Determination of Glycopeptide Primary Structure by 360-MHz Proton Magnetic Resonance Spectroscopy†

Jeremy P. Carver* and Arthur A. Grey

ABSTRACT: A detailed analysis of the proton magnetic resonance spectral parameters for the anomeric and C2 hydrogen resonances of 63 different glycopeptides and oligosaccharides of known structure reveals a general method for the determination of the primary structure of glycopeptides for most currently known classes of structures. In particular, a two-dimensional display formed by plotting mannosyl C1–H vs. C2–H chemical shifts demonstrates that these pairs of values are sensitive to long-range perturbation by remote substitution by hexoses as well as to direct substitution effects. A total of 41 C1–H/C2–H chemical shift clusters have been defined which characterize unique structural microenvironments. On this basis, the sequence and branching pattern for most structures can be derived. Corroborative evidence is obtained from an examination of the chemical shifts of the galactosyl and *N*-acetylglucosaminyl anomeric hydrogens as well as other features of the spectrum.

The chemical determination of the primary structure of the complicated carbohydrates of membrane and serum glycoproteins is time consuming and open to error on several grounds. The advent of high-field proton magnetic resonance spectrometers with their enhanced sensitivity has led to the application of NMR techniques to this problem. The purpose of this paper is to show that for all classes of structures of Asn-linked glycopeptides so far described, NMR provides a rapid, sensitive, and nondestructive alternative to chemical methods. The initial promise of the method, which was suggested by the comparison of glycopeptide spectra with model di- and trisaccharides, has been borne out spectacularly, as witnessed by the pioneering work of the European group (Dorland et al., 1977a,b, 1978a,b, 1979; Fournet et al., 1978; Schut et al., 1978; Strecker et al., 1977), Ballou and colleagues (Lehle et al., 1979; Cohen & Ballou, 1980), and the results from this laboratory (Narasimhan et al., 1980; Atkinson et al., 1981; Carver et al., 1981). In the following, we present the results of an analysis of the spectral parameters of 63 different compounds. Sixty-five percent of these data were obtained in this laboratory and have been (Narasimham et al., 1980; Atkinson et al., 1981; Carver et al., 1981), or are to be, reported in detail elsewhere. The remaining data have been taken from the literature (Dorland et al., 1978a,b, 1979; Fournet et al., 1978; van Halbeek et al., 1980a,b). Since the latter were obtained at room temperature, we present here only the data which we have obtained at that temperature. Cohen & Ballou (1980) have published data obtained at 40 °C. We find that, at 360 MHz, there is a highly significant temperature dependence in the spectral parameters for many glycopeptides (Grey et al., 1979) and have concluded, therefore, that it is only valid to make comparisons at the same temperature. Thus, the data from Cohen & Ballou (1980) have not been included in this analysis.

The major conclusion of this study is that their exist as many as 41 well-defined chemical shift values for specific "microenvironments" which span several hexose residues. This observation has two important implications. First, it establishes a method for determining the *sequence* of sugars and linkages within the glycopeptides rather than simply supplying a table of linkages which are present. Second, it suggests that these glycopeptides have well-defined three-dimensional structures which give rise to very specific long-range perturbations of the chemical shifts of anomeric hydrogens of certain residues. The precise origin of these perturbations and the subsequent utilization of these effects to deduce three-dimensional structural information will clearly require considerable additional investigation.

Materials and Methods

Materials

Glycopeptides were generally obtained by column chromatographic fractionation (Atkinson et al., 1981) or high-voltage paper electrophoresis in borate buffer (Narasimhan et al., 1980) after Pronase digestion of purified glycoproteins. In some cases, homogeneous glycopeptides were subjected to enzymatic degradation with purified endo- and/or exoglycosidases. The details of each preparation and the methods used to characterize each glycopeptide are given in the references indicated.

Nine complex biantennary glycopeptides (structures **1–9**, Table I) from human myeloma IgG(γ1) (Tem) were prepared according to Narasimham et al. (1979) and their NMR spectra characterized (A. A. Grey, S. Narasimhan, H. Schachter, and J. P. Carver, unpublished results). Four bisected, biantennary, complex glycopeptides (structures **64**, **65**, **66**, and **76**, Table I) of human myeloma IgG(γ1) (Tem and Hom) were similarly prepared (Narasimhan et al., 1980; G. Savvidou, K. J. Dorrington, A. A. Grey, and J. P. Carver, unpublished results). An additional neutral biantennary complex glycopeptide (compound **99**) with the opposite asymmetry from the above was also obtained from human myeloma IgG (Tem). Compound **97** was generated from the latter by enzymatic removal of the terminal GlcNAc (J.-R. Brisson, A. E. MacKenzie, and J. P. Carver, unpublished results).

The sialylated biantennary complex glycopeptide from transferrin (compound **78**) and the asialo counterpart (compound **98**) were prepared by Gleeson (P. Gleeson and H. Schachter, unpublished results). Compounds **61–63** were

† From the Departments of Medical Genetics and Medical Biophysics and the Toronto Biomedical NMR Centre, University of Toronto, Toronto, Canada M5S 1A8. *Received January 6, 1981; revised manuscript received June 25, 1981.* Presented, in part, at the Canadian Biochemical Society Symposium on the Biochemistry of Development, April 6–11, 1980, Banff, Alberta, Canada. This research was supported by grants from the Medical Research Council of Canada (MA-6499 and MT-3732) and the National Cancer Institute of Canada.

0006-2960/81/0420-6607$01.25/0

Table I: Structures of Compounds[a]

