

## Function of the CysD domain of the gel-forming MUC2 mucin

Daniel AMBORT\*, Sjoerd VAN DER POST\*, Malin E. V. JOHANSSON\*, Jenny MacKENZIE†, Elisabeth THOMSSON\*, Ute KRENGEL† and Gunnar C. HANSSON\*<sup>1</sup>

\*Department of Medical Biochemistry, University of Gothenburg, Medicinaregatan 9A, SE-413 90 Gothenburg, Sweden, and †Department of Chemistry, University of Oslo, Sem Sælands vei 26, NO-0315 Oslo, Norway

The colonic human MUC2 mucin forms a polymeric gel by covalent disulfide bonds in its N- and C-termini. The middle part of MUC2 is largely composed of two highly O-glycosylated mucin domains that are interrupted by a CysD domain of unknown function. We studied its function as recombinant proteins fused to a removable immunoglobulin Fc domain. Analysis of affinity-purified fusion proteins by native gel electrophoresis and gel filtration showed that they formed oligomeric complexes. Analysis of the individual isolated CysD parts showed that they formed dimers both when flanked by two MUC2 tandem repeats and without these. Cleavages of the two non-reduced CysD fusion

proteins and analysis by MS revealed the localization of all five CysD disulfide bonds and that the predicted C-mannosylated site was not glycosylated. All disulfide bonds were within individual peptides showing that the domain was stabilized by intramolecular disulfide bonds and that CysD dimers were of non-covalent nature. These observations suggest that CysD domains act as non-covalent cross-links in the MUC2 gel, thereby determining the pore sizes of the mucus.

**Key words:** C-mannosylation, disulfide bonds, mass spectrometry (MS), mucus, non-covalent dimer.

### INTRODUCTION

Mucins are large glycoproteins that coat the surfaces of cells in the respiratory, digestive and urogenital tracts, and, in some amphibia, the skin [1,2]. They function to protect epithelial cells from infection, dehydration and physical injury, as well as to aid passage of materials on mucosal surfaces. The mucins have mucin domains with large amounts of O-linked oligosaccharides attached on the protein sequence rich in proline, threonine and serine (PTS domain) [3]. Moreover, these domains often have tandemly repeated amino acid sequences that vary in number, length and sequence in different mucins [3]. There are two types of mucins, membrane-bound and secreted. In humans, six secreted mucins have been confirmed; the gel-forming MUC2 [4], MUC5AC [5], MUC5B [6], MUC6 [7] and MUC19 [8], and the non-gel forming MUC7 [9].

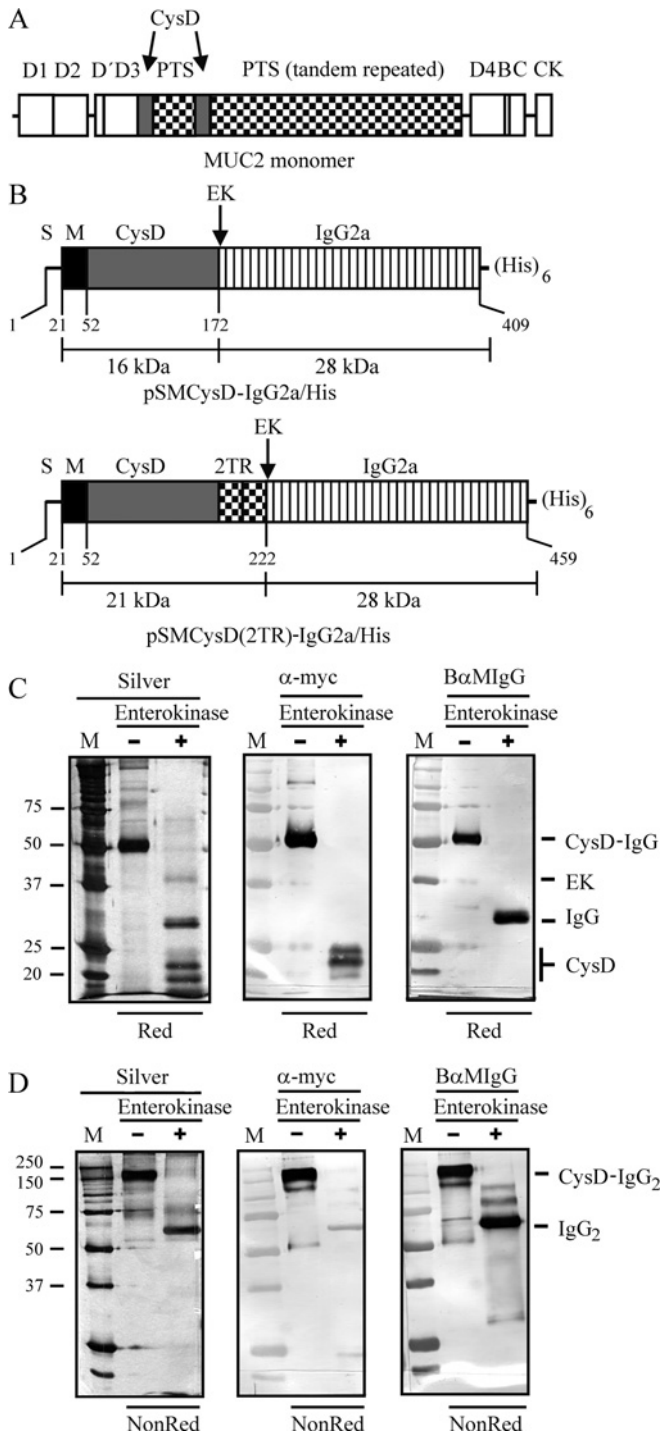
MUC2 is the major gel-forming mucin of the colon and organizes the mucus into two layers, an inner densely packed layer and an outer loosely adherent layer that is colonized by bacteria [10]. MUC2 contains five distinct regions (Figure 1A): an N-terminal part with von Willebrand D1-D2-D'-D3 domains and a CysD domain, a small PTS domain, another CysD domain, a large PTS (tandem-repeated) domain, and a C-terminal part with von Willebrand D4-B-C domains and a CK (cystine-knot) domain [4]. MUC2 forms disulfide-linked dimers via its C-terminal parts in the endoplasmic reticulum [11,12] and disulfide-linked trimers via its N-terminal part in the *trans*-Golgi network [13]. The concerted C-terminal dimerization and N-terminal trimerization of MUC2 monomers leads to the assembly of a polymeric net-like structure that can bind water via its heavily O-glycosylated mucin domains to form a gel. Although the von Willebrand domains and CK domain are involved in polymer formation, the function of the CysD domains remains elusive. Interestingly, multiple CysD

domains have also been identified in other human gel-forming mucins, in MUC5AC [14] and MUC5B [15] respectively. The small 110-residue-long CysD domains with ten invariant cysteine residues are found adjacent to or scattered within the heavily O-glycosylated central region of these mucins, at least nine in MUC5AC and seven in MUC5B. It is worth noting that the genes for human gel-forming mucins MUC2, MUC5AC and MUC5B are localized to the same chromosome locus, 11p15.5 [16]. The high intra- and inter-species homologies suggest that the CysD domains play critical, but as yet undefined, roles in mucus homeostasis. Mucin CysD domains show significant homologies with a protein domain in human cartilage, intermediate layer protein [17], and with a protein domain in the protein Oikosin 1 of the larvacean tunicate *Oikopleura dioica* [18]. Neither the tertiary structure nor the specific functions of these two proteins are known, although it has been proposed that they have structural roles in the organization of articular cartilage and the larvacean mucous houses respectively. A potential C-mannosylation WXXW acceptor site [19] is found in the CysD domains of MUC2, MUC5AC and MUC5B, and in 11 repeats in Oikosin 1. It has been suggested that at least the CysD1 and CysD5 domains of MUC5AC are mannosylated [20]. Moreover, Perez-Vilar et al. [20] reported that the CysD domains CysD1 and CysD5 of MUC5AC and CysD1 and CysD3 of MUC5B were monomers. The mucin domains of different mucins are similar in sequence and the localization of the cysteine residues are identical (see <http://www.medkem.gu.se/mucinbiology/databases/index.html>). More recently, we found the CysD domain in a number of gel-forming mucins from other species such as mammals, frog, fish, fruit fly, sea urchin, sea squirt and lancelet respectively [3].

We therefore aimed at studying the function and biochemical properties of the CysD domain. We chose to analyse the second CysD domain of human MUC2. Two recombinant fusion proteins

Abbreviations used: AGC, automatic gain control; BN-PAGE, blue native PAGE; CHO, Chinese-hamster ovary; CID, collision-induced dissociation; CK domain, cystine-knot domain; EK, enterokinase; ESI, electrospray ionization; FBS, fetal bovine serum; IMDM, Iscove's modified Dulbecco's medium; LC, liquid chromatography; mAb, monoclonal antibody; MS/MS, tandem MS; PTS domain, proline, threonine and serine domain; TCEP-HCl, tris(2-carboxyethyl)phosphine-HCl.

