# Non-C-mannosylable mucin CYS domains hindered proper folding and secretion of mucin

Valérie Gouyer, Bastien Demouveaux, Guillaume Lacroix, Hélène Valque, Frédéric Gottrand, and Jean-Luc Desseyn*

Inserm, Université de Lille, CHU Lille, LIRIC UMR 995, Lille, France

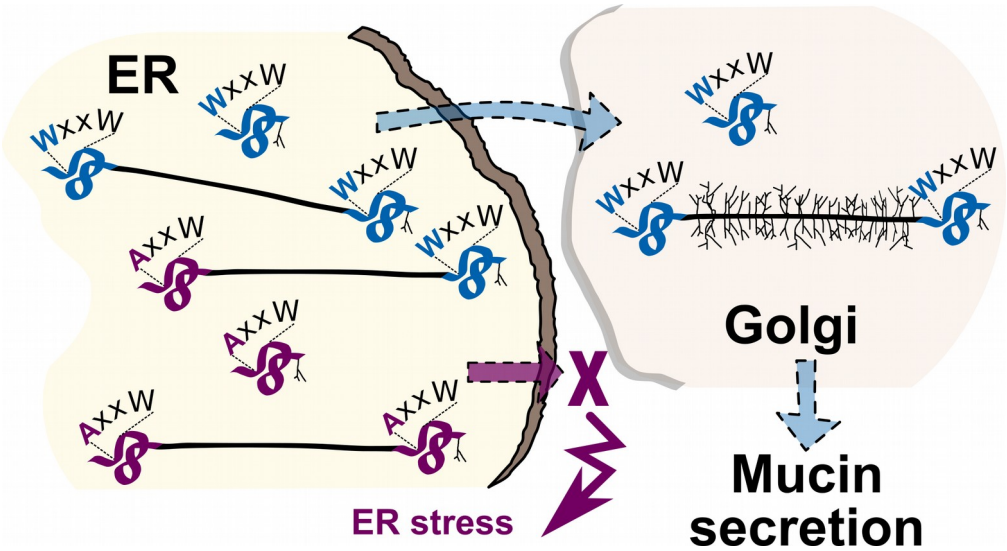
* Corresponding author: Inserm, Université de Lille, CHU Lille, LIRIC UMR 995, F-59000 Lille, France
*Email address:* jean-luc.desseyn@inserm.fr

**ABSTRACT**
The CYS domain occurs in multiple copies in many gel-forming mucins. It is believed that CYS domains can interact with each other in a reversible manner, suggesting a key role of the domain in gel formation. This domain always contains in its amino-terminal sequence the *C*-mannosylation motif WXXW, but whether the CYS domain is *C*-mannosylated is debated, and the putative role of *C*-mannosylation of the domain is unclear. We prepared recombinant CYS domains of the human mucin MUC5B with (WXXW→AXXW) and without a single amino acid mutation and mini-5B mucins made of a large Ser/Thr/Pro region flanked by two CYS domains with the WXXW motif or with the mutated AXXW motif on the first, second or both CYS domains. We found that the single CYS domain and the two CYS domains of mini-5B mucin must be *C*-mannosylable for the efficient maturation and secretion of the recombinant molecules; otherwise, they are retained in the cell and co-localized with a resident enzyme of the endoplasmic reticulum.

**Keywords:** mucin, CYS domain, C-mannosylation

## Highlights

- *C*-mannosylation site WXXW of mucin CYS domains is highly conserved.
- Recombinant CYS domain with mutated *C*-mannosylation site is blocked in the ER.
- Mutation of the WXXW site induces ER stress.
- All CYS domains of a mini-mucin must be *C*-mannosylable.




Biochemical and Biophysical Research Communications 506(4): 812-818 (2018)

https://doi.org/10.1016/j.bbrc.2018.10.138

## 1. Introduction

Among the co- and post-translational modifications of proteins by covalent attachment of carbohydrate, *C*-mannosylation has not been studied thoroughly even though this process appears to be implicated in protein folding and/or trafficking [1]. *C*-mannosylated molecules are found mainly as secreted with variety of cellular functions. *C*-mannosylation involves the covalent binding of an α-mannosylpyranosyl residue to the indole C2 carbon atom of the first Trp in WXXW motifs via a C–C linkage [2]. The replacement of the second Trp with a Phe or Tyr residue of recombinant RNase 2 reduces the efficiency of *C*-mannosylation [3].

Gel-forming mucins harbor in their central structure a large region enriched in Ser/Thr/Pro and carrying numerous oligosaccharidic chains. This region is made of tandemly repeated (TR) sequences [4]. “Naked” domains ~110 amino acid (aa) in length that are rich in Cys residues, called CYS domains, are found upstream of the Ser/Thr/Pro region. These CYS domains also interrupt regions in the secreted mucins MUC2, MUC5B, MUC5AC, their animal orthologues, and numerous secreted mucins in animals [5] and in molecules secreted with *O*-glycoproteins [6]. There are two, seven, and nine CYS domains in human MUC2, MUC5B, and MUC5AC, respectively. Mucin CYS domains always contain in their amino-terminal sequence a WXXW tetrapeptide [7]. Perez-Vilar *et al.* produced recombinant single CYS domains of either MUC5B or MUC5AC to show that this sequence is *C*-mannosylated on the first Trp residue and that the mutation WXXW→AXXW abolishes the *C*-mannosylation and causes retention of the recombinant molecule in the endoplasmic reticulum (ER) [8]. The authors suggested that *C*-mannosylation is probably important in determining the conformation, orientation, and transport at the ER–Golgi interface or in the Golgi, but the precise function of this modification remains unknown. Later, it has been reported that a recombinant CYS domain of MUC2 is not *C*-mannosylated [9]. Today, the role of the C-mannosylation of mucins is not clear and published data suggest that this domain may be engaged in reversible and regulated self-association that may be responsible for the subtle regulated properties of mucus gels [9–16].

Using eukaryotic cell culture and transient expression, we studied here the CYS domain using vectors expressing a recombinant CYS domain of MUC5B and mutated AXXW variant of this CYS domain. Because mucins carrying the CYS domain seem always to possess two or more copies of this domain, we also designed constructs that encode two CYS domains flanking a large Ser/Thr/Pro region of MUC5B (mini-mucin). The CYS domain sequences were the wild type or were mutated on the Trp of the first, second, or both motifs that are *C*-mannosylable. We show that single CYS domain and the two CYS domains of our mini-mucin must be *C*-mannosylable for the proper maturation and secretion; otherwise the recombinant molecules are retained in the ER where they induce ER stress.

## 2. Material and methods

### *2.1. Constructs*

CYS domain-expressing vectors pC and pC* (wild-type CYS domain and mutated WXXW→AXXW CYS domain, respectively) and the mini-mucin mini-5B called hereafter pCtrC (wild-type and mutated pC*trC, pCtrC*, and pC*trC*) were obtained as described in the Supplementary section.

### *2.2. Recombinant molecule analysis*

Transient transfection in COS-7 cells, antibodies used and confocal microscopy are described in the Supplementary materials.

### *2.3. RT-PCR analysis of XBP1 mRNA splicing and real-time PCR analysis of CHOP mRNA*

Methods to study ER stress are described in the Suppl. materials.

## 3. Results

### *3.1. Constructs*

Vectors to express fusion proteins in eukaryotic cells containing the IgK signal peptide followed by the wild-type CYS domain #4 of MUC5B, mutated CYS#4 domain, or mini-5B mucins (mutated and non mutated) were prepared. The schematic representation of the insert of pC encoding the wild-type CYS domain and the nucleotide sequence are depicted in in Suppl. Fig. 1a and b. A Thr/Ser/Ser tripeptide located downstream of the CYS sequence was then mutated into the potential *N*-glycosylation site Asn/Ser/Ser. To study the CYS domain in the context of the mucin structure, the CYS domain must be linked to *O*-glycosylated sequences. In this context, the CYS sequence is flanked by two short peptides that are predicted to be *O*-glycosylated according to the NetOGlyc 3.1 Server [17] (Suppl. Fig. 1b). The alignment of the N-terminal aa sequences of CYS domains from MUC2, MUC5B, MUC5AC, and mouse Muc5b using WebLogo (Suppl. Fig. 1c) point to strong sequence conservation of the *C*-mannosylation sequence, Cys residues, and other aa residues.

