J|A|C|S ARTICLES

Published on Web 01/02/2009

NMR Structure Determination of a Segmentally Labeled Glycoprotein Using in Vitro Glycosylation

Vadim Slynko,[†,‡] Mario Schubert,[†,*] Shin Numao,[‡] Michael Kowarik,[‡,§] Markus Aebi,[‡,*] and Frédéric H.-T. Allain[†,*]

Institute of Molecular Biology and Biophysics, and Institute of Microbiology, ETH Zürich, CH-8093 Zürich, Switzerland

Received November 5, 2008; E-mail: schubert@mol.biol.ethz.ch; aebi@micro.biol.ethz.ch; allain@mol.biol.ethz.ch

Abstract: Although there is great interest in three-dimensional structures of glycoproteins and complex oligosaccharides, their structural determination have been hampered by inhomogeneous and incomplete glycosylation, poor expression, low tendency to crystallize, and severe chemical shift overlap. Using segmental labeling of the glycan and the protein component by in vitro glycosylation, we developed a novel method of NMR structural determination that overcomes some of these problems. Highly homogeneously glycosylated proteins in milligram amounts can be obtained. This allowed the determination of the structure of an N-linked glycoprotein from *Campylobacter jejuni*. The glycosylation acceptor site was found to be in a flexible loop. The presented methodology extends the observable NOE distance limit of oligosaccharides significantly over 4 Å, resulting in a high number of distance restraints per glycosidic linkage. A well-defined glycan structure was obtained.

Introduction

Glycosylation of proteins is a common feature in eukaryotes, archaea,[1] and bacteria[2] that expands the diversity of the proteome. Glycoproteins play crucial roles in a variety of biological processes such as cell growth and differentiation, cell−cell interactions, development, modulation of the immune system, inflammation and cancer, protein folding, quality control and turnover, or pathogenicity and host invasion of bacteria.[3] However, the understanding of the molecular basis of these processes is still quite limited mainly because of a lack of structural data available for glycoproteins. The two main techniques that are used to structurally analyze glycoproteins, X-ray crystallography and NMR spectroscopy, face a variety of obstacles:[4] expression from natural sources is normally poor and glycosylation is often inhomogeneous and incomplete; only a limited number of protein structures with intact glycans have been determined by X-ray crystallography because of a low tendency to crystallize and an absence of electron density for the glycans. Although NMR spectroscopy is frequently used to study oligosaccharides and proteins, only 15 glycoprotein structures have been determined with NMR spectroscopy (PDB search tool at http://www.glycosciences.de), and out of these, only 6 contain a glycan with more than one saccharide.[5−9] The two major obstacles are as follows: first, the severe ^{1}H chemical shift overlap between carbohydrate and protein signals, and second, the production of a sufficient amount of ^{15}N or ^{13}C/^{15}N-labeled glycoconjugate in the predominantly used eukaryotic expression systems. So far, only uniform ^{15}N or ^{13}C/^{15}N labeling of a complete glycoprotein has been possible, e.g. in Chinese hamster ovary (CHO) cells[6] and *Pichia pastoris*.[10]

The N-glycosylation machinery of *Campylobacter jejuni* is the best characterized glycosylation system in bacteria.[11] It is encoded by 12 genes that are organized into a protein glycosylation locus, *pgl*, which can be functionally transferred to *E. coli*.[12] Most of the genes of the *pgl* locus are involved in the synthesis of the lipid-linked oligosaccharide (LLO) and its flipping into the periplasmic space. There, the oligosaccharyltransferase PglB attaches the glycan to the side chain of an asparagine residue in the consensus sequence D/E-X-N-Z-S/T (X and Z are any amino acid residue except proline) on the acceptor protein.[13] The glycan structure was found to be GalNAc-α1,4-GalNAc-α1,4-[Glc-β1,3-]GalNAc-α1,4-GalNAc-

† Institute of Molecular Biology and Biophysics.
‡ Institute of Microbiology.
§ Present address: GlycoVaxyn AG, Grabenstrasse 3, 8952 Schlieren, Switzerland.

(1) Eichler, J.; Adams, M. W. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 393–425.
(2) Benz, I.; Schmidt, M. A. *Mol. Microbiol.* **2002**, *45*, 267–76.
(3) Varki, A. *Glycobiology* **1993**, *3*, 97–130.
(4) Meyer, B.; Moller, H. *Top. Curr. Chem.* **2007**, *267*, 187–251.
(5) Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A. R.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269*, 1273–8.
(6) Metzler, W. J.; Bajorath, J.; Fenderson, W.; Shaw, S. Y.; Constantine, K. L.; Naemura, J.; Leytze, G.; Peach, R. J.; Lavoie, T. B.; Mueller, L.; Linsley, P. S. *Nat. Struct. Biol.* **1997**, *4*, 527–31.
(7) Erbel, P. J.; Karimi-Nejad, Y.; van Kuik, J. A.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. *Biochemistry* **2000**, *39*, 6012–21.
(8) Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D. *Structure* **1994**, *2*, 185–99.
(9) Hashimoto, Y.; Toma, K.; Nishikido, J.; Yamamoto, K.; Haneda, K.; Inazu, T.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1999**, *38*, 8377–84.
(10) Wood, M. J.; Sampoli Benitez, B. A.; Komives, E. A. *Nat. Struct. Biol.* **2000**, *7*, 200–4.
(11) Szymanski, C. M.; Logan, S. M.; Linton, D.; Wren, B. W. *Trends Microbiol.* **2003**, *11*, 233–8.
(12) Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. *Science* **2002**, *298*, 1790–3.

