Carbohydrate Research, 147 (1986) 21–29
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

STRUCTURES OF THE CARBOHYDRATE MOIETY ATTACHED TO ONE SITE IN THE FIRST DOMAIN OF TURKEY OVOMUCOID: ELUCIDATION BY ^{1}H-N.M.R. SPECTROSCOPY

JOHN M. RISLEY AND ROBERT L. VAN ETTEN
Department of Chemistry, Purdue University, West Lafayette, IN 47907 (U.S.A.)
(Received June 7th, 1985; accepted for publication in revised form, August 13th, 1985)

ABSTRACT

^{1}H-N.m.r. spectroscopy was used to elucidate the primary structures of the carbohydrate moiety attached to asparagine at residue 53 in the first domain of turkey ovomucoid, a serine proteinase inhibitor. The carbohydrate moiety is a heterogeneous mixture of three structurally closely related complex-type oligosaccharides. Of the total carbohydrate moiety, 61% is tetra-antennary with terminal galactose and with an intersecting *N*-acetylglucosamine, and containing an additional *N*-acetylglucosamine (10′) attached to mannose (4′). Another 23% is tri-antennary with terminal galactose and with an intersecting *N*-acetylglucosamine. The remaining 16% is tri-antennary with terminal galactose (6 and 8 only), with an intersecting *N*-acetylglucosamine.

INTRODUCTION

Ovomucoids are glycoproteins that constitute a major component of the protein (≈10%) of avian egg-whites. Their structure is three tandem, homologous domains[1]. They specifically inhibit serine proteinases, and thus have been the subject of detailed inquiry for a number of years[2]. The primary focus of these investigations has been the polypeptide portion of the glycoprotein, whereas little work has been reported on the oligosaccharide portion other than the sites of attachment. There are four or five different sites of glycosylation, which is dependent on the species of the bird[3]. The carbohydrate moiety is attached at all sites on the protein through an *N*-glycosylic bond formed from the amide group of an asparagine residue and the reducing end of an *N*-acetylglucosamine-oligosaccharide. The structures of the carbohydrate moieties in hen ovomucoid have been elucidated[4,5], although the specific sites of attachment to the protein for any one of the carbohydrate moieties are not known.

Turkey *(Meleagris gallopavo)* is a Phasianoid bird whose ovomucoid has four sites of asparaginyl-linked oligosaccharide: at residues 10 and 53 in the first domain, at residue 75 in the second domain, and at residue 175 in the third domain; glycosylation at residue 175 is observed[3] at ~50%. The structures of the carbohydrate

0008-6215/86/$ 03.50

moiety attached to these four sites have not been reported. In conjunction with our investigation on the classification of almond glycopeptidylamidase as an amidase enzyme[6], we isolated the carbohydrate moiety attached to residue 53 from the first domain of turkey ovomucoid and determined the primary structure by ^{1}H-n.m.r. spectroscopy; one-dimensional and two-dimensional ^{1}H-n.m.r. spectroscopy were used to make chemical-shift assignments.

RESULTS AND DISCUSSION

Turkey ovomucoid was isolated and purified to homogeneity. Upon controlled proteolytic digestion, a glyco-octapeptide encompassing residues 50–57 from the first domain was isolated[7]: H_2N-Tyr-Gly-Thr-Asn(CHO)-Ile-Ser-Lys-Glu-CO_2H. Almond glycopeptidylamidase catalyzes the hydrolysis of the amide bond between asparagine and the attached *N*-acetylglucosamine (2-acetamido-2-deoxy-D-glucose) in a glycoprotein with release of the complete, intact carbohydrate moiety[6]. The enzyme was used to remove the carbohydrate moiety from the glyco-octapeptide. The free carbohydrate moiety and polypeptide chain were separated by gel-exclusion chromatography. Chemical analysis of the carbohydrate moiety showed the presence of mannose, *N*-acetylglucosamine, and galactose. After equilibration of the carbohydrate moiety in ^2H_2O, the ^{1}H-n.m.r. spectrum and the homonuclear two-dimensional ^{1}H(^{1}H)COSY spectrum were recorded at 470 MHz. The ^{1}H-n.m.r. spectrum for the carbohydrate moiety is shown in Fig. 1, and the resolution-

