Role of Glycosylation on the Conformation and Chain Dimensions of O-Linked Glycoproteins: Light-Scattering Studies of Ovine Submaxillary Mucin†

Randal Shogren,‡ Thomas A. Gerken,* and Neil Jentoft

W. A. Bernbaum Center for Cystic Fibrosis Research, Department of Pediatrics and Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106

Received September 27, 1988; Revised Manuscript Received February 23, 1989

ABSTRACT: The effect of carbohydrate on the conformation and chain dimensions of mucous glycoproteins was investigated by using light-scattering and circular dichroism studies of native, asialo, and deglycosylated (apo) ovine submaxillary gland mucin (OSM). OSM is a large glycoprotein that is extensively O-glycosylated by the disaccharide α-NeuNAc(2–6)α-GalNAc-O-Ser/Thr. Measurements of root mean square radius of gyration, $\langle R_g^2 \rangle^{1/2}$, and hydrodynamic radius, R_h, for OSM and its derivatives were carried out as a function of molecular weight by using static and dynamic light-scattering techniques. The results were fit to the wormlike chain model for describing the dimensions of extended polymer chains. By use of this model, values of h, the length per amino acid residue, and q, the persistence length, which is a measure of chain stiffness, were obtained. These values were then used to assess the conformation and degree of chain extension of intact OSM and its partially and totally deglycosylated derivatives. Native and asialo mucin are found to be highly extended random coils, with asialo mucin having a somewhat less extended structure than intact mucin. Upon the complete removal of the carbohydrate side chains, the extended structure characteristic of intact and asialo mucin collapses to chain dimensions typical of denatured globular proteins. Conformational analyses based on the rotational isomeric state model were also performed by using the probability maps of *N*-acetyl-*O*-(GalNAc)-Thr-*N*-methylamide as starting conformations for native and asialo mucin. The results suggest that both the glycosylated and nonglycosylated residues in native mucin may occupy a small region of conformational space having $-90° < \phi < -60°$ and $60° < \psi < 180°$, while a slightly broader range is found to fit asialo mucin. The proposed conformations obtained for these mucins are consistent with their circular dichroism spectra. Significantly larger ranges of ϕ and ψ values were obtained for apo mucin, as would be expected from its circular dichroism spectra and increased flexibility. These results indicate the expanded mucin structure is the direct result of peptide core glycosylation. These observations together with the results of earlier studies indicate that steric interactions of the O-linked GalNAc residue with the peptide core are primarily responsible for the expanded mucin structure and that these perturbations extend to the nonglycosylated amino acid residues. This expanded mucin conformation must be a significant determinant of the viscoelastic properties of these molecules in solution.

The functions of the carbohydrate moieties in glycoproteins are poorly understood. One class of glycoproteins whose physical properties are likely to be largely determined by their carbohydrate side chains is the mucous glycoproteins (mucins). To date, however, little information describing the effects of glycosylation on mucin properties is available (Carlstedt et al., 1985; Levine et al., 1987; Neutra & Forstner, 1987). To address this gap in our knowledge, we have used light-scattering and NMR spectroscopic methods (Shogren et al., 1985, 1986, 1987; Gerken & Dearborn, 1984; Gerken, 1986; Gerken & Jentoft, 1987) to study the chain dimensions, conformation, and dynamics of a series of native and modified mucins. In this and the following paper (Gerken et al., 1989) we report detailed light-scattering and NMR studies comparing the properties of an extensively deglycosylated mucin with its mono- and disaccharide side chain containing analogues.

Mucins are a unique class of high molecular weight (>10^6), extensively O-glycosylated glycoproteins that are primarily responsible for the viscoelastic properties of mucous secretions (Meyer & Silberberg, 1978; Bell et al., 1982). We have concentrated on ovine submaxillary gland mucin (OSM)[1] for these studies because the O-linked oligosaccharides of this mucin consist almost exclusively of the disaccharide α-NeuNAc(2–6)α-GalNAc-O-Ser/Thr, a relatively simple and easily modified structure common to most mucins (Gottshalk & Bhargava, 1972). The OSM peptide core is typical of salivary mucous glycoproteins, being composed primarily of glycine (19%), serine (18%), threonine (14%), alanine (14%), and proline (11%). Approximately 75–90% of the serine and threonine residues contain the disaccharide (Gerken & Dearborn, 1984), so about one-fourth of the amino acid residues in OSM are glycosylated. The homogeneous structure of the OSM side chains allows the sequential removal of the NeuNAc and GalNAc residues to give asialo OSM containing single sugar side chains and the fully deglycosylated apo OSM. Since products of known structure can be generated, OSM is ideal for studying the role of carbohydrate on the solution structure of mucin-type glycoproteins.

† This work was supported by grants from the Rainbow Chapter of the Cystic Fibrosis Foundation and the United Way of Cleveland, by Research Grants from the Cystic Fibrosis Foundation, and by NIH grants DK 27651, HL 32227, and DK 39918.

* Address correspondence to this author.

‡ Fellow of the Cystic Fibrosis in Pulmonary Disease training grant, NHLBI-HL 07415. Present address: USDA Northern Regional Research Center, Peoria, IL 61604.

[1] Abbreviations: GalNAc, *N*-acetylgalactosamine; NeuNAc, *N*-acetylneuraminic acid (sialic acid); M_w, weight-average molecular weight; $\langle R_g^2 \rangle^{1/2}$, root mean square radius of gyration; R_h, hydrodynamic radius; $\langle R_g^2 \rangle_z^{1/2}$, z-average root mean square radius of gyration; $\bar{R}_h$, inverse of z-average of $1/R_h$; A_2, second virial coefficient; ϕ, peptide C′, N, C_α, C′ dihedral angle; ψ, peptide N, C_α, C′, N dihedral angle; h, length per peptide residue; q, persistence length; C_∞, characteristic ratio; $\langle R^2 \rangle^{1/2}$, root mean square end to end distance; N, number of peptide residues; Gdn-HCl, guanidine hydrochloride.

0006-2960/89/0428-5525$01.50/0

Earlier studies using light-scattering (Shogren et al., 1986, 1987; Carlstedt et al., 1983), rheological (Sheehan & Carlstedt, 1984a), hydrodynamic (Sheehan & Carlstedt, 1984a; Slayter & Codington, 1973; Creeth & Knight, 1967), and NMR approaches to mucin solution structure (Gerken & Dearborn, 1984; Gerken, 1986; Gerken & Jentoft, 1987) indicate that mucins, including OSM, behave as *expanded* random coils in solution. The dimensions of mucins as measured by the root mean square radius of gyration, $\langle R_g^2 \rangle^{1/2}$ (or hydrodynamic radius, R_h), are 2.5–3-fold larger than those of nonglycosylated denatured globular proteins having similar protein core molecular weights (Shogren et al., 1986). Of interest are the recent findings from our light-scattering studies of size-fractionated mucins showing that the chain dimensions of mucins, although dependent on the length of the mucin peptide core, are unaffected by increasing the average oligosaccharide side chain length beyond the first 2 residues [Shogren et al., 1986, 1987; see also Carlstedt and Sheehan (1984)]. These conclusions are consistent with carbon-13 NMR relaxation studies of native OSM and porcine submaxillary mucin (PSM) since these mucins display similar peptide mobilities despite the fact their average side-chain sizes vary from 2 to 3.5 residues [Gerken & Jentoft, 1987; see Gerken et al. (1989)]. Thus, it appears that carbohydrate side chains of 2 or more residues, regardless of structure, have essentially the same effect on the mucin peptide core. The studies of native and sequentially deglycosylated OSM reported here were designed to examine the role of sialic acid (NeuNAc) and the linkage GalNAc on the peptide conformation and chain dimensions of mucins. The results clearly show that glycosylation has profound effects on the conformation of the peptide core and that most of these effects are due to steric interactions between the peptide backbone and the GalNAc residues linked to Ser and Thr.

Materials and Methods

Materials. Frozen ovine submaxillary glands were obtained from Pelfreeze, while neuraminidase type X and trypsin type XI were from Sigma Chemical Co. Sephacryl S-200 and S-1000 were from Pharmacia, while Chelex 100 cation-exchange resin was from Bio-Rad Laboratories. CM-52 cellulose was from Whatman. All other chemicals and materials were from commercial sources and of the highest purity available.

Analytical Methods. Sialic acid was measured by the resorcinol methods of Svennerholm (1958) and *N*-acetylgalactosamine by the method of Morgan and Elson as modified by Reissig et al. (1955); hexose was also monitored by the periodic acid–Shiff assay of Mantle and Allen (1978). Amino acid analyses were performed by reverse-phase HPLC chromatography of the phenylthiocarbamyl (PTC) derivatives of the amino acids by using a modification of the methods of Bidlingmeyer et al. (1984) after glycoprotein hydrolysis (<0.1 mg/mL) in vacuo in 6 M HCl at 110 °C for 24 h.

Mucin Isolation and Modification. OSM was isolated as described previously (Gerken & Dearborn, 1984) omitting hydroxyapatite chromatography except that protease inhibitors (Chelex 100 and phenylmethanesulfonyl fluoride) were added in the initial stages of purification. Enzymatically desialized asialo OSM was obtained by treating OSM with neuraminidase (Gerken & Dearborn, 1984). Asialo OSM was also prepared by acid hydrolysis by incubating solutions of OSM (5 mg/mL OSM in 1.0 M NaCl) at pH 2 and 80 °C for 1.5 h (Carlson, 1969). The removal of sialic acid from both the neuraminidase-treated and acid-hydrolyzed asialo OSM preparations was confirmed by the resorcinol assay (90–95% and 96–98% NeuNAc removed, respectively) and by carbon-13 NMR spectroscopy (Gerken & Dearborn, 1984).

Deglycosylated (apo) OSM was prepared by treating acid-hydrolyzed asialo OSM with trifluoromethanesulfonic acid by using the methods of Edge et al. (1981) as previously applied to OSM (Gerken, 1986). To maximize the removal of GalNAc, reactions proceeded for 7 h at room temperature prior to workup. After dialysis against H_2O, a light yellow precipitate formed and was removed by centrifugation or filtration (0.45-μm filter). About 87% of the GalNAc was removed as determined by HPLC analysis of the benzoylated methyl glycosides (Jentoft, 1985). This was in agreement with the extent of deglycosylation determined by ^{13}C NMR (Gerken et al., 1989). No significant difference in the amino acid composition was found between this material and native or asialo OSM. Deglycosylated OSM was also prepared by incubating enzymatically prepared asialo OSM with α-*N*-acetylgalactosaminidase (Genzyme), giving an apo mucin with a molecular weight of approximately 5×10^5 on the basis of Sephacryl S-500 gel filtration (data not shown). However, quantities of the enzymatically prepared apo mucin were too small for study by light scattering or NMR.