 10.1021/ja808682v CCC: $40.75

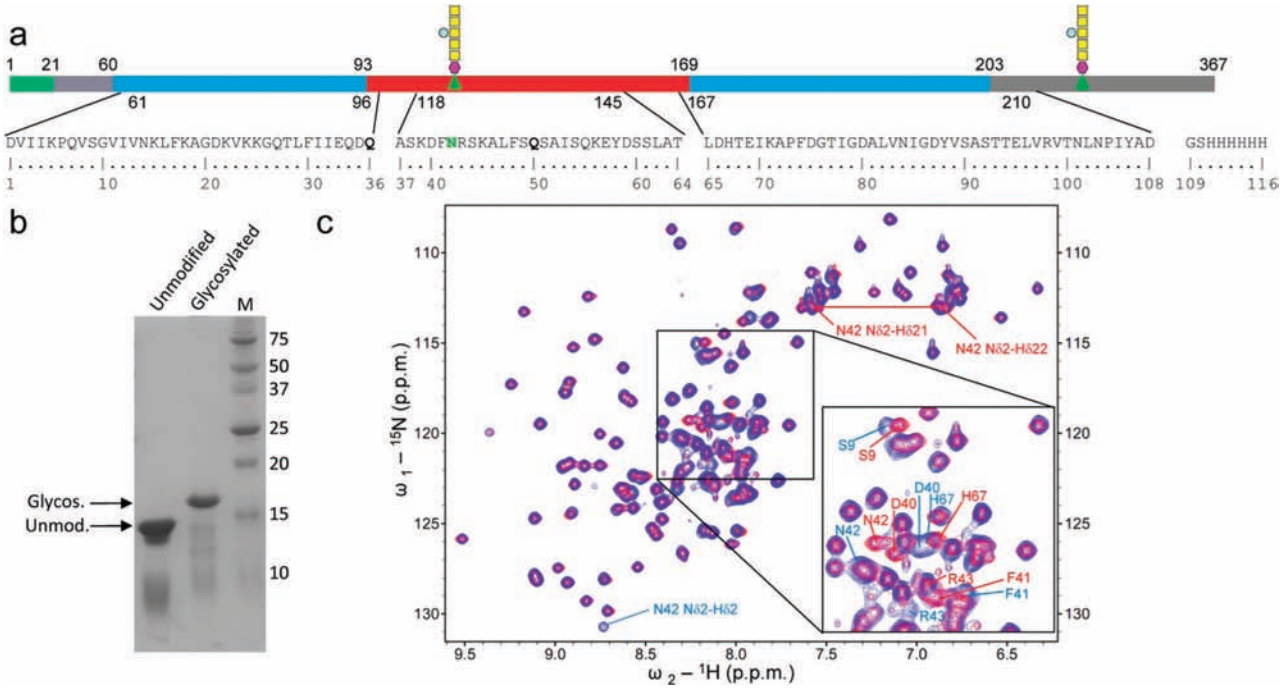

Figure 1. Glycosylation of a truncated version of AcrA. (a) The sequence of the AcrA$^{61-210\Delta\Delta}$ construct used in this study (bottom) and its relation to the AcrA wild type (top). The glycans are shown schematically. Glucose is depicted as a blue circle, GalNac as a yellow square, and bacillosamine as a magenta hexagon. In the protein sequence the Asn of the glycosylation site is highlighted in green and the point mutations K96Q and K131Q in bold. (b) In vitro glycosylation of recombinant AcrA$^{61-210\Delta\Delta}$ monitored by SDS-PAGE. (c) ^{1}H−^{15}N HSQC spectra of the unmodified (red) and glycosylated form (blue) of AcrA$^{61-210\Delta\Delta}$ at ∼1 mM and 303 K.

α1,4-GalNAc-α1,3-Bac-β1,N-Asn, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucopyranose.[14]

Bacterial N-glycosylation can function on folded proteins, independent of the translocation process, as evident from both in vivo and in vitro experiments.[15] Whereas the amino acids at position ±1 of the Asn influence the recognition of pentapeptides by PglB in vitro,[16] this influence is decreased in the context of a protein. The glycosylation efficiency appears to be correlated to the folding properties of the acceptor site. Flexible folding isomers of the recombinant bovine RNase are more efficiently glycosylated than rigid ones,[15] indicating that in vitro glycosylation efficiency depends on the conformational state(s) and dynamics of the acceptor site.

We used the bacterial N-glycosylation system to develop a novel strategy for the study of glycoprotein structures by NMR spectroscopy that overcomes some of the crucial limitations mentioned above. In vitro glycosylation was used to obtain a glycoprotein consisting of a nonlabeled glycan attached to a uniformly ^{13}C/^{15}N labeled protein. We present the 3D structure of the glycoprotein including the oligosaccharide structure and a comparison to the nonglycosylated protein.

Results

The Model Protein and Its in Vitro Glycosylation. As a model system we used the small and efficiently glycosylatable protein AcrA$^{61-210\Delta\Delta}$ from *C. jejuni* (Figure 1a) consisting of residues D61-D210 of the natural AcrA protein with deletions of residues F97-N117, F146-D166 and K96Q, K131Q point mutations.[13] We chose this 13 kDa construct containing the lipoyl domain and an extended loop harboring the glycosylation site because of the good dispersion and line widths in the NMR spectra and because it can be efficiently glycosylated both in vivo and in vitro. An OmpA signal peptide was N-terminally attached for an efficient secretion. This signal sequence was absent in the mature protein. A C-terminal His$_6$-tag was added for purification.

The in vitro glycosylation system consists of three principle components, expressed and isolated separately in dedicated *E. coli* strains: the oligosaccharyltransferase PglB, the lipid-linked oligosaccharide (LLO), and the target protein. The oligosaccharyltransferase PglB was expressed in *E. coli* and purified from solubilized membrane fractions, the LLO was prepared from *E. coli* cells bearing a plasmid containing the pgl locus from *C. jejuni* with an inactivated PglB ORF.[15] Homogeneously ^{13}C/^{15}N-labeled protein containing a His$_6$-tag was expressed in *E. coli*. The purified AcrA$^{61-210\Delta\Delta}$ was glycosylated in vitro and subsequently purified by Ni-NTA affinity chromatography in milligram quantities, which was sufficient for NMR spectroscopy. Typically, a yield of 0.25 μmol glycoprotein was obtained from 2 L of M9 minimal medium culture for expressing the target protein, from 30 L of LB culture for obtaining the LLO, and from 2 L of TB culture for expressing PglB. The degree of glycosylation was ∼90% as judged by SDS-Page (Figure 1b) and NMR spectra.

NMR Spectroscopy, Structure, and Dynamics of the Glycosylated AcrA$^{61-210\Delta\Delta}$. The ^{15}N-HSQC spectrum of the glycosylated protein is shown in Figure 1c. Almost complete resonance assignment was obtained using standard triple-resonance experiments. The side-chain NH signal of the glycosylated N42 was identified by a glycine selective ^{15}N-

(13) Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M. *EMBO J.* **2006**, *25*, 1957–66.
(14) Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St. Michael, F.; Aberg, E.; Szymanski, C. M. *J. Biol. Chem.* **2002**, *277*, 42530–9.
(15) Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M. *Science* **2006**, *314*, 1148–50.
(16) Chen, M. M.; Glover, K. J.; Imperiali, B. *Biochemistry* **2007**, *46*, 5579–85.

