STUDIES ON THE STRUCTURE OF CARBOHYDRATE CHAINS OF GLYCOPROTEINS

J.F.G. Vliegenthart, K. Hård, P. de Waard and J.P. Kamerling

Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, P.O. Box 80.075, NL-3508 TB Utrecht, The Netherlands

SUMMARY

To obtain detailed information on the primary structures of the carbohydrate chains of glycoproteins, it is still necessary to degrade the glycoprotein to partial structures each representing only one glycosylation position. The partial structures suitable for analysis, can be oligosaccharides, oligosaccharide-alditols or glycopeptides. In our approach for the analysis of *N,O*-glycoproteins we cleave first the N-linked chains with the aid of PNGase F. The extent of this reaction has to be checked carefully, in order to make sure that in case the reaction does not lead to complete removal of the carbohydrate chains, the chains that are still linked to the protein do not have a specific structure. For this purpose several methods can be applied. The pool of *N*-glycans is separated from the *O*-glycoprotein, which is then isolated and subjected to alkaline borohydride treatment. The pools corresponding to the N- and O-linked chains, respectively, are fractionated to homogeneous compounds as far as possible. Analysis of the compounds is carried out by 500-MHz ^{1}H-NMR spectroscopy. The occurrence of non-carbohydrate substituents like alkyl, acyl, sulfate and phosphate groups may give rise to serious complications. Often the contents of substituents are far below molar equivalents in a compound, thereby enhancing the (micro)heterogeneity. It should be noted that information about the type and position of the non-carbohydrate substituents can be obtained by 1D- or 2D-^{1}H-NMR spectroscopy. For the analysis of phosphate-containing oligosaccharides we developed a ^{1}H{^{31}P} relayed spin-echo difference spectroscopy technique to characterize the residue to which the phosphate is attached and to detect the C-atom that is substituted.

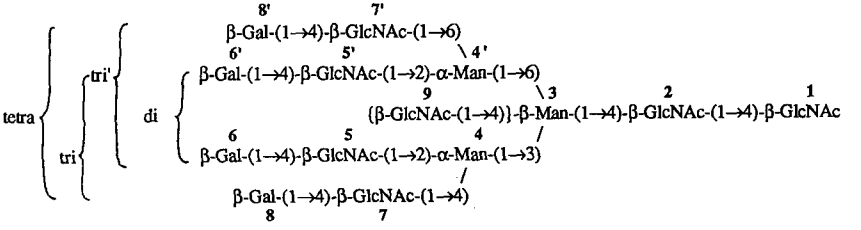
By means of this analytical procedure we were able to determine the structure of glycoprotein-derived carbohydrate chains for quite a number of glycoproteins.

1 INTRODUCTION

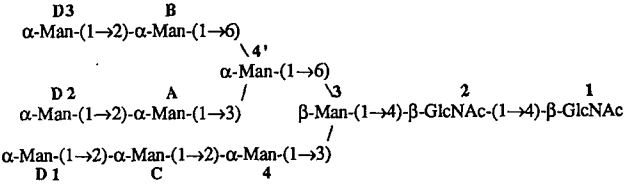
Glycoproteins are biopolymers which consist of a polypeptide backbone, bearing one or more covalently linked carbohydrate chains [1]. The carbohydrate chains of glycoproteins can be classified as *N* - or *O*-glycosidic, depending on the type of linkage to the protein. The N-linked oligosaccharides are bound via the reducing end sugar *N*-acetylglucosamine to the amide nitrogen of Asn. In general, the core element of these chains comprises a trimannosyl-*N*,*N*'-diacetylchitobiose structure:
α-Man-(1→3)-[α-Man-(1→6)-]β-Man-(1→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N-. This core region can be extended in a number of ways, giving rise to several types of N-linked carbohydrate chains. In Fig. 1 examples of the main types are given:

Fig. 1. Examples of the main types of N-linked carbohydrate chains of glycoproteins.

