Structure elucidation of glycoprotein glycans and of polysaccharides by NMR spectroscopy

B.R. Leeftang *, E.J. Faber, P. Erbel, J.F.G. Vliegthart

Bijvoet Center for Biomolecular Research, Utrecht University, P.O. Box 80.075, NL-3508 TB Utrecht, The Netherlands

Received 8 March 1999; received in revised form 5 July 1999; accepted 6 July 1999

Abstract

The applicability of ^1H -NMR spectroscopy for the determination of the primary and tertiary structure of carbohydrate-containing molecules is demonstrated. For classes of known compounds the characterization can be based on chemical shifts observed in 1D NMR spectra with or without the aid of a computer database. For more complex structure determinations 2D NMR techniques are required. Here the application of 2D NMR is demonstrated for the primary structure determination of two bacterial exopolysaccharides, for the spatial structure determination of a disaccharide and a glycoprotein hormone. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Glycoproteins; Exopolysaccharides; NMR spectroscopy; Structure determination

1. Introduction

Carbohydrates are widespread in nature. They occur in plants primarily as polysaccharides such as cellulose, pectins, xylans, and starch. Many bacteria are encapsulated with polysaccharide networks, whereas others secrete polysaccharides, the latter often having strongly gelling properties. Glycans also play an important role in higher animal organisms. Often glycans are involved in recognition processes within the cell or between cells. Although carbohydrates occur as individual molecules, in many cases the carbohydrate chain is covalently bound to a lipid, a peptide or a

protein, thus forming glycolipids, glycopeptides or glycoproteins, respectively.

Several stages of characterization of these glycans can be identified. First of all the primary structure (i.e. monosaccharide composition sequence and substitution pattern) needs to be determined by chemical techniques in combination with ^1H -NMR spectroscopy. In cases where insight into the structure activity relation is sought after, the spatial structure and possibly the dynamics of this spatial structure need to be established.

The primary parameters which NMR spectroscopy yields, are the chemical shifts, i.e. the position in the spectrum where a certain spin resonates, and the J-coupling. The chemical shift is a property that describes the relative magnetic

* Corresponding author. Tel.: +31-30-2533498; fax: +31-30-2540980.

E-mail address: b.r.leeftang@chem.uu.nl (B.R. Leeftang)

field localized at the nucleus. The local magnetic field depends on the attenuation of external magnetic field of the spectrometer caused by the electronic density near the nucleus. The chemical shift is a very informative parameter for the analysis of the primary structure of glycans.

The 1D ^1H -NMR spectra of glycans in $^2\text{H}_2\text{O}$ solutions show that two groups of signals can be distinguished. The so-called bulk signal contains mainly the non-anomeric protons, present in the narrow spectral range from 3.2 to 3.9 ppm. Based on 1D ^1H -NMR spectra, the proton signal assignment in this spectral region is hampered by signal overlap. Outside the bulk area several well-resolved signals appear, that are referred to as 'structural reporter groups'. It has become apparent that it is not necessary to assign all ^1H -NMR signals for a full characterization in terms of a primary structure. Prominent structural reporter groups are summarized in Table 1.

The various chemical shift and coupling patterns are translated into structural information based on a comparison with patterns in a library of relevant reference compounds (Vliegthart et al., 1983; Kamerling and Vliegthart, 1992). The comparison of NMR data for many closely related glycans resulted in empirical rules to correlate chemical shift values with carbohydrate structures. Obviously, the application of this approach requires accurate calibration of the experimental conditions like sample temperature, solvent, and pH.

The large amount of NMR data collected for quite an array of compounds, enabled the development of an NMR database for the computer

assisted identification of carbohydrates (Van Kuik and Vliegthart, 1993). The database, designated SUGABASE, combines ^1H - and ^{13}C -NMR chemical shifts of carbohydrates with structures and literature references as present in the complex carbohydrate structure database (CCSD; Doubet et al., 1989). Starting from an input of chemical shift values, the search program produces a list of partially or completely matching structures. The SUGABASE program is available for UNIX and MS-windows. Also an interface to the database is available on the Internet: <http://www.boc.chem.uu.nl/sugabase/databases.html>.

