Vol. 131, No. 1, 1985 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1985 Pages 507-514

OCCURRENCE OF THE Y DETERMINANT ON THE N-GLYCOSIDIC CARBOHYDRATE UNITS OF HUMAN γ-SEMINOPROTEIN

[a]Herman van Halbeek[a], Gerrit J. Gerwig[a], Johannes F.G. Vliegenthart[a], Ryouicchi Tsuda[b], Mitsuwo Hara[b], Kazuko Akiyama[b,c] and Karl Schmid[c]

[a]Department of Bio-Organic Chemistry, University of Utrecht, NL-3522 AD Utrecht, The Netherlands
[b]Department of Legal Medicine, Kurume University School of Medicine, Kurume 830, Japan
[c]Department of Biochemistry, Boston University School of Medicine, Boston University Medical Center, Boston, MA 02118

Received July 19, 1985

500-MHz ^{1}H-NMR spectroscopy of the oligosaccharides derived from γ-seminoprotein, a human seminal plasma glycoprotein, revealed considerable microheterogeneity both with respect to the degree of branching and with regard to the peripheral sugars. Although the protein possesses only one N-glycosylation site, di-, tri- and tetra-antennary glycans in a ratio of 40:15:45 were found to be present. Moreover, certain branches of the tri- and tetra-antennas contain one or two Fuc residues which form part of the SSEA-1 and Y determinants. It should be noted that this is the first report that describes the occurrence and localization of the Y determinant, i.e. Fucα(1→2)Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→·), in an N-glycan of a glycoprotein.

Human seminal plasma has been reported to contain over 100 proteins (1,2) some of which are associated with reproduction (3) and infertility (4). Many seminal proteins have been identified as blood serum proteins or prostatic enzymes (1,2,5,6). Further, only two of the proteins specific to human seminal plasma, namely β-microseminoprotein and γ-seminoprotein (γ-Sm), have been isolated and partially characterized (7,8, for review see ref. 9). γ-Sm which has recently been purified to homogeneity, has an apparent molecular mass of 23,000 Da and, pertinent to the present study, was reported to contain approximately 12% carbohydrate (9). Moreover, this protein was demonstrated to be an important marker in forensic medicine (10,11) and in the detection of prostate cancer (2,12).

The present paper describes the elucidation of the structures of the N-glycosidic carbohydrate units, including that of the Y determinant, of γ-Sm[1].

[1]A preliminary report of this study was presented at the annual meeting of the Society For Complex Carbohydrates, Lexington, KY, Oct. 9-12, 1983.

0006-291X/85 $1.50

MATERIALS AND METHODS

Preparation of glycopeptides from γ-Sm. γ-Sm (25mg) prepared as described earlier (7) was reduced, alkylated and then exhaustively digested with papain and pronase. The glycopeptide mixture isolated from this digest by gel filtration through Sephadex G-50 was further fractionated by chromatography on a column of immobilized Concanavalin A. Sugar analysis was carried out as reported earlier (13).

^{1}H-NMR Spectroscopy. The resulting glycopeptide fractions were repeatedly exchanged in D_2O (99.96 atom % D, Aldrich, Milwaukee, WI) with intermediate lyophilization and analyzed with a Bruker 500-MHz WM-500 spectrometer (SON NMR-facility, Nijmegen, The Netherlands) operating in the Fourier transform mode at a probe temperature of 27°C. Resolution enhancement of the spectra was achieved by Lorentzian to Gaussian transformation from quadrature phase detection. Chemical shifts are given relative to sodium 4,4-dimethyl-4-silapentane-1-sulfate (indirectly to acetone in D_2O : δ 2.225 ppm) (14).

RESULTS

The carbohydrate composition of human seminal γ-Sm and the glycopeptide mixture derived from this protein (Table I) suggested the presence of N- and O-glycosidically[2] linked oligosaccharides as both Man and GalNAc were found to be constituent monosaccharides. Fractionation of the glycopeptide mixture on Concanavalin A-Sepharose afforded 3 heterosaccharide containing subfractions (Con A-1, Con A-2 and Con A-3) whose compositions are also listed in Table I.