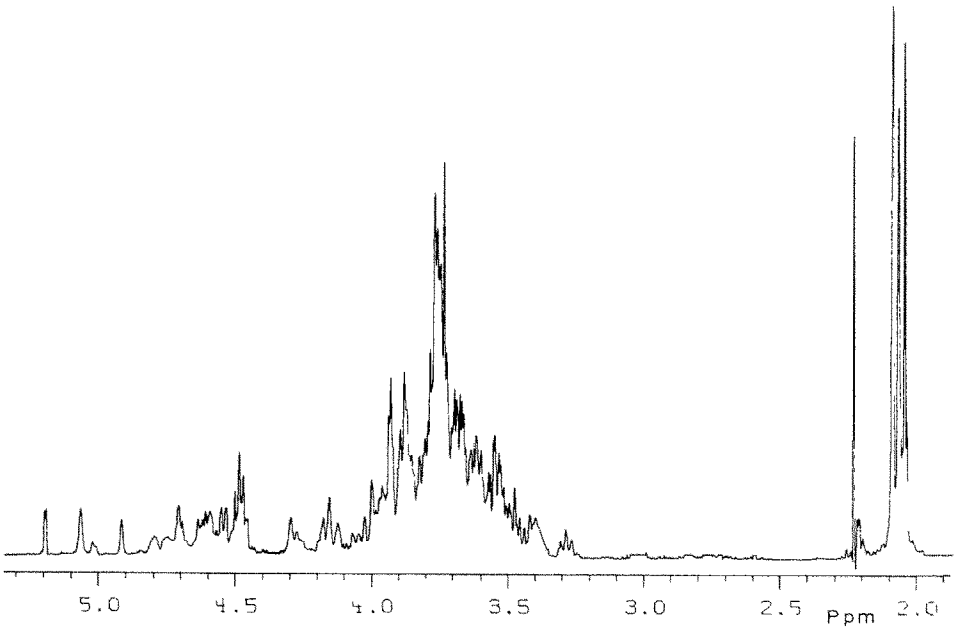


Fig. 1. ^{1}H-n.m.r. spectrum at 470 MHz of the carbohydrate moiety from the first domain of turkey ovomucoid. Internal reference: acetone (2.225 p.p.m.).

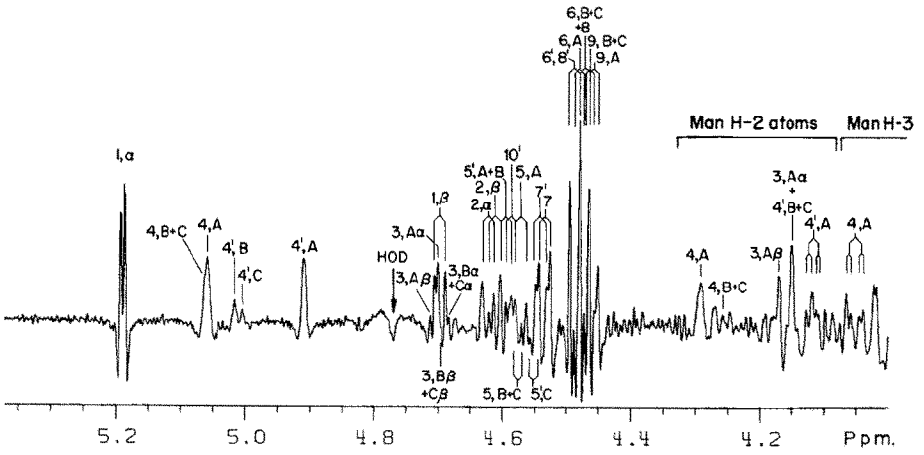


Fig. 2. The resolution-enhanced ^{1}H-n.m.r. spectrum of the anomeric-proton region and mannose H-2,3 region of the carbohydrate moiety from the first domain of turkey ovomucoid with assignment of the signals.