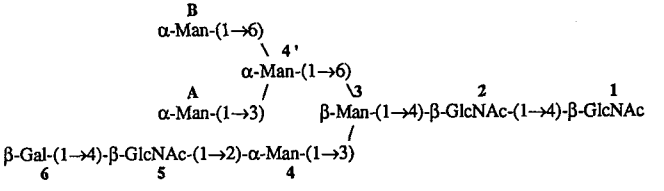
N-acetyllactosamine type



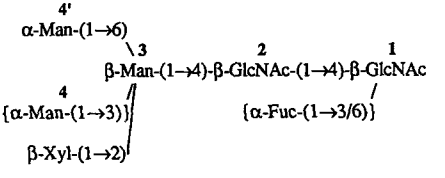
oligomannoside type



hybrid type



xylose type



a. *N-acetyllactosamine type*; the core is extended with β-Gal-(1→4)-β-GlcNAc-(1→ units. The number of units can vary considerably, ultimately leading to polylactosamine type of chains. Furthermore, fucose, sialic acids, and N-acetylgalactosamine may be present. β-GlcNAc-(1→4)-linked to β-Man is designated bisecting GlcNAc. It can occur in the various antennary structures.
b. *Oligomannoside type*; the core is extended with up to six Man residues and additionally with up to three Glc residues.
c. *Hybrid type*; this type shows characteristics of the *N*-acetyllactosamine as well as of the oligomannoside type.
d. *Xylose-containing type*; the core bears β-Xyl (1→2)-linked to β-Man. The type occurring in plants can have α-Fuc (1→3)-linked to GlcNAc-**1** [2]. In lower animal species Xyl has also been found, but in these cases α-Fuc is (1→6)-linked to GlcNAc-**1** [3]. Man-**4** can be absent. Several extensions of the basic structure have been found.

The linkage of the O-chains to the polypeptide backbone exhibits quite some variability. The chains can be attached via GalNAc, Man, or Gal to the hydroxyl group of Ser or Thr, via Xyl to Ser, via Gal to 5-L-hydroxylysine (Hyl) or 4-L-hydroxyproline (Hyp) and via Ara to Hyp. The carbohydrate chains wherein the carbohydrate-protein linkage is formed between GalNAc and Ser or Thr are designated the mucin type. For this kind of chains several core structures have been defined [4] depending on the type of monosaccharide residue(s) attached to C3 and/or C6 of GalNAc, which may be extended with Gal, GlcNAc, GalNAc, Fuc and NeuAc/NeuGc. Quite a number of immuno determinants have been found in these chains, especially in the peripheral parts.

The biosynthetic pathways of protein glycosylation [4,5] imply merely an indirect genetic control over the structures that are finally present in glycoproteins. For the greater part of the glycosylation sites, the process of glycosylation leads to a microheterogeneity in the carbohydrate chains, which is fairly constant and characteristic for individual cells. This creates tissue and species-specific patterns of glycosylation. Therefore, a glycoprotein rarely consists of a single molecular species, but rather of an ensemble of different molecules. If only one glycosylation site occurs the number of molecular species equals the number of different oligosaccharide structures at the site. Obviously, the total number of molecular species increases with the number of glycosylation sites.

2 STRUCTURE ANALYSIS

The analysis of the structure of the carbohydrate chains of glycoproteins is not as straightforward as for the protein part. For each carbohydrate chain the primary structure has to be determined in conjunction with nature and position of non-carbohydrate substituents. Furthermore, the location of the glycosylated amino acid residue in the polypeptide chain has to be identified. For such glycosylation site a clear picture has to be developed as to the individual microheterogeneity. Finally, to understand the function of these carbohydrates at the molecular level, it is essential to have insight into the specific conformational freedom of a particular carbohydrate chain in an intact glycoprotein. The polypeptide chain and/or neighbouring carbohydrate chains may put considerable constraints on the intrinsic flexibility of oligosaccharide and thereby limiting the variety of conformations that it can adopt. It is possible that in the intact glycoprotein, carbohydrate conformations prevail which are energetically less favourable for the free oligosaccharide. It is

also important to investigate for an individual glycosylation site, in which way the conformational freedom is influenced by the natural microheterogeneity. Differences in chain length and/or position of glycosidic linkages may in case when this is reflected in significant conformational alterations, lead to masking or unmasking of recognition sites. This conformational effect can be of importance on top of the change in primary structure which is inherent to microheterogeneity.

