

Isolation and physical characterization of the MUC7 (MG2) mucin from saliva: evidence for self-association

Ravi MEHROTRA, David J. THORNTON and John K. SHEEHAN¹

Wellcome Trust Centre for Cell-Matrix Research, Division of Biochemistry, 2.205 School of Biological Sciences, University of Manchester, Manchester M13 9PT, U.K.

Saliva contains two major families of mucins (MG1 and MG2); the polypeptide of the smaller of these glycoproteins (MG2) has been assigned as the product of the *MUC7* gene. In this study we have devised a rapid two-step procedure that recovers this glycoprotein essentially free of other components and in sufficient quantity to enable physical and self-interaction studies. Raw saliva was solubilized in 4 M guanidinium chloride and thereafter subjected to Sepharose CL-4B chromatography. The MG2-rich fraction was recovered free from the larger MG1 glycoproteins and also smaller proteins/glycoproteins (molecular mass less than 100 kDa). MG2 glycoproteins were finally purified by anion-exchange chromatography on Mono Q. The purity of the preparation was assessed by SDS/PAGE after radiolabelling of the molecules with [¹⁴C]acetic anhydride. Peptide mapping, N-terminal sequencing and amino acid analysis verified the polypeptide of the mucins as the *MUC7* gene product. The isolated molecules were examined by electron microscopy and appeared

as short flexible worm-like structures 30–120 nm in length. The distribution was heterogeneous, containing a major component with number-average and weight-average lengths of 52 and 55 nm respectively and a minor component with number-average and weight-average lengths of 94 and 98 nm respectively. We propose that the two differently sized populations represent monomeric and dimeric species of the mucins. Gel chromatography performed in 0.2 M NaCl indicated the presence of monomers, dimers and tetramers; an average molecular mass for the preparation was 192 kDa. However, in 4 M guanidinium chloride the molecular mass was 158 kDa and a similar molecular mass (155 kDa) was determined for the mucin preparation after reduction. These results suggest that the mucins might self-associate via a protein-mediated interaction. On the basis of the results a model is proposed for the self-association of the MUC7 mucin, which might be important for its biological function.

INTRODUCTION

Salivary mucins are produced by various (sero) mucous salivary glands, i.e. submandibular, sublingual and palatal glands and the minor salivary glands in the lip, cheek and tongue [1]. Previous studies have identified two physically distinct populations of mucins in whole saliva from humans and these have been termed mucin-glycoprotein 1 (MG1) and mucin-glycoprotein 2 (MG2) [2–4]. MG2 is a low-molecular-mass mucin population secreted by the submandibular gland and to a smaller extent the sublingual salivary glands [5]. Currently it is unique among characterized secretory mucins because it is not located at the cell surface and does not form disulphide-linked oligomers. The reported molecular mass of between 125 and 250 kDa [6,7] is the smallest determined so far for a secretory mucin. Ramasubbu et al. [4] have suggested that two isoforms of MG2, designated MG2a and MG2b, are found in submandibular/sublingual saliva. The amino acid compositions of both forms were identical [8] but they differed primarily in their sialic acid and fucose contents, with MG2a containing 14% sialic acid/16% fucose and MG2b containing 27% sialic acid and 7% fucose [4].

The gene encoding the MG2 polypeptide has been cloned and is designated *MUC7* [9]. The cDNA sequence of *MUC7* encodes a 39 kDa protein of 377 amino acid residues. Six almost perfect tandem repeats (TRs), each comprising 23 residues (9 of which are serine/threonine), make up the central region of the protein

and in the mature mucin this region would carry the bulk of the O-glycans. The density of potential O-glycosylation sites decreases in the amino and carboxy regions flanking the tandem repeats. Two cysteine residues are present towards the N-terminus, the first 20 residues of which are hydrophobic and are postulated to be the leader peptide of apo-MUC7 (MG2) [9].

MG2 mucins have been demonstrated to be involved with bacterial binding [10]; however, the small amounts and low quality of purified MG2 mucins have hampered these studies. Thus the primary aim of this work was to devise a protocol for the large-scale purification of the unreduced form of MG2 from human whole saliva and to verify it as the product of the *MUC7* gene. In addition we present a physical description of the molecule in solution that might provide clues as to its interaction with bacteria and with other components of saliva.

EXPERIMENTAL

Materials

[¹⁴C]Acetic anhydride (500 μ Ci in a P1 break-seal ampoule) and the enhanced chemiluminescence (ECL) detection kit were from Amersham (Little Chalfont, Bucks., U.K.), benzyldimethylalkylammonium chloride was from BDH (Poole, Dorset, U.K.) and modified trypsin was purchased from Promega (Madison, WI, U.S.A.). α -Cyano-4-hydroxycinnamic acid, substance P and insulin (bovine pancreas) were from Sigma Chemical Co. (Poole,

Abbreviations used: MG1, high-molecular-mass salivary mucins; MG2, low-molecular-mass salivary mucins; GuHCl, guanidinium chloride; MALDI-TOF-MS, matrix-assisted laser desorption ionization–time-of-flight MS; PEG, poly(ethylene glycol); PAS, periodic acid/Schiff; TFA, trifluoroacetic acid.

¹ To whom correspondence should be addressed (e-mail John.Sheehan@man.ac.uk).

Dorset, U.K.). Trifluoroacetic acid (sequencing grade) was from Applied Biosystems (Warrington, Cheshire, U.K.) and acetonitrile (HPLC grade) was purchased from Rathburn Chemicals Ltd. (Walkerburn, Borders Region, Scotland, U.K.).

Extraction and purification of MG2

Step 1. Saliva was collected from an individual healthy male donor who had refrained from eating, drinking, smoking and oral hygiene for 2 h before collection. Saliva was stimulated by chewing on Parafilm, and collected into an equal volume of 8 M guanidinium chloride (GuHCl) containing 0.1% CHAPS. The solution was mixed for 2 h and was subsequently centrifuged at 4400 g_{av} for 10 min at 4 °C. The supernatant was decanted and stored at 4 °C until processed further.