### *3.2. The recombinant CYS domain is secreted into the cell culture medium as N– and O-glycosylated molecules*

COS-7 cells were transfected transiently with pC and the cell culture media and cell lysates were analyzed by Western blotting. pC was secreted as expected as different glycoproteins of molecular weights ~42kDa (Suppl. Fig. 2a). The calculated molecular weight of the secreted naked pC was 16kDa, indicating posttransductional modification of the domain that is compatible with glycosylation. pC was also detected in cell lysates as one major protein with molecular mass of 22kDa, which corresponds to the calculated weight of the *N*-glycosylated form. This was confirmed by tunicamycin treatment of pC-transfected COS-7 cells showing a band of 16kDa corresponding to the non *N*-glycosylated pC molecule (Suppl. Fig. 2a). Confocal microscopy analysis showed that pC localized at the cell membrane surface (Suppl. Fig. 2b). Similar images were obtained without and under a permeable condition confirming the localization of pC outside of the cell (Suppl. Fig. 2c). Altogether, these data demonstrated the secretion of pC.

### *3.3. The CYS domain must be C-mannosylable for proper secretion of the domain*

To assess if CYS domain for which its *C*-mannosylation site is abolished is still secreted, a vector expressing mutated **W**XXW→**A**XXW CYS domain (pC*) was prepared. Its schematic structure is depicted in Fig. 1a. The mutation of the first Trp in the *C*-mannosylation site prevented its secretion into the cell culture medium as shown by Western blot (Fig. 1b) and in agreement with others [8]. The non-C-mannosylable pC* was retained in the cell as two proteins with molecular mass of 16 and 22kDa (Fig. 1c); these probably correspond to the naked peptide and the *N*-glycosylated peptide, respectively. Confocal microscopy confirmed that pC was secreted while pC* was not with a perinuclear localization (Fig. 1d). Nonpermeabilized cells transfected with pC* did not stain with the specific antibody (Ab; Fig. 1e). Colocalization experiments by immunofluorescence showed that pC*, but not pC, colocalized with calnexin (Fig. 1f), a resident lectin of the ER. In contrast, both pC* and pC do not colocalized with ST3GalI, a resident glycosyltransferase of the trans-Golgi network (Fig. 1g). Taken together, these findings

demonstrate that the mutation WXXW→AXXW blocks the CYS domain in the ER, and prevents its transport into the Golgi apparatus.

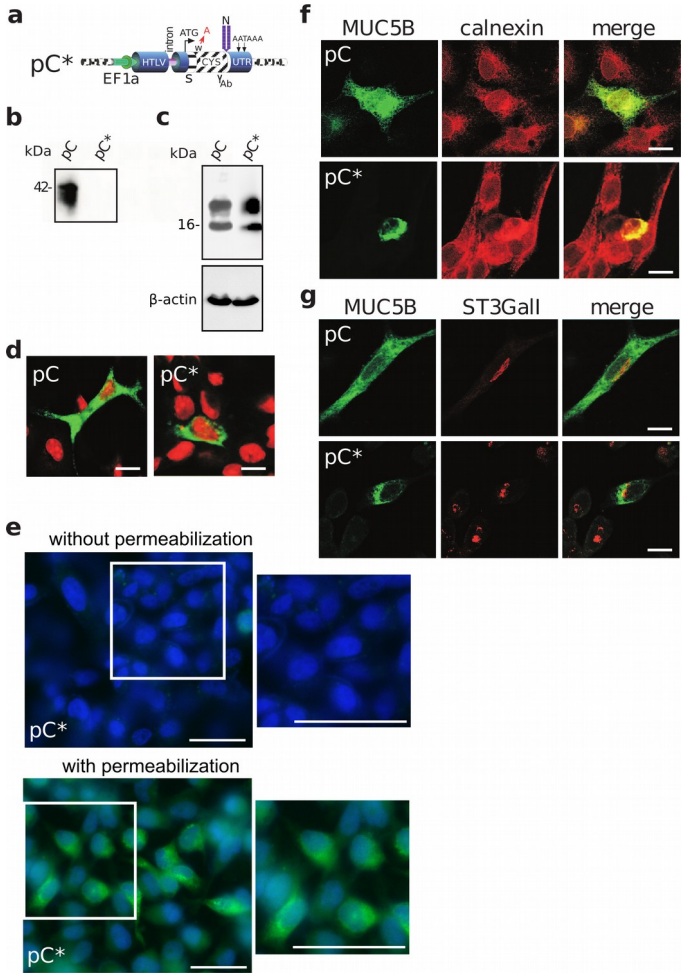



**Figure 1. Requirement of the *C*-mannosylation sequence for the proper folding and secretion of the CYS domain.** (**a**) Schematic representation of the sequences of interest of pC*. ATG initiating Met, *C*-mannosylation sequence (W) and two AATAAA sequences are indicated. S, secretion peptide; ORF, open reading frame; UTR, untranslated region; $Y_{Ab}$, Ab-recognition peptide sequence of MUC5B. N, *N*-glycosylation site. Mutated aa residue is indicated in red (W→A). (**b, c**) Western blot analysis in of cell culture media of COS-7 cells transiently transfected with pC and pC*. **b**-media, pC* was not secreted. **c**-cell lysates, pC was found in two forms. (**d-g**) Transfected cells were studied by confocal microscopy. The nuclei were counterstained by propidium iodide (red) or with Hoechst 33258 (blue). (**d**) Immunostaining with anti-MUC5B antibody (green) showing that pC* exhibited likely perinuclear localization. (**e**) Confocal analysis of transiently transfected COS-7 cells with pC* and analyzed with anti-MUC5B Ab (green) without (top) or with (below) permeabilization confirming that pC* is retained within cells. Nuclei were counterstained with Hoechst 33258 (blue). Scale bar, 50μm. (**f**) Colocalization experiments using Abs directed against calnexin (ER marker) showing that the mutant construct pC* was blocked in the ER. (**g**) pC was secreted and localized to the cell membrane while pC* was retained in the cells and do not colocalized with ST3GalI in the Golgi apparatus. d,f,g: scale bar, 10 μm. e, scale bar, 50μm.

The ER localization of the mutated molecules prompted us to look for possible ER stress. Accumulation of unfolded or misfolded proteins in the ER triggers the unfolded protein response (UPR), also called ER stress. This process is marked by the induction of protein expression or splicing activation of several ER stress-sensor proteins [18]. We looked for activation of two among the three UPR pathways known to be activated after ER stress [19]. The first one is the inositol-requiring protein 1a (IRE1a)-X-box-binding protein 1 (XBP1) pathway where a specific endoribonuclease activity cleaves a small nucleotide segment of *XBP1* mRNA, which we examined using reverse transcript-polymerase chain reaction in COS-7 cells transiently transfected with pC and pC*. The cleaved form of *XBP1* mRNA was found at a higher levels in COS-7 cells producing pC* than in cells transfected producing pC, indicating the activation of ER stress in cells expressing pC* (Suppl. Fig. 3a and b). The second UPR pathway we looked at is the activating transcription factor (ATF)6a, which activates the transcription of CCAAT/enhancer-binding protein C/EBP homologous protein (CHOP) [20].*CHOP* mRNA expression was measured by quantitative PCR. Consistent with an activated unfolded protein response for pC*, CHOP mRNA expression was found to be 4-fold higher (P=0.02) in COS-7 cells transfected with pC* compared with COS-7 cells transfected with pC (Suppl. Fig. 3c).

*3.4. The mini-5B mucin is secreted into the cell culture medium in two major glycoforms*

Because the CYS domain is never found as a single copy in mucin, we constructed an expression plasmid encoding a secreted fusion protein containing two MUC5B CYS-domains flanking a tandem array of 11 irregular repeats from MUC5B followed by a Rend domain (111 aa polypeptide enriched in Ser/Thr/Pro), which we named pCtrC. This recombinant mini-5B/pCtrC (Suppl. Fig. 4a) is representative of a mini-MUC5B that lacks the globular amino- and carboxy-terminal regions necessary for the mucin polymerization. The deduced aa sequence of the Ser/Thr/Pro/ region (Suppl. Fig. 4b) and aa composition of the Ser/Thr/Pro region and the full pCtrC are given (Suppl. Fig. 4c and d) outlining the high content of Ser and Thr in the Ser/Thr/Pro region and the high content of Cys residues in the CYS domain. The pCtrC mini-5B construct depicted in Suppl. Fig. 4e.