enhanced anomeric-proton region and Man H-2,3 region are shown in Fig. 2; acetone (2.225 p.p.m.) was used as the internal reference for the assignment of chemical shifts. The ^{1}H(^{1}H)COSY spectrum for the region 3.2–5.2 p.p.m. is shown in Fig. 3. These spectra were used to elucidate the structure of the carbohydrate moiety by comparison of the chemical shifts with those reported for other structures[4,8–13]. The anomeric mixture of *N*-acetyl-α- and β-glucosamine residues at the reducing end of the free carbohydrate moiety were retained for n.m.r. analysis and not reduced to the alditol. The reduction was unnecessary for the n.m.r. analysis because anomeric mixtures of oligosaccharides generated from enzymic hydrolysis of the *N,N'*-diacetylchitobiose of the core structure are routinely used for n.m.r. analysis, and the assignment of n.m.r. signals for these anomeric mixtures is well documented[8].

The core structure for an asparagine-linked oligosaccharide is α-Man(1→3)[α-Man(1→6)]β-Man(1→4)β-GlcNAc(1→4)α,β-GlcNAc. The attachment of specific sugar residues at the non-reducing end of the core structure defines the oligosaccharide as high mannose-, hybrid-, or complex-type. Each of these types of oligosaccharides has a distinctive set of ^{1}H-n.m.r. signals[8]. Examination of the ^{1}H-n.m.r. spectrum in Fig. 1 and Fig. 2 reveals immediately that the carbohydrate moiety is a complex-type of oligosaccharide because the n.m.r. signals for the anomeric protons of only the three mannose residues in the core structure are present; no anomeric-proton n.m.r. signal for the additional mannose residues found in high mannose- or hybrid-type oligosaccharides are observed[8]. In addition, the "triplet" at 3.281 p.p.m. is characteristic for an intersected oligosaccharide[10], namely an *N*-acetylglucosamine residue attached to the β-Man in a (1→4) linkage. The two regions of the spectrum that were analyzed closely in order to elucidate

