

---

## Studies on glycoprotein-derived carbohydrates

Johannes F. G. Vliegthart

Department of Bio-Organic Chemistry, Bijvoetcenter, Utrecht University, P.O. Box 80.075 NI-3508 TB Utrecht, The Netherlands

---

### Introduction

Within the past decade, it has been demonstrated that the carbohydrate chains of glycoproteins can exert a large variety of functions *in vivo*. Most fascinating is the involvement of these chains in molecular recognition processes [1-4]. Considering that the glycans are located at the outer-surface of the glycoprotein and that they can exhibit a huge variety in structure, it has often been stated that carbohydrate chains are very well suited for such biological roles. But it is rarely possible to identify one particular carbohydrate chain of a glycoprotein that can be held responsible for a specific function. The heterogeneity of the carbohydrate chains in combination with the difficulty to set up bio-assays that allow the unambiguous characterization of effects in terms of specific carbohydrate chains are the main reasons for the complexity of the problem.

In the analytical sense, enormous progress has been made in the study of glycoproteins. Adequate methods are available to isolate the glycoproteins free from contaminants, to degrade glycoproteins to partial structures chemically and/or enzymically, as well as to fractionate the partial structures. The improvements in fractionation techniques such as high-pH anion-exchange chromatography with pulsed amperometric detection [5,6] and capillary electrophoresis have revealed that the heterogeneity of the glycans of a defined glycoprotein can be

much larger than anticipated so far. The number of compounds that can be traced now has grown enormously even though some occur only at low concentrations. This gives rise to important questions regarding the origin and the relevance of these minor constituents. On the one hand, a tendency can be recognized to conceive part of the collection of glycans as artifactual because they are formed along catabolic routes, or arise from different cell types, or more trivially, arise from the pooling of material, thereby reflecting natural variability. On the other hand, there is the doubt that not all glycans of a glycoprotein have been isolated and identified. This last aspect invariably plays a role in cases wherein a particular biofunction is ascribed to a glycan. Synthesis of the compounds is almost indispensable in verifying the correctness of the proposed action.

In our work, we have encountered a few new interesting examples of heterogeneity of the *N*- and/or *O*-linked carbohydrate chains of glycoproteins. In all cases, the glycoproteins were first purified rigorously. The *N*-glycans were excized by means of peptide-*N*<sup>4</sup>-(*N*-acetyl- $\beta$ -glucosaminy)-asparagine amidase F and the degree of *N*-deglycosylation was carefully checked by PAGE with glycan detection. For each glycoprotein, the conditions for quantitative removal of the *N*-glycans have to be optimized. For research purposes, we prefer this enzymic approach over hydrazinolysis. Subsequently, the *O*-glycans are cleaved from the obtained or authentic *O*-glycoprotein by alkaline borohydride treatment. In this way, pools of glycans

---

Abbreviations used: CHO, Chinese hamster ovary; eCG, equine chorionic gonadotropin; rEPO, recombinant erythropoietin; ZP, zona pellucida.

are obtained that were either *N*- or *O*-linked to the protein. After thorough fractionation via various chromatographic procedures, the carbohydrates are identified. We have used one- and two-dimensional  $^1\text{H}$ -n.m.r. spectroscopy for this purpose, if necessary and possible in conjunction with methods affording additional evidence. The  $^1\text{H}$ -n.m.r. methodology is based on the structural reporter group concept [7,8], allowing the translation of clearly identifiable  $^1\text{H}$ -n.m.r. signals into primary structures. Recently, we developed a database of  $^1\text{H}$ -n.m.r. data that makes it possible to connect observed n.m.r. parameters with structures [9]. In this way, parts of structures can also be retrieved, demonstrating the suitability of this method in analysing unknown compounds. Below, we give an overview of some of the glycoproteins that we have investigated recently.

### Human Tamm–Horsfall glycoprotein

This glycoprotein has been purified from the urine of one male. This PtdIns-anchored membrane protein [10] is localized in the thick ascending limb of the loop of Henle and the early convoluted tubule segments of the nephron. After cleavage from the membrane, this protein appears in the urine. Its physiological function is still unclear. The glycans of this protein have been proposed to be involved in the prevention of urinary tract and urinary bladder infections. Furthermore, they could possibly play a role in immunosuppressive features [11].

