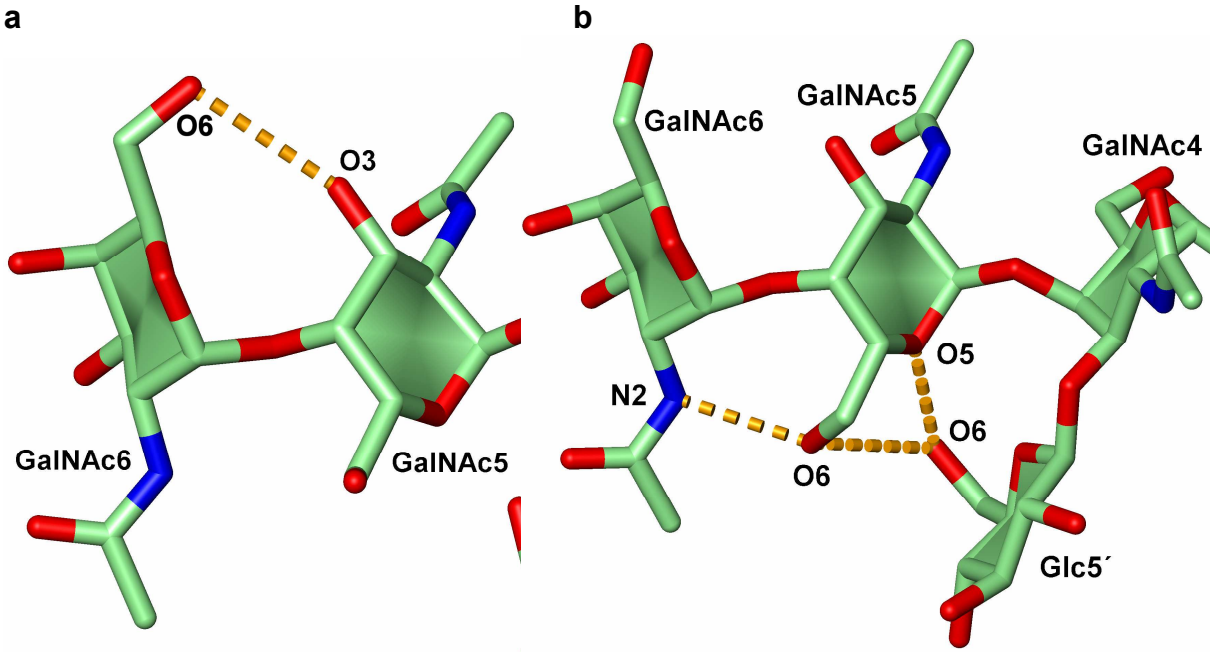

Figure S6: The *C. jejuni* glycan is likely be stabilized by hydrogen bonds. Potential hydrogen bonds observed in the NMR ensemble of 20 structures. Only in few structures oxygens and nitrogens are in hydrogen bond distance because the exo-cyclic hydroxymethyl groups are populating different rotameric states (ω torsion angle defined as O5-C5-C6-O6). The NOE data do not allow to quantify the populations of the three possible rotamers. Due to inter-residue NOE upper distance restraints to both H6 and H6' the gt rotamer (ω= 60) might be overrepresented in the ensemble. However, the gt rotamer is most populated in gluco and galacto pyranoside monosaccharides [6] and the three rotamers are likely interconverting between each other within the NMR time scale. **(a)** potential hydrogen bond stabilizing the conformation of a α-GalNAc-(1-4)-GalNAc linkage. The gt rotamer of GalNAc6 enables a potential hydrogen bond between O6 of GalNAc6 and O3 of GalNAc5. **(b)** potential hydrogen bonds stabilizing the orientation of the branched Glc5' for which both the Glc5' and the GalNac5 need to adopt the less common gg rotamers (ω= –60). In addition a potential hydrogen bond between N2 of GalNAc6 and O6 of GalNAc5 can form if GalNAc5 adopts the less common gg rotamer (ω= –60) as shown.

