Protein Modifications | Hot Paper

Deutsche Ausgabe: DOI: 10.1002/ange.201502093
Internationale Ausgabe: DOI: 10.1002/anie.201502093

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

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

Abstract: *Posttranslational modifications (PTMs) are an integral part of the majority of proteins. The characterization of structure and function of PTMs can be very challenging especially for glycans. Existing methods to analyze PTMs require complicated sample preparations and suffer from missing certain modifications, the inability to identify linkage types and thus chemical structure. We present a direct, robust, and simple NMR spectroscopy method for the detection and identification of PTMs in proteins. No isotope labeling is required, nor does the molecular weight of the studied protein limit the application. The method can directly detect modifications on intact proteins without sophisticated sample preparation. This approach is well suited for diagnostics of proteins derived from native organisms and for the quality control of biotechnologically produced therapeutic proteins.*

Posttranslational modifications (PTMs) expand the diversity of proteomes by two to three orders of magnitude.[1] Most mammalian proteins are posttranslationally modified[1] with glycosylation being the most abundant modification.[2] Roles of glycosylation encompass protein folding, stability, sorting, secretion, and molecular recognition,[3] while abnormalities in glycosylation patterns can lead to severe disorders, such as cancer[4] or congenital disorders.[5] Glycans are also involved in bacterial, fungal, and viral infections.[6] Further, PTMs are of central importance for therapeutic proteins, since they can directly influence their function but may also lead to severe side effects, for example, because of their immunogenicity. For example, inappropriate blood-group antigens on therapeutic proteins—mainly glycans—constitute a severe problem for biologics. To date, the characterization of intact glycans attached to glycoproteins requires a detour through chemical or enzymatic sample modifications; to our knowledge a direct method does not exist.

Typically PTMs in proteins are analyzed indirectly by a combination of mass spectrometry (MS) and enzymatic digestion or degradation, liquid chromatography, enrichment, and affinity separation.[7] For direct analysis, PTMs or parts thereof are released from the protein, in many cases derivatized, and analyzed according to their retention time in liquid chromatography and their molecular weight determined by MS. Another strategy of a MS-assisted approach for protein characterization—the bottom-up strategy—uses proteolytic digestion and analysis of the resulting peptides carrying modifications, for example, glycopeptides whose chemical structure can be determined by MS combined with other techniques. In this approach again enzymatic digestion, chemical cleavage, or chemical modifications are often used to analyze such glycopeptides. For example, the detection of mass shifts upon sequential treatment with glycosidases specific for certain glycans and linkages allows an indirect determination of the glycan structure. Alternatively, tandem mass spectrometry (MS^2)[8] involving different fragmentation techniques is a very efficient method for the characterization of glycopeptide composition and glycan structure. More recently, MS analysis of intact proteins containing PTMs is becoming popular owing to the ability to elucidate the compositions of different PTMs, such as different glycoforms.[8,9a] However, the exact stereochemistry of a glycan, including the type of glycosidic linkage, cannot be determined by MS. Therefore it is not possible to distinguish between the hexoses glucose (Glc), galactose (Gal), and mannose (Man), or the *N*-acetylhexoamines *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc) because of their identical molecular mass. Conclusively, MS-based analysis of proteins containing PTMs is a very sensitive and efficient method, but it suffers from substantial shortcomings, for example, the instabilities of the samples in the gas phase, non-ionizable components of the molecule, and inability to recognize the type of linkages and the stereochemistry within the modifications.[8,9]

Herein, we present a new method for fast detection and analysis of PTMs attached to the protein of interest without complicated sample preparation, based on simple NMR spectroscopy measurements. The NMR experiments are independent of the size of macromolecule, do not require any isotope labeling, and provide structural information on modifications complementary to mass spectrometry.