Despite the urgent need for reliable information on the three dimensional structure of intact glycoproteins in the crystalline state as well as in solution, only limited data are available. By X-ray analysis it is still difficult to define entire carbohydrate chains in glycoproteins. In particular the peripheral parts of the oligosaccharides, which are interesting as to interaction features, can only be described with a large degree of uncertainty. Also NMR spectroscopy has inherent limitations in unravelling the structure of intact glycoproteins. It is evident that the occurrence of multiple glycosylation sites gives rise to a high signal density in the spectra that can hardly be interpreted into any detail. In other words, the study of the three dimensional structure of glycoproteins is not trivial and cannot yet be carried out on a routine like basis.

The analysis of the primary structure of the carbohydrate chains of glycoproteins is predominantly carried out on partial structures, preferably representing a single glycosylation locus. Several procedures exist for the liberation of *N*- or *O*-linked carbohydrates from glycoproteins as summarized in Table I.

Table I. *Procedures to liberate carbohydrate chains from the glycoprotein*

N-glycosidic	*O*-glycosidic
proteolytic digestion[a]	proteolytic digestion[a]
hydrazinolysis	alkaline borohydride treatment[b]
alkaline borohydride treatment[b]	enzymatic hydrolysis
enzymatic hydrolysis	

[a]This procedure leads to formation of glycopeptides
[b]Different conditions are applied for the release of *N*- and *O*-glycosidic carbohydrate chains

Proteolytic degradation of glycoproteins to glycopeptides may in the ideal situation lead to compounds corresponding to an individual glycosylation site. Several proteases can be applied for this purpose depending on the protein structure. Often exhaustive pronase digestion is used but it is difficult to end with glycopeptides that are homogeneous in the protein part as well as in the carbohydrate part. Complex fractionation problems may result from this approach. The use of this method can also be hampered by the distribution of the carbohydrate chains over the protein because hindrance imposed by the carbohydrates may impair complete degradation thereby leaving adjacent carbohydrate chains attached to one peptide. This is specially relevant for the analysis of mucin type of *O*-linked structures. Hydrazinolysis [6] is the most commonly applied chemical cleavage procedure for the *N*-glycosidic GlcNAc-Asn linkage . The method has some disadvantages, like partial modification of the reducing end which introduces further heterogeneity, partial cleavage of the mucin type chains and removal of several essential non-carbohydrate substituents (e.g. from sialic acids). These complications may explain the continuous redefining of the optimal conditions for cleavage. *O*-glycosidic chains involving the GalNAc-Ser/Thr linkage are usually split off by alkaline borohydride treatment. Unfortunately, there are not yet alternative procedures available which are

generally applicable . Like hydrazinolysis, the β-elimination reaction leads to a loss of information as to the glycosylation sites in case of the occurrence of multiple glycosylation sites and to loss of some non-carbohydrate substituents.

For the release of *N*-glycosidic carbohydrate chains two important classes of enzymes are known, namely the endo-β-N-acetylglucosaminidases [7], which hydrolyse the β-GlcNAc-(1→4)-GlcNAc linkage in the *N,N'*-diacetylchitobiose unit and the peptide-N^4-(*N*-acetyl-β-glucosaminyl) asparagine amidases [8] (PNGases) which hydrolyse the β-aspartyl glycosylamine linkage. The specificities of the endoglycosidases are rather rigid, which limit their applicability as a general tool in structural studies. Two different PNGases have been characterized extensively so far, PNGase A (E.C. 3.5.1.52) isolated from almond emulsin and PNGase F (E.C. 3.5.1.52) from the culture filtrate of *Flavobacterium meningosepticum.* A prominent feature of PNGases is their resistance to denaturing reagents like chaotropic compounds, detergents and disulfide reductants. This feature is a major advantage since unfolding of the proteins enhances the efficiency of the oligosaccharide release. Although the PNGases A and F have in common the ability to release most of the Asn-linked carbohydrates, there are differences. It has been observed that PNGase A can split off chains bearing α-Fuc-(1→3)-linked to GlcNAc-1 [9], whereas PNGase F is unable to do so.