<sup>1</sup> To whom correspondence should be addressed (email [gunnar.hansson@medkem.gu.se](mailto:gunnar.hansson@medkem.gu.se)).



**Figure 1** Domain organization of MUC2 and the two CysD fusion constructs

(A) MUC2 is made up of D1-D2-D'-D3-CysD-PTS-CysD-PTS (tandem repeated)-D4-B-C-CK domains. (B) The two CysD fusion constructs named pSMCysD-IgG2a/His and pSMCysD(2TR)-IgG2a/His are made up by an N-terminal Myc tag, CysD and a C-terminal IgG-Fc and His<sub>6</sub> tag. The second fusion construct included two repeats (2TR) of the general tandem repeat of the PTS domain. Both fusion proteins included an EK-cleavage site. (C and D) Analysis of Protein G-purified CysD-IgG fusion protein by SDS/PAGE and Western blotting under reducing (Red) and non-reducing (NonRed) conditions before (-) and after (+) EK cleavage. M, molecular-mass standards in kDa. The anti-Myc tag antibody labelled the CysD part, whereas the B $\alpha$ MlgG antibody recognized the IgG part. Silver, Silver staining.

were constructed, in which the CysD domain alone or including two copies of the general tandem repeat from the PTS domain was fused to a removable IgG-Fc region. The two fusion proteins expressed in CHO (Chinese-hamster ovary) cells were secreted into the culture medium. The purified fusion proteins were studied by gel filtration, and denaturing and native gel electrophoresis, and the disulfide bond pattern was established by MS. The CysD part formed pH-independent non-covalent dimers stabilized by intramolecular disulfide bonds.

## EXPERIMENTAL

### Construction of expression vector pSMCysD-IgG2a/His and pSMCysD(2TR)-IgG2a/His

A PCR product termed EK (enterokinase)-IgG2a/His was generated from pcDNA3/MUC1-IgG [21] using primers 5'-CGTC-TAGAGGATCCGATGATGATAAAGCAGAGC-3' and 5'-CGTCGCTAGCATGATGGTGTATGGTGTATGGTGTAG-3'. This encoded XbaI and NheI sites, an EK-cleavage site, exons 1-3 of the murine IgG2a Fc domain and a C-terminal histidine tag. This amplicon inserted into XbaI and NheI sites of pSM vector [12] encoding a N-terminal mouse Ig  $\kappa$ -chain signal sequence and a Myc tag was named pSM-IgG2a/His. The CysD domain was amplified from cDNA made from mRNA of LS174T cells of human MUC2 with primers 5'-GACTAGTGCTAGCC-CATGCGTGCCTCTCTGC-3' and 5'-GGTCTAGAGTTCTCT-GTGGTGGTGGTTGTCATG-3'. This amplicon encoding XbaI and NheI sites, the second CysD of human MUC2 (bases 5368-5703) and a N-terminal part of the adjacent PTS domain (bases 5661-5703) was inserted in-frame into the XbaI site of pSM-IgG2a/His and was called pSMCysD-IgG2a/His.

The annealed phosphorylated primer pair with primers 5'-GC-TAGCCCAACAACACACCCATCAGCACCACCACCATGG-TGACCCCAACCCCAACACCCACTGGAACACAGACCT-3' and 5'-GCTAGAGGTCTGTGTTCCAGTGGGTGTTGGGGT TGGGGTCAACATGGTGGTGGTGTCTGATGGGTGTCGTTG TTGGG-3' encoding the general tandem repeat of the PTS domain of human MUC2 was inserted sequentially twice in-frame into the XbaI site of pSMCysD-IgG2a/His and was called pSMCysD(2TR)-IgG2a/His.

### Expression of CysD-IgG and CysD(2TR)-IgG

CHO-K1 and CHO-K1 Lec 3.2.8.1 [22] cells grown in 10% FBS (fetal bovine serum) IMDM (Iscove's modified Dulbecco's medium; Lonza) with 1% penicillin/streptomycin were transfected with pSMCysD-IgG2a/His or pSMCysD(2TR)-IgG2a/His using Lipofectamine<sup>TM</sup> 2000 (Invitrogen) and stable clones were generated 2-3 days later by adding 250  $\mu$ g/ml Geneticin (Invitrogen). Clones were selected based on high secretion of fusion protein into culture medium as described previously [23]. One high-level expression clone for each construct was selected, recloned and used for protein production.

### Protein production

Adherent CHO-K1 Lec 3.2.8.1 cells producing CysD-IgG were resuspended at  $0.3 \times 10^6$  cells/ml in 2% FBS ProCHO-4 medium (Lonza) with 4 mM L-glutamine, 1  $\times$  ProHT supplement and 250  $\mu$ g/ml Geneticin (Gibco). Cells were adapted to serum-free suspension growth within 6 weeks at 90 rev./min at 37  $^{\circ}$ C in 5% CO<sub>2</sub>. A 1.5 litre perfusion culture with spin filter separation (10  $\mu$ m) was performed in a 3 litre bioreactor (Applikon)

controlled by an ADI 1010 Bio Controller and an ADI 1025 Bio Console at 37°C, pH 6.9, 40% dissolved O<sub>2</sub> and 150–300 rev./min. Oxygen and CO<sub>2</sub> were introduced by bubble-free aeration using 9 m of 1-mm thick silicone tubing. Perfusion rate was 0.3–0.8 dilutions/day of culture volume. The CysD(2TR)–IgG fusion protein was produced in T-175 flask cultures grown in 10% IMDM with 250 µg/ml Geneticin (Invitrogen).

### SDS/PAGE, BN-PAGE (blue native PAGE) and Western blotting

SDS/PAGE was performed as described previously [24]. Precision protein standards (Bio-Rad Laboratories) were used as markers. Silver staining was carried out as described previously [25]. For LC (liquid chromatography)-ESI (electrospray ionization) MS/MS (tandem MS) non-reducing SDS gels were stained with Imperial Protein Stain (Thermo Scientific). Western blotting was performed as described previously [26]. After transfer the membrane was blocked in 5% skimmed milk powder in PBS with 0.1% Tween 20 or in 10 mM Tris/HCl (pH 7.5), 1% BSA, 100 mM NaCl, 0.1% Tween 20 and 0.05% sodium azide (for biotin-labelled antibody) and incubated with an anti-Myc mAb (monoclonal antibody) (clone 9E10.2, American Type Culture Collection, CRL-1729) or biotin-labelled rat anti-mouse IgG2a (R19–15, BD Biosciences). The anti-Myc mAb was cultured by the Mammalian Protein Expression core facility at the University of Gothenburg. The membrane was then treated with AP (alkaline phosphatase)-conjugated goat anti-mouse IgG1 (Southern Biotech) or streptavidin (Southern Biotech) and developed with NBT (Nitro Blue Tetrazolium)/BCIP (5-bromo-4-chloroindol-3-yl phosphate) (Promega).

BN-PAGE allows studies of proteins under native conditions [27]. BN-PAGE was performed as described previously [27] on ready-made NativePAGE™ Novex® 4–16% BisTris Gels (Invitrogen). NativeMark™ Unstained Protein Standards (Invitrogen) were used as markers. BN gels were stained using silver [25].

### Protein G purification of CysD–IgG, CysD(2TR)–IgG or EK-cleaved CysD–IgG

Spent culture medium (9.5 litres) containing CysD–IgG fusion protein was filtered (0.45 µm Mini Capsule, PALL) and concentrated by Tangential Flow Filtration (Pellicon™-2 system, Millipore) with two 5 kDa PLCC filters (four times in 1 litre of PBS reduced to 500 ml). For Protein G purification of CysD(2TR)–IgG fusion protein, serum-containing medium (130 ml) was dialysed (Spectra/Por® Dialysis Membrane, molecular-mass cut-off of 6–8 kDa, Spectrum Laboratories) three times against 20 mM sodium phosphate buffer (pH 7) and filtered (Durapore® Membrane Filter, 0.22 µm GVWP, Millipore). The concentrate or filtered medium was loaded on to a HiTrap Protein G HP column (1.6 cm × 2.5 cm, Amersham Biosciences) equilibrated with 20 mM sodium phosphate buffer (pH 7) at 1 ml/min. The column was rinsed with the same buffer. Bound components were eluted [100 mM glycine/HCl (pH 2.7)], collected in 1-ml neutralized fractions [100 mM Tris/HCl (pH 9)] and analysed by SDS/PAGE and silver staining. EK-cleaved CysD–IgG was directly loaded on to a Protein G column and the flow-through and eluate fractions were analysed.

### Sialidase and EK treatment

Purified CysD–IgG and CysD(2TR)–IgG fusion proteins were cleaved with EKMax™ for 15 h at 37°C in EKMax™ Reaction

Buffer [50 mM Tris/HCl (pH 8.0), 1 mM CaCl<sub>2</sub> and 0.1% Tween 20 (Invitrogen)] at 1:100 (enzyme/substrate ratio) and the reaction was stopped with 2 mM PMSF. The CysD(2TR)–IgG fusion protein was treated with sialidase [1:100 (Sigma)] and EK-cleaved (1:100) for 18 h at 37°C in 50 mM sodium phosphate buffer (pH 6.0). Desialylated and/or EK-cleaved samples were analysed by gel filtration, SDS/PAGE and BN-PAGE.