The SUGABASE program on the Internet has a JAVA based structure builder tool that enables the user to construct a primary structure fragment, which can be passed to the database search program together with the experimentally observed chemical shift values. A snap shot of the structure builder is given in Fig. 1. The structural reporter group approach can be applied manually or in a computer-aided fashion. It will be evident that the former requires training or specialist skills on the part of the spectroscopist, whereas the latter approach requires less of these skills.

For certain classes of glycans that have been well-studied the chemical shifts of a few well-resolved protons in the NMR spectra are indicative for the primary structure. In other cases more advanced (2D and 3D) NMR techniques need to be performed, first to assign the various signals and secondly to establish the primary structure. NMR spectroscopy has also proven to be a powerful tool for determining molecular conformations.

Here, the application of NMR spectroscopy to the determination of the primary and tertiary structure of glycans from glycoprotein and polysaccharides will be demonstrated.

1.1. Primary structure analysis of bacterial exopolysaccharides

Bacterial exopolysaccharides (EPS) are widely applied in the food industry as gelling and thickening agents. A strong interest has arisen in exopolysaccharides produced by lactic acid bacteria, because these strains are food grade (generally

Table 1
Structural reporter groups for glycoprotein derived glycans

•	Anomeric protons
•	Alkyl and acyl substituents like methyl, acetyl, glycolyl
•	Sialic acid H3 atoms
•	GalNac-ol H2, H3, H4 and H5, Man-H2 and Gal-H4 atoms
•	Protons shifted out of the bulk-region due to glycosylation shifts or under influence of substituents, such as sulfate, phosphate, and acyl groups
•	6-Deoxy sugar H6 atoms

Fig. 1. Snap shot of the primary structure element builder tool of the WWW interface to the SUGABASE carbohydrate NMR database.

recognized as safe (GRAS)). Recently, our lab published the primary structure determination of the EPSs of two *Streptococcus thermophilus* strains, Rs and Sts (Faber et al., 1998).

1.2. Conformational analysis of a disaccharide

As an introductory example of the potential of NMR spectroscopy in conformational analysis, the conformation of β -Me-cellobioside will be described.

1.3. Spatial structure determination of a glycoprotein hormone subunit

The spatial structure of the α -subunit of human chorionic gonadotropin (α hCG) has been determined using solution NMR spectroscopy. Attention has been focused on the potential influence of glycosylation on the protein structure. An X-ray structure of this glycoprotein is known (Lapthorn et al., 1994; Wu et al., 1994).

2. Materials and methods

Three examples are given to demonstrate the applicability of NMR spectroscopy on glycans to biotechnology. All the experimental information with respect to the three examples have been published elsewhere. Here the experimental description is limited to the strategic essentials. For details the reader is referred to the original publications. All NMR experiments were performed on Bruker AMX spectrometers at 500 or 600 MHz. Processing of the NMR spectra was either done using Bruker xwinnmr software or using the in house TRITON software

2.1. Primary structure analysis of bacterial exopolysaccharides

The exopolysaccharides produced by two strains of *S. thermophilus* lactic acid bacteria, referred to as Rs and Sts, were dissolved in 500 μ l $^2\text{H}_2\text{O}$. NMR spectra were recorded at an elevated temperature of 353 K in order to lower the viscosity and thus to narrow the line width of the NMR

signals. Besides the 1D NMR spectrum, homonuclear 2D TOCSY and NOESY spectra were recorded. Furthermore, heteronuclear (^1H , ^{13}C) HSQC, HSQC-TOCSY, and HSQC-NOESY spectra were recorded. The monosaccharide composition and the information on the substitution patterns were obtained using chemical and chromatographic methods (Gerwig et al., 1978, 1979).