To obtain glycosylated mucins of similar molecular weight as the chemically prepared apo mucin ($\sim 15 \times 10^3$), both native and acid-hydrolyzed asialo OSM were treated with trypsin, producing Trp OSM and Trp asialo OSM. These were prepared by treating mucin (10 mg/mL in 0.05 M Tris buffer, pH 8.5) with trypsin (0.024 mg total/mg of mucin) for 8 h at room temperature with enzyme added at 0- and 4-h time points. The solutions were dialyzed, filtered through 0.45-μm filters, made 10 mM sodium phosphate, pH 7, and passed through CM-cellulose to remove trypsin. Trp OSM and Trp asialo OSM were then dialyzed exhaustively and lyophilized.

Gel Filtration. Prior to light-scattering studies, native and modified mucins were chromatographed by gel filtration on either Sephacryl S-1000 or S-200 depending on the molecular weight of the sample (10^5–10^7 or 10^3–10^5, respectively). This permitted us to obtain fractions with different molecular weights so that $\langle R_g^2 \rangle_z^{1/2}$ and $\bar{R}_h$ could be measured as functions of M_w. Native OSM was fractionated into four fractions on S-1000 in 5 M guanidine hydrochloride (Gdn-HCl) as described earlier (Shogren et al., 1987). The enzymatically and hydrolytically prepared asialo mucins were also fractionated on S-1000 in 5 M Gdn-HCl as detailed in Figure 1.

Fractions of asialo OSM subjected to rechromatography on S-1000 gave unimodal, nearly Gaussian distributions (data not shown). Values of $M_w/M_n = 3.4 \pm 0.7$ are calculated on the basis of $\mu_2/\bar{\Gamma}^2 = 0.5 \pm 0.1$ (Shogren et al., 1986). Values of the distribution coefficient, K_D $[(V_{peak} - V_0)/(V_t - V_0)]$, were found to fit the universal model of Porath (1963), $K_D^{1/3} = 1 - BR_h$ by using a value of $B = 2200$ Å^{-1} determined previously (Shogren et al., 1987). As shown in Table I, values of $\bar{R}_h$ for asialo OSM based on S-1000 chromatography are in agreement with values of $\bar{R}_h$ determined by light scattering. Fractions of trypsin-treated native and asialo OSM and apo OSM also gave single, Gaussian peaks on S-200 chromatography (data not shown). Values of $M_w/M_n = 1.6 \pm 0.3$ are calculated on the basis of based on $\mu_2/\bar{\Gamma}^2 = 0.15 \pm 0.05$ (Shogren et al., 1986). Monodisperse, globular proteins having known values of R_h were used for the S-200 calibration,[2] giving a value of 85 Å^{-1} for B. Since the $\bar{R}_h$ values obtained from

[2] To facilitate comparison with the light-scattering data, $\bar{R}_h$ was estimated from the S-200 data by using $\bar{R}_h = R_h Y (M_w/M_n)^{0.25}$, where Y and M_w/M_n are calculated as described previously from $\mu_2/\bar{\Gamma}^2$ (Shogren et al., 1986). This correction is necessary to account for the z-average weighting of the light-scattering data of polydisperse materials.

gel filtration and light scattering are nearly identical (Table I), no aggregation of these samples is apparent at the high (10–40 mg/mL) concentrations used for light scattering.

Light Scattering. Solutions of asialo OSM were prepared by dissolving weighed proportions of fractionated, lyophilized mucin in 5 M Gdn-HCl and 10 mM phosphate buffer, pH 7, by slow rocking overnight at 3 °C. Mucin concentrations were 1–6 mg/mL and were corrected for their residual 5% w/w water content present after lyophilization (Shogren et al., 1986). Solutions of enzymatically and hydrolytically prepared asialo OSM were filtered (5- and 0.45-μm Millipore MF filters, respectively) directly into light-scattering cells. No detectable amounts of mucin were lost on filtration, as estimated by determining hexose content of the filtered and unfiltered solutions. The solutions were diluted with filtered buffer and centrifuged for 1 h at 5000 rpm immediately prior to light-scattering analysis. Solutions of fractionated apo mucin and trypsin-treated mucins in 0.5 M NaCl and 5 mM sodium cacodylate, pH 7.5, were concentrated to 10–40 mg/mL by ultrafiltration using Amicon YM-2 membranes and then filtered by using 0.22-μm Millipore filters. No solute was lost on the filter as estimated by UV absorbance at 220 nm of the filtered and unfiltered solutions. It was necessary to concentrate apo mucin by ultrafiltration because lyophilized apo OSM often proves difficult to redissolve. A solution of apo OSM (fraction 2, Figure 2C) in 6 M Gdn-HCl was concentrated as described above and filtered through a 0.1-μm filter. A large amount of apo mucin was lost on the filter, indicating extensive aggregation had occurred. The soluble fraction gave a M_w of 8000 (Table I), which is about half of the M_w of the original fraction in 0.5 M NaCl. Because of this, 0.5 M NaCl was used as the solvent for the light-scattering studies of all of the low molecular weight mucins.

Light-scattering measurements were performed by using laser light of wavelength λ = 6328 Å and a BI240 photogoniometer with BI2020 autocorrelator (Brookhaven Instruments Corp.) Values of weight-average molecular weight, M_w, z-average radius of gyration, $\langle R_g^2\rangle_z^{1/2}$, and the second osmotic virial coefficient, A_2, were determined from Zimm plots of average intensities of scattered light at scattering angles, θ, of 20–90°. The refractive index increments at constant solvent chemical potential, $(dn/dc)_\mu$, were determined for the following solutions after extensive dialysis: OSM in 5 M Gdn-HCl, 0.125 mL/g; OSM and trypsin-treated OSM in 0.5 M NaCl, 0.155 mL/g; asialo OSM in 5 M Gdn-HCl, 0.115 mL/g; asialo OSM and trypsin-treated asialo OSM in 0.5 M NaCl, 0.155 mL/g; apo OSM in 0.5 M NaCl, 0.16 mL/g; apo OSM in 5 M Gdn-HCl, 0.105 mL/g. The intensity autocorrelation function of the scattered light was analyzed by using the method of cumulants (Kopple, 1972) to derive the first moment, $\bar{\Gamma}$, and the second moment, μ_2. $\bar{\Gamma}$ was measured as a function of mucin concentration, c, and scattering angle over the same range as described above. The z-average translational diffusion constant, Dt_z^0, was determined by plotting $\bar{\Gamma}/Q^2$ against Q^2 and c and extrapolating to $Q^2 = 0$ and $c = 0$, where $Q = 4\pi n/\lambda \sin(\theta/2)$ and n is the refractive index of the solvent. The hydrodynamic radius, $\bar{R}_h = 1/\langle R_h^{-1}\rangle_z$, was derived from Dt_z^0 by using the Stokes–Einstein equation, $\bar{R}_h = kT/(6\pi\eta_0 Dt_z^0)$, where η_0 is the viscosity of the solvent. The number of amino acid residues per glycopeptide chain, N, was calculated by using the glycopeptide molecular weight, percent protein, and an average molecular weight per amino acid residue of 90. Values used for the percent protein were 37% for OSM (Hill et al., 1977a), 54% for asialo OSM, and 90% for apo OSM.

Prior to plotting in Figure 3, values of $\langle R_g^2\rangle_z^{1/2}$ and $\bar{R}_h$ were roughly corrected for polydispersity from the values of $\mu_2/\bar{\Gamma}^2$ as described previously (Shogren et al., 1986). The corrections are not very large; for example, for $\mu_2/\bar{\Gamma}^2 = 0.5 \pm 0.1$, $\langle R_g^2\rangle^{1/2} = (0.79 \pm 0.02)\langle R_g^2\rangle_z^{1/2}$ and $R_h = (0.92 \pm 0.01)\bar{R}_h$, while for $\mu_2/\bar{\Gamma}^2 = 0.15 \pm 0.05$, $R_h = (0.98 \pm 0.01)\bar{R}_h$. The cited errors in the corrections reflect uncertainties in $\mu_2/\bar{\Gamma}^2$. Values of $\langle R_g^2\rangle_z^{1/2}$ could not be determined for the low molecular weight trypsin-treated mucins and apo mucin by light scattering since $Q\langle R_g^2\rangle^{1/2} \ll 1$ for these samples. Values of $\langle R_g^2\rangle^{1/2}$ for these preparations were calculated by using the wormlike chain model and experimental data as described below and are given in parentheses in Table I.

Excluded volume effects for mucins appear to be small and on the order of the experimental error. For example, the expansion factor, $\alpha_s = \langle R_g^2\rangle^{1/2}/\langle R_g^2\rangle_0^{1/2}$ (where the subscript 0 represents $\langle R_g^2\rangle^{1/2}$ in the absence of excluded volume effects) is estimated from Figure 3 of Shogren et al. (1986) to be 1.05 ± 0.1 for PSM in 6 M Gdn-HCl with a M_w of 10^6. Further evidence that α_s is nearly 1 for PSM and most other mucins, including OSM, is found in the plots of log $\langle R_g^2\rangle^{1/2}$ vs log N (Shogren et al., 1986, 1987). In these plots $\langle R_g^2\rangle^{1/2}$ values measured for PSM in 6 M Gdn-HCl ($A_2 \simeq 2 \times 10^{-4}$) and 0.1 M NaCl ($A_2 \simeq 0$) are indistinguishable and fall on a line that fits the $\langle R_g^2\rangle^{1/2}$ data obtained for other mucins. Such agreement strongly indicates that the excluded-volume effects are small for all mucins studied to date.[3] Excluded-volume corrections were not necessary for apo mucin since $A_2 \simeq 0$ (see Table I). Thus, the $\langle R_g^2\rangle^{1/2}$ values plotted in Figure 3, corrected only for polydispersity, were used in the studies described below.

Wormlike Chain Calculations. The wormlike chain model of Kratky and Porod (1949) is characterized by three parameters: the pitch, h, the persistence length, q, and the diameter, d. Equation 17 of Benoit and Doty (1953) was used to calculate $\langle R_g^2\rangle$, where L is the contour length, Nh. R_h was obtained from the Stokes equation, $R_h = f(6\pi\eta_0)^{-1}$, where the frictional coefficient, f, was calculated from eq 49 and 51 of Yamakawa and Fujii (1973) for $L > \sigma$ and $L < \sigma$, respectively, where L is the reduced contour length, $L/2q$. In these equations of Yamakawa and Fujii, σ = 2.278 and the coefficients A_i and C_i are given by eq 50 and 52, where d is the reduced chain diameter, $d/2q$.