### Gel filtration

Purified CysD–IgG, CysD, IgG or sialylated and desialylated EK-cleaved CysD(2TR)–IgG fusion protein (5 µg of each) was applied to a Superose 6 or Superose 12 PC 3.2/30 column (Amersham Biosciences) and eluted at 0.02 ml/min using an Ettan system (Amersham Biosciences) with UV absorbance at 215 nm, 254 nm and 280 nm. Standards were chromatographed at pH 7 (50 mM phosphate buffer and 150 mM NaCl). The purified CysD–IgG and CysD(2TR)–IgG (EK- or sialidase-treated) were either directly injected or were buffer-exchanged by ultrafiltration (10 000 g at 4°C; Vivaspin 6 PES, molecular-mass cut-off of 10 kDa, Sartorius) to pH 5.2 (100 mM acetic acid), pH 7 (50 mM phosphate buffer), pH 7.4 (100 mM Hepes) or pH 8.3 (100 mM sodium bicarbonate) containing 150 mM NaCl with or without 50 mM CaCl<sub>2</sub>. The CysD-containing (flow-through) or IgG-containing (eluate) protein parts isolated by Protein G purification of EK-treated fusion protein were either directly injected or buffer-exchanged by ultrafiltration (10 000 g at 4°C; Vivaspin 6 PES, molecular-mass cut-off of 10 kDa, Sartorius) to pH 5 (20 mM acetic acid), pH 6 (20 mM Mes), pH 7 (20 mM Tris/HCl), pH 7.4 (10 mM phosphate buffer) or pH 8 (20 mM Tris/HCl) containing 150 mM NaCl and 5 mM EDTA or 10 mM CaCl<sub>2</sub>.

### LC-ESI MS/MS, in-gel digestion and MS data analysis

CysD–IgG- and CysD(2TR)–IgG-containing bands excised from non-reducing SDS gels were destained [three times in 50% acetonitrile and 25 mM Tris/HCl (pH 7.8)], dried and digested with Asp-N [0.01 µg/µl in 25 mM Tris/HCl (pH 7.8) (Roche)]. Peptides were extracted and partly dried by centrifugation under vacuum. The digest of one sample per condition was reduced {50 mM TCEP-HCl [tris(2-carboxyethyl)phosphine-HCl]} after which the peptides from both conditions were extracted using C<sub>18</sub> ZipTips (Millipore), eluted (60% acetonitrile, 40% H<sub>2</sub>O and 0.2% trifluoroacetic acid), partly dried and solubilized (0.2% formic acid). Samples were analysed by LC-ESI MS/MS (LTQ Orbitrap XL, Thermo Scientific). Sample injection and LC were performed as described previously [28]. Data were acquired in a data-dependent mode automatically switching between MS and MS/MS acquisition. MS scans were obtained in the Orbitrap at *m/z* 400–2000, two microscans, maximum 500 ms injection and an AGC (automatic gain control) of 500 000. MS/MS was performed in linear ion-trap on the six most abundant multiple charged ions for each scan (one microscan, maximum 200 ms injection and an AGC of 30 000) using CID (collision-induced dissociation) fragmentation at 30% normalized collision energy. After fragmentation, peptides were excluded for 10 s for further acquisition. Peaklists were generated from raw data using extract\_msn.exe (Thermo Scientific). Data were interpreted using Mascot (version 2.2, Matrix Science) searched against IPI human database (version 3.52) with addition of the CysD–IgG or CysD(2TR)–IgG fusion construct. A second search was performed against a specific cross-linked database generated for CysD–IgG or Cys(2TR)–IgG using xComb (version 1.1) [29].

Search parameters for reduced samples against IPI human were as follows: (i) one missed cleavage Asp-N; (ii) tolerance 5 p.p.m. (precursor), 0.5 Da (fragment ions); (iii) charge state 2<sup>+</sup>, 3<sup>+</sup>, 4<sup>+</sup>; and (iv) oxidation cysteine, methionine (variable). Search parameters for non-reduced samples were as follows: (i) 'do not cleave' after amino acid 'J' for non-cleavage; (ii) tolerance 5 p.p.m. (precursor), 0.5 Da (fragment ions); (iii) charge state 2<sup>+</sup>, 3<sup>+</sup>, 4<sup>+</sup>; and (iv) dehydro cysteine (fixed), oxidation methionine (variable), gain of water asparatate (variable). The Mascot cut-off score was set to 25.

## RESULTS

### Construction and expression of a recombinant CysD-IgG fusion protein

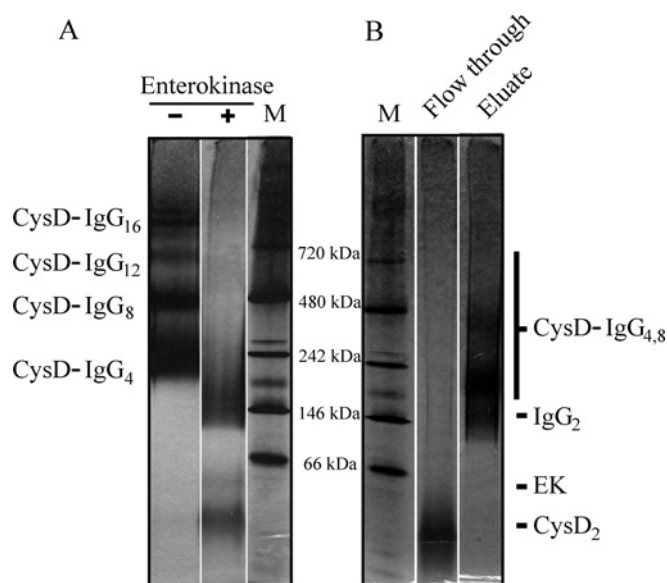
A plasmid expressing a Myc tag, the second CysD of human MUC2 (residues 1782–1878; [4]), an EK-cleavage site, exons 1–3 of the murine IgG-Fc region and a C-terminal histidine tag was constructed and called pSMCysD-IgG2a/His (Figure 1B). Fusion of the CysD domain to IgG was chosen as this gives high protein expression. The CysD domain contains 97 amino acids and lies between the small and large PTS domains of human MUC2. The plasmid was transfected into CHO-K1 Lec 3.2.8.1 cells [22]. Stable clones were generated and selected for the secretion of maximum levels of CysD-IgG, as determined by immunoassay [23] and Western blotting (results not shown). One high-level expression clone was selected and adapted to suspension culture in protein-free medium.

### Purification of the CysD-IgG fusion protein

To obtain the CysD-IgG fusion protein, the supernatant from a perfusion culture was concentrated by ultrafiltration and then loaded on to a Protein G column. The bound proteins were eluted and fractions were analysed by SDS/PAGE and silver stained (Figures 1C and 1D). The fusion protein migrated at ~50 kDa under reducing conditions (Figure 1C, -EK) and at ~150 kDa under non-reducing conditions on SDS gels corresponding to the CysD-IgG monomer and dimer respectively (Figure 1D, -EK). To separate the CysD from IgG, the fusion protein was cleaved with EK. The IgG part showed a ~30 kDa band under reducing conditions (Figure 1C, +EK) and a dimeric ~60 kDa band under non-reducing conditions that were both labelled by the IgG-specific antibody (Figure 1D, +EK). Three distinct bands of ~20–22 kDa were found under reducing SDS/PAGE that were identified as the CysD part with the anti-Myc tag antibody (Figure 1C, +EK). When the EK-cleaved samples were analysed by non-reducing SDS/PAGE most of the CysD-containing material disappeared (Figure 1D, +EK). The disappearance of the CysD protein under non-reducing conditions on SDS-treated samples led us to speculate that it had a high tendency to form insoluble aggregates and the CysD was studied further under native conditions.

