Subunit Structure of Porcine Submaxillary Mucin†

Rekha Gupta and Neil Jentoft*
Departments of Pediatrics and Biochemistry, Rainbow Babies and Children's Hospital, Case Western Reserve University, Cleveland, Ohio 44106

Received September 26, 1988; Revised Manuscript Received March 8, 1989

ABSTRACT: The structure of a high molecular weight fraction of porcine submaxillary mucin was studied by using degradative techniques. Reduction of disulfide linkages released mucin subunits together with an associated protein(s) of approximately 140 kDa. The molecular weights of the subunits ranged from approximately 0.5×10^6 to 2.5×10^6. Trypsinization of subunits generated glycosylated domains and small, poorly glycosylated or nonglycosylated tryptic peptides. The glycosylated domains, which have an average molecular weight of approximately 270K, possess an unusual amino acid composition containing only nine different amino acids. The minor amino acids which are absent from the glycosylated domains but which are consistently present in both the mucin and the mucin subunits were recovered in the tryptic peptides. Pronase digestion of the glycosylated domains generated smaller fragments of approximately 17 kDa. Comparing these results to the partial cDNA sequence for porcine submaxillary mucin reported by Timpte et al. [(1988) *J. Biol. Chem. 263*, 1081-1088] suggests that the glycosylated domains consist of variable numbers of the 81 amino acid tandem repeat observed in the cDNA sequence. Further, the fact that porcine submaxillary mucin contains subunits, link proteins, and glycosylated domains suggests that its structure is similar to that described for cervical and intestinal mucins. Intact mucin, mucin "subunits", and the glycosylated domains are all polydisperse with respect to molecular weight, indicating that mucin polydispersity is due to variability in the number of units linked together as well as to variability in the size of the units.

Mucous glycoproteins contain a peptide core rich in serine, threonine, glycine, alanine, and proline that is heavily substituted with oligosaccharides linked O-glycosidically to the serine and threonine residues (Carlstedt et al., 1985; Neutra & Forstner, 1987). Reported molecular weights for mucins range from less than 1×10^6 to greater than 20×10^6, with the higher values coming from light-scattering studies of materials isolated by using protease inhibitors and minimal shear forces (Carlstedt et al., 1985; Shogren et al., 1986, 1987). Physical studies indicate that mucins are linear, unbranched molecules which have a relatively stiff random-coil conformation (Carlstedt et al., 1985; Shogren et al., 1986, 1987) while electron microscopy studies (Marianne et al., 1987; Sheehan et al., 1987) confirm the linear nature of mucins. Mucins are characteristically polydisperse with respect to molecular weight; cervical mucins, for example, possess molecular weights ranging from 6×10^6 to 24×10^6 when separated by rate zonal centrifugation (Sheehan & Carlstedt, 1987), and ovine submaxillary mucin (OSM)[1] can be fractionated by gel filtration on Sephacryl S-1000 into pools having average molecular weights ranging from 0.8×10^6 to 5.2×10^6 (Shogren et al., 1987).

Many studies suggest that mucins possess some type of subunit structure; reduction of disulfide linkages in mucins releases subunits while proteolytic cleavage releases fragments of ca. 300-500 kDa from intestinal (Allen, 1983), cervical (Carlstedt et al., 1985; Meyer, 1983), and tracheobronchial (Ringler et al., 1987) mucins. In addition to depolymerizing mucins, reducing agents also release a polypeptide(s) which has (have) been termed the "link protein" (Allen, 1983). Link proteins of 60-120 kDa have been described for intestinal (Fahim et al., 1987) and tracheobronchial mucins (Ringler et al., 1987). It has generally been assumed that these link proteins bridge mucin subunits via disulfide bonds since reduction of disulfide linkages releases the link protein with a concomitant decrease in mucin size. However, it is also possible that this reduction in size is due to breaks in the peptide chain which remain cryptic until disulfide bonds which bridge the portion containing the break are reduced (Carlstedt et al., 1985).

Other studies on mucin structure have used enzymatic deglycosylation of mucins. Applied to salivary mucins, this method yields apomucins of 60-100 kDa (Hill et al., 1977; Eckhardt et al., 1987) corresponding to fully glycosylated materials of 150-300 kDa. Interestingly, if the varying ratios of carbohydrate to protein in different mucins are taken into account, both the chemical degradation and enzymatic deglycosylation studies point to a basic structure in mucins consisting of a ca. 100-kDa peptide. In spite of the basic agreement between these two approaches as to the sizes of the products that are generated, the results of enzymatic deglycosylation studies imply that the monomeric units of mucins are assembled via noncovalent interactions involving carbohydrate-carbohydrate or carbohydrate-protein interactions (Hill et al., 1977) while the results of studies using reduction and/or proteolysis suggest that covalent linkages must be cleaved in order to produce the monomeric unit. Since mucins are not depolymerized into monomers by exposure to 5 M guanidine hydrochloride (Shogren et al., 1984, 1986; Carlstedt et al., 1985) and since they can be fractionated into subclasses of differing molecular weights and maintain those molecular weights over time (Sheehan & Carlstedt, 1987; Shogren et al., 1987), it is clear that any noncovalent linkages that are present must be stable to denaturing solvents and that these linkages must also be sufficiently strong to preclude monomer/polymer equilibria.

† Supported by NIH Grants AM27651 and HL32227 and by grants from the Cystic Fibrosis Foundation.

* To whom correspondence should be addressed at the Department of Pediatrics, Case Western Reserve University.

[1] Abbreviations: PSM, porcine submaxillary mucin; OSM, ovine submaxillary mucin; DTT, dithiothreitol; SDS, sodium dodecyl sulfate.

0006-2960/89/0428-6114$01.50/0

One possible explanation for the dichotomy between the results from the deglycosylation studies and those using biochemical fragmentation of mucins may be that different classes of mucins have different structures. The biochemical studies have mainly been performed with cervical and intestinal mucins while enzymatic deglycosylation has only been applied to salivary mucins (Hill et al., 1977; Eckhardt et al., 1987) which have generally been thought to lack disulfide linkages and link proteins. The present studies were undertaken to determine whether PSM resembles the intestinal and cervical mucins in its response to disulfide bond cleavage and proteolysis. The results demonstrate that PSM is analogous to cervical mucins in terms of the products released by reduction and proteolysis and also that it possesses a "link protein" in common with other mucins. Moreover, striking similarities are noted between the mucin fragments released by sequential reduction, alkylation, and trypsinolysis and the apomucin generated by deglycosylation. Finally, investigation into the structural basis of mucin polydispersity reveals that PSM and its cleavage products are polydisperse at every level of structure. A preliminary report of these findings has been presented (Gupta & Jentoft, 1987).