Step 2. Aliquots of the supernatant (50 ml) were chromatographed on a column of Sepharose CL-4B (68 cm \times 5 cm) eluted with 4 M GuHCl containing 0.05% CHAPS at a flow rate of 48 ml/h and at room temperature. Fractions were pooled as shown in Figure 1.

Step 3. The pooled fractions were dialysed against 6 M urea saturated with poly(ethylene glycol) (PEG) until a decrease in volume to approximately one-tenth had been reached.

Step 4. The solution was then dialysed three times against 1 litre of 6 M urea and finally against 1 litre of 6 M urea containing 0.05% CHAPS (buffer A). The MG2-containing pool in buffer A was then chromatographed on a Pharmacia Mono Q HR 10/10 column. The column was eluted at a flow rate of 2 ml/min with the following gradient program: an initial linear gradient from 0% to 35% buffer B [0.4 M lithium perchlorate/6 M urea/10 mM piperazine (pH 5) containing 0.05% CHAPS] over 28 min, then isocratic elution at 35% buffer B for 10 min, followed by a linear gradient from 35% to 55% buffer B over 27 min and finally a linear gradient from 55% to 100% buffer B over 5 min. The purified MG2 preparation was concentrated by rechromatography on the Mono Q HR 10/10 column eluted at a flow rate of 1 ml/min with 100% buffer B. Stock MG2 preparations were dialysed against 4 M GuHCl and stored at 4 °C.

Acetylation of MG2 with [14 C]acetic anhydride

MG2 preparations (2.5 mg/ml) in 4 M GuHCl were dialysed against disodium tetraborate buffer, pH 9, containing 0.2 M CaCl_2 at 4 °C. The suspension of MG2 (7 ml) was removed from the dialysis tubing and placed in the top half of a P1 break-seal ampoule. The seal was broken, allowing the [14 C]acetic anhydride (500 μCi) to react with the mucin solution. After 60 min the mucin preparation was dialysed against 4 M GuHCl at 4 °C and stock ^{14}C -labelled MG2 solutions (specific radioactivity 4.2×10^6 d.p.m./mg) were stored at 4 °C.

SDS/PAGE

Samples were prepared in 0.0625 M Tris/HCl/0.5% SDS; before electrophoresis they were reduced with 1% (v/v) 2-mercaptoethanol at 95 °C for 5 min. Electrophoresis was performed with a discontinuous gel system in a 7.5% (w/v) resolving gel with a 4% (w/v) stacking gel. Gels were either analysed with Brilliant Blue G-colloidal staining or dried and radioactivity quantified with a Fujix Bas 2000 Bioimager.

Electron microscopy

Electron microscopy was performed as described previously [11]. In brief, carbon films were made by evaporation of carbon onto

freshly cleaved mica. The films were floated off on water and deposited on 600-mesh grids. Specimens were prepared by spreading molecules (1 $\mu\text{g}/\text{ml}$) in a benzyltrimethylammonium chloride monolayer on a hypophase of 50 mM magnesium acetate. The molecules were picked up on carbon-coated grids and subsequently stained with uranyl acetate and rotary-shadowed with platinum/tungsten. Molecules were examined with a JEOL 1200 electron microscope at 120 keV. Contour lengths on photographically enlarged electron micrographs (final magnification $\times 150000$) were measured with a Planix digital planimeter (Hall and Watts, London, U.K.).

Preparation and analysis of mucin peptides and glycopeptides

Mucin glycopeptides and low-molecular-mass peptides were prepared with the strategy previously described by Thornton et al. [12]. An aliquot of the MG2 stock solution (approx. 1 mg of mucin) was dialysed into 1 ml of 0.1 M ammonium hydrogen carbonate, pH 8.0, and incubated with 0.5 μg of trypsin for 24 h at 37 °C. The digestion products were freeze-dried and subsequently chromatographed on a Pharmacia Superose 12 PC 3.2/30 column eluted with 0.1 M ammonium hydrogen carbonate, pH 8.0, at a flow rate of 0.2 ml/min. Fractions from the column were taken into two pools (see Figure 5), a glycopeptide-containing pool that was subjected to amino acid analysis and a low-molecular-mass peptide pool that was subjected to reverse-phase chromatography.

Reverse-phase HPLC of MG2 peptides

The MG2-peptide pool was dried and dissolved in 500 μl of 0.1% (v/v) trifluoroacetic acid (TFA). The resulting solution was chromatographed on a $\mu\text{RPC C}_2/\text{C}_{18}$ PC 3.2/3 column with the Pharmacia Smart system at a flow rate of 0.25 ml/min. The column elution programme was 0.1% (v/v) TFA (5 min) followed by a linear gradient of 0–70% (v/v) acetonitrile in 0.1% (v/v) TFA (25 min) and finally a second linear gradient of 70–100% (v/v) acetonitrile in 0.1% (v/v) TFA (5 min). Major peaks were analysed by matrix-assisted laser desorption/ionization–time-of-flight MS (MALDI–TOF–MS) and peptides were purified to homogeneity by rechromatography with shallower gradients centred on their elution point. Selected fractions were subjected to N-terminal sequencing.

MS

MS was performed by MALDI–TOF in a Micromass TOFSpec-E. Samples (1 μl) in 0.1% (v/v) TFA containing various proportions of acetonitrile were mixed with an equal volume of 50 mM α -cyano-4-hydroxycinnamic acid and applied to a TOFSpec target. Samples were analysed in positive-ion mode with substance P (M^+ 1348.7) and bovine insulin (M^+ 5734.5) as internal standards. The data generated were processed with the OPUS[®] peak-detection program.

N-terminal sequencing

N-terminal amino acid sequencing was performed on isolated peptides with an Applied Biosystems 476A protein microsequencer.