[*] Dr. M. Schubert,[§] [+] Dr. M. J. Walczak,[+] Prof. Dr. G. Wider
Institute of Molecular Biology and Biophysics
ETH Zürich, 8093 Zürich (Switzerland)
E-mail: mario.schubert@sbg.ac.at
michal.walczak@mol.biol.ethz.ch
Prof. Dr. M. Aebi
Institute of Microbiology
ETH Zürich, 8093 Zürich (Switzerland)
[§] Present address:
Department of Molecular Biology, University of Salzburg
Billrothstrasse 11, 5020 Salzburg (Austria)
[+] These authors contributed equally to this work.

[**] We thank Prof. Dr. Thomas Lütteke (University Giessen) for providing access to glycosciences.de when it was out of order. We are grateful to Prof. Dr. Dario Neri (ETH Zürich) for providing human Tumor Necrosis Factor, Dr. Andreas Nägeli (ETH Zürich) for bacterial glycoprotein and Dr. Therese Wollschlager (University of Salzburg) for her valuable comments on the manuscript. This work was supported by ETH Zürich (Grant ETH-23 10-2).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502093.

The basis for the NMR spectroscopic detection and analysis of PTMs are two-dimensional (2D) ^{1}H–^{13}C (at natural abundance of ^{13}C) and 2D ^{1}H–^{1}H correlation spectra that show distinct fingerprint patterns typical for each saccharide type and saccharide linkage. We investigated glycoproteins under denaturing conditions, eliminating size limitations for the proteins and simplifying the spectra. In entirely denatured proteins, only random coil chemical shifts corresponding to the 20 amino acids are observed (Figure S1 in the Supporting Information). Deviations from random-coil positions indicate modifications. Usually carbohydrates do not adopt structures of higher order, even less so in denatur-

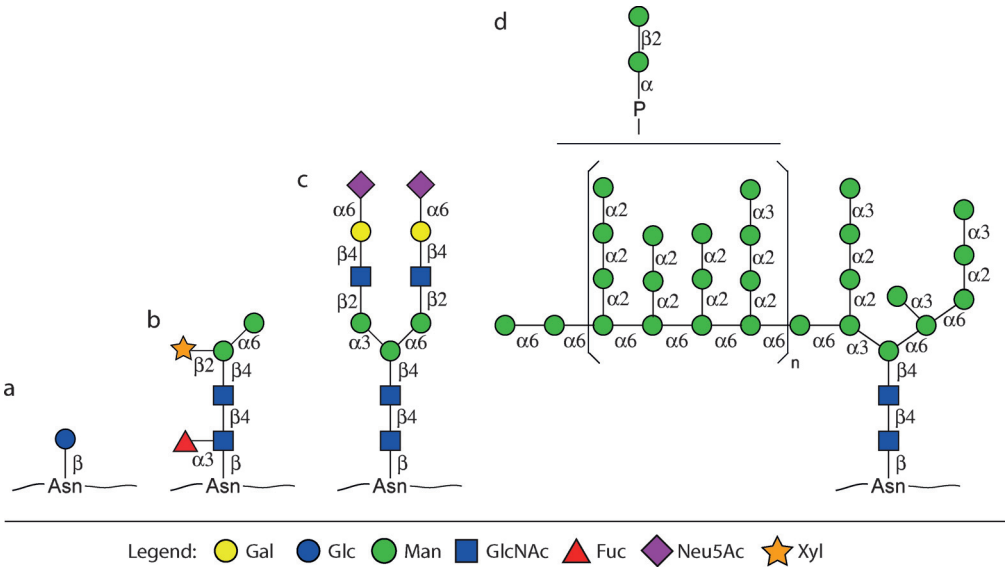

Figure 1. Schematic representations of the previously reported glycans of the proteins investigated. a) AtaC$^{1866–2428}$, b) bromelain, c) albumin, d) invertase (glycan model). α and β denote the type of linkages and P = phosphate. The legend (bottom) defines symbols for carbohydrates with the following abbreviations: Gal = galactose, Glc = glucose, Man = mannose, GlcNAc = *N*-acetylglucosamine, Fuc = fucose, Neu5Ac = *N*-acetylneuraminic acid (sialic acid), Xyl = xylose.

