Membrane-bound mucin modular domains: from structure to function.

Nicolas Jonckheere, Nicolas Skrypek, Frédéric Frénois, Isabelle Van Seuningen

Membrane-bound mucin modular domains: from structure to function

Nicolas Jonckheere[1, 2, 3], Nicolas Skrypek[1, 2, 3], Frédéric Frénois[1, 2, 3] and Isabelle Van Seuningen[1, 2, 3]

1 Inserm, UMR837, Jean Pierre Aubert Research Center, Team #5 "Mucins, epithelial differentiation and carcinogenesis", rue Polonovski, 59045 Lille Cedex, France

2 Université Lille Nord de France, 1 Place de Verdun, 59045 Lille cedex, France

3 Centre Hospitalier Régional et Universitaire de Lille, Place de Verdun, 59037 Lille cedex, France.

Corresponding author : Dr Nicolas Jonckheere, Inserm UMR837/JPARC, rue Polonovski, 59045 Lille Cedex, France, Phone: +33 3 20 29 88 50, Fax: +33 3 20 53 85 62, E-Mail: nicolas.jonckheere@inserm.fr

Abstract

Mucins belong to a heterogeneous family of large *O*-glycoproteins composed of a long peptidic chain called apomucin on which are linked hundreds of oligosaccharidic chains. Among mucins, membrane-bound mucins are modular proteins and have a structural organization usually containing Pro/Thr/Ser-rich *O*-glycosylated domains (PTS), EGF-like and SEA domains. *Via* these modular domains, **the membrane-bound** mucins participate in cell signalling and cell interaction with their environment in normal and pathological conditions. Moreover, the recent knowledge of these domains and their biological activities led to the development of new therapeutic approaches involving mucins. In this review, we show 3D structures of EGF and SEA domains. We also describe the functional features of the evolutionary conserved domains of **membrane-bound** mucins and discuss consequences of splice events.

Keywords: mucin, evolution, structure-function, EGF-like, 3D structure

1. Introduction

Mucins belong to a heterogeneous family of large *O*-glycoproteins composed of a long peptidic chain called apomucin on which are linked hundreds of oligosaccharidic chains. Based on biochemical and molecular biology studies, mucins were separated into two structurally distinct **categories**: the secreted and the membrane-bound mucins. Mucins have a cell- and tissue-specific pattern of expression that is profoundly altered in epithelial cancers (loss of expression, over-expression, aberrant expression, neo-expression) suggesting their implication in tumourigenesis [1, 2]. They play roles in cell signalling, cell proliferation, tumour progression or cell polarity, and mediate immune evasion. Mucins are considered as potent new therapeutic targets in mucosal biology, in malignant and inflammatory diseases of the epithelial tissues [3-6].

Secreted mucins, gel-forming components of viscoelastic mucus gels protecting the epithelia, include **the 11p15 secretory mucins**: MUC2, MUC5AC, MUC5B, MUC6. **Next to that family of genes, a fifth secretory mucin was described that is MUC19.** Their main function is to participate in mucus formation by forming a three-dimensional network *via* oligomerization domains in order to protect underlying epithelia against various injuries (inflammation, bacteria, virus, pollutants, pH, etc). MUC7 and MUC9 are smaller secreted mucins that do not oligomerize and do not form gels [3, 7].

The membrane-bound mucins belong to an ever increasing **group** of type I membrane-anchored proteins. Based on their structure and localization at the cell surface they are thought to act in cell-cell and cell-matrix interactions and in cell signalling. The biological roles they play in cellular interactions and in cell signalling indicate that they are involved in regulating biological properties of epithelial cells [4, 5]. **Among the membrane-bound mucins, some conform to the mucin definition (presence of a Pro/Thr/Ser (PTS) region in the peptidic sequence) that are** MUC1, MUC3A/3B, MUC4, MUC12, MUC16, MUC17, MUC21 and

MUC22 [8-11]. **Others were assigned the term MUC despite the absence of that PTS domain that are MUC13, MUC15 and MUC20**. Because of their specific pattern of expression during the different steps of carcinogenesis, membrane-bound mucins stay under intense investigation as both potent new biomarkers and therapeutic targets in epithelial cancers [2-4].