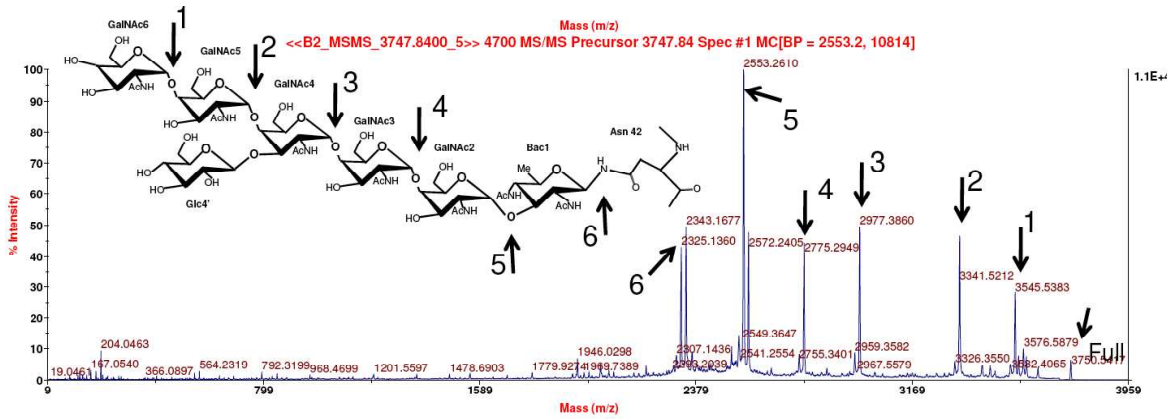

Figure S7: Glycosylated AcrA$^{61\text{-}210\Delta\Delta}$ was analyzed by MALDI-MS/MS. The fragmentation spectrum of the ion at m/z= 3747.84 corresponds to the AspN-glycopeptide of AcrA$^{61\text{-}210\Delta\Delta}$ bearing the *C. jejuni* N-glycan. The inset illustrates the *C. jejuni* N-glycan attached to the expected peptide. As indicated with the arrows, sequential loss of HexNAc (203 Da), Hex (162 Da) and Bac (228 Da) residues confirms the anticipated structure.

Table S1: NMR acquisition parameters for glycosylated AcrA$^{61\text{-}210\Delta\Delta}$.