Experimental Procedures

Mucin Purification. PSM was isolated from one pair of porcine submaxillary glands from a single animal as described previously (Shogren et al., 1986) except that 5 mM *N*-ethylmaleimide was added to the homogenization buffer (0.01 M Tris-HCl, pH 7.6) to inhibit possible interchange of disulfide bonds (Oike et al., 1980). Solid guanidine hydrochloride was added to the unfractionated mucin to give a final concentration of 1–1.5 mg/mL in 5 M guanidine hydrochloride. This solution was applied to a 9 × 80 cm column of Sephacryl S-1000 and eluted with 5 M guanidine hydrochloride containing 10 mM sodium phosphate buffer, pH 7.0. The broad peak obtained was arbitrarily divided into three pools as shown in Figure 1. The material in pool 1 was further purified by density gradient centrifugation (Bhaskar & Creeth, 1974) in 4 M guanidine hydrochloride/CsCl adjusted to a density of 1.4 g/mL. Samples were centrifuged at 4 °C for 48 h at 42 000 rpm in a Beckman Model L centrifuge using a type 60 fixed-angle rotor and the gradients fractionated using a Model 185 density gradient fractionator from Isco. The fractions were assayed by the periodic acid/Schiff assay (Mantle & Allen, 1978) and by their absorbance at 220 and 280 nm. The density of the fractions was measured by weighing 25-μL aliquots. Fractions containing mucin were dialyzed and stored at −20 °C. Purified mucin was also typed for blood group activity by the procedure of Carlson (1968) and was found to be A positive.

Reduction and Alkylation. Disulfide bonds were cleaved by reducing PSM under nitrogen at 2.5 mg/mL final concentration with 15 mM DTT in 6 M guanidine hydrochloride, 5 mM EDTA, and 0.2 M Tris, pH 7.0. Reductions were carried out under different conditions in different experiments; conditions tested included 1 h at 45 °C, 5 h at room temperature, and overnight at 4 °C. Identical gel filtration patterns were observed for mucin samples reduced under these different conditions. For carboxymethylation with unlabeled iodoacetamide, the pH was raised to 8.5, iodoacetamide added to a final concentration of 60 mM, and the solution incubated overnight in the dark at 4 °C (Konigsberg, 1972). For carboxymethylation with [^{14}C]iodoacetamide, the concentration of DTT was first reduced either by dialysis vs 1 mM DTT/7 mM EDTA or by complexing the DTT with $NaAsO_2$ (Zahler & Cleland, 1968).

In order to determine whether reduction was reversible, reduced PSM was dialyzed against water to remove DTT and then reoxidized overnight at 4 °C in 10 mM Cu^{2+} and oxygen (Ristow & Wetlaufer, 1973). In separate experiments, the reoxidation was performed in the presence or absence of 6 M guanidine hydrochloride. The mucin samples were then alkylated with [^{14}C]iodoacetamide as described above and molecular weight distributions assessed by gel filtration on Sephacryl S-1000. In experiments using sodium borohydride rather than DTT as the reducing agent, reduction was carried out using 1 mg/mL $NaBH_4$ in 6 M guanidine hydrochloride, 0.5 M Tris, and 2 mM EDTA, pH 7.0.

Polyacrylamide gel electrophoresis of the link protein was carried out on a 10–22% acrylamide gradient gel and stained with Coomassie Brilliant Blue. Preparative SDS–polyacrylamide gel electrophoresis was carraied out on 7.5% gels (Laemmli, 1970), and the bands were cut out and electroeluted (Hunkapillar et al., 1983). SDS was removed by ion pair extraction (Henderson et al., 1979).

Protease Treatment of PSM Subunits. L-(Tosylamino)-2-phenylethyl chloromethyl ketone treated trypsin (1.5 μg) was added to a 2.0-mg sample of the [^{14}C]iodoacetamide-labeled PSM subunit fraction from S-1000 chromatography (Figure 3) in 80 mM NH_4HCO_3, pH 8.5. The solution was incubated at 37 °C for 24 h, a second 1.5-μg portion of trypsin added, and the incubation repeated. The solution was lyophilized, redissolved in 0.5 M NH_4HCO_3, pH 7.5, and chromatographed on a 1 × 110 cm column of Sephacryl S-500 eluted with the same solvent. The high and low molecular weight peaks were pooled for Pronase treatment and peptide mapping, respectively. In a second experiment, trypsin was added at an enzyme:substrate ratio of 0.36:1, and the products were chromatographed on Sephacryl S-500 (1 × 115 cm) in 5 M guanidine hydrochloride. Identical chromatographic patterns were obtained at both trypsin concentrations.

The high molecular weight material from the tryptic digest (Figure 6) was digested with Pronase using, in separate experiments, enzyme:substrate ratios of 1:10 and 1:100 in 0.2 M ammonium acetate buffer, pH 7.0, containing 1 mM $CaCl_2$ and a drop of toluene. After incubation at 37 °C for 72 h, the digests were lyophilized and then chromatographed on columns of Sephacryl S-200 equilibrated in either 0.5 M ammonium bicarbonate buffer, pH 7.5, or 5 M guanidine hydrochloride. In all cases, similar chromatographic patterns were observed.

Analytical Procedures. Amino acid analyses were carried out by HPLC of the phenythiocarbamyl derivatives using the method of Heinrikson and Meredith (1984) modified to allow analysis of (carboxymethyl)cysteine and hexosamines (Gupta & Jentoft, 1989). Column effluents were assayed for carbohydrate by the method of Mantle and Allen (1978) while carbohydrate analyses were done as described by Jentoft (1985). Disulfide linkages were quantified by using the procedure of Thannhauser et al. (1984).