Structure of the Glycans of Fraction Con A-1. The 500 MHz ^{1}H-NMR spectrum of this fraction (Fig. 1) revealed the main components to be tri- and tetra-antennary N-glycosidic heteroglycans. This could be deduced from the pattern of the Man H-2 signals, particularly from the major signal at δ 4.214 which is attributed to the coinciding H-2 resonances of Man $\underline{3}$ and $\underline{4}$ (compare ref. 14). Fuc was found to occur in α(1→6) linkage of GlcNAc $\underline{1}$ of all glycans as revealed by the singularity of the NAc signal of GlcNAc $\underline{2}$ at δ 2.093. Furthermore, this α(1→6)Fuc has its H-1 signal at δ 4.866. Combination of the sugar analysis and the ^{1}H-NMR spectroscopic data shows the ratio of tri- to tetra-antennas to be approximately 1:3 which is in agreement with the amount of Gal reported in Table I and with the intensity ratio of the Man $\underline{4}'$ H-1 signals at δ 4.920 (tri-antenna) and δ 4.866 (tetra-antenna) and H-1 of Fucα(1→6) being 1:7. In ad-

[2]The structure of the O-glycosidic glycans present in fraction Con A-1 could not be deduced from the ^{1}H-NMR spectrum, but will be described at a later date.

Table I

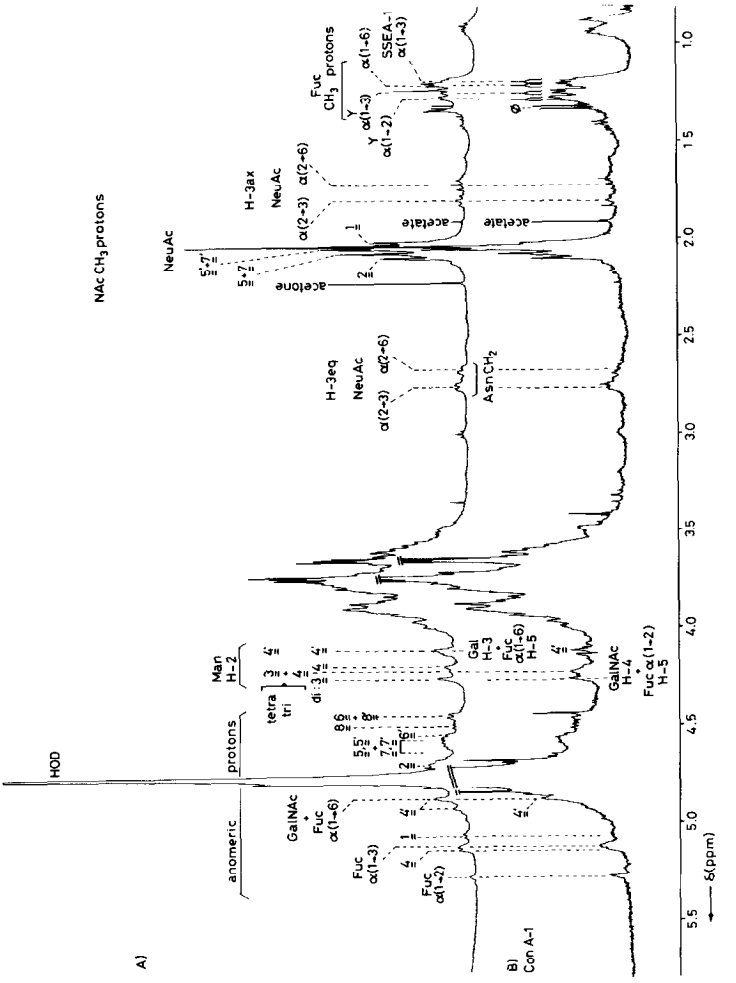
THE CARBOHYDRATE COMPOSITION OF HUMAN γ-SEMINOPROTEIN, THE GLYCOPEPTIDE MIXTURE AND THE CONCANAVALIN A-SEPARATED GLYCOPEPTIDE FRACTIONS DERIVED FROM THIS PROTEIN