COS-7 cells were transiently transfected with the pCtrC construct, and the cell culture medium and cell lysate were analyzed as before by Western blotting. Mini-5B was secreted in at least two major glycoforms of high molecular weight (>160kDa) (Fig. 2a). Mini-5B was also found in an immature form in cell lysates with an apparent molecular weight of ~98kDa corresponding likely to the non-glycosylated form (Fig. 2a). The analysis of the secreted proteins with different lectins showed that the two main bands exhibited different glycosylation patterns: the higher molecular weight glycoform was sialylated as shown by the MAA lectin (Fig. 2b) while the two bands correspond to molecules carrying mucin type carbohydrate T-antigens as shown by the PNA lectin. The secretion of mini-5B was confirmed by confocal microscopy in transiently transfected cells for both permeabilized and nonpermeabilized cells (Fig. 2c and d).

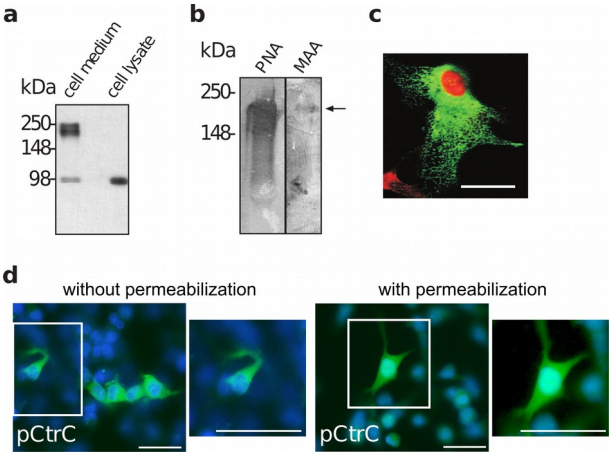



**Figure 2. Secretion of mini-5B (pCtrC) as a highly *O*-glycosylated molecule.** (**a**) Western blot analysis showing that pCtrC was secreted in at least two forms with *Mr* >160kDa. pCtrC was found with an *Mr* ~98kDa in cell lysates. (**b**) pCtrC analyzed by Western blotting using PNA and MAA lectins showing that pCtrC was secreted as two main glycoforms, one of which was sialylated. (**c**) Confocal microscopy showing that pCtrC (green) is secreted. Scale bar 10 µm. (**d**) Confocal analysis of pCtrC (green) without (left panel) or with (right panel) permeabilization confirming that pCtrC is secreted. Nuclei were counterstained with Hoechst 33258 (blue). Scale bar, 50µm.

*3.5. All CYS domains must be C-mannosylable for efficient mucin secretion*

To further evaluate whether the two CYS domains must be C-mannosylable for the mini-mucin secretion, we prepared different mutants of pCtrC, which were mutated on the *C*-mannosylation site of the first CYS domain (pC*trC) or on the *C*-mannosylation site of the second CYS domain (pCtrC*) or on both *C*-mannosylation sites (pC*trC*) (Fig. 3a). Plasmid vectors were transiently transfected into COS-7 cells. Western blotting was first used to analyze the cell culture media and cell lysates. A higher volume/concentration of supernatant from COS-7 cells transfected with constructs encoding all mutated mini-5B, which was estimated to be at least five times higher than the wild-type mini-5B, was loaded in order to confirm that COS-7 cells were efficiently transfected. pCtrC was secreted with a MM >160kDa as previously shown (Fig. 2a). A thin band at >160kDa and a major product at ~98kDa were detected for pC*trC and pCtrC* and corresponds to an immature non-glycosylated mini-5B (Fig. 3b). The mutation of both *C*-mannosylation sites (pC*trC*) fully abolished mucin production (Fig. 3b). All mutant products were found as an immature form of ~98kDa in cell lysates (Fig. 3c) suggesting that the 98kDa band observed in media corresponds to released mini-mucin by cell death. The secretion of pCtrC but not mutated mini-mucins was next confirmed by confocal microscopy, which showed that the pCtrC product localized at the external cell surface membrane (Fig. 3d). However, all mutated mini-5B exhibited mainly a perinuclear localization (Fig. 3d). Absence of secretion of pC*trC was confirmed using nonpermeabilized transfected cells (Fig. 3e, compared with pCtrC in Fig. 2d).

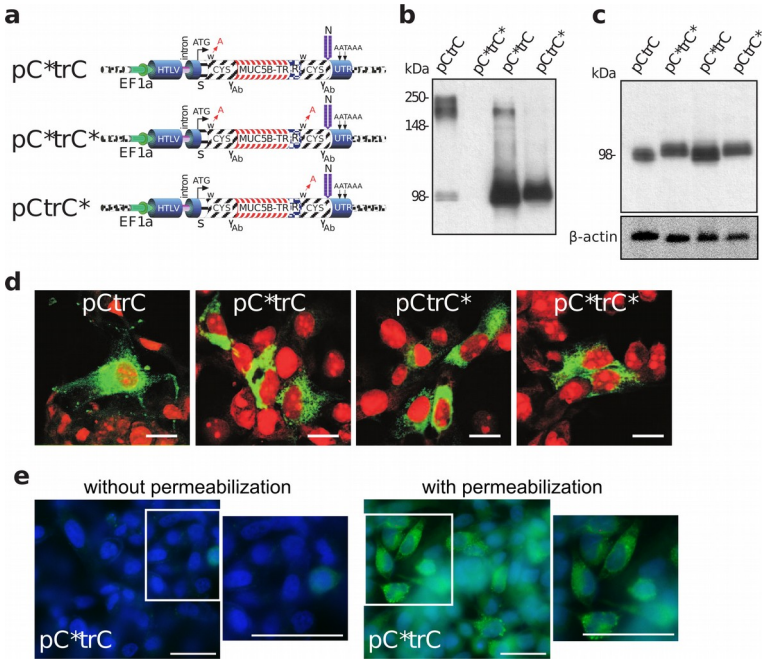



**Figure 3. Both CYS domains of pCtrC must be C-mannosylable for the proper secretion of the recombinant mucin.** (**a**) Schematic representation of the sequences of interest of mutated mini-mucins. Western blot analysis of (**b**) cell culture media and (**c**) cell lysates of COS-7 cells transiently transfected with the expression plasmids pCtrC, pC*trC*, pC*trC, or pCtrC*. In **b**, the 3 mini-5B with mutation were overloaded on purpose. (**d,e**) Confocal microscopy of COS-7 cells transiently transfected with mini-5B constructs. (**d**) Mutated mini-5B showed perinuclear localization. The cells were immunostained with anti-MUC5B Ab (green), and the nuclei were counterstained with propidium iodide (red). Scale bars, 10 µm. (**e**) pC*trC (green) was confirmed to be not secreted as revealed without (left panel) permeabilization of COS-7 cells. Nuclei were counterstained with Hoechst 33258 (blue). Scale bar, 50µm.

To determine where the mutated mni-5B localized within the cell, immunofluorescence experiments were conducted using antibodies directed against RE and Golgi resident proteins. pCtrC did not co-localized with calnexin while each single and the double mutated mini-5B colocalized with calnexin in the ER (Suppl. Fig. 5a) but not with ST3GalI in the Golgi apparatus as shown for pC*trC and pC*trC* (Suppl. Fig. 5b).

## 4. Discussion

The CYS domain is the best-conserved domain in gel-forming mucins, suggesting that it plays a key role [5,7,21,22]. Its sequence seems to have undergone concerted evolution with a high selective pressure to conserve cysteine residues and few other aa, particularly on the *C*-mannosylation consensus sequence located in the amino-terminal part of the domain [5].