Experiment solvent and sample	field (MHz)	nucleus	acq. pts (complex)	spectral width (ppm)	carrier freq. (ppm)	number of scans	mixing time (ms)
^{1}H-^{15}N HSQC	900	t_1=^{15}N	128	24.0	119.0	4	
(H_2O; ^{15}N)		t_2=^{1}H	1024	10.0	4.70		
^{1}H-^{15}N HSQC	500	t_1=^{15}N	35	49.3	117.0	4	
(D_2O; ^{13}C/^{15}N)		t_2=^{1}H	512	12.0	4.70		
^{1}H-^{13}C HSQC	900	t_1=^{13}C	128	70.0	40.0	4	
(aliph., D_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	13.9	4.70		
^{1}H-^{13}C HSQC	600	t_1=^{13}C	128	60.0	120.0	4	
(arom., D_2O; ^{13}C/^{15}N)		t_2=^{1}H	512	13.9	4.70		
^{1}H-^{13}C HSQC	500	t_1=^{13}C	240	99.9	65.0	288	
(nat. abund., D_2O; ^{15}N)		t_2=^{1}H	512	12.0	4.70		
^{15}N NOESY-HSQC	900	t_1=^{1}H	115	11.0	4.70	8	120
(H_2O; ^{15}N)		t_2=^{15}N	47	25.0	119.2		
		t_3=^{1}H	512	11.0	4.70		
^{13}C NOESY-HSQC	900	t_2=^{1}H	113	11.1	4.70	8	120
(H_2O; ^{13}C/^{15}N)		t_1=^{13}C	50	35.1	38.0		
		t_3=^{1}H	512	11.1	4.70		
HNCA	600	t_1=^{15}N	30	24.0	119.2	32	
(H_2O; ^{13}C/^{15}N)		t_2=^{13}C	70	26.5	53.0		
		t_3=^{1}H	512	16.0	4.70		
HN(CO)CA	600	t_1=^{15}N	30	24.0	119.0	32	
(H_2O; ^{13}C/^{15}N)		t_2=^{13}C	70	26.5	53.0		
		t_3=^{1}H	512	16.0	4.70		
CBCACONH	500	t_1=^{13}C	28	66.3	35.0	16	
(H_2O; ^{13}C/^{15}N)		t_2=^{15}N	20	29.9	119.0		
		t_3=^{1}H	512	16.0	4.70		
^{1}H-^{1}H NOESY	900	t_1=^{1}H	395	9.77	4.70	70	150
(D_2O; ^{15}N)		t_2=^{1}H	1024	9.79	4.70		
Gly (i+1) ^{1}H-^{15}N HSQC	500	t_1=^{15}N	32	29.9	119.0	64	
(H_2O; ^{13}C/^{15}N)		t_2=^{1}H	512	16.0	4.70		
^{1}H{^{15}N} HNOE	900	t_1=^{15}N	70	30.2	119.0	32	
(H_2O; ^{15}N)		t_2=^{1}H	512	12.0	4.70		
^{13}C filt.-filt. NOESY	900	t_1=^{1}H	200	11.1	4.70	96	150
(D_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	11.1	4.70		
^{13}C filt.-ed. NOESY	900	t_1=^{1}H	250	11.1	4.70	96	150
(D_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	11.1	4.70		
^{13}C ed.-filt. NOESY	900	t_1=^{1}H	100	12.0	4.70	16	150
(D_2O; ^{13}C/^{15}N)		t_2=^{13}C	29	79.8	38.0		
		t_3=^{1}H	512	12.0	4.70		
^{15}N filt.-filt. NOESY [a]	900	t_1=^{1}H	312	11.1	4.70	96	150
(H_2O; ^{15}N)		t_2=^{1}H	1024	11.1	4.70		

[a] see **Figure S3** for the pulse sequence and more experimental details.

Table S2: NMR acquisition parameters for unmodified AcrA$^{61\text{-}210\Delta\Delta}$.

Experiment solvent and sample	field (MHz)	nucleus	acq. pts (complex)	spectral width (ppm)	carrier freq. (ppm)	number of scans	mixing time (ms)
^{1}H-^{15}N HSQC	600	t_1=^{15}N	128	38.0	116.0	4	
(H_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	13.0	4.70		
^{1}H-^{15}N HSQC	900	t_1=^{15}N	64	40.0	116.0	4	
(D_2O; ^{15}N)		t_2=^{1}H	1024	13.9	4.70		
^{1}H-^{13}C HSQC	600	t_1=^{13}C	128	80.0	39.3.0	4	
(aliph., H_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	12.0	4.70		
^{1}H-^{13}C HSQC	900	t_1=^{13}C	64	26.0	125.0	4	
(arom., H_2O; ^{13}C/^{15}N)		t_2=^{1}H	1024	13.9	4.70		
^{15}N NOESY-HSQC	900	t_1=^{1}H	128	11.1	4.70	8	120
(H_2O; ^{15}N)		t_2=^{15}N	48	32.9	118.5		
		t_3=^{1}H	1024	11.1	4.70		
^{13}C NOESY-HSQC	900	t_2=^{1}H	110	11.1	4.70	8	120
(H_2O; ^{13}C/^{15}N)		t_1=^{13}C	55	69.1	37.0		
		t_3=^{1}H	1024	11.1	4.70		
HNCA	600	t_1=^{15}N	33	24.9	118.4	8	
(H_2O; ^{13}C/^{15}N)		t_2=^{13}C	45	30.1	55.6		
		t_3=^{1}H	1024	16.0	4.70		
HN(CO)CA	600	t_1=^{15}N	33	24.9	119.0	16	
(H_2O; ^{13}C/^{15}N)		t_2=^{13}C	45	22.0	53.0		
		t_3=^{1}H	1024	16.0	4.70		
HNCACB	600	t_1=^{15}N	33	24.9	118.4	32	
(H_2O; ^{13}C/^{15}N)		t_2=^{13}C	39	66.3	43.3		
		t_3=^{1}H	1024	16.7	4.70		
HCCH TOCSY	500	t_1=^{1}H	150	12.0	4.70	4	
(H_2O; ^{13}C/^{15}N)		t_2=	44	66.0	38.0		
		t_3=	512	12.0	4.70		
^{1}H-^{1}H NOESY	900	t_1=^{1}H	361	13.4	4.70	64	60
(D_2O; ^{15}N)		t_2=^{1}H	1024	13.4	4.70		
^{1}H{^{15}N} HNOE	900	t_1=^{15}N	70	30.2	119.0	32	
(H_2O; ^{15}N)		t_2=^{1}H	512	12.0	4.70		