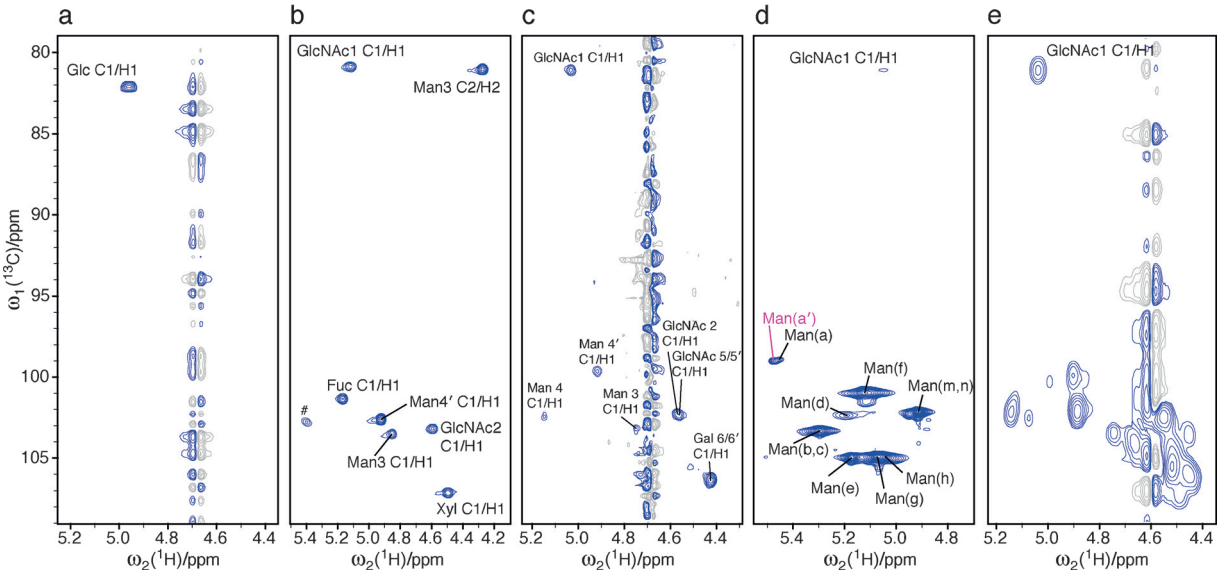

Figure 2. Glycosylation detected by [^{1}H, ^{13}C]-HSQC spectra in denatured bacterial, fungal, plant, and mammalian proteins. The five spectral regions each show the chemical shift range of the anomeric C1 carbon and of the attached proton of five different proteins. a) in vitro glycosylated bacterial protein AtaC$^{1866–2428}$ [10] displays the chemical shifts as of one glucose linked to an asparagine.[11b] b) Bromelain displays the previously reported chemical shifts;[15] an additional signal of unknown origin is indicated by a hash sign. c) Human serum albumin; the observed signals mainly correspond to previously reported signals of a complex biantennary structure with two sialic acid residues (nomenclature as previously reported[16]). d) Invertase from *S. cerevisae*; the observed signals mainly correspond to previously reported signals of isolated mannans from *Candida glabrata*;[14a] we used the same nomenclature for the mannoses as in Jawhara et al.[14a] The cross-peak with the magenta label originates from Manα1,2-Man*α-phosphate (* indicates saccharide corresponding to signal). e) Human TNFα expressed in HEK cells clearly shows a signal characteristic of *N*-glycans: the characteristic C1–H1 correlation of the first GlcNAc that is linked to Asn.

ing conditions which allows comparison of the observed chemical shifts with database entries. To avoid interference between the water and carbohydrate signals, we performed the NMR measurements in D_2O solution. The glycoproteins were denatured with 7 M urea which does not increase the ionic strength and thus not impair the sensitivity of the NMR experiments. As urea is highly hygroscopic and its deuteration is not straightforward, we used commercially available [D_4]-urea (Armar Chemicals) to obtain a high deuteration level.