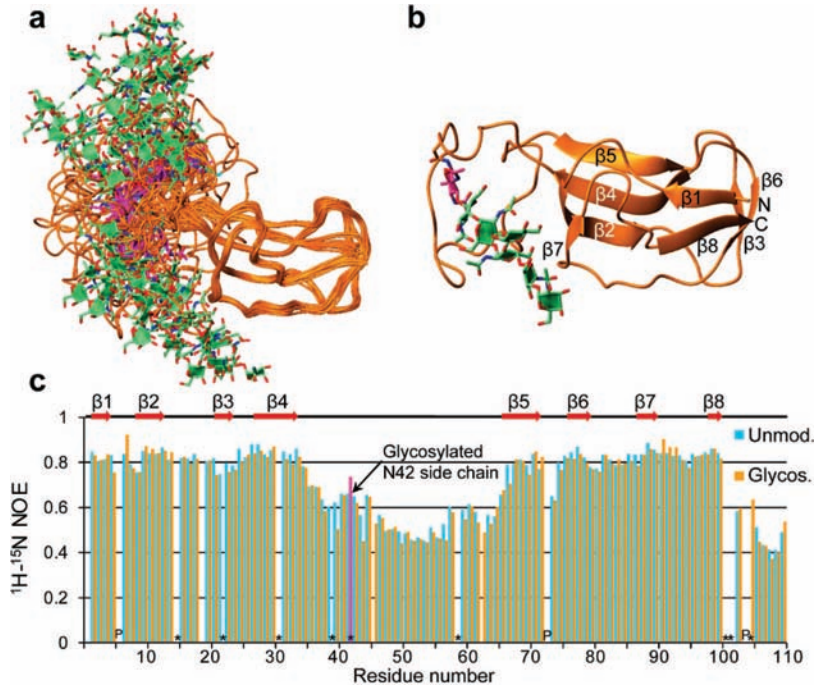

Figure 2. The structure and dynamics of the glycosylated and the unmodified AcrA$^{61-210\Delta\Delta}$. (a) Ensemble of the glycosylated AcrA$^{61-210\Delta\Delta}$ with the glycan. Structures were superimposed on the Cα atoms of residues 2−33 and 67−100. The first carbohydrate ring Bac-1 is colored in magenta and the side chain of N42 in cyan. (b) Representative structure of the glycoprotein. (c) Backbone heteronuclear ^{1}H{^{15}N}-NOE relaxation values of the unmodified (blue) and glycosylated AcrA$^{61-210\Delta\Delta}$ (orange) acquired at 900 MHz as a function of the residue number. The secondary structure is depicted on top. Missing bars are due to Pro residues (P) or spectral overlap (*). The data point of the N42 backbone is missing, but the N42 side-chain NH is shown in magenta.

HSQC specific for $CH_2-CO-NH$ moieties[17] and confirmed by triple resonance spectra. The ^{1}H and ^{15}N resonances are similar to those of previously reported glycosylated asparagines.[10] Using a total of 1932 NOE distance restraints, an ensemble of 20 best conformers of the glycoprotein was obtained (Figure 2a,b). The construct adopts a lipoyl domain fold consisting of four N-terminal and four C-terminal intertwined β-strands forming a β-sandwich,[18] very similar to AcrA from *E. coli* and to MexA from *P. aeruginosa*.[19−21] A flexible loop harboring the glycosylation site is attached between the two lipoyl half-motifs consisting of residues D1−D35 and H67−N100 (Figure 2b). For details of the applied constraints and structural statistics, see Table 1. To obtain experimental evidence for this flexibility, we recorded steady-state heteronuclear ^{1}H{^{15}N}-NOE data (Figure 2c), a sensitive indicator of motions on the subnanosecond time scale. Residues 35 to 65 including the glycosylation site N42 exhibit reduced ^{1}H{^{15}N}-NOE values indicative of dynamics. The NH of the glycosylated N42 side chain appears to be less flexible than the backbone of the surrounding amino acid residues probably because of the reduced conformational degree of freedom imposed by the attached glycan.

The Structure of the Glycan and Its Interaction with the Protein. Choosing ^{13}C/^{15}N labeling for the protein and no labeling for the carbohydrate enabled us to separate signals from the two components by filtering and editing techniques (Figure 3). The constitution of the *C. jejuni* N-glycan (Figure 4a) was previously determined by mass spectrometry and NMR spectroscopy using a glycopeptide.[14] With the help of the chemical shifts reported,[14] we could obtain complete resonance assignment of the glycan attached to AcrA$^{61-210\Delta\Delta}$ using a 2D ^{13}C-filtered-filtered NOESY[22] recorded at 900 MHz (Figure 3b and details in Figure S1 in the Supporting Information) and a natural abundance ^{13}C-HSQC spectrum (Figure S2 in the Supporting Information). In the former, all protein resonances are suppressed resulting in a 2D NOESY spectrum of the unlabeled carbohydrate. In addition signals of an impurity, likely a (1−4)-β-D-Xylp oligomer, were detected (see Supporting Information results). The glycan assignments were further confirmed by a 2D ^{15}N-filtered-filtered NOESY that contains NOEs to the sugar amides while suppressing protein amide signals (Figure 3a). This spectrum was very helpful to unambiguously assign the acetamido groups present in the glycan because the sugar amide resonances show strong intraresidual NOEs to both the H2 and to the acetamido CH_3 group. The pulse sequence is depicted in Figure S3 in the Supporting Information.

The high field (900 MHz) and a mixing time of 150 ms used for the NOESY spectra helped to extract a total of 125 interproton distance constraints within the attached glycan (depicted schematically in Figure 4b). The high sensitivity, high resolution, and favorable tumbling time lead to the observation of this unexpectedly large number of NOEs. Half of those are inter-residue NOEs, on average 11 per glycosidic linkage. The 2D ^{15}N-filtered-filtered NOESY was particularly useful since it provided 15 out of the 66 inter-residue restraints. Torsion angle dynamics calculations[23] with upper distance constraints derived from the NOE intensities resulted in a well-defined and rod-

(17) Schubert, M.; Oschkinat, H.; Schmieder, P. *J. Magn. Reson.* **2001**, *148*, 61–72.
(18) Johnson, J. M.; Church, G. M. *J. Mol. Biol.* **1999**, *287*, 695–715.
(19) Akama, H.; Matsuura, T.; Kashiwagi, S.; Yoneyama, H.; Narita, S.; Tsukihara, T.; Nakagawa, A.; Nakae, T. *J. Biol. Chem.* **2004**, *279*, 25939–42.
(20) Higgins, M. K.; Bokma, E.; Koronakis, E.; Hughes, C.; Koronakis, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9994–9.
(21) Mikolosko, J.; Bobyk, K.; Zgurskaya, H. I.; Ghosh, P. *Structure* **2006**, *14*, 577–87.
(22) Peterson, R. D.; Theimer, C. A.; Wu, H.; Feigon, J. *J. Biomol. NMR* **2004**, *28*, 59–67.