Since the data are insufficient to estimate h, q, and d uniquely, the hydrodynamic diameter, d, was first estimated theoretically by using the porous sphere approach (Bloomfield,

[3] It was pointed out in the review process that the positive values of A_2 obtained for native and asialo OSM and PSM (Table I; Shogren et al., 1986, 1987) appear to be inconsistent with values of the excluded volume expansion factor of $\simeq$1. We do not fully understand these discrepancies except to note that mucins, being highly extended molecules, would be expected to show the onset of significant excluded-volume effects at higher chain lengths when compared to more flexible random coiled peptides. Our findings are not inconsistent with those of Huber et al. (1985). These workers report for polystyrene ($M_r \leq 10^4$) in toluene (a good solvent) that corrected values of α_s of $\sim\leq$1.1 and values of A_2 on the order of 10^{-4}–10^{-3} cm^3 mol/g^2 are consistent with theory. In toluene, polystyrene molecules of $M_r \sim 10^4$ have contour lengths equal to about 18 persistence lengths. For mucins this number of persistence lengths is equivalent to $\simeq$1100 peptide residues [18 × 145 Å/(2.5 Å/residue)], which corresponds to a mucin peptide molecular weight of approximately 10^5 and native mucin molecular weights on the order of 2–3 × 10^5 for OSM and PSM. This comparison suggests that it is reasonable to find small values of α_s with positive values of A_2 for the mucins under study. Finally, we note that the theoretical basis for describing the effects of chain stiffness on A_2 and the relationship between A_2 and α_s for stiff chains is not yet fully understood (Huber & Stockmayer, 1987).

1983). The OSM chain was modeled as a concentric cylinder having a dense peptide core of radius $r_1 \simeq$ 3-Å thickness and a less dense outer layer containing the disaccharides of $r_2 \simeq$ 8-Å thickness. By use of eq 2 of Bloomfield the outer layer sugar density was estimated to be $\simeq 8 \times 10^{-4}$ Å^{-3} for a side-chain length of 2 and fraction of glycosylated peptide residues of 0.3. The inverse hydrodynamic shielding length, κ ($\simeq$0.2 Å^{-1}) was calculated by using eq 3 of Bloomfield (1983) using 2.5 Å as an estimate of the Stokes radius of a sugar residue. From Figure 3 of Bloomfield (1983), a plot of $d/2(r_1 + r_2)$ versus $\kappa(r_1 + r_2)$, one obtains $d/2(r_1 + r_2)$ of $\sim$0.45 so that the effective hydrodynamic diameter, d, for OSM is $\simeq 10 \pm 4$ Å. The principal uncertainty in this calculation resides in r_2 and depends on the orientation of the disaccharide side chain with respect to the protein core. Fortunately, R_h and $\langle R_g^2 \rangle^{1/2}$ are rather insensitive to changes in d, so the errors involved in the calculation of q and h will be smaller than for d. Next, h was calculated from d and the R_h data for low molecular weight trypsin-treated mucins. This is justified since these samples have a nearly rodlike configuration (see Results and Discussion), and hence R_h is essentially independent of q. Finally, q was calculated by using the high molecular weight $\langle R_g^2 \rangle^{1/2}$ data that are dependent on q and h. For apo mucin, apparent values of h were obtained from the rotational isomeric state model (see below). This was necessary because the apo mucin fractions were of molecular sizes that were significantly larger than the rodlike behavior limit of approximately 10 Å based on the apo mucin persistence length.

Rotational Isomeric State (RIS) Calculations. The RIS method Flory (1969) was used to calculate the mean square unperturbed radius of gyration $\langle R_g^2 \rangle$ and the mean square unperturbed end to end distance, $\langle R^2 \rangle$ for mucin and apo mucin polypeptide chains. When the conformational energies of the ith peptide residue, $E(\phi_i,\psi_i)$, are assumed to be independent of $\phi_{j\neq i}$ and $\psi_{j\neq i}$, $\langle R_g^2 \rangle$ and $\langle R^2 \rangle$ are given by matrix eq 45 and 41 in Chapter 1 of Flory (1969). For apo OSM we have used the average coordinate transformation matrices, $\langle T \rangle$, of Brant et al. (1967) for Ala and Gly residues and those of Schimmel and Flory (1968) for Pro. The restricted conformation for the residue preceding a Pro residue was represented by the matrix, $\langle T \rangle_{ala(pro)}$, for Ala preceding Pro as described by Schimmel and Flory (1968). The Ala matrix of Brant has been shown to give values of $\langle R_g^2 \rangle$ in agreement with experimental values determined for four synthetic homopolypeptides with side chains larger than Ala (Brant & Flory, 1965). Therefore, the Ala matrix was used to represent amino acid residues in apo OSM having side chains of one or more carbon atoms, with the exception of Pro. To simplify the calculations, a random amino acid sequence model, composed of 59% Ala, 19% Gly, 11% Pro, and 11% Ala preceding Pro, was used to calculate an average transformation matrix, $\langle \bar{T} \rangle$, as follows: $\langle \bar{T} \rangle = 0.59\langle T \rangle_{ala} + 0.19\langle T \rangle_{gly} + 0.11\langle T \rangle_{pro} + 0.11\langle T \rangle_{ala(pro)}$. The effects of amino acid sequence on $\langle R_g^2 \rangle$ have been shown to be small for many proteins (Miller & Goebel, 1968). As a test of the Ala matrix approximation, the available transformation matrices for the amino acids in apo mucin modeled by Ala [i.e., Ser, Thr, Glx, and Val (Miller & Goebel, 1968)] were included in the calculations, giving essentially the same results as obtained by using the Ala matrix.

The average ϕ,ψ probability map (expressed as percentage) for apo OSM (Figure 4A) was obtained by averaging the probability maps for each amino acid using the composition described above: $\bar{P}(\phi,\psi) = 0.59P_{ala}(\phi,\psi) + 0.19P_{gly}(\phi,\psi) + 0.11P_{pro}(\phi,\psi) + 0.11P_{ala(pro)}(\phi,\psi)$, where $P(\phi,\psi)$ values were calculated from the energy maps in the above references by using a Boltzman distribution at 293 K: $P(\phi,\psi) = \exp[-E(\phi,\psi)/RT]/[\sum_\phi \sum_\psi \exp(-E(\phi,\psi)/RT)]$. Conformational energies, $E(\phi,\psi)$, were read from the above published surfaces at 30° intervals on ϕ and ψ. The values of C_∞ based on these $E(\phi,\psi)$ readings accurately (±5%) reproduced the original values of C_∞.

The parameters h, q, and C_∞ were calculated from the RIS model to facilitate comparison with the wormlike chain model. The characteristic ratio, C_∞, was obtained by the RIS model by using eq 44 in Chapter 1 Flory (1969), which is equivalent to $\langle R^2 \rangle / Nl^2$ in the limit of large N, where l is the distance between peptide α-carbons, 3.8 Å (note that as N approaches infinity $\langle R^2 \rangle = 6\langle R_g^2 \rangle$). The persistence length, q, is given by $q = C_\infty l^2/2h$. For OSM and asialo OSM the pitch, h, was obtained by extrapolating to $N = 1$ the linear plots of $\langle R^2 \rangle^{1/2}/N$ versus N (for values of N in the range 4–20). For apo mucin the plot of $\langle R^2 \rangle^{1/2}/N$ vs N is highly curved, and as N approaches 1 an intercept of $h = 3.8$ Å is obtained. This behavior reflects the flexibility of the apo OSM peptide core (in contrast to the rodlike nature of the short mucin chains with similar N). The apparent value of h obtained for apo OSM in this manner represents the distance between α-carbons, 3.8 Å, as expected for random coiled polypeptides.

For native and asialo OSM the probability maps for *N*-acetyl-*O*-(GalNAc)-Thr-*N'*-methylamide (Duben & Bush, 1983) were used to obtain the apparent average transformation matrix, $\langle \bar{T} \rangle_{mucin}$. $\langle \bar{T} \rangle_{mucin}$ was calculated by using probabilities taken from Table VI and Figure 5 of Duben and Bush (1983) at 10° intervals of ϕ,ψ as follows: $\langle \bar{T} \rangle_{mucin} = \sum_\phi \sum_\psi P(\phi,\psi) T(\phi,\psi) / \sum_\phi \sum_\psi P(\phi,\psi)$, where $T(\phi,\psi)$ is the coordinate transform matrix. This average matrix $\langle \bar{T} \rangle_{mucin}$ was used to represent both glycosylated and nonglycosylated residues since no energy data are available for residues adjacent to glycosylated residues. Hence, this model is for a homopolymer. $\langle R^2 \rangle$, $\langle R_g^2 \rangle$, and C_∞ were obtained from $\langle \bar{T} \rangle_{mucin}$ as describe above. By use of the complete probability map of Duben and Bush for $\langle \bar{T} \rangle_{mucin}$, a $C_\infty = 3.1$ was obtained. This is much smaller than the experimental value of $C_\infty = 21$ for asialo OSM. To bring the calculated and experimental values of C_∞ into better agreement, the region of high probability centered at $\phi = -75°$ and $\psi = -55°$, representing the right-handed α-helix, was discarded. This is justified since the values of h calculated from this region, $\sim$1.2 Å, are inconsistent with the experimental value of 2.5 ± 0.5 Å. After this adjustment, C_∞ 11.7, in better agreement with experiment. Next, probabilities in the region of $-180° < \phi < -120°$ were decreased from about 10% to 5% of the total probability to give $C_\infty = 21$. This change was favored since much larger changes in the region $-120° < \phi < -60°$ would have been necessary to give a C_∞ in agreement with experiment. To obtain a similar map for native OSM, it was necessary to eliminate outlying probabilities altogether ($-180° < \phi < -100°$) and to further narrow the major probability regions.[4] In light of the CD data, which suggest a higher content of polyproline II structure in OSM than in asialo OSM, the region near $\phi = -80°$ and $\psi = -150°$ was favored for OSM.

Circular Dichroism Studies. Solutions of Trp OSM, Trp asialo OSM, and apo OSM in 0.10 M NaF and 5 mM PO_4, pH 7, each $\sim$50 μg/mL in protein in 0.1-cm cells, were used

[4] The average transformation matrices for asialo and native OSM obtained from these distributions (shown in Figure 4B,C) are respectively

0.27	−0.16	0.80	0.48	−0.60	0.62
−0.48	−0.64	−0.04	0.53	−0.29	−0.73
0.75	−0.45	−0.20	0.63	0.67	0.19

Table I: Light-Scattering Parameters for Modified OSM

sample[a]	solvent	M_w (×10⁻⁶)	$\langle R_g^2 \rangle_z^{1/2}$ [b,h]	A_2[c] (×10⁴)	$\bar{R}_h$(L)[b,d]	$\bar{R}_h$(G)[b,e]	$\mu_2/\bar{\Gamma}^2$
E-asialo P1	6 M Gdn-HCl[f]	8.0 ± 10%	2200 ± 10%	2.1 ± 0.5	1010 ± 5%	1010 ± 15%	0.5 ± 0.1
E-asialo P2	6 M Gdn-HCl[f]	2.9	1130	3.0	650	710	0.5
E-asialo P3	6 M Gdn-HCl[f]	1.6	860	3.0	480	530	0.5
H⁺-asialo P1	6 M Gdn-HCl[f]	1.2	570	1.3	420	360	0.5
H⁺-asialo P2	6 M Gdn-HCl[f]	0.45	520	2.6	260	270	0.5
H⁺-asialo P3	6 M Gdn-HCl[f]	0.27	370	4.9	160	230	0.5
apo P1	0.5 M NaCl[g]	0.037	(100)	−3.0	58	53	0.5
apo P2	0.5 M NaCl[g]	0.014	(57)	0	33	36	0.5
apo P3	0.5 M NaCl[g]	0.0084	(43)	6.0	25	21	0.5
apo P1 + P2 + P3	0.5 M NaCl[g]	0.017	(60)	0	35	36	0.5
apo P2	6 M Gdn-HCl[g]	0.0062	(32)	−24	19	–	0.5
Trp P2	0.5 M NaCl[g]	0.0160	(42)	12	23	29	0.15 ± 0.05
Trp P3	0.5 M NaCl[g]	0.0094	(30)	11	19	21	0.20
Trp asialo P2	0.5 M NaCl[g]	0.0120	(40)	8.5	23	29	0.20
Trp asialo P3	0.5 M NaCl[g]	0.0078	(30)	9.8	19	21	0.15