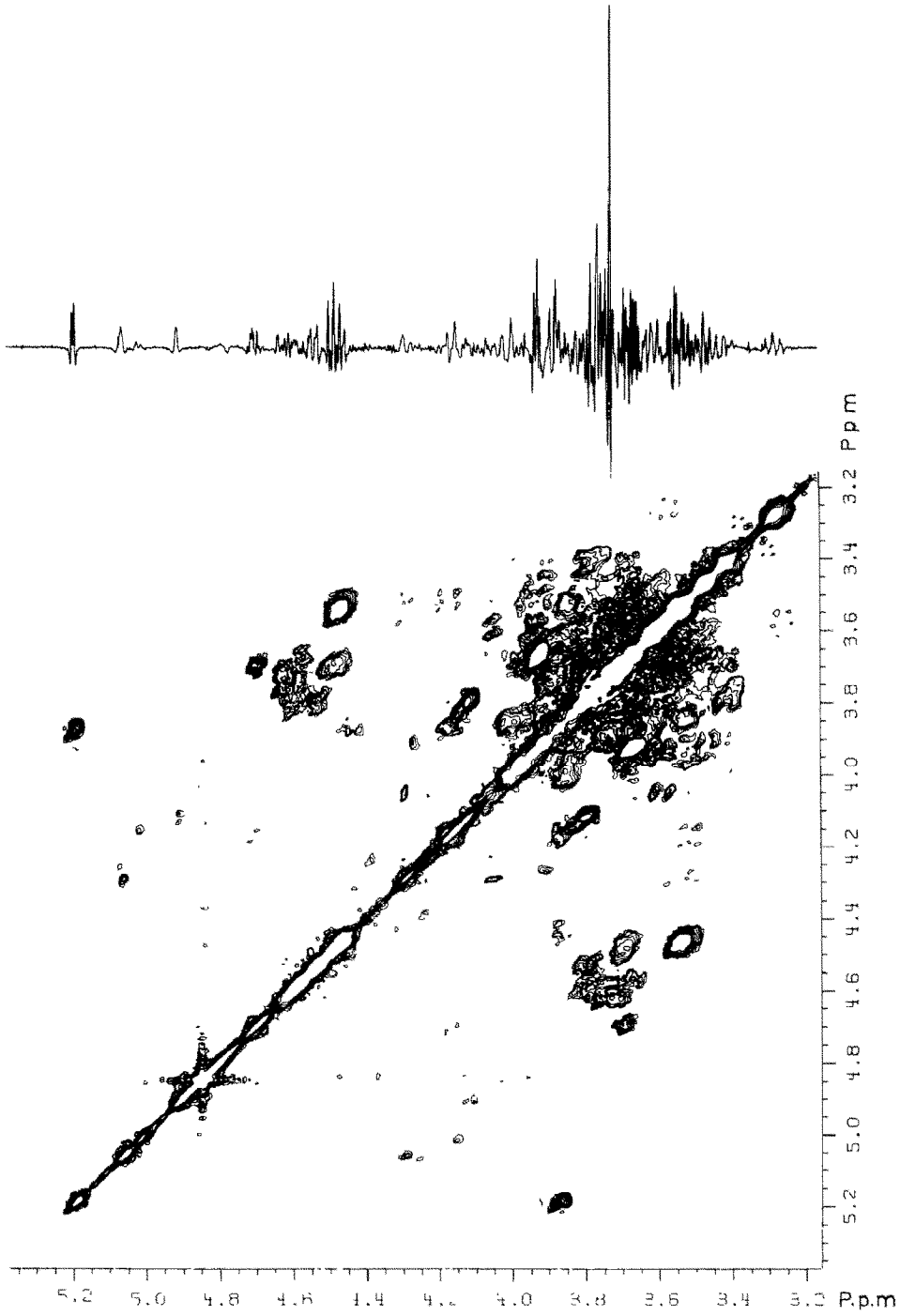
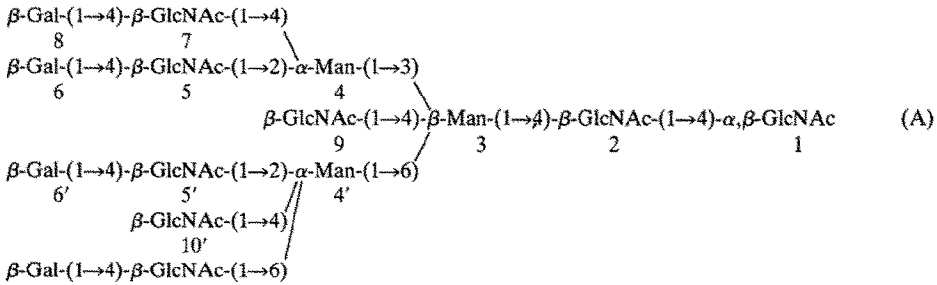
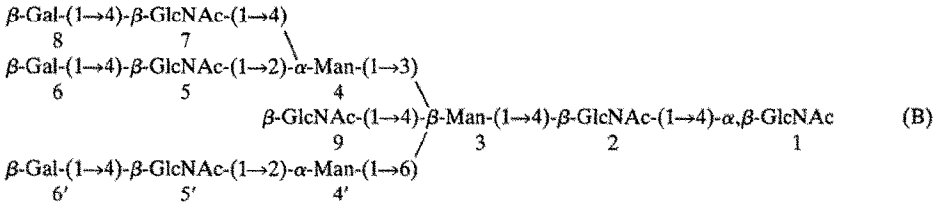