Monosaccharide	Native Glycoprotein[c,d]	Glycopeptide Mixture[c]	Subfraction[c]		
			Con A-1	Con A-2	Con A-3
Man[a]	3.0	3.0	3.0	3.0	3.0
Gal	2.3	2.7	3.8	1.8	1.8
GlcNAc[b]	3.6	3.3	4.5	3.7	3.5
Fuc	1.3	2.2	3.6	1.0	0.9
GalNAc[b]	0.4	0.6	1.8	0.3	-
NeuAc	2.2	1.4	2.0	2.1	1.9

[a]Taken as 3.0. [b]Uncorrected. [c]Each preparation contained small amounts of Glc as impurity. [d]The total carbohydrate content of the protein accounted for approximately 12%.

dition to the mentioned two glycans small amounts of a di-antennary glycopeptide was also present in fraction Con A-1. Evidence for the latter compound stems from the set of the low intensity signals at δ 4.248 [Man $\underline{3}$] and 4.198 [Man $\underline{4}$] (Fig. 1). Independent proof of the presence of this compound was obtained by its removal from this fraction by affinity chromatography on lentil lectin. The ^{1}H-NMR spectrum (not shown) of the non-retained glycopeptide fraction was found to lack specifically the signals characteristic of a di-antenna (see below for fractions Con A-2 and Con A-3).

As to the peripheral region of the tri- and tetra-antennary glycans, a striking microheterogeneity of the NeuAc and Fuc linkages and branch locations was noted. NeuAc occurs in α(2→3) and α(2→6) bonds to Gal in a ratio of 1:1. Fuc residues occur, in addition to that linked to the core mentioned above, in two other types of linkages, namely α(1→2) linked to Gal and α(1→3) linked to peripheral GlcNAc residues and, thus, form part of either one of the following two oligosaccharide structures: 1. only one Fuc is present and is linked to a peripheral GlcNAc residue constituting the so-called SSEA-1 (X or Le^x) determinant, i.e. Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→·) (15,16), and 2. two Fuc residues are present and are linked to the same NAc-lactosamine residue constituting the Y (Le^y) determinant (for structure see Summary) (17). Evidence for the presence of SSEA-1 is afforded by the Fuc H-1 signal at δ 5.109 in combination with its CH_3 signal at δ 1.176 (14). The occurrence of the Y determinant is


A)
B) Con A-1
HOD
anomeric protons
NAc CH3 protons
NeuAc
H-3ax NeuAc
H-3eq NeuAc
Man H-2
tetra
tri
di
Gal H-3
Fuc α(1-6) H-5
GalNAc H-4
Fuc α(1-2) H-5
GalNAc + Fuc α(1-6)
Fuc α(1-3)
Fuc α(1-2)
α(2-3)
α(2-6)
acetone
acetate
AsnCH2
Fuc CH3 protons
α(1-6)
SSEA-1 α(1-3)
α(1-3)
α(1-2)
δ(ppm)