Table 1. NMR Structure Determination Statistics of Glycosylated and Unmodified AcrA$^{61-210\Delta\Delta}$

	glycosylated protein	unmodified protein
NMR Distance and Dihedral Constraints		
distance restraints		
total NOE (protein)	1784	1681
intraresidue	410	402
inter-residue	1374	1279
sequential ($i - j = 1$)	568	471
nonsequential ($i - j > 1$)	806	808
hydrogen bonds	26	16
total NOE (glycan)	125	
intraresidue	59	
inter-residue	66	
sequential ($i - j = 1$)	65	
nonsequential ($i - j > 1$)	1	
hydrogen bonds	0	
protein−glycan NOE	23	
total dihedral angle restraints	108	110
protein		
ϕ	51	55
ψ	50	55
glycan		
HN−CO peptide bonds of acetamido	7	
sugar pucker	0	
Structure Statistics		
violations (mean and SD)		
number of distance constraint violations >0.2 Å	21 ± 6	1.35 ± 1.0
max. distance constraint violation (Å)	0.65 ± 0.17	0.29 ± 0.08
deviations from idealized geometry		
bond lengths (Å)	0.005	0.013
bond angles (deg)	0.7	2.2
impropers (deg)		
average pairwise rms deviation[a] (Å)		
protein (res. V2-E33, H67-T100)		
heavy	0.86 ± 0.12	0.83 ± 0.12
backbone	0.46 ± 0.10	0.32 ± 0.07
glycan		
all glycan heavy	0.56 ± 0.10	
protein and glycan heavy	7.0 ± 1.1	
average pairwise rms deviation[a] (Å) between the glycosylated and unmodified protein (res. V2-E33, H67-T100)	1.05 ± 0.19	

[a] Pairwise rms deviation was calculated among 20 refined structures.

like structure with an rmsd of 0.56 ± 0.10 Å (Figure 4c and Table 1). The torsion angles φ and ψ of the glycosidic linkages cluster well and display only small deviations pointing toward a single stable conformation (Figure S4 in the Supporting Information). The glycosidic angles are located in energetically favored regions of theoretically calculated energy landscapes[24] described in detail in the Supporting Information. It is noteworthy that the glycan structure was calculated with torsion angle dynamics solely using experimentally derived upper distance NOE restraints and thus without any force field. The fact that the resulting ensemble of structures lies in a theoretical energy minimum demonstrates that the NMR data provide enough constraints to accurately determine a precise structure of the attached sugar.

Twenty-three NOE constraints of the glycan/protein interface derived from a 3D ^{13}C-edited-filtered NOESY (Figure S5 in the Supporting Information) were applied to orient the glycan in respect to the protein. As shown in Figure 2a,b the orientation of the glycan rod covered a broad range because of the flexibility of the protein backbone around the glycosylation site. However, a common feature in all structures is that the glycan is kinked due to the $\beta(1-3)$ linked Bac and tends to orient toward the residues preceding the glycosylated N42. This is in agreement with the fact that the largest chemical shift perturbations occur around residues 35 to 42 (Figure 5). However, no specific interactions between the protein and the carbohydrate rings beyond α-GalNac2 were observed. NOEs between the glycan and the protein are only observed to β-Bac1 and α-GalNac2 resulting in the backward orientation of the glycan arm kinked at the β-(1−3)-Bac1 linkage (Figure 4c,d).

Structural Comparison Between the Glycoprotein and the Nonglycosylated Protein. In order to address the issue of whether the target glycosylation site is required to form a predefined structure,[25] we determined the structure of the nonglycosylated AcrA$^{61-210\Delta\Delta}$. Using a total of 1681 NOE distance restraints an ensemble of 20 best conformers from 100 calculated structures was obtained (Figure 5a,b). The structure of the lipoyl domain is virtually the same in the unmodified and the glycosylated protein, and the rmsd between the two mean structures of both ensembles is 1.05 Å for backbone atoms (residues 2−33 and 67−100).

As a sensitive indicator for differences in the chemical environment, backbone amide chemical shift deviations

(23) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–98.
(24) Frank, M.; Lutteke, T.; von der Lieth, C. W. *Nucleic Acids Res.* **2007**, *35*, 287–90.
(25) Imperiali, B.; Shannon, K. L.; Rickert, K. W. *J. Am. Chem. Soc.* **1992**, *114*, 7942–7944.

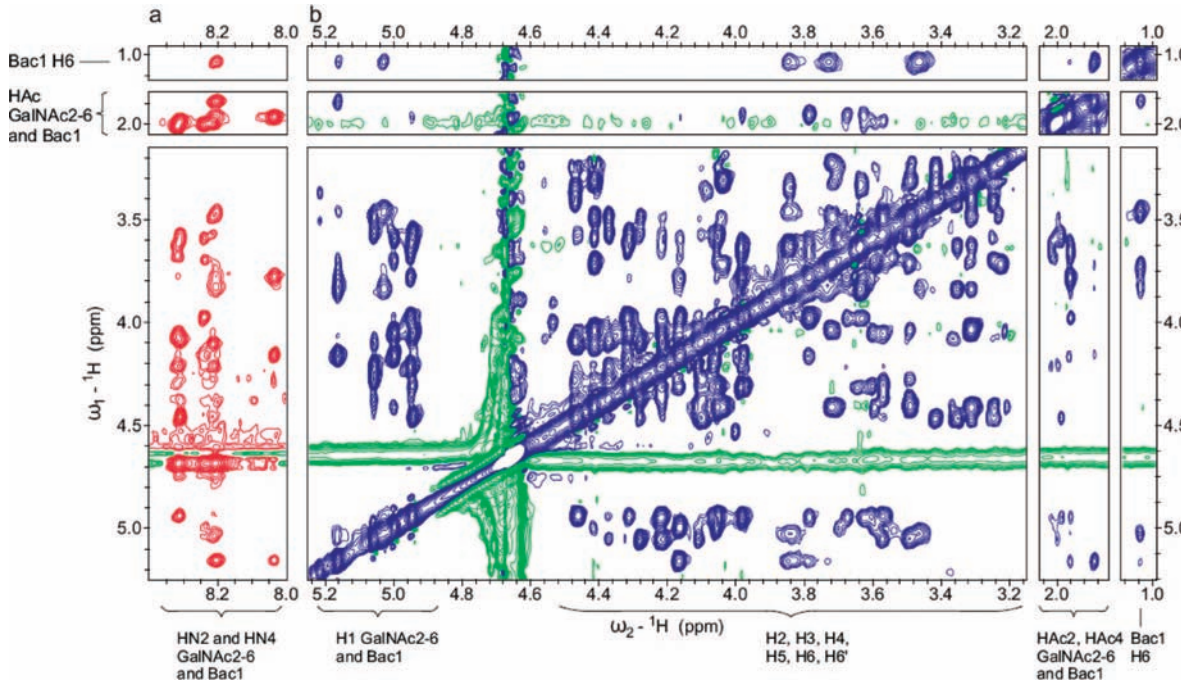

Figure 3. 2D NOESY spectra of glycosylated AcrA$^{61-210\Delta\Delta}$ used to derive NOE restraints for the glycan, measured at 303 K and 900 MHz. (a) ^{15}N-Filtered-filtered 2D NOESY of the glycoprotein (^{15}N-labeled protein) showing NOEs to the glycan HN2 and HN4 resonances recorded with 96 scans in ~27 h. (b) ^{13}C-filtered-filtered 2D NOESY of the glycoprotein (^{13}C/^{15}N-labeled protein) recorded with 96 scans in ~25 h. Here only NOEs within the glycan are observable. Assignments of the NOEs in selected regions are shown in Figure S1 in the Supporting Information.