Light-Scattering Studies. Light-scattering studies were performed as previously described (Shogren et al., 1986, 1987). Briefly, stock solutions of PSM were made up by adding solid guanidine hydrochloride to a final concentration of 6 M followed by exhaustive dialysis vs 6 M guanidine hydrochloride. The concentration of PSM in the stock solution was determined by differential refractometry using a value of 0.135 mL/g (Shogren et al., 1984) for the refractive index increment of PSM in 6 M guanidine hydrochloride. A series of concentrations were made for each of the samples by diluting the stock solutions. The solutions were then filtered through 5-μm

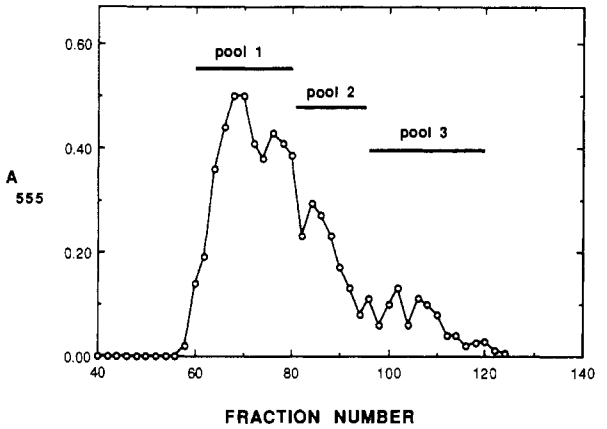

FIGURE 1: Gel filtration of PSM on Sephacryl S-1000. Partially purified mucin from one pair of porcine submaxillary glands was dissolved in 5 M guanidine hydrochloride, pH 7.0, applied to a 9 × 80 cm Sephacryl S-1000 column, and eluted with the same buffer. The eluate was assayed for carbohydrate by the periodic acid/Schiff assay (Mantle & Allen, 1978). The fractions were combined into pools as indicated by the bars.

(native mucin), 0.45-μm (carboxymethylated mucin), or 0.22-μm (protease-treated mucin) filters. Light-scattering measurements were performed by using laser light of wavelength 632.8 nm and a BI 240 photogoniometer with a 2030 autocorrelator (Brookhaven Instruments, Ronkonkoma, NY). The average intensities of scattered light were measured as a function of scattering angle for high molecular weight material. Zimm plots of intensity data yield ⟨MW⟩ and the second virial coefficient A_2. The intensity autocorrelation of the scattered light was analyzed in order to get the normalized second moment μ_2/γ^2 in the limits of $\theta \rightarrow 0$, $c \rightarrow 0$. This parameter is equal to the normalized variance of Dt^0 and reflects the polydispersity in the hydrodynamic radius of the solute.

Materials. Guanidine hydrochloride, absolute grade, was purchased from Schwarz/Mann. Technical-grade guanidine hydrochloride was treated with activated charcoal and partially purified by dialysis through a hollow-fiber dialyzer (Bio-Rad) before use in the 9 × 80 cm preparative S-1000 column. Ultrapure CsCl was from Bethesda Research Laboratories. L-(Tosylamino)-2-phenylethyl chloromethyl ketone treated trypsin was from Cooper Biochemicals, and Pronase was obtained from Calbiochem. [^{14}C]Iodoacetamide was obtained from Amersham. Sephacryl S-1000, S-500, and S-200 and Sephadex G-10 were from Pharmacia. Molecular weight standards were from Bio-Rad. Other reagents used were of reagent grade.

Results

Purification and Characterization of Mucin. PSM was purified by methods described previously (Shogren et al., 1986) in which the tissues are homogenized for 30 s at half-speed and in which protease inhibitors are added during the initial stages of the purification. These measures are necessary to obtain mucin samples of high molecular weight. In these experiments, *N*-ethylmaleimide was also added to the homogenate in order to block free sulfhydryl groups and thus inhibit possible disulfide interchange. The unfractionated mucin preparation was then applied to a column of Sephacryl S-1000 and eluted with 5 M guanidine hydrochloride in order to separate the mucin into fractions of differing molecular weight. The elution pattern obtained is shown in Figure 1. Fractions were pooled as shown by the bars. Aliquots from these pools were rechromatographed on an analytical Sephacryl S-1000 column as shown in Figure 2. Pools 1–3 eluted from the

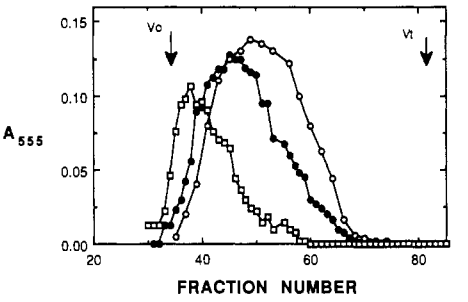

FIGURE 2: Rechromatography of PSM pool 1 (□), pool 2 (●), and pool 3 (○) on a 1 × 115 cm column of Sephacryl S-1000. The column was eluted at a flow rate of 8 mL/h with 5 M guanidine hydrochloride/10 mM phosphate, pH 7.0, and fractions (1 mL) were assayed for carbohydrate by the periodic acid/Schiff assay (Mantle & Allen, 1978). λ phage DNA and AMP were used as void and total volume markers.

Table I: Molecular Weights of PSM and PSM Fragments

sample	MW × 10^6 (light scattering)	$R_{g,z}$ (Å)	μ_2/γ^2	MW × 10^6 (gel filtration)[a]
PSM pool 1	8.6	2340	0.4	8.4
PSM pool 2	ND[b]	ND	ND	3.9
PSM pool 3	ND	ND	ND	2.8
pool 1 subunits	2.1	1110	0.3	1.8
pool 2 subunits	ND	ND	ND	1.3
pool 3 subunits	ND	ND	ND	1.0
glycosylated domains	0.27	330	0.2–0.3	0.36
Pronase glycopeptides	0.017	ND	ND	ND

[a] Molecular weights were determined from K_d values as described in Shogren et al. (1987). [b] ND, not determined.

analytical columns with K_d values of 0.05, 0.20, and 0.32 correlating to average molecular weight values of 8.4 × 10^6, 3.9 × 10^6, and 2.8 × 10^6, respectively, as shown in Table I. The molecular weight for pool 1 mucin was also measured by static light-scattering techniques, and the results are listed in Table I. PSM isolated by this procedure lacks detectable amounts of nucleic acids as determined by the UV absorption at 260 nm, proteoglycans as determined by the absence of sulfate and glucosamine, and non-mucin proteins as determined by SDS gel electrophoresis on 7.5% gels (Laemmli, 1970) using both Coomassie blue and silver staining. Further, density gradient centrifugation in CsCl/4 M guanidine hydrochloride demonstrated the presence of a single peak banding at 1.4–1.5 g/mL and the absence of other materials (data not shown). Compositional analyses of the mucin pools are given in Table II. PSM from pool 1 was analyzed for the presence of disulfide linkages by using the sodium 2-nitro-5-thiosulfobenzoate assay as described by Thannhauser et al. (1984); the results of this experiment indicated a cysteine content of 0.5 residue per 100 amino acid residues.

