THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 266, No. 4, Issue of February 5, pp. 2622-2631, 1991
Printed in U.S.A.

Bovine Glomerular Basement Membrane

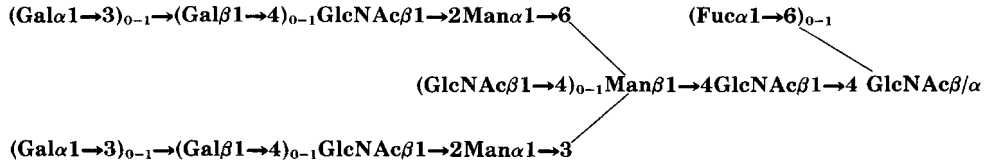
LOCATION AND STRUCTURE OF THE ASPARAGINE-LINKED OLIGOSACCHARIDE UNITS AND THEIR POTENTIAL ROLE IN THE ASSEMBLY OF THE 7 S COLLAGEN IV TETRAMER*

(Received for publication, May 4, 1990)

Jan P. M. Langeveld‡, Milton E. Noelken‡, Karl Hård§, Parvin Todd‡, Johannes F. G. Vliegenthart§, Jim Rouse¶, and Billy G. Hudson‡‖

From the ‡Department of Biochemistry, University of Kansas Medical Center, Kansas City, Kansas 66103, the §Department of Bio-Organic Chemistry, Utrecht University, Utrecht, The Netherlands, and the ¶Veterans Administration Hospital, Kansas City, Missouri 64128

Collagen IV contains an amino-terminal tetramerization domain (7 S) that is involved in aggregation and cross-linking as part of the process of self-assembly of the collagen IV matrix of basement membranes. We determined the structure and location of the Asn-linked oligosaccharides of the 7 S tetramer. Two glycopeptides, GP-1 and GP-2, were isolated from tryptic digests of the 7 S tetramer and were characterized. GP-1 and GP-2 are derived from the α1(IV) chain and the α2(IV) chain, respectively. Each glycopeptide contained one sequence, -Asn-Xaa-Thr-, which was shown to be *N*-glycosylated at Asn, corresponding to position 126 of the α1 chains and 138 of the α2 chain. ^{1}H NMR spectroscopic analysis of the oligosaccharide moiety from each glycopeptide indicated essentially identical structures. Each oligosaccharide is a biantennary *N*-acetyllactosamine type of *N*-linked oligosaccharide with a broad heterogeneity in the presence of the sugar residues at their nonreducing termini as indicated.



The location of the Asn-linked oligosaccharide units and Hyl-linked disaccharide units and their orientation with respect to the surface of the triple helix were calculated using two models. We conclude that both units are important determinants in the assembly of the 7 S tetramer.

Collagen IV is a glycoprotein, as are many extracellular matrix constitutents. It is considered to be an important structural component of basement membranes because of its ubiquitous presence in basement membranes throughout the body and its binding sites for other basement membrane constituents. The collagen IV protomer is a three-chained molecule about 400–420 nm long, and it is characterized by three distinct structural domains that have roles in formation of the supramolecular structure. These are the 7 S domain, which is the approximately 30-nm-long amino-terminal section of the collagenous domain; the NC1 domain, which is a trimeric 9-nm diameter globular moiety at the carboxyl-terminal end; and the triple helical domain, which is 360 nm long and is located between the terminal domains (1–4). Assembly of the collagen IV supramolecular structure involves tetramerization of the 7 S domain, dimerization of the NC1 domain, and noncovalent interactions between dimerized NC1 domains and the triple helical domain (5–8).

The carbohydrate moiety of collagen IV is distributed between Hyl-linked disaccharide units and putative Asn-linked oligosaccharide units. The pioneering work of Spiro (9) provides unambiguous evidence for the presence of both types of units in glomerular basement membrane. Other studies of this membrane suggest that both units are found on a single collagen chain (10–13). Structural studies on the 7 S domain tetramer from human placenta revealed that Hyl-linked disaccharide units and a putative Asn-linked oligosaccharide unit are located within the 7 S domain (tetramer) on both the α1 and α2 chains (14, 15). Evidence for the latter unit in placenta 7 S tetramer was obtained by partial characterization of a glycopeptide fraction from a Pronase digestion (16). The potential Asn glycosylation sites of the complete chains are Asn-126 of the α1 chain and Asn-138 and Asn-1270 of the α2 chain, as recently deduced from cDNA clones (17–20). Together, these findings indicate that the 7 S tetramer,[1] particularly that from placental tissue, contains both Hyl-linked

* This work was supported in part by National Institutes of Health Grant AM 18381, American Heart Association (Kansas Affiliate) Fellowship KS-85-F-3 (to J. P. M. L.), the Electron Microscopy Research Center of the University of Kansas Medical Center, the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO) and the Innovation Directed Research Program on Carbohydrates (IOP-K) (to J. F. G. V.), and the Suomen Kultuurirahasto Foundation (to K. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ To whom correspondence should be addressed.

[1] The 7 S domain is defined as residues 28–162 of the α1 chain and residues 29–171 of the α2 chain on the basis of cDNA analysis (17, 18). This region is referred to as the amino-terminal 7 S domain of the triple helical protomer. The 7 S domain of the tetramer is referred to herein as the 7 S tetramer. The tetramer is resistant to collagenase digestion and can be excised from basement membrane for study.

disaccharide units and Asn-linked oligosaccharide units. However, the detailed structure and whether a specific Asn residue is glycosylated remain undetermined.

The carbohydrate moiety of glycoproteins generally plays an important role in the maintenance of normal protein physicochemical properties and stabilization against proteolysis. The carbohydrate moiety also has specific roles, including involvement in biosynthesis and secretion of proteins (21), aggregation of IgG molecules in rheumatoid arthritis (22), tumor cell adhesion to laminin and collagen IV (23), and signal transduction of glycoprotein hormone (24, 25). Furthermore, the structure of the carbohydrate moiety of laminin, another major basement membrane glycoprotein, is under investigation because of its potential role in the function of laminin in cellular transformation, activity, and malignancy (26–28). Likewise, the carbohydrate units of the 7 S domain of collagen IV may have biological roles such as proper alignment of collagen IV molecules for the assembly of the supra-structure (7 S tetramer), attachment sites for other basement membrane constituents and cell surface receptors, masking of antigenic sites, and protection against proteolysis.

The purpose of the present study was 2-fold: 1) to determine the location and structure of the putative Asn-linked oligosaccharide units of the 7 S tetramer of GBM[2] collagen IV and 2) to determine the spacial positions of the oligosaccharide units and the Hyl-linked disaccharide units with respect to the surface of the triple helix. Our results revealed potential roles for both types of units in the mechanism of tetramer assembly, which thus extended the model of the assembly process described by Siebold *et al.* (15). We found that the Asn-linked units are likely to have an important role in determining the extent of longitudinal overlap of protomers in the 7 S tetramer and that the oligosaccharide and disaccharide units constitute a hydrophilic face that is likely to contribute to the geometry and stability of the 7 S tetramer.

The tetramer from GBM was chosen for study because of 1) the important physiological role of GBM in renal function (29), 2) the involvement of GBM collagen in the pathogenesis of Goodpasture and Alport syndrome (30), and 3) the molecular heterogeneity of basement membrane of various tissues (31).

MATERIALS AND METHODS

Preparation of the 7 S Domain of Collagen IV—The long form 7 S domain (7 Sl) was prepared from bovine renal cortex by a procedure based on established methods (32, 33). The short form of the 7 S domain (7 Ss) was purified from the previously described pool 1 material (34) obtained from bovine GBM which had been consecutively subjected to digestion with bacterial collagenase at 37 °C, anion-exchange chromatography, and gel filtration. Solutions originating from 4 g of GBM were dialyzed against 0.5 M acetic acid and concentrated to 100 ml by ultrafiltration (Amicon YM-10). At this stage, about 2% of the starting material was recovered on a dry weight basis. Eight mg of pepsin was added to the solution, which was then stirred for 20 h at 4 °C, after which the protein was precipitated by adding 20 g of NaCl. The pellet was dissolved in 40 ml of 50 mM Tris-HCl, 2 M urea, pH 7.4, 20 mM NaCl and subjected to anion-exchange column chromatography, gel filtration over agarose A-5m, cation-exchange chromatography, desalting, and lyophilization (32, 33). In one experiment, the 7 Ss domain was isolated from renal cortex by following the published procedures for the 7 Sl domain (32, 33) except that the digestion with bacterial collagenase was performed at 37 instead of 20 °C. The identity of the 7 Ss and 7 Sl domains was confirmed by electron microscopy.

Preparation of Tryptic Peptides—Before digestion with trypsin, 7 Sl and 7 Ss preparations were first subjected to reduction and *S*-aminoethylation under denaturing conditions. Ten mg of protein was reduced with 20 μl of β-mercaptoethanol and *S*-aminoethylated with 1.6 g of Aminoethyl-8® (Pierce Chemical Co.) according to the method of Schwartz *et al.* (35). At the end of the treatment, the solvent was replaced with 0.2 M NH_4HCO_3 on Bio-Gel P-6DG and concentrated to 10 ml by ultrafiltration (Amicon YM-10). The sample was heated at 90 °C for 5 min and then cooled in a water bath to 40 °C, at which temperature trypsin (tosylphenylalanylchloromethane-treated, 212 units/mg; Worthington) was added (enzyme:substrate ratio = 1:50, dry weight basis). The digestion was continued at 37 °C for 20 h, and the sample was then stored at 4 °C in the presence of 0.05% NaN_3 until use.