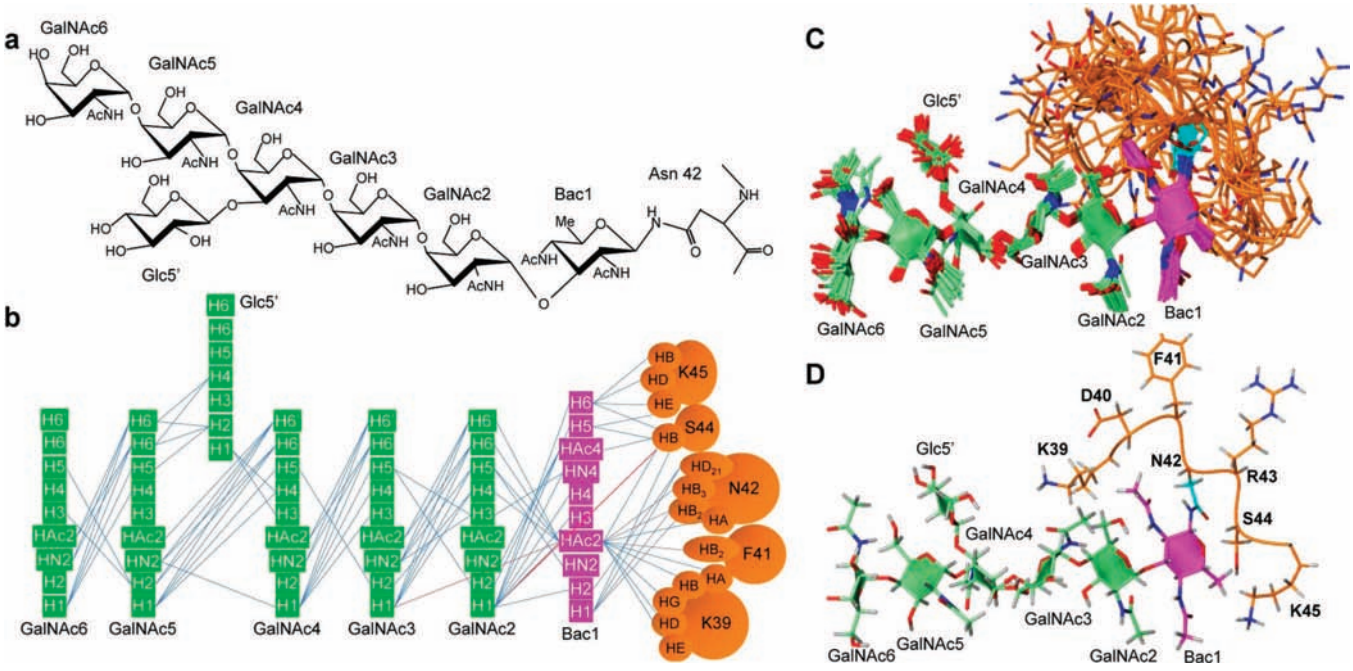

Figure 4. The structure of the Asn-linked heptasaccharide glycan of *C. jejuni*. (a) The chemical structure. (b) Schematic presentation of the inter-residue NOE constraints within the glycan and between the glycan and the protein. Distance constraints between nonconsecutive sugar rings and between GalNAc2 and the protein are indicated by red lines. (c) Structural ensemble of the glycan superimposed on the heavy atoms of all saccharides. The rmsd is 0.56 ± 0.10 Å. (d) Representative structure of the glycan.

between the unmodified and glycosylated AcrA$^{61-210\Delta\Delta}$ are displayed in Figure 5c. Not surprisingly the largest deviations are found for residues around the glycosylation site N42 at the beginning of the flexible loop (residue 35 to 46). The resonances of the lipoyl domain are generally unaffected. Figure 5d shows differences in the Cα chemical shift between the unmodified and the glycosylated AcrA$^{61-210\Delta\Delta}$. Cα chemical shift deviations from random coil values, indicative of secondary structures, show differences for residues D35−N42. For the unmodified AcrA$^{61-210\Delta\Delta}$ deviations between 1.0 and 1.7 ppm are found for residues Q36−S44 (deviations larger than +1.5 ppm are indicative of an α-helix[26]) whereas for the glycosylated AcrA$^{61-210\Delta\Delta}$ the values are systematically lower for this region. The data indicate a tendency for the formation of an α-helix around the glycosylation site Q36−S44 in unmodified AcrA$^{61-210\Delta\Delta}$