Fig. 3. $^1H(^1H)$COSY spectrum of the carbohydrate moiety from the first domain of turkey ovomucoid between 3.2–5.2 p.p.m.

the primary structure of the carbohydrate moiety were the anomeric proton and mannose H-2,3 region between 4.0–5.3 p.p.m., and the *N*-acetyl group region between 2.0–2.1 p.p.m. The chemical shifts of the n.m.r. signals in Fig. 1 and Fig. 2 were tabulated and compared to the published chemical shifts for complex-type oligosaccharides for assignment of the n.m.r. signals to specific sugar residues[4,8–13]. The spectrum in Fig. 1 was integrated to quantitate the area under each signal. The details of the analysis are outlined next.

The carbohydrate moiety isolated from residue 53 of the first domain of turkey ovomucoid is heterogeneous. It is a mixture of three structurally closely related oligosaccharides whose structures are:





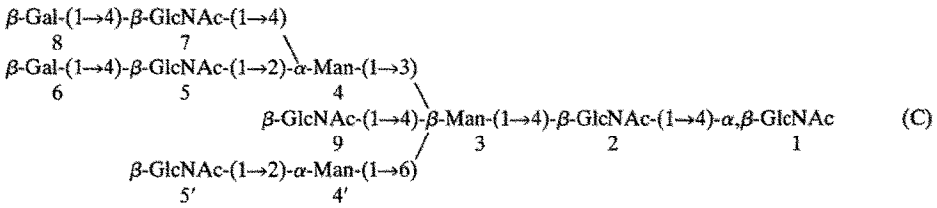


Table I lists the chemical-shift data for such structure.

The anomeric proton of GlcNAc-1 has two resonance signals at 5.188 p.p.m. ($J_{1,2}$ = 2.7 Hz) and at 4.696 p.p.m. ($J_{1,2}$ = 8.0 Hz) which correspond to the α and β anomers of the sugar[11] in a ratio of ~2:1, respectively. Two signals are also observed for the anomeric proton of GlcNAc-2 at 4.620 and 4.610 p.p.m. ($J_{1,2}$ = 8.6 Hz) for the α anomer and the β anomer (of GlcNAc-1), respectively[11]. Four signals are found for the anomeric proton of Man-3: 4.710, 4.699, 4.694, and 4.682 p.p.m. The proton signals at 4.699 and 4.710 p.p.m. are those of the anomeric