2.2. Conformational analysis of a disaccharide

β -Me(-d3)-cellobioside was made from cellobiose and methanol(-d4) using a conventional synthesis route. The sample (1 mg) was dissolved in 500 μ l $^2\text{H}_2\text{O}$. A series of ROESY experiments was recorded with mixing times between 75 and 400 ms. Furthermore, a GROMOS Molecular Dynamics simulation was done. The outcome of both was analyzed using the CROSREL program (Leeftang and Kroon-Batenburg, 1992; Kroon-Batenburg et al., 1993). This program calculates NOESY or ROESY peak intensities from molecular models, such as a MD trajectory.

2.3. Spatial structure determination of a glycoprotein hormone subunit

NMR studies were carried out on native α hCG (glycosylated at Asn52 and Asn78), as well as on partially deglycosylated (glycosylated at Asn78) and deglycosylated α hCG up to the first GlcNAc residues at Asn52 and Asn78. 2D TOCSY and NOESY spectra were recorded at an elevated temperature of 328 K with various mixing times. From the NOESY data, models of the protein structure were generated using an adapted version of the X-PLOR program (Nilges et al., 1988, 1991; Kuszewski et al., 1992). The modification to the program was needed because native X-PLOR does not know how to deal with the complex disulfide bonding topology and with intact glycoproteins. Solution state NMR structure information has been published by our group (De Beer et al., 1996; Thijssen-van Zuylen et al., 1998; Erbel et al., 1999).

Table 2

Monosaccharide composition, branching information and primary structure of the repeating unit of the exopolysaccharides of *S. thermophilus* strains, Rs and Sts^a

Monosaccharide composition		Branching information	
Gal	5	3,4-Rha	0.9
Rha	2	2-Rha	1.1
		2,3,4,6-Gal	0.9
		2,3,6-Gal	1.0
		2,4,6-Gal	2.2
Primary structure			
β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp			
1			
↓			
4			
$\rightarrow$ 3)- α -D-Galp-(1 $\rightarrow$ 3)- α -L-Rhap-(1 $\rightarrow$ 2)- α -L-Rhap-(1 $\rightarrow$ 2)			
- α -D-Galp-(1 $\rightarrow$ 3)- α -D-Galp-(1 $\rightarrow$			

^a The numbers preceding the monosaccharide in the column for branching information identify the positions of free hydroxyl groups.

3. Results and discussion

3.1. Primary structure analysis of bacterial exopolysaccharides

The first noteworthy result of the 1D NMR spectra of the Rs and Sts exopolysaccharide samples was that the spectra were virtually identical. This was indeed remarkable because the physical properties (viscosity) differ considerably. The difference in molecular mass or the degree of polymerization turned out to be decisive in this respect. The monosaccharide compositions and the substitution patterns, as shown in Table 2, are also identical. The homonuclear 2D spectra could not be analyzed in sufficient detail to obtain the linkage information, because of signal overlap in the bulk area. Using heteronuclear HSQC, HSQC-TOCSY, and HSQC-NOESY spectra the overlap problem could be solved, due to the larger chemical shift dispersion in the ¹³C dimension. Combination of the chemically obtained information with the linkage data from the HSQC-NOESY spectra the following identical primary structure repeating unit were determined for both samples. From the HSQC the one-bond heteronu-

clear scalar coupling constants were obtained that were used to characterize the anomeric configuration of the various linkages.

NMR spectroscopy could be used here to characterize the primary structure of two exopolysaccharides from different origins. The TOCSY and HSQC spectra were above all needed to be able to assign the signals in the NMR spectra. The NOESY and especially the HSQC-NOESY spectra, that have cross peaks depending on short inter proton distances could be used to establish the exact linkage information for the polysaccharide repeating unit. NOESY spectra are not unquestionably reliable to obtain the linkage information. The observation of a NOESY peak only identifies a short distance, and is not the unambiguous proof for the chemical linkage. HMBC spectra are also sometimes used to obtain the linkage position information via a through-bond interaction. Here the combination of chemical and qualitative spectroscopic data was sufficient for the primary structure determination.