From this *N*-glycoprotein, we could obtain more than 150 carbohydrate-containing fractions, some which still contain more than one oligosaccharide. The structure of 30 glycans could be elucidated; these varied from nonfucosylated mono-sialylated diantennary to fucosylated tetrasialylated tetra-antennary compounds. Interestingly, the antennae can be terminated with  $(\text{SO}_4^-)$ -3Gal $\beta$ 1-4GlcNAc $\beta$ 1-,  $(\text{SO}_4^-)$ -4GalNAc $\beta$ 1-4GlcNAc $\beta$ 1- and NeuAc $\alpha$ 2-3[GalNAc $\beta$ 1-4)Gal $\beta$ 1-4GlcNAc $\beta$ 1-. The last terminus represents the Sd<sup>a</sup> determinant, which has now been localized precisely. In contrast to previous studies [12–14], we did not observe significant amounts of oligomannose-type carbohydrates. This result might indicate that the collection of glycans occurring in Tamm–Horsfall glycoprotein shows a considerable individual variation. The issue of the immunosuppressive activity of glycans of this protein is still not answered. It is not even clear whether glycans that would exhibit such an effect are derived from it. In cases when the immunosuppressive activity of glycan chains would really exist,

it is an interesting point to establish its (patho)physiological function.

Upon the discovery of novel glycan structures, it has often been assumed that these structures are unique and specific for the glycoprotein involved. Our finding of the  $(\text{SO}_4^-)$ -4GalNAc $\beta$ 1-4GlcNAc $\beta$ 1-sequence in Tamm–Horsfall glycoprotein and in the urinary type plasminogen activator [15] proves that we have to be very careful in this respect and that its occurrence is by no means restricted to pituitary glycoprotein hormones [16]. This observation complicates the study of structure–function relationships.

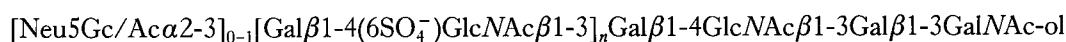
### Recombinant human erythropoietin

This *N,O*-glycoprotein is synthesized by the kidney and stimulates the proliferation and differentiation of erythroid progenitor cells. In view of its application as a drug for the treatment of anemia associated with renal failure [17], there is great interest in the recombinant human glycoprotein, expressed in heterologous cell systems. In addition to the studies carried out on recombinant erythropoietin (rEPO) expressed in Chinese hamster ovary (CHO) or baby hamster kidney cells [18–23], we carried out a detailed structure analysis of the glycans of EPO expressed in CHO cells, in order to unravel the native sialylation pattern and to locate the extra *N*-acetylglucosamine units [24,25]. The isolated and purified glycans turned out to be of the di-, tri-, tri'- and tetra-antennary *N*-acetylglucosamine type. The majority of these chains is completely sialylated and a minority partially. The sialic acids are invariably  $\alpha$ 2-3 linked and occur as *N*-acetylneuraminic acid (95%), *N*-glycolylneuraminic acid (2%) or as *N*-acetyl-9-*O*-acetylneuraminic acid. The pattern of sialylation in only partially sialylated antennae is in agreement with the known antenna-specificity of the sialyl transferase. One or two extra *N*-acetylglucosamine units may occur in each of the antennae attached to the  $\alpha$ 1-6 linked Man residue. Such repeats of *N*-acetylglucosamine units are observed in completely or partially sialylated di-, tri'- and tetra-antennary compounds. Two *O*-linked chains have been identified, namely Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3GalNAc-ol and Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-3[Neu5Ac $\alpha$ 2-6]GalNAc-ol. Without claiming that all constituting glycans of the *N*-type have been isolated and identified, it can be said that for 36 of them the structures have been elucidated in this study. This result illustrates nicely the formidable heterogeneity that can occur. We consider this to be an authentic feature for recombinant glycoproteins from CHO cells because the overall degree of sialylation is high and

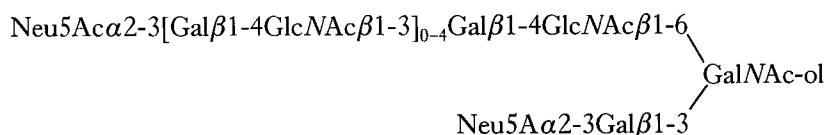
for several types of structures complete. Furthermore, in case of partial sialylation, the distribution of sialic acids over the antennae is not random. As a consequence, the sialylation pattern results from biosynthesis rather than from catabolism or from isolation artefacts. It should be noted that there are no indications that sulphate residues occur in rEPO from CHO cells.

