

Structural studies on glycoprotein glycans

Johannis P. Kamerling

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

Abstract - A minireview is presented on methodological approaches for the primary structure analysis of oligosaccharide ensembles of N,O-glycoproteins, and examples from the recombinant glycoprotein field are discussed in more detail.

INTRODUCTION

In recent years the interest in glycoprotein glycans has grown dramatically. This is mainly due to the general acceptance that these carbohydrate chains exert essential biological and physico-chemical functions in a great variety of organisms. In principle, progress in the biological and physico-chemical studies of the carbohydrate chains of glycoproteins can only be made on the basis of a team-work of structure analysis including conformational analysis, synthesis and biosynthesis of this type of molecules.

The development of advanced methods to unravel the structures of the ensembles of carbohydrate chains, present at specific amino acid residues in polypeptide backbones, has been the first step in the glycoprotein glycan revival. Although in recent years several approaches based on $^1\text{H-NMR}$ spectroscopy (1, 2) and mass spectrometry (3, 4) have become available, and much attention is paid to the development of profiling techniques based on high performance separation procedures, the primary structure analysis of glycoprotein glycans has still not reached the level of a routine analysis. Even nowadays, completely novel oligosaccharide elements are identified. Due to the complexity of the carbohydrate chains, not only in terms of monosaccharide constituents and sequences, but also in terms of non-carbohydrate substituents, the current approaches have to be applied with great care. This paper will focus on general strategies developed for the primary structure analysis of the oligosaccharide ensembles of N,O-glycoproteins, and will discuss some examples of recombinant glycoproteins.

GENERAL FEATURES OF GLYCOPROTEIN GLYCANS

Glycoproteins are biopolymers consisting of a polypeptide backbone with covalently attached carbohydrate side chains (5). In the context of this minireview, two major classes will be discussed, namely, N-glycans linked via *N*-acetyl-D-glucosamine (GlcNAc) to the amide nitrogen of L-asparagine (Asn), and O-glycans linked via *N*-acetyl-D-galactosamine (GalNAc) to the hydroxyl group of L-serine (Ser) or L-threonine (Thr). Glycoproteins may contain more than one glycosylation site, and N- as well as O-linked chains can occur at one protein backbone. Moreover, microheterogeneity of the oligosaccharide structures attached to one amino acid residue is a known phenomenon, yielding a series of distinct glycoforms for each glycoprotein.

In general, N-glycans share a common Asn-linked pentasaccharide core element $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1-$ (Man, D-mannose), which can be extended in several ways. Based on typical

elongations of this core element, three major types of N-glycans have been distinguished, namely (i) the oligomannose type; (ii) the *N*-acetylglucosamine type; and (iii) the hybrid type, showing characteristics of both the oligomannose and the *N*-acetylglucosamine type. However, recent structural data will make a new discussion about this classification necessary (adding of the xylose and the *N,N'*-diacetylglucosamine type, redefining of the hybrid type). As an illustration, in Fig. 1 two composite N-linked structures are presented.