Direct structural analysis of the partial structures in a complex mixture is usually not very meaningful. Fractionations based on differences in mass, charge, hydrophylic properties and affinity for lectins or antibodies are essential to arrive at single components or at relatively simple mixtures. For the structure determination, a variety of analytical techniques are available, involving sugar analysis, methylation analysis [10], specific degradation techniques, like periodate/Smith oxidation [11], deamination with nitrous acid [12], and acetolysis [13], specific enzymatic degradation by endo- or exoglycosidases [7], electron impact-, chemical ionization-, and fast atom bombardment mass spectrometry [10, 14, 15] whether or not in combination with GC, LC or SCF as well as ^{1}H-NMR [16] and ^{13}C-NMR spectroscopy [17]. The unambiguous determination of the structure of novel carbohydrate chains often requires a combination of two or more of the aforementioned methods.

3 ^{1}H-NMR SPECTROSCOPY

During the last 15 years we have developed the application of high-resolution ^{1}H-NMR spectroscopy to a very powerful method for the structure determination of glycoconjugate derived carbohydrate chains. For the identification of N- as well as O-chains, we introduced the structural-reporter group concept [18]. The greater part of the sugar skeleton protons of the constituent residues resonate between δ= 3.5 and 3.9. This region of overlapping resonances is called the bulk region. The signals present outside the bulk region are designated the structural-reporter-groups signals, since they stem from protons that provide information on the primary structure. The most important structural-reporter groups are listed in Table 2.

Table 2. Structural-reporter groups

- Anomeric protons
- H-3 atoms of sialic acids
- Man H-2 atoms
- Gal H-4 atoms
- GalNAc-ol H-2, H-3, H-4 and H-5 atoms
- Protons shifted out of the bulk region due to glycosylation shifts
- Protons shifted out of the bulk region due to the presence of non-carbohydrate substituents like acyl, sulfate and phosphate groups
- Protons belonging to non-carbohydrate substituents like *O*-methyl, *N,O*-acetyl, *N*-glycolyl groups

The translation of the 1D ^{1}H-NMR spectra into structures is based on extensive libraries of reference compounds [16, 19]. Since the concept of structural-reporter groups is simple and adequate, it is generally taken as the first step in the interpretation of the ^{1}H-NMR spectra. If not all structural-reporter groups can be assigned that are necessary to define the primary structure unambiguously, 2D-NMR experiments can be performed to unravel the interconnection of proton signals. A very useful method is 2D homonuclear Hartmann Hahn (2D HOHAHA) spectroscopy [20] that provides subspectra of the constituting monosaccharide residues. However, the transfer of magnetization to a proton in a residue through scalar coupling may be hampered by the occurrence of small couplings. NOE type of spectra can be very valuable for sequential assignments. Owing to the higher sensitivity for smaller compounds, 2D rotating-frame NOE spectroscopy (2D ROESY) is more suited for assignment purposes than 2D NOESY [21].

4 ANALYSIS OF *N,O*-GLYCOPROTEINS

Recently, we developed a suitable strategy for the analysis of *N,O*-glycoproteins wherein the O-linked chains are attached via Ser or Thr. It is based on the release of the N-linked chains with PNGase (F or A depending on the glycoprotein) followed by alkaline borohydride cleavage of the O-linked chains from the recovered *O*-glycoprotein [22]. The procedure is summarized in Fig. 2.

A few points deserve further attention. It is important to monitor the enzymatic cleavage of the N-linked chains carefully. Several methods can be employed to this end, including SDS-PAGE with silver staining, Glycan-kit detection [23], Con A test [24] or a test with other lectins on a similar basis. It is important that the reaction leads to complete *N*-deglycosylation. This may require the optimization of the reaction conditions as to pH, chaotropic reagents, etc. Various chromatographic techniques like FPLC on Mono Q, HPLC on Lichrosorb-NH_2 and HPLC on CarboPac PA1 are useful to obtain compounds of sufficient purity to be analysed successfully.