[a] Abbreviations: E-asialo, enzymatically prepared asialo OSM; H⁺-asialo, hydrolytically prepared asialo OSM; apo, apo OSM; Trp, trypsin-treated OSM; Trp asialo, trypsin-treated asialo OSM. Fractions (P) indicated are those pooled as shown in Figures 1 and 2. [b] Values in angstroms. [c] Values in cm³ mol/g². [d] Values obtained from light scattering. [e] Values obtained from gel filbration on S-1000 or S-200, see text. [f] Fractions obtained from S-1000 chromatography in 5 M Gdn-HCl with 10 mM Na_2HPO_4, pH 7.0. [g] Fractions obtained from S-200 chromatography in 0.5 M NaCl with 5 mM sodium cacodylate, pH 7.5. [h] Values in parentheses are calculated from $\bar{R}_h$ values by using the wormlike chain model (see text).

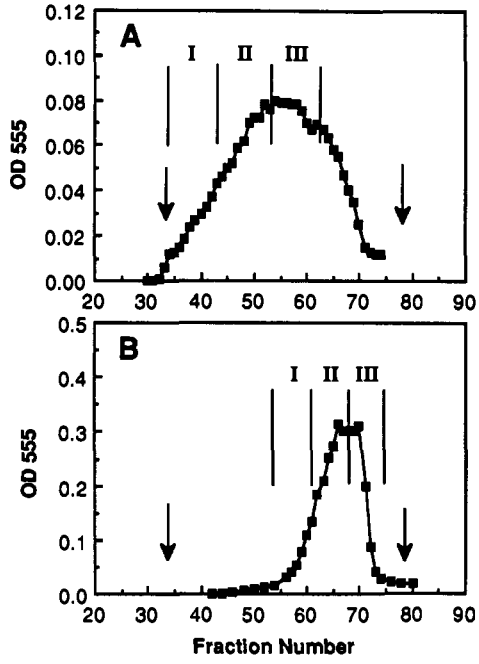

FIGURE 1: Gel filtration profiles on Sephacryl S-1000 of asialo OSM prepared by neuraminidase treatment (A) and by acid hydrolysis (B). The indicated fractions were pooled and subsequently studied by light scattering. Mucin was monitored by the carbohydrate periodic acid–Shiff assay at 555 nm (Mantel & Allen, 1978). Chromatographic conditions: 120 × 2.5 cm column, 6.9-mL fraction volume, 5 M Gdn-HCl in 10 mM phosphate buffer, pH 7, elutant. The arrows indicate the excluded and included column volumes, respectively.

for CD studies. A Jasco J-40 spectrometer was used to collect ambient temperature spectra that were calibrated with (+)-10-camphorsulfonic acid by using $[\theta]_{290.5}$ = 7800 deg cm²/dmol and $[\theta]_{192.5}$ = −1560 deg cm²/dmol (Yang et al., 1986). Reported spectra are the differences between the average of two repetitive sample scans and two solvent scans.

Results and Discussion

Mucin Molecular Weight and Solubility. A few comments are in order regarding the effects of mucin modification on mucin molecular weight. The chromatographic tracings in Figures 1 and 2 and the data in Table I show that asialo mucin prepared by acid hydrolysis has a lower M_w than enzymatically prepared asialo OSM. This is probably due to limited peptide cleavage occurring during the acid hydrolysis step. The molecular weights reported earlier for native OSM (Shogren et al., 1987) are also lower (1–6 × 10⁶) than those reported in the present study for the enzymatically prepared asialo OSM. These differences are attributed to the normal variations in molecular weight in different mucin preparations. Tri-

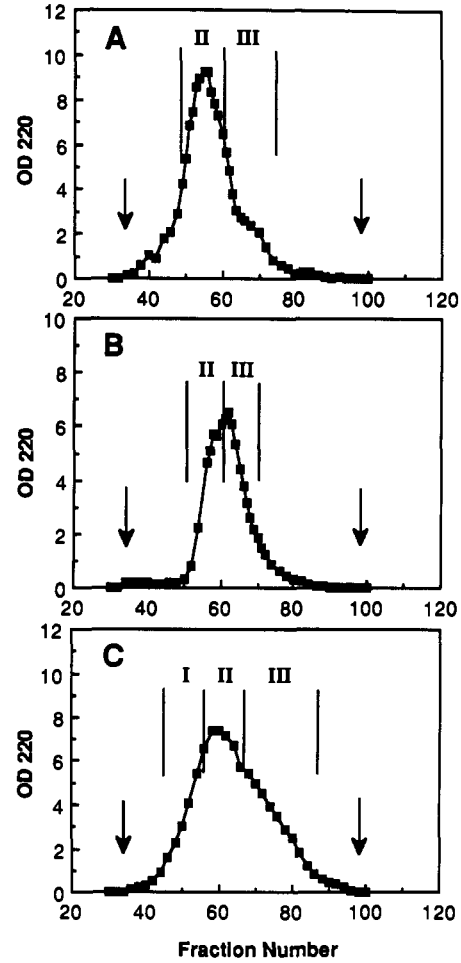

FIGURE 2: Gel filtration profiles of trypsin-treated native (A) and asialo (B) OSM and deglycosylated apo OSM (C) on Sephacryl S-200. The indicated fractions were pooled and subsequently studied by light scattering. Mucin was monitored by amide absorbance at 220 nm. Chromatographic conditions: 40 × 2.7 cm column, 2.2-mL fraction volume, 0.5 M NaCl in 5 mM sodium cacodylate buffer, pH 7.5, elutant. The arrows indicate the excluded and included column volumes, respectively.

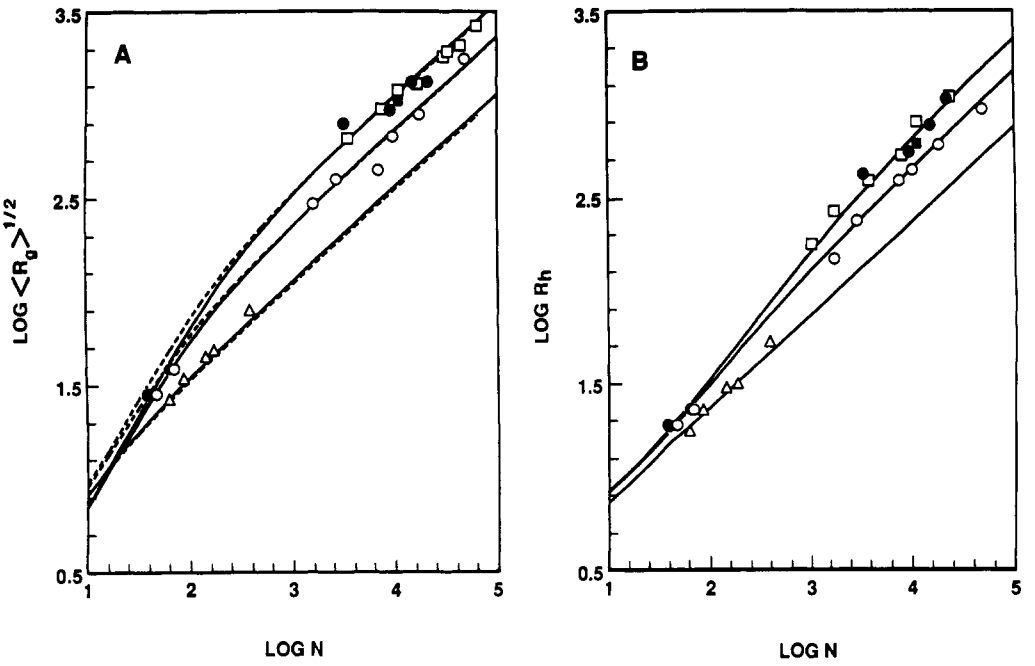

FIGURE 3: Plots of log $\langle R_g^2 \rangle^{1/2}$ vs log N (A) and log R_h vs log N (B) for fractionated native and modified mucins: native OSM (●) (Shogren et al., 1987); asialo OSM (○); apo OSM (△); native (□) and asialo PSM (■) (Shogren et al., 1985, 1986). The plotted values of $\langle R_g^2 \rangle^{1/2}$ and R_h from Table I were corrected for polydispersity as described in the text. Uncertainties in the points are roughly the size of the symbols. The solid curves represent the fit of the wormlike chain model (as described under Materials and Methods) using the parameters given in Table II for native OSM, asialo OSM, and apo OSM. The dashed lines adjacent to the curves in (A) represent calculated results from the rotational isomeric state model using the transformation matrices described in the text, footnote 3, and Figure 4.

fluoromethanesulfonic acid treatment of asialo mucin causes further degradation of the peptide core, giving molecular weights on the order of 10^4 for apo OSM. Since the amino acid compositions of OSM, asialo OSM, and apo OSM are nearly identical (data not shown), the apo mucin, in spite of its low molecular weight, is representative of the native mucin peptide core. Trypsin treatment of OSM and asialo OSM gives glycopeptides of similar size and amino acid composition as the apo mucin. Fractionated native, asialo, and apo OSM are polydisperse, as indicated by the broad chromatographic profiles and values of $\mu_2/\bar{\Gamma}^2 = 0.5 \pm 0.1$. The fractionated trypsin-treated mucins, however, are more nearly monodisperse with values of $\mu_2/\bar{\Gamma}^2 = \sim 0.15 \pm 0.05$.

Table I also demonstrates that values of the second virial coefficient, A_2, are positive for asialo OSM and the trypsin-treated mucins[3] and near zero or even highly negative for apo OSM depending on solvent. Positive values for A_2 are reported for native PSM and OSM (Shogren et al., 1986, 1987) and human cervical mucin (Sheehan & Carlstedt, 1987) in 4–6 M Gdn-HCl, while values of approximately zero are reported for PSM in 0.1 M NaCl (Shogren et al., 1986). The value of the second viral coefficient is thought to reflect solute size and solute–solvent interactions. Positive values indicate good solvent conditions and minimal solute–solute interaction, while negative values of A_2 indicate poor solute–solvent interaction leading to solute self-association. The data in Table I and discussed above indicate that aqueous salt solutions are relatively good solvents for glycosylated mucins but are relatively poor solvents for the deglycosylated mucin protein core. This points to the importance of carbohydrate in maintaining the hydration and solubility of mucins. The poor solubility of apo OSM in aqueous solvents is not unexpected since homopolypeptides containing the amino acids that are dominant in apo OSM, i.e., Gly, Ser, Thr, and Ala, are very poorly soluble in aqueous solvents. Recent studies of apo mucins from canine and human tracheal mucins show similar solubility problems (Woodward et al., 1987).