To highlight the capabilities of our method we examined proteins from four different kingdoms of life comprising bacteria, fungi, plants, and animals using the in vitro glycosylated bacterial protein AtaC$^{1866-2428}$ [10], invertase (EC 3.2.1.26) from *Saccharomyces cerevisiae*, pineapple stem bromelain, and human serum albumin, respectively. In addition we examined the human Tumor Necrosis Factor (TNF) expressed in mammalian cells. This selection covers a wide range of glycans (Figure 1) and allows the usefulness of the method to be assessed. The [^{1}H, ^{13}C]-HSQC 2D NMR experiment provides a fingerprint of PTMs and in particular glycans. In the anomeric region (^{13}C chemical shifts between $\delta = 80$ and 110 ppm) each saccharide is represented by one signal as illustrated for the five investigated proteins in Figure 2. The C1–H1 cross-peaks in that region are typically well dispersed and not disturbed by protein signals. From the correlations in the [^{1}H, ^{13}C]-HSQC spectrum (Figure 2) *N*-glycans can be identified by a specific signal in the chemical shift range $\delta = 5.0$–5.2 ppm for ^{1}H and $\delta = 80$–82 ppm for ^{13}C (referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS))—the characteristic C1–H1 correlation for Asn-linked saccharides.[11] The anomeric carbons C1 for O-linkages typically resonate in the range $\delta = 98$–106 ppm and the attached proton between $\delta = 5.8$ and 4.3 ppm.[12] For the identification and assignment of glycan components a set of five 2D NMR experiments was used: a [^{1}H, ^{13}C]-HSQC, a [^{1}H, ^{13}C]-HMBC, a [^{1}H, ^{13}C]-HMQC-COSY, a [^{1}H, ^{1}H]-TOCSY with a short and one with a long mixing time. 2D TOCSY signals correlate the most disperse