Conclusions from published works on the *C*-mannosylation of mucin CYS domains are conflicting. A report using recombinant CYS domains produced by the fibroblastic cell line CHO-K1 Lec 3.2.8, which is deficient in the enzyme GnT-I transferase and the UDP-galactose transporter [23], suggested that the CYS domain is not *C*-mannosylated [9]. By contrast, other data suggest that all

mucin CYS domains are probably *C*-mannosylated. First, the *C*-mannosylation consensus sequence (WXXW/F/Y) is found in all CYS domains reported previously (135 copies in seven species) as shown in Suppl. Fig. 6 according to the consensus sequences of the amino-terminal parts of CYS domains found in all species. In addition, all secreted molecules that have a WXXW/Y/F sequence conserved between species, such as properdin, thrombospondin, and F-spondin, have been shown to be *C*-mannosylated (for review, see [1]). The extreme selective pressure upon the *C*-mannosylation sequence, not forgetting that to the two tryptophan residues correspond a single TGG codon, indicates that the sequence has likely been conserved for the *C*-mannosylation event. Second, at least two MUC5AC (#1 and #5) and three MUC5B CYS domains (#1–3) are *C*-mannosylated according to a report using recombinant CYS domains from these two mucins [8], although the conclusion of the authors relied on the characterization of mutated recombinant CYS domains without any direct biochemical proof concerning the *C*-mannosylation. Mass spectrometry analysis of MUC5B and MUC5AC isolated from human tracheobronchial epithelial (HTBE) cells and Calu-3 cells (airway epithelial cells), respectively, suggested that the first three CYS domains of MUC5B are likely not *C*-mannosylated but the authors could not rule out the presence of the modification at low level and identification of *C*-mannose in other CYS domains of MUC5B and in CYS domains of MUC5AC was not possible with the method used [24]. Third, MUC2 contains three sequences that may be *C*-mannosylated: one in each of its two CYS domains (WSDW and WTGW in MUC2 CYS#1 and #2, respectively) and a third (WGNF) in its amino-terminal region. Gas chromatography/MS analysis of colonic mucins containing mainly or only MUC2 showed that there are three *C*-mannosylation events for one MUC2 molecule, supporting the concept that the three *C*-mannosylation sequences of MUC2 are *C*-mannosylated [25]. This is consistent with our conclusion from our observations of mini-5B that the two CYS domains must be *C*-mannosylable for the secretion of the mini-mucin supporting that all CYS domains are likely *C*-mannosylated for the proper secretion of the mucin. The inconsistency of conclusions drawn on independent studies about the *C*-mannosylation of mucins might suggest that, depending on the biological sample used, *C*-mannosylation of the CYS domain may be required. Coexistence of both *C*-mannosylated and un-*C*-mannosylated molecule has already been reported [26].

Unfortunately, our many attempts to demonstrate using mass spectrometry that our recombinant molecules are or not *C*-mannosylated were unsuccessfull. The main hurdles are the low production of recombinant molecules by transient expression, the low efficiency of purification of the molecules, the predicted high hydrophobicity of the CYS domain and the *N*- and *O*- glycosylation of recombinant molecules that greatly hinder the mass spectrometry analysis.

Obtaining a full recombinant gel-forming mucin is a difficult task because of the extremely large size of mucin mRNA (>15kb). To study the CYS domain in the context of the Ser/Thr/Pro region, we designed a construct encoding a full Ser/Thr/Pro region of MUC5B (440aa) flanked by two CYS domains (110aa). A strong point of all our constructs was the introduction of an *N*-glycosylation site because gel-forming mucins are *N*-glycosylated in their *C*-terminal region [27] and several potential *N*-glycosylation sites are conserved in mucins [28]. *N*-glycans are probably required to maintain a slow folding rate for mucins, which allows proper maturation. *N*-glycosylation is also important for correct protein folding, quality control, sorting, and transport [29,30]. In the first published work on *C*-mannosylation of mucin, recombinant His-tagged CYS domains of MUC5B and MUC5AC were visualized only after cell radiolabeling followed by absorption and purification using the poly-His tag [8] indicating a poor efficiency of production/secretion using COS-7 cells, CHO-Lec31.1 and CHO-K1 cells. Furthermore, when these constructs were fused to the green fluorescent protein sequence, recombinant molecules were found mainly in the Golgi complex. In contrast, our wild-type CYS molecule and mini-5B were secreted exclusively. Taken together, these results show that the tools we designed have all the features required to study the biological role of the CYS domain and the mini-5B mucin. In this line, we reported that mini-5B promotes MCF7 cell proliferation and invasion *in vitro* and tumor growth and tumor cell dissemination *in vivo* using stable MCF7 transfected clones [31].

*C*-mannose may play a structural role in stabilizing *C*-mannosylated molecules and in the proper transport from the ER to the Golgi apparatus via putative lectin recognition [1,8]. Our results demonstrate that recombinant mutated molecules are blocked in the ER. This result supports a key role in the quality control system of nascent protein in the ER. This suggests that mutated recombinant molecules tend to misfold and then to fail the quality control [32], which induces ER stress. If our CYS domains are C-mannosylated, then it is possible that *C*-mannose stabilizes the CYS domain. It is also possible that the *C*-mannose, as suggested for *N*-glycans [33], may contribute substantially to increasing the hydrophilicity of the highly hydrophobic CYS domain, which then inhibits aggregation. An attractive hypothesis is that the *C*-mannose participates actively to the mucin intra- and/or inter-chain bridges as suggested for the *C*-mannosylation of the receptor of IL-21 [34].

## Author's contributions

JLD designed the constructs. HV, BD, GL and VG performed *in vitro* experiments. VG and JLD wrote the manuscript with the critical reading of FG. All authors approved the final manuscript.

## Funding

BD is the recipient of a fellowship from the Région Hauts-de-France and the Centre Hospitalo-Universitaire de Lille. GL is the recipient of a fellowship from the Université de Lille. This work was supported by a the Région Hauts-de-France [grant ARCIr Volet Dynamique 1400324].

## Acknowledgements

We thank D. Swallow and K. Rousseau for the MUC5B monclonal Ab, U. Mandel for the ST3GalI Ab and I. Carlstedt for the anti-MUC5B pAb LUM5B-2.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2018.10.138.

## References

[1] A. Furmanek, J. Hofsteenge, Protein C-mannosylation: facts and questions, Acta Biochim. 47 (2000) 781–789.

[2] J. Hofsteenge, D.R. Muller, T. de Beer, *et al*., New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us, Biochemistry. 33 (1994) 13524–13530.

[3] J. Krieg, S. Hartmann, A. Vicentini, *et al*., Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp, Mol. Biol. Cell. 9 (1998) 301–9.

[4] B. Demouveaux, V. Gouyer, F. Gottrand, *et al*., Gel-forming mucin interactome drives mucus viscoelasticity, Adv. Colloid Interface Sci. 252 (2018) 69–82. doi:10.1016/j.cis.2017.12.005.