It should also be noted that values for A_2 for OSM and asialo OSM given in Table I and Shogren et al. (1987) are dependent on the molecular weight of the sample. A plot of log A_2 vs log N, where N is the number of peptide residues, for native and asialo OSM is found to fit a straight line given by $A_2 = 3.2 \times 10^{-3} N^{-0.3}$ (log A_2 vs log N correlation coefficient of 0.86). Huber et al. (1985) and Huber and Stockmayer (1987) have recently described a similar molecular weight dependence for linear synthetic polymers. These workers found for polystyrene in toluene (a good solvent) values of A_2 that are proportional to $M^{-0.2}$ in the random coil limit of $M > 10^5$. For intermediate size polystyrene ($10^4 < M < 10^5$) A_2 scales as $M^{-0.38}$, while for lower molecular weight species, where the molecules are more rodlike, A_2 becomes independent of M. Although we do not observe these changes in slope in our plot of log A_2 vs log M for mucins, probably due to insufficient data at low M, our slope of −0.3 is closest to the intermediate size range, corresponding to a semiflexible chain in a good solvent.

Peptide Chain Length Dependence of R_g and R_h and Wormlike Chain Model of Mucin Peptide Core Structure. The dependence of $\langle R_g^2 \rangle^{1/2}$ and R_h on peptide core chain length for native, asialo, and apo mucin was examined by plotting the logarithms of $\langle R_g^2 \rangle^{1/2}$ and R_h versus the logarithm of the number of peptide core amino acid residues, N. These results, corrected for polydispersity as described under Materials and Methods, are shown in Figure 3A,B. Data points for native OSM and native and asialo PSM are also included for comparison. Solid lines in Figure 3 represent the best fit of the data to the wormlike chain model (see Materials and Methods) using the parameters listed in Table II.

The wormlike chain model can be visualized as a curved thin cylinder whose properties are characterized by three parameters: the pitch, h, the persistence length, q, and the diameter, d. The pitch, h, is the chain length per peptide residue and varies as a function of peptide secondary structure. For example, h is 1.5 Å in an α-helix and 3.6 Å in a β-strand, while for typical random coiled peptides (absence of secondary

Table II: Experimental and Modeled Molecular Parameters for Native and Modified OSM

		wormlike chain model[a]			rotational isomeric state model[a]		
sample	C_∞	h (Å)	q (Å)	d (Å)	C_∞	h (aÅ)	q (Å)
native OSM	50 ± 6	2.5 ± 0.5	145 ± 20	10 ± 4	47	2.9	120
asialo OSM	21 ± 3	2.5 ± 0.5	62 ± 10	10 ± 4	21	2.9	53
apo OSM	5.5 ± 0.7	3.8[b]	10 ± 1	6 ± 2	4.9	3.8	9.3

[a] Models described in text. [b] Value chosen from RIS model as described in text.

structure) h is approximately 3.8 Å, the distance between α-carbons. The persistence length, q, is the average projection of the end to end distance vector onto the first bond of the chain in the limit of infinite chain length (Cantor & Schimmel, 1980). It is, therefore, a measure of the distance that a polymer chain traverses before a significant change in direction occurs and, therefore, reflects chain stiffness. For example, double-stranded DNA molecules, which are quite rigid, have persistence lengths on the order of 500 Å [see Yamakawa and Fujii (1973) and Cantor and Schimmel (1980)], while amylose, which is considerably more flexible, has a value of 28 Å (Yamakawa, 1984). Denatured proteins are even more flexible, displaying values of q on the order of 10 Å.

For values of N greater than $\sim 10^3$ the data points in Figure 3 lie on lines with slopes of approximately 0.5, indicating that both the native and modified mucins have random coil configurations (i.e., $\langle R_g^2 \rangle \propto N$). In this region the limiting value of the characteristic ratio, C_∞, can be obtained directly from the experimental data by using $C_\infty = 6\langle R_g^2 \rangle / Nl^2$ where l is the distance between peptide α-carbons, 3.80 Å. The characteristic ratio is the actual mean square end to end distance divided by the mean square end to end distance calculated for a theoretical freely joined chain (i.e., a chain with complete flexibility about its bonds). Thus, the more extended the chain, the larger the value of C_∞ becomes. Values of C_∞ thus obtained are listed in Table II. For small values of N (<50 amino acid residues), the slope of log $\langle R_g^2 \rangle^{1/2}$ vs log N approaches 1 for OSM and asialo OSM, reflecting a transition to a rodlike configuration as the chain length becomes less than the persistence length. Note that this transition is barely observed for apo OSM since q is small for apo mucin.

In the random coil limit of large N, however, it is apparent that asialo OSM has somewhat smaller values of $\langle R_g^2 \rangle^{1/2}$ and R_h than does OSM, indicating that removal of sialic acid causes a decrease in chain dimensions and a concomitant increase in peptide core flexibility. Apo OSM has much smaller values of R_h than asialo OSM, indicating that the removal of the GalNAc residue results in a much more pronounced increase in flexibility. Similar patterns are observed for changes in the persistence length, q (Table II). The persistence length decreases by a factor of approximately 2 upon removal of sialic acid but decreases by a factor of 6 when GalNAc is removed from asialo OSM. These data indicate that the extended chain dimensions and restrictions in chain flexibility of the OSM peptide core are predominantly due to the peptide-linked GalNAc residue. These findings are confirmed by the ^{13}C NMR studies reported in the following paper (Gerken et al., 1989).

$\langle R_g^2 \rangle^{1/2}$ and R_h values for native PSM (Shogren et al., 1986) also lie on the same curve as native OSM (Figure 3). From these points it is estimated that native PSM has a C_∞ of 50 ± 6, suggesting that PSM has similar chain dimensions and flexibilities as native OSM. A single data point for asialo PSM falls slightly below the line for PSM and OSM, giving a C_∞ of 40 ± 7, indicating that asialo PSM is only slightly more flexible than PSM. The larger C_∞ for asialo PSM relative to asialo OSM suggests that the Gal residue linked to C-3 of the linkage GalNAc as well as other (Fuc, GalNAc) residues must play a similar role in chain extension in PSM as the 2–6-linked sialic acid in OSM. Since many of the oligosaccharide side chains in PSM consist of a single GalNAc residue, substituted GalNAc-O-Ser/Thr oligosaccharides are likely to restrict the mobility of the peptide core over a longer range than an unsubstituted GalNAc residue. Molecular models of the OSM and PSM glycopeptides (Gerken & Jentoft, 1987) indeed show that carbohydrate residues substituted at the C-3 and C-6 positions of the peptide-linked GalNAc residue extend in opposite directions from the GalNAc ring and are both capable of sterically interacting with neighboring peptide core residues. On this basis it is tempting to speculate that the GalNAc residue and its C-3 and C-6 substituents alone determine the extent of the mucin peptide core expansion.

Since large values of $\langle R_g^2 \rangle^{1/2}$ and R_h were obtained in the presence of the strong denaturant 5 M Gdn-HCl, it appears that steric interactions between sugar and peptide residues, rather than H-bond, electrostatic, or hydrophobic interactions, are the dominant factor causing the observed extension and stiffening of the peptide core. Steric interactions must also dominate for mucins in dilute salt solutions as well since, as discussed below and in the following paper (Gerken et al., 1989), the CD and NMR data suggest that both native and asialo OSM have little α or β secondary structure and possess no long-lived amide H-bonds in 0.1 M salt. This is supported by previous studies (Shogren et al., 1986) showing that PSM possesses the same conformation in both 6 M Gdn-HCl and in 0.1 M NaCl and by the observation that values of C_∞ for asialo PSM and PSM are nearly identical (above). Preliminary light-scattering experiments carried out on native and asialo OSM in 0.5 M NaCl also give similar or only slightly smaller values of C_∞ compared to those obtained in 5 M Gdn-HCl (data not shown).[5] These latter results also imply that for native OSM in 0.5 M NaCl the electrostatic interactions between the carboxylic acid groups of the sialic acid residues are sufficiently shielded so as not to significantly alter mucin conformation compared to 5 M Gdn-HCl. The effects of reducing the ionic strength on the charge interactions of the sialic acid residues can be further examined by comparing the Debye electrostatic shielding length with the estimated distance between sialic acid residues. The Debye length is the distance at which solute electrostatic interactions are effectively screened by the presence of salt (Debye & Huckle, 1923). In 0.15 M NaCl the Debye length is ~7 Å (Mikkelsen et al., 1985), a distance less than the estimated minimum separation of sialic acid residues of 8 Å [on the basis of 1 NeuNAc per 3 amino acid residues using h = 2.5 Å/residue (Table II)]. Since the NeuNAc residues would be expected to be even farther separated due to internal rotations of the peptide core, it appears that electrostatic interactions would only affect the

[5] Similar effects, small (15%) increases in measured $\langle R_g^2 \rangle^{1/2}$, have been reported by Sheehan and Carlstedt (1984b) for human cervical mucin when Gdn-HCl concentration was increased from 0.2 to 6 M. These workers have proposed that the effects of guanidine may be due to the unfolding of unglycosylated regions of the peptide core.

conformation of OSM at ionic strengths well below 0.15 M. Thus, the widely accepted concept [see Montreuil (1984)] that electrostatic repulsions between sialic acid and other anionic residues play a major role in the expanded mucin conformation appears to be incorrect under physiologic ionic strengths. Likewise, the notion that sialic acid removal leads to the near complete collapse of mucin structure (Montreuil, 1984) is also clearly incorrect.

Values of $\langle R_g^2 \rangle^{1/2}$ and R_h for apo OSM in 0.5 M NaCl and 5 M Gdn-HCl fit the same theoretical curve (Figure 3) with a C_∞ of 5.5 ± 0.8. Thus, apo OSM has the same random coil configuration in both solvents. For comparison, values of C_∞ for denatured globular proteins in 6 M Gdn-HCl range between 4 and 8 (Tanford et al., 1966; Lapanje & Tanford, 1967; Miller & Goebel, 1968). Using gel filtration, Eckhardt and co-workers (Eckhardt et al., 1987) measured an R_h of approximately 69 Å for apo PSM (in 0.25 M NaCl) and a molecular weight of 96 500 (in 6 M Gdn-HCl). On the assumption that the molecular weights were the same in both solvents, C_∞ for apo PSM can be estimated from their data to be 4.2 ± 2. Thus, as expected from their amino acid compositions, the core proteins from both PSM and OSM seem to have similar random coil conformations.