Enzymatic Release of N-Linked Saccharides—Glycopeptides from 10 mg of 7 S domain were dissolved in 0.75 ml of 50 mM Tris-HCl, 10 mM EDTA, pH 8.3, 0.02% NaN_3. One unit of peptide *N*-glycosidase F (EC 3.2.2.18, Boehringer Mannheim) was added at each of three time points: 0, 4, and 24 h, and the digestion was performed at ambient temperature for a total time of 48 h. The extent of *N*-glycosylation was measured with the direct ConA test. In this test, the capacity of the intact glycopeptide to bind ConA is used as a 100% control. Cleavage of saccharides from peptides will result in a decrease of this value.

Chromatographic Procedures for Peptides and Oligosaccharides—For purification of peptides from tryptic digests and separation of oligosaccharide and peptide moieties after peptide *N*-glycosidase F digestion of purified peptides, reversed-phase HPLC was carried out at 30 °C with a C18 column (Vydac, 201TP5μ, 0.46 × 15 cm) in a Varian LC model 5500, which can be programmed for linear gradients composed of 1% solvent increments and spectrophotometric detection at 230 nm. Solvent A consisted of water, 0.1% (v/v) trifluoroacetic acid, and solvent B consisted of acetonitrile, 0.1% (v/v) trifluoroacetic acid. After acidification to pH 2 with trifluoroacetic acid, samples were injected at a solvent ratio (A:B) of 90:10. After injection, the solvent ratio was first kept for 2 min at 90:10 and then consecutively increased in 1 min to 85:15, in 30 min to 75:25, and in 20 min to 60:40 and maintained at 60:40 for 2 min. At the end of the run, the solvent ratio was returned to 90:10 with the flow rate at 3 ml/min for 4 min and after that 1 ml/min. Fractions were monitored for carbohydrates by gas-liquid chromatography and the direct ConA binding test.

Anion-exchange chromatography of peptide moieties and test samples was carried out on a fast protein liquid chromatography system with Mono-Q HR 5/5 column (Pharmacia LKB Biotechnology Inc.) using spectrophotometric detection at 214 nm. The chromatographic conditions were modified from Damm *et al.* (36). Salt-free peptide samples, derived from 10 mg of 7 Sl or 7 Ss domain dissolved in 1 ml of HPLC-quality water (Millipore system, pH 6–7), were applied to the column in 50- or 500-μl aliquots per run. NeuAc and acetate were used as charge markers. The flow rate was kept at 2 ml/min. Solvents consisted of water (A) and 0.5 M NaCl (B). The runs were developed as follows. At the start, solvent ratio A:B was kept for 2 min at 100:0, changed to 90:10 in 8 min, and changed further to 0:100 in 8 min, after which time the ratio was maintained for 2 min and then returned at once to starting conditions.

Gel Filtration of Peptides and Oligosaccharides—Lyophilized peptide or oligosaccharide preparations were dissolved in 0.75 ml of water and run over 1 × 30-cm Bio-Gel P2 (Bio-Rad) columns at a flow rate of 9–12 ml/h using HPLC-grade water as eluent. The effluent was monitored by absorbance at 206 nm. Gel filtration on a 1.2 × 46-cm Bio-Gel P6 (Bio-Rad) column was carried out with 0.2 M ammonium acetate as the eluent. The flow rate was maintained at 10 ml/h with a peristaltic pump. The elution position of peptides was monitored by absorbance at 230 nm, and that of oligosaccharides was monitored with the indirect ConA test.

Chemical Analyses—Protein concentration was estimated by the Lowry method using bovine serum albumin as a standard. To remove interfering substances when necessary, protein was first precipitated (37). Hydrolysis of protein and analysis of amino acids were performed as before (38), with norleucine as internal standard.

Monosaccharide analysis was carried out by gas-liquid chromatography analysis of trimethylsilyl derivatives prepared from 0.1–1 nmol of desalted and lyophilized glycopeptide or glycoprotein samples. The procedure was that of Chaplin (39) with modifications to methanolysis (6 h at 70 °C in 1 M methanolic HCl) and silylation volume (50 μl of reagent). After dissolving samples in hexane, 3 μl was analyzed on a Hewlett-Packard model 5830 gas chromatograph using a splitless mode for 0.5 min during injection, followed by a split mode with a split ratio of 1:100. A 0.32 mm × 25-m Chrompack CP Sil5 CB

[2] The abbreviations used are: GBM, glomerular basement membrane; 7 Sl and 7 Ss, long and short forms of the 7 S domain, respectively; ConA, concanavalin A; GP, glycopeptide; HPLC high perormance liquid chromatography.

capillary column was used. The conditions of separation were the following: injector temperature, 240 °C; detection temperature, 270 °C; oven temperature program, 100 °C initially, with an 8 °C/min rise for 3 min, a 4 °C/min rise for the next 11 min, and then a 10 °C/min rise to a final temperature of 240 °C, which was maintained for 5 min; oxygen-free helium as carrier gas; and a column head pressure of 1 $kg \cdot cm^{-2}$. A calibration curve with carbohydrate standards was included in each series of analyses, and myoinositol was used as internal standard.

Electron Microscopy—Samples (25 μg/ml) were sprayed in 25 mM acetic acid, 50% (v/v) glycerol and further treated for rotary shadowing following established procedures (34, 40). A carbon replica with waffle pattern (Pelco) was used to determine actual magnification.

^1H NMR Spectroscopy—Prior to ^1H NMR analysis, the samples were repeatedly exchanged in ^2H_2O (finally using 99.96 atom % ^2H, Aldrich) with intermediate lyophilization. ^1H NMR spectra were recorded on Bruker AM-500 spectrometers operating at 500 MHz at a probe temperature of 300 K. Prior to Fourier transformation, the resolution was enhanced by multiplication with a Gaussian modification function. Chemical shifts (δ) are given relative to internal acetone (δ = 2.225 ppm in ^2H_2O) (41). The two-dimensional HOHAHA spectrum was obtained using a 120-ms Malcolm Levitt-17 mixing sequence (42). 512 spectra of 4,000 data points were recorded, with 64 scans/t_1 value. The 90° ^1H pulse width was 27.8 μs, and the residual HO(^2H) signal was suppressed by presaturation for 1.0 s. A phase-sensitive Fourier transformation was performed after zero filling to a 2,000 × 4,000-data point matrix and multiplying with a phase-shifted sine-bell function. In such a two-dimensional spectrum, cross-peaks can occur between all protons in a scalar coupling network, depending on the applied mixing time and the magnitude of the coupling constants in the spin system. For oligosaccharides, subspectra of the individual monosaccharides can be derived by taking cross-sections of the two-dimensional spectrum at clearly resolved diagonal peaks of (mostly anomeric or H-2) protons (43, 44). The assignments in this study are thus based on both the high resolution one-dimensional spectra and the two-dimensional HOHAHA spectral analysis.

Amino Acid Sequence Analysis—Amino-terminal amino acid sequence analyses were based on automated Edman degradation and were performed (45) with a gas-liquid, solid-phase peptide sequenator (Applied Biosystems). Identification of phenylthiohydantoin derivatives was performed on a Waters HPLC using a Nova-pak C18 column for reversed-phase HPLC (46).

Concanavalin A Binding Assay (ConA Test)—This assay was used to measure the presence of saccharides that contain mannose and is analogous to an enzyme-linked immunosorbent assay. A direct and an indirect ConA test were designed. The direct test measures glycoconjugates that can be coated onto the walls of reaction vials. The test was carried out in flat-bottom polystyrene microtiter plates with high adsorbing capacity (NUNC, Denmark). For coating, samples were diluted either with 0.05 M sodium carbonate buffer, pH 9.6, 0.02% (w/v) NaN_3 or, if calcium was present, with 50 mM Tris-HCl, pH 7.5, 0.15 M NaCl, 0.02% (w/v) NaN_3. Two hundred μl of serial dilutions were pipetted into each well and incubated overnight. After washing, the wells were incubated for 4 h with ConA (40 μg/ml, Sigma type IV) in 20 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.05% (w/v) Tween 20, 0.2% bovine serum albumin, 0.02% (w/v) NaN_3. The plates were washed, incubated for 1 h with 200 μl of horseradish peroxidase (10 μg/ml; Sigma type VI) in the incubation buffer, and washed again. Plates were developed with 200 μl of substrate solution containing *o*-phenylenediamine dihydrochloride (0.4 mg/ml) in 24.3 mM citric acid, 51.4 mM Na_2HPO_4, 0.012% (v/v) H_2O_2. The reaction was stopped, usually after 30 min, by adding 50 μl of 2 M H_2SO_4. Absorbance was measured at 495 nm. For blanks, wells were used that had gone through the substrate step only.

The indirect test was used for free saccharides, which can inhibit the binding of ConA to a coated glycoconjugate. For the indirect test, the direct test was modified as follows. The coating step was carried out with chicken ovalbumin at 0.4 μg/ml. For the incubation step with ConA, 100 μl of saccharide solution diluted in incubation buffer was combined in each well with an equal volume of ConA solution (80 μg/ml).

In both tests, all steps were carried out at ambient temperature. Washes were performed with 0.15 M NaCl, 0.05% (w/v) Tween 20, 1 mM $CaCl_2$, 1 mM $MnCl_2$. Under these conditions, optimal coat concentrations of 7 S domain and ovalbumin were 1 and 0.4 μg/ml, respectively. The specificity of the lectin in the assay was confirmed by testing the competitive effect of several sugars in the indirect test. A 50% inhibition of ConA binding was reached for methyl α-D-mannoside, methyl α-D-glucoside, D-mannose, D-glucose, N-acetyl-D-glucosamine, and methyl β-D-glucoside at concentrations of 55, 227, 268, 1,250, 2,062, and 5,140 nmol/ml, respectively. Gal, Fuc, GalNAc, and NeuAc did not inhibit at concentrations less than 0.01 M. The high sensitivity of the direct test (10 pmol of ovalbumin/ml is sufficient) is probably due to the tetravalent binding property of the lectin (47). The peroxidase is a glycoprotein itself and has good affinity for ConA.