Analysis of the peptidic sequences of mucins allowed description of their modular organization. Membrane-bound mucins are modular proteins which share conserved domains such as epidermal growth factor-like (EGF) **or** Sea urchin sperm protein Enterokinase and Agrin (SEA) **domains** (Figure 1) [8, 9, 12]. **The PTS domain, which is a common feature between mucins is the only domain not conserved at the genomic level although similarities exist at the amino acid level.** The human MUC1, MUC3A/3B, MUC12, MUC13, MUC16, and MUC17 mucins contain the SEA domains. Some domains are specific to one mucin. For **example**, the MUC4 mucin contains NIDO, AMOP, and VWF-D domains that are unique in the apomucin family. MUC22 (also called as Panbronchiolitis-related mucin-like protein 1, PBMUCL1) has recently been discovered in the disease-susceptibility locus for diffuse panbronchiolitis and contains both the typical PTS and a transmembrane (TM) domains [10]. Existence of other conserved domains in MUC22 remains to be demonstrated. **At this time, MUC15 is considered as a mucin-like since it lacks the characteristic tandem repeat (TR) region.**

Understanding the structure and the function of membrane-bound mucin domains will bring new information about their biological roles in epithelium homeostasis as well as in pathological situations such as carcinogenesis or inflammatory processes in which membrane-bound mucin expression **and localization are** often altered. In this review, we will discuss the structural and functional features of the evolutionary conserved domains of **membrane-bound** mucins and their abilities to modulate the biological properties of epithelial cells.

2. Mucins and evolution

Phylogenetic analyses have shown that the membrane-bound mucins are only found in mammals with the exception of MUC16 and MUC4 [13]. Mucin domains are shared among most of the MUC family members, suggesting common ancestors **or adoption of functional modules during evolution**. MUC1, MUC3, MUC12, MUC13 and MUC17 **share a peptidic sequence similar to** heparan sulfate proteoglycan of basement membrane (HSPG2) (Figure 2). *MUC3*, *MUC12* and *MUC17* genes are present contiguously on chromosome 7 (7q22) supporting the hypothesis of a process of duplication. *MUC16* evolved separately from agrin. MUC4 arose from two evolutionary events involving (i) NIDO and EGF-like domains from a common ancestor to nidogen and (ii) AMOP and VWF-D domains from a common ancestor to the Susd2 [13]. Beside its homology through the SEA domain, MUC1 has sequence similarities with other membrane-bound mucins and in fact evolved from repeated sequences of MUC5B secreted mucin [14]. Based on available sequence information, one can hypothesize that membrane-bound mucins evolved from several distinct ancestor genes and could be divided in different subgroups based on their conserved domains. Interestingly, mucins arising from a common ancestor gene are frequently clustered in a chromosome locus as illustrated on figure 2.

3. The membrane-bound mucin prototype domains

A prototype membrane-bound mucin is **characterized** by the presence of an extracellular *O*-glycosylated PTS domain, a TM domain and a cytoplasmic tail. Two EGF-like and one SEA domains are commonly found in most of membrane-bound mucins (Figure 1A).

3. 1 The PTS domain

The hallmark of membrane-bound mucins is the large extracellular subunit that protudes at the cell surface mainly composed by the PTS domain. The PTS domain is the substrate of important post-translational *O*-glycosylation modifications. This central domain is encoded by a large intronless genomic DNA sequence (>10 kb), typically characterized by a variable number of tandem repeat (VNTR) polymorphism [15]. PTS domain harbours extensive *O*-glycosylation (up to 80% of the total weight of the mature mucin) that forms chains of varying lengths, sequences, and compositions. The biosynthesis of mucin *O*-glycan chains is a step-by-step process occurring in the Golgi apparatus, involving specific glycosyl- and sulfo-transferases expressed in a tissue-specific manner [16, 17]. The *O*-glycans on PTS domain play major roles in the conformation and stability of the mucin and participate in defense of mucosae covering and protecting the underlying epithelium against various types of aggression (mechanical and chemical stress). In diseases such as cancer and inflammation, mucin *O*-glycosylation is altered, modifying the antigenic and adhesive properties of the glycan epitopes they bear [18]. Tumour-associated carbohydrate antigens (TAAs) are produced *via* incomplete synthesis of carbohydrate chains. TAA notably involve Tn and T antigens and their sialylated counterparts, sialyl-Tn and sialyl-T antigens. Cancer-associated *O*-glycans are often highly sialylated and less sulfated compared with *O*-glycans from normal mucins [19, 20].