The values for the persistence length and pitch reported here appear to be in agreement with values previously reported or calculated from the electron micrographic and physical studies of others. Using electron microscopy, Rose and co-workers (Rose et al., 1984) examined the architecture of native, asialo, and apo OSM. At low mucin concentrations, the molecules of OSM appear as long, narrow, parallel rods having average dimensions of 1560 × 12 Å. A value for h of 2.4 Å can be calculated, which is in good agreement with our value of 2.5 Å. The overall parallel rodlike appearance of the mucin is, however, likely to be an artifact of the adherence of mucin to mica since the measured contour lengths are about 10 times the persistence length reported here. Lamblin and co-workers (Lamblin et al., 1979) observed similar rodlike images on electron microscopy of reduced human tracheobronchial mucin (HTBM) from which we have calculated a value of h = 2.6 Å. Stokke and co-workers (Stokke et al., 1987), using electron microscopy, have recently reported a value for q of 150 Å for HTBM, in agreement with our results for native OSM. Taken together, these data support the concept that all native mucins will probably have similar values of h and q and that these are independent of oligosaccharide structure and charge (Shogren et al., 1986). [Note, however, that values of h calculated from other electron micrographic studies on mucins and mucin-like glycoproteins range from 0.5 to 3.6 Å (Sheehan et al., 1986; Slayter & Codington, 1973).] For asialo OSM, Rose and co-workers (Rose et al., 1984) reported somewhat less extended structures compared to native OSM, in qualitative agreement with our findings of a smaller value of q for asialo OSM. Their finding of a very compact, globular form for apo OSM, however, is not totally in keeping with a random coil structure, which would have had a more expanded appearance. The collapsed apo mucin structure observed may represent a drying artifact.

Rotational Isomeric State (RIS) Analysis. The effect of O-glycosylation on the conformation of the mucin peptide core was examined by modeling the mucin chain dimensions by the RIS approach. In the RIS model the actual peptide bond lengths and ϕ,ψ probability (or energy) maps are used with a statistical mechanical approach to calculate the mean square radius of gyration, $\langle R_g^2 \rangle$, and mean square end to end distance, $\langle R^2 \rangle$ [see Flory (1969) and Cantor and Schimmel (1980)].

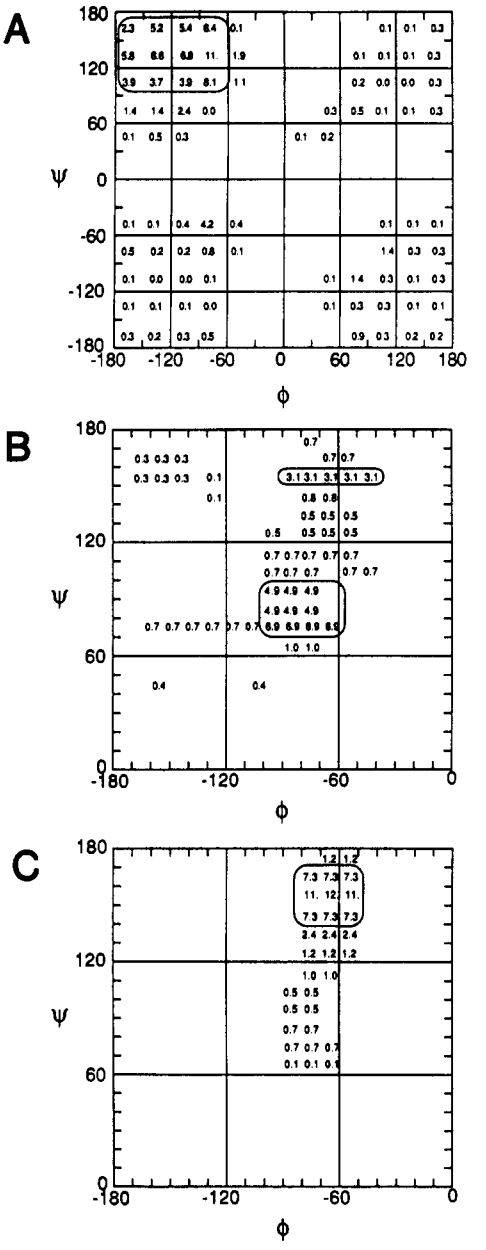

FIGURE 4: Rotational isomeric state ϕ,ψ Boltzman distribution probability maps (expressed in percent) for apo OSM (A), asialo OSM (B), and native OSM (C). These distributions were used to calculate the physical constants given in Table II for the rotational isomeric state model. Regions of high probability are circled (see Materials and Methods for the origins of these distributions). Note that the maps for native and asialo mucin may not be unique, although they are consistent with the experimental data (see Results and Discussion). Also note that only the upper left quadrants are shown for asialo and native OSM. The average transformation matrices for asialo and native OSM obtained from these distributions are given in footnote 3.

Since ϕ,ψ probability maps for O-glycosylated Ser/Thr residues in glycopeptides are unavailable, likely conformational maps were derived from the calculated probability maps of *N*-acetyl-*O*-GalNAc-Thr-*N'*-methylamide (Duben & Bush, 1983) after adjustments were made so as to reproduce the experimental values of C_∞ and h as described under Materials and Methods. For apo OSM, the published amino acid ϕ,ψ maps were used without alteration. Dashed lines in Figure 3A represent fits of the data using the RIS model for the probability maps shown in Figure 4. Molecular parameters derived from the RIS model are listed in Table II for comparison to the wormlike chain model.

For apo OSM, values of C_∞, h, and q obtained from the RIS calculations (Table II) are in very good agreement with the experimentally derived characteristic ratio and the persistence length obtained from the wormlike chain model. This result, together with the experimental finding that d(log $\langle R_g^2\rangle^{1/2}$)/d(log N) = 0.5, as well as the spectroscopic CD and NMR data, further supports the concept that apo OSM has a flexible random coil conformation. This high degree of flexibility is represented in Figure 4 as a very broad distribution of probable ϕ,ψ peptide dihedral angles.

For native and asialo OSM, plots of log $\langle R_g^2\rangle^{1/2}$ vs log N in Figure 3 show that the RIS and wormlike chain models are capable of giving very similar values for the radius of gyration. At low values of N, however, $\langle R_g^2\rangle^{1/2}$ (RIS) values are slightly larger than $\langle R_g^2\rangle^{1/2}$ (WC) values since h (RIS) = 2.9 Å while h (WC) = 2.5 Å. See Table II for a comparison of the RIS and wormlike chain models for native and asialo mucins. The derived probability map for asialo OSM (Figure 4B) shows that likely values of ϕ and ψ are restricted to a narrow range ($-180° < \phi < -60°$, $40° < \psi < 170°$). This is a much narrower distribution than that calculated for apo OSM (Figure 4B) and reflects the greater stiffness of asialo OSM compared to apo OSM. The range of probable values of ϕ and ψ for native OSM, given in Figure 4C ($-100° < \phi < -60°$, $60° < \psi < 170$) is even narrower than for asialo OSM and reflects the greater stiffness of OSM. This range corresponds to a semiflexible homopolypeptide helix with 2–3 residues per turn. Although these ranges of ϕ and ψ are not necessarily unique, the region specified in Figure 4B is the only area that satisfies both the energy calculations of Duben and Bush (1983) and the experimental requirement that $h = 2.5 \pm 0.5$ Å [see Figure 6.4B in Fraser and MacRae (1973)]. This range, furthermore, is consistent with the proposed conformation of the antifreeze glycoprotein (see below). It is important to note that proline, a residue with very restricted conformational space, is well accommodated by the above ϕ and ψ values. It appears that the relatively restricted conformation holds not only for the glycosylated serine and threonine residues but also for the remaining nonglycosylated residues. This can be demonstrated, for example, by replacing the Ser and Thr residues in apo OSM (modeled by Ala) with the newly derived transformation matrix obtained from the modeling of asialo OSM. Such RIS calculations give $C_\infty = 4.9$, which is no larger than originally obtained for apo OSM and 4-fold less than the experimental values obtained for asialo OSM (Table II).

It should be noted that the presence of any residues with a broad distribution of ϕ and ψ will tend to significantly reduce the calculated value of C_∞. Thus, the nonglycosylated residues must have much narrower ϕ and ψ distributions in native and asialo OSM than in apo OSM to account for the large C_∞ values of 20–50 obtained for these mucins. These results indicate that O-glycosylation significantly alters the conformation of both the glycosylated and nonglycosylated residues. Molecular models indeed demonstrate that the bulky GalNAc and NeuNAc residues restrict the adjacent amino acid residues, a finding consistent with carbon-13 NMR relaxation studies described in the following paper (Gerken et al., 1989).

From ^{1}H NMR, energy calculations, and CD studies Bush and Feeney (1986) and Rao and Bush (1987) have concluded that the antifreeze glycoprotein from antarctic fish, which is a repeating glycopolypeptide of Ala-Ala[β-Gal(1–3)α-GalNAc-O]-Thr, assumes an extend polyproline II like conformation ($\phi \simeq -80°$, $\psi \simeq 160°$) with ~3 residues per turn. This conformation is included in our probability map and that of Duben and Bush (1983). Although by the nature of the approach the RIS calculation cannot readily distinguish different conformations accurately (i.e., the solutions are not necessarily unique), our results do suggest that OSM may have a somewhat wider range of acceptable conformations than the antifreeze glycoprotein. Another region of high probability centered near $\phi = -80°$ and $\psi = 70°$ in our plot for asialo OSM, and to a lesser extent for that of native OSM, resembles that of the γ-turn [$\phi = -70$ to -85, $\psi = 60$–70 (Smith & Pease, 1980)], which is stabilized by H-bonds between the C=O of the ith residue and the α-NH of the $(i + 3)$th residue. These results will be discussed in light of the CD spectroscopic studies below.

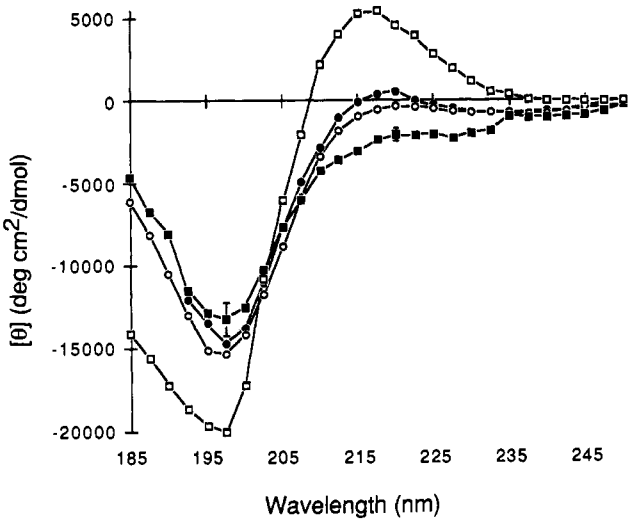

FIGURE 5: Circular dichroism spectra of the peptide core of apo OSM (■), asialo OSM (○), and native OSM (●). The peptide core spectrum of the antifreeze glycoprotein (Bush et al., 1981) is also shown (□). The CD spectra are corrected for the presence of carbohydrate by subtracting the CD spectrum of model oligosaccharides as described in the text. The spectra were obtained by using the conditions described under Materials and Methods. Representative error bars are given at 197 and 227 nm in the apo OSM spectrum, while errors at 185 nm are nearly twice those at 197 nm.