Supplementary Methods

Plasmid preparation. All enzymes for cloning were purchased from MBI Fermentas, Tartu, Estonia unless stated otherwise. The pIH1 plasmid[7] has been used as a template for amplification of the AcrA$^{61\text{-}210\Delta\Delta}$ gene by polymerase chain reaction (PCR). The 5'-phosphorylated forward primer **p**-CCCCCCCAT**ATG**AAAAAGACAGCTATCGCGATTGCAGTGGC contained the NdeI cleavage site (underlined) and the START-codon (in bold). 5'phosphorylated reverse primer **p**-ACGCGTCGAC**TCA***GTGGTGGTGGTGGTGGTG* contained a hexa-histidine tag (in italics), a STOP-codon (in bold) and a SalI cleavage site (underlined). Both amplified DNA fragment and the pET24b (Novagen) vector were treated with NdeI and SalI restriction endonucleases for 2 hours at 37 ºC. After this the vector was treated with shrimp alkaline phosphatase (SAP) for 30 minutes at 37 ºC. The ligation by the T4 DNA ligase (New England Biolabs) was carried out at 16 ºC overnight. After this the DNA was transformed into *E. coli* DH5α cells. Selected colonies from the ligation plate were inoculated into 10 ml of LB culture. The cells were incubated overnight at 37 ºC with shaking. The miniprep was performed from obtained cell cultures using the QIAprep Spin Miniprep Kit (Qiagen). The cloned DNA fragment was sequenced (Synergene) confirming the correct position and sequence of the ORF that corresponds to the protein: MKKTAIAIAVALAGFATVAQADVIIKPQVSGVIVNKLFKAGDKVKKGQTLFIIEQDQASKDF NRSKALFSKSAISQKEYDSSLATLDHTEIKAPFDGTIGDALVNIGDYVSASTTELVRVTNLN PIYADGSHHHHHH, in which the OmpA signal peptide attached for an efficient secretion is underlined.

Protein Expression & Purification. AcrA$^{61\text{-}210\Delta\Delta}$ was overexpressed in *E. coli* BL21(DE3) cells carrying the plasmid growing at 37 ºC on M9 minimal medium supplemented with ^{13}C labeled glucose and ^{15}N labeled ammonium chloride. At mid-log phase (OD_{600}=0.4-0.5) cells