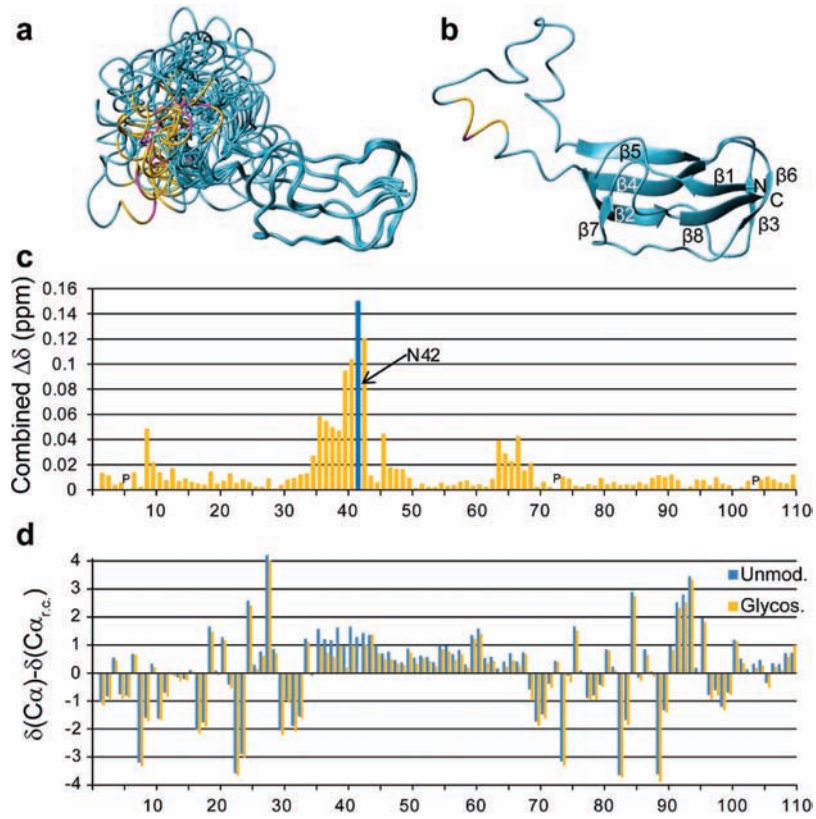

Figure 5. The structure of the unmodified AcrA$^{61-210\Delta\Delta}$ and chemical shift differences to the glycosylated protein. (a) Ensemble of the nonglycosylated AcrA$^{61-210\Delta\Delta}$. The glycosylation site N42 is colored in magenta, neighboring residues D40-S44 in yellow. (b) Representative structure of the nonglycosylated AcrA$^{61-210\Delta\Delta}$. (c) Combined chemical shift deviations ($\Delta\delta = [(\delta_{HN})^2 + (\delta_N/6.51)^2]^{1/2}$) are plotted versus the residue number. (d) Cα chemical shift difference to random coil values of the unmodified (red) and glycosylated form (blue) of AcrA$^{61-210\Delta\Delta}$.

but do not support the presence of a stable helix. No significant changes of the dynamics in the subnanosecond time scale around the glycosylation site are observed upon glycosylation indicated by reduced ^{1}H{^{15}N}-NOE values of ~0.6 (Figure 2c). Taking together the chemical shift data, the NOESY spectra and the ^{1}H{^{15}N}-NOE data, the glycosylation site in the unmodified protein appears to be partially flexible with a tendency to form an α-helix. Upon glycosylation, this preference to form an α-helix is lost.

Discussion

A New Strategy for the NMR Study of Glycoproteins. In order to overcome limitations of structure determination of glycoproteins and complex oligosaccharides, we present here a novel strategy using NMR spectroscopy that includes attaching an unlabeled oligosaccharide to a ^{13}C/^{15}N-labeled recombinant expressed protein and filtered/edited NOESY techniques at high field (900 MHz). The recently introduced in vitro N-glycosylation technology[15] was up-scaled and optimized to yield highly homogeneously glycosylated proteins in milligram amounts. This has been a long time goal in glycobiology and represents a major advance to structural biology. The protein can be uniformly ^{13}C/^{15}N labeled using standard recombinant techniques, enabling the selective observation of NMR signals of either the protein or the glycan alone, or of signals at the interface between the glycan and the protein using advanced filtering and editing NMR techniques.

The attachment of the oligosaccharide to a protein has the advantage that the NOE transfer within the carbohydrate becomes very efficient because of the increased overall tumbling time. The NOE build up and efficiency is then comparable to those observed in proteins of a similar size. In addition we use high field NMR spectroscopy at 900 MHz to maximize the sensitivity and resolution of the NOESY spectra. This enabled us to observe long-range NOEs significantly further than 4 Å, and we classified the weakest NOEs into 6 Å upper distance restraints in analogy to protein structure calculations. In contrast NOEs observed from isolated oligosaccharides in solution range usually only up to ~4 Å[27] mainly because of the unfavorable tumbling regime and the smaller maximum theoretical NOE of the usually used ROESY experiments.[28] The improved NOE transfer results in a crucially enhanced quality of oligosaccharide 3D structures. The conformation of the glycosidic linkages can be well defined with ~10 inter-residue restraints per linkage whereas for isolated oligosaccharides in solution the low number of restraints, typically 2−4, are often not sufficient to define the oligosaccharide conformation.[29]

The use of a ^{13}C/^{15}N-labeled protein was crucial for the presented approach because severe chemical shift overlapping between the protein and the glycan could be resolved using filtered/edited techniques. In particular ^{1}H signals of Ser, Lys, Arg, and Asp side chains and various Hα atoms can thus be

(26) Luginbuhl, P.; Szyperski, T.; Wuthrich, K. *J. Magn. Reson. Ser. B* **1995**, *109*, 229–233.
(27) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371–86.
(28) Bothnerby, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811–813.
(29) Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–16.

separated from glycan resonances. Such overlap severely hampers the resonance assignment and structure determination of unlabeled or ^{15}N-labeled glycoproteins.[30] In addition to the improved quality of the oligosaccharide structure, the glycan−protein interface can be defined by NOEs from a ^{13}C-edited-filtered NOESY experiment. This will be of great advantage in cases of a well-defined glycan−protein interface like in the case of human CD2[5] or human chorionic gonadotropin.[7]

While in vitro glycosylation is at present only applicable for attaching bacterial glycans, there are alternatives to obtain such a homogeneous and segmental isotopically labeled glycoprotein consisting of a ^{13}C/^{15}N-labeled protein and an unlabeled oligosaccharide including transglycosylation, and synthetic or semisynthetic approaches.[31−33] Such alternatives would make a wider variety of glycans accessible.

Implications to Our Understanding of the Glycosylation Mechanism. We applied the presented strategy to gain insights into the structural and dynamic requirements of the acceptor protein that is recognized by the oligosaccharyltransferase PglB. Biochemical evidence suggested that a functional glycosylation acceptor site is located in a flexible region of a folded protein.[15] Our data are in agreement with this hypothesis. We demonstrate that the glycosylation site is located in a moderately flexible region that partially adopts an unstable α-helix, and random coil conformations making the site easily accessible to the oligosaccharyltransferase. Whether the flexibility was required for the recognition and/or for catalytic activity remains unclear. However, we cannot exclude that an acceptor site with a defined rigid conformation can also be recognized and glycosylated. Our results stand in contrast to studies of small peptides indicating that the target glycosylation site is predominately adopting a predefined Asn-turn that get destabilized upon glycosylation.[34,35] An Asn-turn is characterized by a hydrogen bond between the Asn side chain Oδ1 and both the NH and OH of Ser/Thr in position (i+2). We do not observe any inter-residue NOEs supporting an Asn-turn in the unmodified AcrA$^{61-210\Delta\Delta}$. On the contrary, we observe an α-helix in exchange with unstructured conformations that gets destabilized upon glycosylation. However, we cannot exclude that a transient Asn-turn conformation induced by the enzyme might be required for efficient glycosylation.