TABLE I

^{1}H-N M R CHEMICAL-SHIFT DATA FOR THE THREE STRUCTURES OF THE CARBOHYDRATE MOIETY FROM TURKEY OVOMUCOID FIRST DOMAIN

Residue	*Proton*	*Structure A* 8-7, 6-5-4, 9-3-2-1, 6′-5′-4′, 10′, 8′-7′	*Structure B* 8-7, 6-5-4, 9-3-2-1, 6′-5′-4′	*Structure C* 8-7, 6-5-4, 9-3-2-1, 5′-4′
GlcNAc-1	H-1α	5 188	5.188	5 188
	H-1β	4.696	4 696	4.696
	H-2α	3.875	3.875	3 875
	H-2β	3 694	3 694	3 694
	NAc	2.037	2.037	2.037
GlcNAc-2	H-1α	4.620	4.620	4.620
	H-1β	4.610	4 610	4.610
	H-2	3.715	3.715	3.715
	NAc	2.080	2.080	2.080
Man-2	H-1α	4 699	4.682	4 682
	H-1β	4.710	4 694	4 694
	H-2α	4 148		
	H-2β	4.172		
	H-3α	3.832		
	H-3β	3.875		
Man-3	H-1	5.057	5 061	5.061
	H-2	4 289	4.255	4.255
	H-3	4.050	3 875	3 875
	H-4	3 591		
	H-5	(3 281)[a]		
Man-4′	H-1	4 908	5.015	5.003
	H-2	4.115	4 148	4 148
	H-3	3.793		
GlcNAc-5	H-1	4 570	4.576	4.576
	H-2	3.755	3.755	3.755
	NAc	2 060	2.060	2.060
GlcNAc-5′	H-1	4 593	4.593	4 547
	H-2	3 781	3 781	3 781
	NAc	2.037	2.047	2.052
Gal-6	H-1	4.479	4 472	4.472
	H-2	3.544	3 544	3 544
Gal-6′	H-1	4.485	4.485	
	H-2	3 557	3.557	
GlcNAc-7	H-1	4 533	4.533	4 533
	H-2	4 789	3.789	3.789
	NAc	2.080	2.080	2.080
GlcNAc-7′	H-1	4.538		
	H-2	3 822		
	NAc	2 039		
Gal-8	H-1	4.471	4 471	4 471
	H-2	3.549	3 549	3.549
Gal-8′	H-1	4.486		
	H-2	3 534		
GlcNAc-9	H-1	4 456	4.461	4.461
	H-2	3.676	3.676	3 676
	NAc	2.060	2.060	2.060
GlcNAc-10′	H-1	4 585		
	H-2	3.669		
	NAc	2.084		

[a]Tentative assignment

proton of Man-3 in structure A for the α anomer and β anomer (of GlcNAc-1), respectively. The proton signals at 4.682 and 4.694 p.p.m. are those of the anomeric proton of Man-3 in structures B and C for the α anomer and β anomer (of GlcNAc-1), respectively. The proton signals for the α anomers are shifted upfield from the β anomers in the sugar residue that is two sugars removed from the anomeric sugar, whereas the opposite is observed in the sugar residue attached to the anomeric sugar[8]. The broad singlet at 5.057 p.p.m. (Fig. 1) appears as a shouldered singlet upon resolution enhancement (Fig. 2) and as two signals in the COSY spectrum (Fig. 3) with chemical shifts of 5.057 and 5.061 p.p.m. These signals are for the Man-4 anomeric protons in structure A and in structures B and C, respectively; the signal at 5.057 p.p.m. for structure A is shifted upfield from the signal observed in the agalacto-oligosaccharide[4]. Three signals are observed for the anomeric proton of Man-4′: 4.908 p.p.m. for structure A, 5.015 p.p.m. for structure B, and 5.003 p.p.m. for structure C. A downfield shift of the Man-4′ anomeric proton in structure A is observed compared to the agalacto-oligosaccharide[4]. The upfield shift of 0.012 p.p.m. in the proton signal for Man-4′ of structure C from structure B clearly indicates[8] the absence of Gal-6′. The integration of the three resonance signals for the anomeric proton of Man-4′ shows that the ratio of the three species is 61:23:16 (A:B:C). The signals for the anomeric protons of the mannose sugars are all characteristic[4,8,10] of an intersecting GlcNAc-9.

Two proton signals for the anomeric proton of GlcNAc-5 are at 4.570 p.p.m. for structure A and at 4.576 p.p.m. for structures B and C. The large downfield shifts of ~0.030 p.p.m. from the agalacto-oligosaccharides[4,8,10] are indicative[8] of the presence of Gal-6. There are also two proton signals for the anomeric proton of GlcNAc-5′: 4.593 p.p.m. for structures A and B, and 4.547 p.p.m. for structure C. The upfield shift of ~0.045 p.p.m. for the anomeric proton of GlcNAc-5′ in structure C from structures A and B is characteristic of the agalacto-oligosaccharide[4]. One anomeric-proton signal for GlcNAc-7 is detected at 4.533 p.p.m. for structures A, B, and C, and one proton signal for the anomeric proton of GlcNAc-7′ is found at 4.538 p.p.m. The magnitude of the downfield shift (approximately 0.017 p.p.m.) from the agalacto-oligosaccharides[4,8] indicates[8] the presence of Gal-8 and Gal-8′. Two anomeric proton signals are observed for the intersecting GlcNAc-9 residue: 4.456 p.p.m. for structure A and 4.461 p.p.m. for structures B and C. The n.m.r. signal at 4.456 p.p.m. is downfield from the shift observed in the agalacto-oligosaccharide[4]. The anomeric-proton signal for GlcNAc-10′ at 4.585 p.p.m. is the same in the galacto- and agalacto-oligosaccharides[4].