were induced by 1 mM IPTG and growth continued for 3 hours. The following procedure was performed at 4 ºC or on ice if not stated otherwise. The cell pellet was resuspended in Polymyxin B (Sigma-Aldrich) solution (26 mg/ml in 30 mM Tris HCl pH 7.5 and 300 mM NaCl) and incubated at 4 ºC for 2 hours with gentle mixing. Ni-NTA purification was performed using the following buffer 30 mM Tris HCl pH 7.5 and 300 mM NaCl supplemented with 10 mM, 20 mM and 250 mM imidazole for the binding, washing and elution buffers, respectively. To adjust the binding conditions the appropriate amount of the elution buffer was added to the soluble periplasm fraction until the final imidazole concentration of 10 mM was reached. Then the periplasm was loaded onto a Hi-Trap Ni-NTA column (volume 1 ml, Invitrogen) equilibrated with 5 ml of binding buffer, followed by a washing step with 10 ml of washing buffer, and elution by 5 ml of elution buffer. The eluted protein was dialyzed against 50 mM Potassium Phosphate buffer (slyde-a-lyzer, 3000 MWCO, Pierce), pH 6.4 and concentrated to 250 µl using an Amicon Ultra 5000 MWCO concentrator (Amicon). The concentration was measured by UV spectrometry and was typically ~ 1 mM from 2 l culture of M9 medium.

MS analysis. Peptides from AspN in-gel digested proteins were resuspended in 10 µl 0.1% trifluroacetic acid, desalted with a µC18 ZipTip (Millipore, Bedford, MA) and directly eluted on a MALDI-plate with αCCA (alpha-cyano-4-hydroxycinnamic acid; 4 mg/ml) in 70% acetonitrile with 0.1% trifluroacetic acid. MALDI-MS/MS was carried out in positive ion reflectron mode using an ABI 4700 MALDI-TOF/TOF Analyzer Instrument (Applied Biosystems).

Structure calculation and refinement of unmodified AcrA$^{61\text{-}210\Delta\Delta}$. Initial structures of unmodified protein were obtained by the programs ATNOS and CANDID[8,9] using the chemical shift assignments of the protein and a 3D ^{15}N-edited NOESY, a 3D ^{13}C-edited NOESY and a homonuclear 2D NOESY in D_2O. 20 structures generated by ATNOS and

CANDID were used as starting structures for refinement in AMBER 8.0 with the AMBER force field 99 along with an implicit water model[10]. In addition to the structural constraints list obtained from ATNOS and CANDID, hydrogen-bond constraints based on the deuterium exchange data and TALOS generated dihedral angle constraints[11] were added in the final calculations. The Ramachandran plot of the folded domain (res. 2-33, 67-100) shows 79.0% of the residues in the most favorable regions, 18.9% in the additionally allowed regions, 1.2% in the generously allowed regions and 0.9% in the disallowed regions. The corresponding values of the entire protein are 74.3%, 22.6%, 1.8% and 1.2%, respectively.

Illustrations. Molecular graphics were generated by MOLMOL[12].

Supplementary Results

Analysis of impurity signals. In addition to the expected signals, resonances of another unlabeled carbohydrate ring were present in the NMR spectra of the glycoprotein. The observed ^{1}H and ^{13}C chemical shifts are identical to those reported for (1-4)-β-D-Xylp oligomers[13,14] and therefore suggest the presence of this sugar oligomer in the glycoprotein sample. No NOEs linking this carbohydrate to either *C. jejuni* heptasaccharide or to protein resonances could be identified. Strong positive intra-sugar NOEs within this unexpected sugar moiety were observed indicating that it is either a repetitive oligomer or that it tumbles together with the glycoprotein. The source, molecular weight and possible interaction of this xylan with the glycoprotein is still awaiting further elucidation.

Glycosidic torsion angles of the NMR ensemble. For the involved glycosidic linkages energy landscapes of disaccharides are available at the GlycoMapsDB database [5] derived from molecular dynamics (MD) calculations. The angles of the presented ensemble are located in energetically favored regions within these energy landscapes (Figure S4). All GalNAc-α(1,4)-GalNAc linkages of the ensemble cluster at the energetically favored region (φ,ψ)=(–40°,–20°). However, the energy landscape shows an additional favorable region at (φ,ψ)=(–10°,40°). The observed medium strong NOEs between HAc2 (i) and H6/H6' (i–1) and strong NOEs between H1 (i) and H6/H6' (i–1) (Figure S1) are only compatible with the conformation at (φ,ψ)=(–40°,–20°). For the other conformation at (φ,ψ)=(–10°,40°) we would expect those NOEs to be weaker than observed. The Glc5'-β(1,3)-GalNAc4 linkage of the ensemble is well defined despite the larger and shallower energy minimum in the energy landscape from GlycoMapsDB. The energy landscape of a GalNAc-α(1,4)-Bac linkage displays two distinct energetic minima at (φ,ψ)=(–40°,–40°) and (–20°,50°). The GalNAc2-α(1,4)-Bac1 linkage of the NMR ensemble clusters at (φ,ψ)=(–40°,–12°) located in the first