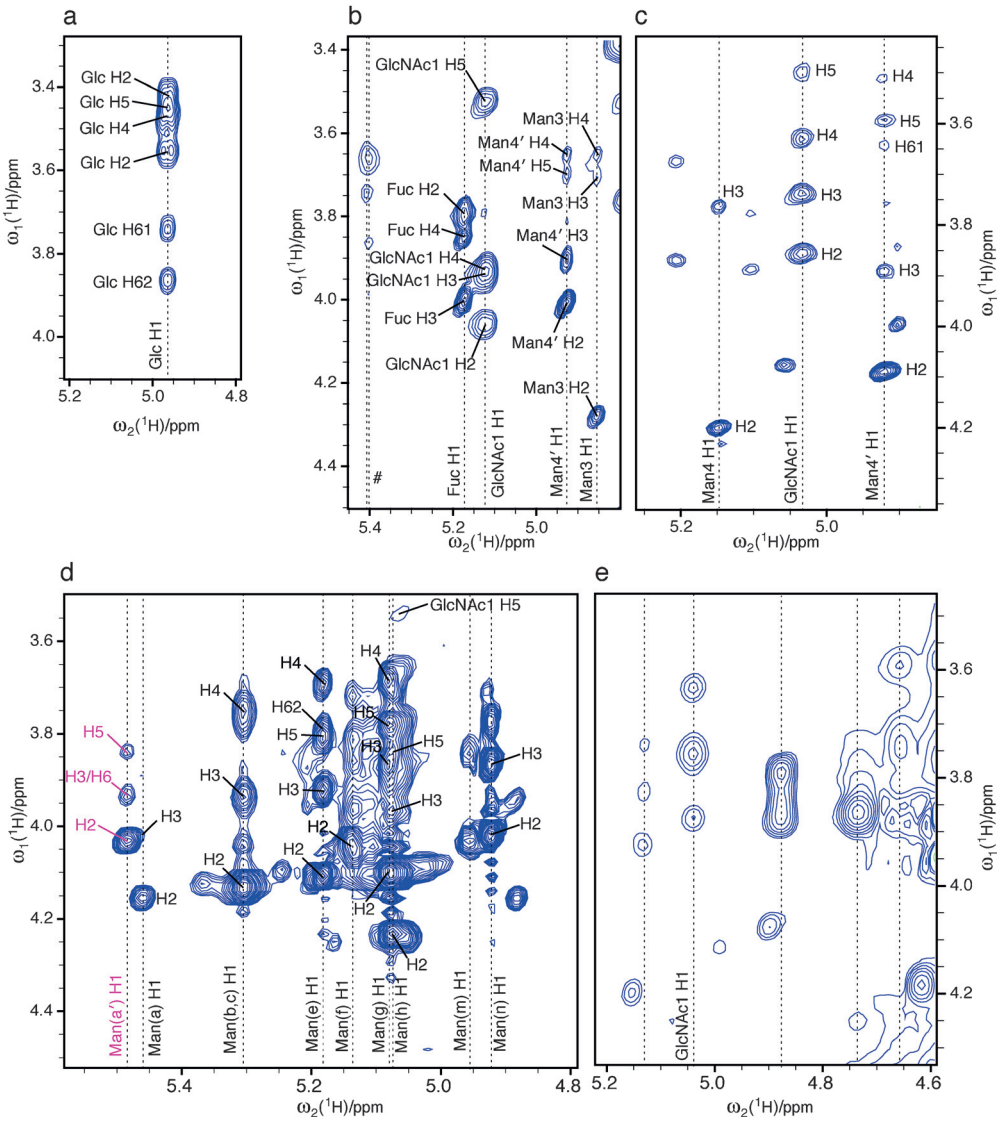

Figure 3. Glycosylation detected in denatured bacterial, fungal plant and mammalian proteins using 2D [^{1}H, ^{1}H]-TOCSY spectra (120 ms mixing time). The spectral regions show the chemical shift range of the anomeric H1 proton (ω_2) and correlated proton signals. a) in vitro glycosylated bacterial protein AtaC$^{1866-2428}$ [10] displays correlations to all chemical shifts of the glucose spin system. b) Bromelain displays the previously reported chemical shifts;[15] an additional resonance of unknown origin is indicated by a hash sign at $\omega_2 = 5.4$ ppm. c) Spectrum of human serum albumin, the observed signals mainly correspond to previously reported signals of a complex biantennary structure with two sialic acid residues;[16b] the same nomenclature as previously reported is used. d) Invertase from *S. cerevisae*; the observed signals mainly correspond to previously reported signals of isolated mannans from *Candida glabrata*;[14a] we used the same nomenclature for the mannoses as in Ref. [14a], the dispersion in the anomeric proton dimension allows detection of previously reported spin systems as well as novel ones; the to date from mannans not reported resonances labeled with Man(a′) are clearly different from Man(a); they can be assigned to Manα1,2-Man*α-phosphate based on the identical chemical shifts reported for a bacterial cell surface polysaccharide[20] whereas Man(a) corresponds to the expected Manβ1,2-Man*α-phosphate linked branches.[21] e) Human TNF expressed in HEK cells.

^{1}H chemical shifts of the anomeric protons with other ^{1}H resonances of the spin system illustrated for the five investigated proteins in Figure 3. A detailed description of the resonance assignment is given in the Supporting Information.

The approach works well for glycoproteins with glycans composed of possibly up to 20 saccharides. The gigantic *N*-glycans of yeasts, called mannans, having sizes of ten to several hundreds of kDa are particularly complex so that a step-by-step approach linking the spin systems did not work. In the case of the invertase of *S. cerevisiae* that contains over eight *N*-glycosylation sites[13] we therefore had to rely on previous studies of isolated mannans for which chemical shift lists were reported.[14] Although the most advanced NMR spectroscopic studies of mannans[14a] were not from *S. cerevisiae* but from related yeasts, such as *Candida glabrata*, our spectra look very similar and contain most of the described signals whose chemical shifts within the individual spin system match to ours. Homogenous glycosylation as in the case of AtaC and bromelain provides spectra with a single set of glycan resonance signals that can be interpreted in a straightforward manner. Moreover, also inhomogeneities such as Neu5Acα2,3 in addition to Neu5Acα2,6 linkages in albumin are clearly detectable. However, it is not possible with the presented method to assign these to a specific glycosylation site. Owing to the high dynamic range of state-of-the-art NMR spectrometers, less-abundant modifications can even be detected if the population is as low as 0.5 % as demonstrated for phosphorylated glucose in starch hydrolysates in which approximately one out of 200 Glc units are modified.[17]