Location and Orientation of Saccharides and Hydrophobic and Cysteine Residues with Respect to the Surface of the 7 S Triple Helical Aggregation Domain—We calculated the three-dimensional positions of the β-carbons of *X* and *Y* residues of the $(GlyXY)_n$ repeating sequence by use of a model developed for triple helical collagen by Hoffmann *et al.* (48) and applied to the 7 S domain by Siebold *et al.* (15). In this model, the α2(IV) chain is the reference chain for relative chain location. Pro-55 (Pro-19 (15)) of the α2(IV) chain is the first residue of the triple helical domain (although the first Gly*XY* triplet begins with residue 22) and is considered to be a hypothetical Gly residue. Its C_α atom is assigned the following cylindrical coordinates: z = 0 nm and ϕ = 0°, where z is the axial distance and ϕ is the azimuthal angle, measured anticlockwise when viewing the triple helix from its carboxyl terminus. We calculated the positions of the C_β atoms of the *X* and *Y* residues relative to the position of C_α of α2(IV) Pro-55 (Pro-19 (15)) from the atomic coordinates of $(GlyProPro)_n$ (49, 50). Equivalent atoms in the other two chains are located at z + 0.287, ϕ − 108° and at z + 0.574, ϕ − 216°, respectively (51). To determine potential interaction between hydrophobic residues of different protomers, we developed a computer program, which used the conditions described by Hoffmann *et al.* (48), to calculate distances between β-carbons of hydrophobic residues of different associated protomers at various relative orientations of the protomers. By using the program, we verified a putative hydrophobic reaction edge (15) and located the saccharide-bearing residues and cysteine residues with respect to it. In these calculations, we assumed 30 residues/turn of the triple helix because Fraser *et al.* (52) found that value for type I collagen and Siebold *et al.* (15) found that 1) its use yielded maximum hydrophobic interaction, and 2) increasing the pitch to 39 made no significant change in the location of the reaction edge. In our description of the resulting extension of the model of Siebold *et al.* (15), we used residue numbers that correspond to those of the cDNA-derived sequence (17–20). However, to facilitate reference to the paper of Siebold *et al.* (15), we also note the numbers that they used. The sets of numbers differ because of the inclusion of gaps in the sequence of α1(IV) to provide maximum alignment with the α2(IV) sequence (15) and because of the signal sequences included in the cDNA-derived sequences.

RESULTS

Characterization of Long and Short Forms of 7 S Domain—These domains were characterized with respect to yield, purity, morphological appearance, and chemical composition. The yield of 7 Sl domain was 22 mg, dry weight/kg, wet weight of cortex (range, 19–32 mg/kg; four experiments) or 0.47 mg/g, dry weight of extracted cortex. The yield of 7 Ss domain from GBM on a dry weight basis was 6.7 mg/g (range, 6.2–7.3 mg/g; three experiments). Electron micrographs of rotary-shadowed samples confirmed the identity of the preparations by the morphological appearance and size of the 7 Sl and 7 Ss molecules as reported by others (15, 53).

The amino acid and carbohydrate composition of the 7 Sl and 7 Ss domains are typical for collagen IV as indicated by the presence of 3-Hyp, the relatively high Cys and Hyl content, and the presence of carbohydrates other than Glc and Gal (Table I). The Cys content is highly enriched in the short form compared with the long form. Carbohydrates are present in roughly equal amounts in both preparations. The sugar composition of both forms points to the presence of Hyl- and Asn-linked oligosaccharides.

In subsequent paragraphs, the determination of the structure and location of the Asn-linked oligosaccharides is described. First, glycopeptides of defined size were prepared by tryptic digestion and chromatographic separation. Then, Asn-linked saccharide and peptide moieties were released from

TABLE I

Amino acid and carbohydrate composition of long and short form 7 S domain from bovine collagen IV

Values are based on residues/1,000 amino acid residues and are given as means ± S.D. with the number of preparations analyzed shown in parentheses.

Component	7 S domain of collagen IV	
	7 Sl[a] (4)	7 Ss[a] (3)
	residues/1,000 amino acid residues	
Amino acid		
3-Hyp[b]	10.0 ± 5.6	10.9 ± 4.7
4-Hyp[b]	117.0 ± 7.2	118.3 ± 3.4
Asx	48.0 ± 1.5	49.9 ± 2.2
Thr	25.7 ± 2.6	27.6 ± 1.1
Ser	26.1 ± 2.3	26.0 ± 2.0
Glx	101.6 ± 8.0	88.1 ± 3.0
Pro	69.9 ± 2.4	69.2 ± 2.2
Gly	331.2 ± 12.5	332.6 ± 8.9
Ala	20.0 ± 5.8	21.4 ± 3.9
Val	30.5 ± 3.0	33.7 ± 3.6
½Cys[c]	14.9 ± 3.7	24.2 ± 4.3
Met[d]	9.5 ± 4.3	11.0 ± 2.3
Ile	19.9 ± 2.1	26.2 ± 3.7
Leu	50.2 ± 1.6	44.6 ± 3.8
Tyr	11.5 ± 1.9	11.4 ± 1.8
Phe	20.8 ± 4.1	17.3 ± 2.0
His	7.9 ± 2.0	9.9 ± 1.1
Hyl	46.3 ± 1.9	37.9 ± 2.7
Lys	6.8 ± 3.1	3.6 ± 1.0
Arg	32.4 ± 1.5	36.3 ± 1.8
Carbohydrate		
Fuc	9.1 ± 4.0	11.2 ± 4.8
Man	17.6 ± 0.9	17.2 ± 3.8
Gal	46.8 ± 2.1	44.4 ± 7.5
Glc	43.9 ± 2.2	35.6 ± 5.2
GalNAc	1.5 ± 0.7	0.7 ± 0.5
GlcNAc	16.4 ± 4.0	17.4 ± 2.8
NeuAc	1.4 ± 0.7	1.3 ± 0.7

[a] 7 Sl was isolated from cortex and 7 Ss from GBM.
[b] 3-Hyp and 4-Hyp were both calculated using 4-Hyp as standard.
[c] ½Cys reflects the sum of half-cystine and cysteic acid.
[d] Met reflects the sum on methionine and methionine sulfoxides.

each other, purified, and analyzed to determine their structure.

Isolation of Glycopeptides Containing Mannose—For the detection of glycopeptides containing *N*-linked oligosaccharides, Man served as a marker. Two Man-rich glycopeptide fractions (A and B) were obtained upon reversed-phase HPLC of tryptic digests of reduced and *S*-aminoethylated 7 S domain preparations (Fig. 2). By carbohydrate analysis, A and B contain more than 70% of the Man (and also of GlcNAc and Fuc) of all fractions together and were the only two with more Man than Glc (Fig. 1, Table II). Fractions A and B are designated GP-1 and GP-2, respectively, and as shown below, GP-1 is derived from the α1(IV) collagen chain and GP-2 from the α2(IV) collagen chain.

Deglycosylation and Purification of Saccharide and Peptide Moieties—GP-1 and GP-2 were incubated with peptide *N*-glycosidase F, and the products were resolved by reversed-phase HPLC (Fig. 2). After 48 h of incubation, deglycosylation was >99% complete. The deglycosylated peptides (Fig. 2, *pools B*) from GP-1 and GP-2 were further purified for sequence analysis by anion-exchange chromatography with Mono-Q resin (see "Materials and Methods," data not shown). The peptide moiety of GP-2 bound to the column, whereas that of GP-1 did not. The oligosaccharide moieties of GP-1 and GP-2 (Fig. 2, *pools A*) were used for structural analysis. In the ^{1}H NMR analyses (see below), no signals were detected that would be typical of a reducing GlcNAc-2 residue (54), which shows that the peptide *N*-glycosidase F preparation used in these studies was free of contaminating endo-F activity.

Localization of N-Glycosylation Sites—To identify the site(s) where saccharides are attached to Asn, the amino acid sequence of GP-1 and GP-2 was determined before and after deglycosylation with peptide *N*-glycosidase F. The advantage of using peptide *N*-glycosidase F is not only that it releases *N*-linked saccharides, leaving the carbohydrate moiety intact, but it also converts Asn to Asp and thus adds a negative charge to the peptide moiety, providing further evidence for *N*-linkage of the oligosaccharide (36, 54–57).

GP-1 exhibits a completely collagenous amino acid sequence with Gly at every third position, and it contains several 4-Hyp residues (Fig. 3). At position 33, a blank occurred, representing a potential site for *N*-glycosylation by its occurrence in the first position of the amino acid sequence -Asn-Xaa-Thr/Ser- (58). In the de-*N*-glycosylated form, residue 33 was identified as Asp, which indicates that this residue is an *N*-glycosylated Asn in the intact peptide. At position 36, another blank occurred both before and after enzymatic deglycosylation; it is considered to be an *O*-glycosylated Hyl residue, since Glc and Gal were detected in GP-1 (Table II) and Hyl was identified in the compositional analysis of a hydrolysate from de-*N*-glycosylated peptide (Table III). On the basis of the amino acid composition data and the tryptic origin of GP-1, Arg was assigned as the carboxyl-terminal amino acid residue (residue 39). Comparison of the GP-1 sequence with the reported sequences of the α1 and α2 chains reveals that GP-1 represents residues 94–132 of the α1(IV) collagen chain (14, 17, 20, 59). Hence, the *N*-linked oligosaccharide in the 7 S domain is attached to Asn-126 (Asn-102 (15)) of the α1(IV) chain.