1	G4GN2M6[S6G4GN2M3]M4GN4[F6]GN-Asn
2	G4GN2M6[G4GN2M3]M4GN4[F6]GN-Asn
3	GN2M6[GN2M3]M4GN4[F6]GN-Asn
4	M6[M3]M4GN4[F6]GN-Asn
5	M6[GN2M3]M4GN4[F6]GN-Asn
6	M6[G4GN2M3]M4GN4[F6]GN-Asn
7	M6[S6G4GN2M6]M4GN4[F6]GN-Asn
8	GN2M6[S6G4GN2M3]M4GN4[F6]GN-Asn
9	GN2M6[G4GN2M3]M4GN4[F6]GN-Asn
10, 22, 98	G4GN2M6[G4GN2M3]M4GN4GN-Asn
11, 78	S6G4GN2M6[S6G4GN2M3]M4GN4GN-Asn
12	S3G4GN2M3M4GN
13	S6G4GN2M3M4GN
14	M6[S6G4GN2M3]M4GN
15	G4GN2M6[S3G4GN2M3]M4GN
16	G4GN2M6[S6G4GN2M3]M4GN
17	S3G4GN2[S6G4GN2]M3M4GN
18	S3G4GN2M6[S3G4GN2M3]M4GN
19	S3G4GN2M6[S6G4GN2M3]M4GN
20	S6G4GN2M6[S6G4GN2M3]M4GN
21	S6G4GN2M6[S6G4GN2[S3G4GN4]M3]M4GN
23–27	G4GN2M6[G4GN2[G4GN4]M3]M4GN4GN-Asn
28	G4GN2M6[G4GN2[G4[F3]GN4]M3]M4GN4GN-Asn
29–33	G4GN2[G4GN6]M6[G4GN2[G4GN4]M3]M4GN4GN-Asn
34–37	G4GN2[G4GN6]M6[G4GN2[G4[F3]GN4]M3]M4GN4GN-Asn
38, 39	M3[M6]M6[M2M3]M4GN
40, 41	M3[M6]M6[M3]M4GN
42, 43	M6[M3]M6M4GN4GN
44, 45	M3M6M4GN
46, 88, 89	M2M2M3M4GN
47, 86, 87	M2M3M4GN
48, 84, 85	M3M4GN
49, 50	M4GN
51	M4M
52	M6M6M
53	M3M2M2M
54, 83	M3[M6]M6[M3]M4GN4GN-Asn
55	M3[M6]M6[M2M3]M4GN4GN-Asn
56	M3M6[M3]M4GN4GN-Asn
57	M3[M6]M6[GN2[G4GN4]M3][GN4]M4GN4GN-Asn
58	M3[M6]M6[GN2M3][GN4]M4GN4GN-Asn
59, 60	M3[M6]M6[GN2[G4GN4]M3][GN4]M4GN
61, 62	M3[M6]M6[GN2M3]M4GN4GN-Asn
63	M6[GN2M3]M4GN4[F6]GN-Asn
64	S6G4GN2M6[S6G4GN2M3][GN4]M4GN4[F6]GN-Asn
65	G4GN2M6[S6G4GN2M3][GN4]M4GN4[F6]GN-Asn
66	M6[GN2M3][GN4]M4GN4[F6]GN-Asn
67, 68, 77	GN2M6[GN2[GN4]M3][GN4]M4GN4GN-Asn
69, 75	GN2M6[GN2M3][GN4]M4GN
70, 71	M2M6[M3]M6[M2M3]M4GN4GN-Asn
72, 92, 93	M2M6[M2M3]M6[M2M2M3]M4GN
73	M3[M6]M6[GN4][GN2[GN4]M3]M4GN4GN-Asn
74	G4GN2M6[G4GN2[G4GN4]M3]M4GN4GN-Asn
76	GN2M6[GN2M3][GN4]M4GN4[F6]GN-Asn
79	GN2M6[GN2M3][GN4]M4GN4GN-Asn
80	M3M6[GN4][GN2M3]M4GN4GN-Asn
81	M6M4GN4GN-Asn
82, 94	M3[M6]M6M4GN4GN-Asn
90, 91	M6[M2M3]M4GN
95	M3M6M4GN4GN-Asn
96	GN2M6[GN2M3][GN4]M4GN4[F6]GN-Asn
97	G4GN2M6[M3]M4GN4[F6]GN-Asn
99	G4GN2M6[GN2M3]M4GN4[F6]GN-Asn

[a] S3 = αNeuNAc(2,3), S6 = αNeuNAc(2,6), G4 = βGal(1,4), GN2 = βGalNAc(1,2), GN4 = βGlcNAc(1,4), GN6 = βGlcNAc(1,6), M3 = αMan(1,3), M4 = βMan(1,4), M6 = αMan(1,6), F3 = αFuc(1,3), and F6 = αFuc(1,6).

prepared enzymatically as described by Harpaz & Schachter (1981) and their structures confirmed by NMR (Narasimhan et al., 1980). Compound **96** was also prepared enzymatically (S. Narasimhan and H. Schachter, unpublished results).

Compounds **54**, **55**, **57**, and **58**, as well as the endoglycosidase H derivatives of the first three (**41**, **39**, and **59** (**60**)), were prepared from chicken ovalbumin as described by Atkinson et al. (1981). Compound **57** was digested with β-galactosidase to produce compound **73** (Atkinson et al., 1981) and compound **58** was exhaustively digested with α-mannosidase to yield **80** (P. H. Atkinson, unpublished results). Three minor components (compounds **56**, **70** (**71**), and **77**) were also obtained from Pronase digests of ovalbumin (Atkinson et al., 1981; Narasimhan et al., 1980). Compounds **56** and **70** (**71**) have been previously reported and designated III-B and VI, respectively (Tai et al., 1977a,b). A previously undescribed component was tentatively assigned to be either **77** or **79** [Table I, Narasimhan et al. (1980)]. The former bisected triantennary complex structure was subsequently found to be the correct one by comparison with the data obtained for the chicken ovotransferrin glycopeptide (Dorland et al., 1979). Compound **55** was subjected to partial degradation with purified α-mannosidase, and the resulting $Man_4GlcNAc_2Asn$ (compound **94**), $Man_3GlcNAc_2Asn$ (compound **95**), and the corresponding endoglycosidase H products (compounds **43** and **45**) were obtained (P. H. Atkinson, J. P. Carver, C. Ceccarini, A. A. Grey, J. Hakimi, and J. Tsang-Lee, unpublished results).

The linear oligosaccharides from mannosidosis urine [compounds **46**–**48**, described by Norden et al. (1974)] were obtained from A. Lundblat (courtesy of H. Schachter). βMan(1,4)GlcNAc (compound **50**) was obtained after exhaustive α-mannosidase digestion of the endoglycosidase H product from ovalbumin D3 [compound **55**; Carver et al. (1981)]. Other mannose-containing oligosaccharides were obtained from various sources: βMan(1,4)Man (compound **51**) from A. Perlin and αMan(1,3)αMan(1,2)αMan(1,2)Man (compound **53**) and αMan(1,6)αMan(1,6)Man (compound **52**) from C. Ballou (courtesy of P. H. Atkinson and H. Schachter). The oligosaccharide from the urine of a patient with Sandhoff disease [compound **69** (**75**); Strecker et al., (1977)] was obtained from G. Strecker (courtesy of H. Schachter). The unit A oligosaccharide from bovine thyroglobulin [compound **72**; Ito et al. (1977)] and the fetuin glycopeptide [compound **74**; Nilsson et al. (1979)] were kindly donated by A. Adamany. The former was further fractionated by P. H. Atkinson. Also added to the data set were chemical shift data for an additional 16 structures (compounds **12**–**21**, **28**–**37**, **81**, **84**, and **90**) reported by Vliegenthart and his colleagues (Dorland et al., 1978a,b, 1979; Fournet et al., 1978; van Hallbeek et al., 1980a,b). The chemical shifts for the corresponding α anomers (compounds **85**, **87**, **89**, **91**, and **93**) were not used in this analysis. Data, from the same papers, for eight other structures (**10**, **22**–**27**, **67** (**68**), **82**, **83**, **86**, **88**, and **92**) were included. The chemical shift values for these compounds had also been determined as part of our studies; however, both sets of data have been included since they represent independent measurements.

Methods

A. Sample Preparation. Divalent heavy metal ions were removed from aqueous solutions of glycopeptides by column chromatography on Chelex-100 (200–400 mesh, Bio-Rad Laboratories). Samples were exchanged twice with deuterium oxide (99.7 atom % D; Merck, Sharp and Dohme, Montreal, Canada) and were redissolved in an appropriate volume of deuterium oxide (100.0 atom % D; Stohler Isotope Chemicals, Waltham, MA) to give a concentration between 1 and 20

mM: 400–500 μL for examination in a 5-mm tube (catalog 528PP; Wilmad Glass Co., Inc., Buena, NJ) or 100 μL for examination in a microcell (catalog 529-E; Wilmad Glass Co.). Acetone was added to roughly equimolar concentration as an internal chemical shift reference. Carbon tetrachloride was used as the spectrally transparent medium for suspension of microcells in the 5-mm tubes.

B. Instrumental Conditions. Proton magnetic resonance spectroscopy was performed mostly by using the 360-MHz facility at the Toronto Biomedical NMR Centre. Some early spectra were obtained by using the 360-MHz facilities of the Stanford Magnetic Resonance Laboratory, Brookhaven National Laboratories (with the kind permission of Dr. Alan McLaughlin), and the Purdue University Biological Magnetic Resonance Laboratory. Spectra were obtained by Fourier transformation of data: 16–512 pulses, 16K data points, 5–6-s pulse delays, and 2500-Hz spectral widths. Samples were examined at 23 ± 1 °C. Chemical shifts were calculated by setting the resonance from internal acetone at 2.225 ppm (relative to internal DSS[1]).