### Equine chorionic gonadotropin

In the framework of our studies on gonadotropins, the structures of the glycans of equine chorionic gonadotropin (eCG) have been investigated. Previously, we identified for this heterodimeric glycoprotein ( $\alpha$  and  $\beta$  subunits) the major *N*- and *O*-linked glycans of the  $\beta$  subunit [26]. The *N*-glycosidic chains consist of mono-, tri- and tri'-antennary *N*-acetylglucosamine type of chains, partly fucosylated at the Asn-bound GlcNAc. The



sialic acid residues can be *N*-acetyl or *N*-acetyl-4-*O*-acetyl neuraminic acid in  $\alpha 2-3$  or  $\alpha 2-6$  linkage. The presence of an acetyl group at position 4 of sialic acid increases the heterogeneity of the glycan population. Upon administration of eCG to animals that normally do not possess sialic acids with a 4-*O*-acetyl substituent, it can be expected that the hormone has a prolonged lifetime when compared to the compounds without the *O*-acetyl substituent. It remains to be established how this would affect the bioactivity. The *O*-glycosidic chains would consist mainly of tri-, penta- and hexasaccharides. Further investigations [27] proved that the proposal [26], that polylactosamine units would also occur in this class of glycans, is correct. The newly identified oligosaccharide alditols have the structure:



The finding that the repeating *N*-acetylglucosamine units are exclusively located in the branch linked to C-6 of GalNAc is in keeping with the data showing that, in core type 2, this branch is preferentially elongated with *N*-acetylglucosamine units.

### Porcine zona pellucida glycoproteins

The zona pellucida (ZP) comprises the glycoprotein matrix surrounding the mammalian oocyte, and is involved in some critical steps in the fertilization process. There are strong indications that the glycans of the ZP glycoproteins, which belong to

the *N,O*-glycoproteins, play a role in the sperm recognition event [3]. For the porcine system, however, there is still ambiguity as to whether the *N*- or the *O*-chains are involved in that process [28,29]. Structural studies on the *N*-glycans [28,30] as well as on the neutral *O*-glycans [31] have been reported.

To gain insight into the structures of the acidic *O*-glycans, these chains were released from the porcine ZP glycoproteins by alkaline borohydride treatment, after removal of the *N*-glycans by PNGase F. An extremely heterogeneous mixture of *O*-linked oligosaccharide alditols was obtained. After several chromatographic steps, the primary structures of 32 *O*-glycans were determined by one- and two-dimensional  $^1\text{H}$ -n.m.r. spectroscopy. The major part of the analysed oligosaccharide alditols [25,32] belongs to a collection of compounds with the structure:

where  $n = 0-6$ .

It is interesting to notice that in *O*-linked glycans with core type 1, long poly(*N*-acetylglucosamine) chains can be formed. It is now important to investigate whether one or more of the characterized carbohydrate chains have indeed the presumed role in the sperm-oocyte interaction. Because the *N*- and *O*-linked chains have so much in common at the non-reducing termini, it is tempting to propose that the core regions are perhaps involved in the epitope.

### Concluding remarks

The study of various glycoproteins has shown that the heterogeneity of the glycans of the *N*- as well as

of the *O*-type can be very large. To describe the heterogeneity in detail, it is essential that the working-up procedures give as good as possible a quantitative insight into all structures present. This is the more important in cases wherein a bioactivity is presumed to be associated with one or more specific carbohydrate chain(s). N.m.r. spectroscopy is a suitable method with which to identify oligosaccharide structures. It is, however, essential that the amount of starting glycoprotein is sufficient also

to enable the glycan occurring in the lowest concentration to be analysed unambiguously.

The author gratefully acknowledges the contributions of Drs A. A. Bergwerff, J. B. L. Damm, C. van Dedum, C. H. Hokke, K. Hård, J. P. Kamerling, B. Penninkhof and G. van Zadelhoff. The investigations were supported by grants from the Netherlands Foundation for Chemical Research (SON/NWO) and the Netherlands Program for Innovation Oriented Carbohydrate Research (IOP-k).