### Gel filtration and BN-PAGE of the CysD-IgG fusion protein

The intact fusion protein displayed four major bands at ~200 kDa, ~400 kDa, ~600 kDa and ~800 kDa for the tetra-, octa-, dodeca- and hexadeca-mers respectively, on silver-stained BN gels (Figure 2A, -EK). The same material analysed in reducing SDS gels gave single bands of ~50 kDa (Figure 1C, -EK). The EK-cleaved fusion protein showed two strong bands of ~35 kDa for the CysD dimer and ~120–180 kDa for the IgG2



**Figure 2** BN-PAGE of purified CysD-IgG fusion protein

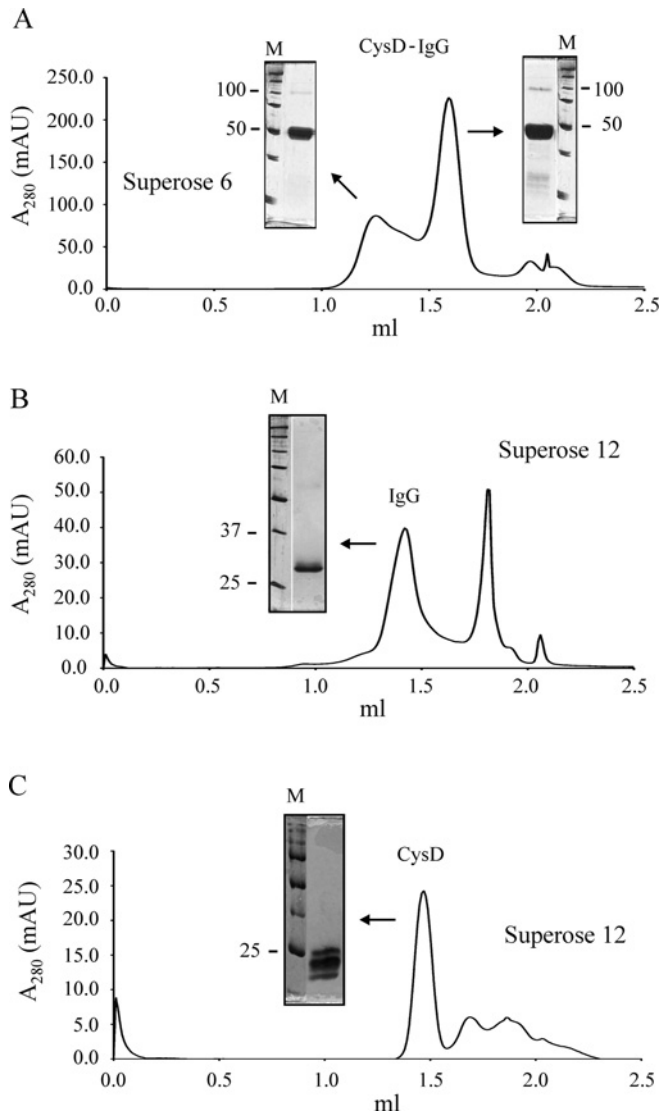
(A) BN gel of purified fusion protein before (-) and after (+) EK cleavage. (B) EK-cleaved fusion protein was loaded on to a Protein G column and the flow-through and eluate was analysed by BN-PAGE. M, molecular-mass standards in kDa. The gel lanes, including standards, shown together come from the same gel.

dimer respectively, in BN gels (Figure 2A, +EK). Next, EK-cleaved fusion protein was loaded on to a Protein G column and analysed by BN-PAGE. The CysD fusion part found in the flow-through migrated as a ~35 kDa CysD dimer on BN gels as the calculated CysD monomeric mass is 16 kDa (Figure 2B, Flow through). When the IgG-containing eluate was separated on native gels, three bands of ~120–180 kDa, ~200 kDa and ~400 kDa respectively, were found (Figure 2B, Eluate), the first band due to dimeric IgG that was cleaved from its fusion partner, and the two latter ones due to oligomeric forms of remaining non-cleaved fusion protein.

Gel filtration was then used to further investigate the oligomeric state of CysD and its fusion protein. The CysD-IgG fusion protein eluted as two major peaks at ~200 kDa and ~800 kDa respectively (Figure 3A). This suggests the higher oligomeric tetra and hexadeca forms as observed on the native gels as analysis of these two peaks on reducing SDS gels showed the expected ~50 kDa CysD-IgG band in both peaks. When the IgG part was analysed by gel filtration, a single peak eluted at ~35 kDa (Figure 3B). Reducing SDS/PAGE and silver staining confirmed the presence of the ~30 kDa IgG form. The CysD protein eluted at 30 kDa suggesting a dimeric form as it migrated with an apparent mass of 20–22 kDa on reducing SDS gels (Figure 3C).

### Gel filtration of CysD(2TR)-IgG fusion protein

To further prove the dimeric nature of the CysD domain and that this was not caused by the absence of a flanking mucin domain with a typical glycosylation, a second fusion protein with two human MUC2 tandem repeats (23 amino acids each) were also analysed (Figures 1A and 1B). The plasmid pSMCysD(2TR)-IgG2a/His was transfected into CHO-K1 cells which give normal sialylated N-glycans and largely sialyl-T O-glycans (NeuAc $\alpha$ 2-3Gal $\beta$ 1-3GalNAc-) [21]. One high-level expression clone was selected for protein production. The Protein G-purified fusion



**Figure 3** Gel filtration of the CysD-IgG fusion protein

(A) Purified fusion protein was analysed on a Superose 6 gel-filtration column. The Superose 6 column had a void volume of 0.843 ml, and the standards thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa) and ovalbumin (43 kDa) eluted at 1.315 ml, 1.494 ml, 1.629 ml and 1.753 ml respectively. (B and C) EK-cleaved fusion protein was loaded on to a Protein G column and the bound material (B) and flow-through (C) were analysed on a Superose 12 gel-filtration column. The Superose 12 column had a void volume of 0.831 ml, and the standards aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa) eluted at 1.236 ml, 1.315 ml, 1.382 ml and 1.584 ml respectively. UV absorbance units are given in mAU. The content of collected fractions from major peaks (indicated by arrows) was analysed on reducing SDS gels (shown as insets) with silver staining. M, molecular-mass standards in kDa. The gel lanes shown together come from the same gel.

protein migrated at ~75 kDa on a reducing SDS gel (results not shown). The IgG part of the EK-cleaved fusion protein showed as for the first protein a ~30 kDa band under reducing and a ~60 kDa band under non-reducing SDS/PAGE (Figure 4A). Three distinct bands of ~40–45 kDa owing to the CysD(2TR) part were labelled by the Myc tag antibody in a reducing SDS gel (Figure 4A). When the same EK-cleaved material was analysed under non-reducing SDS/PAGE, most of the CysD-containing material was absent as before (Figure 4A). The EK-cleaved fusion protein eluted from a gel-filtration column as two distinct peaks at ~113 kDa and

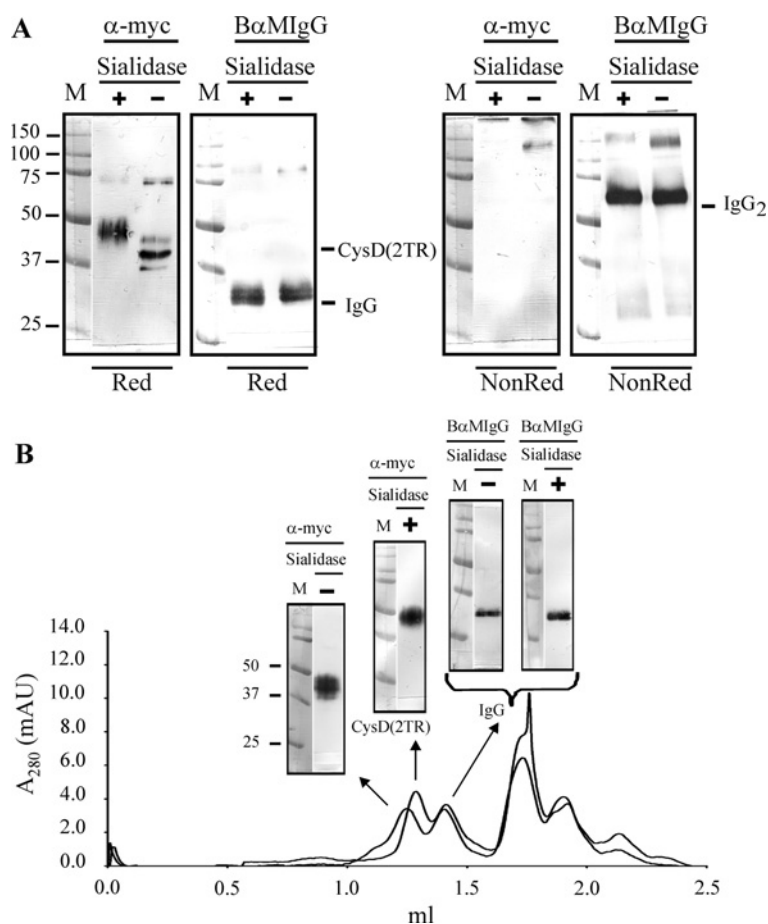
~35 kDa respectively. The 113 kDa peak was due to dimeric CysD(2TR), whereas the 35 kDa protein form was due to IgG as confirmed by SDS/PAGE and Western blotting (Figure 4B, – sialidase). Desialylation had no influence on the elution time of the IgG part from the gel-filtration column or migration behaviour in reducing and non-reducing gels (Figures 4A and 4B). Desialylated dimeric CysD eluted at ~95 kDa from the gel-filtration column (Figure 4B, + sialidase) and migrated as a monomer at ~50 kDa, slightly larger than sialylated, when analysed by reducing SDS/PAGE (Figure 4B, + sialidase). The sialic acids are probably attached to the N-linked glycans as the CysD domain of MUC2 contains three potential N-glycosylation sites and this also explains the presence of three separate bands on reduced SDS gels when the protein was sialylated. It can thus be concluded that the CysD had an estimated mass inferred from gel filtration and BN-PAGE experiments as for a dimer. The IgG formed covalent dimers and when fused to the CysD that also formed dimers, larger oligomeric ladders were found. The CysD dimer were not quantitatively revealed on non-reduced SDS gels, something that might suggest disulfide stabilization. This was not the case as shown below.