Mucin Subunits. In order to investigate the role of these disulfide linkages in the structure of PSM, a sample of PSM from pool 1 was reduced with DTT and then carboxymethylated with [^{14}C]iodoacetamide as described under Experimental Procedures. Gel filtration of the reduced and alkylated sample on Sephacryl S-1000 (Figure 3) revealed two major peaks; these were pooled for further analysis as shown by the bars. The higher molecular weight peak eluted from the column at a K_d value of 0.34 and contained both carbohydrate, as measured by the periodic acid/Schiff assay (Mantle & Allen, 1978), and radioactivity derived from [^{14}C]iodoacetamide. This material is termed the mucin subunit. A second peak containing radioactivity from [^{14}C]-iodoacetamide eluted at the included volume. Upon SDS–polyacrylamide gel electrophoresis (Figure 4), this material

Table II: Comparison of Amino Acid Compositions of PSM and PSM Fragments to the Calculated Composition of the 81 Amino Acid Repeat of Timpte et al. (1988)[a]

amino acid	pool 1	subunits	link protein	glycosylated domains	tryptic peptides	Pronase glycopeptides	repeat[b]
Asx	2.8	2.3	10.0	0.2	10.3	0.2	0
Glx	7.1	7.0	7.4	4.3	10.9	4.4	4.9
CM-Cys	0.5[c]	0.3	0.2	0	3.6	0	0
Ser	20.3	20.1	14.1	23.9	12.7	24.7	24.7
Gly	16.6	16.0	11.6	22.2	14.0	21.8	22.2
His	0.7	0.5	1.7	0.1	6.6	0	0
Arg	3.8	3.9	8.6	3.0	6.7	4.4	2.5
Thr	13.5	13.7	10.8	14.3	6.1	14.3	13.6
Ala	14.7	12.9	7.1	14.5	3.1	13.8	16.0
Pro	7.3	7.3	5.3	6.8	5.3	7.4	6.2
Tyr	0.4	0.6	0	0.1	2.7	0	0
Val	9.0	9.0	6.7	7.1	6.3	6.8	7.4
Ile	2.3	2.5	0.7	2.8	3.6	1.8	2.5
Leu	1.6	1.5	7.4	0.2	4.3	0	0
Phe	0.8	1.1	4.4	0.2	2.0	0.1	0
Lys	1.2	1.5	4.1	0	5.6	0	0

[a] Values in mole percent. [b] Values calculated from the cDNA sequence data of Timpte et al. (1988). [c] Disulfide linkages determined colorimetrically by the method of Thannhauser et al. (1984).

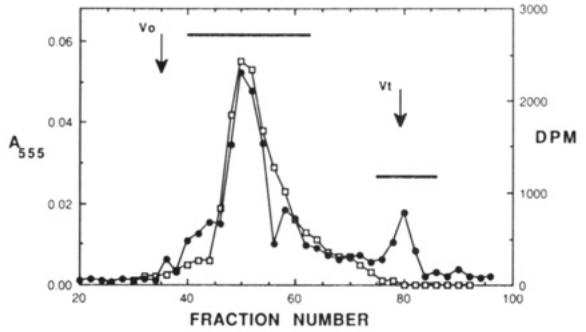

FIGURE 3: Chromatography of ^{14}C-carboxymethylated PSM. Mucin from pool 1 was reduced and carboxymethylated with [^{14}C]iodoacetamide as described under Experimental Procedures and then fractionated on a Sephacryl S-1000 column. Chromatographic conditions were as described in Figure 2. Fractions were assayed for carbohydrate (□) by using the periodic acid/Schiff assay (Mantle & Allen, 1978) and for radioactivity (●) by scintillation counting. Fractions were pooled for further analysis as shown by the bars.

gave two closely spaced bands of approximately 140 kDa. Thus, PSM appears to contain associated peptides that are presumably linked to the subunits via disulfide bonds and which appear to be analogous to the 70-kDa "link protein" described by Allen (1983) and the 118-kDa material described by Fahim et al. (1987). The amino acid composition of the link protein from PSM is similar to that from rat intestinal mucin (Fahim et al., 1987) with the most striking difference being its relatively low content of cysteine. Detection of galactosamine in amino acid analyses suggests that the porcine material, like the intestinal link protein, contains carbohydrate. The reduction in the molecular weight of the mucin that accompanies cleavage of disulfide bonds suggests that this link protein may serve as a bridge between subunits, but other interpretations of these data cannot be ruled out.

Molecular weight values for the PSM subunit determined from gel filtration data and light-scattering measurements are shown in Table I, and the Zimm plot for the subunits is shown in Figure 5. These values are in close agreement with molecular weights determined after reduction of PSM with β-mercaptoethanol using PSM samples purified by methods similar to those used here (Shogren et al., 1986) but are somewhat higher than the values determined in earlier work in which mucin was purified in the absence of protease inhibitors and using greater shear forces (Shogren et al., 1983, 1984). Both the gel filtration and light-scattering data demonstrate that this subunit preparation is polydisperse. Com-

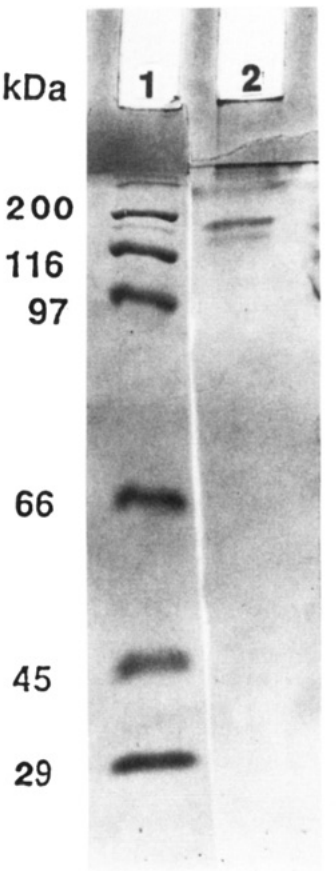

FIGURE 4: Electrophoresis of the link protein. The included volume fraction from Figure 3 was dialyzed and an aliquot subjected to SDS gel electrophoresis (lane 2) and compared to molecular weight standards (lane 1). The gel was stained with Coomassie blue.