GP-2 also has a completely collagenous amino acid sequence, but the sequence is different from that of GP-1 (Fig. 3). At position 9, a blank occurred, followed by a Thr residue at position 11, indicative of a potential glycosylated Asn analogous to GP-1. After de-*N*-glycosylation, residue 9 was identified as Asp, which indicates that this residue is *N*-glycosylated Asn in GP-2. Moreover, enzymatic deglycosylation altered the charge from neutral to negative, as indicated by the change in elution position of the peptide on anion-exchange chromatography, which is consistent with conversion of Asn to Asp. Amino acid composition analysis of the peptide moiety of GP-2 further indicated that four amino acids, *i.e.* Leu, Gly, Phe, and Arg, were not accounted for by the sequence analysis (Table III). These are considered to represent the last four carboxyl-terminal residues, which could not be detected because of low amounts of available peptide (Table III). Because trypsin was used to generate the glycopeptide, Arg represents the ultimate carboxyl-terminal residue (Fig. 3). Thus, from analysis of sequence and composition, the total length of GP-2 is 24–27 amino acid residues. Comparison of the sequence of GP-2 with that of the α1(IV) and α2(IV) chains reveals that GP-2 represents residues 129–153 of the α2(IV) chain (15, 18, 19, 60–62). Hence, the α2(IV) chain of the 7 S domain contains an oligosaccharide *N*-linked to Asn-138 (Asn-102 (15)).

The presence of Glc in GP-2 suggested the presence of a glycosylated Hyl residue. However, no Hyl was found in the compositional analysis of the de-*N*-glycosylated peptide of GP-2. This suggested that a contaminating peptide containing glycosylated Hyl had been removed from GP-2 after de-*N*-glycosylation and subsequent reversed-phase HPLC.

The amino acid sequence data of both glycopeptides were the same whether the glycopeptides were derived from the long or short form of the 7 S domain. Therefore, the 7 S domain of collagen IV from GBM contains at least two

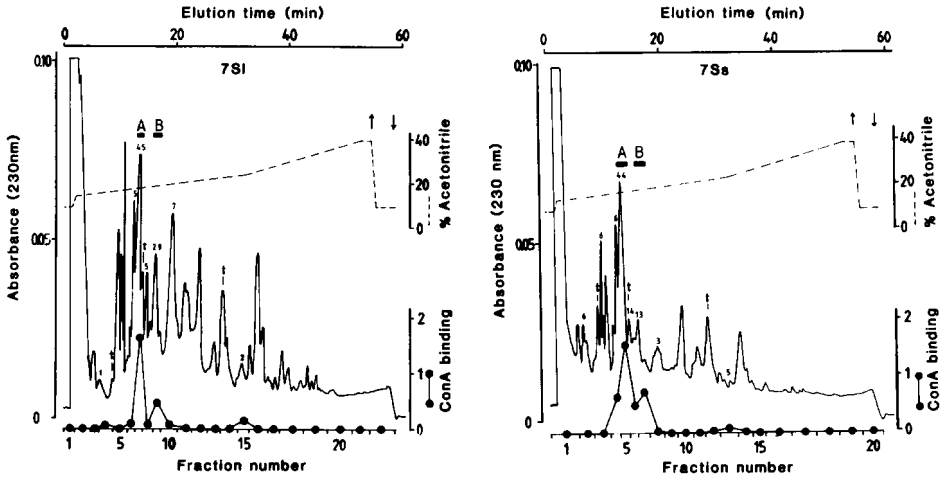


FIG. 1. **Isolation of glycopeptides 1 and 2 from tryptic digests of 7 S domain by reversed-phase HPLC.** Tryptic digests of 2.5 mg of reduced and *S*-aminoethylated 7 Sl (*left*) and 7 Ss (*right*) were subjected to reversed-phase HPLC as described under "Materials and Methods." The size of collected fractions depended on the shape of peaks in the 230-nm absorbance profile and is indicated at the *bottom line*. Elution time is given at the *top line*. Aliquots of each fraction were taken for analysis of sugar composition by gas-liquid chromatography and the direct ConA test. Above the absorbance profile, the amount of Man is expressed as a percentage of the total amount of this sugar detected in all fractions together. No sugars were found in the unbound fraction (fractions *1* and *2*). The direct ConA test was performed on equally diluted samples. The selected pools *A* and *B* contain GP-1 and GP-2, respectively, and were used for further study of primary structure. The shape of the acetonitrile gradient is depicted by the *dashed line*. *Arrows up* and *down* indicate that the flow rate was increased from 1 to 3 ml/min and decreased to 1 ml/min, respectively; *t* shows the position of peaks from trypsin only, obtained from a separate run with a control sample.

TABLE II

Carbohydrate composition of glycopeptides 1 and 2 from the 7 S domain of bovine collagen IV

Glycopeptides 1 and 2 were obtained from reversed-phase HPLC of 7 S preparations as described in Fig. 1. Values are presented as residues relative to Man residues set at 3 and are given as means ± S.D. of three preparations each.

Sugar	Glycopeptide 1	Glycopeptide 2
Fuc	1.3 ± 0.3	1.0 ± 0.7
Man	3	3
Gal	2.5 ± 0.2	2.4 ± 0.6
Glc	1.4 ± 0.1	1.6 ± 0.6
GalNAc	0.2 ± 0.06	0.2 ± 0.08
GlcNAc[a]	(1+) 2.7 ± 0.6	(1+) 2.2 ± 0.4
NeuAc	0.2 ± 0.10	0.2 ± 0.06

[a] The numerical values in parentheses mean that provided one oligosaccharide unit/peptide is present, 1 GlcNAc residue has theoretically to be added to correct for the core GlcNAc, *N*-linked to Asn, which is not released in the assay.

different α chains with *N*-linked oligosaccharides: one is derived from the α1(IV) chain, and the other from the α2(IV) chain.

Structural Analysis of the N-Linked Oligosccharides—NMR analysis revealed that the *N*-linked saccharides from both GP-1 and GP-2 (Fig. 1, *pools A*) are mixtures of similar structures derived from the biantennary *N*-acetyllactosamine type, varying in substitution at their nonreducing ends. In the case of GP-1 (Fig. 4, Table IV), 80% is substituted by Fucα at the C6 of GlcNAc-1, 80% by bisecting GlcNAc-9 at C4 of Man-3, and 50 and 60% by Gal-6 and -6′ at the C4 of the GlcNAc-5 and -5′ residues, respectively. Substitution by Galα at the C3 of Gal-6 and -6′ occurs in 10% of the oligosaccharides. The distribution of this Galα over the two branches remains unclear. In the NMR spectra, there is no evidence for the presence of sialic acid. From these data, it can be concluded that at least 64% of the *N*-linked saccharides in GP-1 existed with Fuc and bisecting GlcNAc-9 in the same structure, of which 20–90% was substituted by Gal-6 or -6′ (with and without Galα) at C4 of GlcNAc-5 and -5′. The heterogeneity was also evident from subsequent separation of the saccharide mixture by HPLC over a polar amino column (63), in which four main peaks and at least four minor peaks were observed (eluent, 35% H_2O in acetonitrile), but yields were too low for NMR analysis. The NMR spectra obtained for the saccharide moiety of GP-2 were very similar to those of GP-1, indicating that the saccharide moieties have essentially identical structures.

Location and Orientation of Saccharides and Hydrophobic and Cysteine Residues with Respect to the Surface of the 7 S Triple Helical Domain—In Fig. 5 are plotted values of *z*, the axial translation in nm, and ϕ, the azimuthal angle in degrees for the β-carbons of hydrophobic residues, residues containing saccharides, and Cys residues. The calculated disaccharide positions are based on the assumption that every Hyl residue contains a disaccharide (see "Discussion"). In Fig. 5, we include the hydrophobic reaction edge with two hydrophobic faces at each end (*shaded areas A* and *B*) that Siebold *et al.* (15) determined were the most likely to be involved in tetramerization of the protomers.[3] Contact between A faces and between B faces of different protomers leads to aggregation driven by interprotomer interactions between hydrophobic residues in the reaction edges; the approximately 90° difference between the boundaries of a reaction edge limits the aggregation to tetramer formation. We calculated the distances between the β-carbons of hydrophobic residues of 7 S

[3] This was based on the extent of interaction with their counterparts on other protomers and the overlap of protomers in the tetramer which could give rise to stabilizing interprotomer cross-links. The hydrophobic faces, with interprotomer contact points at ϕ values of 36 and 306° constitute a self-interacting reaction edge of approximate width 90°. The boundaries were determined by rotation of protomers in 18° increments and measuring the number of interprotomer hydrophobic interactions at each orientation as determined by the distance between hydrophobic residues. The 90° width is approximate (see also Ref. 15) because of the relatively small rate of change of calculated interresidue distance with ϕ.