The results obtained with our approach confirm the presence and structures of the N-glycans of the bacterial protein AtaC,[11] the pineapple stem bromelain,[15] and human serum albumin[16] (Figure 4). Our results of yeast invertase support the model of mannans deduced from *C. glabrata*.[14a] Intriguingly our NMR spectroscopic method generated new information on the glycosylation of the well-characterized proteins invertase and albumin: in addition to previously reported Manαβ1,2-Man-phosphate linkages in mannans[14a] (Figure 1 d), we observed Manα1,2-Man-phosphate linkages in the invertase spectra (Figure 3 d and Figure 4 d). In the case of human serum albumin previously unreported signals could be assigned to α2,3-linked sialic acid (Figure S2) based on chemical shifts which are identical to previously reported resonance signals of sialic acids α2,3-linked to β-D-galactose.[18] Human Tumor Necrosis Factor (TNF), known to be *O*-glycosylated,[19] shows also *N*-glycosylation (Figure 2 e and Figure 3 e) which has not been reported so far. The presence of the *N*-glycan is further supported by digestion of TNF with the *N*-glycan-specific enzyme, PNGase F (Figure S3).

A limitation of the presented method is the amount of sample required. We used protein quantities between 0.15 μmol and 1.8 μmol which is several orders of magnitude more sample than required for MS. However, our approach is not destructive and the sample can be recovered. For certain applications such as quality control of therapeutic proteins the quantities required are not limiting. The herein described NMR method for the characterization of PTMs has great potential to complement other approaches, such as mass spectrometry: NMR spectroscopy can clearly detect certain glycan fragments with a defined stereochemistry and linkage type, but cannot assign the modification sites on the protein sequence. On the other hand, MS provides information on the