Table III: Carbohydrate Analysis of PSM and PSM Fragments

	carbohydrate composition[a] (mole ratio relative to GalNAc)			
sugar	mucin pool 1	mucin subunits	glycosylated domains	Pronase glycopeptides
GalNAc	1.00	1.00	1.00	1.00
fucose	0.41	0.44	0.44	0.42
sialic acid	0.45	0.45	0.34	0.35
Gal	0.49	0.46	0.44	0.44

[a] Values represent the average of three determinations.

positional analyses (Tables II and III) reveal only small differences between the intact mucin and the subunits derived

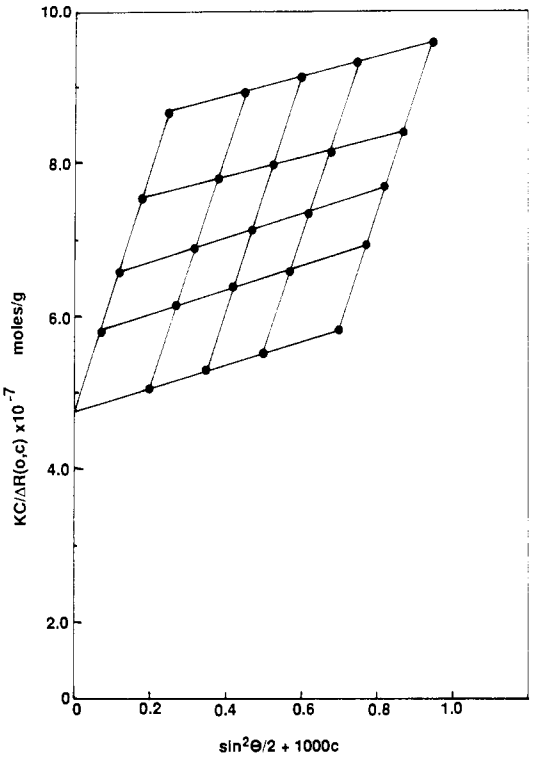

FIGURE 5: Zimm plot for mucin subunits in 6 M guanidine hydrochloride/10 mM phosphate buffer, pH 7.0. Extrapolation to $c = 0$ and $\theta = 0$ gives MW = 2.1×10^6 and $R_{g,z}$ = 1110 Å. The values for the second virial coefficient were positive, indicating good solvent conditions and minimal solute–solute interactions.

from the mucin. Amino acid analysis of the reduced and alkylated mucin subunit demonstrated a (carboxymethyl)-cysteine content of 0.3 mol %, a value in reasonable agreement with the colorimetric determination of disulfide bonds in unreduced PSM described above. Since label incorporated into the subunit fraction from [^{14}C]iodoacetamide was recovered as (carboxymethyl)cysteine in amino acid analysis, this incorporation was due to reaction with cysteine residues rather than being due to side reactions with other amino acids.

Reduction of PSM with DTT followed by dialysis and reoxidation of sulfhydryls in the presence of Cu^{2+} and O_2 and labeling of the remaining free sulfhydryls with [^{14}C]iodoacetamide was attempted in order to determine whether the molecule could be reassembled into a high molecular weight species. Gel filtration on S-1000 (data not shown) indicated that the reoxidized product chromatographed with the same K_d value as reduced and alkylated subunits. Thus, under these conditions, the changes in mucin structure that accompany disulfide bond reaction appeared to be irreversible.

Trypsin Treatment of Subunits. The subunit peak from Figure 3 was pooled and treated with 0.075% w/w trypsin for 24 h at 37 °C followed by a second addition of the same amount of trypsin and another 24-h incubation. Gel filtration of an aliquot of the trypsin digest on Sephacryl S-1000 (data not shown) demonstrated that the major carbohydrate-containing peak eluted with a K_d value of 0.7. A second aliquot of the tryptic digest was placed on a column of Sephacryl S-500 and eluted with 5 M guanidine hydrochloride. Fractions were assayed for both carbohydrate and radioactivity. As shown in Figure 6, two peaks were eluted. The first, termed the glycosylated domains, contained carbohydrate but no radioactivity while the second contained no detectable carbohydrate but all of the radioactivity that had been present in the subunit fraction. The width of the glycosylated domain peak demonstrates that it is polydisperse. Fractions from the gel

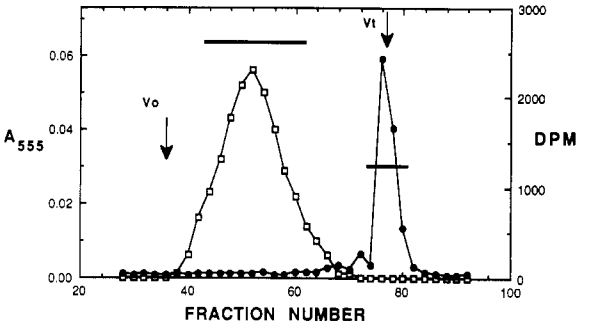

FIGURE 6: Gel filtration of trypsinized mucin subunits. The subunit fraction from Figure 3 was dialyzed and an aliquot treated with trypsin as described under Experimental Procedures. The digest was applied to a 1 × 115 cm column on Sephacryl S-500 and eluted with 5 M guanidine hydrochloride/10 mM phosphate, pH 7.0. Chromatographic conditions were as described in Figure 2. Fractions were assayed for carbohydrate (□) by using the periodic acid/Schiff assay (Mantle & Allen, 1978) and for radioactivity (●) by scintillation counting and pooled for further analysis as shown by the bars.

filtration column were pooled for further analysis as indicated by the bars in Figure 6. In a separate experiment, the PSM subunits were treated with 36% (w/w) trypsin and the products chromatographed on the S-500 column. The glycosylated domain peak from this experiment eluted at the same position as that shown in Figure 6, indicating that more stringent trypsinolysis does not give lower molecular weight products. Similar results were also obtained when intact mucin rather than mucin subunits was treated with trypsin.

The absence of radioactivity in the glycosylated domain fraction indicates that the cysteine residues reside in trypsin-sensitive sequences. Amino acid analysis of the glycosylated domains (Table II) confirms this and further demonstrates that several of the other minor amino acids of PSM are located within trypsin-sensitive sequences since the glycosylated domain fraction consists almost entirely of nine amino acids with only trace amounts of Asp, Cys, His, Tyr, Leu, Phe, and Lys. The amino acid composition of the glycosylated domains is, therefore, quite similar to that of the apomucin described by Eckhardt et al. (1987) and to the 81 amino acid repeat from the partial cDNA sequence for PSM reported by Timpte et al. (1988). The former study demonstrated that apoPSM contains only nine different amino acids with the minor amino acids all being present at less than 0.1 mol %. The data in Table II, however, demonstrate that these minor amino acids are reproducibly present both in the intact, purified mucin and in the reduced and alkylated subunits.