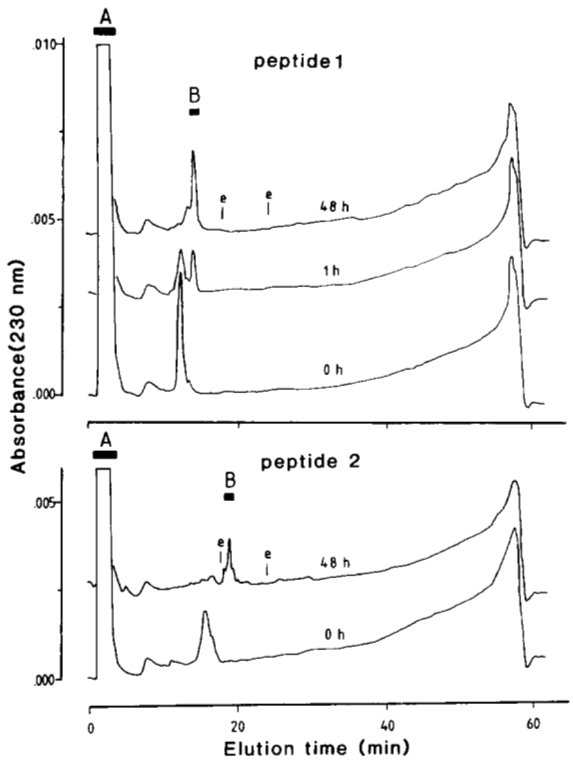


FIG. 2. **Separation of peptide and saccharide moieties.** Glycopeptides 1 and 2 were each incubated with peptide *N*-glycosidase F for 48 h under conditions described under "Materials and Methods." Following this treatment, the samples were acidified to pH 2 with trifluoroacetic acid and directly subjected to reversed-phase HPLC using the same acetonitrile gradient system used for preparing GP-1 and GP-2 (see "Materials and Methods" and "Results"). To illustrate the change in elution by the de-*N*-glycosylation of the peptides, the elution profile of the samples before and after adding the enzyme is shown for both peptides; glycopeptide 1 after 1 h of digestion is also shown. The elution position of the enzyme run separately at high concentration is indicated by *e* (two peaks). *Bars* in the run of the samples after 48 h of digestion indicate the pools selected for further investigation, with *A* as the saccharide pool and *B* as the deglycosylated peptide pool. Pools *A* were lyophilized and desalted with Bio-Gel P2 ad indicated under "Materials and Methods."

protomers and verified the finding (15) that these hydrophobic faces could interact to cause tetramer formation.

DISCUSSION

In the present study, the 7 S tetramer of GBM collagen IV was excised by collagenase digestion and characterized with respect to its dimensions, composition, and location and structure of the putative Asn-linked oligosaccharide units. The overall dimensions and composition of the tetramer were found to be similar to that of Engelbreth-Holm-Swarm sarcoma-tumor matrix and placental basement membrane (32, 53). This revealed that collagen IV of GBM is also assembled into a suprastructure by tetramerization of the 7 S domain and that the tetramer is heavily glycosylated. As deduced herein, its carbohydrate units are likely to play an important role in the mechanism of tetramer assembly.

The carbohydrate units of the 7 S tetramer are of two types, Asn-lined oligosaccharide units and Hyl-linked disaccharide units. Our studies of glycopeptides GP-1 and GP-2 provide unambiguous evidence for the presence and location of Asn-linked oligosaccharide units within the tetramer and reveal the detailed structure of these units. Each chain contains one oligosaccharide unit, which is attached to Asn-126 of the α1 chain and Asn-138 of the α2 chain. The oligosaccharide is of the biantennary type and is heterogeneous with respect to the number of monosacharides at the nonreducing termini. Its structure on the α1 chain is essentially identical to that on the α2 chain. Thus, a total of three Asn-linked oligosaccharide units is present in the 7 S domain of the triple helical protomer. The 7 S domain also contains 18 disaccharide units/protomer (6 each per each α1 and α2 chain, as deduced from chemical sequence analyses (14, 15)).[4]

Thus, the 7 S tetramer contains a total of 12 oligosaccharide and 72 disaccharide units. These values are in good agreement with the values of 10 and 63, respectively, calculated from the composition data in Table I.[5] Potentially, these units have important roles in the mechanism of tetramer assembly, as revealed by their strategic distribution with respect to the surface of the triple helical protomer and 7 S tetramer and their physicochemical properties.

The oligosaccharides are expected to play a key role in determining the geometry of the tetramer because of their size and location. Each triple helical protomer has three bulky oligosaccharides (one per chain) at essentially the same axial distance and distributed almost uniformly around the circumference of the protomer. Thus, it is likely that one of the oligosaccharides, or two acting together, would restrict the longitudinal overlap of the triple helical domains of monomers in the tetramer because the oligosaccharides have a size considerably larger than the diameter of a triple helix. Any model for tetramerization must include the steric effects resulting from each protomer chain containing a bulky oligosaccharide.

To illustrate the steric effect of the oligosaccharides, we considered the 7 S tetramer model recently published by Siebold *et al.* (15). It can be seen in Fig. 5 that one of the oligosaccharides, with a z value of 24.5 nm and a ϕ value of 23.6 ° (linked to the second α1 chain in the $\alpha 2\alpha 1\alpha 1$ order of the model), is located in the ϕ range of the reaction edge near the *B* zone. This oligosaccharide would sterically restrict the longitudinal overlap of the triple helical domains of monomers in the tetramer.[6] Thus, the maximum value for the triple helical overlap would be 84 amino acid residues for a chain, corresponding to a length of 24 nm.

[4] The disaccharides of the 7 S domain are attached to Hyl in the α1 chain at positions 45, 48, 78, 90, 129, and 156 and in the α2 chain at positions 57, 87, 90, 102, 165, and 168. The position of attachment is based on chemical sequence analyses (14, 15), and the position numbers are based on the complete sequences of the α1 and α2 chains recently from cDNA clones (17, 18).

[5] The experimental value of 10 oligosaccharides was calculated from the Man content of the 7 Ss form (Table I) using 3 Man residues/unit, and the value of 63 disaccharide units/unit was calculated from the Glc content (Table I) using 1 Glc residue/unit. However, these are approximate values because they are based on the assumption that each chain of the 7 S domain, which represents a collagenase-derived fragment, is 145 residues long. This length was reported by Siebold *et al.* (15) for the 7 S fragment of the α2 chain. However, the length of the α2 chain may vary with the conditions of collagenase digestion, and the length of the α1 chain is unknown.

[6] The same conclusion was reached by consideration of the model for triple helical $(GlyProHyp)_n$ based on x-ray diffraction studies on collagen I located in kangaroo tail tendon (52). This model differs from the $(GlyProPro)_n$ model because of differences in the geometry of the residues at the X and Y positions, leading to ϕ values of 2.4 and $-45.2°$ for the β-carbons of the X and Y residues of the first triplet of the α2 chain, instead of 38.8 and $-12.6°$, respectively. This would result in a shift in ϕ of $-36.4°$ for the boundaries of the reaction edge because the interprotomer hydrophobic interactions involve residues in the X position. However, the position of the β-carbon of the Asn linked to the oligosaccharide of the second α1 chain (a Y residue) would be shifted by a similar amount, $-32.6°$, resulting in a net shift of only 3.8° for the Asn β-carbon relative to the reaction edge.

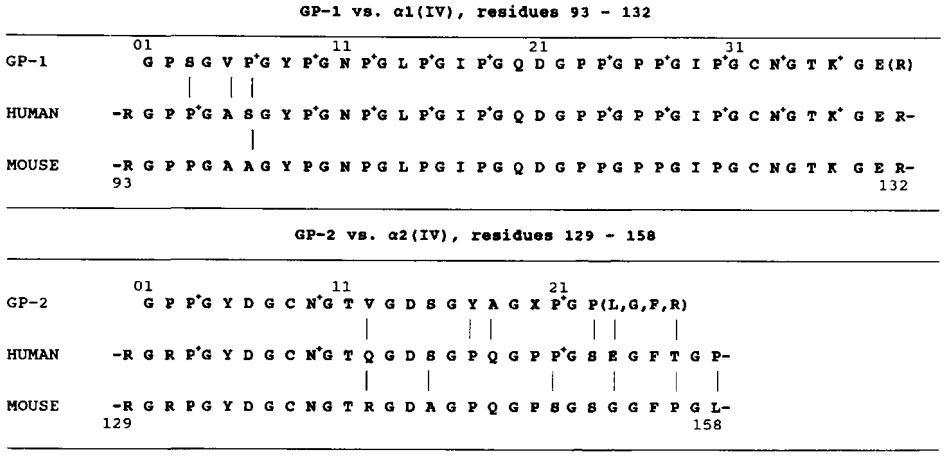


FIG. 3. **Identification of the *N*-linked glycosylation sites of the 7 S domain of bovine collagen IV.** The amino acid sequences presented were obtained from intact GP-1 and GP-2 (two experiments each), as isolated in Fig. 1, and from de-*N*-glycosylated GP-1 (one experiment) and GP-2 (two experiments), isolated as described under "Materials and Methods." P^+, hydroxyproline; K^+, glycosylated Hyl; N^+, *N*-glycosylated Asn; *C*, detected as *S*-aminoethylcysteine. For comparison and α chain identification, the known sequences of the putative *N*-glycosylated region of collagen IV α chains from human placenta (obtained by peptide and nucleotide sequencing (14, 15, 20)) and from mouse (by nucleotide sequencing (17, 18)) are given. The Asn in position 126 in GP-1 and position 138 in GP-2 initially appeared in the sequence analysis as a blank but, after deglycosylation, appeared as Asp, which is characteristic for a *N*-glycosylated Asn residue. Residue numbering starts from the amino terminus and is based upon the amino acid sequence of the complete α1 and α2 chains, including the signal peptides obtained by nucleotide sequencing.

TABLE III

Amino acid composition of glycopeptides 1 and 2

The data were obtained by amino acid analysis of hydrolyzed samples of purified de-*N*-glycosylated peptide samples. Total amount of amino acid residues analyzed in each sample was 305 and 37 nmol for glycopeptides 1 and 2, respectively. The values are expressed in residues/peptide, assuming that glycopeptides 1 and 2 are 39 and 24 residues long, respectively. If differences were encountered, values in parentheses are given which reflect the balance in residues between the amino acid composition of hydrolysate and the sequence data. Values in parentheses were determined by subtracting values from the sequence analysis (Fig. 3) from those of hydrolysate and rounding off to the closest integer.