Structure of the Glycan. In contrast to the general belief that glycans are flexible molecules without a defined three-dimensional structure, we found that the bacterial N-linked glycan of *C. jejuni* forms a well-defined rod-like structure, suggesting that structured glycans might be more common than expected. Few comparable well-defined and extended oligosaccharide structures have been reported: of a chondroitin sulfate pentasaccharide (rmsd of 0.53 Å for residue 1 to 4),[36] an octasaccharide isolated from chondroitin sulfate[37] and fucosylated penta- and hexasaccharides.[38] In these studies, residual dipolar coupling (RDC) and chemical shift anisotropy (CSA) restraints were used to complement the insufficient amount of NOE restraints. However, NMR structures present a time average over several hundred milliseconds, and there is likely to be local flexibility present at each of the glycosidic linkages at a time scale faster than the NMR time scale. Oligosaccharides containing bulky substitutions like acetamido or SO_4^{2-} groups or multiple branch points can form well-defined structures due to steric hindrance but also potential hydrogen bonds. Molecular dynamics (MD) simulations of $Man_9GlcNAc_2$ suggest that the high local flexibility of the glycosidic linkages is not translated into gross changes in the structure resulting in a moderately defined overall topology.[39] As suggested by the Almond group, weak transient but 'long-lived' hydrogen bonds and water bridges may be the key in understanding the three-dimensional structure of oligosaccharides.[40] The *C. jejuni* glycan can potentially be stabilized by hydrogen bonds that are observed in a few conformers as depicted in Figure S6 in the Supporting Information.

What might be the role of a structured glycan? In glycan−protein interactions large oligosaccharides compared to smaller ones are recognized by more hydrogen bonds, giving rise to a favorable binding enthalpy. However, this often comes with a high entropic cost that is generated by the loss in degrees of flexibility upon binding.[41] This is the reason why most lectin−oligosaccharide interactions are restricted to an affinity in the mM to μM range. However, if the unbound oligosaccharide is already structured, the entropy cost is low and very high affinities can be reached. This is exemplified by the recognition of the rigid trisaccharide moiety Lewis(a) by PA-IIL with an affinity of 220 nM.[42] However, there is no protein known so far that binds the N-glycan of *C. jejuni* with high affinity.

Conclusion and Perspective. We developed a novel method to determine the 3D structure of glycoproteins and oligosaccharides by NMR spectroscopy using segmental labeling and in vitro N-glycosylation. The method was applied to a glycoprotein containing the N-glycan of *C. jejuni*. The development of a method to replace this glycan by eukaryotic glycans or to attach eukaryotic glycans to folded proteins is on its way and will make our proposed strategy generally applicable to a wide range of glycoproteins. This will pave the way to study the 3D structure of a wide variety of glycoproteins and of oligosaccharides.

Methods

Plasmid Preparation, Protein Expression, and Purification. The details of molecular cloning, expression, and purification are described in the Supporting Information. In short, AcrA$^{61-210\Delta\Delta}$ was expressed by using a modified pIH1 plasmid[13] that contained a C-terminal His_6-tag and a N-terminal OmpA signal peptide attached for an efficient secretion. The recombinant protein was purified using Ni-NTA chromatography.

(30) Wyss, D. F.; Choi, J. S.; Wagner, G. *Biochemistry* **1995**, *34*, 1622–34.
(31) Bennett, C. S.; Wong, C. H. *Chem. Soc. Rev.* **2007**, *36*, 1227–38.
(32) Brik, A.; Ficht, S.; Wong, C. H. *Curr. Opin. Chem. Biol.* **2006**, *10*, 638–44.
(33) Yamamoto, N.; Tanabe, Y.; Okamoto, R.; Dawson, P. E.; Kajihara, Y. *J. Am. Chem. Soc.* **2008**, *130*, 501–10.
(34) Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421–5.
(35) OConnor, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **1997**, *119*, 2295–2296.
(36) Yu, F.; Wolff, J. J.; Amster, I. J.; Prestegard, J. H. *J. Am. Chem. Soc.* **2007**, *129*, 13288–97.
(37) Blanchard, V.; Chevalier, F.; Imberty, A.; Leeflang, B. R.; Basappa; Sugahara, K.; Kamerling, J. P. *Biochemistry* **2007**, *46*, 1167–75.
(38) Almond, A.; Petersen, B. O.; Duus, J. O. *Biochemistry* **2004**, *43*, 5853–63.
(39) Woods, R. J.; Pathiaseril, A.; Wormald, M. R.; Edge, C. J.; Dwek, R. A. *Eur. J. Biochem.* **1998**, *258*, 372–86.
(40) Almond, A. *Carbohydr. Res.* **2005**, *340*, 907–20.
(41) Imberty, A.; Mitchell, E. P.; Wimmerova, M. *Curr Opin Struct Biol* **2005**, *15*, 525–34.
(42) Perret, S.; Sabin, C.; Dumon, C.; Pokorna, M.; Gautier, C.; Galanina, O.; Ilia, S.; Bovin, N.; Nicaise, M.; Desmadril, M.; Gilboa-Garber, N.; Wimmerova, M.; Mitchell, E. P.; Imberty, A. *Biochem. J.* **2005**, *389*, 325–32.

In Vitro Glycosylation. The reaction conditions for the in vitro glycosylation of AcrA$^{61-210\Delta\Delta}$ was similar to that reported previously (100 mM HEPES, pH 7.5; 1 mM $MnCl_2$; 1% Triton X100; 100 nM PglB; 16 h reaction at 30 °C).[15] As the LLO was the limiting reagent, the reaction was initially done in small scale (100 μL reaction size) to determine the minimal amount of LLO required for full conversion. Depending on the batch, 40 to 60 μL of LLO was required to convert 1 nmol of purified AcrA$^{61-210\Delta\Delta}$. Assuming the 100% usage of the LLO in the crude LLO (cLLO) extract, we estimated the concentration of the LLO to be 20 μM. To scale up, every component was increased proportionally to a total volume of 80 mL. The reaction was performed in Falcon tubes shaken in an orbital shaker overnight at 30 °C. The mixture was centrifuged, and the protein was purified by Ni-NTA affinity chromatography and concentrated as described for the unmodified AcrA$^{61-210\Delta\Delta}$. About 90% of the initial AcrA$^{61-210\Delta\Delta}$ was recovered in the glycosylated form as judged by UV. Typical yields of 0.25 μmol glycoprotein were obtained from 2 L of M9 minimal medium culture for expressing the protein and 30 L of LB culture for obtaining the LLO. The degree of glycosylation was ~90% as judged from SDS-PAGE (Figure 1b) and NMR spectra. The glycan composition was validated by MS-MS (see Supporting InformationFigure S7).