#### Effect of pH and calcium on the dimeric state of the CysD domain

To reveal whether the CysD dimer was affected by pH and calcium, recombinant CysD was isolated and analysed by gel filtration at different pH values in the presence or absence of calcium. The isolated CysD-containing material eluted as a 30 kDa dimer in buffers at low or high pH (Figure 5A) and when analysed in the presence or absence of calcium (Figure 5B). A tiny shift towards a smaller size, less than expected for a change in oligomeric status, was observed at low pH and high calcium as for the conditions in the secretory vesicles. Different pH and calcium conditions were tested to dissociate the 30 kDa CysD dimers into monomers (described in the Experimental section), but no conditions were able to do this (results not shown).

#### Disulfide bond pattern of the CysD domain determined by LC-ESI MS/MS

To reveal whether the CysD dimers were held together by disulfide bonds, the band of interest containing the CysD-IgG fusion protein was excised from a non-reducing SDS gel, in-gel digested with Asp-N and then analysed by LC-ESI MS/MS and database searching as described previously [29]. All ten cysteine residues of the CysD domain formed five intramolecular disulfide bonds within four peptides (Figure 6A). Three unambiguous cysteine pairs were allocated as follows: Cys<sup>32</sup>-Cys<sup>36</sup>, Cys<sup>63</sup>-Cys<sup>73</sup> and Cys<sup>92</sup>-Cys<sup>100</sup> respectively. MS analysis revealed that the non-cysteine-containing peptide at  $m/z$  934.96<sup>2+</sup> was partly acetylated at Lys<sup>46</sup>. The last peptide was observed as a triply charged ion at  $m/z$  924.07<sup>3+</sup>. This peptide included the four remaining cysteine residues that formed two internal disulfide bonds as the observed molecular mass (2769.17 Da) had lost four hydrogen atoms as for two disulfide bonds. There were no fragmentation spectra that could reveal the exact disulfide bond pattern, but the most likely disulfide bond arrangement in this peptide was Cys<sup>116</sup>-Cys<sup>126</sup> and Cys<sup>125</sup>-Cys<sup>128</sup> respectively. The peptide DVPIGQLGQTVVCDVSVGLICKNE, including an Asp-N cleavage site at Asp<sup>93</sup> within the Cys<sup>92</sup>-Cys<sup>100</sup> disulfide bridge (in bold), eluted as two distinct peaks due to a partial cleavage at this site (Figure 6B). The first peptide at  $m/z$  835.09<sup>3+</sup> eluted at 37.5 min from the column, whereas the



**Figure 4** Analysis of EK-cleaved CysD(2TR)-IgG fusion protein by SDS/PAGE and gel filtration

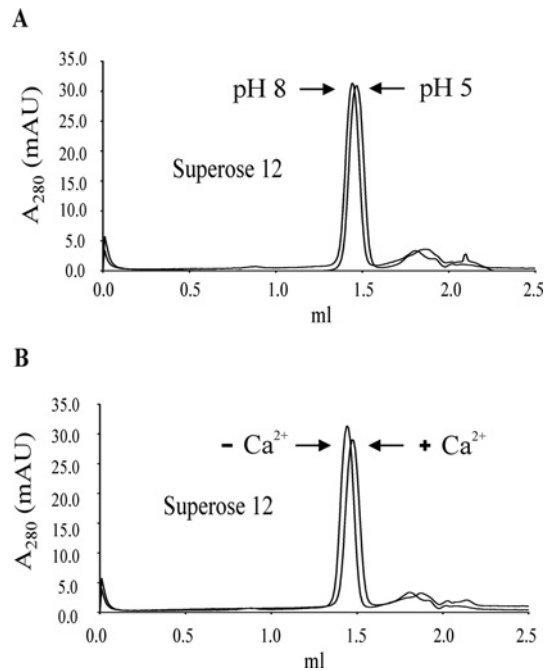
(A) EK-cleaved CysD(2TR)-IgG fusion protein was analysed by SDS/PAGE and Western blotting under reducing (Red) and non-reducing (NonRed) conditions before (–) and after (+) sialidase treatment. The same material as in (A) was also analysed on a Superose 12 gel-filtration column (B). Elution times of standards are as specified in the legend for Figures 3(B) and 3(C). UV absorbance units are given in mAU. The content of collected fractions from major peaks (indicated by arrows) before (–) and after (+) sialidase treatment was analysed by reducing SDS/PAGE and Western blotting (shown as insets). The anti-Myc tag antibody labelled the CysD(2TR) part, whereas the B $\alpha$ MiGg antibody recognized the IgG part. M, molecular-mass standards in kDa. The gel lanes shown together come from the same gel.

second peptide at  $m/z$  1242.63<sup>2+</sup> eluted at 52.5 min. The first peak originated from cleavage at Asp<sup>93</sup> within the loop formed by the Cys<sup>92</sup>-Cys<sup>100</sup> disulfide pair. In the second peak, the non-cleaved form of the same peptide was found with an intact loop. Proteolytic cleavage within the loop of Cys<sup>92</sup>-Cys<sup>100</sup> had a profound effect on the hydrophobic properties of this peptide, as the elution times for the two different forms when separated on the hydrophobic C<sub>18</sub> column shifted dramatically upon opening the loop.

In the peptide DLSSPCVPLCNWTGWL at  $m/z$  894.90<sup>2+</sup>, the intramolecular disulfide bridge (Cys<sup>32</sup>-Cys<sup>36</sup>) was still intact after fragmentation, and peptide sequence information was only obtained for the neighbouring amino acids that were flanking the intact cysteine pair (Figure 7A). Upon reduction of the samples with the reducing agent TCEP-HCl the intact disulfide bridge was chemically cleaved (Figure 7B) and almost complete *y* and *b* ion series were obtained. The WXXW peptide motif in this peptide was not C-mannosylated as predicted on the first tryptophan residue when analysed in its reduced or non-reduced form (Figure 7). Thus all disulfide pairs were within the CysD and no indication of any intermolecular disulfide bonds could be found.

## DISCUSSION

The function of CysD domains found in gel-forming mucins has not been understood to date. However, we have been able to show that the CysD domain of human MUC2 forms non-covalent dimers. The ~50 kDa form of the CysD-IgG fusion protein observed in reducing SDS gels migrated at ~200 kDa, ~400 kDa, ~600 kDa and ~800 kDa by BN-PAGE, and eluted at ~200 kDa and ~800 kDa by gel filtration due to IgG covalent and CysD non-covalent dimers. Furthermore, the CysD part with a calculated mass without glycans of 16 kDa eluted at 30 kDa by gel filtration and at 35 kDa by BN-PAGE, suggesting that CysD forms dimers. The Fc region of IgG from mouse is known to form three intermolecular disulfide bridges via residues Cys<sup>237</sup>, Cys<sup>240</sup> and Cys<sup>242</sup> respectively, in the hinge core by connecting two IgG chains to a covalent dimer [30]. In line with this, we observed that the ~30 kDa IgG part formed a ~60 kDa dimer in non-reducing SDS gels. The dimeric nature of the CysD domain was confirmed by analysing a second recombinant CysD fusion protein that included two 23-amino acid-tandem repeats of the MUC2 mucin. The CysD(2TR) protein formed three bands at ~40–45 kDa on



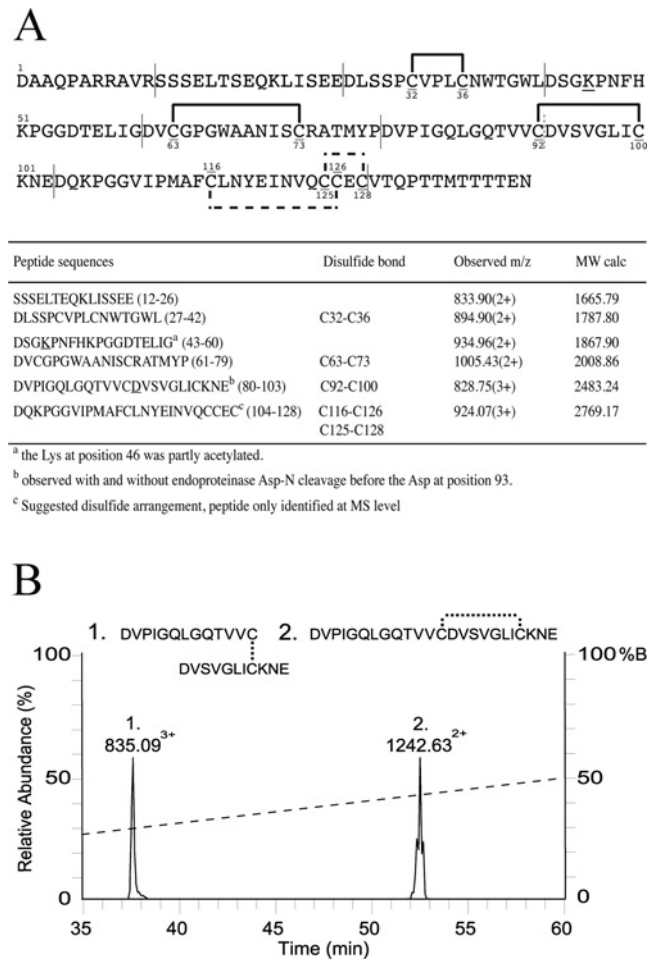
**Figure 5** Effect of pH and calcium on the dimeric state of the CysD domain