C. Nomenclature. Although the glycopeptide structures which we shall be discussing are complicated, they can be classified into relatively few characteristic classes on the basis of the branching pattern which is present. As more and more different structures have been elucidated, the original classification of glycopeptide structures into complex (or acidic) and high mannose (or neutral) types has been expanded. Thus, subcategories of bi-, tri-, and tetraantennary structures have been established within the complex class (Fournet et al., 1978), and an additional class termed "hybrid" has been identified (Yamashita et al., 1978).

In our laboratory we have been primarily concerned with the glycopeptides from hen ovalbumin and human myeloma IgG. Among these occur several structures in which the β-linked mannose of the trimannosyl core[2] is substituted at C4 by βGlcNAc. Because the usual mannose substituents of the β-linked mannose are at C3 and C6, we propose that the βGlcNAc at C4 be termed "bisecting" and structures bearing this substitution be termed "bisected". Thus, one may identify structures as "bisected triantennary complex", "bisected hybrid", etc. The complete set of classes reported to date is summarized in Table II along with schematic representations of the structures.

In the following we will be presenting evidence that at least 41 different hexose microenvironments can be identified from their characteristic chemical shift values. By "hexose microenvironments" we mean the particular linkage and anomeric configuration *together with* those substituent hexoses which are required to be present to generate the particular chemical shift values which we have associated with that microenvironment. Thus, it has been possible to identify which substituent patterns lead to significant perturbations of the chemical shifts for the substituted residue and therefore to detect the presence of these microenvironments. To aid discussion of these microenvironments, we have devised a shorthand notation which describes the linkage and substitution of the residue under discussion. We confine this notation to the mannose residues since the branching pattern of a glycopeptide is defined by the substituents of these residues alone. The linkage is represented by a single number, i.e., 4 for β(1,4), 6 for α(1,6), etc., to which a "t" or "i" is added to indicate whether the residue is terminal (nonreducing) or internal, respectively; i.e., 4i, internal β(1,4)-linked, etc. The anomeric configuration of the hexose has been omitted to simplify the symbolism. In the cases discussed below it is clear from the context whether the linkage is α or β. Finally, since the mode of substitution is often important, a parenthetical modifier is added as follows: 4i(3,6), 4i(3,4,6), and 3i(2M) to represent respectively a β(1,4)-linked residue which is substituted at the 3 and 6 positions, a β(1,4)-linked residue substituted at the 3, 4, and 6 positions (nature of the substitution unspecified), and an α(1,3)-linked residue substituted at the 2 position by an α-linked mannose. The latter distinction is necessary because of the special effects of this substitution.

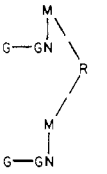
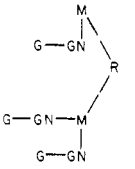
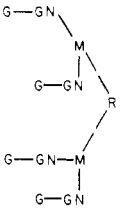
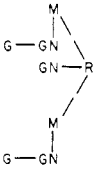
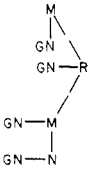
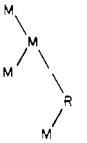
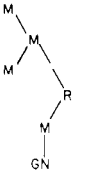
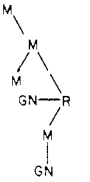
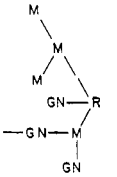
Results and Discussion

Determination of the Anomeric Linkages. The use of high-resolution NMR for the determination of the structures of various synthetic and naturally occurring compounds has a long history. It has generally been confined to relatively low molecular weight compounds for which complete assignment of the spectrum can be made. For increasingly complex molecules, such as the glycopeptides dealt with in this study, the approach is to seek out those features of the spectrum which are "identical" with those of the lower molecular weight compounds of known structure (di- or trisaccharides, for example). When this can be done, a partial structure consistent with part of the spectrum can be assigned. Usually a unique structure emerges once all of the pieces are fitted together; however, in some instances redundancy remains. For the final structure to be deemed to have any credibility, it is necessary to know how unique individual chemical shifts are, and therefore how great is the certainty that other structures cannot give the same spectrum. The uniqueness of the chemical shifts could only be determned with certainty if the spectra of all possible sugar structures were available. Since they cannot be, it is necessary to generalize on what factors have been found to be important influences on chemical shifts in those models which have been studied so far. However, before entering into such considerations, it is necessary to clarify what the experimental errors are and hence how close two chemical shifts must be in order to be considered "identical". The precision with which a chemical shift can be measured depends on many factors, some of which are under operator control (digital resolution, field homogeneity, temperature stability). For the data reported in this paper, the accuracy is ±0.0014 ppm, corresponding to the limitations of digital resolution. The ability to reproduce any of the shifts quoted, however, is more like ±0.003 ppm. Fournet et al. (1978), in a comparison of the chemical shifts for the anomeric protons of mannosyl residues in a variety of structures, found a similar reproducibility. Thus, one is justified in considering two chemical shifts distinct if they differ by more than about 0.006 ppm. As will become clear, such precision is more than enough to make most assignments. For example, anticipating for a moment some of the results to be presented below (Table IV), the chemical shifts of the anomeric and C2 hydrogens of the core α(1,6) linked mannosyl residues, in 12 different structural environments, range from 4.866 to 5.015 ppm (anomeric) and 3.969

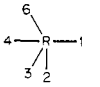
[1] Abbreviations used: DSS, sodium 4,4-dimethyl-4-silapentane-1-sulfonate; Man, D-mannopyranose; GlcNAc, 2-acetamido-2-deoxy-D-glucopyranose; Gal, D-galactopyranose; NeuNAc, *N*-acetyl-D-neuraminic acid; Fuc, L-fucopyranose. In many of the compounds examined the optical configuration of the hexoses has not been confirmed.

[2] All the Asn-linked glycopeptide structures found in Table I have a common structural element: $Man_3GlcNAc_2Asn$. We refer to the Man_3 component as the "trimannosyl core", the GlcNAc to which the Man_3 is linked as the "core GlcNAc", and the GlcNAc which is linked to the Asn as the "asparagine GlcNAc".

Table II: Saccharide Structures[a]

	class		example
I	biantennary complex		human serotransferrin[b] human IgG[c]
II	triantennary complex		fetuin;[d] VSV-G[e]
III	tetraantennary complex		human α_1-acid glycoprotein[f]
IV	bisected biantennary complex		human IgG[c] human Sandhoff urine[g]
V	bisected triantennary complex		chicken ovotransferrin[h]
VI	high mannose		chicken ovalbumin[i] calf thyroglobulin[j]
VII	biantennary hybrid		bovine rhodopsin[k]
VIII	bisected biantennary hybrid		chicken ovalbumin[l]
IX	bisected triantennary hybrid		chicken ovalbumin[m]

[a] Symbolism: M = αMan; GN = βGlcNAc; G = βGal; R = βMan(1,4)βGlcNAc(1,4)βGlcNAc-Asn or βMan(1,4)βGlcNAc(1,4)[αFuc(1,6)]-βGlcNAc-Asn. The linkages are represented by the orientation as shown



Several other possibilities have not been listed (e.g., bisected high mannose) because the corresponding structures have not been found. In addition, many variations of these structures occur which have residues deleted and/or fucosyl and/or NeuNAc residues attached at various points. [b] Dorland et al. (1977a). [c] A. A. Grey, S. Narasimhan, H. Schachter, and J. P. Carver, unpublished results. [d] Nilsson et al. (1979). [e] Reading et al. (1978). [f] Fournet et al. (1978). [g] Strecker et al. (1977). [h] Dorland et al. (1979). [i] Tai et al. (1975). [j] Ito et al. (1977). [k] Liang et al. (1979). [l] Tai et al. (1977a). [m] Yamashita et al. (1978).