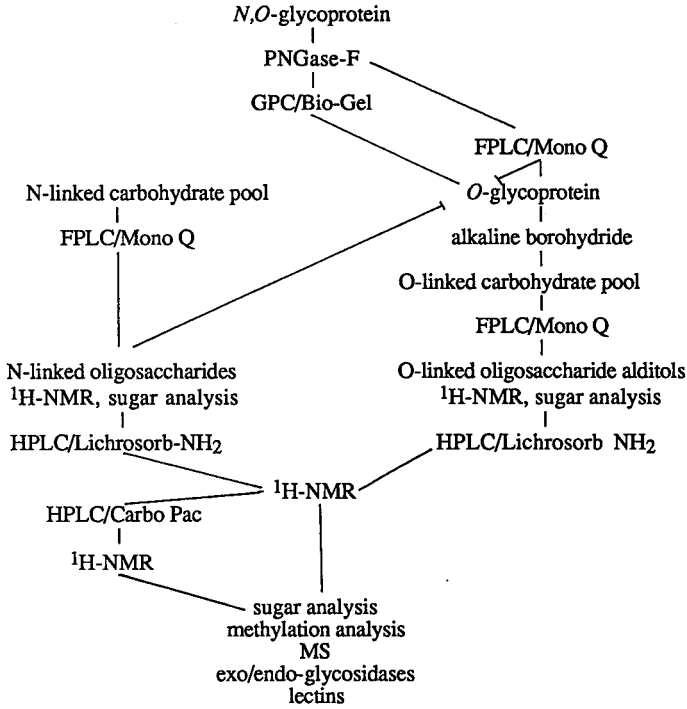
5 ANALYSIS OF RECOMBINANT DNA GLYCOPROTEINS

The first recombinant glycoprotein that we investigated was human γ-interferon expressed in Chinese hamster ovary (CHO) cells [25]. We observed the presence of a mixture of mono- and disialo diantennary

structures that are partially fucosylated at GlcNAc-1. α-NeuAc occurs exclusively in (2→3) linkage, which is a normal feature for CHO cells. In this respect the recombinant form differs from the native human IFN-γ wherein α-NeuAc is mainly (2→6) linked. We carried out studies on the glycosylation patterns of human chorionic gonadotropin (hCG) [unpublished] and human follitropin (hFSH) [26], both expressed in CHO cells. The analysis proved an impressive similarity in types of structure occurring in the recombinant glycoproteins when compared to the corresponding native compounds. Apart from variations in relative amounts of the various types of lactosamine structures, it is interesting to note that the CHO cells do not introduce bisecting GlcNAc residues, nor α-NeuAc in (2→6) linkage.

Recently, we studied the carbohydrate chains of the human insulin-like growth factor (IGF-1) expressed in *Saccharomyces cerevisiae* [27]. By mild alkaline treatment of the glycoprotein all oligosaccharides, which consisted of Man only, could be split off. By ^{1}H-NMR spectroscopy the released oligosaccharides could be identified as Manα(1→2)Man as the major chain, and Manα(1→3)Manα(1→2)Manα(1→2)Man as the minor constituent.

Fig. 2 Working up procedure for the release, fractionation and analysis of glycoprotein glycans.