[5] J.-L. Desseyn, Mucin CYS domains are ancient and highly conserved modules that evolved in concert, Mol. Phylogenet. Evol. 52 (2009) 284–292. doi:10.1016/j.ympev.2009.03.035.
[6] F. Spada, H. Steen, C. Troedsson, *et al*., Molecular patterning of the oikoplastic epithelium of the larvacean tunicate Oikopleura dioica, J Biol.Chem. 276 (2001) 20624–20632.
[7] J.-L. Desseyn, J.-P. Aubert, N. Porchet, A. Laine, Evolution of the large secreted gel-forming mucins, Mol. Biol. Evol. 17 (2000) 1175–1184.
[8] J. Perez-Vilar, S.H. Randell, R.C. Boucher, C-Mannosylation of MUC5AC and MUC5B Cys subdomains, Glycobiology. 14 (2004) 325–337.
[9] D. Ambort, S. van der Post, M.E.V. Johansson, *et al*., Function of the CysD domain of the gel-forming MUC2 mucin, Biochem. J. 436 (2011) 61–70. doi:10.1042/BJ20102066.
[10] R. Bansil, B.S. Turner, Mucin structure, aggregation, physiological functions and biomedical applications, Curr. Opin. Colloid Interface S. 11 (2006) 164–170.
[11] L.E. Bromberg, D.P. Barr, Self-association of mucin, Biomacromolecules. 1 (2000) 325–334.
[12] R. Brunelli, M. Papi, G. Arcovito, *et al*., Globular structure of human ovulatory cervical mucus, FASEB J. 21 (2007) 3872–3876.
[13] X. Cao, R. Bansil, K.R. Bhaskar, *et al*., pH-dependent conformational change of gastric mucin leads to sol-gel transition, Biophys. J. 76 (1999) 1250–1258.
[14] Z. Hong, B. Chasan, R. Bansil, *et al*., Atomic force microscopy reveals aggregation of gastric mucin at low pH, Biomacromolecules. 6 (2005) 3458–3466.
[15] A. Maleki, G. Lafitte, A.L. Kjoniksen, *et al*., Effect of pH on the association behavior in aqueous solutions of pig gastric mucin, Carbohydr. Res. 343 (2008) 328–340.
[16] J. Perez-Vilar, R.C. Boucher, Reevaluating gel-forming mucins' roles in cystic fibrosis lung disease, Free Radic. Biol. Med. 37 (2004) 1564–1577.
[17] K. Julenius, A. Mølgaard, R. Gupta, S. Brunak, Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites, Glycobiology. 15 (2005) 153–64. doi:10.1093/glycob/cwh151.
[18] H. Yoshida, M. Oku, M. Suzuki, K. Mori, pXBP1(U) encoded in XBP1 pre-mRNA negatively regulates unfolded protein response activator pXBP1(S) in mammalian ER stress response, J. Cell Biol. 172 (2006) 565–75. doi:10.1083/jcb.200508145.
[19] M. Wang, R.J. Kaufman, Protein misfolding in the endoplasmic reticulum as a conduit to human disease, Nature. 529 (2016) 326–35. doi:10.1038/nature17041.
[20] H. Zinszner, M. Kuroda, X. Wang, *et al*., CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum, Genes Dev. 12 (1998) 982–95.
[21] J.-L. Desseyn, M.-P. Buisine, N. Porchet, *et al*., Evolutionary history of the 11p15 human mucin gene family, J. Mol. Evol. 46 (1998) 102–106. doi:10.1007/PL00006276.
[22] T. Lang, G.C. Hansson, T. Samuelsson, Gel-forming mucins appeared early in metazoan evolution, Proc. Natl. Acad. Sci. USA. 104 (2007) 16209–16214.
[23] S. Oelmann, P. Stanley, R. Gerardy-Schahn, Point mutations identified in Lec8 Chinese hamster ovary glycosylation mutants that inactivate both the UDP-galactose and CMP-sialic acid transporters, J. Biol. Chem. 276 (2001) 26291–300. doi:10.1074/jbc.M011124200.
[24] R. Cao, T.T. Wang, G. DeMaria, *et al*., Mapping the protein domain structures of the respiratory mucins: a mucin proteome coverage study, J.Proteome.Res. 11 (2012) 4013–4023.
[25] J.-P. Zanetta, A. Pons, C. Richet, *et al*., Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins, Anal. Biochem. 329 (2004) 199–206.
[26] Y. Goto, Y. Niwa, T. Suzuki, *et al*., C-mannosylation of human hyaluronidase 1: possible roles for secretion and enzymatic activity, Int. J. Oncol. 45 (2014) 344–50. doi:10.3892/ijo.2014.2438.
[27] D.J. Thornton, K. Rousseau, M.A. McGuckin, Structure and function of the polymeric mucins in airways mucus, Annu. Rev. Physiol. 70 (2008) 459–486.
[28] J.-L. Desseyn, J.-P. Aubert, I. Van Seuningen, *et al*., Genomic organization of the 3' region of the human mucin gene MUC5B, J. Biol. Chem. 272 (1997) 16873–16883. doi:10.1074/jbc.272.27.16873.
[29] S.L. Bell, G. Xu, I.A. Khatri, *et al*., N-linked oligosaccharides play a role in disulphide-dependent dimerization of intestinal mucin Muc2, Biochem. J. 373 (2003) 893–900. doi:10.1042/BJ20030096.
[30] A. Helenius, M. Aebi, Intracellular functions of N-linked glycans, Science. 291 (2001) 2364–9.
[31] H. Valque, V. Gouyer, F. Gottrand, J.-L. Desseyn, MUC5B Leads to Aggressive Behavior of Breast Cancer MCF7 Cells, PLoS One. 7 (2012) e46699. doi:10.1371/journal.pone.0046699.
[32] E.S. Trombetta, A.J. Parodi, Quality Control and Protein Folding in the Secretory Pathway, Annu. Rev. Cell Dev. Biol. 19 (2003) 649–676. doi:10.1146/annurev.cellbio.19.110701.153949.
[33] L.W. Ruddock, M. Molinari, N-glycan processing in ER quality control, J. Cell Sci. 119 (2006) 4373–4380. doi:10.1242/jcs.03225.
[34] O.J. Hamming, L. Kang, A. Svensson, *et al*., Crystal structure of interleukin-21 receptor (IL-21R) bound to IL-21 reveals that sugar chain interacting with WSXWS motif is integral part of IL-21R, J. Biol. Chem. 287 (2012) 9454–60. doi:10.1074/jbc.M111.311084.

**Supplementary Materials**

**Constructs**

All vector construction PCRs were performed using Pfu Turbo (Stratagene/Agilent Technologies, Massy, France). Primers included appropriate restriction cleavage sites to allow further insertion of the PCR products into the expression vectors. PCR products were cloned into the pMosBlue Blunt vector (Amersham biosciece/GE Healthcare Velizy-Villacoublay, France). Cloned PCR fragments were excised from pMosBlue Blunt with the appropriate restriction enzymes (New England Biolabs, Evry, France) to be subcloned. CYS domain-expressing vectors pC and pC* (wild-type CYS domain and mutated CYS domain on the *C*-mannosylation site, respectively) and the mini-mucin mini-5B called hereafter pCtrC (wild-type and mutated pC*trC, pCtrC*, and pC*trC*) were obtained using the following strategy. Briefly, the cytomegalovirus immediate early promoter of the pcDNA3.1 Neo+ (Invitrogen/Life technologies, Villebon sur Yvette, France) carried by the 940bp genomic sequence *Bgl*II-*Eco*RI was replaced by a 845bp PCR-amplified EF1α-HTLV composite promoter from pMG (InvivoGen, Toulouse, France) flanked by the two unique *Bam*HI and *Eco*RI restriction sites, producing the expressing vector pcMG. The BEN2 genomic clone containing the large central exon of MUC5B [1] was used to amplify the two intronless fragments coding either for the CYS#4 domain alone (fragment C) or for CYS#4 followed by a long Ser/Thr/Pro-rich region made of 11 irregular TRs of 29 amino acid (aa) residues and a Rend domain of 111 aa residues (fragment Ctr). The CYS#4 of MUC5B was chosen because specific antibodies against a peptide sequence belonging to this domain (see Antibodies section below) were available. We mutated the sequence encoding the tripeptide TSS of the CYS sequence to the *N*-glycosylation site NSS using the QuickChange Site-Directed Mutagenesis protocol from Stratagene using the two oligonucleotide sequences 5'-CCAGCACCCCGGCCA**A**CAGCTCTACGGCCAC-3' (forward) and 5'-TGGCCGTAGAGCTG**T**TGGCCGGGGTGCTGGG-3' (reverse). The fragment carrying the CYS sequence was subcloned into the pSecTagB vector (Invitrogen) in frame with the IgK leader chain. The IgK-CYS sequence was then subcloned into the pcMG vector (expression vectors pC and $pCH_6$, respectively).

The *C*-mannosylation site (WSEW) of pC was changed to ASEW (vector pC*) using two oligonuceotides with the following sequence: 5'-TGAGCCCCAGTGTGCC**GC**GTCAGAGTGGCT-3' (forward) and 5'-AGCCACTCTGAC**GC**GGCACACTGGGGCTCA-3' (reverse). The amplified genomic fragment Ctr was also mutated on the *C*-mannosylation site (WXXW→AXXW; fragment C*tr). The wild-type and mutated fragments were subcloned in frame between the signal peptide and the CYS sequence from the pC and the pC* vectors, producing the four expression vectors pCtrC, pC*trC, pCtrC*, and pC*trC*.

**Antibodies**

Polyclonal antibodies (pAb) against human calnexin (H70) and β-actin (ab8227) were purchased from Santa Cruz (Santa Cruz Biotechnologies, Heidelberg, Germany) and abcam (Cambridge, UK), respectively. Mouse monoclonal antibodies (mAbs) against MUC5B (Eu1 and Eu2) and ST3GalI (4B10) were generous gifts from D. Swallow and K. Rousseau [2] from the European Consortium (Concerted Action contract number BMH4-CT98-3222) and U. Mandel [3] (School of Dentistry, Copenhagen, Denmark), respectively. The rabbit anti-MUC5B pAb LUM5B-2 [4] was a gift from I. Carlstedt (Lund University, Sweden).