Because of its extracellular localization **at the apical pole of the cell, the** MUC_{PTS} domain is mainly thought to act **as a sensor of the microenvironment. In the normal polarized epithelial cell, it may interact with immune cells, antibodies, viruses, bacteria to help maintain the epithelial homeostasis. In a depolarized cancer cell, the membrane-bound mucin is then expressed circumferentially and is going to then play different roles in cell-cell or cell-extracellular matrix interactions.** For example, **the** $MUC1_{PTS}$ domain **is able** to interact with adhesion molecules such as the endothelial protein ICAM-1 and

consequently may assist **the** cancer cell during heterotopic adhesion and invasion into the endothelium and reattachment at distant metastasis sites [21, 22].

Galectins belong to a family of carbohydrate binding proteins (β-galactoside-specific lectins). Galectin-1 was shown to interact with MUC16. Galectin-3 interacts with MUC1, MUC4 and MUC16. *O*-glycans on $MUC4_{PTS}$ interact with galectin-3 at the cell surface and mediate docking of tumour cells on endothelial cells [22, 23]. Galectin-3-MUC1 interaction alters MUC1 cell surface polarization, enhances tumour cell homotypic aggregation, prevents anoikis [24, 25], and regulates EGFR internalization, subcellular localization and ERK signalling pathway activation [26].

3. 2 The epidermal growth factor (EGF)-like domain

EGF-like domains are 30-40 residue-long and evolutionarily well-conserved. The EGF-like domain contains six cysteine residues that form three different disulfide bonds within the domain (C1–C3, C2–C4, and C5–C6). The first cysteine exhibits a mild conservation between the different mucins whereas the five others are highly conserved suggesting that the resulting tertiary structure leads to biological functions (Figure 3A/3B).

The EGF-like domains of membrane-bound mucins are believed to direct the interactions with different proteins. Mucin EGF-like domains are thought to act as ligands with membrane receptor such as those of the ErbB family mostly based on the work conducted on MUC4/rMuc4 biological roles. It has been speculated that EGF domains located in **the** released extracellular domain of mucins resulting from cleavage could also act as growth factors and serve as indicators of alteration of epithelial surfaces [5].

Experimental evidence suggests that rat homologues of Muc4 and ErbB2 are regulators of signalling related to growth, motility or differentiation of the cell *via* the EGF-like domains [27]. Human MUC4 and ErbB2 were shown to physically interact *via* a peptidic region

including the three EGF-like and intermediate domains of MUC4 [28]. Recombinant human MUC3A TM-EGF1/EGF2 protein led to reduced apoptosis induced either by TNF-α or Fas receptor stimulation. MUC17 EGF-like domains (also designated as Cys-rich domains, CRD) were shown to have anti-apoptotic and pro-migratory activity *via* ERK phosphorylation but did not stimulate cell proliferation [29, 30]. *In vivo* rectally administered MUC17 CRD, as well as its mouse homologue, accelerated healing in an experimental mouse model of colitis [29], suggesting a potential therapeutic use of EGF-like domains as treatment of mucosal inflammatory diseases.

The deduced amino-acid (aa) sequences of the EGF-like domains of MUC4, MUC3A, MUC12, MUC13, MUC16 and MUC17 (Table1) were used to predict their 3D structures using the Phyre 2 server (Protein Homology/analogy Recognition Engine V2.0) [31]. The predicted 3D structures of each EGF-like domain was compared with the resolved 3D structure of one monomer of the human EGF available through RCSB (Research Collaboratory for Structural Bioinformatics) protein data bank (pdb) (http://www.pdb.org/pdb/home/home.do) (pdb code : 1jl9) and with the other EGF domains 3D predicted structures using the molecular visualization system, PyMOL (The PyMOL Molecular Graphics System), version 1.5.0.1 Schrödinger, LLC). PyMOL did not allow us to predict the secondary structure of the third EGF domain of MUC4 ($MUC4_{EGF3}$) nor of the EGF domain of MUC12. To determine whether the EGF-like domains of two membrane-bound mucins were structurally homologous, PyMOL was used to measure the root-mean-square-deviation (RMSD) of superimposed 3D structures. Human EGF resolved structure consists notably of two anti-parallel β sheets. This topology is conserved in MUC_{EGF} predicted structures as illustrated on Figure 4A. The RMSD illustrates the average distance value between the backbone alpha carbons of two structures and consequently, the RMSD of two aligned structures indicates their divergence from one another. If the RMSD value is