5 NON-CARBOHYDRATE CONSTITUENTS IN CARBOHYDRATE CHAINS

In carbohydrate chains of glycoproteins various types of non-carbohydrate substituents can occur. The stability of the linkages of the substituents under the conditions used for the structure determination may be quite different. Some remain intact during the whole analysis procedure e.g. methyl ethers. The presence of such methyl groups can be established via sugar analysis. The location of the methyl groups in the carbohydrate chain can be inferred by means of ^{1}H-NMR spectroscopy [28, 29, 30]. The acyl groups in a carbohydrate chain can easily be lost during the analysis. In studies on equine fibrinogen [31] and on equine chorionic gonadotropin [32], we could demonstrate the occurrence in N-linked chains of *N*-acetyl-4-*O*-acetylneuraminic acid residues. The position of the acetyl group in sialic acid and the location of the sialic acid residue(s) in the N-linked chain could unambiguously by derived from the ^{1}H-NMR spectra. The mild isolation procedure of the N-linked chains that were split off with PNGase F, allowed the determination of the occurrence of Neu4,5Ac$_2$ residues in the free and intact oligosaccharides. The same approach is suitable for the identification of the N-substituent of sialic acids as we could show for the structure analysis of *N*-acetyl/*N*-glycolylneuraminic acid-containing N-linked carbohydrate chains of bovine fibrinogen [33].

Sulfates often occur as substituents in N- as well as in O-linked carbohydrate chains. It seems advisable to keep the conditions for cleavage and isolation of sulfated carbohydrate chains as mild as possible. Usually the location of the sulfate group in the constituting monosaccharides and the position of the sulfated monosaccharides in the chain can be deduced from ^{1}H-NMR spectra [34-36]. Sulfate groups induce a large downfield shift (approximately 0.5 ppm) on the proton(s) attached to the C-atom bearing the sulfate group. By consequence, the signals of these protons resonate outside the bulk region. The also often occurring phosphates are slightly more difficult to localize because the shifts induced by phosphates can be much smaller. Therefore, the proton(s) attached to a C-atom that bears a phosphate group, may still resonate in the bulk region. For this purpose we developed the heteronuclear 1D ^{1}H{^{31}P} relayed spin-echo Difference Spectroscopy (RESED) [37]. This novel pulse sequence can be used for the observation of subspectra of phosphorylated residues in carbohydrates [38].

ACKNOWLEDGEMENT

This investigation was supported by the Netherlands Foundation for Chemical Research (SON) and the Netherlands Program for Innovation Oriented Carbohydrate Research (IOP-k).

REFERENCES

[1] N. Sharon, H. Lis in H. Neurath (ed.), The proteins, Vol. 5, Academic Press, New York, 1982, pp. 1-144.

[2] J. Kérekgyártó, J.P. Kamerling, J.B. Bouwstra, J.F.G. Vliegenthart, A. Lipták, Carbohydr. Res. 186 (1989) 51-62.

[3] J.A. van Kuik, H. van Halbeek, J.P. Kamerling, J.F.G. Vliegenthart, J. Biol. Chem. 260 (1985) 13984-13988.

[4] K.L. Carraway, S.R. Hull, BioEssays 10 (1989) 117-121.

[5] R. Kornfeld, S. Kornfeld, Ann. Rev. Biochem. 54 (1985) 631-664.

[6] S. Takasaki, T. Mizuochi, A. Kobata, Methods Enzymol. 83 (1982) 263-268.

[7] A. Kobata, Anal. Biochem. 100 (1979) 1-14.

[8] F. Maley, R.B. Trimble, A.L. Tarentino, T.H. Plummer Jr., Anal. Biochem. 180 (1989) 195-204.

[9] N. Takahashi, H. Hitotsuya, H. Hanzawa, Y. Arata, Y. Kurihara, J. Biol. Chem. 265 (1990) 7793-7798.

[10] J.P. Kamerling, J.F.G. Vliegenthart in A.M. Lawson (ed.), Clinical Biochemistry - Principles, Methods, Applications, Vol. 1, Mass Spectrometry, Walter de Gruyter, Berlin (1989) 175-263.

[11] T. Krusius, J. Finne, J. Carbohydr. Res. 90 (1981) 203-214.

[12] G. Strecker, A. Pierce-Crétel, B. Fournet, G. Spik, J. Montreuil, Anal. Biochem. 111 (1981) 17-26.

[13] B. Bayard, J. Montreuil, J. Carbohydr. Res. 124 (1972) 427-443.

[14] A. Dell, J.E. Oates in C.J. Bierman and G. McGinnis (eds.) Analysis of carbohydrates by GLC and MS, CRC Press, Boca Raton, Florida (1988).