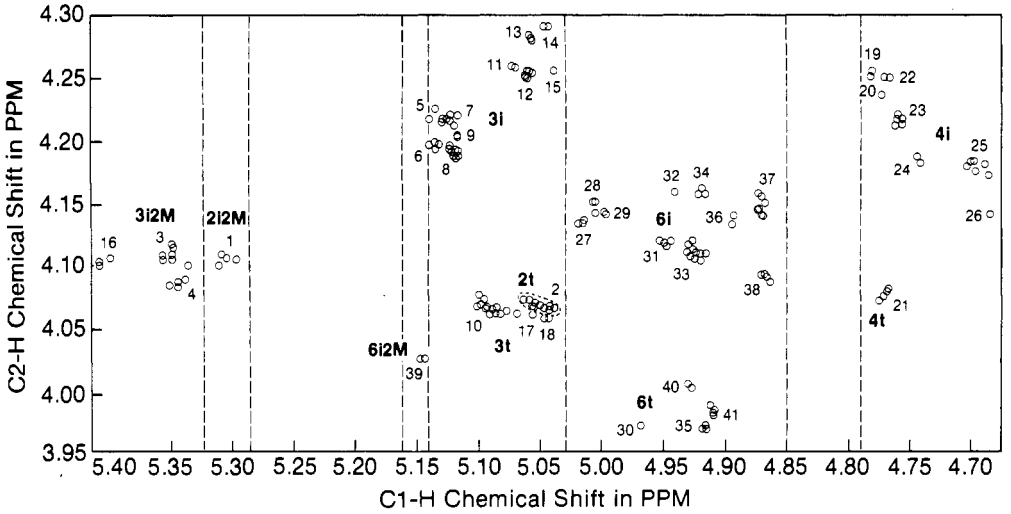


FIGURE 1: Mannose chemical shifts for C1–H plotted vs. those for C2–H. The C1–H axis is divided into nonoverlapping strips corresponding to 2i or 3i (2M) −5.41–5.03 ppm, 6i (2M) −5.14 ppm, 3i/3t −5.14–5.03 ppm, 6i/6t −5.02–4.85 ppm, and 4i/4t −4.79–4.67 ppm (for the meaning of the labels see the text). The 2t region falls in a pocket (5.05–5.02 ppm) within the 3i/3t strip and is shown by a dashed line. All clusters within the strips are numbered, and the corresponding data are summarized in Table IV. All chemical shifts are given in ppm relative to internal DSS using internal acetone set to 2.225 ppm (D_2O, 23 ± 1 °C).

Table III: Broad Regions of Chemical Shift Values Associated with Different Linkages of Mannosyl Residues[a]

	2-substituted[b]		others[c]		
				C2-H	
linkage	C1-H	C2-H	C1-H[d]	terminal	internal
α(1,2)	5.29–5.31 (4)[e]	4.10–4.11 (4)	5.03–5.07 (14)	4.06–4.08 (14)	
α(1,3)	5.33–5.41 (11)	4.09–4.12 (11)	5.03–5.14 (75)	4.05–4.08 (18)	4.18–4.29 (57)
α(1,6)	5.14–5.15 (4)	4.03 (4)	4.86–5.02 (75)	3.96–4.01 (15)	4.10–4.17 (60)
β(1,4)			4.68–4.79 (36)	4.07–4.08 (6)	4.14–4.26 (30)

[a] Chemical shifts are in ppm relative to internal DSS (using internal acetone at 2.225 ppm) in D_2O at 23 ± 1 °C. [b] 2-Substituted by α-mannose. [c] Substitution other than α(1,2) with mannose. [d] Both internal and terminal except those substituted α(1,2) with mannose. [e] The numbers in parentheses indicate the number of chemical shift values from the data base which fall into this category.

to 4.159 ppm (C2 hydrogen). The mean anomeric chemical shift value for each environment differs from each of the others by more than 0.006 ppm for all but 2 out of the 66 possible comparisons. For these two environments, however, the C2 hydrogen chemical shifts differ by more than 0.006; hence, an assignment can still be made unequivocally. Finally, it should be noted that the chemical shift range for the anomeric hydrogen resonance of these α(1,6)-linked mannosyl residues (0.149 ppm), although large compared with the precision of chemical shift measurement, still represents only 28% of the entire range of chemical shifts found in this study for the anomeric protons of α-linked mannosyl residues (0.539 ppm). It is for this reason that, despite the variability within each range due to substitution effects, the chemical shift ranges of the anomeric hydrogens of differently linked mannosyl residues are distinct one from another. The only exception to this statement is the effect of 2-substitution with α-linked mannosyl residues (see below).

Analysis of Mannosyl Chemical Shifts. C2–H mannose resonances are known to be shifted from their normal 3.9–4.0-ppm position to 4.10–4.30 ppm when certain substitution patterns are present (Carver et al., 1981; Gray & Ballou, 1971; Cohen & Ballou, 1980; Dorland et al., 1978b). Moreover, we had observed that in many cases *either* the C1–H *and/or* the C2–H resonance of a mannose is shifted depending upon the nature of the immediate or distal substituent hexoses, their number, and configuration as well as position of substitution. Dorland et al. (1978b) had already pointed out the value of these chemical shift comparisons in defining the branching pattern of glycopeptides of the complex type. We therefore examined all the available data on mannose C1–H/C2–H chemical shifts in an unbiased way in order to discover as many correlations between structural environments and C1–H/C2–H values as possible. The method which we have used is to plot all 230 C1–H/C2–H pairs of values obtained for the compounds listed in Table I in a two-dimensional plot. The structures of the compounds which generated points falling within 0.006 ppm of each other, on both axes, were then examined for common features. We report below the 41 correlations which we have found.

A. General Features of the C1–H vs. C2–H Plot. A graph of the C1–H chemical shift vs. the corresponding C2–H shifts is shown in Figure 1. The two-parameter correlation reveals detail at three different levels. First, with few exceptions, the C1–H chemical shift axis can be divided into nonoverlapping strips (separated by vertical dashed lines) corresponding to C1–H chemical shift values for mannosyl residues in a specific linkage, e.g., α(1,3), α(1,6), or β(1,4). Second, when one examines the distribution of C2–H chemical shift values within each strip, one finds that they fall into two broad nonoverlapping regions which correspond to either terminal or substituted mannosyl residues. Taken together with the three regions corresponding to the residues which are 2-*O*-substituted by mannose, a total of ten distinct regions can be identified. These are labeled 2t, 3i, 3t, 6i, 6t, 4i, 4t, 2i(2M), 3i(2M), and 6i(2M) in Figure 1. The corresponding chemical shift ranges are summarized in Table III. Finally, within each region, clusters of points are found with C1–H/C2–H chemical shift pairs which are identical within experimental error. The 41 clusters which have been identified are labeled in Figure 1 and

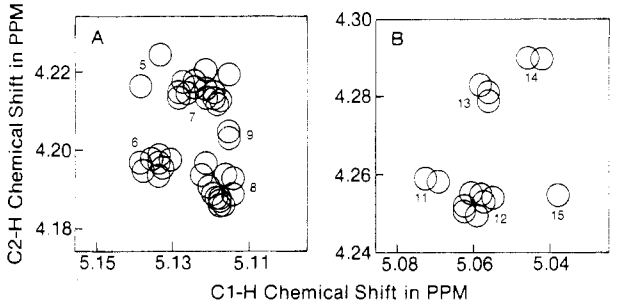


FIGURE 2: Expanded 3i region of Figure 1. Clusters are numbered according to Table IV. (A) Region corresponding to unbisected structures; (B) region corresponding to bisected structures.

the corresponding chemical shift data are given in Table IV. The 3i region is expanded in Figure 2. In each case, common structural features (microenvironments) can be found for the residues generating the C1–H/C2–H pairs falling within a single cluster.