minimum. A strong NOE between GalNAc2 H1 and Bac1 HAc4 and a weaker NOE between GalNAc2 H1 and Bac1 HAc2 are only compatible with this conformation. We would expect swapped intensities for the other conformation.

Supporting information on the secondary structure of the glycosylation site. A detailed analysis of the backbone chemical shifts by the program TALOS[11] that predicts backbone ψ/ϕ angles indicates differences between the unmodified and glycosylated AcrA$^{61\text{-}210\Delta\Delta}$. TALOS predicts α-helical angles for residues A37-K39 and F41-K45 for the unmodified AcrA$^{61\text{-}210\Delta\Delta}$ whereas in the glycosylated form only residues S38, K39 and K45 are predicted to adopt α-helical angles. The presence of an α-helix in the unmodified form is supported by typical $H\alpha_{(i)}$-$H\beta_{(i+3)}$ NOEs present for residues A37-S44. However, strong exchange/NOE peaks to the H_2O resonance and very weak or missing diagonal peaks in a ^{15}N edited NOESY indicate water exposed amides typical of partially unfolded conformations in this region. Further confirmation of an α-helix by typical $H\alpha_{(i)}$-$HN_{(i+3)}$ NOEs is hampered by severe chemical shift overlap and NH exchange.

References:

1. Geen, H. & Freeman, R. Band-Selective Radiofrequency Pulses. *Journal of Magnetic Resonance* **93**, 93-141 (1991).
2. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance Series A* **102**, 241-245 (1993).
3. Marion, D., Ikura, M., Tschudin, R. & Bax, A. Rapid Recording of 2d Nmr-Spectra without Phase Cycling - Application to the Study of Hydrogen-Exchange in Proteins. *Journal of Magnetic Resonance* **85**, 393-399 (1989).
4. Lutteke, T., Frank, M. & von der Lieth, C.W. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res* **33**, D242-6 (2005).
5. Frank, M., Lutteke, T. & von der Lieth, C.W. GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Res* **35**, 287-90 (2007).
6. Thibaudeau, C. et al. Correlated C-C and C-O bond conformations in saccharide hydroxymethyl groups: parametrization and application of redundant 1H-1H, 13C-1H, and 13C-13C NMR J-couplings. *J Am Chem Soc* **126**, 15668-85 (2004).
7. Kowarik, M. et al. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* **25**, 1957-66 (2006).
8. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* **24**, 171-89 (2002).
9. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-27 (2002).
10. Case, D.A. et al. *AMBER 8*, (University of California, San Francisco, 2004).
11. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302 (1999).
12. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-5, 29-32 (1996).
13. Hoffmann, R.A., Leeflang, B.R., de Barse, M.M., Kamerling, J.P. & Vliegenthart, J.F. Characterisation by 1H-n.m.r. spectroscopy of oligosaccharides, derived from arabinoxylans of white endosperm of wheat, that contain the elements -4)[alpha-L-Araf-(1-3)]-beta-D-Xylp-(1- or -4)[alpha- L-Araf-(1-2)][alpha-L-Araf-(1-3)]-beta-D-Xylp-(1-. *Carbohydr Res* **221**, 63-81 (1991).
14. Bock, K., Pedersen, C. & Pedersen, H. C-13 Nuclear Magnetic-Resonance Data for Oligosaccharides. *Advances in Carbohydrate Chemistry and Biochemistry* **42**, 193-225 (1984).