The CysD-containing fraction of EK-cleaved CysD-IgG fusion protein was analysed on a Superose 12 gel-filtration column at pH 8 and pH 5 with 5 mM EDTA (**A**) or at pH 8 with and without 10 mM  $\text{CaCl}_2$  (**B**). UV absorbance units are given in mAU. Elution times of standards are as specified in the legend for Figures 3(**B**) and 3(**C**).

reducing SDS gels and eluted as a  $\sim 113$  kDa peak in gel filtration. Mass determination by gel filtration depends on the hydrodynamic (Stokes) radius of the protein and overestimation of mass is not atypical for glycoproteins [31]. Nevertheless, comparative gel-filtration analysis of glycoproteins and their desialylated forms revealed that the apparent molecular masses of desialylated forms were smaller than expected just from removal of the mass of sialic acid [32]. Thus the aberrant behaviour of glycoproteins by gel filtration is largely contributed to by the negative charges of sialic acids [32]. Desialylated CysD(2TR) protein migrated at  $\sim 50$  kDa in reducing SDS gels and eluted as a  $\sim 95$  kDa peak by gel filtration. These findings strongly supported that the CysD domain formed dimers. The three bands of  $\sim 20$ – $22$  kDa of recombinant CysD and  $\sim 40$ – $45$  kDa of the CysD(2TR) protein respectively, found in reducing SDS gels originated from differentially sialylated glycoprotein forms. For the latter protein we showed that the three distinct bands were shifted to one  $\sim 50$  kDa band by sialidase treatment. The sialic acids are probably attached to the N-linked glycans as the CysD domain of MUC2 contains three N-glycans.

Perez-Vilar et al. [20] have previously studied two of the CysD domains in the MUC5AC mucin and they concluded that these are monomeric. In contrast with [20], we have made quantitative biochemical analyses at the protein level instead of only using radioactive traces. Although we have not worked with the same mucin CysD, we have not found evidence for the conclusions made by Perez-Vilar et al. [20]. Cross-linking experiments as reported by Perez-Vilar et al. [20] after simple purification does not prove that CysD are monomers. In fact, there are substantial amounts of dimers in their gels.

In conclusion, CysD forms non-covalent dimers and the  $\sim 200$  kDa tetrameric complexes of the CysD-IgG2a/His fusion

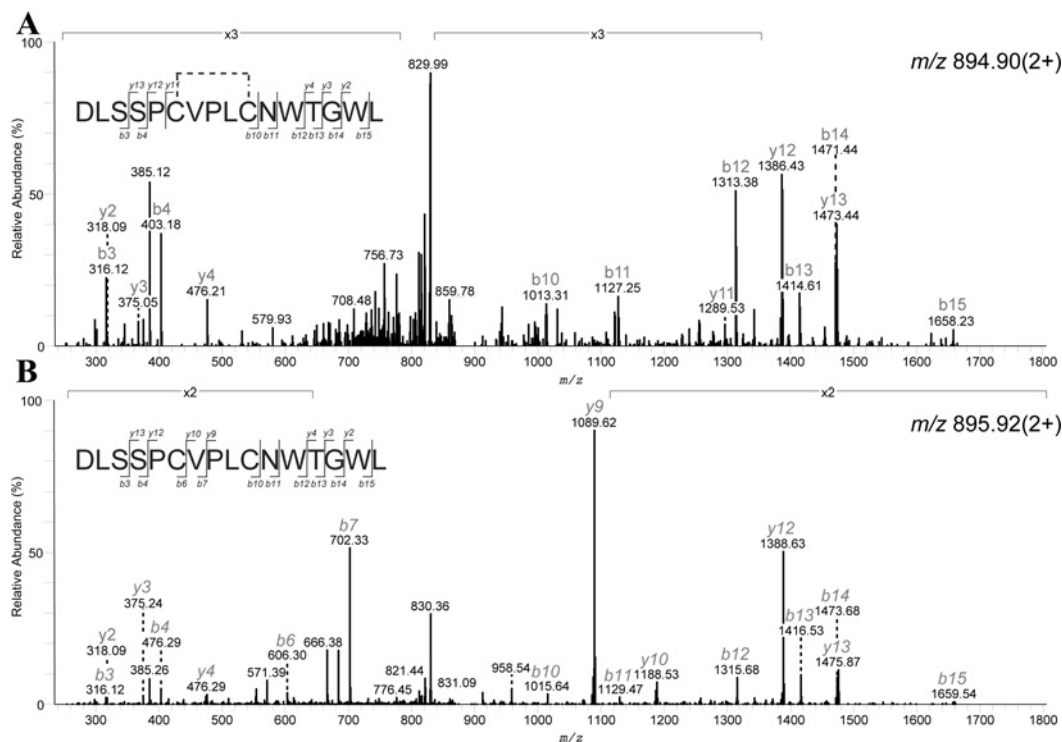


**Figure 6** Analysis of disulfide bonds in the CysD domain by LC-ESI MS/MS

(**A**) Disulfide bond pattern in the CysD domain. Underlined numbers indicate cysteine residues forming intramolecular disulfide bonds. Solid horizontal lines indicate unambiguous disulfide bonds, whereas broken horizontal lines indicate theoretical disulfide bonds. Solid vertical lines represent the position of cleavage sites. (**B**) LC-MS separation of the peptide DVPIGQLGQTVVCDVSVGLICKNE with an internal disulfide bond. Peak number one represented the extracted ion chromatogram of the peptide cleaved N-terminal to the aspartate residue position 93. The second peak belonged to the non-cleaved form causing an increase in hydrophobicity.

protein are built up by disulfide-bonded IgG dimers that interact via its fused CysD domain(s) with another identical dimer (Figure 8A). Further oligomerization (dimerization) of the  $\sim 200$  kDa complexes is mediated via non-occupied free CysD domains thus forming stable  $\sim 400$  kDa octameric (Figure 8B),  $\sim 600$  kDa dodecameric and  $\sim 800$  kDa hexadecameric complexes.

To further understand the nature of the CysD dimer and the bonds formed by the ten cysteine residues in the CysD domain, we cleaved the CysD-IgG fusion protein with Asp-N and analysed the peptides by LC-ESI MS/MS. All ten cysteine residues of the CysD domain were involved in intramolecular disulfide bonds as follows: Cys<sup>32</sup>-Cys<sup>36</sup>, Cys<sup>63</sup>-Cys<sup>73</sup>, Cys<sup>92</sup>-Cys<sup>100</sup>, Cys<sup>116</sup>-Cys<sup>126</sup> and Cys<sup>125</sup>-Cys<sup>128</sup> respectively. Although the exact disulfide pattern for the peptide with four cysteine residues could not be elucidated from the MS data, the disulfide bonds Cys<sup>116</sup>-Cys<sup>126</sup> and Cys<sup>125</sup>-Cys<sup>128</sup> are most likely. As all of the five disulfide bonds were pair wise within four peptides and no other cysteine-containing peptides were found, it can be



**Figure 7** LC-ESI MS/MS analysis of the CysD peptide with a C-mannosylation motif

A CysD-IgG-containing band from a non-reduced SDS gel was excised, in-gel digested with Asp-N and analysed by LC-ESI MS/MS. CID-fragmentation spectra of the CysD peptide DLSSPCVPLCNWTGWL in (A) with an intact disulfide bridge and in (B) after reduction with TCEP-HCl. Both fragmentation spectra showed the absence of C-mannosylation on the first tryptophan residue of the WXXW peptide motif.

concluded that the disulfide bonds were intramolecular and that the CysD dimers were held together with non-disulfide bonds.