below 1.5Å, two 3D structures whose sequence alignment is over 30% can be considered as almost homologous. The most significant result is then correlated with the highest number of residues aligned. Regarding the mucin EGF domains, there is strong structural homology between the 3D structure of the human EGF and the 3D predicted structure of $MUC4_{EGF1}$, $MUC3A_{EGF}$ and $MUC13_{EGF1}$ (Table 1). We also found high homology between $MUC13_{EGF1}$ and $MUC3A_{EGF}$ and $MUC12_{EGF}$. The 3D predicted structure of $MUC4_{EGF1}$ was shown **to be** homologous to $MUC16_{EGF}$.

3. 3 The Sea urchin sperm protein Enterokinase and Agrin (SEA)

The SEA module is a 120 aa domain found in many membrane-associated proteins at the cell surface. SEA domains were initially described as extracellular domains associated with *O*-glycosylation but more recently they have also been implicated in both cleavage events and association of the subunits [32-34]. Mucins usually contain one SEA domain in their extracellular region. However, MUC16 contains multiple SEA domains that were repeated up to 56 times through duplication events [13]. Alignment analysis indicated a mild conservation of the SEA domains of the different membrane-bound mucins (Figure 3C). We confirmed that MUC3, MUC12 and MUC17, that are located in a cluster on the human 7q22 chromosome, share the highest identity of SEA sequences (Figure 3D).

The $MUC1_{SEA}$ module is a self-cleaving domain [35]. Computer moduling of $MUC1_{SEA}$ domain initially suggested that it consists in three α-helixes and six β-strands forming an α/β sandwich fold [33]. Macao *et al.* later determined that SEA consists **of** four-stranded antiparallel β-sheets and four α-helixes occurring in the β1-α1-α2-β2↑β3-α3-α4-β4 order and with the helixes packed against the concave surface of the β-sheet [35]. Moreover, determination of the SEA domain structure from the murine Muc16 homologue using multi-dimensional NMR spectroscopy allowed Maeda *et al.* [36] to propose the GSVVV motif as a

proteolytic cleavage site located in the short loop connecting β2 and β3 sheets. Levitin et *al.* proposed that this module likely functions as a site for proteolytic cleavage in all SEA module-containing proteins [37]. Cleavage of membrane-bound mucins in the SEA domain is not entirely sequence dependent, but may be related to structural features of the SEA domain. Exposure of the cleavage site protruding outward from a compact structure enhances its recognition and cleavage by cellular proteases [33]. In addition to its cleavage involvement, MUC17 EGF1-linker-EGF2 recombinant protein induces anti-apoptotic activity that requires an intact SEA module [29].

The deduced aa sequences of the SEA domains of MUC1, MUC3A, MUC12, MUC13, MUC16 and MUC17 (Table 2) were used to predict their 3D structures using the Phyre 2 server. The 3D predicted structures of each SEA domain was compared with the 3D resolved structure of the human mucin MUC1 available in the protein data bank (pdb code : 2ACM) and with the 3D predicted structures of the other SEA domains using the molecular visualization system, PyMOL. In the same manner and with the same criteria as for the 3D prediction structures of the EGF domains, we looked at the 3D structural homology thanks to the RMSD calculation. Bioinformatic analysis indicates that the 3D resolved structure of $MUC1_{SEA}$ is homologous with $MUC17_{SEA}$. $MUC17_{SEA}$ is itself structurally close to $MUC3A_{SEA}$ and $MUC12_{SEA.}$ Superimposition of the 3D structures showed that the α/β sandwich fold of resolved $MUC1_{SEA}$ is conserved in predicted structures of $MUC3A_{SEA,}$ $MUC12_{SEA,}$ $MUC17_{SEA}$ (Figure 4B) [33].