Amino acid	Glycopeptide 1	Glycopeptide 2
	residues/peptide	
4-Hyp	7.8	2.1
Asx	4.0 (+1)	3.1
Thr	1.2	1.0
Ser	1.2	0.9
Glx	2.4	0.1
Pro	2.7	2.3
Gly	11.9 (−1)	9.2 (+1)
Ala	ND[a]	1.0
Val	1.0	1.0
Met	ND	ND
Ile	1.9	ND
Leu	1.0	0.8 (+1)
Tyr	1.0	1.8
Phe	ND	1.0 (+1)
His	ND	ND
HOLys[b]	1.0	ND
Lys	ND	0.1
Cys[c]	1.0	1.0
Arg	0.9 (+1)	0.7 (+1)

[a] ND, not detected.
[b] HOLys, hydroxylysine.
[c] Cys was analyzed as *S*-aminoethylcysteine.

The location of the α1 chain oligosaccharide on the surface of the triple helix, relative to the *A* and *B* zones of the reaction edge, its size relative to the diameter of the triple helix, and its orientation and location on the tetramer are depicted by the schematic in Fig. 6. The schematic is based on the tetramer model described by Siebold *et al.*. (15) in which four protomers are arranged in an antiparallel fashion with their amino-terminal triple helical domains overlapping. In the tetramer structure, the oligosaccharide of one protomer is positioned next to the amino-terminal end of a second protomer. Steric hindrance due to the oligosaccharide is predicted to restrict the longitudinal overlap of two protomers as shown in Fig. 6 and in the three-dimensional model in Fig. 7*A*. Although this result is based on the model in which the chain order is α2α1α1, an α1 chain oligosaccharide would be positioned within the approximately 90° azimuthal span of the reaction edge regardless of the chain order,[7] thus still restricting the longitudinal overlap of the protomers.

The total length of the maximal overlap of two protomers is the sum of the length of the triple helical overlap zone, 24 nm on the basis of steric hindrance due to the oligosaccharide unit on an α1 chain, and that of the non-triple helical domain that precedes the triple helical domain (Fig. 7*A*). In the isolated 7 Sl domain, the non-triple helical regions of the α1 and α2 chains contain 15 and 18 residues, respectively (14, 15), corresponding to 48 residues/protomer. The diameter of the non-triple helical domain, assuming a spherical shape, is 2.4 nm.[8] Thus, the total overlap has a theoretical length of 26–27 nm. These features and dimensions are consistent with those of the 7 Ss domain revealed by electron microscopy. We observed that the GBM collagen IV 7 S domain is a rod with a globule at each end and has an overall length of 30.1 ± 1.4 nm, in good agreement with the value 28 nm reported by Kühn *et al.* (53) for the 7 S domain of placental collagen IV. The diffference between the predicted maximal overlap and

[7] An azimuthal angle of 239° was found for the Asn 126-linked oligosaccharide of the first α1 chain in the model in which the chain order is α1α2α1, between the edges of the *A* and *B* hydrophobic faces at 198 and 288°, respectively (15, and the corresponding values for the model with chain order α1α1α2 are 131, 90, and 180°.

[8] The calculation is based on the assumption that the amino acid residues have a mean residue weight of 110–115 and a partial specific volume of 0.70–0.75 cm^3/g and thus a partial molecular volume of 1.3×10^{-22} cm^3.

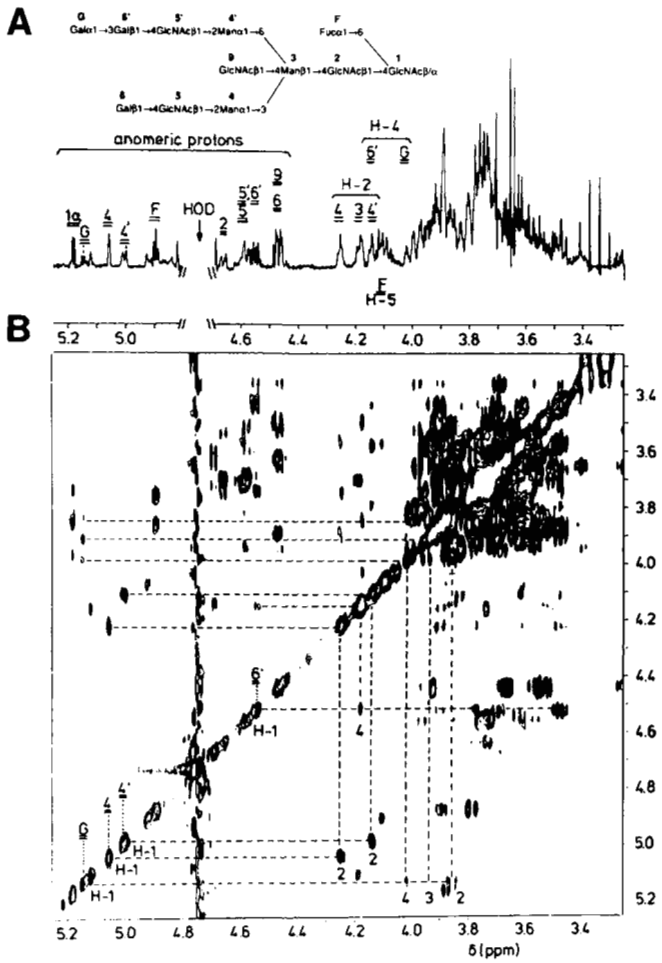


FIG. 4. **500-MHz ^{1}H NMR spectra of *N*-linked oligosaccharides from glycopeptides 1 and 2.** ^{1}H NMR analyses were performed on mixtures of saccharides obtained from glycopeptide 1 (three independent experiments) and glycopeptide 2 (two independent experiments). The spectra of the two oligosaccharide mixtures revealed close structural similarities between them. By combining the available material from glycopeptide 1 into one pool, the recording of a two-dimensional HOHAHA spectrum was possible. The figure shows the one-dimensional spectrum (*A*) and the two-dimensional HOHAHA spectrum (*B*), which were recorded for the combined material from glycopeptide 1. The *inset* in *panel A* shows the most extended structure in the mixture derived from the analyses, for which the ^{1}H chemical shift values are presented in Table IV. In the two-dimensional spectrum, the cross-peaks between the *H-1* and *H-2* signals of Man-4 and -4′ are indicated. These were used to confirm the branching pattern of the proposed structure in the *inset*. The track drawn from Galα(G) *H-1* parallel to the ω_2-axis shows the cross-peaks to *H-2*, *H-3*, and *H-4*. The cross-peak between Gal-6′ *H-1* and *H-4* is characteristic for a terminal Galα1→3 substitution. The trimannosylchitobiose core structure is present in all oligosaccharides in the mixture. Based upon the analyses of the combined material, the frequencies of occurrence of certain terminal sugar residues, expressed as percentages of that of the core structure, are as follows. The degree of fucosylation at GlcNAc-1 is approximately 80%, a value obtained by comparing the integral for the Fuc *H-1* signals with the sum of the integrals for the Man-4 *H-1* signals at δ 5.057 and δ 5.121; the percentage of structures with a bisecting GlcNAc-9 residue is approximately 80%, as is revealed by the ratio of integrated areas of the Man-4 *H-1* signals at δ 5.057 (with GlcNAc-9) and δ 5.121 (without GlcNAc-9); roughly equal areas of the Man-4′ *H-1* signals at δ 5.011 and δ 4.999 are interpreted to mean that there are as many bisecting GlcNAc-9 containing structures terminating in GlcNAc-5′ as those bearing a Gal-6′. Likewise, the difference in areas of signals at δ 4.464 and δ 4.554 points to the presence of structures with intersecting GlcNAc-9 of which 40% terminate with GlcNAc-5 and 60% with Gal-6 (41). About 10% of the branches (20% in the case of glycopeptide 2) are terminated by Galα1→3 linked mainly to Gal-6′; however, the exact distribution of the terminal Galα residues over the two branches remains unclear.

TABLE IV

^{1}H Chemical shifts of assigned protons for the largest structure in the mixture of N-linked saccharides from the α1 (IV) 7 S domain of bovine collagen IV, together with those from reference compound F

Chemical shifts are given at 300 K relative to acetone (δ 2.225) and were derived from the one-dimensional and two-dimensional ^{1}H NMR spectra of Fig. 4. —, not reported; ND, not determined. Analyses of the mixture further showed that the other components were smaller forms of this structure by the absence of one or more of the sugars at the nonreducing termini, as discussed under "Results" and in Fig. 4.

Residue	Proton	G-6′-5′-4′ \ F 9-3-2-1 / 6-5-4	Component E[a]
GlcNAc-1	H1	(1α) 5.180, (1β) 4.694	—
	H2	(1α) 3.88, (1β) ND	4.188
	NAc	2.038	—
GlcNAc-2	H1	4.66	4.641
	H2	3.74	—
	NAc	2.094	—
Man-3	H1	ND	4.709
	H2	4.18	4.188
Man-4	H1	5.057	5.062
	H2	4.252	4.251
Man-4′	H1	5.011	5.003
	H2	4.14	4.146
GlcNAc-5	H1	4.582	4.583
	NAc	2.051	—
GlcNAc-5′	H1	4.578	4.583
	NAc	2.049	—
Gal-6	H1	4.464	4.475
Gal-6′	H1	4.547	—
GlcNAc-9	H1	4.470	4.479
	H2	ND	3.69
	NAc	2.063	—
Galα(G)	H1	5.146	5.149
	H2	3.86	3.865
	H3	3.95	—
	H4	4.02	—
Fucα(F)	H1	(1α) 4.891, (1β) 4.899	4.903
	H5	4.100	4.072
	CH_3	(1α) 1.207, (1β) 1.219	1.225

[a] Component E is used as a reference compound as previously described (67).

the observed value may reflect contributions to the latter from the oligosaccharide units, the magnitude of which would depend on whether they interact with amino acid residues, as is the case for the IgG immunoglobulin molecule (64), or instead project from the 7 S tetramer as shown in Fig. 7, *A* and *B*. In addition, the platinum in the replica used for electron microscopy and potential additional amino acid residues on the arms of the tetramer will contribute to the length of the overlap. Therefore, we conclude that the maximum length of the overlap zone between protomers is determined by steric hindrance from the oligosaccharide units.