3.2. Conformational analysis of a disaccharide

In the framework of a research program aimed at a better understanding of cellulose structure on a molecular level the choice was made to analyze the crystallographic repeating unit, cellobiose. The β -Me-(d3) glycoside of cellobiose was used to simplify the NMR spectra, by excluding the α -glycosidic form that does not occur in cellulose. In Fig. 2, a model including relevant atom labels is depicted. The ROESY cross-peak intensities were compared with calculated intensities from MD simulations. The CROSREL program takes the ROESY offset dependence and the direct TOCSY transfer into account. These effects make the quantitative analysis of the ROESY data more difficult than of NOESY data. Special attention has been given to glycosidic bond conformations and to the rotamer population distributions of the exocyclic hydroxymethyl groups of both glucose moieties. Although calculations indicated that up to four potential energy minimal conformations could be possible, each characterized by typical short inter-proton distances, only one conformation is observed with NMR spectroscopy. This is

the same conformation that is found in the solid state as determined by fiber X-ray diffraction. With respect to the hydroxymethyl group a clear difference between the solid and solvated state is being observed. In native cellulose the hydroxymethyl group is oriented towards the next glucose residue in the polymer. This conformation is referred to as the *tg* conformation. From the analysis of the J-couplings between the H5 and H6 atoms it can be concluded that in solution the *tg* conformation hardly occurs.

3.3. Spatial structure determination of α glycoprotein hormone subunit

A great challenge in applying NMR spectroscopy in structure determination still is the structure determination of proteins, especially for glycoproteins, because only in rare cases does one have suitably isotope-labelled compounds. In comparison to the analysis of the disaccharide, the number of NOE cross peaks in protein spectra is considerably larger. Because of this, the spectral

assignment process is much more tedious. Fortunately the chemical shift dispersion in protein spectra is much larger than in carbohydrate spectra.

Once the NMR spectra for the α hCG samples were assigned as completely as possible, lists of NOE peaks were made. The intensities were classified qualitatively as 'strong', 'medium', 'weak', and 'very weak'. This data was the input for the modified X-PLOR program. It would be beyond the scope of this article to go into the technical details of this program and of the modification. X-PLOR generates a series of structure models that fit within 0.5 Å deviation from the NMR distance information (NOEs). As striking feature of these generated structures is that some regions are very well-defined, whereas one region is poorly defined. This is depicted in Fig. 3, where a single structure is shown. Backbone atoms are displayed for well-defined regions, whereas a ribbon is displayed for the flexible region. This non-structured region interacts with the β -subunit of hCG in the heterodimer. We have evidence that

Fig. 2. Molecular model of β -Me-cellobioside taken from a MD simulation trajectory. Hydrogen atom label conforming the IUPAC rules for carbohydrates are shown. The conformation of both hydroxymethyl groups is the *gt* conformation. In the *tg*, *gg*, *gt* naming convention both characters refer to a torsional angle being either gauche (*g*) or trans (*t*). The first character relates to the O6(')-C6(')-C5(')-O5(') angle, whereas the second relates to the O6(')-C6(')-C5(')-C4(') angle.

Fig. 3. Structure of α hCG deglycosylated beyond the first Glc-Nac, as determined from NMR spectroscopy using the X-PLOR program. The well-defined regions of the structure are displayed using the actual backbone atom positions, whereas for the flexible region a ribbon was used.

the absence of this subunit allows for the increased flexibility of this peptide region. Currently our group is occupied establishing the influence of the glycosylation on the protein structure.

4. Conclusions

NMR spectroscopy can be used in various stages in the structure determination of biomolecules. This was illustrated in three examples. In all examples several experiments need to be done to be able to assign the peaks in the NMR spectra to the different atoms in the molecules. The analysis of distance information can be done qualitatively, semi-quantitatively or quantitatively.

In the exopolysaccharide qualitative interpretation provided the information to determine the primary structure repeating unit. For the characterization of the disaccharide conformation the J-coupling and ROESY cross-peak intensities have been analyzed quantitatively to obtain sufficiently detailed conformational information.

In this last example it was shown that NOESY NMR data provided in a qualitative way are sufficient to determine the structure of a (glyco)-protein.