The anomeric proton signals for the galactose residues in structure A are 4.479 (Gal-6), 4.485 (Gal-6′), 4.471 (Gal-8), and 4.486 p.p.m. (Gal-8′). In structure B the anomeric-proton signals of the galactose residues are found at 4.472 (Gal-6), 4.485 (Gal-6′), and 4.471 p.p.m. (Gal-8). In structure C the anomeric proton signals for the two galactose residues are at 4.472 (Gal-6) and 4.471 (Gal-8). These signals are the same as have been reported[8,10].

The chemical shifts for the *N*-acetyl groups on the eight GlcNAc sugar

residues are listed in Table I, and are approximately those observed in other studies[8,10]. However, we did not observe the large downfield shifts for two *N*-acetyl groups of 2.122 and 2.093 p.p.m. in structure A as were reported for the agalacto-oligosaccharide[4]. Instead, we found the *N*-acetyl groups for GlcNAc-10′ at 2.084 p.p.m. and for GlcNAc-7′ at 2.039 p.p.m. Apparently the galactose sugar residues do have a considerable influence on these n.m.r. signals.

The data from the homonuclear two-dimensional COSY spectrum (Fig. 3) allowed for the assignments of all of the H-2 protons, which generally agree with the chemical shifts reported[9,12,13]. The cross-peaks of the COSY spectrum indicate the chemical shifts of the protons that are coupled[14]. Thus, the chemical shift for the H-2 proton coupled to the anomeric proton in each residue could be assigned, and the assignments for the mannose H-2,3 residues could be verified. In addition, it was possible to observe other cross-peaks and to make chemical-shift assignments for other coupled protons. The H-2 proton of GlcNAc-10′ has approximately the same chemical shift as the H-2 proton of GlcNAc-9, and both cross-peaks in the COSY spectrum (Fig. 3) are distinct from the other GlcNAc sugar residues. These data indicate that the GlcNAc-10′ sugar may itself be considered to be "intersecting". The three mannose sugars yielded the most data from the COSY experiment. There is some indication from this data that the "triplet" characteristic of the intersected oligosaccharide[10] may be the H-5 proton of Man-4, however this is far from certain and further analysis of the system is necessary.

The structures of the asparaginyl-linked carbohydrate moiety isolated from residue 53 in the first domain of turkey ovomucoid determined here are not surprising, based on the work that has been done on the elucidation of the structures of the carbohydrate moieties from hen ovomucoid[4,5]. The specific sites of attachment for the carbohydrate moieties in hen ovomucoid are not known. In the case of turkey ovomucoid, we found that a heterogeneous mixture of three oligosaccharide structures comprises the carbohydrate moiety located on a specific site, residue 53. However, they all had in common a complex structure with an intersecting *N*-acetylglucosamine residue.

EXPERIMENTAL

Materials. — Turkey ovomucoid, turkey ovomucoid glyco-octapeptide from the first domain, and almond glycopeptidylamidase enzyme were obtained as previously described[6]. The carbohydrate moiety from the hydrolysis reaction of turkey ovomucoid glyco-octapeptide with almond glycopeptidylamidase was isolated as described earlier[6]. The intact carbohydrate moiety was equilibrated in deuterium oxide with intermediate lyophilization: three times in 99.8% ^2H_2O (Baker), two times in 99.96% ^2H_2O (SIC), and two times in 99.996% ^2H_2O (SIC).