Amino acid analysis

Samples were subjected to aqueous-phase acid hydrolysis (6 M HCl at 105 °C for 16 h). The resulting amino acids were derivatized with phenyl isothiocyanate and the derivatives were

separated by reverse-phase HPLC with a 3M ODS2 column with gradient elution (0–50% acetonitrile in 12 mM sodium phosphate buffer, pH 6.4).

Light scattering

Molecular mass determination of MG2 was accomplished as described previously [12]. In brief, preparations were chromatographed on a Pharmacia Superose 6 HR 10/30 column eluted with 0.2 M NaCl or 4 M GuHCl, at a flow rate of 0.5 ml/min. The column effluent was passed through an in-line Dawn® DSP laser photometer and a Wyatt/Optilab 903 interferometric refractometer (Optichem, Mold, Clywd, U.K.) to measure light scattering and sample concentration respectively. Light-scattering measurements were taken continuously at 18 angles between 15° and 151° and the data were analysed by the method of Zimm [13] with proprietary software.

Analytical methods

Slot-blotting and periodic acid/Schiff (PAS) staining were performed as described previously [14]. Sialic acid was measured by an automated procedure [15].

RESULTS

Mucin (MG2) purification

Saliva was extracted with 4 M GuHCl/0.1% CHAPS, insoluble material was removed by centrifugation, and the supernatant was fractionated by Sepharose CL-4B chromatography (Figure 1). Two carbohydrate-rich populations of macromolecules were observed as monitored with the PAS reagent and these were deemed to be MG1 (void volume) and MG2 (included volume) respectively. The MG2-containing peak had a higher sialic acid content than the MG1-containing peak. The column eluate was

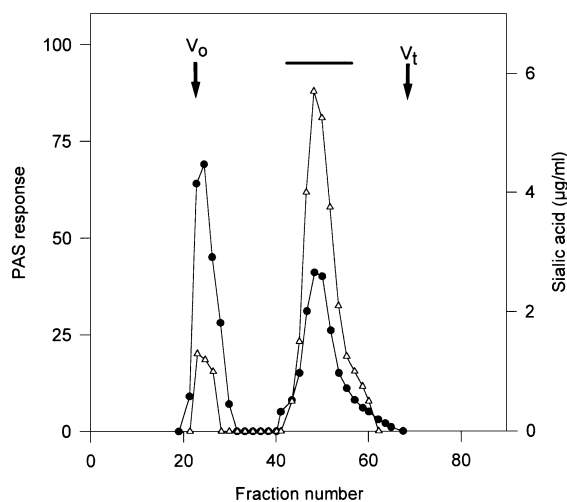


Figure 1 Preparative Sepharose CL-4B gel chromatography of human saliva extract

Clarified human saliva extract (50 ml) in 4 M GuHCl containing 0.05% CHAPS was chromatographed, at a flow rate of 48 ml/h, on a column of Sepharose CL-4B eluted with the same solvent. Fractions (16 ml) were analysed for carbohydrate with the PAS reagent (●) and sialic acid (△). MG2-containing fractions were pooled as indicated by the horizontal bar. The void and totally included volumes were determined with Dextran Blue and sodium dichromate respectively.

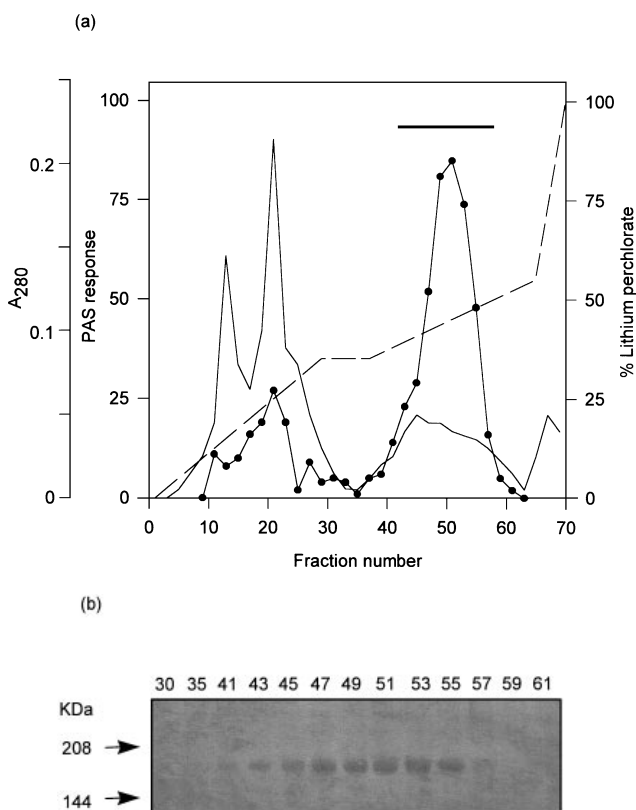


Figure 2 Preparative anion-exchange chromatography of the MG2 mucin-containing pool

(a) The included volume pool from Sepharose CL-4B was concentrated by dialysis against PEG in 6 M urea containing 0.05% CHAPS and chromatographed on a Mono Q HR 10/10 column as described in the Experimental section. Fractions (2 ml) were assayed for carbohydrate with the PAS reagent (●) and proteins were determined with A_{280} measurements (solid line). The nominal gradient is indicated by the broken line. (b) Fractions from across the distribution were analysed by SDS/PAGE [7.5% (w/v) gel] under reducing conditions and molecules were revealed with Brilliant Blue G-colloidal staining. The final purified MG2 preparation was pooled as indicated in (a) by the horizontal bar.

also monitored at A_{280} (results not shown) and most of the absorbance was found between fractions 60 and 70. The MG2-containing fractions were pooled as indicated (Figure 1), dialysed against PEG in 6 M urea containing 0.5% CHAPS to decrease the volume to approximately one-tenth and then subjected to preparative anion-exchange chromatography on Mono Q (Figure 2a). Two groups of molecules were observed a lower-charged population with a substantial A_{280} and a low PAS response, and a more highly charged population with a higher PAS reactivity. Aliquots of fractions across the chromatogram were subjected to SDS/PAGE and the gel was stained with colloidal Brilliant Blue G (Figure 2b). Fractions 35–57 contained a component with an apparent molecular mass of 176 kDa; no other components were seen. Fractions were pooled as indicated (Figure 2a) and the MG2 preparation was concentrated by rechromatography on Mono Q by isocratic elution (results not shown).