Circular Dichroism Studies. The CD spectra for the peptide core of trypsin-treated native and asialo OSM and apo OSM are shown in Figure 5. The spectra were corrected for the contributions of the carbohydrate residues by subtracting the CD spectra of neuraminyl(2–6)lactose and α-methyl-*N*-acetylgalactosamine (Codutis et al., 1977; Dickinson & Bush, 1975). Corrections were small. For instance, at 200 nm, corrections for NeuNc and GalNAc were 0 and +400 deg cm^2/dmol, respectively, and at 197 nm, −800 and −3000 deg cm^2/dmol, respectively.

The CD spectrum for apo OSM has a negative minimum at 197 nm with $[\theta]$ = −13 000 deg cm^2/dmol and a slight negative shoulder at 227 nm. This closely resembles spectra obtained for various denatured proteins and polypeptides that have a single negative minimum at 200 ± 5 nm and $[\theta]$ = −8000 to −22 000 deg cm^2/dmol (Tiffany & Krimm, 1969). The magnitude and position of this minimum depend on the type of amino acid residue as well as the distribution of ϕ and ψ angles (Ronish & Krimm, 1972). The small negative shoulder in apo OSM may reflect a small amount (<10%) of β-sheet (antiparallel) or β- or γ-turn (Yang et al., 1986). The turns are particularly difficult to estimate since there are structurally many different turns and their CD spectra are variable (Yang et al., 1986). Thus, no specific conformations can be identified on the basis of the negative shoulder at 227 nm. In addition, the high content of Pro, Gly, Ser, and Thr that typically favors coil formation over α-helix formation

makes the presence of α-helical structures unlikely.

Eckhardt et al. (1987) recently reported a very similar CD spectra for apo PSM except that the magnitude of the minimum at 197 nm was larger ($[\theta] = -26\,000$ deg cm^2/dmol). This difference could be due to compositional or sequence differences. These workers have calculated on the basis of the CD that apo PSM contains 42% aperiodic random coil structure, 40% β-turn, 10% antiparallel β-sheet, and 8% α-helix using reference spectra obtained from globular proteins. Since most globular proteins contain little random coil structure, it is very likely that the aperiodic reference CD spectrum used may not truly represent the aperiodic structure typical of denatured random coiled proteins. Thus, the aperiodic structure in apo PSM could easily be underestimated, thereby inflating the proportions of β-turn, β-sheet, and α-helix structures.

The CD spectrum of the trypsin-treated native and asialo OSM peptide cores show a deepening of the minimum at 197 nm and the appearance of a small maximum at 220 nm, the latter being slightly more pronounced for OSM than asialo OSM. Such features are observed in extended polypeptides such as collagen and aqueous polyproline that have the polyproline II (pPII) conformation ($\phi = -80°$, $\psi = 160°$). The CD spectrum of peptides with the pPII conformation exhibits a maximum at 218–225 nm ($[\theta] \simeq 4500$ deg cm^2/dmol) and a minimum at 195–206 nm ($[\theta] \simeq -40\,000$ deg cm^2/dmol) (Ronish & Krimm, 1974). Thus, OSM may contain some pPII structure, although these domains are probably very limited in length and have a significant degree of flexibility. These conclusions are in keeping with the conformations obtained from the RIS calculations described above. The antifreeze glycoprotein, by comparison, is thought to contain a larger proportion of pPII structure on the basis of its CD spectrum ($[\theta]_{218} = 5000$ and $[\theta]_{197} = -20\,000$ deg cm^2/dmol; Bush et al., 1981). Its CD spectrum is also reproduced in Figure 5. Thus, all evidence points to native OSM having a somewhat more flexible structure than the antifreeze glycoprotein. This is consistent with the apparent lack of short-range sequence repeat (Hill et al., 1977B) and the presence of large amounts of glycine in OSM compared to the well-defined sequence repeat and absence of glycine in antifreeze glycoprotein.

Mucin Aggregation. It has been suggested (Hill et al., 1977a) that mucin oligosaccharides are involved in noncovalent carbohydrate–carbohydrate or carbohydrate–protein interactions and that these interactions are responsible for the high molecular weights of mucins through the formation of aggregates. This study and those of others (Sheehan & Carlstedt, 1987; Gupta & Jentoft, 1989) suggest that carbohydrate–carbohydrate interactions are unlikely to be involved in this type of aggregation since mucin fragments, which retain their carbohydrate structure, do not re-form species of high molecular weight. This is supported by the values of the second virial coefficients, A_2 (see Table I), which indicate that relatively good (nonassociative) solvent conditions, rather than associative solvent conditions prevail for the glycosylated mucins in Gdn-HCl and NaCl. Furthermore, in our hands enzymatically deglycosylated apo OSM is found to retain a rather high molecular weight. It is more difficult to ascertain whether carbohydrate–protein interactions occur since it can be difficult to differentiate strong noncovalent bonds from covalent linkages. However, the high molecular weight mucins used in these studies were purified in the presence of 5 M Gdn-HCl so that any noncovalent linkages present must be quite stable. It should be noted that the present studies demonstrate that OSM undergoes a minor relaxation of chain dimensions upon removal of sialic acid and a major conformational change upon complete deglycosylation. This series of modified mucins was also used in the studies of Hill et al. (1977a) from which it was proposed that OSM aggregates via its carbohydrate residues. We do not fully understand these discrepancies except to note that the significant changes in peptide core conformation accompanying deglycosylation may have been interpretated as an aggregation phenomenon.

Conclusions

Earlier light-scattering studies (Shogren et al., 1986, 1987) demonstrated that chain dimensions of mucins, as measured by $\langle R_g^2 \rangle^{1/2}$ or R_h, are dependent on the length of the mucin peptide core and independent of mucin carbohydrate side chain structure. Thus, all mucins for which data are available possess the same expanded random coil conformation regardless of their average carbohydrate side chain length. The studies described here were designed to extend these findings by directly examining how sialic acid and the linkage GalNAc residue influence the solution structure of OSM. OSM provided an ideal system for these studies since it has the same expanded conformation as the more complex mucins (Shogren et al., 1987).

The results clearly demonstrate that glycosylation dramatically restricts the conformation of the mucin peptide core. This is indicated by the large values of C_∞ and persistence length obtained for native and asialo OSM as compared to apo mucin. We have, therefore, attempted to further characterize the conformation of the mucin peptide core using the RIS model in its simplest form. Although the conformations obtained from the RIS analysis are not exact nor necessarily unique, they are consistent with the CD studies. The results suggest that the most probable conformations of the glycosylated mucin peptide core lie within the relatively narrow range of $-90° < \phi < -60°$ and $60° < \psi < 180°$, which includes that of polyproline II. In contrast, the deglycosylated protein core can be fit to the less restricted random coil conformations typical of denatured proteins. This points to the importance of carbohydrate, rather than amino acid sequence, in maintaining the extended mucin conformation. Since H-bonds do not stabilize the mucin conformation to a significant effect, it has a greater degree of flexibility than, for example, an α-helix. For comparison, polyproline in its II form, which is not stabilized by H-bonds, is relatively flexible, with C_∞ calculated to be 30 for the rotational isomer having $\psi = 150 \pm 50°$ (Mattice et al., 1973).

Steric interactions between the GalNAc and peptide core residues appear to be the major source of the extended mucin conformation, while interactions of sialic acid (2–6 linked to GalNAc) have lesser effects on the peptide core conformation. These interactions, furthermore, appear to extend to amino acid residues adjacent to the glycosylated Ser/Thr residues. Suprisingly, electrostatic interactions between sialic acid residues appear to play only a minor role at best in affecting mucin conformation at physiological ionic strengths. Other sugars linked to the GalNAc residue, including the 1–3-linked Gal residue found in PSM, may be at least as effective as sialic acid in restricting the peptide core conformation, while substitution of GalNAc at both carbons 3 and 6 may further restrict the mobility of the peptide core to a smaller extent. Carbon-13 NMR measurements of the dynamics of the peptide core carbons in native, asialo, and apo mucin are consistent with the findings reported here (Gerken et al., 1989).

The same steric interactions between carbohydrate and protein that dictate the conformation of mucins are also likely

to be operative in other glycoproteins that contain O-linked oligosaccharides. Membrane glycoproteins in particular frequently contain relatively short (20–50 amino acid residues) sequences that are heavily O-glycosylated. Our results on mucin conformation predict that these sequences would exist as highly extended semiflexible rods with a length of about 2.5 Å per amino acid residue. Although this conformation cannot, by definition, be considered a type of secondary structure, it does appear to represent a new structural element in glycoproteins that is specified by amino acid sequence and by the actions of specific glycosyltransferases.

Role of Mucin Carbohydrate Side Chains. From the above it is readily apparent that the carbohydrate residues attached to the mucin protein core act primarily to expand and stiffen the peptide core. In addition, it has been shown that the oligosaccharide side chains must also be important for hydrating and solubilizing a relatively insoluble mucin core protein. For mucins with two or more carbohydrate residues per side chain, the resulting nearly 3-fold expansion in chain dimensions, as measured by $\langle R_g^2 \rangle^{1/2}$ and R_h, represents a 20–25-fold increase in the volume that a glycosylated mucin occupies in solution compared to the nonglycosylated peptide core. With such an expanded structure and with a sufficiently long peptide core, mucin molecules will overlap extensively and entangle at low concentrations, thus giving rise to the viscoelastic properties characteristic of mucous gels. Although direct evident is lacking, it is tempting to speculate that while oligosaccharide side chains longer than two sugars may not have significant additional effects on mucin conformation, they may contribute to viscoelastic properties through their contribution to the frictional properties of the mucin molecules.

ACKNOWLEDGMENTS

We thank Drs. J. Blackwell and A. Jamieson of the Department of Macromolecular Science at Case Western Reserve University for the use of their light-scattering apparatus.