Spectroscopy. — The equilibrated carbohydrate moiety was dissolved in 0.5 mL of 99.996% ^2H_2O for ^1H-n.m.r. analysis. A Nicolet NTC-470 spectrometer operating at 469.9 MHz was fitted with a 5-mm probe which was equilibrated at

23°. A sweep width of ±1000 Hz (quadrature phase detection) and a 90° pulse-angle were used. The ^{1}H-n.m.r. spectrum was recorded using a 16K data-block, and the f.i.d. was transformed using a resolution-enhancement factor and zero-filling to 32K data points. Homonuclear, two-dimensional quadrature-detected correlated spectra (2D-COSY)[14] were recorded using a 512 × 1024 data-block, which was transformed after zero-filling to 2048 data points. Chemical shifts were determined relative to sodium 4,4-dimethyl-4-silapentane-1-sulfonate using acetone as the internal reference (2.225 p.p.m.)

ACKNOWLEDGMENTS

The work was supported by USPHS Research Grant GM 27003 from the National Institute of General Medical Sciences and by NIH Grant RR 01077 from the Division of Research Resources. We thank Dr. Milo Westler for helpful discussions.

REFERENCES

1 M. Laskowski, Jr and I. Kato, *Ann. Rev. Biochem.*, 49 (1980) 593–626.
2 M. Laskowski, Jr., M. Tashiro, M. W. Empie, S. J. Park, I. Kato, W. Ardelt, and M. Wieczorek, in N. Katunuma, H. Umezawa, and H. Holzer (Eds.), *Proteinase Inhibitors: Medical and Biological Aspects*, Japan Sci. Soc. Press, Tokyo, 1983, pp. 55–68.
3 I. Kato, W. J. Kohr, and M. Laskowski, Jr., *Proc. FEBS Meet.*, 47 (1978) 197–206.
4 J. P. Parente, J.-M. Wieruszeski, G. Strecker, J. Montreuil, B. Fournet, H. van Halbeek, L. Dorland, and J. F. G. Vliegenthart, *J. Biol. Chem.*, 257 (1982) 13173–13175.
5 K. Yamashita, Y. Tachibana, A. Hitoi, and A. Kobata, *Carbohydr. Res.*, 130 (1984) 271–288.
6 J. M. Risley and R. L. Van Etten, *J. Biol. Chem.*, in press
7 E. M. Taga, A. Waheed, and R. L. Van Etten, *Biochemistry*, 23 (1984) 815–822.
8 J. F. G. Vliegenthart, L. Dorland, and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209–374.
9 S. N. Bhattacharyya, W. S. Lynn, J. Dabrowski, K. Trauner, and W. E. Hull, *Arch. Biochem. Biophys.*, 231 (1984) 72–85.
10 A. Pierce-Cretel, H. Debray, J. Montreuil, G. Spik, H. van Halbeek, J. H. G. M. Mutsaers, and J. F. G. Vliegenthart, *Eur J. Biochem.*, 139 (1984) 337–349.
11 R. C. Bruch, M. D. Bruch, J. H. Noggle, and H. B. White, III, *Biochem. Biophys. Res. Commun.*, 123 (1984) 555–561.
12 R. Geyer, H. Geyer, S. Stirm, G. Hunsmann, J. Schneider, U. Dabrowski, and J. Dabrowski, *Biochemistry*, 23 (1984) 5628–5637.
13 H. Paulsen, T. Peters, V. Sinnwell, R. Lebuhn, and B. Meyer, *Liebigs Ann. Chem.*, 1985 (1985) 489–509.
14 G. Wider, S. Macura, A. Kumar, R. R. Ernst, and K. Wuthrich, *J. Magn. Reson.*, 56 (1984) 207–234.