**Cell culture and transient transfection**

COS-7 cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal calf serum, 200mM L-glutamine, and 100U penicillin/streptomycin (Invitrogen). Transient transfection of COS-7 cells was performed using Effectene (Qiagen, Courtaboeuf, France). Cells were seeded in standard medium in 10mm Petri dishes. Cells were transfected using 2µg of plasmid according to the manufacturer's instructions when cells reached 70–80% confluency. Twenty-four hours after transfection, cells were cultured in serum-free medium containing 0.01% bovine serum albumin. The cell culture media and cells were collected 48 h after transfection, and the cell culture media were concentrated in a Centricon-3 concentrator (Amicon, Bedford, MA). For tunicamycin treatment, cells were transfected with the pC expression plasmid, and 24 h later, the cells were treated with 10µg/mL tunicamycin for 1 h before harvest.

**Confocal microscopy**

Confocal microscopy was performed on transiently transfected COS-7 cells. The cells were fixed in 4% paraformaldehyde for 20 min, quenched for 30 min with 50mM $NH_4Cl$ in phosphate-buffered saline (PBS), and permeabilized with 0.2% saponin in PBS for 20 min. The saturation step was performed for 30 min in PBS containing 1% bovine serum albumin and 0.2% saponin. To detect MUC5B, ST3GalI, and calnexin, mAbs Eu1 and Eu2 (diluted 1/600), 4B10 (1/4), and rabbit pAb H-70 (1/200), respectively, were added overnight. LUM5B was used for double labeling of MUC5B/calnexin and MUC5B/ST3GalI. For single labeling, the cells were incubated with fluorescein isothiocyanate-conjugated goat anti-mouse Ab for 90 min, and the nuclei were counterstained with propidium iodide or Hoechst 33258. For double labeling, fluorescein isothiocyanate-conjugated goat anti-rabbit Ab and Texas-Red conjugated goat anti-mouse Ab (Jackson Immunoresearch, Interchim, Montluçon, France) were used. In order to confirm the secretion and the localization at the membrane of the recominant molecules, immunolabeling was performed as described above but without permeabilization, i.e. in absence of saponin. Confocal microscopy was performed using a DMIRBE microscope (model TCS-NT; Leica) with a 40x 1.32 Plan-Apochromat oil-immersion objective lens. Acquisition was performed using PowerScan software (Leica) and processed with Adobe Photoshop CS.

**Protein detection by Western blot and lectin blot analysis**

The cell culture medium was collected (5mL) and concentrated to 200µL by centrifugation in a Centricon-3 tube (Amicon Millipore, Molsheim, France). Cells were washed five times with PBS and resuspended in 200µL PBS containing 0.2mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride. The cells were freeze-thawed and sonicated, and the protein concentration was measured using the BCA protein assay (Pierce Biotechnology, Rockford, IL). Protein lysates (usually 30 µg) and supernatants (usually 20 µL) were loaded onto an 8% or 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using a vertical gel apparatus (Bio-rad, Marnes-la-Coquette, France), electrophoresed, and then transferred to a nitrocellulose membrane. To analyze the mini-mucins, the supernatant of mutated mini-5B was concentrated to be at least five times the

concentration of the wild-type mini-5B. The membrane was blocked with 5% powdered milk in PBS/0.1% Tween 20 (TPBS) overnight at 4°C, washed, and probed with mAb against MUC5B diluted 1:600 in TPBS. After washing, the membrane was incubated for 45 min with horseradish peroxidase-conjugated goat anti-mAb (Santa Cruz Biotechnology) diluted 1:2000 in TPBS. The membrane was then washed for 30 min in TPBS, and an ECL™-Plus chemiluminescent detection kit (Amersham Biosciences, GE Healthcare) was used for visualization (5 min). To normalize protein expression the membrane with cell lysates was then probed with pAb against β-actin diluted 1:3000 in TPBS. For lectin blots, after blotting, the nitrocellulose membrane was treated for 2 h with 2% polyvinylpyrrolidone K 30 in 10mM Tris-HCl, pH 7.4, containing 0.15M NaCl (TBS). The membrane was incubated in digoxigenin-labeled MAA lectins, which recognize the oligosaccharide species Neu5Acα2-3Gal-R or peanut *Arachis hypogaea* (PNA), which recognizes Galα1-3GalNAc-R (Antigen-T) at a concentration of 5 μg/mL in TBS. The nitrocellulose membrane was incubated for 1 h with alkaline phosphatase-labeled anti-digoxigenin Fab fragment (1μg/mL in TBS). The labeled glycoproteins were revealed by 4-nitroblue tetrazolium chloride/5-bromo-4-chloro- 3-indolyl phosphate staining.

**RT-PCR analysis of XBP1 mRNA splicing and real-time PCR analysis of CHOP mRNA**

Total RNA was extracted and reverse transcribed as before [5]. To detect spliced *XBP1* mRNA, the PCR conditions were 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min for 30 cycles using the primer pair 5′-CCTTGTAGTTGAGAACCAGG-3′ and 5′-GGGGCTTGGTATATATGTGG-3′, which yielded a 424bp unspliced *XBP1* fragment and a 442bp spliced *XBP1* fragment. G3PDH amplification was used as endogenous control (Clontech, Saint-Germain-en-Laye, France). PCR amplifications were analyzed by electrophoretic migration on an 8% acrylamide gel and stained with ethidium bromide. Band intensity was analyzed using ImageJ (NIH, Bethesda, MD, USA) and the intensity of the spliced-form was normalized to the intensity of the unspliced form. For quantitative PCR analysis of *CHOP* mRNA, primer and TaqMan probe sequences were selected using the Primer3 Output program (MIT) (Primer3 freeware v0.4.0) with the Macaca *CHOP* cDNA sequence (*Macaca mulatta* gene ENSMMUG00000011286 (DDIT3)). The specific primers and probe for *CHOP* were as follows: macacaCHOP forward primer, 5'-GCCAAAATCAGAGCTGGAAC-3' belonging to exon 1 and macacaCHOP reverse primer, 5'-GCTTTCAGGTGTGGTGATGT-3' belonging to exon 2; and macacaCHOP probe, 5'-GGAGAGAGTGTTCAAGAAGGAAGTGTA-3' overlapping the exon 1–exon 2 junction to avoid amplifying contaminating genomic DNA. The 18s rRNA was chosen as the internal positive control. PCR analysis was performed as described previously. All samples were measured in triplicate. The cycle threshold values of all samples were measured using the ABI Prism 7700 sequence detector system (Applied Biosystems), and results are expressed using the $2^{-\Delta\Delta Ct}$ method relative to the changes in gene transcription. For each sample, the ΔΔCt value was obtained by subtracting the ΔCt value for 18s rRNA from that of the *CHOP* mRNA, and the ΔΔCt was obtained by subtracting the ΔCt value of the pC-transfected COS-7 cells from that of the pC mutant-transfected COS-7 cells. Data are expressed as mean ± standard deviations of triplicate determination. Statistical significance was evaluated using the Student's t-test for unpaired comparison. Difference with $p < 0.05$ was considered to be statistically significant.

**Supplementary References**

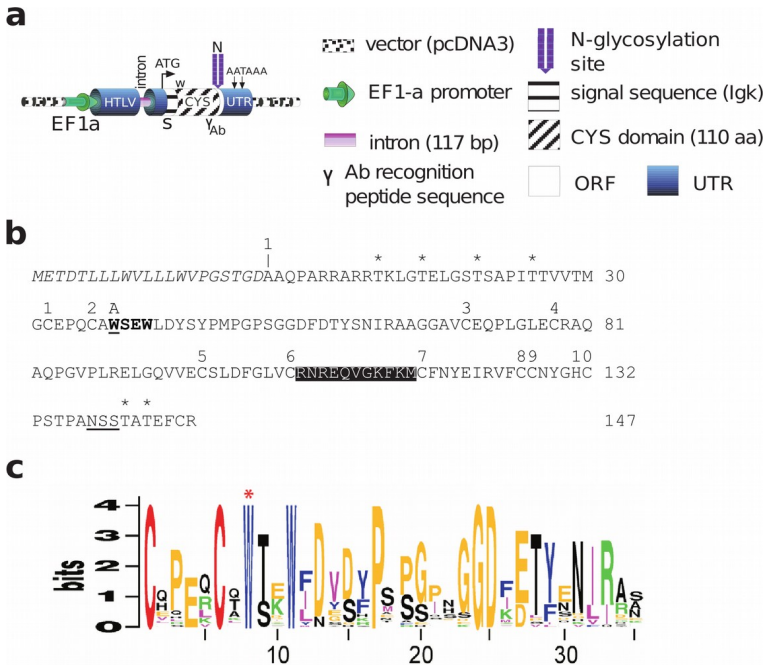

[1] J.-L. Desseyn, V. Guyonnet-Dupérat, N. Porchet, *et al*., Human mucin gene MUC5B, the 10.7-kb Large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family, J. Biol. Chem. 272 (1997) 3168–3178. doi:10.1074/jbc.272.6.3168.