REFERENCES

Bell, A. E., Allen, A., Morris, E., & Rees, D. A. (1982) *Adv. Exp. Med. Biol. 144*, 3–28.
Benoit, H., & Doty, P. (1953) *J. Phys. Chem. 57*, 958–963.
Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.
Bloomfield, V. (1983) *Biopolymers 22*, 2141–2154.
Brant, D. A., & Flory, P. J. (1965) *J. Am. Chem. Soc. 87*, 2788–2791.
Brant, D. A., Miller, W. G., & Flory, P. J. (1967) *J. Mol. Biol. 23*, 47–65.
Bush, C. A., & Feeney, R. E. (1986) *Int. J. Pept. Protein Res. 28*, 386–397.
Bush, C. A., Feeney, R. E., Osuga, D. T., Ralapati, S., & Yeh, Y. (1981) *Int. J. Pept. Protein Res. 17*, 125–129.
Cantor, C., & Schimmel, P. (1980) *Biophysical Chemistry, Part III, The Behavior of Biological Macromolecules*, pp 979–1018, Freeman, New York.
Carlson, D. M. (1969) *J. Biol. Chem. 243*, 616–626.
Carlstedt, I., & Sheehan, J. K. (1984) *Biochem. Soc. Trans. 12*, 615–617.
Carlstedt, I., Lindgren, H., & Sheehan, J. K. (1983) *Biochem. J. 213*, 427–435.
Carlstedt, I., Sheehan, J. K., Corfield, A. P., & Gallagher, J. T. (1985) *Essays Biochem. 20*, 40–76.
Codutis, P. L., Gordon, E. C., & Bush, C. A. (1977) *Anal. Biochem. 78*, 9–20.
Creeth, J. M., & Knight, C. G. (1967) *Biochem. J. 105*, 1135–1145.
Debye, P., & Huckel, E. (1923) *Phys. Z. 24*, 185.
Dickinson, H. R., & Bush, C. A. (1975) *Biochemistry 14*, 2299–2304.
Duben, A., & Bush, C. A. (1983) *Arch. Biochem. Biophys. 225*, 1–15.
Eckhardt, A. E., Timpte, C. S., Abernathy, J. L., Toumadge, A., Johnson, C. W., & Hill, R. L. (1987) *J. Biol. Chem. 262*, 11339–11344.
Edge, A. S. B., Falthynek, C. R., Hof, L., Reichert, L. E., & Weber, P. (1981) *Anal. Biochem. 118*, 131–137.
Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, pp 1–29, 250–304, Wiley, New York.
Fraser, R. D. D., & MacRae, R. P. (1973) *Conformation of Fibrous Proteins*, p 136, Academic Press, New York.
Gerken, T. A. (1986) *Arch. Biochem. Biophys. 247*, 239–253.
Gerken, T. A., & Dearborn, D. G. (1984) *Biochemistry 23*, 1485–1497.
Gerken, T. A., & Jentoft, N. (1987) *Biochemistry 26*, 4689–4699.
Gerken, T. A., Butenhof, K., & Shogren, R. (1989) *Biochemistry* (following paper in this issue).
Gottschalk, A., & Bhargava, A. (1972) in *Glycoproteins: Their Composition, Structure and Function* (Gottschalk, A., Ed.) Part B, pp 810–829, Elsevier, New York.
Gupta, R., & Jentoft, N. (1989) *Biochemistry* (in press).
Hill, H. D., Reynolds, J. A., & Hill, R. L. (1977a) *J. Biol. Chem. 252*, 3791–3798.
Hill, H. D., Schwyzer, M., Steinmen, H. M., & Hill, R. L. (1977b) *J. Biol. Chem. 252*, 3799–3804.
Huber, K., & Stockmayer, W. H. (1987) *Macromolecules 20*, 1400–1402.
Huber, K., Bantle, S., Lutz, P., & Burchard, W. (1985) *Macromolecules 18*, 1461–1467.
Jentoft, N. (1985) *Anal. Biochem. 148*, 424–433.
Koppel, D. E. (1972) *J. Chem. Phys. 57*, 4814–4820.
Kratky, O., & Porod, G. (1949) *Recl. Trav. Chim. 68*, 1106.
Lamblin, G., Lhermitte, M., Degand, P., & Roussel, P. (1979) *Biochimie 61*, 23–43.
Lapanje, S., & Tanford, C. (1967) *J. Am. Chem. Soc. 89*, 5030–5033.
Levine, M. J., Reddy, M. S., Tabrak, L. A., Loomis, R. E., Bergey, E. J., Jones, P. C., Cohen, R. E., Stimson, M. W., & Al-Hashimi, I. (1987) *J. Dent. Res. 66*, 436–441.
Mantle, M., & Allen, A. (1978) *Biochem. Soc. Trans. 6*, 607–609.
Mattice, W. L., Nishikawa, K., & Ooi, T. (1973) *Macromolecules 6*, 443–446.
Meyer, F. A., & Silberberg, A. (1978) *Respiratory Tract Mucins, Ciba Found. Symp. 54*, 203–218.
Mikkelsen, A., Stokke, B. T., Christensen, B. E., & Elgsaeter, A. (1985) *Biochemistry 24*, 1683–1704.
Miller, W. G., & Goebel, C. V. (1968) *Biochemistry 7*, 3925–3935.
Montreuil, J. (1984) *Biol. Cell. 51*, 115–132.
Neutra, M. R., & Forstner, J. F. (1987) in *Physiology of the GI Tract* (Johnson, L. R., Ed.) pp 975–1009, Raven Press, New York.
Porath, J. (1963) *Pure Appl. Chem. 6*, 233–244.
Rao, B. N. N., & Bush, C. A. (1987) *Biopolymers 26*, 1277–1244.
Reissig, J. L., Strominger, J. L., & Leloir, L. F. (1955) *J. Biol. Chem. 217*, 959–966.
Ronish, E. W., & Krimm, S. (1972) *Biopolymers 11*, 1919–1928.
Ronish, E. W., & Krimm, S. (1974) *Biopolymers 13*, 1635–1651.

Rose, M. C., Voter, W. A., Sage, H., Brown, C. F., & Kaufman, B. (1984) *J. Biol. Chem. 259*, 3167-3172.
Schimmel, P. R., & Flory, P. J. (1968) *J. Mol. Biol. 34*, 105-120.
Sheehan, J. K., & Carlstedt, I. (1984a) *Biochem. J. 217*, 93-101.
Sheehan, J. K., & Carlstedt, I. (1984b) *Biochem. J. 221*, 499-504.
Sheehan, J. K., & Carlstedt, I. (1987) *Biochem. J. 245*, 757-762.
Sheehan, J. K., Oates, K., & Carlstedt, I. (1986) *Biochem. J. 239*, 147-153.
Shogren, R. L. (1985) Ph.D. Thesis, Case Western Reserve University.
Shogren, R. L., Jamieson, A. M., Blackwell, J., & Jentoft, N. (1986) *Biopolymers 25*, 1505-1517.
Shogren, R. L., Jentoft, N., Gerken, T. A., Jamieson, A. M., & Blackwell, J. (1987) *Carbohydr. Res. 160*, 317-327.
Slayter, H. S., & Codington, J. F. (1973) *J. Biol. Chem. 248*, 3405-3410.
Smith, J. A., & Pease, L. G. (1980) *CRC Crit. Rev. Biochem. 8*, 315.
Stokke, B. T., Elgsaeter, A., Akjak-Braek, G., & Smidsrod, O. (1987) *Carbohydr. Res. 160*, 13-28.
Svennerholm, L. (1958) *Acta Chem. Scand. 12*, 547-554.
Tanford, C., Kawahara, K., & Lapanje, S. (1966) *J. Biol. Chem. 241*, 1921-1922.
Tiffany, M. L., & Krimm, S. (1969) *Biopolymers 8*, 347-359.
Woodward, H. D., Ringler, N. J., Selvakumar, R., Simet, I. M., Bhavanandan, V. P., & Davidson, E. A. (1987) *Biochemistry 26*, 5315-5322.
Yang, J. T., Chuen-Shang, C. W., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208-269.
Yamakawa, H. (1984) *Annu. Rev. Phys. Chem. 35*, 23-47.
Yamakawa, H., & Fujii, M. (1973) *Macromolecules 6*, 407-415.

Effects of Glycosylation on the Conformation and Dynamics of O-Linked Glycoproteins: Carbon-13 NMR Studies of Ovine Submaxillary Mucin†

Thomas A. Gerken,* Kenneth J. Butenhof,‡ and Randal Shogren‡,§
W. A. Bernbaum Center for Cystic Fibrosis Research, Department of Pediatrics and Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106

Received September 27, 1988; Revised Manuscript Received February 23, 1989

ABSTRACT: Carbon-13 NMR spectroscopic studies of native and sequentially deglycosylated ovine submaxillary mucin (OSM) have been performed to examine the effects of glycosylation on the conformation and dynamics of the peptide core of O-linked glycoproteins. OSM is a large nonglobular glycoprotein in which nearly one-third of the amino acid residues are Ser and Thr which are glycosylated by the α-NeuNAc(2-6)α-GalNAc- disaccharide. The β-carbon resonances of glycosylated Ser and Thr residues in intact and asialo mucin display considerable chemical shift heterogeneity which, upon the complete removal of carbohydrate, coalesces to single sharp resonances. This chemical shift heterogeneity is due to peptide sequence variability and is proposed to reflect the presence of sequence-dependent conformations of the peptide core. These different conformations are thought to be determined by steric interactions of the GalNAc residue with adjacent peptide residues. The absence of chemical shift heterogeneity in apo mucin is taken to indicate a loss in the peptide–carbohydrate steric interactions, consistent with a more relaxed random coiled structure. On the basis of the ^{13}C relaxation behavior (T_1 and NOE) the dynamics of the α-carbons appear to be unique to each amino acid type and glycosylation state, with α-carbon mobilities decreasing in the order Gly > Ala = Ser > Thr $\gg$ monoglycosylated Ser/Thr $\gtrsim$ disaccharide linked Ser/Thr. The α-carbons of glycosylated Ser and Thr are considerably more constrained than their nonglycosylated counterparts in apo mucin, while the effects of carbohydrate side chain length (i.e., asialo vs native mucin) on the dynamics of the Ser and Thr residues is relatively small. The nonglycosylated Gly residue also exhibits an increase in motion upon removal of GalNAc; thus, the effects of glycosylation extend to residues beyond the amino acids directly bound to carbohydrate. These results are consistent with the changes in molecular dimensions determined by light-scattering techniques for the same series of modified mucins [Shogren et al. (preceding paper in this issue)]. Taken together, these results further demonstrate that mucins possess a highly expanded conformation that is dominated by steric interactions between the peptide core and the O-linked GalNAc residue.

It has been suspected for some time that the presence and nature of the O-linked oligosaccharide side chains in mucous glycoproteins (mucins) play an important role in governing their viscoelastic properties. Presently, however, very little information is available that describes the effects of O-glycosylation on the conformation and dynamics of O-linked glycoproteins. Likewise, little is known of the mechanism of how the carbohydrate or peptide structure may contribute to the observed physical properties of mucin glycoproteins. To obtain a more complete understanding of the effects of mucin glycosylation on mucin solution structure at both the atomic and macromolecular levels, we have undertaken parallel

† This work was supported by grants from the Rainbow Chapter of the Cystic Fibrosis Foundation and the United Way of Cleveland, by a Research Grant from the Cystic Fibrosis Foundation, and by NIH Grants DK 39918 and DK 27651.
* Address correspondence to this author.
‡ Fellow of the Cystic Fibrosis in Pulmonary Disease training grant, NHLBI-HL 07415.
§ Present address: USDA, Northern Regional Research Center, Peoria, IL 61604.

0006-2960/89/0428-5536$01.50/0