To assess the efficiency of the purification protocol an aliquot of MG2 preparation was radiolabelled with [14 C]acetic anhydride. Aliquots of the labelled and unlabelled preparation were subjected to SDS/PAGE and the gel was analysed by Brilliant Blue G-colloidal staining followed by autoradiography (Figure 3). A

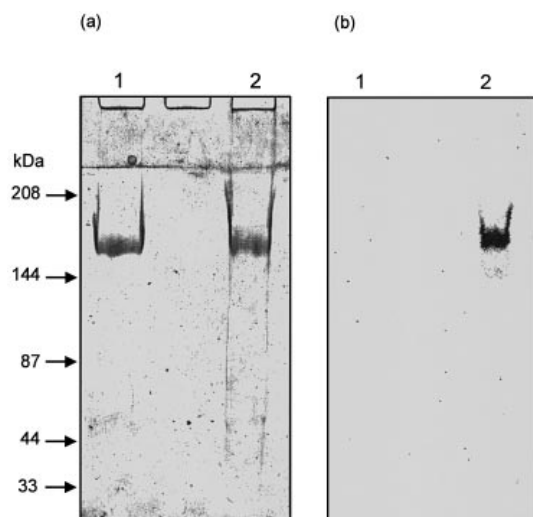


Figure 3 SDS/PAGE of MG2 and ^{14}C -labelled MG2 preparations

Unlabelled MG2 (5 μg) (lane 1) and ^{14}C -labelled MG2 (5 μg) (lane 2) were subjected to SDS/PAGE [7.5% (w/v) gel] under reducing conditions. Molecules were detected by Brilliant Blue G-250 staining (a) and by autoradiography of the dried gel with a Fujix Bas 2000 Bioimager (5 h exposure) (b).

Table 1 Yield of the MG2 mucin during purification

^{14}C -labelled MG2 (10 μg) (specific radioactivity 4.2×10^6 d.p.m./mg) was added to 50 ml of saliva before purification. During various steps in the purification protocol the radioactivity was determined by liquid-scintillation counting. For details of each step see the description in the Experimental section. For step 5, the MG2 preparation from step 4 was dialysed extensively against water and freeze-dried. The purification protocol described by Ramasubbu et al. [4] employed dialysis against water before freeze-drying between steps. Our results indicate that this procedure was responsible for the greatest loss of material, probably through surface binding to the dialysis tubing, and should be avoided.

Step	^{14}C -labelled MG2 recovered (%)
1 (extraction and low-speed centrifugation)	100
2 (gel-filtration chromatography)	79
3 (concentration)	65
4 (anion-exchange chromatography)	43
5	26

single band was found by both methods, indicating a high level of purity with no other detectable proteins.

To assess the efficiency of the purification protocol and to estimate the concentration of MG2 in raw saliva, an aliquot of the purified ^{14}C -labelled material was added to a saliva extract. The radioactivity was determined at each stage of the purification and the results are presented in Table 1. Extensive dialysis of the final preparation against water and subsequent freeze-drying yielded 1.9 mg of mucin. Therefore, assuming that this 1.9 mg contained pure MG2, one can calculate that the original 50 ml of unfractionated saliva contained 7.3 mg of MG2 (146 $\mu\text{g}/\text{ml}$).

Identification of MG2 as the MUC7 gene product

Confirmation of the polypeptide of this glycoprotein as the product of the *MUC7* gene was sought by tryptic peptide mass analysis. The reduced MG2 preparation was subjected to digestion with trypsin; an aliquot of the fragments generated was

analysed by MALDI-TOF-MS. A complex spectrum was observed and only four of the peptides (M^+ 814, 875, 1027 and 2385 Da) had masses corresponding to predicted tryptic peptides from the deduced amino acid sequence of the MUC7 polypeptide (Figure 4).

As another approach to mucin identification, amino acid analysis and N-terminal sequencing of peptides were performed. An aliquot of the fragments produced by treatment with trypsin was fractionated by gel chromatography on Superose 12 (Figure 5). After chromatography, material was taken into two main pools, TR-I and TR-II. Fraction TR-I contained fragments resulting from the major glycosylated portion of the mucin polypeptide and these were taken for amino acid analysis. MALDI-TOF-MS indicated that fraction TR-II contained the lower-molecular-mass peptide and glycosylated peptides (results not shown); these were fractionated and subjected to N-terminal sequencing.

Amino acid analysis

The mucin glycopeptide-containing fraction (TR-I) was subjected to amino acid compositional analysis: the data are presented in Table 2. Comparison with the amino acid composition of the major glycopeptide inferred from a theoretical trypsin digestion of the MUC7 polypeptide (i.e. the 19353 Da peptide illustrated in Figure 4) demonstrates good agreement with the predicted values. The only significant discrepancy was a smaller amount of threonine than the predicted value; this might be due to the level of oligosaccharide substitution of this residue.

Peptide analysis

Tryptic peptides in TR-II were fractionated by reverse-phase chromatography (Figure 6) and major peaks were analysed by MALDI-TOF-MS (results not shown). A major peptide (M^+ 2382 Da) was present in fraction 18; this was purified to homogeneity by rechromatography (results not shown). Automated Edman degradation of this peptide yielded the sequence SHFELPHYPGLLAHQK. This sequence is identical with a 16-residue sequence within the MUC7 polypeptide that has a theoretical mass of 2388 Da (Figure 4). A search of the protein sequence revealed no identity with other protein sequences.