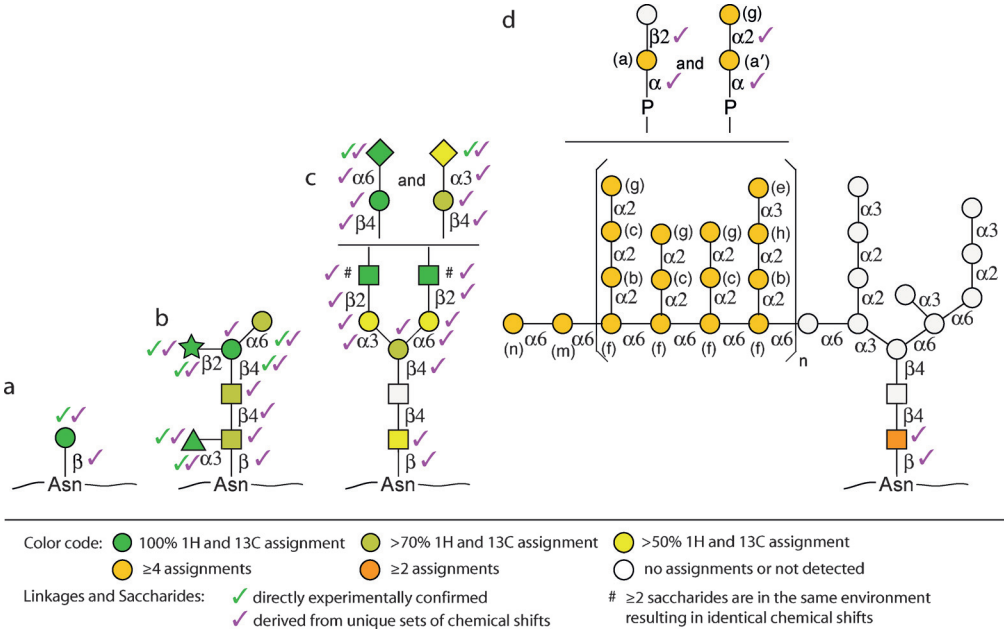

Figure 4. Schematic representations of the NMR spectroscopy detected glycans of the analyzed proteins. a) AtaC$^{1866-2428}$, b) bromelain, c) albumin, and d) invertase (mannan model). The color code reports on the assignment completeness for certain saccharide types, the symbols are identical to Figure 1. The saccharide types were either established from characteristic spin-system features, such as, methyl or methylene groups (green check mark) or/and characteristic chemical shifts (magenta check mark). Connections were either confirmed by H1–CX long range correlations ($^3J_{H1CX}$) across the glycosidic linkages (green check mark) or derived from the unique chemical shift combinations of a certain spin system (magenta check mark). The letters in parenthesis within the mannan structure correspond to the spin system labels previously reported[14a] and used in the invertase spectra.

exact location of the modification on the protein, the exact mass of the intact modification, and to some extent glycan structure but does not reveal the exact stereochemistry or the type of linkage.

In summary, our approach to analyze intact proteins allows detection of PTMs, represented in this case by glycans, their composition, and their types of linkages irrespective of glycan heterogeneity and the molecular weight of a protein. In contrast to other NMR methods, the macromolecules do not require isotope labeling and can originate from natural sources. The herein described direct NMR method is ideally suited for quality control of therapeutic proteins, for example, to exclude any immunogenicity arising from blood-group antigens. This method has great potential for the detection of posttranslational modifications other than glycosylation, for example, methylation,[22] acetylation, phosphorylation,[23] propionylation, butyrylation,[2] or lipidations.

Experimental Section

Sample preparation: A detailed description is given in the Supporting Information. In brief, all proteins were dissolved in D_2O with 50 mM [D_{11}]Tris (Sigma–Aldrich, cat. no. 486248) and lyophilized. The proteins were re-suspended in D_2O and twice lyophilized. For the NMR spectroscopy measurements all lyophilized proteins were dissolved in 7 M [D_4]Urea (98% D, Armar Chemicals, cat. no. 049500.3041), 2 mM DTT ([D_{10}]DTT Cambridge Isotope, cat. no. DLM-2622), pD 5.5 (adjusted with [D_4]acetic acid, Sigma–Aldrich cat. no. 233315) with a concentration as indicated in Table S1.

A full description of the NMR spectroscopy and chemical shift assignments is given in the Supporting Information.

Keywords: glycomics · glycoproteins · NMR spectroscopy · protein modifications · proteomics

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 7096–7100
Angew. Chem. **2015**, *127*, 7202–7206