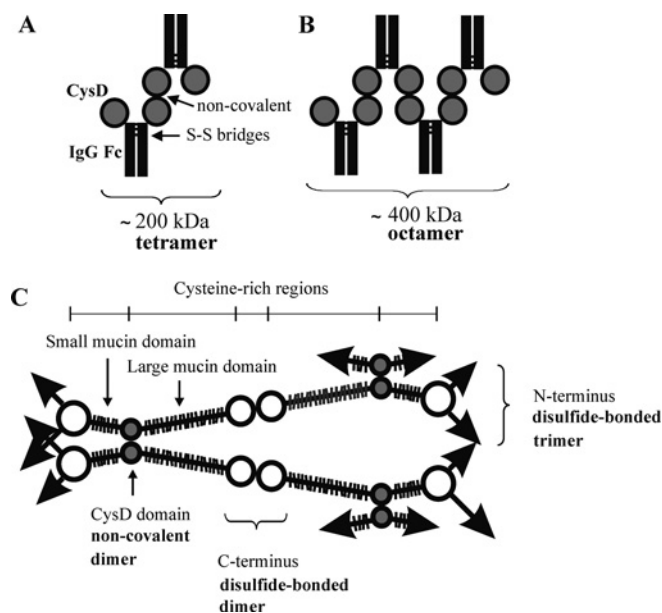
A high hydrophobicity of the loop generated by the Cys<sup>92</sup>–Cys<sup>100</sup> disulfide bridge (CDVSVGLIC) was observed as long as the loop was intact. This hydrophobicity was much higher than that of the open peptide itself and a cleavage between the two cysteine residues (Figure 6B) gave the same loss of hydrophobicity as reducing the disulfide bond of the intact peptide. A similar conserved hydrophobic amino acid sequence stretch between these two cysteine residues can be found in all CysD domains of gel-forming mucins, indicating that the formation of a hydrophobic loop at this site is important for the CysD function. It may be speculated that the ‘disappearance’ of the CysD protein under non-reducing conditions upon SDS/PAGE is due to this (and maybe other) hydrophobic patch(es) present in a highly stabilized globular protein. It is tempting to speculate that the hydrophobic loop stabilized by the Cys<sup>92</sup>–Cys<sup>100</sup> pair is exposed at the surface of the molecule, thereby forming the very stable CysD dimers that were observed by gel filtration. The observation that one of the lysine residues could be acetylated would make the CysD even more hydrophobic.

Besides gel-forming mucins, the CysD domain has also been identified in the protein Oikosin 1 of the larvacean tunicate *O. dioica* [18]. This protein contains 13 tandemly repeated CysD domains. Larvaceans feed on dissolved organic carbon and microorganisms by filtering seawater through a transparent structure called the house. Importantly, Oikosin 1 is the major structural component [18] and is probably involved in the assembly of the secreted house. Owing to the high number of adjacent tandemly repeated CysD domains, it may be speculated that they function in ‘sticking’ the house together.

A potential C-mannosylation WXXW acceptor site [19] is found in the CysD domains of MUC2, MUC5AC and MUC5B, and in 11 repeats in Oikosin 1. Although it was postulated based on mutational studies that at least the CysD1 and CysD5 domains of MUC5AC are mannoseylated [20], we did not find such a modification on the tryptophan residue of the WXXW peptide motif in the peptide DLSSPCVPLCNWTGWL. In our analysis the tryptophan residue was not modified as determined by ESI-MS/MS. Mannosylation is easily identified by MS [33]. The two different CysD variants investigated in the present study were produced in CHO-K1 and CHO-K1 Lec 3.2.8.1 cells respectively. Other studies on C-mannosylated proteins have expressed the protein in the CHO-K1 cell line where the tryptophan residue in the WXXW peptide motif was modified [34]. Perez-Vilar et al. [20] claim CysD mannoseylation, something we could not verify, at least in the CysD investigated in the present study. In their paper [20], indirect experimentation using mutagenesis was done and no biochemical proof supporting such a conclusion was provided.

The pH along the secretory pathway shifts gradually, from 7.2 in the endoplasmic reticulum [35], to 6.0 in the *trans*-Golgi network [36], to 5.2 in the secretory granules [37] and upon release rises up to 8.0 in the colon lumen [38]. In addition, mucin packing into storage granules is accompanied by an increase in intragranular calcium concentration [39]. Gel-filtration analysis of isolated CysD dimer indicated that it did not dissociate at different pH conditions (pH values of 5, 6, 7 or 8) or calcium concentrations (0, 10 or 50 mM) and is probably due to the hydrophobic nature of the disulfide-bond-stabilized loop. The mucins stored in the mucus granulae are expanded 1000-fold upon release into the lumen. It is difficult to envisage how CysD dimers could exist in the secretory granulae and still allow this large volume





**Figure 8** Proposed model of CysD dimers in the fusion protein and the MUC2 gel

The CysD-IgG fusion protein is made up of two subunits and each is 50 kDa in size. These two subunits are covalently linked via the IgG Fc part (disulfide bridges) and form a covalent 100 kDa dimer. (A) Two covalent 100 kDa CysD-IgG dimers form a non-covalent 200 kDa tetramer via non-covalent interactions of the CysD domain. (B) Four covalent 100 kDa CysD-IgG dimers form a non-covalent 400 kDa octamer via non-covalent interactions of the CysD domain. In both cases the CysD domains form non-covalent dimers. (C) The MUC2 mucin forms a covalent gel via its N- and C-terminal domains. The N-terminus of MUC2 forms covalent trimers, whereas the C-terminus forms covalent dimers. The CysD domain forms non-covalent dimers. The CysD domain inserts non-covalent cross-links into the MUC2 gel thereby determining its pore size and gel properties.

expansion. Although low pH or high calcium conditions do not dissociate the CysD dimers, it is tempting to speculate that the CysD domains exist in a monomeric form within the secretory granules. The putative monomeric state of the CysD domains may be achieved by a 'capping' protein.

As the second CysD domain is localized to the large middle part of MUC2 flanked by the small and large mucin domains [4], this CysD domain will insert additional cross-links into the net-like gel (Figure 8C). Although such a role is theoretical and based on *in vitro* experiments of recombinant proteins, it is tempting to speculate that the CysD may act as a biomolecular 'glue' sticking neighbouring polymer chains together. By glueing adjacent CysD domains together the overall pore size of the mucus gel meshwork will decrease. Interestingly, the three gel-forming mucins MUC2 [40], MUC5AC [14] and MUC5B [15] do not differ in their N-terminal and C-terminal parts, but they differ in the number of CysD domains and how these are distributed within the central part of the molecules. The MUC2 mucin has two CysD domains, whereas MUC5AC has at least nine CysD domains [14] and MUC5B has seven CysD domains [15], interspersed by relatively short mucin domains. The MUC2 mucin has one large mucin domain (approximately 2300 amino acids), which does not contain any additional CysD domains [4]. As the number of and the distance between the CysD domains varies among different gel-forming mucins, gels with different porosity will be produced. This suggests a much denser polymer in MUC5AC and MUC5B than in MUC2. Whereas MUC2 is the major gel-forming mucin of the intestine [4], MUC5AC and MUC5B are the main

components of secreted mucin in the stomach (only MUC5AC) and airways [41]. The different expression patterns of MUC2 and MUC5AC/MUC5B may point to different functions of the gels formed. Analysis of mouse colonic mucus [10] showed that it consists of two layers: a densely packed firm layer devoid of bacteria and a movable loose layer. As both layers are mainly composed of Muc2, these findings indicate a barrier function of the Muc2 colonic mucus probably controlled by the pore sizes. However, the MUC2 mucin still needs to have a relatively large pore size as it should still allow nutrients to pass in the small intestine. In contrast, the mucus of the stomach should only allow ions to pass and in the lungs only gases. It may thus be proposed that denser MUC5AC/MUC5B gels are important for trapping small particles in the lungs before being removed by the mucociliary system. Interestingly, in a previous review by Hollingsworth and Swanson [42] it was discussed that the mucus gel may act like a molecular sieve allowing passage of small molecules, but excluding large molecules and organisms due to steric hindrance.

In conclusion, the CysD domain of human MUC2 forms non-covalent dimers and probably has an important role in the assembly and properties of a mucus gel. However, additional biochemical studies are necessary to shed light on the role of the many CysD domains of other gel-forming mucins.

#### AUTHOR CONTRIBUTION

Daniel Ambort planned, conducted and interpreted the experiments, and wrote the manuscript; Sjoerd van der Post conducted and interpreted the MS experiments; Malin Johansson and Jenny MacKenzie initiated the project and interpreted results; Elisabeth Thomsson produced the recombinant protein; Ute Krenzel supervised Jenny MacKenzie; Gunnar Hansson planned and interpreted the results, and wrote the manuscript together with Daniel Ambort.

#### ACKNOWLEDGEMENTS

We thank The Mammalian Protein Expression (MPE) Core facility for expression and production of recombinant fusion proteins.

#### FUNDING

This work was supported by the Swedish Research Council [grant numbers 7461, 21027, 342-2004-4434]; the Swedish Cancer Foundation; the Swedish Cystic Fibrosis Foundation; the Knut and Alice Wallenberg Foundation [grant number KAW2007.0118]; the IngaBritt and Arne Lundberg Foundation; Sahlgren's University Hospital (LUA-ALF); EU-FP7 IBDase [grant number 200931]; Wilhelm and Martina Lundgren's Foundation; Torsten och Ragnar Söderbergs Stiftelser; and the Swedish Foundation for Strategic Research – The Mucosal Immunobiology and Vaccine Center (MIVAC) and the Mucus-Bacteria-Colitis Center (MBC) of the Innate Immunity Program (2010–2014).