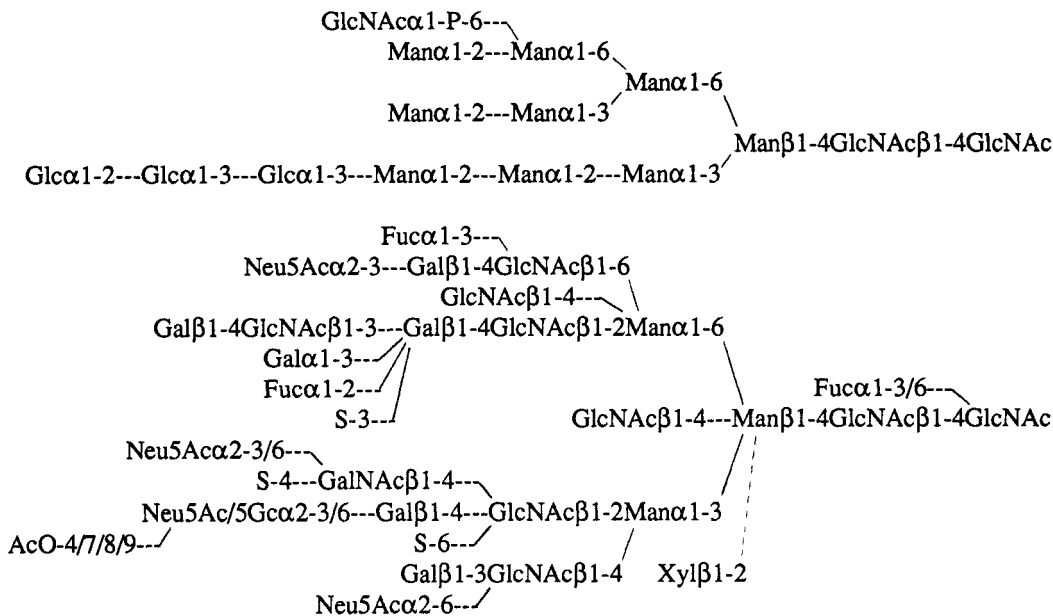


Fig. 1 Composite N-glycan structures (Glc, D-glucose; Gal, D-galactose; Fuc, L-fucose; Xyl, D-xylose; Neu5Ac, *N*-acetyl-D-neuraminic acid; Neu5Gc, *N*-glycolyl-D-neuraminic acid; AcO, *O*-acetyl; P, phosphate; S, sulfate).

For mucin-type O-glycans, in which α GalNAc is attached to Ser or Thr, nine core elements can be defined: GalNAc, Gal β 1-3GalNAc, GlcNAc β 1-3GalNAc, GalNAc α 1-3GalNAc, GlcNAc β 1-6GalNAc, GalNAc α 1-6GalNAc, GlcNAc β 1-6(Gal β 1-3)GalNAc, GlcNAc β 1-6(GlcNAc β 1-3)GalNAc, and Gal β 1-6(Gal β 1-3)GalNAc. These cores can be extended in several ways, comprising Gal β 1-4GlcNAc β or Gal β 1-3GlcNAc β elements, and terminated by monosaccharides like Fuc, Gal, GalNAc or Neu5Ac/5Gc. Moreover, *O*-acetyl and sulfate groups can occur, as demonstrated for the N-glycans.

STRUCTURE ANALYSIS OF GLYCOPROTEIN GLYCANS

In view of the presented complexity of the analytical problem, it may be evident that determination of the primary structure of Asn-linked and Ser- or Thr-linked carbohydrate chains on intact glycoproteins is not possible. Therefore, preparation of partial structures of the protein, having a single glycosylation site, or cleavage of the glycan moieties from the glycoprotein, is a prerequisite. In Fig. 2 the major routes to prepare mixtures of glycopeptides or oligosaccharides from native glycoproteins are summarized. In a next step these mixtures have to be fractionated, and different combinations are in use. Frequently applied combinations are high-voltage paper electrophoresis and Bio-Gel P-4 gel-permeation chromatography (6), or Mono Q, Lichrosorb-NH₂ and CarboPac liquid chromatography (7). Concerning the structural analysis of isolated carbohydrate chains, also different routes are followed. Two major strategies form the combination of Bio-Gel P-4 chromatography with methylation analysis and exoglycosidase studies (5, 6), and the combination of medium- and high-pressure chromatography with ¹H-NMR spectroscopy (7) and/or mass spectrometry (4), eventually completed with methylation analysis or glycosidases. But also other approaches have been reported. Nowadays, great effort is put into the development of profiling procedures, which can be applied to get an

impression of the carbohydrate chains present in unknown glycoproteins, or will be of use in controlling possible glycan microheterogeneity shifts in known glycoproteins (batch control). In this respect, more-dimensional HPLC mapping techniques (8), high-pH anion-exchange chromatography (9), capillary electrophoresis (10), fluorophore-assisted-carbohydrate electrophoresis (11) and so-called reagent-array analyses (12) seem to be highly promising.