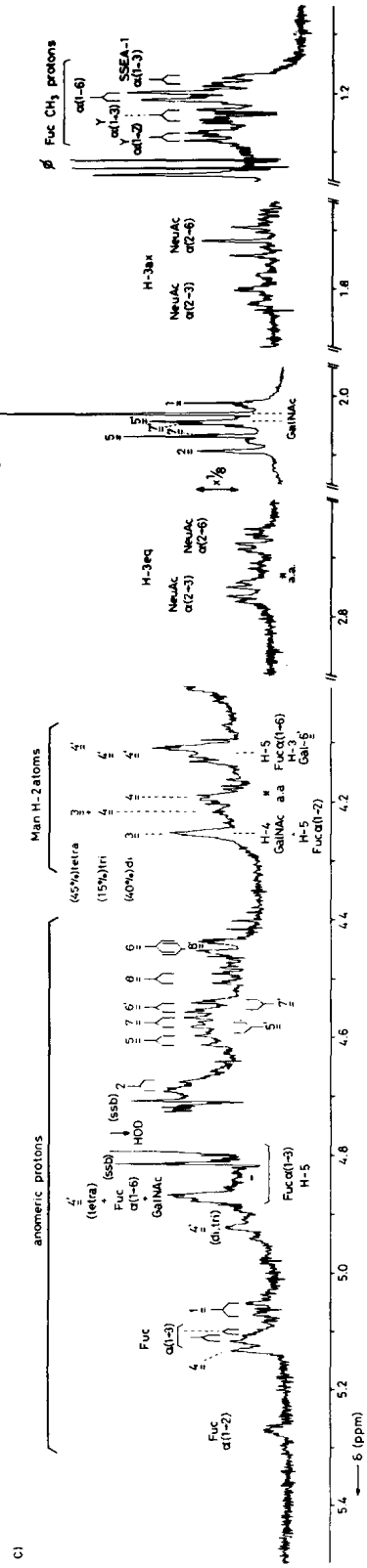


Fig. 1. 500-MHz ^{1}H-NMR spectra (D_2O, 27°C, pD 7) of the glycopeptide mixture (A) and fraction Con A-1 (B) derived from human γ-seminoprotein. In (C) the resolution-enhanced expansion of spectrum (A) shows the structural-reporter-group resonances in detail. The numbers indicated in the spectra refer to the corresponding residues of the structures given in Fig. 2. Although the tri- and tetra-antennas are microheterogeneous in the branch location of the peripheral sugars, we indicated here(for reasons of clarity) only the assignments for one particular isomer, namely that containing the Y determinant at Gal $\underline{8}$'-GlcNAc $\underline{7}$', NeuAcα(2→6) at Gal $\underline{6}$, NeuAcα(2→3) at Gal 6' $\underline{\equiv}$ and SSEA-1 Fuc at GlcNAc $\underline{7}$. The doublet at γ 1.32, marked by ϕ, stems from a non-protein, non-carbohydrate contaminant.

revealed by the Fuc H-1 signals that are located at δ 5.267 [Fucα(1→2)] and 5.109 [Fucα(1→3)] together with the CH_3 signals at δ 1.273 and 1.234, respectively. The intensity ratio of these four signals was found to be 1:1:3:3. Moreover, one of the Gal H-1 signals was observed at δ 4.497 confirming the presence of one Y determinant per molecule. The chemical shift values of the reporter-group signals typical of the Y determinant agree with those reported by Hindsgaul et al. (18). It should be noted that the H determinant sequence, i.e. Fucα(1→2)Galβ(1→4)GlcNAcβ(1→·), could be ruled out as no Fuc H-1 signal was found at δ 5.31 (compare ref. 19). On the average, each of the tri- and tetra-antennas possesses one SSEA-1 and one Y determinant along with a Fuc residue at GlcNAc 1 agreeing with the carbohydrate analysis (Table I). As to the branch location of the various terminating residues, it is of interest to note that the Gal 8-GlcNAc 7 branch has either the SSEA-1 or the Y determinant as evidenced by the resonance position of the NAc signal of GlcNAc 7 (δ 2.069). The Gal 6-GlcNAc 5 branch of almost all compounds possesses a NeuAc residue in α(2→6) linkage to Gal. This is based on the position of the H-1 signal of Man 4 (δ 5.129) in conjunction with the position of the NAc signal of GlcNAc 5 (δ 2.067). Regarding the other two branches, NeuAc is α(2→3) linked to Gal 6' or 8' and Fuc to GlcNAc 7' or 5', respectively. On the basis of the NMR data we are not able to distinguish between these two possibilities, but the simultaneous presence of NeuAcα(2→3) and Fucα(1→3) in one branch can be excluded (compare ref. 20). The results discussed above suggest the comprehensive structure for the tri- and tetra-antennary glycans of the Con A-1 glycopeptides presented in Fig. 2. If the 8-7 branch possesses the SSEA-1 determinant, the Y determinant is found on the 6' - 5' or 8' - 7' branch and vice versa.