In addition, it is likely that the oligosaccharide and disaccharide units contribute to the stability of the tetramer because of (*a*) their occurrence within the overlap region and at its boundary, (*b*) their restricted distribution within the overlap region, (*c*) their hydrophilic nature. All three oligosaccharides are located in a narrow region which circumscribes the triple helical domain of the protomer and demarcates the boundary of the overlap region (Figs. 5 and 7*A*). Twelve disaccharides of the protomer are located within the overlap region (Fig. 5). In the tetramer, eight of these are distributed about the exterior and four are buried inside (Figs. 5 and 7, *A* and *B*). Thus, the exterior surface of the tetramer, in the overlap region, contains 32 disaccharide units and 12 oligosaccharide units, which corresponds to 220 monosaccharide

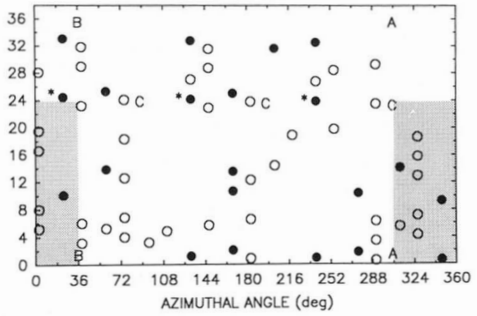


FIG. 5. **Location and orientation of saccharides and hydrophobic and cysteine residues with respect to the surface of the 7 S triple helical domain.** Shown is a two-dimensional representation of the surface that shows the calculated axial distance in nm, z, and the azimuthal angle in degrees, ϕ, for the β-carbons of four classes of residues. The β-carbons of residues with covalently bound disaccharides are indicated by *filled circles*, Asn-linked oligosaccharides by *asterisks* alongside *filled circles*, hydrophobic residues by *open circles*, and cysteine residues by *C* alongside *open circles*. See text for positions within the α chains. The calculations were carried out as described under "Materials and Methods." The *shaded areas*, *A* (*right*) and *B* (*left*) mark the approximate ends of the self-interacting hydrophobic reaction edge involved in tetramer formation, with two boundaries at ϕ values of 306 and 36 ° (15). The upper boundary of the reaction edge has a maximum value of about 24 nm and is determined by the large hydrophilic oligosaccharide unit attached to an Asn residue whose β-carbon is at $z = 24.5$ nm and $\phi = 23.6$ °.

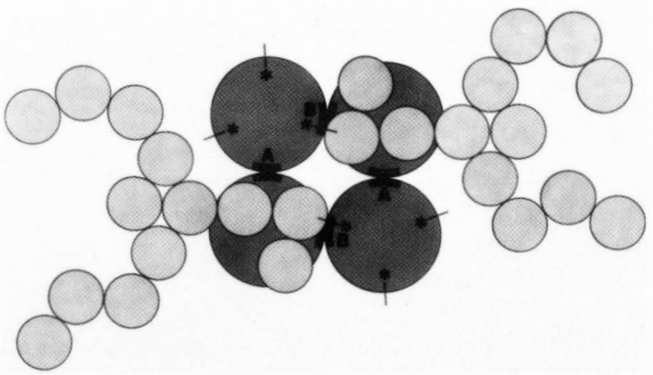


FIG. 6. **Schematic (to scale) depicting an end view of the 7 S tetramer and the steric restrictions on assembly caused by *N*-linked oligosaccharides.** The four 7 S triple helical aggregation domains (*large circles*) interact by means of the *A* and *B* hydrophobic faces as shown. The diameter of the triple helix is 1.3 nm, as calculated from the M_r/length ratio and the partial specific volume (65). Adjacent triple helices interact through either *A* faces or *B* faces and are antiparallel to each other (15). The *upper left* and *lower right* triple helices are parallel to each other; relative azimuthal positions of oligosaccharides *N*-linked to Asn-126 (Asn-102 (15)) of each α1(IV) chain and to Asn-138 (Asn-102 (15)) of the α2(IV) chain in these two helices are indicated by *asterisks*. Schematics of two of the oligosaccharides are shown, with their monomer residues indicated by *small circles* of diameter 0.67 nm, the value calculated for Glc residues from the density of Glc solutions as a function of concentration (66), assuming a spherical shape for Glc. Each of the two oligosaccharides depicted is bound to an α1(IV) chain corresponding to the second one in the $\alpha 2\alpha 1\alpha 1$ order of the triple helix model (see "Materials and Methods" and Fig. 5). The actual conformation of the oligosaccharides is unknown. The oligosaccharides restrict the longitudinal overlap of interacting triple helices, requiring the ends of the *shaded* triple helices to lie below the oligosaccharides.

residues, assuming no heterogeneity, whereas the interior contains 16 disaccharides, which corresponds to 32 residues. This asymmetrical distribution of carbohydrate units about the surface of the triple helix and the hydrophobic faces of the interior surface confers to the protomer an amphipathic character which contributes to the stability of the tetramer.

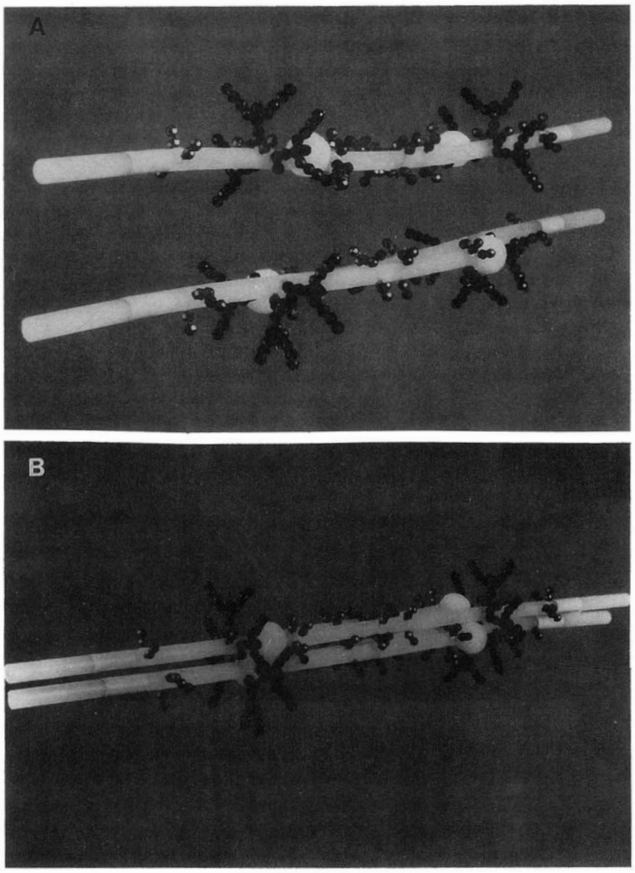


FIG. 7. **Three-dimensional model (to scale) depicting the locations of disaccharides and *N*-linked oligosaccharides of the 7 S domain in its dimeric and tetrameric forms.** *A*, dimers that are formed through interaction of *B* hydrophobic faces of two 7 S monomers (see Figs. 5 and 6 and Ref. 15). The *Y-shaped structures* on the cylinders (triple helices) represent Asn-linked oligosaccharides; a *bead* represents a hexose residue. The *two-beaded structures* represent Glc-Gal- units *O*-linked to Hyl residues. The *white spheres* represent the non-triple helical region, preceding the triple helical region, that consists of the amino-terminal 15 residues of the α1(IV) chains plus the 18 residues of the α2(IV) chain (14, 15). It is assumed that the non-triple helical region is a sphere of (calculated) diameter 2.4 nm. An oligosaccharide on each triple helix determines the maximum extent of longitudinal overlap of the helices when they form an antiparallel dimer, as shown by a *white sphere* in contact with an oligosaccharide. The *A* hydrophobic faces are at the *top* and *bottom sides*, respectively, of the lower and upper dimers. *B*, tetramer formed from interaction of the *A* hydrophobic faces of the dimers described for *panel A*. Note that the oligosaccharides and amino-terminal non-triple helical regions also have the same steric functions in tetramer formation that they do in dimer formation.

Furthermore, the bulky oligo- and disaccharides, by steric hindrance, may restrict the number of ways in which protomers can associate, thus promoting their antiparallel arrangement in the tetramer, as previously postulated (15). This feature may also be important in limiting the association of protomers to tetramerization. Likewise, the interior disaccharide units, because of their bulk, may limit the number of ways in which the dimers can associate through their *B* faces and thus be a determinant of the geometry of the tetramer.

To discover the potential ways in which the oligosaccharide and disaccharide units determine some of the features of the 7 S tetramer, we used the model of Siebold *et al.* (15) based on triple helical $(GlyProPro)_n$. However, the same roles for the saccharides would be retained in a model based on type I collagen, even though the two models would differ in overall three-dimensional structure.[6] Thus, our findings do not allow

a choice between these models but instead illustrate the importance of saccharides in determining the geometry of the 7 S tetramer.

In summary, we conclude that both the oligosaccharide and disaccharide units of the 7 S domain are likely to have a role in determining the geometry and stability of the 7 S tetramer and, therefore, are important elements of the structure-function relationship of basement membrane collagen. These conclusions are based on their abundance, structure, hydrophilic properties, and strategic distribution about the surface of the triple helix.

Acknowledgments—The fruitful discussions about the lectin binding assays with Dr. Joaquin Timoneda and the contributions of Dr. Jon Barr in the amino acid analyses are greatly appreciated. We are indebted to Dr. Julie Hudson and Billie Wisdom for their critical reading of the manuscript.