Physical characterization of MUC7 (MG2) mucin

Electron microscopy

The mucins were seen under electron microscopy as small flexible worm-like structures (Figure 7a). Contour length measurements (Figure 7b) showed molecules to be between 30 and 120 nm in length. The distribution indicates the presence of two components, the major one with a number-average length of 52 nm and a weight-average length of 55 nm, and a minor component with a number-average length of 94 nm and a weight-average length of 98 nm.

Determination of molecular mass

The size and homogeneity of the mucin preparation were analysed by light scattering and refractive index measurements after gel chromatography on Superose 6. An example of the results, in this instance after chromatography of 200 μg of the unreduced MUC7 (MG2) mucin preparation, is shown in Figure 8(a). The refractive index measurement profile demonstrated that the mucin preparation was eluted as a major included peak with shoulders extending towards the void volume of the column. The light-scattering profile (results not shown) mirrored that of the

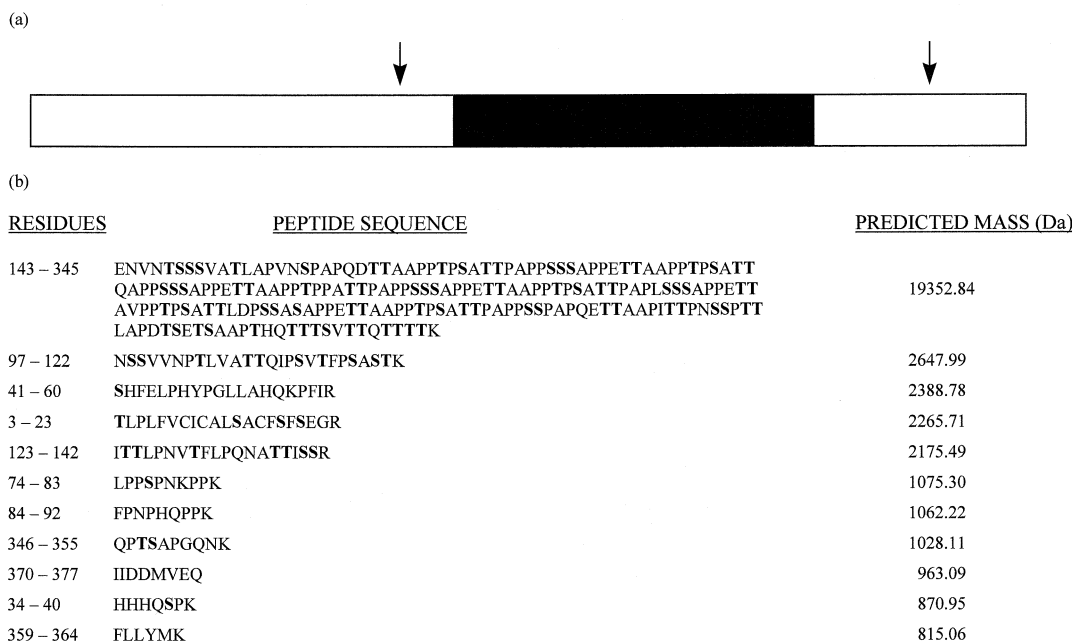


Figure 4 Theoretical trypsin digestion of the MUC7 polypeptide

(a) Diagrammatic representation of the MUC7 polypeptide based on the sequence data presented by Bobek et al. [9]. The filled region represents the central tandem repeat region and the arrows indicate trypsin cleavage sites flanking this region that would generate a putative glycopeptide in the mature mucin. (b) The peptides and their molecular masses (above 750 Da) generated from a theoretical trypsin digest of the MUC7 polypeptide. Possible O-glycosylation sites (serine and threonine residues) are in bold. The 19351 Da peptide represents the protein backbone of the putative glycopeptide shown in (a).

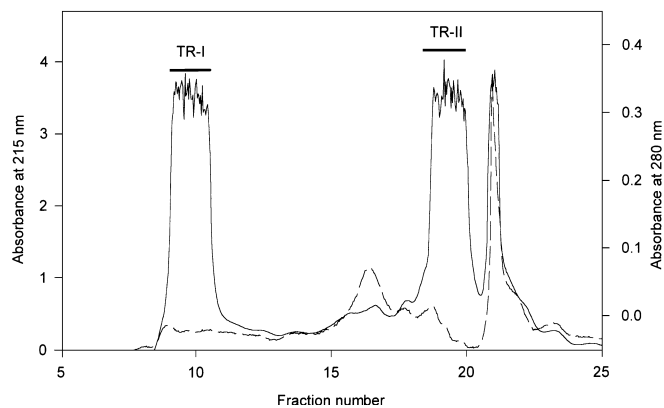


Figure 5 Superose 12 chromatography of trypsin digestion products of reduced MG2

The reduced MG2 preparation was treated with trypsin in with 0.1 M ammonium hydrogen carbonate, pH 8.0, and subsequently chromatographed on a Superose 12 PC 3.2/30 column. The eluent was monitored for absorbance at 215 nm (solid line) and 280 nm (broken line) and the fractions from the column were pooled as follows: TR-I, 8–11; TR-II, 15–20. These poolings are indicated by the horizontal bars.

refractive index; the average molecular mass determined across the chromatographic distribution was 198 kDa (Figure 8a). The chromatographic heterogeneity is reflected in the molecular mass measurements: there is evidence of components of molecular masses 600, 302 and 172 kDa. No measurable value for the radius of gyration could be obtained across the entire molecular

Table 2 Amino acid analysis of MG2 glycopeptides (TR-I)

Abbreviation: n.d., not determined. The predicted amino acid compositions are calculated from a theoretical tryptic peptide fragment of the deduced sequence of MUC7 mucin containing the tandem repeat region shown in Figure 4. Figures are expressed per 100 residues.