NMR Data Acquisition and Processing. All spectra were recorded at 30 °C on either a Bruker DRX/Avance III 500 MHz (with CryoProbe), DRX/Avance III 600 MHz, or Avance 900 MHz spectrometer. Protein samples (1 mM) dissolved in 50 mM KH_2PO_4, pH 6.4, were used for all measurements. All acquired spectra were processed using either XWinNMR or TopSpin. ^{1}H, ^{13}C, ^{15}N chemical shifts of the protein were assigned by standard methods.[43] Details of all experiments are given in the Tables S1 and S2 in the Supporting Information. 3D ^{15}N-edited and 3D ^{13}C-edited NOESY spectra with mixing times of 150 ms were recorded. Slowly exchanging amide protons were detected by measuring a ^{15}N-HSQC 1 h after dissolving a lyophilized ^{15}N-labeled sample in D_2O. The N42 side chain amide group to which the glycan is attached was assigned by a glycine selective ^{15}N-HSQC experiment[44] and a CBCA(CO)NH spectrum[45] containing cross-peaks of N42 ^{13}Cα and ^{13}Cβ linked to the Asn side chain NH as well as to the R43 NH backbone. A 2D ^{13}C-filtered-filtered NOESY[22] and a 2D ^{15}N-filtered-filtered NOESY were recorded with 150 ms mixing time to obtain NOEs of the glycan. For the ^{13}C assignment of the carbohydrate, a natural abundance ^{13}C-HSQC was measured. Steady state ^{1}H{^{15}N}-NOE[46] spectra of the unmodified and glycosylated protein were recorded at 900 MHz.

Structure Calculation and Refinement of Glycosylated AcrA$^{61-210\Delta\Delta}$. An initial structure of the protein part alone was obtained by the programs ATNOS and CANDID[47,48] using the chemical shift assignments of the glycosylated protein and a 3D ^{15}N-edited NOESY and a 3D ^{13}C-edited NOESY. The structure of the glycan was generated and refined using CYANA 2.1[23] starting from 200 randomized structures and applying 125 distance constraints manually derived from the ^{13}C-filtered-filtered NOESY and the ^{15}N-filtered-filtered NOESY. Six categories of upper distance constraints were used: 6 Å, 5.5 Å, 5.0 Å, 4.0 Å, 3.2 Å, and 2.8 Å based on signal/to noise ratios of 2 to 5, 5 to 10, 10 to 20, 20 to 50, 50 to 80, and >80, respectively. NOEs involving Bac1 displayed an overall reduced signal-to-noise ratio and constraints were calibrated according to intraresidual peak intensities. Since the geminal protons H6 and H6′ were not stereospecifically assigned, floating stereospecific assignment is used during the CYANA algorithm.

The structure of the glycoprotein was generated and refined using CYANA 2.1 using the distance constraints list obtained from ATNOS and CANDID for the protein part and the glycan distance constraints mentioned above. In addition 25 protein−glycan constraints were extracted from a 3D ^{13}C-edited-filtered NOESY and grouped into the above-mentioned categories. Hydrogen bond constraints based on hydrogen exchange data and dihedral angle constraints generated by TALOS[49] were added during the refinement. The CYANA library was extended by an Asn residue lacking the Hδ2 for N42 and carbohydrate residues for the bacillosamine, unbranched GalNAc, branched GalNAc, and the Glc. Two linkers connecting the protein C-terminus with the C1 of bacillosamine and GalNAc6 with the C3 of Glc were added to obtain a single chain that is required by CYANA. For the attachment of the glycan to N42, the N−C1 distance and corresponding angles (no dihedrals) were derived from a representative glycoprotein crystal structure at 1.4 Å resolution containing β-D-GlcNAc-Asn linkages and hydrogen positions (PDB: 1LLF).[50] A refinement with AMBER was not applied to avoid bias of a force field. The Ramachandran plot of the folded domain (res. 2−33, 67−100) shows 75.3% of the residues in the most favorable regions, 22.8% in the additionally allowed regions, 1.2% in the generously allowed regions, and 0.7% in the disallowed regions. The corresponding values of the entire protein are 62.0%, 30.9%, 5.0%, and 2.1%, respectively.

Accession Code. The atomic coordinates, chemical shifts, and restraints used for the structure calculations are available from the Protein data Bank, www.pdb.org (PDB ID codes 2K33 and 2K32) and the BioMagRes Bank, www.bmrb.wisc.edu (BMRB ID codes 15737 and 15735).

Acknowledgment. We thank Robert J. Woods for kindly providing us a pdb template of the glycan, Reto Buser for recording MS-MS spectra, Thomas Peters, Jesus Angulo, and Torsten Herrmann for beneficial discussions, and Martin Frank and Thomas Lütteke for generating the GlycoMap of the bacillosamine linkage and beneficial suggestions regarding glycosidic torsion angles. This investigation was supported by a grant from the ETH Zürich (TH-Fonds Nr. 0-20126-03) to F.H.-T.A. and M.A. and grants from the Swiss National Science Foundation to M.A. and from the Structural Biology National Center of Competence in Research to F.H.-T.A.

Supporting Information Available: Detailed views of the ^{13}C-filtered-filtered 2D NOESY spectrum of glycosylated AcrA$^{61-210\Delta\Delta}$ (Figure S1). ^{13}C-HSQC spectra of the glycoprotein (Figure S2). Pulse sequence of the 2D ^{15}N-filtered-filtered NOESY (Figure S3). Glycosidic torsion angles φ and ψ of the glycoprotein ensemble (Figure S4). Planes of the 3D ^{13}C-edited-filtered NOESY spectrum (Figure S5). Potential hydrogen bonds observed in the NMR ensemble (Figure S6). MALDI-MS/MS spectrum of glycosylated AcrA$^{61-210\Delta\Delta}$ (Figure S7). NMR acquisition parameters (Tables 1 and 2). Supplementary Methods and Results. This material is available free of charge via the Internet at http://pubs.acs.org.

JA808682V

(43) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
(44) Schubert, M.; Smalla, M.; Schmieder, P.; Oschkinat, H. *J. Magn. Reson.* **1999**, *141*, 34–43.
(45) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 6291–6293.
(46) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972–9.
(47) Herrmann, T.; Guntert, P.; Wuthrich, K. *J Biomol NMR* **2002**, *24*, 171–89.
(48) Herrmann, T.; Guntert, P.; Wuthrich, K. *J. Mol. Biol.* **2002**, *319*, 209–27.
(49) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289–302.
(50) Pletnev, V.; Addlagatta, A.; Wawrzak, Z.; Duax, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 50–6.