Light-scattering experiments on the glycosylated domains demonstrated an average molecular weight of 270K. This is very similar to the value of 275K calculated by Eckhardt et al. (1987) from the molecular weight of apomucin after adjusting for an assumed protein content of 35%. Thus, the peptide core of the glycosylated domains closely resembles the apomucin described by Eckhardt et al. (1987) in both size and composition. How these very different approaches generate what appears to be the same material is not entirely clear.

The second peak from S-500 gel filtration chromatography of the trypsin-treated mucin subunits appears to consist of small tryptic peptides and includes the minor amino acids missing from the glycosylated domains. This material appeared in the included volume when chromatographed on Sephadex G-10 (data not shown) and was not retained in 1000 molecular weight cutoff dialysis tubing. The [^{14}C]iodoacetamide label originally present in the mucin subunits was quantitatively recovered in this low molecular weight fraction after trypsinolysis. The number of peptides released, their

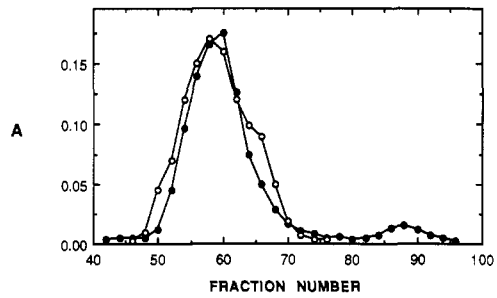

FIGURE 7: Gel filtration of Pronase-digested mucin subunits on Sephacryl S-200. Subunits prepared from pool 1 mucin by reduction and carboxymethylation were treated with Pronase (1:100 substrate:enzyme ratio) at 37 °C for 72 h. The products of digestion were separated on a 1 × 114 cm column of Sephacryl S-200 and eluted with 5 M guanidine hydrochloride/10 mM phosphate, pH 7.0. Fractions were monitored for the UV absorbance at 228 nm (●) and were also assayed for carbohydrate (○) as described in Figure 2.

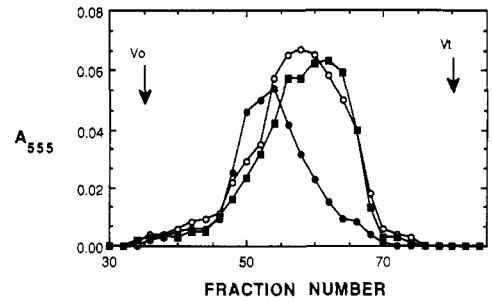

FIGURE 8: Chromatography of PSM subunits from pool 1 (●), pool 2 (○), and pool 3 (■) on Sephacryl S-1000. Subunits were prepared by reducing and carboxymethylating the three pooled mucin fractions from Figure 1. These were applied separately to a column of Sephacryl S-1000, eluted, and assayed for carbohydrate as described in Figure 2.

recoveries, and their sequences are under further investigation.

Pronase digestion of either the glycosylated domains or the reduced and alkylated PSM yields a single polydisperse peak on Sephacryl S-200 (Figure 7) with an average molecular weight of 17K as determined by static light scattering. On the basis of its size, this material may be related to the 81 amino acid repeat in the partial sequence of PSM reported by Timpte et al. (1988) which would have a calculated molecular weight of 23K.

PSM Polydispersity. Mucins are characteristically polydisperse; i.e., they display a range of molecular weights varying about a mean value (Carlstedt et al., 1985). OSM, for example, can be separated into fractions having average molecular weights ranging from 0.8×10^6 to 5.2×10^6 (Shogren et al., 1987) while cervical mucins display molecular weights ranging from 6×10^6 to 24×10^6 (Sheehan & Carlstedt, 1987). This phenomenon is rather puzzling since biosynthetic processes are normally geared to producing monodisperse products. Taking advantage of the fact that gel filtration on Sephacryl S-1000 can be used to characterize the size distributions of mucins and their subunits, the relationship between the size of the parent mucin and the size of subunits derived from it was investigated. Thus, aliquots of intact PSM from pools 1, 2, and 3 were separately reduced with DTT and alkylated with iodoacetamide, and the sizes of the resulting subunits were compared by gel filtration. The chromatographic patterns, shown in Figure 8, suggest that the larger mucins, on average, contain larger subunits. However, the subunit peaks are clearly polydisperse and show considerable overlap, suggesting that the correlation between the size of the mucin and the size of the subunits derived from it is relatively weak. Approximate molecular weight averages for the subunits ranged from 2×10^6 for pool 1 mucin to 1×10^6 for pool 3 mucin as judged by light-scattering or retention times on S-1000 (Table I) while the number of subunits per molecule varied from approximately 4.1 in pool 1 material to approximately 2.8 in mucin from pool 3. These results suggest that the origins of mucin polydispersity are complex; i.e., mucin polydispersity is not due solely to variability in the number of subunits present in a molecule or to the assembly of a fixed number of subunits of different sizes but rather to a combination of both factors. Qualitatively similar results were seen when $NaBH_4$ rather than DTT was used to reduce the mucin except that the K_d values were somewhat greater while increased tailing of the peaks to lower molecular weights was observed. This would be consistent with earlier studies indicating that $NaBH_4$ cleaves peptide linkages (Crestfield & Moore, 1963).

Discussion

PSM, like other mucins, is polydisperse, containing molecules of widely varying molecular weight. Since it appeared likely that polydispersity of the starting material would interfere with our ability to interpret structural studies, the mucin was first fractionated on Sephacryl S-1000 to give material with a narrower range of molecular weights. The mucin pool of highest molecular weight was chosen for further study. This material, which had an average molecular weight of 8.6×10^6, was reduced with DTT and alkylated with iodoacetamide, releasing mucin subunits of 2×10^6 Da together with a link protein fraction which gave two closely spaced bands of ca. 140 kDa on SDS gels. Treating either the mucin subunits or the intact mucin with trypsin produced glycosylated domains having an average molecular weight of 270K and a unique amino acid composition in which nine different amino acids comprise 99% of the total. Also released were tryptic peptides containing all of the cysteine residues and most of the other minor amino acids that are present in the intact mucin or in mucin subunits but which are essentially absent from the glycosylated domains.

The mucin subunit, then, appears to consist of a variable number of glycosylated domains together with poorly glycosylated or nonglycosylated regions that are susceptible to proteases, releasing peptides when treated with trypsin. Since the overall structure of mucins is linear, these elements must be assembled in an end to end fashion to form the mucin subunit. Although these data would be consistent with a repeating structure in which glycosylated domains alternate with a trypsin-sensitive sequence or sequences, other configurations are also possible. For example, the glycosylated domains could be separated by very short trypsin-sensitive sequences with longer peptide sequences present at the ends of the subunit. However, preliminary results on separating the tryptic peptides suggest the presence of only four major species, a result more consistent with a simple repeating structure.