#### REFERENCES

- Perez-Vilar, J. and Hill, R. L. (1999) The structure and assembly of secreted mucins. *J. Biol. Chem.* **274**, 31751–31754
- Thornton, D. J., Rousseau, K. and McGuckin, M. A. (2008) Structure and function of the polymeric mucins in airways mucus. *Annu. Rev. Physiol.* **70**, 459–486
- Lang, T., Hansson, G. C. and Samuelsson, T. (2007) Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16209–16214
- Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J. Biol. Chem.* **269**, 2440–2446
- Li, D., Gallup, M., Fan, N., Szymkowski, D. E. and Basbaum, C. B. (1998) Cloning of the amino-terminal and 5'-flanking region of the human MUC5AC mucin gene and transcriptional up-regulation by bacterial exoproducts. *J. Biol. Chem.* **273**, 6812–6820

- 6 Desseyn, J. L., Aubert, J. P., van Seuning, I., Porchet, N. and Laine, A. (1997) Genomic organization of the 3' region of the human mucin gene MUC5B. *J. Biol. Chem.* **272**, 16873–16883
- 7 Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W. L., Gum, E., Hicks, J. W., Gum, Jr, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) Human gastric mucin. Identification of a unique species by expression cloning. *J. Biol. Chem.* **268**, 5879–5885
- 8 Zhu, L., Lee, P., Yu, D., Tao, S. and Chen, Y. (2010) Cloning and characterization of human MUC19 gene. *Am. J. Respir. Cell Mol. Biol.*, doi:10.1165/rcmb.2010-03120C
- 9 Bobek, L., Tsai, H., Biesbrock, A. and Levine, M. (1993) Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC7). *J. Biol. Chem.* **268**, 20563–20569
- 10 Johansson, M. E., Phillipson, M., Petersson, J., Velcich, A., Holm, L. and Hansson, G. C. (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15064–15069
- 11 Asker, N., Axelsson, M. A., Olofsson, S. O. and Hansson, G. C. (1998) Dimerization of the human MUC2 mucin in the endoplasmic reticulum is followed by a N-glycosylation-dependent transfer of the mono- and dimers to the Golgi apparatus. *J. Biol. Chem.* **273**, 18857–18863
- 12 Lidell, M. E., Johansson, M. E., Mörgelin, M., Asker, N., Gum, Jr, J. R., Kim, Y. S. and Hansson, G. C. (2003) The recombinant C-terminus of the human MUC2 mucin forms dimers in Chinese-hamster ovary cells and heterodimers with full-length MUC2 in LS 174T cells. *Biochem. J.* **372**, 335–345
- 13 Godt, K., Johansson, M. E., Lidell, M. E., Mörgelin, M., Karlsson, H., Olson, F. J., Gum, Jr, J. R., Kim, Y. S. and Hansson, G. C. (2002) The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J. Biol. Chem.* **277**, 47248–47256
- 14 Escande, F., Aubert, J. P., Porchet, N. and Buisine, M. P. (2001) Human mucin gene MUC5AC: organization of its 5'-region and central repetitive region. *Biochem. J.* **358**, 763–772
- 15 Desseyn, J. L., Guyonnet-Dupérat, V., Porchet, N., Aubert, J. P. and Laine, A. (1997) Human mucin gene MUC5B, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family. *J. Biol. Chem.* **272**, 3168–3178
- 16 Pigny, P., Guyonnet-Dupérat, V., Hill, A. S., Pratt, W. S., Galiegue-Zouitina, S., d'Hooge, M. C., Laine, A., Van-Seuning, I., Degand, P., Gum, J. R. et al. (1996) Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. *Genomics* **38**, 340–352
- 17 Lorenzo, P., Neame, P., Sommarin, Y. and Heinegård, D. (1998) Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) identifies a proform including a nucleotide pyrophosphohydrolase. *J. Biol. Chem.* **273**, 23469–23475
- 18 Spada, F., Steen, H., Troedsson, C., Kallesoe, T., Spriet, E., Mann, M. and Thompson, E. M. (2001) Molecular patterning of the oikoplasic epithelium of the larvacean tunicate *Oikopleura dioica*. *J. Biol. Chem.* **276**, 20624–20632
- 19 Löffler, A., Doucey, M. A., Jansson, A. M., Müller, D. R., de Beer, T., Hess, D., Meldal, M., Richter, W. J., Vliegenthart, J. F. and Hofsteenge, J. (1996) Spectroscopic and protein chemical analyses demonstrate the presence of C-mannosylated tryptophan in intact human RNase 2 and its isoforms. *Biochemistry* **35**, 12005–12014
- 20 Perez-Vilar, J., Randell, S. H. and Boucher, R. C. (2004) C-Mannosylation of MUC5AC and MUC5B Cys subdomains. *Glycobiology* **14**, 325–337
- 21 Bäckström, M., Link, T., Olson, F. J., Karlsson, H., Graham, R., Picco, G., Buchell, J., Taylor-Papadimitriou, J. T., Noll, T. and Hansson, G. C. (2003) Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *Biochem. J.* **376**, 677–686
- 22 Stanley, P. (1981) Selection of specific wheat germ agglutinin-resistant (WgaR) phenotypes from Chinese hamster ovary cell populations containing numerous lecR genotypes. *Mol. Cell. Biol.* **1**, 687–696
- 23 Bäckström, D., Hansson, G. C., Nilsson, O., Johansson, C., Gendler, S. and Lindholm, L. (1991) Purification and characterization of a membrane-bound and a secreted mucin-type glycoprotein carrying the carcinoma-associated sialyl-Lea epitope on distinct core proteins. *J. Biol. Chem.* **266**, 21537–21547
- 24 Lämml, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
- 25 Swain, M. and Ross, N. W. (1995) A silver stain protocol for proteins yielding high resolution and transparent background in sodium dodecyl sulfate-polyacrylamide gels. *Electrophoresis* **16**, 948–951
- 26 Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- 27 Schägger, H. and von Jagow, G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**, 223–231
- 28 Andersch-Björkman, Y., Thomsson, K. A., Holmén Larsson, J. M., Ekerhovd, E. and Hansson, G. C. (2007) Large scale identification of proteins, mucins, and their O-glycosylation in the endocervical mucus during the menstrual cycle. *Mol. Cell. Proteomics* **6**, 708–716
- 29 Panchaud, A., Singh, P., Shaffer, S. A. and Goodlett, D. R. (2010) xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J. Proteome Res.* **9**, 2508–2515
- 30 Harris, L. J., Larson, S. B., Hasel, K. W. and McPherson, A. (1997) Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* **36**, 1581–1597
- 31 Andrews, P. (1965) The gel-filtration behaviour of proteins related to their molecular weights over a wide range. *Biochem. J.* **96**, 595–606
- 32 Alhadeff, J. A. (1978) Gel filtration of sialoglycoproteins. *Biochem. J.* **173**, 315–319
- 33 Wang, L. W., Leonhard-Melief, C., Haltiwanger, R. S. and Apte, S. S. (2009) Post-translational modification of thrombospondin type-1 repeats in ADAMTS-like 1/punctin-1 by C-mannosylation of tryptophan. *J. Biol. Chem.* **284**, 30004–300015
- 34 Krieg, J., Gläsner, W., Vicentini, A., Doucey, M. A., Löffler, A., Hess, D. and Hofsteenge, J. (1997) C-Mannosylation of human RNase 2 is an intracellular process performed by a variety of cultured cells. *J. Biol. Chem.* **272**, 26687–26692
- 35 Wu, M. M., Llopis, J., Adams, S., McCaffery, J. M., Kulomaa, M. S., Machen, T. E., Moore, H. P. and Tsien, R. Y. (2000) Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7**, 197–209
- 36 Demaurex, N., Furuya, W., D'Souza, S., Bonifacio, J. S. and Grinstein, S. (1998) Mechanism of acidification of the trans-Golgi network (TGN). *In situ* measurements of pH using retrieval of TGN38 and furin from the cell surface. *J. Biol. Chem.* **273**, 2044–2051
- 37 Miesenböck, G., De Angelis, D. A. and Rothman, J. E. (1998) Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192–195
- 38 Guyton, A. C. and Hall, J. E. (2000) *Medical Physiology*. WB Saunders, Elsevier, Philadelphia
- 39 Paz, H. B., Tisdale, A. S., Danjo, Y., Spurr-Michaud, S. J., Argüeso, P. and Gipson, I. K. (2003) The role of calcium in mucin packaging within goblet cells. *Exp. Eye Res.* **77**, 69–75
- 40 Gum, Jr, J. R., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E. and Kim, Y. S. (1992) The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. *J. Biol. Chem.* **267**, 21375–21383
- 41 Davies, J. R. and Carlstedt, I. (2001) Respiratory tract mucins. *Cilia and Mucus: from Development to Respiratory Defense*. (Salthe, M., ed.), pp. 167–178, Marcel-Dekker, New York
- 42 Hollingsworth, M. A. and Swanson, B. J. (2004) Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60

Received 8 December 2010/10 February 2011; accepted 22 February 2011

Published as BJ Immediate Publication 22 February 2011, doi:10.1042/BJ20102066