[2] K. Rousseau, C. Wickstrom, D.B. Whitehouse, *et al*., New monoclonal antibodies to non-glycosylated domains of the secreted mucins MUC5B and MUC7, Hybrid.Hybridomics. 22 (2003) 293–299.

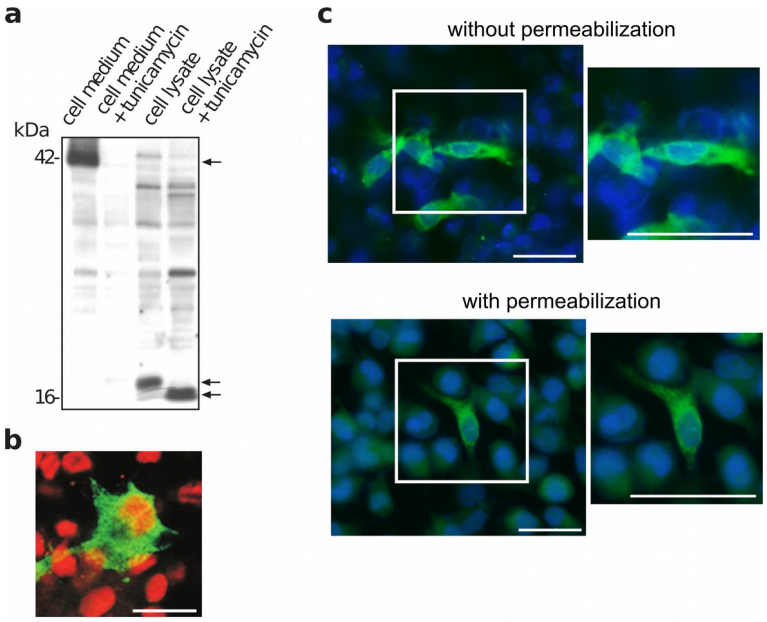
[3] V. Vallejo-Ruiz, R. Haque, A.M. Mir, *et al*., Delineation of the minimal catalytic domain of human Galbeta1-3GalNAc alpha2,3-sialyltransferase (hST3Gal I), Biochim. Biophys. Acta. 1549 (2001) 161–73.

[4] C. Wickstrom, J.R. Davies, G. V Eriksen, *et al*., MUC5B is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage, Biochem.J. 334 (1998) 685–693.

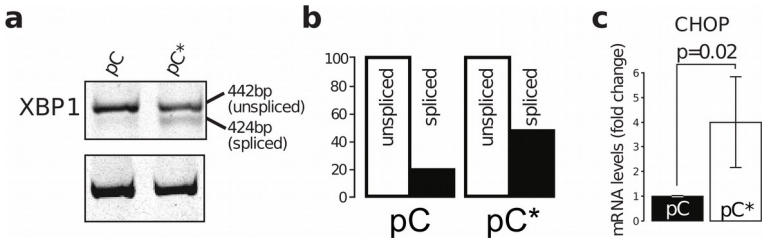
[5] J.-L. Desseyn, D. Tetaert, V. Gouyer, Architecture of the large membrane-bound mucins, Gene. 410 (2008) 215–222. doi:10.1016/j.gene.2007.12.014.



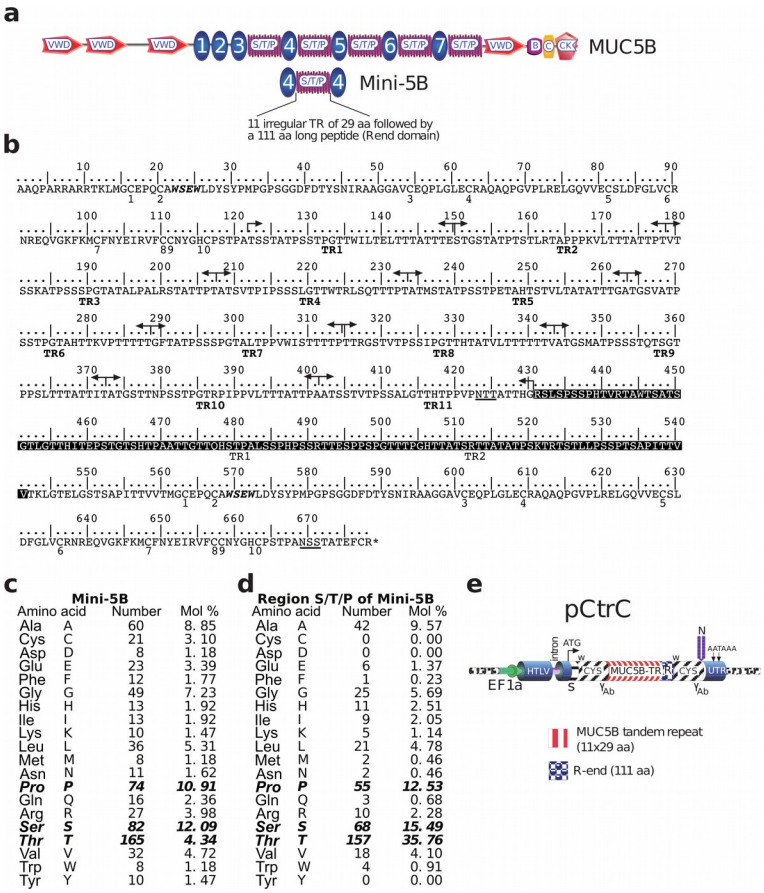

**Suppl. Figure 1. Amino acid sequence of pC and pC*.** (**a**) Schematic representation of the sequence of interest in the expression vector pC. ATG initiating Met, *C*-mannosylation sequence (W) and two AATAAA sequences are indicated. S, secretion peptide; ORF, open reading frame; UTR, untranslated region; $Y_{Ab,}$ Ab-recognition peptide sequence of MUC5B. (**b**) The deduced aa sequence of the recombinant CYS domain (vector pC) was 168aa long with a 21-residue peptide signal (in italics). The predicted released peptide is 147 residues (1–147). The consensus *C*-mannosylation sequence (WSEW) is in bold, and the first Trp in the wild-type sequence (vector pC) that was replaced by Ala using directed mutagenesis (vector pC*) is indicated. The Cys residues are numbered 1–10 in bold. The Ab-recognition peptide sequence (MUC5B antigen, 106–116) is within black box. The potential *N*-glycosylation site (NSS) introduced by directed mutagenesis (TSS→NSS) is underlined. Asterisks, predicted *O*-glycosylated aa. (**c**) The alignment sequences of amino-terminal peptides of the two CYS domains of MUC2, seven CYS domains of MUC5B, nine CYS domains of MUC5AC, and 10 CYS domains of mouse Muc5b were used to generate a WebLogo outlining the strong conservation of Cys and Trp residues and showing the W residue mutated in this work (red asterisk).