- 1 Berger, E. G., Buddecke, E., Kamerling, J. P., Kobata, A., Paulson, J. C. and Vliegthart, J. F. G. (1982) *Experientia* **38**, 1129–1162
- 2 Rademacher, T. W., Parekh, R. B. and Dwek, R. A. (1988) *Annu. Rev. Biochem.* **57**, 785–838
- 3 Wassarman, P. M. (1988) *Annu. Rev. Biochem.* **57**, 415–422
- 4 Varki, A. (1993) *Glycobiology* **3**, 97–130
- 5 Townsend, R. R. and Hardy, M. R. (1991) *Glycobiology* **1**, 139–147
- 6 Hermentin, P., Witzel, R., Vliegthart, J. F. G., Kamerling, J. P., Nintz, M. and Conradt, H. S. (1992) *Anal. Biochem.* **203**, 281–289
- 7 27th IUPAC Congress, Helsinki, 27–30 August 1979 (1980) *Pure Appl. Chem.*, pp. 253–262, Pergamon Press, Oxford and New York
- 8 Vliegthart, J. F. G., Dorland, L. and van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* **41**, 209–374
- 9 Van Kuik, J. A., Hård, K. and Vliegthart, J. F. G. (1992) *Carbohydr. Res.* **235**, 53–68
- 10 Rindler, M. J., Naik, S. S., Li, N., Hoops, T. C. and Peraldi, M. N. (1990) *J. Biol. Chem.* **265**, 20784–20789
- 11 Pennica, D., Kohr, W. J., Kuang, W. J., Glaister, D., Aggarwal, B. B., Chen, E. Y. and Goeddel, D. V. (1987) *Science* **236**, 83–88
- 12 Serafini-Cessi, F., Dall'Olio, F. and Malagolini, N. (1984) *Biosci. Rep.* **4**, 269–274
- 13 Dall'Olio, F., de Kanter, F. J. J., van den Eynden, D. H. and Serafini-Cessi, F. (1988) *Carbohydr. Res.* **178**, 327–332
- 14 Smagula, R. M., van Halbeek, H., Decker, J. M., Muchmore, A. V., Moody, C. E. and Sherblom, A. P. (1990) *Glycoconjugate J.* **7**, 609–624
- 15 Bergwerff, A. A. (1994) Thesis, Utrecht University
- 16 Baenziger, J. U. and Green, E. D. (1988) *Biochem. Biophys. Acta* **974**, 287–306
- 17 Takeuchi, M. and Kobata, A. (1991) *Glycobiology* **1**, 337–346
- 18 Sasaki, H., Bothner, B., Dell, A. and Fukuda, M. (1987) *J. Biol. Chem.* **262**, 12059–12076
- 19 Sasaki, H., Ochi, N., Dell, A. and Fukuda, M. (1988) *Biochemistry* **27**, 8618–8626
- 20 Takeuchi, M., Takasaki, S., Miyazaki, H., Kato, T., Hoshi, S., Kochibe, N. and Kobata, A. (1988) *J. Biol. Chem.* **263**, 3657–3663
- 21 Tsuda, E., Goto, M., Murakami, A., Akai, K., Ueda, M., Kawanishi, G., Takahashi, N., Sasaki, H., Chiba, H., Ishihara, H., Mori, M., Tejima, S., Endo, S. and Arata, Y. (1988) *Biochemistry* **27**, 5646–5654
- 22 Inoue, N., Takeuchi, M., Asano, K., Shimizu, R., Takasaki, S. and Kobata, A. (1993) *Arch. Biochem. Biophys.* **301**, 375–378
- 23 Nimtz, M., Marin, W., Wray, V., Klöppel, K.-D., Augustin, J. and Conradt, H. S. (1993) *Eur. J. Biochem.* **213**, 39–56
- 24 Hokke, C. H., Kamerling, J. P., van Dedum, G. W. K. and Vliegthart, J. F. G. (1991) *FEBS Lett.* **286**, 18–24
- 25 Hokke, C. H. (1993) Thesis, Utrecht University
- 26 Damm, J. B. L., Hård, K., Kamerling, J. P., van Dedum, G. W. K. and Vliegthart, J. F. G. (1990) *Eur. J. Biochem.* **189**, 175–183
- 27 Hokke, C. H., Roosenboom, M. J. H., Thomas-Oates, J. E., Kamerling, J. P. and Vliegthart, J. F. G. (1994) *Glycoconjugate J.*, in the press
- 28 Noguchi, S., Hatanaka, Y., Tobita, T. and Nakano, M. (1992) *Eur. J. Biochem.* **204**, 1089–1100
- 29 Yurewicz, E. C., Pack, B. A. and Sacco, A. G. (1991) *Mol. Reprod. Dev.* **30**, 126–134
- 30 Noguchi, S. and Nakano, M. (1992) *Eur. J. Biochem.* **209**, 883–894
- 31 Hirano, T., Takasaki, S., Hedrick, J. L., Wardrip, N. J., Amano, J. and Kobata, A. (1993) *Eur. J. Biochem.* **224**, 763–769
- 32 Hokke, C. H., Damm, J. B. L., Kamerling, J. P. and Vliegthart, J. F. G. (1993) *FEBS Lett.* **329**, 29–34

---

Received 24 December 1993