REFERENCES

1. Timpl, R., Wiedemann, H., Van Velden, V., Furthmayr, H., and Kühn, K. (1981) *Eur. J. Biochem.* **120,** 203–211
2. Siebold, B., Deutzmann, R., and Kühn, K. (1988) *Eur. J. Biochem.* **120,** 203–211
3. Engel, J., and Furthmayr, H. (1987) *Methods Enzymol.* **145,** 3–78
4. Martin, G. R., Timpl, R., and Kühn, K. (1988) *Adv. Protein Chem.* **39,** 1–50
5. Yurchenco, P. D., Tsilibary, E. C., Charonis, A. S., and Furthmayr, H. (1986) *J. Histochem. Cytochem.* **34,** 93–102
6. Tsilibary, E. C., and Charonis, A. S. (1986) *J. Cell Biol.* **103,** 2467–2473
7. Yurchenco, P. D., and Ruben, G. C. (1987) *J. Cell Biol.* **105,** 2559–2568
8. Yurchenco, P. D., and Ruben, G. C. (1988) *Am. J. Pathol.* **132,** 278–291
9. Spiro, R. G. (1967) *J. Biol. Chem.* **242,** 1923–1932
10. Hudson, B. G., and Spiro, R. G. (1972) *J. Biol. Chem.* **247,** 4239–4247
11. Sato, T., and Spiro, R. G. (1976) *J. Biol. Chem.* **251,** 4062–4070
12. Levine, M. J., and Spiro, R. G. (1979) *J. Biol. Chem.* **254,** 8121–8124
13. Dean, D. C., Peczon, B. D., Noelken, M. E., and Hudson, B. G. (1981) *J. Biol. Chem.* **256,** 7543–7548
14. Glanville, R. W., Qian, R.-Q., Siebold, B., Risteli, J., and Kühn, K. (1985) *Eur. J. Biochem.* **152,** 213–219
15. Siebold, B., Qian, R.-Q., Glanville, R. W., Hoffmann, H., Deutzmann, R., and Kühn, K. (1987) *Eur. J. Biochem.* **168,** 569–575
16. Leushner, J. R. A. (1987) *Biochem. Cell Biol.* **65,** 501–506
17. Muthukumaran, G., Blumberg, B., and Kurkinen, M. (1989) *J. Biol. Chem.* **264,** 6310–6317
18. Saus, J., Quinones, S., and MacKrell, A., Blumberg, B., Muthukumaran, G., Pihlajaniemi, T., and Kurkinen, M. (1989) *J. Biol. Chem.* **264,** 6318–6324
19. Hostikka, S. L., and Tryggvason, K. (1988) *J. Biol. Chem.* **263,** 19488–19493
20. Soininen, R., Haka-Risku, T., Prockop, D. J., and Tryggvason, K. (1987) *FEBS Lett.* **225,** 188–194
21. Dubé, S., Fisher, J. W., and Powell, J. S. (1988) *J. Biol. Chem.* **263,** 17516–17521
22. Rademacher, T. W., Parekh, R. B., and Dwek, R. A. (1988) *Annu. Rev. Biochem.* **57,** 785–838
23. Dennis, J. W., Waller, C. A., and Schirrmacher, V. (1984) *J. Cell Biol.* **99,** 1416–1423
24. Matzuk, M. M., Keene, J. L., and Boime, I. (1989) *J. Biol. Chem.* **264,** 2409–2414
25. Sairam, M. R. (1989) *FASEB J.* **3,** 1915–1926
26. Fujiwara, S., Shinkai, H., Deutzmann, R., Paulsson, M., and Timpl, R. (1988) *Biochem. J.* **252,** 453–461
27. Knibbs, R. N., Perini, F., and Goldstein, I. J. (1989) *Biochemistry* **28,** 6379–6392
28. Arumugham, R. G., Hsieh, T.-C. Y., Tanzer, M. L., and Laine, R. A. (1986) *Biochim. Biophys. Acta* **883,** 112–126
29. Farquhar, M. G., Courtoy, P. J., Lemkin, M. C., and Kanwar, Y. S. (1982) in *New Trends in Basement Membrane Research* (Kuehn, K., Schoene, H., and Timpl, R., eds) pp. 9–29, Raven Press, New York
30. Hudson, B. G., Wieslander, J., Wisdom, B. G., Jr., and Noelken, M. E. (1989) *Lav. Invest.* **61,** 256–269
31. Langeveld, J. P. M., Wieslander, J., Timoneda, J., McKinney, P., Butkowski, R. J., Wisdom, B. J., Jr., and Hudson, B. G. (1988) *J. Biol. Chem.* **263,** 10481–10488
32. Risteli, J., Bächinger, H. P., Engel, J., Furthmayr, H., and Timpl, R. (1980) *Eur. J. Biochem.* **108,** 239–250
33. Dixit, S. N., Mainardi, C. L., Beachey, E. H., and Kang, A. H. (1983) *Collagen Relat. Res.* **3,** 263–270
34. Wieslander, J., Langeveld, J., Butkowski, R., Jodlowski, M., Noelken, M., and Hudson, B. G. (1985) *J. Biol. Chem.* **260,** 8564–8570
35. Schwartz, W. E., Smith, P. K., and Royer, G. P. (1980) *Anal. Biochem.* **106,** 43–48
36. Damm, J. B. L., Kamerling, J. P., van Dedem, G. W. K., and Vliegenthart, J. F. G. (1987) *Glycoconjugate J.* **4,** 129–144
37. Bensadoun, A., and Weinstein, D. (1976) *Anal. Biochem.* **70,** 241–250
38. Butkowski, R. J., Wieslander, J., Wisdom, B. J., Barr, J. F., Noelken, M. E., and Hudson, B. G. (1985) *J. Biol. Chem.* **260,** 3739–3747
39. Chaplin, M. F. (1982) *Anal. Biochem.* **123,** 336–341
40. Shotton, D. M., Burke, B. E., and Branton, D. (1979) *J. Mol. Biol.* **131,** 303–329
41. Vliegenthart, J. F. G., Dorland, L., and van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* **41,** 209–374
42. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* **65,** 355–360
43. Homans, S. W., Dwek, R. A., Boyd, J., Soffe, N., and Rademacher, T. W. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1202–1205
44. Inagaki, F., Kohda, D., Kodama, C., and Suzuki, A. (1987) *FEBS Lett.* **212,** 91–97
45. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) *J. Biol. Chem.* **256,** 7990–7997
46. Butkowski, R. J., Langeveld, J. P. M., Wieslander, J., Hamilton, J., and Hudson, B. G. (1987) *J. Biol. Chem.* **262,** 7874–7877
47. Goldstein, I. J., and Poretz, R. D. (1986) in *The Lectins* (Liener, I. E., Sharon, N., and Goldstein, I. J., eds) pp. 33–247, Academic Press, Orlando, FL
48. Hoffmann, H., Fietzek, P. P., and Kühn, K. (1978) *J. Mol. Biol.* **125,** 137–165
49. Traub, W., Yonath, A., and Segal, D. M. (1969) *Nature* **221,** 914–917
50. Traub, W., and Piez, K. A. (1971) *Adv. Protein Chem.* **25,** 243–252
51. Harrington, W. F., and Von Hippel, P. H. (1961) *Adv. Protein Chem.* **16,** 1–138
52. Fraser, R. D. B., MacRae, T. P., and Suzuki, E. (1979) *J. Mol. Biol.* **129,** 463–481
53. Kühn, K., Wiedemann, H., Timpl, R., Risteli, J., Dieringer, H., Voss, T., and Glanville, R. W. (1981) *FEBS Lett.* **125,** 123–128
54. Damm, J. B. L., Voshol, H. Hård, K., Kamerling, J. P., van Dedem, G. W. K., and Vliegenthart, J. F. G. (1988) *Glycoconjugate J.* **5,** 221–233
55. Plummer, T. H., Jr., Elder, J. H., Alexander, S., Phelan, A. W., and Tarentino, A. L. (1984) *J. Biol. Chem.* **259,** 10700–10704
56. Risley, J. M., and Van Etten, R. L. (1985) *J. Biol. Chem.* **260,** 15488–15494
57. Carr, S. A., and Roberts, G. D. (1986) *Anal. Biochem.* **157,** 396–406
58. Struck, D. K., and Lennarz, W. J. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W., ed) pp. 35–83, Plenum Publishing Corp., New York
59. Schuppan, D., Glanville, R. W., Timpl, R., Dixit, S. N., and Kang, A. H. (1984) *Biochem. J.* **220,** 227–233
60. Sakurai, Y., Sullivan, M., and Yamada, Y. (1986) *J. Biol. Chem.* **261,** 6654–6657
61. Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R. W., Deutzmann, R., Timpl, R., and Kühn, K. (1986) *Eur. J. Biochem.* **157,** 49–56
62. Blumberg, B., MacKrell, A. J., Olson, P. F., Kurkinen, M., Monson, J. M., Natzle, J. E., and Fessler, J. H. (1987) *J. Biol. Chem.* **262,** 5947–5950
63. Blanken, W. M., Bergh, M. L. E., Koppen, P. L., and Van den Eijnden, D. H. (1985) *Anal. Biochem.* **145,** 322–330
64. Huber, R., Diesenhofer, J., Colman, P. M., and Matsushima, M. (1976) *Nature* **264,** 415–420
65. Rice, R. V., Casassa, E. F., Kerwin, R. E., and Maser, M. D. (1964) *Arch. Biochem. Biophys.* **105,** 409–423
66. Wolf, A. V., Brown, M. G., and Prentiss, P. B. (1975) *in Handbook of Chemistry and Physics* (Weast, R. C., ed) 55th Ed., p. D-205, Chemical Rubber Publishing Co., Cleveland, OH
67. Geyer, R., Geyer, H., Stirm, S., Hunsmann, G., Schneider, J., Dabrowski, U., and Dabrowski, J (1984) *Biochemistry* **23,** 5628–5637