The Glycans of Fractions Con A-2 and Con A-3. The 500-MHz spectra of these fractions showed the typical features of a di-antennary glycopeptide of the N-glycosidic type. This di-antenna (data not shown) has the same structure as that of horse pancreatic ribonuclease (14). It should be added that the separation of these two fractions from each other is probably due to differences in the peptide moiety of these glycopeptides.

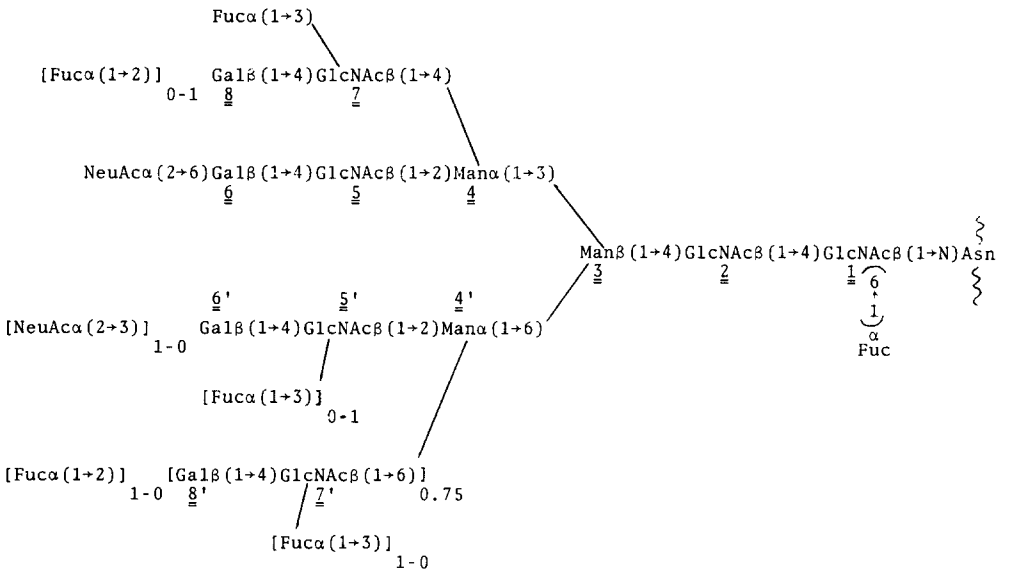


Fig. 2. The structures of the N-glycosidic tri- and tetra-antennary carbohydrate chains of human γ-seminoprotein. Note, the tri-antennary structure lacks the 8' - 7' branch and accounts for 25% of the mixture of these carbohydrate chains.

DISCUSSION

From the ^{1}H-NMR data presented above it could be deduced that approximately 40% of the glycan molecules of γ-Sm possesses a di-antennary, 15% a tri-antennary and 45% a tetra-antennary structure. Further, the carbohydrate content and the molecular weight of this protein together with the molecular weights of the glycans indicate the presence of one N-glycosylation site per molecule only. That glycans of different sizes are attached to a single glycosylation site of a molecular species is a finding that is very similar to the observation we reported earlier for human blood plasma α_1-acid glycoprotein (21,22). As to the Y determinant which has so far been reported to be present in glycolipids (17), mucins (23), milk (24) and urinary (25) oligosaccharides, it is of interest to note that this paper describes for the first time the occurrence of this determinant in N-glycans of a glycoprotein.