[15] A. Dell, Adv. Carbohydr. Chem. Biochem. 45 (1989) 19-72.

[16] J.F.G. Vliegenthart, L. Dorland, H. van Halbeek, Adv. Carbohydr. Chem. Biochem. 41 (1981) 209-374.

[17] K. Dill, E. Berman, A.A. Pavia, Adv. Carbohydr. Chem. Biochem. 43 (1985) 1-49.

[18] J.F.G. Vliegenthart, H. van Halbeek, L. Dorland, Pure Appl. Chem. 53 (1981) 45-77.

[19] J.P. Kamerling, J.F.G. Vliegenthart, Biol. Magn. Reson. (1990) in press.

[20] A. Bax, D.G. Davis, J. Magn. Reson. 65 (1985) 355-360.

[21] J. Breg, D. Romijn, J.F.G. Vliegenthart, G. Strecker, J. Montreuil, Carbohydr. Res. 183 (1988) 19-34.

[22] J.B.L. Damm, J.P. Kamerling, G. van Dedem, J.F.G. Vliegenthart, Glycoconj. J. 4 (1987) 129-144.

[23] A. Haselbeck, W. Hösel, Glycoconj. J. 7 (1990) 63-74.

[24] J.P.M. Langeveld, M.E. Noelken, K. Hård, P. Todd, J.F.G. Vliegenthart, J. Rouse, B.G. Hudson, submitted for publication.

[25] J.H.G.M. Mutsaers, J.P. Kamerling, R. Devos, Y. Guisez, W. Fiers, J.F.G. Vliegenthart, Eur. J. Biochem. 156 (1986) 651-654.

[26] K. Hård, A. Mekking, J.B.L. Damm, J.P. Kamerling, W. de Boer, R.A. Wijnands, J.F.G. Vliegenthart, Eur. J. Biochem., in press.

[27] K. Hård, W. Bitter, J.P. Kamerling, J.F.G. Vliegenthart, FEBS Lett. 248 (1989) 111-114.

[28] J.A. van Kuik, R.P. Sijbesma, J.P. Kamerling, J.F.G. Vliegenthart, E.J. Wood, Eur. J. Biochem. 160 (1986) 621-625.

[29] J.A. van Kuik, R.P. Sijbesma, J.P. Kamerling, J.F.G. Vliegenthart, E.J. Wood, Eur. J. Biochem. 169 (1987) 399-411.

[30] G.J. Gerwig, P. de Waard, J.P. Kamerling, J.F.G. Vliegenthart, E. Morgenstern, R. Lamed, E.A. Bayer, J. Biol. Chem. 264 (1989) 1027-1035.

[31] J.B.L. Damm, H. Voshol, K. Hård, J.P. Kamerling, J.F.G. Vliegenthart, Eur. J. Biochem. 180 (1989) 101-110.

[32] J.B.L. Damm, K. Hård, J.P. Kamerling, G.W.K. van Dedem, J.F.G. Vliegenthart, Eur. J. Biochem. 189 (1990) 175-183.

[33] J.B.L. Damm, A.A. Bergwerff, K. Hård, J.P. Kamerling, J.F.G. Vliegenthart, Recl. Trav. Chim. Pays-Bas 108 (1989) 351-359.

[34] J.A. van Kuik, J. Breg, C.E.M. Kolsteeg, J.P. Kamerling, J.F.G. Vliegenthart, FEBS Lett. 221 (1987) 150-154.

[35] P. de Waard, A. Koorevaar, J.P. Kamerling, J.F.G. Vliegenthart, J. Biol. Chem. (in press).

[36] K. Sugahara, I. Yamashina, P. de Waard, H. van Halbeek, J.F.G. Vliegenthart, J. Biol. Chem. 263 (1988) 10168-10174.

[37] P. de Waard, J.F.G. Vliegenthart, J. Magn. Reson. 81 (1989) 173-177.

[38] P. de Waard, J.F.G. Vliegenthart, Y. Kozutsumi, T. Kawasaki, I. Yamashina, J. Biol. Chem. 264 (1989) 12141-12144.