Amino acid	Content (%)	
	Observed	Predicted
Pro	30	26
Ala	17	18
Thr	14	24
Ser	13	15
Glx	7	6
Asx	5	4
Val	4	3
Leu	3	1
Ile	2	1
Arg	1	0
Gly	1	0
Lys	1	1
Phe	1	0
Met	0	0
Tyr	0	0
Trp	n.d.	0
Met	n.d.	0
Cys	n.d.	0
His	n.d.	1

distribution (results not shown). On reduction, the preparation chromatographed as a single sharp peak as monitored by refractive index (Figure 8b) and light scattering (results not

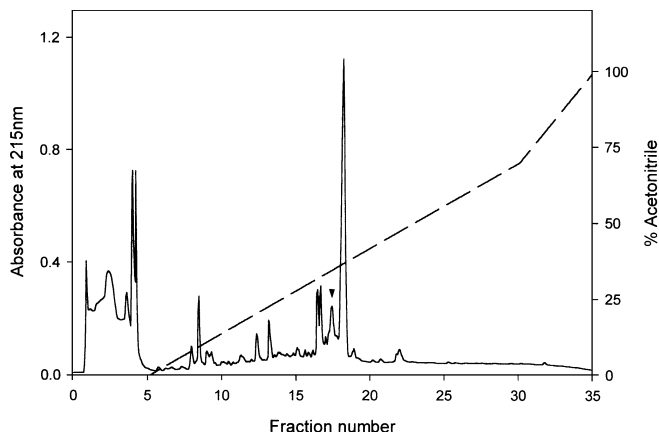
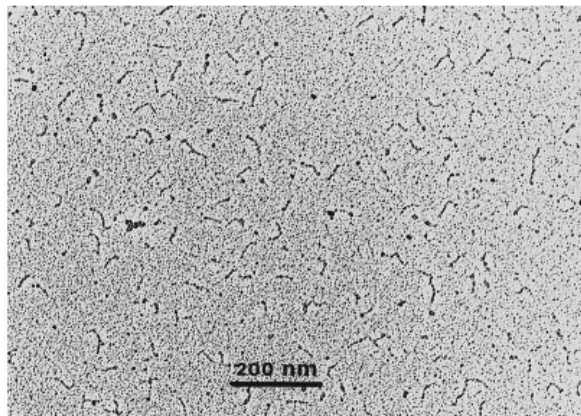


Figure 6 Reverse-phase chromatography of tryptic peptide fraction TR-II

Fraction TR-II was chromatographed on a μ RPC C₂/C₁₈ PC 3.2/3 column. The eluent was monitored for absorbance at 215 nm (solid line); the nominal gradient is indicated by the broken line. Fractions were analysed by MALDI-TOF-MS; a major peptide in fraction 18 (arrowhead) was purified by rechromatography.

(a)



(b)

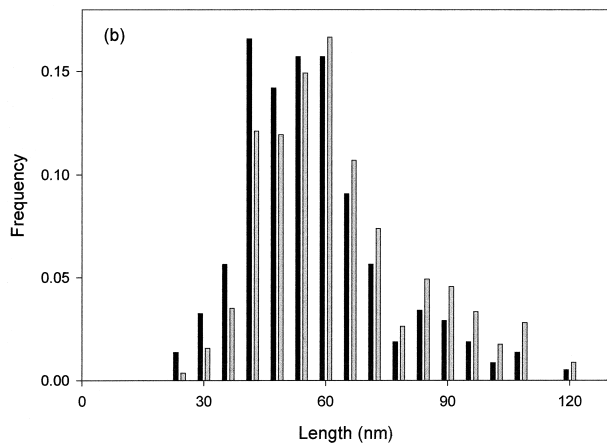


Figure 7 Electron microscopy of the MUC7 (MG2) mucin

(a) The molecules were spread and shadowed as described in the text. (b) Histograms of the contour length measurements of the mucins. A total of 150 molecules were measured and the frequencies of the number distributions (black bars) and weight distributions (grey bars) are shown.

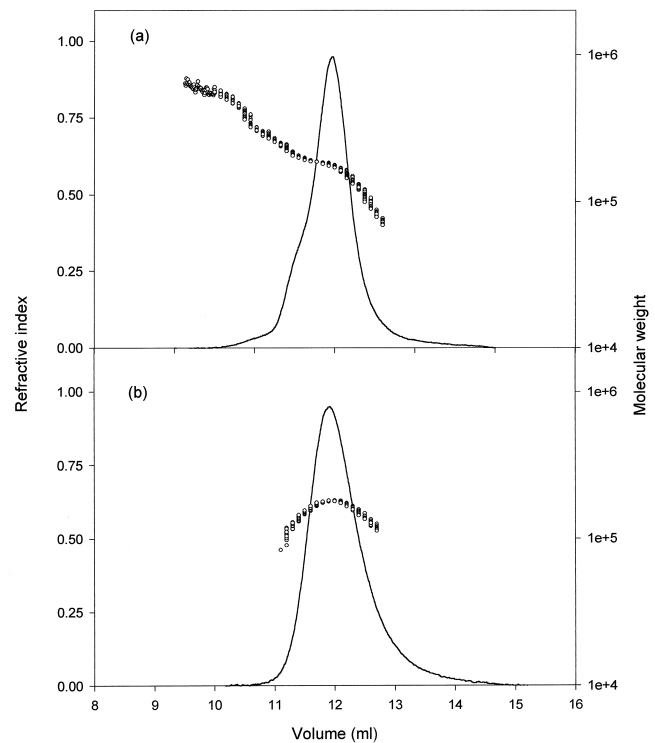


Figure 8 Superose 6 chromatography of the MUC7 (MG2) mucin

Unreduced mucins (200 μ g) (a) and reduced mucins (85 μ g) (b) were chromatographed on a Superose 6 HR 10/30 column eluted with PBS at a flow rate of 0.5 ml/min. The column eluate was monitored on-line for light scattering to determine molecular mass (\circ) and the refractive index was measured to determine sample concentration (solid line). In (a) molecular mass measurements plotted across the distribution indicate the presence of monomeric, dimeric and tetrameric mucin species, while in (b) only monomers were observed. The void and totally included volumes of the column were 7.8 and 19 ml respectively.

shown) and yielded an average molecular mass for the mucins of 155 kDa. A similar profile was observed after chromatography of the unreduced mucins in 4 M GuHCl and the average molecular mass was determined as 158 kDa (results not shown).