The complete mucin molecule is a complex consisting of several subunits and one or more link proteins all interconnected through disulfide bonds. The subunits must be connected end to end via disulfide linkages, but it is not clear whether they are directly linked to one another or whether the link protein serves as a bridge between the subunits.

These findings on the structure of PSM parallel those reported by Carlstedt and Sheehan (1984a,b) on the degradation of cervical and gastrointestinal mucins in which reduction yielded subunits of 2×10^6 to 3×10^6 Da while trypsin treatment produced glycosylated "T-domains" of 400 kDa. Link proteins have been observed in gastrointestinal mucins

by Fahim et al. (1987), and by Mantle et al. (1981), and in tracheobronchial mucins by Tabachnik et al. (1981) and by Ringler et al. (1987). The molecular weights reported for the link proteins vary from approximately 65–70K (Ringler et al., 1987; Mantle et al., 1981) to 118K (Fahim et al., 1987).

A very different view of mucin structure comes from studies on enzymatically deglycosylated salivary mucins. Deglycosylation of sheep submaxillary mucin produced an apomucin with a reported molecular weight of 58 300 which contained only 11 different amino acids (Hill et al., 1977). Similar studies with PSM gave an apomucin with a molecular weight of 96 500 in which only nine different amino acids were present at levels greater than 0.1 mol % (Eckhardt et al., 1987). Assuming a carbohydrate content of 35%, this corresponds to a molecular weight of 275K for the fully glycosylated unit. These observations are consistent with a model of mucin structure (Eckhardt et al., 1987) in which individual 275-kDa units self-associate through carbohydrate–carbohydrate or carbohydrate–peptide interactions to produce high molecular weight aggregates. Other peptides containing the minor amino acids present in the intact mucin but absent from the apomucin are also believed to be noncovalently associated with the mucin.

In part, the results reported here are consistent with this model; the glycosylated domains obtained by trypsin treatment of the mucin subunits are strikingly similar in both size and composition to the apomucin produced by deglycosylation. The fact that these very different approaches are in agreement strongly argues that this structure is an important element in the structure of PSM. In addition, the fact that PSM shares elements of structural organization such as subunits and link proteins with mucins from other sources suggests that the results from cDNA studies on PSM (discussed below) may be true for other mucins. However, other aspects of the present studies are in disagreement with the model. Studies on intact mucins demonstrate that their high molecular weights are retained in the presence of 5 M Gdn·HCl (Carlstedt et al., 1985; Shogren et al., 1986), that fractions of different average molecular weights can be isolated, and that these fractions retain their characteristic sizes in the presence of denaturants (Sheehan & Carlstedt, 1987; Shogren et al., 1987), while SDS gel electrophoresis of purified mucin shows the absence of noncovalently linked peptides. Reduction and alkylation of disulfide linkages reduces mucin molecular weights, but the products are still much larger than the apomucin. Further, the fact that the glycosylated domains from PSM and equivalent proteolytic fragments from other mucins (Carlstedt et al., 1985) do not reassociate to yield high molecular weight materials would appear to eliminate the possibility that carbohydrate–carbohydrate interactions are responsible for the high molecular weights of mucins. In addition, the content of minor amino acids remains essentially unchanged after manipulations such as reduction and alkylation in the presence of denaturing agents (Table I). In our hands, these minor amino acids are only releasable by proteolytic cleavage. In sum, these observations are more consistent with a covalent than with a noncovalent structure for mucins. This is supported by data from studies on mucin biosynthesis in which nonglycosylated mucin precursors of greater than 400 kDa have been observed (Marianne et al., 1987). This size would be consistent with a glycosylated mucin subunit of greater than 1×10^6 Da. However, it should be emphasized that it is experimentally difficult to differentiate between covalent and very strong noncovalent linkages and the possibility that there are noncovalent linkages resistant to disruption by denaturants cannot be eliminated.

Recently, a partial cDNA sequence for PSM has been reported (Timpte et al., 1988). It consists of approximately 3.5 contiguous 81 amino acid tandem repeats terminated by an unrelated C-terminal sequence. The number of repeats in the complete sequence is unknown, and it is also not known whether a corresponding unique N-terminal sequence is also present. The amino acid composition of the repeat is virtually identical with that of the glycosylated domains isolated in this study and with that of the apomucin described by Eckhardt et al. (1987). All are composed of the same nine amino acids, and all lack the minor amino acids discussed above. the C-terminal sequence, on the other hand, contains significant quantities (1–4 mol %) of these minor constituents. Presumably, the glycosylated domains generated in this study lack the minor amino acids present in the C-terminal region because the latter has been cleaved by trypsin, but it is not clear why they are also absent from the apomucin. One possibility is that the apomucin is much larger than the reported 96.5 kDa so that the C-terminal region represents a proportionally smaller part of the total molecule. Alternatively, proteolytic cleavage may have occurred during the deglycosylation step since it would be expected that the apomucin, which is in a highly accessible random-coil conformation (Gerken, 1986; Eckhardt et al., 1987), would be much more susceptible to proteases than the native, globular protein used to assess the protease content of the glycosidases.

The similarity in the amino acid compositions of the glycosylated domains and the tandem repeat (Table II) suggests that they are closely related. Since it is known from the cDNA sequence that several tandem repeats occur consecutively, since it is believed that at least 8 of these repeats occur in the apomucin (Timpte et al., 1988), and since the average size of the glycosylated domains (270 kDa) would represent 11 or 12 tandem repeats, it is tempting to speculate that trypsin cleaves at sequences that separate stretches of contiguous tandem repeats. Thus, the size of the glycosylated domains would reflect the number of tandem repeats assembled consecutively. However, the tandem repeat contains an Arg–Ile bond that is theoretically susceptible to trypsin. Although most of these linkages must be protected from trypsin by glycosylation of neighboring residues, it is possible that glycosylation of specific Ser and Thr residues is random so that some of these Arg–Ile bonds may be susceptible to cleavage. If so, it follows that the measured sizes of the glycosylated domains would represent minimum values for the lengths of the tandem repeat regions.

PSM is highly polydisperse as shown by its elution profile from Sephacryl S-1000. This is due both to variability in the number of subunits per molecule and to variability in subunit size. Overall, subunit sizes range from approximately 0.5×10^6 to 2.5×10^6 Da as estimated from elution position on S-1000. PSM from pool 1 contains subunits that are approximately twice as large as subunits isolated from the other pools. Interestingly, this range in subunit molecular weight covers the range of sizes for mucin subunits reported by Carlstedt and Sheehan, 2×10^6 (Carlstedt et al., 1985), and by Allen (1983), who describes subunits of 500 kDa.