- * Isolation of glycoproteins
- * Preparation of carbohydrate chains (5)
 - * protease digestion → glycopeptides
 - * hydrazinolysis for GlcNAc-Asn type → oligosaccharides/alditols/PA derivatives
 - * PNGase F or A for GlcNAc-Asn type → oligosaccharides/alditols/PA derivatives
 - * endo-H and other endo-*N*-acetylglucosaminidases → oligosaccharides/alditols/PA derivatives for GlcNAc-Asn type (mainly oligomannose type)
 - * alkaline borohydride treatment for GalNAc-Ser/Thr type → oligosaccharide-alditols
- * Fractionation and purification, if possible to homogeneity (5-7)
 - * gel-permeation chromatography (Bio-Gel P-4)
 - * high-voltage paper electrophoresis
 - * high-pH anion-exchange chromatography (CarboPac)
 - * lectin chromatography
 - * high-pressure chromatography (e.g. Lichrosorb-NH₂)
 - * fast-protein liquid chromatography (Mono Q)
- * Degradation procedures (5)
 - * exo-glycosidases (α -D-sialidases, α -D-galactosidases, β -D-galactosidases, *N*-acetyl- β -D-hexosaminidases, α -D-glucosidase, α -D-mannosidases, β -D-mannosidases, α -L-fucosidases, β -D-xylosidase)
 - * endo-glycosidases (endo- β -D-galactosidases)
 - * partial solvolysis
 - * periodate oxidation
- * Monosaccharide analysis, including absolute configuration determination (GLC-MS) (3)
- * Methylation analysis (GLC-MS) (3)
- * 1D/2D NMR Spectroscopy (¹H, ¹³C, ³¹P) (1, 2)
- * Mass Spectrometry (electron impact, chemical ionization, fast-atom bombardment, electrospray/ion spray, matrix-assisted laser desorption) (3, 4)
- * Profiling (LC mapping techniques, high-pH anion-exchange chromatography, capillary electrophoresis, fluorophore-assisted-carbohydrate electrophoresis, reagent-array analyses; combination with exoglycosidases or mass spectrometry) (8-12)

Fig. 2 Major methodologies in the structure analysis of glycoprotein glycans (PNGase, peptide-*N*⁴-(*N*-acetyl- β -glucosaminyl)asparagine amidase; PA, pyridylamino).

RECOMBINANT GLYCOPROTEINS

In the context of the therapeutic administration of recombinant glycoproteins, the analysis of the glycosylation patterns of the engineered proteins has got much attention. Typical examples of structure analysis as carried out in our research group comprise γ -interferon (13), follitropin (FSH) (14), chimeric plasminogen activator (K₂tu-PA) (15) and erythropoietin (EPO) (16), expressed in CHO cells. In all these cases the analyses, using ¹H-NMR spectroscopy as the major technique, were performed on the level of the sialylated carbohydrate chains (always α 2-3-linked sialic acid), giving also information about the sialylation patterns. In the case of FSH it was found that, in contrast to the native form (17), the recombinant form contains no carbohydrate chains with a bisecting GlcNAc residue (GlcNAc β 1-4Man β), in accordance with the incapability of CHO cells to produce such chains. Furthermore, native FSH has both α 2-3- and α 2-6-linked sialic acid. K₂tu-PA contains the kringle-2 domain of native tissue plasminogen activator (t-PA) and the protease domain of native urinary plasminogen activator (u-PA). Comparison of the glycosylation patterns at Asn-302 in u-PA (18) and at Asn-247 in K₂tu-PA (originally Asn-302 of u-PA) shows an important change in types of carbohydrate chains shifting from *N,N'*-diacetyllactosediamine to *N*-acetyllactosamine. With respect to erythropoietin (16, 19-24), comparison of the structures of the desialylated N-glycans of urinary, CHO and BHK

EPO's shows qualitatively the same branching patterns for the *N*-acetylglucosamine-type of chains. However, the finding of a tendency that CHO and BHK EPO's contain a higher amount of N-linked carbohydrate chains with Gal β 1-4GlcNAc β 1-3 repeats than the urinary form is not generally supported. Furthermore, conflicting results have appeared concerning the structures of the O-linked chains of urinary EPO. It may be clear that the EPO-findings support the statement in the Introduction that the primary structure analysis of glycoprotein glycans has still not reached the level of a routine analysis.