DISCUSSION

The initial objective of this work was to establish a simple efficient protocol for the large-scale isolation of undegraded pure MG2 that could be employed for bacterial binding studies. Previous studies [3,4,6] have isolated this molecule to homogeneity but their methods were more complex and required reduction. The application of chaotropic solvents such as GuHCl for the dispersion of the mucus gel has been demonstrated to be of value in a number of cases including saliva and was the approach adopted here. Initial fractionation of solubilized saliva by gel-filtration chromatography separated the larger MG1 mucins from the MG2 population. In addition the MG2 mucins were separated from the bulk of smaller proteins. This procedure had advantages over the approach presented by Ramasubbu et al. [4], which used Sepharose CL-2B gel chromatography resulting in the co-elution of MG2 and the low-molecular-mass components. Therefore the second gel-filtration step on Sephacryl S-300 described by these authors was not required in the present study.

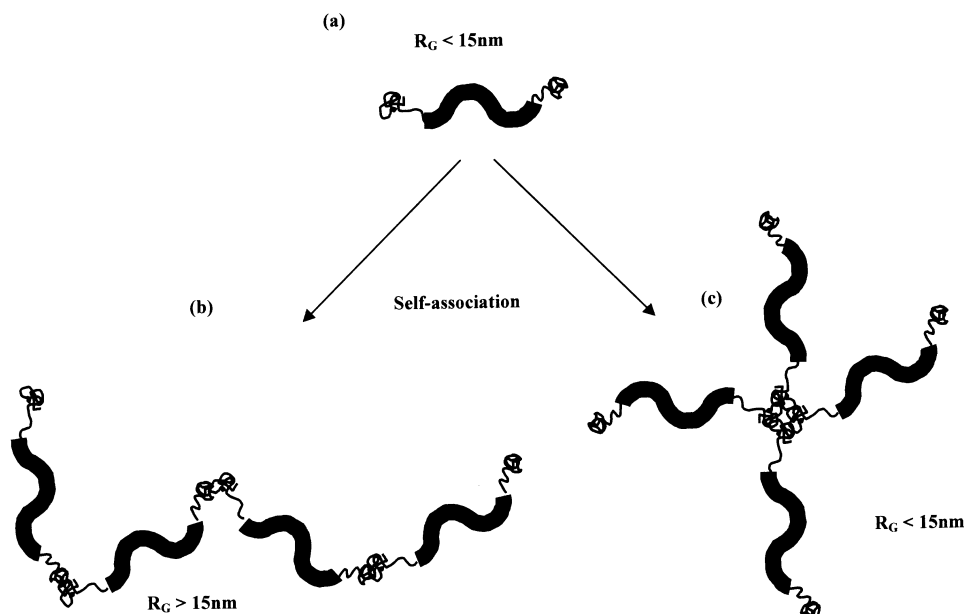


Figure 9 Model for the shape and self-interaction of the MUC7 (MG2) mucin in solution

Electron microscopy and light scattering results suggest that the molecules are random coil-like structures with a radius of gyration (R_G) below 15 nm. This value is the lowest detectable by the light-scattering technique employed and indeed there was no measurable value for R_G across the entire distribution of oligomers. These results suggest that the mucins most probably associate at only one end because an end-to-end association would yield a measurable value for R_G of the species above dimers. Reduction seems to abolish self-interaction and suggests that the N-terminus of the molecule (the site of the two cysteine residues) might be involved. The diagram shows two proposed models for the self-association of the MUC7 (MG2) mucin. The monomeric mucin (a) can either form end-to-end associations (b) or clusters via interaction at a single end (c). The results presented favour self-association as indicated in (c).

The critical step in the isolation procedure presented here was anion-exchange chromatography and in particular the use of lithium perchlorate as the eluting salt. When NaCl was used the MG2 was inseparable from another component of molecular mass 54 kDa (results not shown). This contaminating component might have been amylase, as deduced from its molecular mass on SDS/PAGE [4]. Lithium perchlorate is better than NaCl because it is a chaotropic agent and will therefore aid the dissociation of hydrophobically bound components from MG2 and the column matrix. The non-specific binding of MG2 to salivary components and experimental apparatus is probably due to the hydrophobic regions on MG2. Indeed, numerous hydrophobic domains in the regions flanking the tandem repeat region are present in a Kyte–Doolittle hydropathy plot of the MUC7 protein sequence (presented in [9]).

In summary, the purification protocol presented allows the purification of MG2 in two separation steps, both of which can be employed on a large scale; the efficiency of recovery is approx. 40%. The concentration of MG2 calculated to be in stimulated human whole saliva from results in this study ($146\ \mu\text{g}/\text{ml}$) is higher than that from a previous investigation ($77.4 \pm 55.1\ \mu\text{g}/\text{ml}$), which calculated the concentration of MG2 by immunological methods [16]. The purity of the preparation was assessed by SDS/PAGE with colloidal Brilliant Blue G staining and autoradiography after non-specific ^{14}C -labelling of lysine residues. No significant protein contamination was detected in either case. In addition, no contaminating material of any kind could be detected by on-line refractive index measurements made during gel chromatography.