The glycosylated domains are also polydisperse with molecular weights as high as 0.8×10^6. It is not clear whether the apoPSM isolated by Eckhardt et al. (1987) is similarly polydisperse, but the gel filtration patterns shown for apoOSM (Hill et al., 1977) indicative equivalent or greater polydispersity. In PSM, at least, this is not due to genetic variation among individual animals since these experiments used mucin isolated from a single animal. This finding of polydispersity in the glycosylated domains was initially somewhat surprising

since they appear to represent a basic structural unit in the mucin molecule and it was expected that monodisperse products would be generated. However, several recent molecular biology studies (Swallow et al., 1987; Timpte et al., 1988) on mucins demonstrate repetitive sequences in mucin cDNAs and indicate that the number of repeats may be quite variable. Thus, the variability in size of the glycosylated domains may reflect variable numbers of the 81 amino acid tandem repeat. In any event, these results demonstrate that PSM in polydisperse at each level of structure from the intact mucin to mucin subunits and finally to the products of trypsin digestion.

ACKNOWLEDGMENTS

We thank Charles Blackwell for his excellent technical assistance and Drs. A. M. Jamieson and J. Blackwell for the use of their light-scattering facilities. In addition, we thank Dr. R. L. Shogren for informative discussions and Drs. J. E. Jentoft and T. A. Gerken for help in preparing the manuscript.

REFERENCES

Allen, A. (1983) *Trends Biochem. Sci. 8*, 59–62.
Bhaskar, K. R., & Creeth, J. M. (1974) *Biochem. J. 143*, 669–679.
Carlson, D. M. (1968) *J. Biol. Chem. 243*, 616–626.
Carlstedt, I., & Sheehan, J. K. (1984a) *Biochem. Soc. Trans. 12*, 615–617.
Carlstedt, I., & Sheehan, J. K. (1984b) *Ciba Found. Symp. 109*, 157–166.
Carlstedt, I., Sheehan, J. K., Corfield, A. P., & Gallagher, J. T. (1985) *Essays Biochem. 20*, 40–76.
Crestfield, A. M., & Moore, S. (1963) *J. Biol. Chem. 238*, 622–627.
Eckhardt, A. E., Timpte, C. S., Abernethy, J. L., & Hill, R. L. (1987) *J. Biol. Chem. 262*, 11339–11344.
Fahim, R. E. F., Specian, R. D., Forstner, G. G., & Forstner, J. F. (1987) *Biochem. J. 243*, 631–640.
Gerken, T. A. (1986) *Arch. Biochem. Biophys. 247*, 239–253.
Gupta, R., & Jentoft, N. (1987) Annual Meeting of the Society for Complex Carbohydrates, Bethesda, MD, Nov, 5–7, Abstr. 49.
Gupta, R., & Jentoft, N. (1989) *J. Chromatogr.* (in press).
Heinrikson, R. L., & Meredith, S. C. (1984) *Anal. Biochem. 136*, 65–74.
Henderson, L. E., Oroszlan, S. E., & Konigsberg, W. (1979) *Anal. Biochem. 93*, 153–157.
Hill, H. D., Reynolds, J. A., & Hill, R. L. (1977) *J. Biol. Chem. 252*, 3791–3795.
Hunkapillar, M. W., Ostrander, E. F., & Hood, L. E. (1983) *Methods Enzymol. 91*, 227–236.
Jentoft, N. (1985) *Anal. Biochem. 185*, 424–433.
Jentoft, N., & Dearborn, D. G. (1979) *J. Biol. Chem. 254*, 4359–4365.
Konigsberg, W. (1972) *Methods Enzymol. 25*, 185–188.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
Mantle, M., & Allen, A. (1978) *Biochem. Soc. Trans. 6*, 607–609.
Mantle, M., Mantle, D., & Allen, A. (1981) *Biochem. J. 195*, 277–285.
Marianne, T., Perini, J. M., Lafitte, J. J., Houdret, N., Pruvot, F. R., Lamblin, G., Slayter, H. S., & Roussel, P. (1987) *Biochem. J. 248*, 189–195.
Meyer, F. A. (1983) *Biochem. J. 215*, 701–704.
Neutra, M. R., & Forstner, J. F. (1987) in *Physiology of the Gastrointestinal Tract* (Johnson, L. R., Ed.) pp 975–1009, Raven Press, New York.
Oike, Y., Kimata, K., Shinomura, T., & Suzuki, S. (1980) *Biochem. J. 191*, 193.
Ringler, N., Selvakumar, R., Woodward, H. D., Simet, I. M., Bhavanandan, V. P., & Davidson, E. A. (1987) *Biochemistry 26*, 5322–5328.
Ristow, S., & Wetlaufer, D. B. (1973) *Biochem. Biophys. Res. Commun. 50*, 544–550.
Sheehan, J. K., & Carlstedt, I. (1987) *Biochem. J. 245*, 757–762.
Sheehan, J. K., Oates, K., & Carlstedt, I. (1987) *Biochem. J. 239*, 147–153.
Shogren, R. L., Jamieson, A. M., Blackwell, J., Cheng, P. W., Dearborn, D. G., & Boat, T. F. (1983) *Biopolymers 22*, 1657–1675.
Shogren, R. L., Jamieson, A. M., Blackwell, J., & Jentoft, N. (1984) *J. Biol. Chem. 259*, 14657–14662.
Shogren, R. L., Jamieson, A. M., Blackwell, J., & Jentoft, N. (1986) *Biopolymers 25*, 1505–1517.
Shogren, R. L., Jentoft, N., Jamieson, A. M., Gerken, T. A., & Blackwell, J. (1987) *Carbohydr. Res. 160*, 317–327.
Swallow, D. M., Gendler, S., Griffiths, B., Taylor-Papadimitrou, J., & Bramwell, M. E. (1987) *Nature 328*, 82–84.
Tabachnik, N. F., Blackburn, P., & Cerami, A. (1981) *J. Biol. Chem. 256*, 7161–7165.
Thannhauser, T. W., Konishi, Y., & Scheraga, H. A. (1984) *Anal. Biochem. 138*, 181–188.
Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., & Hill, R. L. (1988) *J. Biol. Chem. 263*, 1081–1088.
Zahler, W. L., & Cleland, W. W. (1968) *J. Biol. Chem. 243*, 716–719.