REFERENCES

1. J.F.G. Vliegthart, L. Dorland and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.* **41**, 209-374 (1983).
2. J.P. Kamerling and J.F.G. Vliegthart, *Biol. Magn. Reson.* **10**, 1- 194 (1992).
3. J.P. Kamerling and J.F.G. Vliegthart, in *Clin. Biochem. - Principles, Methods, Applications, Vol.1* (A.M. Lawson, ed.), pp.175-263, Walter de Gruyter, Berlin (1989).
4. J.A. McCloskey (ed.), *Methods Enzymol.* **193**(III), 539-768 (1990).
5. H.J. Allen and E.C. Kisailus (eds.), *Glycoconjugates*, Marcel Dekker Inc., New York (1992).
6. K. Yamashita, T. Mizuochi and A. Kobata, *Methods Enzymol.* **83**, 105-126 (1982).
7. K. Hård, J.B.L. Damm, M.P.N. Spruyt, A.A. Bergwerff, J.P. Kamerling, G.W.K. van Dedem and J.F.G. Vliegthart, *Eur. J. Biochem.* **205**, 785-798 (1992).
8. N. Tomiya, J. Awaya, M. Kuroono, S. Endo, Y. Arata and N. Takahashi, *Anal. Biochem.* **171**, 73-90 (1988).
9. P. Hermentin, R. Witzel, J.F.G. Vliegthart, J.P. Kamerling, M. Nimtz and H.S. Conradt, *Anal. Biochem.* **203**, 281-289 (1992).
10. P. Hermentin, R. Doenges, R. Witzel, C.H. Hokke, J.F.G. Vliegthart, J.P. Kamerling, M. Nimtz, H.S. Conradt and D. Brazel, *Anal. Biochem.* in press (1994).
11. C.M. Starr, I. Masada, C. Hague and B. Seid, *Glycobiology* **3**, 511 (1993).
12. C.J. Edge, T.W. Rademacher, M.R. Wormald, R.B. Parekh, T.D. Butters, D.R. Wing and R.A. Dwek, *Proc. Nat. Acad. Sci. USA* **89**, 6338-6342 (1992).
13. J.H.G.M. Mutsaers, J.P. Kamerling, R. Devos, Y. Guisez, W. Fiers and J.F.G. Vliegthart, *Eur. J. Biochem.* **156**, 651-654 (1986).
14. K. Hård, A. Mekking, J.B.L. Damm, J.P. Kamerling, W. de Boer, R.A. Wijnands and J.F.G. Vliegthart, *Eur. J. Biochem.* **193**, 263-271 (1990).
15. A.A. Bergwerff, J. van Oostrum, F.A.M. Asselbergs, R. Bürgi, C.H. Hokke, J.P. Kamerling and J.F.G. Vliegthart, *Eur. J. Biochem.* **212**, 639-656 (1993).
16. C.H. Hokke, A.A. Bergwerff, G.W.K. van Dedem, J.P. Kamerling and J.F.G. Vliegthart, manuscript in preparation (1994).
17. A. Stockell Hartree and A.G.C. Renwick, *Biochem. J.* **287**, 665-679 (1992).
18. A.A. Bergwerff, J.E. Thomas-Oates, J. van Oostrum, J.P. Kamerling and J.F.G. Vliegthart, *FEBS Lett.* **314**, 389-394 (1992).
19. H. Sasaki, B. Bothner, A. Dell and M. Fukuda, *J. Biol. Chem.* **262**, 12059-12076 (1987).
20. E. Tsuda, M. Goto, A. Murakami, K. Akai, M. Ueda, G. Kawanishi, N. Takahashi, R. Sasaki, H. Chiba, H. Ishihara, M. Mori, S. Tejima, S. Endo and Y. Arata, *Biochemistry* **27**, 5646-5654 (1988).
21. M. Takeuchi, S. Takasaki, H. Miyazaki, T. Kato, S. Hoshi, N. Kochibe and A. Kobata, *J. Biol. Chem.* **263**, 3657-3663 (1988).
22. K.G. Rice, N. Takahashi, Y. Namiki, A.D. Tran, P.J. Lisi and Y.C. Lee, *Anal. Biochem.* **206**, 278-287 (1992).
23. N. Inoue, M. Takeuchi, K. Asano, R. Shimizu, S. Takasaki and A. Kobata, *Arch. Biochem. Biophys.* **301**, 375-378 (1993).
24. M. Nimtz, W. Martin, V. Wray, K.-D. Klöppel, J. Augustin and H.S. Conradt, *Eur. J. Biochem.* **213**, 39-56 (1993).