The C1–H vs. C2–H plot clearly shows that a specific linkage and general substitution pattern [t, i, or i(2M)] can be assigned to a mannose with a high degree of certainty from its C1–H/C2–H coordinates. A precise "cluster" match allows further microenvironmental features about the mannose to be elucidated. For example, 5.123/4.216 ppm falls in the general 3i region, while the exact position matches cluster 7, that for a 3i(2,4) in a tri- or tetraantennary structure [i.e., 2,4-di-substituted with β-linked GlcNAc].

It is apparent from Figure 1 and Table IV that about half the clusters span more than 0.006 ppm on one or both axes. In some cases this arises from larger than normal experimental errors; for example, when peaks overlap, it is not possible to measure their chemical shift as accurately. In other cases this does not appear to be the cause, and one must assume that it is differences in structure which give rise to the range of chemical shifts observed. In order to identify what structural features of the microenvironments are responsible for these differences, it is necessary to have a series of closely related model compounds. In those cases where data for such a series are not available, the points in question have been left in a single cluster, despite the large range that ensues.

B. Patterns within Broader Regions. A systemic examination of each labeled region of Figure 1 reveals a pattern of cluster positions according to the branching and the terminal sialic acid substitution of the associated microenvironment. This is more evident in the expanded plot of the 3i region (Figure 2). The pattern is not precisely reproduced from one broad region to another: in some cases sufficient data are not available, and in others the corresponding structures are not possible. Moreover, identical patterns should not necessarily be expected for mannoses of different anomeric configuration, linkage, or even position within the structure. Some correlation could be expected if the substitution alters the chemical shift through a direct electrostatic effect and if the perturbations in secondary structures are sufficiently small to retain the overall three-dimensional structure. A detailed discussion of all the effects manifested in Figure 1 would appear premature at this time. The comments below are therefore confined to the 3i region, as an illustration of the general types of effects found.

In the area corresponding to unbisected structures (Figure 2A), six clusters (5–9) can be distinguished. For clusters 7 and 8, the ranges of chemical shift values included are large. For cluster 7, the data points arise from both tri- and tetraantennary structures. Although for most of these compounds the C1–H/C2–H pairs fall into two separate clusters according to their branching pattern, the two resulting ranges differ by only 0.006/0.005 ppm and therefore have not been reported separately. A similar situation exists for cluster 8 which arises from βGal and αNeuNAc(2,3) terminating biantennary structures. In Figure 2B, an expansion of the 3i region is shown which contains the points for the bisected structures. Five clusters (11–15) are found. There are not as many data points, but the same dependence on branching pattern and sialylation can be seen as in Figure 2A.

Several interesting conclusions arise from these data. First of all, it is clear that many factors influence the C1–H and C2–H chemical shift values. Substitution of the mannosyl residues, in general, results in a downfield shift of the C2–H resonance (clusters 10, 17, and 18 compared to clusters 5–9 and 11–15), but the extent varies with the microenvironment. In addition, 2,4-disubstitution has a larger effect than 2-substitution (clusters 5 and 7 compared to 6 and 8 or clusters 13 and 14 compared to 11 and 12). Secondly, the direct substitution pattern alone does not determine the chemical shift values since remote substitutions, such as the presence of a bisecting GlcNAc, have a large effect on the C1–H chemical shifts of the trimannosyl core anomeric hydrogen resonances (Figure 2, A vs. B). In addition, the chemical shifts of mannosyl residues show a microenvironment-dependent sensitivity to remote substitution with NeuNAc, the magnitude of which depends on whether the linkage is $\alpha(2,3)$ or $\alpha(2,6)$. For example, clusters 6 and 8, 5 and 7, or 11 and 12 differ by the presence of an αNeuNAc(2,6) terminus. Finally, the reproducibility of the perturbations in chemical shift values suggests that the microenvironment identified with each cluster corresponds to an ordered three-dimensional structure which extends over several hexose residues.

The general 6i region encompasses 4.85–5.02 and 4.08–4.16 ppm on the C1–H and C2–H axes, respectively. While some of the chemical shift perturbations described above for the 3i region are also observed here, different sensitivities to distal substitution arise, reflecting the inherently different nature of the linkage.

Finally, the mannoses involved in $\alpha(1,2)$ linkages form a special class. In 2t configurations they can be seen at the coordinates 5.049/4.068 ppm (cluster 2) in Figure 1 and Table IV. Although most are distinguishable by the experimental error criterion, some may be easily mistaken for 3t units (clusters 10, 17, and 18). Fortunately, substitution with an $\alpha(1,2)$-linked mannosyl residue on another mannose causes a dramatic anomeric hydrogen deshielding on the 2-*O*-substituted residue. As can be seen from Figure 1 and Table IV, all i(2M) residues studied to date, regardless of linkage, fall into unique clusters (1, 3, 4, 16, and 39). Thus the presence of a 2t unit can always be detected.

Several of the compounds used to generate the C1–H/C2–H chemical shift data base are oligosaccharides. In all these compounds a reducing terminal GlcNAc residue is present which exists in both α and β forms. In such cases two different anomeric and/or C2–H resonances may be observed for 3t and 4i units of some high mannose (class VI) oligosaccharides. The mannosyl resonances in question are present in the ratio of the α and β anomers of the terminating core GlcNAc. In those cases where such effects have been observed, the data were not included in the above analysis.

That the C1–H/C2–H resonances of 4i residues should be sensitive to the anomeric configuration of the core GlcNAc is not suprising. However, that 3t residues some 12–15 bonds distant from the anomeric of the core GlcNAc should be affected is unexpected. In a previous report (Grey et al., 1979)

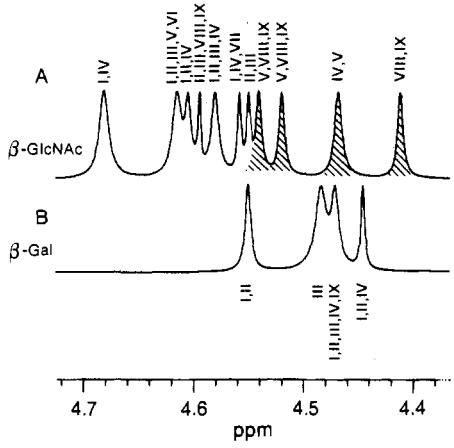


FIGURE 3: Representation of the anomeric chemical shifts of (A) βGlcNAc and (B) βGal. Peaks represent the center of the 7–9-Hz doublets of the β anomeric hydrogen resonances. Half-widths correspond to calculated standard deviations. Roman numerals associated with each peak indicate the structural classes which might give rise to the resonance position. The code for these classes is given in Table II. Chemical shift data are summarized in Tables V and VI. Cross-hatched peaks correspond to the resonances of those structures which contain a bisecting GlcNAc. All chemical shifts are relative to internal DSS calculated by using internal acetone set to 2.225 ppm (D_2O, 23 ± 1 °C).

we have postulated, on the basis of the temperature dependence of the C1–H/C2–H chemical shift values of ovalbumin D3 (compound **55**), the existence of a specific orientation for the 3t residue in the αMan(1,6) arm relative to the core GlcNAc residue which brings these residues almost into contact. This model would explain the sensitivity of the 3t anomeric hydrogen chemical shift to the anomeric configuration of the reducing GlcNAc.