It has been reported that the core protein of the MG2 mucin is encoded by the *MUC7* gene [9]. Peptide mass mapping results

alone were insufficient to make this identification for our MG2 preparation. Only four of the fragments predicted from a theoretical trypsin digestion of the deduced primary sequence of the MUC7 polypeptide were found. This discrepancy might be accounted for by substitution on many of the peptides in the mature mucin with O-linked glycans. However, the amino acid compositional and N-terminal sequence results are able to identify the polypeptide underlying our mucin preparation as the MUC7 gene product.

The physical dimensions of mucins are often dominated by extended protein domains induced by the dense substitution on serine and/or threonine residues with O-linked glycans [17]; this is presumably true for MUC7 as well. The central region of the MUC7 polypeptide is rich in potential O-glycosylation sites, including a tandem repeat region of 138 residues. If this central region were extensively O-glycosylated and extended, for example as a β structure, it would have an estimated length of approx. 65 nm. The measured weight-average length for the mucin of 55 nm is consistent with this approximation. In addition, the contour length measurements suggest the presence of two populations of differently sized molecules. At the concentrations employed for electron microscopy (more than $1\ \mu\text{g}/\text{ml}$) we would not expect significant self-association effects (see below) and thus the majority of the molecules are assumed to be monomeric. However, a small proportion might exist as dimers and these might be accounted for by the larger species we observe (weight-average length 98 nm). If these were dimers the results would be consistent with an end-to-end association of the mucins.

Gel chromatography of the MUC7 (MG2) mucin in 0.2 M NaCl showed evidence of heterogeneity that was abolished by chromatography in 4 M GuHCl. Therefore we infer that the

heterogeneity arises from self-association of the mucins and is consistent with the presence of predominantly monomeric species but also small amounts of dimer and tetramer. The molecular mass of the monomeric mucin (155–158 kDa) is in agreement with previous estimates [6,7]. Interestingly we could not obtain a measurable value for the radius of gyration, even for the tetramer. This finding is not consistent with four or more mucins being linked end-to-end; we therefore propose that they interact at a single end to form clusters. Interestingly there are two cysteine residues in the deduced MUC7 primary sequence and these are towards the N-terminus. The effect of reduction of the mucins is to yield pure monomeric species, which suggests some role for disulphides in the self-association. A model for mucin self-association consistent with both the electron microscopy and light-scattering results is presented in Figure 9.

The self-interaction of this mucin might have an important functional role in saliva. Indeed there is considerable evidence for the interaction of the MG2 mucin with oral bacteria: a number of workers have reported the aggregation of bacteria by this molecule [18,19]. If the mucin interacts via its O-linked oligosaccharides with lectins on the bacterial surface, its less-glycosylated N-terminal domain might remain free to interact with other mucins, which might be interacting with another bacterium. Thus mucin self-associations could thereafter promote bacterial aggregation.

We thank Ms. Nagma Khan for the amino acid analysis, and the Wellcome Trust and Oxford Glycosystems Ltd. for financial support. R. M. was the holder of a BBSERC/CASE award in collaboration with Oxford Glycosystems Ltd.

Received 31 March/2 June 1998; accepted 30 June 1998

REFERENCES

- Milne, R. W. and Dawes, C. (1973) *Vox Sang.* **25**, 298–307
- Tabak, L. A., Levine, M. J., Mandel, I. D. and Ellison, S. A. (1982) *J. Oral Pathol.* **11**, 1–17
- Loomis, R. E., Prakobphol, A., Levine, M. J., Reddy, M. S. and Jones, P. C. (1987) *Arch. Biochem. Biophys.* **258**, 452–464
- Ramasubbu, N., Reddy, M. S., Bergey, E. J., Haraszthy, G. G., Soni, S. D. and Levine, M. J. (1991) *Biochem. J.* **280**, 341–352
- Cohen, R. E., Aguirre, A., Neiders, M. E., Levine, M. J., Jones, P. C., Reddy, M. S. and Haar, J. G. (1991) *Arch. Oral Biol.* **36**, 347–356
- Prakobphol, A., Levine, M. J., Tabak, L. A. and Reddy, M. S. (1982) *Carbohydr. Res.* **108**, 111–122
- Biesbrock, A. R., Reddy, M. S. and Levine, M. J. (1991) *Infect. Immun.* **59**, 3492–3497
- Reddy, M. S., Levine, M. J. and Prakobphol, A. (1985) *J. Dent. Res.* **64**, 33–36
- Bobek, L. A., Tsai, H., Biesbrock, A. R. and Levine, M. J. (1993) *J. Biol. Chem.* **268**, 20563–20569
- Tabak, L. A. (1995) *Annu. Rev. Physiol.* **57**, 547–564
- Sheehan, J. K. and Carlstedt, I. (1990) *Biochem. J.* **265**, 169–177
- Thornton, D. J., Howard, M., Khan, N. and Sheehan, J. K. (1997) *J. Biol. Chem.* **272**, 9561–9566
- Zimm, B. H. (1948) *J. Chem. Phys.* **16**, 1099–1116
- Thornton, D. J., Holmes, D. F., Sheehan, J. K. and Carlstedt, I. (1989) *Anal. Biochem.* **182**, 160–164
- Jourdain, G. W., Dean, L. and Roseman, S. (1971) *J. Biol. Chem.* **246**, 430–435
- Aguirre, A., Testa-Weintraub, L. A., Banderas, J. A., Haraszthy, G. G., Reddy, M. S. and Levine, M. J. (1993) *Crit. Rev. Oral Biol. Med.* **4**, 343–350
- Gerken, T. A., Butenhof, K. J. and Shogren, R. (1989) *Biochemistry* **28**, 5536–5543
- Murray, P. A., Levine, M. J., Tabak, L. A. and Reddy, M. S. (1982) *Biochem. Biophys. Res. Commun.* **106**, 390–396
- Koop, H. M., Valentijn-Benz, M., Nieuw Amerongen, A. V., Roukema, P. A. and de Graaff, J. (1990) *Antonie Van Leeuwenhoek* **57**, 245–252