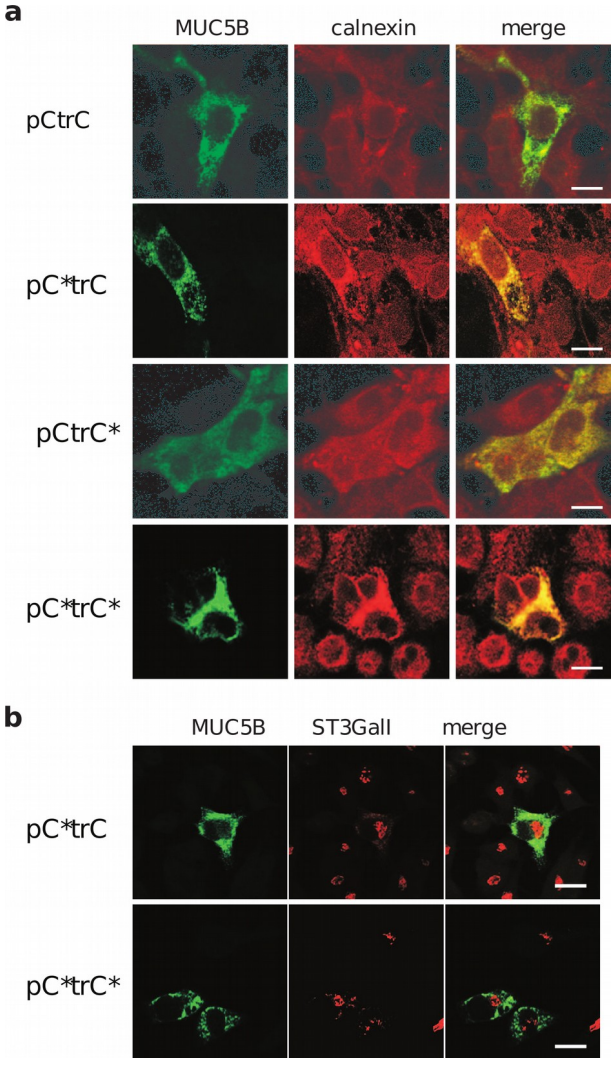

**Suppl. Figure 2. Secretion of recombinant CYS domain as *N*- and *O*-glycosylated molecules.** (**a**) Cell culture media and cell lysates of transiently transfected COS-7 cells with the plasmid pC were analyzed by Western blotting and 15% SDS-PAGE, and show that the molecule was secreted with a *Mr* of ~42kDa. After 1 h tunicamycin treatment, the molecule was blocked in the ER with a *Mr* of ~16kDa, which is consistent with the calculated *Mr*. Without tunicamycin treatment, the CYS domain in the cell lysate had a *Mr* of ~22 kDa, which is consistent with the predicted molecular weight of the *N*-glycosylated pC molecule. (**b**) Confocal analysis of transiently transfected COS-7 cells with pC. Cells were immunostained with anti-MUC5B Ab (green), and the nuclei were counterstained with propidium iodide (red). Scale bar 10 µm. (**c**) Confocal analysis of transiently transfected COS-7 cells with pC and analyzed with anti-MUC5B Ab (green) without (top) or with (below) permeabilization of cells confirming that pC is secreted and visualized as sticking on the whole cell surface. Nuclei were counterstained with Hoechst 33258 (blue). Scale bar, 50µm.



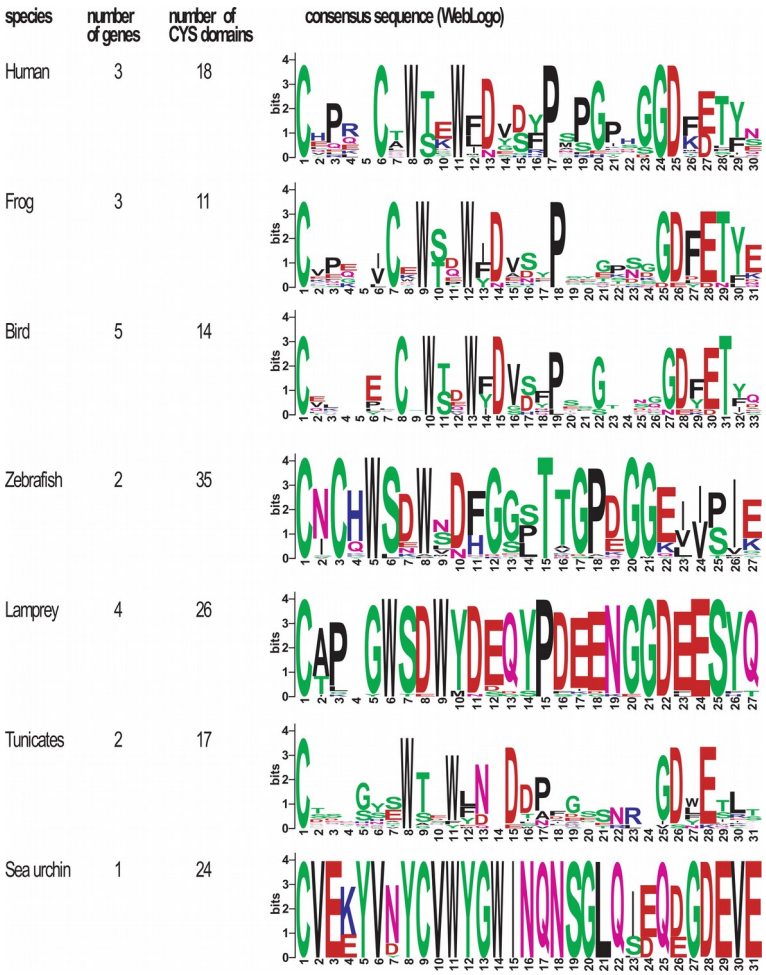

**Suppl. Figure 3. ER stress induced by mutation of the CYS domain.** Total RNA from cells transfected with pC or pC* was analyzed. (**a,b**) The results show an increase of the amount of the spliced form of XBP1 (424bp) in cells transfected with the mutated construct (pC*). The efficiency of the cDNA synthesis was estimated by PCR using G3PDH-specific primers, which produced one band at 931bp. (**c**) mRNA levels of CHOP were measured by RT-qPCR (Taqman). Results are means ± standard deviation.




**Suppl. Figure 4. Mini-5B mucin.** (**a**) Modular organization of the complete mucin structure of MUC5B compared with the mini-5B mucin. vWD, von Willebrand D domain. B and C are two domains found in the von Willebrand factor; CK, cystine knot domain; the ovals numbered 1–7 represent the CYS domains ; Ser/Thr/Pro, highly *O*-glycosylated region. (**b**) Amino acid sequence of the predicted mature mini-5B polypeptide (pCtrC). Two identical CYS domains (CYS#4 of MUC5B) with their *C*-mannosylable sequences (bold and italic) flank a large Ser/Thr/Pro region of MUC5B made of 11 irregular TRs (numbered in bold and delimited by arrows) followed by a 111 aa Rend domain (black box). The 10 Cys residues of each CYS domain are numbered. The *N*-glycosylation site NSS introduced by mutagenesis and the NTT tripeptide sequence found in the 11th TR and the one added by mutagenesis are underlined. The mature recombinant mini-5B has a calculated molecular weight of 69.4 kDa. (**c**) Amino acid composition of the full mini-5B in comparison with (**d**) of its Ser/Thr/Pro region. Ser, Thr Pro contents are in italic and bold. (**e**) Schematic representation of the expression vector construct for the Mini-5B. For explanation of the symbols, see the legend of Suppl. Fig. 1a.




**Suppl. Figure 5. Mutated Mini-5B are blocked in the ER.** (**a**) Confocal analysis of transiently transfected COS-7 cells with pCtrC, pC*trC, pCtrC* and pC*trC* using an anti-MUC5B (green) and anti-calnexin antibodies (in red). All mutated mini-5B were retained within the cell and co-localized (yellow) with calnexin in the ER while the control wild-type mini-5B pCtrC was secreted by cells. (**b**) Pictures illustrating that mini-5B (in green) with a **W**XXW→**A**XXW mutation in its first CYS domain or the mutation within its two CYS domains were retained in the cells but did not co-localized with ST3Gall (red) in the Golgi apparatus. Scale bars, 10µm.




**Suppl. Figure 6. CYS domains containing the *C*-mannosylation sequence WXXW in their amino-terminal region.** CYS sequences were used to generate a WebLogo for each species. The number of CYS domains found in each species and the number of genes (human and one tunicate gene) or putative genes encoding multiple copies of CYS domains are indicated. Two CYS sequences from oikosine-1 secreted by a tunicate contained the WXXY and WXXF peptide sequences that could be *C*-mannosylated. Two CYS sequences from the zebrafish contained the WXXP/A sequences that could not be *C*-mannosylated.