Analysis of Gal and GlcNAc Chemical Shifts. The β-anomeric hydrogen resonances with J = 7–8 Hz (i.e., not mannosyl) in these compounds are all found in the 4.4–4.7-ppm region. The only exception is for a βGlcNAc unit whose C1 carbon bears an asparagine which produces a downfield shift to about 5.1 ppm (J = 9–10 Hz; Tanaka & Yamashina, 1973; Dorland et al., 1977b). Approximately 180 βGlcNAc and 100 βGal shifts were included in the analysis.

The data were analyzed by constructing a histogram of frequency of occurrence vs. chemical shift. The microenvironments of the hexoses giving rise to the chemical shift values within each peak of the histogram were examined for common features. A graphical representation was then constructed by using Gaussian peaks with positions and half-widths corresponding to the means and standard deviations of the chemical shift values belonging to that microenvironment. The summarized data are given in Tables V and VI and graphically represented in Figure 3.

While the Gal and GlcNAc C1–H chemical shift values overlap, certain other features in the spectrum usually permit these resonances to be assigned with certainty. For example, the presence of a bisecting βGlcNAc(1,4) can be unequivocally determined from the Man C1–H/C2–H values. Once this is done, only those GlcNAc chemical shift values for the appropriate classes of structures need be considered. In Figure 3 the peaks corresponding to structures with a bisecting GlcNAc are shaded. It can be seen immediately that half the peaks no longer need to be considered. As another example, structures containing terminal αNeuNAc(2,3) and αNeuNAc(2,6) residues can be identified because of the existence of characteristic resonances in other parts of the spectrum (see Other Regions). Thus, having detected the presence of a terminal NeuNAc residue, the range of microenvironments associated with a given chemical shift is significantly reduced.

Clearly, specific structural environments can create overlap between these two families, making assignments based on this region alone rather perilous. However, with or without overlap, when used in conjunction with the C1–H/C2–H plot, the Gal and GlcNAc anomeric hydrogen chemical shifts can substantiate assignments. An irreconciable difference in the interpretation of the chemical shift data from the two approaches would then suggest the existence of a new structure not represented in the basis set.

Other Regions. As mentioned previously, other distinguishing features in the spectrum are often available as an aid to the identification of the presence and linkage of certain residues. Used in conjunction with the anomeric hydrogen resonance shifts of the Man, Gal, and GlcNAc residues, these features frequently allow further specific assignments to be made.

A. N-Acetyl CH_3 Region. The methyl resonances of *N*-acetyl groups present in these compounds are found as singlets in the 2.00–2.12-ppm range. The quantitative aspects of the magnetic resonance method permit integration of this region relative to signals of known origin and number, thereby allowing the number of GlcNAc and sialic acid residues to be assessed. The chemical shifts of these resonances can also be correlated with their microenvironment in many cases (Dorland et al., 1978a, 1979; Fournet et al., 1978; Carver et al., 1981; Narasimhan et al., 1980).

B. N-Acetylneuraminic Acid. In addition to the presence of its *N*-acetyl methyl resonance mentioned above, axial and equatorial ring hexose hydrogens of C3 of sialic acid show multiplet resonances in the 1.7- and 2.7-ppm regions, respectively. The magnitudes of the coupling constants associated with these multiplets reflect their configuration within the molecule and are approximately as follows: J(3a,3e) = 11–12 Hz, J(3a,4a) = 11–12 Hz, and J(3e,4a) = 4–5 Hz. Thus, the low-field H3e resonance appears as a quartet while its axial geminal counterpart appears as a "triplet" at higher fields. Linkage information may be directly extracted from their chemical shifts as shown by Dorland et al. (1978) and Schut et al., (1978) and corroborated in our laboratory (A. A. Grey, K. A. Kronis, and J. P. Carver, unpublished results): αNeuNAc(2,6), C3–H_e 2.670 (0.005), C3–H_a 1.720 (0.001), and αNeuNAc(2,3), C3–H_e 2.758 (0.001), C3–H_a 1.799 (0.015) ppm. The values in parentheses are the ranges of observed chemical shifts.

With quantitative integration, the number of NeuNAc residues and their linkages may be quickly evaluated from the observed intensity of these resonances. Since the resonances at 2 ppm may arise from both NeuNAc and GlcNAc, an independent determination of the number of NeuNAc residues permits the number of GlcNAc residues to be deduced. Once this has been done, the number of Gal residues can be obtained since the anomeric resonances in the region 4.40–4.55 ppm arise from these two residues alone.

C. Fucose. Fucosyl residues have been described in two different linkages in Asn-linked glycopeptides (Montreuil & Vliegenthart, 1979): α(1,6) linked to the Asn-GlcNAc of the core and α(1,3) linked to an arm GlcNAc. The mean chemical shift values found in this study for the α(1,6)-linked fucose are 4.873 ± 0.007 (C1–H), 4.128 ± 0.004 (C5–H), and 1.203 ± 0.006 (C6–H) ppm. Fournet et al. (1978) have reported data for the α(1,3)-linked fucose of α_1-acid glycoprotein. The means of the reported chemical shifts are 5.111 ± 0.004

Table IV: Mannose Cl-H/C2-H Chemical Shift Pairs [a]

	carbon substitution [b]					Cl-H/C2-H shifts	class/type/ bisected [c]	Cl-H/C2-H [d] ranges	compound no.
	2	3	4	6					
					αMan(1,2)				
	arm units [e]								
1	M	H	H	H	2i (2M)	5.304/4.106	hm, ln	0.014/0.009	46, 72, 88, 90, 92
2	H	H	H	H	2t	5.049/4.068	hm, ln	0.025/0.008	46, 47, 55, 71, 72, 86, 88, 90, 92
					αMan(1,3)				
	core mannose [e]								
3	M	H	H	H	3i (2M)	5.352/4.110	hm, ln	0.011/0.012	47, 55, 71, 86, 90
4	M2M	H	H	H	3i (2M2M)	5.341/4.089	hm, ln	0.015/0.018	46, 72, 88, 92
5	S6G4GN	H	S3G4GN	H	3i (2, 4)	5.137/4.221	cx/tri	0.005/0.008	17, 21
6	S6G4GN	H	H	H	3i (2)	5.135/4.197	cx/bi	0.008/0.005	1, 7, 11, 13, 14, 16, 19, 20, 78
7	G4GN	H	G4GN	H	3i (2, 4)	5.123/4.216	cx/tri, tet	0.013/0.009	23–27, 29–34, 36, 37, 74
8	(S3G4)GN	H	H	H	3i (2)	5.119/4.190	cx/bi	0.008/0.010	2, 3, 5, 6, 9, 10, 12, 15, 18, 22, 63, 98, 99
9	GN	H	H	H	3i (2)	5.116/4.205	hy/bi	0.000/0.001	61, 62
10	H	H	H	H	3t	5.089/4.067	cx, hm, hy, ln	0.032/0.016	4, 43, 54, 55, 61, 62, 71, 80, 82, 83, 94, 95, 97
11	S6G4GN	H	H	H	3i (2)	5.071/4.259	cx/bi/bis	0.003/0.001	64, 65
12	(G4)GN	H	H	H	3i (2)	5.059/4.253	cx/bi/bis	0.007/0.006	66, 69, 75, 76, 79, 80, 96
13	GN	H	GN	H	3i (2, 4)	5.057/4.281	cx/tri/bis	0.002/0.004	67, 68, 77
14	GN	H	(G4)GN	H	3i (2, 4)	5.044/4.290	hy/tri/bis	0.004/0.000	57, 73
15	GN	H	H	H	3i (2)	5.038/4.255	hy/bi/bis	0.000/0.000	58
	arm units								
16	M2M	H	H	H	3i (2M2M)	5.404/4.103	hm	0.009/0.006	72, 92
17	H	H	H	H	3t	5.055/4.061	hy/bi/bis	0.000/0.000	58
18	H	H	H	H	3t	5.044/4.058	hy/tri/bis	0.004/0.000	57, 73
					βMan(1,4)				
	core units								
19	H	M	H	M	4i (3, 6)	4.781/4.253	hm	0.001/0.004	54, 83
20	H	M	H	M	4i (3, 6)	4.772/4.236	hm [f]	0.000/0.000	55
21	H	H	H	(M)	4t, 4i (6)	4.771/4.076	ln	0.008/0.010	50, 51, 81, 82, 94, 95
22	H	(G4)GN2M	H	G4GN2M	4i (3, 6)	4.767/4.250	cx/bi	0.004/0.000	22, 98, 99
23	H	G4GN2[G4GN4]M	H	G4GN2([G4GN6])M	4i (3, 6)	4.757/4.216	cx/tri, tet	0.006/0.009	23–27, 29–34, 36, 37
24	H	GN2M	GN	(M6)[M3]M	4i (3, 4, 6)	4.742/4.186	hy/bi/bis	0.007/0.007	58, 80
25	H	(S6)(G4)GN2M	GN	(S6)(G4)GN2M	4i (3, 4, 6)	4.695/4.180	cx/bi/bis	0.018/0.011	64, 65, 69, 75, 76, 79
26	H	GN2[GN4]M	GN	GN2M	4i (3, 4, 6)	4.683/4.142	cx/tri/bis	0.001/0.002	67, 68, 77
					αMan(1,6)				
	core units								
27	(S6)G4GN	H	H	H	6i (2)	5.015/4.135	cx/bi/bis	0.005/0.002	64, 65, 76
28	GN	H	H	H	6i (2)	5.005/4.150	cx/bi/bis	0.003/0.006	69, 75, 79, 96
29	GN	H	H	H	6i (2)	4.999/4.142	cx/tri/bis	0.009/0.002	67, 68, 77
30	H	H	H	H	6t	4.967/3.971	cx/bi/bis	0.000/0.000	66
31	S6G4GN	H	H	H	6i (2)	4.947/4.118	cx/bi, tri	0.009/0.005	11, 20, 21, 78