ACKNOWLEDGMENTS

The authors thank Dr. J.H.G.M. Mutsaers for excellent assistance in the lentil-lectin purification. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands

Organization for Advancement of Pure Research (ZWO), grant UUKC-OC 83-13 from the Netherlands Foundation of Cancer Research (KWF) and a grant from the National Institutes of Health (GM-10374), USPHS.

REFERENCES

1. Edwards, J.J., Tollaksen, S.L., and Anderson, N.G. (1981) Clin. Chem. 27, 1335-1340.
2. Wang, M.C., Kuriyama, M., Papsidero, L.D., Loor, R.M., Valenzuela, L.A., Murphy, G.P., and Chu, T.M. (1982) Meth. Cancer Res. 19, 179-197.
3. deJong, F.H. (1979) Mol. Cell Endocrinol. 13, 1-10.
4. Kovary, P.M., Dykers, A., and Niermann, H. (1977) Int. J. Fertil. 22, 251-254.
5. Lulac, J., and Koren, E. (1979) J. Reprod. Fertil. 56, 501-506.
6. Blake, E.T., and Sensabaugh, G.F. (1978) J. Forensic Sci. 23, 717-729.
7. Koyanagi, Y. (1974) Igaku Kenkyo (Acta Medica) 44, 69-88.
8. Tsuda, R., Inoue, T., and Hara, M. (1982) Jap. J. Legal Med. 36, 703-709.
9. Gaensslen, R.E. (1983) Sourcebook in Forensic Serology, Immunology and Biochemistry, U.S. Department of Justice, National Institute of Justice, Washington, D.C.
10. Tsuda, R., Hara, M., and Inoue, T. (1983) Jap. J. Legal Med. 37, 336-341.
11. Matsuzawa, S., Itoh, Y., Miyauchi, C., Hara, M., Inoue, T., and Tsuda, R. (1982) J. Forensic Sci. 27, 848-854.
12. Okabe, T., and Eto, K. (1983) Jap. J. Urol. 74, 1313-1319.
13. Kamerling, J.P., and Vliegenthart, J.F.G. (1982) Cell Biol. Monogr. 10, 895-898.
14. Vliegenthart, J.F.G., Dorland, L., and van Halbeek, H. (1983) Adv. Carbohydrate Chem. Biochem. 41, 209-374.
15. Hakomori, S., Nudelman, E., Levey, S.B., and Kannagi, R. (1984) J. Biol. Chem. 259, 4672-4680.
16. Gooi, H.C., Feizi, T., Kapadia, A., Knowles, B.B., Solter, D., and Evans, M.J. (1981) Nature 292, 156-158.
17. Abe, K., McKibbin, J.M., and Hakomori, S. (1983) J. Biol. Chem. 258, 11793-11797.
18. Hindsgaul, O., Norberg, T., LePendu, J., and Lemieux, R.U. (1982) Carbohydr. Res. 109, 109-142.
19. Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G., Kochetkov, N.K., Arbatsky, N.P., and Derevitskaya, V.A. (1982) Eur. J. Biochem. 127, 21-29.
20. Lamblin, G., Boersma, A., Klein, A., Roussel, P., van Halbeek, H., and Vliegenthart, J.F.G. (1984) J. Biol. Chem. 259, 9051-9058.
21. Fournet, B., Montreuil, J., Strecker, G., Dorland, L., Haverkamp, J., Vliegenthart, J.F.G., Binette, J.P., and Schmid, K. (1978) Biochem. 17, 5206-5214.
22. Schmid, K., Binette, J.P., Dorland, L., Vliegenthart, J.F.G., Fournet, B., and Montreuil, J. (1979) Biochim. Biophys. Acta 581, 356-359.
23. Hounsell, E.E., and Feizi, T. (1982) Med. Biol. 60, 227-236.
24. Kobata, A., Yamashita, K., and Tachibana, Y. (1978) Methods Enzymol. 50, 216-220.
25. Strecker, G., and Montreuil, J. (1979) Biochimie 61, 1199-1246.