[1] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372; *Angew. Chem.* **2005**, *117*, 7508–7539.
[2] F.-X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M.-K. Yoon, R. Kriwacki, I. Landrieu, G. Lippens, P. Selenko, *J. Biomol. NMR* **2012**, *54*, 217–236.
[3] A. Helenius, M. Aebi, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049.
[4] a) S. Hakomori, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10231–10233; b) T. Y. Chou, G. W. Hart, *Adv. Exp. Med. Biol.* **2001**, *491*, 413–418.
[5] H. H. Freeze, M. Aebi, *Curr. Opin. Struct. Biol.* **2005**, *15*, 490–498.
[6] M. Matrosovich, G. Herrler, H. D. Klenk, *Top. Curr. Chem.* **2013**, *337*, 9023–9030.
[7] a) M. Černý, J. Skalák, H. Cerna, B. Brzobohatý, *J. Proteomics* **2013**, *92*, 2–27; b) A. R. Farley, A. J. Link, *Methods Enzymol.* **2009**, *463*, 725–763.
[8] M. R. Larsen, M. B. Trelle, T. E. Thingholm, O. N. Jensen, *BioTechniques* **2006**, *40*, 790–797.
[9] a) A. D. Catherman, O. S. Skinner, N. L. Kelleher, *Biochem. Biophys. Res. Commun.* **2014**, *445*, 683–693; b) A. Silva, R. Vitorino, M. R. M. Domingues, C. M. Spickett, P. Domingues, *Free Radical Biol. Med.* **2013**, *65*, 925–941.
[10] 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.
[11] a) V. Slynko, M. Schubert, S. Numao, M. Kowarik, M. Aebi, F. H. T. Allain, *J. Am. Chem. Soc.* **2009**, *131*, 1274–1281; b) F. Schwarz, Y.-Y. Fan, M. Schubert, M. Aebi, *J. Biol. Chem.* **2011**, *286*, 35267–35274.
[12] a) O. Zerbe, S. Jurt, *Applied NMR Spectroscopy for Chemists and Life Scientists*, Wiley, Hobkoken, **2013**; b) J. Ø. Duus, C. H. Gotfredsen, K. Bock, *Chem. Rev.* **2000**, *100*, 4589–4614.
[13] V. Reddy, R. Johnson, K. Biemann, R. Williams, F. Ziegler, R. Trimble, F. Maley, *J. Biol. Chem.* **1988**, *263*, 6978–6985.
[14] a) S. Jawhara, E. Mogensen, F. Maggiotto, C. Fradin, A. Sarazin, L. Dubuquoy, E. Maes, Y. Guérardel, G. Janbon, D. Poulain, *J. Biol. Chem.* **2012**, *287*, 11313–11324; b) E. Maes, C. Mille, X. Trivelli, G. Janbon, D. Poulain, Y. Guérardel, *J. Biochem.* **2009**, *145*, 413–419; c) N. Shibata, A. Suzuki, H. Kobayashi, Y. Okawa, *Biochem. J.* **2007**, *404*, 365–372; d) E. Vinogradov, B. O. Petersen, J. Ø. Duus, *Carbohydr. Res.* **2000**, *325*, 216–221.
[15] 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.
[16] a) Y. Sakamoto, K. Kitamura, J. Madison, S. Watkins, C.-B. Laurell, M. Nomura, T. Higashiyama, F. W. Putnam, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1995**, *1252*, 209–216; b) J.-M. Wieruszeski, J.-C. Michalski, J. Montreuil, G. Strecker, *Glycoconjugate J.* **1989**, *6*, 183–194.
[17] D. Santelia, O. Kötting, D. Seung, M. Schubert, M. Thalmann, S. Bischof, D. A. Meekins, A. Lutz, N. Patron, M. S. Gentry, *Plant Cell* **2011**, *23*, 4096–4111.
[18] D. Machytka, R. A. Klein, H. Egge, *Carbohydr. Res.* **1994**, *254*, 289–294.
[19] R. Takakura-Yamamoto, S. Yamamoto, S. Fukuda, M. Kurimoto, *Eur. J. Biochem.* **1996**, *235*, 431–437.
[20] N. Paramonov, M. Rangarajan, A. Hashim, A. Gallagher, J. Aduse-Opoku, J. M. Slaney, E. Hounsell, M. A. Curtis, *Mol. Microbiol.* **2005**, *58*, 847–863.
[21] J. Rahkila, F. S. Ekholm, R. Panchadhayee, A. Ardá, F. J. Cañada, J. Jiménez-Barbero, R. Leino, *Carbohydr. Res.* **2014**, *383*, 58–68.
[22] F.-X. Theillet, S. Liokatis, J. O. Jost, B. Bekei, H. M. Rose, A. Binolfi, D. Schwarzer, P. Selenko, *J. Am. Chem. Soc.* **2012**, *134*, 7616–7619.
[23] S. Liokatis, A. Dose, D. Schwarzer, P. Selenko, *J. Am. Chem. Soc.* **2010**, *132*, 14704–14705.

Received: March 5, 2015
Published online: April 29, 2015