32	H	M	H	6i (3)	4.940/4.159	hy/bi/bis	0.000/0.000	80
33	(S3)(G4)GN	H	H	6i (2)	4.926/4.111	cx/bi, tri	0.015/0.016	2, 3, 9, 10, 15, 16, 18, 19, 22–27, 74, 97–99
34	H	M	H	M	6i (3, 6)	4.917/4.158	hy/bi, tri/bis	0.006/0.005	57–60, 73
35	H	H	H	H	6t	4.915/3.969	cx/bi, ln	0.004/0.004	4–7, 81
36	H	M	H	H	6i (3)	4.893/4.137	ln	0.001/0.007	45, 95
37	H	M	H	M	6i (3, 6)	4.872/4.147	hm	0.013/0.018	41, 43, 54, 55, 71, 82, 83, 92, 94
38	G4GN	H	H	G4GN	6i (2, 6)	4.866/4.090	cx/tet	0.008/0.006	29–37
arm units								
39	M	H	H	H	6i (2M)	5.144/4.025	hm	0.005/0.000	72, 73, 92
40	H	H	H	H	6t	4.928/4.004	hy/bi, tri/bis	0.003/0.003	58, 73
41	H	H	H	H	6t	4.909/3.986	hm, ln	0.004/0.009	39, 41, 54, 55, 82, 83, 94

[a] Chemical shifts are in ppm relative to internal DSS (using internal acetone at 2.225 ppm) in D_2O at 23 °C. [b] S3 = αNeuNAc(2,3); S6 = αNeuNAc(2,6); G4 = βGal(1,4); GN2 = βGlcNAc(1,2); M = αMan; M2M = αMan(1,2)αMan. Substituents shown in parentheses do not affect the chemical shift values. [c] cx = complex; ln = linear (i.e., one of the core mannoses is absent); hm = high mannose; hy = hybrid; bi = biantennary; tri = triantennary; tet = tetraantennary; bis = bisected. [d] The values are the ranges found for clusters which are defined as similar microenvironments. The ranges are quoted as 0.000/0.000. The numbers following the errors are those corresponding to the structures found in Table I from which the averages were extracted. The large ranges for clusters 2, 10, 33, and 37 may arise from structural effects or experimental error. The latter occurs in some cases because of overlapping peaks. [e] Core = chemical shift of those mannoses which are part of the trimannose core; arm = chemical shift of those mannoses external to the core. [f] Cluster coordinates (not plotted) for 3αMan-substituted β(1,4)Man oligosaccharides are (4.774/4.225) and (4.786/4.239). The former group have an additional αMan(1,2) substituent on the terminal nonreducing mannose of the linear structures.

Table V: Chemical Shifts of Anomeric Hydrogen Resonances [a] of βGlcNAc and βGal Residues

	R-αMan(1,3) arm			R-αMan(1,6) arm		
R [b]	G4	GN2	class	G4	GN2	class
S6G4GN2	4.444	4.602	I, II, IV	4.444	4.602	I, II, IV
S3G4GN2	4.548	4.578	I	4.548	4.578	I
G4GN2	4.469	4.578	I, IV	4.469	4.578	I, II, III, IV
G4GN2	4.469	4.591	II, III			
GN2		4.555	I, IV, VII		4.538	V
GN2		4.532	V, VIII			
GN2		4.517	IX			
		GN4			GN6	
S3G4GN4	4.548	4.547	II			
G4GN4	4.469 [c]	4.547	II, III			
G4GN6				4.482	4.547	III
G4GN4	4.471	4.535	XI			
GN4		4.517	V, XI			

[a] All values are given in ppm relative to internal DDS (using internal acetone at 2.225 ppm) in D_2O at 23 ± 1 °C. Errors were calculated as standard deviations and range from ±0.002 to ±0.006 ppm with the average centered at ±0.003 ppm. [b] S3 = αNeuNAc(2,3); S6 = αNeuNAc(2,6); G4 = βGal(1,4); GN2 = βGlcNAc(1,2); GN4 = βGlcNAc(1,4), and GN6 = βGlcNAc(1,6). [c] Fournet et al. (1978) also quote 4.444 ppm.

Table VI: Chemical Shifts of Anomeric Hydrogen Resonances of βGlcNAc in Core Residues [a]

	shift	class
	Core GlcNAc	
+fucose [b]	4.678	I, IV
–fucose	4.612	I, II, III, V, VI
–fucose	4.590	VIII, IX
	Bisecting GlcNAc	
	4.466	IV, V
	4.410	VIII, IX

[a] All values are given in ppm relative to internal DSS (using internal acetone at 2.225 ppm) in D_2O at 23 ± 1 °C. Errors were calculated as standard deviations and range from ±0.002 to ±0.005 ppm with the average centered at ±0.003 ppm. [b] αFuc(1,6) linkage to Asn-GlcNAc.

(C1–H), 4.829 ± 0.011 (C5–H), and 1.174 ± 0.004 (C6–H) ppm. In all cases the GlcNAc to which the fucosyl residue is attached bears an additional βGal(1,4) substitution, and it has been suggested that the downfield shift observed for the C5–H resonance arises from a specific deshielding effect of the C1, C4, and C5 oxygens of the galactosyl residue (Fournet et al., 1978). Lemieux had reported this effect in vicinally substituted, branched oligosaccharides containing α(1,2) and α(1,4)-linked Fuc residues (Lemieux et al., 1975). Thus, because the C5–H hydrogen is sensitive in many cases to steric effects arising from vicinal substitution and can be easily located by decoupling the C6 methyl protons, such branched structures can be readily detected.