References

- De Beer, T., Van Zuylen, C.W.E.M., Leeftang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P., Vliegthart, J.F.G., 1996. NMR studies of the free α -subunit of human chorionic gonadotropin; structural influences of *N*-glycosylation and the β -subunit on the conformation of the α -subunit. *Eur. J. Biochem.* 241, 229–242.
- Doubet, S., Bock, K., Smith, D., Darvill, A., Albersheim, P., 1989. *Trends Biochem. Sci.* 14, 475–477.
- Erbel, P.J.A., Karimi-Nejad, Y., De Beer, T., Boelens, R., Kamerling, J.P., Vliegthart, J.F.G., 1999. Solution structure of the α -subunit of human chorionic gonadotropin. *Eur. J. Biochem.* 260.
- Faber, E.J., Zoon, P., Kamerling, J.P., Vliegthart, J.F.G., 1998. The exopolysaccharides produced by *Streptococcus thermophilus* Rs and Sts have the same repeating unit but differ in viscosity of their milk cultures. *Carbohydr. Res.* 310, 269–276.
- Gerwig, G.J., Kamerling, J.P., Vliegthart, J.F.G., 1978. *Carbohydr. Res.* 62, 349–357.
- Gerwig, G.J., Kamerling, J.P., Vliegthart, J.F.G., 1979. *Carbohydr. Res.* 77, 1–7.
- Kamerling, J.P., Vliegthart, J.F.G., 1992. High-resolution ^1H -nuclear magnetic resonance spectroscopy of oligosaccharide-alditols released from mucin-type *O*-glycoproteins. *Biol. Magn. Reson.* 10, 1–194.
- Kroon-Batenburg, L.M.J., Kroon, J., Leeftang, B.R., Vliegthart, J.F.G., 1993. Conformational analysis of methyl

- β -cellobioside by ROESY NMR spectroscopy and MD simulations in combination with the CROSREL method. *Carbohydr. Res.* 245, 21–42.
- Kuszewski, J., Nilges, M., Brünger, A.T., 1992. Sampling and efficiency of metric matrix distance geometry: a novel partial metrication algorithm. *J. Biomol. NMR* 2, 33–56.
- Lapthorn, A.J., Harris, D.C., Littlejohn, A., Lustbader, J.W., Canfield, R.E., Machin, K.J., Morgan, F.J., Isaacs, N.W., 1994. Crystal structure of human chorionic gonadotropin. *Nature* 369, 455–461.
- Leeftang, B.R., Kroon-Batenburg, L.M.J., 1992. CROSREL: full relaxation matrix analysis for NOESY and ROESY NMR spectroscopy. *J. Biomol. NMR* 2, 495–518.
- Nilges, M., Clore, G.M., Gronenborn, A.M., 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* 229, 317–324.
- Nilges, M., Kuszewski, J., Brünger, A.T., 1991. Computational aspects of the study of biological macromolecule by NMR. In: Hochm, J.C. (Ed.). Plenum Press, New York.
- Thijssen-van Zuylen, C.W.E.M., De Beer, T., Leeftang, B.R., Boelens, R., Kaptein, R., Kamerling, J.P., Vliegthart, J.F.G., 1998. Mobilities of the inner three core residues and the Man(α 1–6) branch of the glycan at Asn78 of the α -subunit of human chorionic gonadotropin are restricted by the protein. *Biochemistry* 37, 1933–1940.
- Van Kuik, J.A., Vliegthart, J.F.G., 1993. Databases of complex carbohydrates. *Trends Food Sci. Technol.* 4, 73–77.
- Vliegthart, J.F.G., Dorland, L., Van Halbeek, H., 1983. High-resolution ^1H -nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins. *Adv. Carbohydr. Chem. Biochem.* 41, 209–374.
- Wu, H., Lustbader, J.W., Liu, Y., Canfield, R.E., Hendrickson, W.A., 1994. Structure of human chorionic gonadotropin at 2.6 Å resolution from AMD analysis of the selenomethionyl protein. *Structure* 2, 545–558.