D. Bisecting GlcNAc. The presence of a βGlcNAc(1,4)-βMan(1,4) sequence is always associated with well-defined mannose shifts (Table IV) and a 7-8 Hz doublet at either 4.440 or 4.466 ppm (Table VI). In addition, a 3.3-ppm multiplet has always been found in the NMR spectrum. This resonance has not yet been assigned; however, it is probably due to some specific interaction of the bisecting β(1,4)-linked GlcNAc arising from the particular stereochemistry associated with this substitution.

Conclusion

We have demonstrated that, with an appropriate analysis of the spectral parameters, the 360-MHz ^1H NMR spectra

of Asn-linked glycopeptides can be used to determine the primary structure for compounds from each of the nine structural classes so far described. A particularly powerful aspect of this analysis is the use of a plot of the C1–H chemical shift vs. the corresponding value for C2–H. Forty-one unique C1–H/C2–H chemical shift pairs have been correlated with specific structural microenvironments by this technique. Thus sequence information extended over many residues can be obtained immediately upon inspection of the mannose C1–H/C2–H values of an unknown. We suggest that this correlation arises from the existence of well-defined three-dimensional structures for the sequence of hexoses making up the microenvironment.

Previous reports have established the usefulness of C1–H and C2–H chemical shifts in determining the branching pattern of complex structures (Dorland et al., 1978b). In addition, the same group have reported the effects of αNeuNAc(2,3) and αNeuNAc(2,6) substitution in a biantennary complex structure (Schut et al., 1978). Ballou and colleagues (Lehle et al., 1979; Cohen & Ballou, 1979, 1980) have reported mannosyl C1–H/C2–H chemical shift values at 40 °C in structures drawn from the high mannose series.

In addition to the analysis of the Man chemical shifts, we have demonstrated that corroborative information can be gained from a consideration of the Gal and GlcNAc anomeric chemical shifts. Microenvironments generating distinguishable chemical shift values were found for the arm GlcNAc (seven), core GlcNAc (three), bisecting GlcNAc (two), and arm Gal residues (five).

The present report represents the first study in which the entire range of nine structural classes have been compared. This analyses establishes the feasibility of differentiating structures from each of these classes and lays the groundwork for a systemic approach to the determination of glycopeptide primary structure by 360-MHz proton magnetic resonance.

Acknowledgments

We thank our many collaborators for kindly supplying samples and encouragement, Paul Atkinson and Pamela Stanley for constructive criticism of the manuscript, and Alan Lee for excellent technical assistance.

References

Atkinson, P. H., Grey, A. A., Carver, J. P., Hakimi, J., & Ceccarini, C. (1981) *Biochemistry 20*, 3979–3986.

Carver, J. P., Grey, A. A., Ceccarini, C., Hakimi, J., & Atkinson, P. H. (1981) *Biochemistry* (preceding paper in this issue).

Cohen, R. E., & Ballou, C. E. (1979) Abstracts of Papers, ACS/CSJ Chemical Congress, Honolulu, HI, April 1–6, American Chemical Society, Washington, DC, No. CARB 52.

Cohen, R. E., & Ballou, C. E. (1980) *Biochemistry 19*, 4345–4358.

Dorland, L., Haverkamp, J., Schut, B. L., Vliegenthart, J. F. G., Spik, G., Strecker, G., Fournet, B., & Montreuil, J. (1977a) *FEBS Lett. 77*, 15–20.

Dorland, L., Schut, B. L., Vliegenthart, J. F. G., Strecker, G., Fournet, B., Spik, G., & Montreuil, J. (1977b) *Eur. J. Biochem. 73*, 93–97.

Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Strecker, G., Michalski, J.-C., Fournet, B., Spik, G., & Montreuil, J. (1978a) *Eur. J. Biochem. 87*, 323–329.

Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Fournet, B., Strecker, G., Spik, G., Montreuil, J., Schmid, K., & Binette, J. P. (1978b) *FEBS Lett. 89*, 149–152.

Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Montreuil, J., Spik, G., & Fournet, B. (1979) *Eur. J. Biochem. 100*, 569–574.

Fournet, B., Montreuil, J., Strecker, G., Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Binette, J. P., & Schmid, K. (1978) *Biochemistry 17*, 5206–5214.

Gray, G. R., & Ballou, C. E. (1971) *J. Biol. Chem. 246*, 6845–6842.

Grey, A. A., Atkinson, P. H., Ceccarini, C., & Carver, J. P. (1979) Complex Carbohydrates Symposium, Toronto, Canada, Abstr 15.

Harpaz, N., & Schachter, H. (1980) *J. Biol. Chem. 255*, 4885–4893.

Ito, S., Yamashita, K., Spiro, R. G., & Kobata, A. (1977) *J. Biochem. (Tokyo) 81*, 1621–1631.

Lehle, L., Cohen, R. E., & Ballou, C. E. (1979) *J. Biol. Chem. 254*, 12209–12218.

Lemieux, R. U., Bundle, D. R., & Baker, D. A. (1975) *J. Am. Chem. Soc. 97*, 4076–4083.

Liang, C. J., Yamashita, K., Muellenberg, C. G., Sichi, H., & Kobata, A. (1979) *J. Biol. Chem. 254*, 6414–6418.

Montreuil, J., & Vliegenthart, J. F. G. (1979) in *Glycoconjugate Research* (Gregory, J. D., & Jeanloz, R. W., Eds.) Vol. I, pp 35–78, Academic Press, New York.

Narasimhan, S., Wilson, J. R., Martin, E., & Schachter, H. (1979) *Can. J. Biochem. 57*, 83–96.

Narasimhan, S., Harpaz, N., Longmore, G., Carver, J. P., Grey, A. A., & Schachter, H. (1980) *J. Biol. Chem. 255*, 4876–4884.

Nilsson, B., Norden, N. E., & Svensson, S. (1979) *J. Biol. Chem. 254*, 4545–4553.

Norden, N. E., Lundblat, A., Svensson, S., & Autio, S. (1974) *Biochemistry 13*, 871–874.

Reading, C. L., Penhoet, E. E., & Ballou, C. E. (1978) *J. Biol. Chem. 253*, 5600–5612.

Schut, B. L., Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., & Fournet, B. (1978) *Biochem. Biophys. Res. Commun. 82*, 1223–1228.

Strecker, G., Herlant-Peers, M. C., Fournet, B., Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., & Farriaux, J.-P. (1977) *Eur. J. Biochem. 81*, 165–171.

Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Iwashita, S., Inoue, Y., & Kobata, A. (1975) *J. Biol. Chem. 250*, 8569–8575.

Tai, T., Yamashita, K., Ito, S., & Kobata, A. (1977a) *J. Biol. Chem. 252*, 6687–6694.

Tai, T., Yamashita, K., & Kobata, A. (1977b) *Biochem. Biophys. Res. Commun. 78*, 434–441.

Tanaka, M., & Yamashina, J. (1973) *Carbohydr. Res. 27*, 175–183.

van Halbeek, H., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Michalski, J.-C., Montreuil, J., Strecker, G., & Hull, W. E. (1980a) *FEBS Lett. 121*, 65–70.

van Halbeek, H., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Strecker, G., Michalski, J.-C., Montreuil, J., & Hull, W. E. (1980b) *FEBS Lett. 121*, 71–77.

Yamashita, K., Tachibana, Y., & Kobata, A. (1978) *J. Biol. Chem. 253*, 3862–3869.