4. The unique domains of membrane-bound mucin MUC4

Among membrane-bound mucins, MUC4 contains NIDO, AMOP, and VWF-D domains that are unique in the apomucin family.

4. 1. The Nidogen-like (Nido) domain

NIDO nidogen-like domain is an extracellular domain of unknown function found in nidogen and other hypothetical proteins such as sushi nidogen and EGF-like domains 1 (SNED1), plexin domain containing-1/-2 (PLXDC-1/-2) and tectorin alpha (TECTA). Phylogenetic analysis revealed the origin of the $MUC4_{NIDO}$ domain from an ancestor common to the NIDO protein [13]. Recent work demonstrated for the first time that $MUC4_{NIDO}$ does not alter the motility of pancreatic cancer cells but promotes invasion and extravasation. The $MUC4_{NIDO}$ domain, which is similar to the G1-domain of fibulin-2 or extracellular matrix protein Nidogen/entactin, contributes to the interaction properties of MUC4. Senapati *et al*. hypothesized that $MUC4_{NIDO}$ disrupts the interaction between NIDO/entactin and fibulin-2 [38], a major component of murine liver blood vessels [39], in a competitive inhibition manner and thus may create a favorable environment in liver for the extravasation of metastatic pancreatic cancer cells. $MUC4_{NIDO}$ domain is thus proposed as a mediator of protease-independent cell invasion in pancreatic cancer metastasis [38].

4. 2 The Adhesion-associated domain in MUC4 and Other Proteins (AMOP)

AMOP and vWF-D domains originated from a common ancestor to the Sushi-domain containing protein 2 (SUSD2) [13]. AMOP domain is uncommon in the human genome as it is described only in four proteins (Isthmin-1, -2, SUSD2 and MUC4) using Simple Modular Architecture Research Tool (SMART) database. The role of AMOP domain is unclear. Presence of AMOP in cell adhesion molecules could be indicative of a role for this domain in adhesion. Indeed, mouse Isthmin, an angiogenesis modulator, disrupts endothelial cell (EC) capillary network formation on Matrigel through its C-terminal AMOP domain [40]. From there, one could hypothesize such role for MUC4 in angiogenesis.

4. 3 The Von Wildebrand Factor-D (vWF-D) domain

The human gel-forming mucins contain vWF-D and C-terminal cysteine-knot domains that are responsible for their oligomerization [9, 41, 42]. Only MUC4 membrane-bound mucin contains a vWF-D in its intracellular region [9]. VWF domains usually participate in forming disulfide bonds. However, the cysteine residues, conserved in secreted mucins, are not conserved in MUC4 suggesting a loss of function of this domain in membrane-bound mucins during evolution [43]. The crystal structures of three vWF domains are known. All three domains share identical three-dimensional fold with a α-β-α sandwiched model [44].

5. The cytoplasmic tail

Cytoplasmic tails (CT) of mucins are poorly conserved ruling out a potential common evolution. Size and aa sequence greatly vary among membrane-bound mucins leading to a great variety of functions in cell signalling (Table 3). Indeed, MUC4 acts as a receptor partner at the membrane through its extracellular domain because of its short CT whereas MUC1 acts as an intracellular docking protein for signalling molecules. Among membrane-bound mucins, MUC1 has been under intense investigation.

$MUC1_{CT}$, is 72 aa long, contains seven tyrosine residues, an Src Homology 2 (SH2) interaction domain, and is phosphorylated by several kinases [4, 22, 45, 46]. The $MUC1_{CT}$ allows the direct interaction of MUC1 with a wide array of signalling pathways in tumour cells. $MUC1_{CT}$ was shown to interact with Src family kinase including c-Src, Lyn and Lck and promotes phosphorylation of Tyr46 involved in protein-protein interaction. $MUC1_{CT}$ is also targeted by kinases ζ chain associated protein kinase of 70 kDa (ZAP70), and the δ isoform of protein kinase C (PKCδ) [22, 47]. Interaction between $MUC1_{CT}$ and proapoptotic kinase cAbl leads to phosphorylation of Tyr60 by its binding to the c-Abl SH2 domain [48]. Binding of scaffolding protein Grb2 on $MUC1_{CT}$ activates MAPK pathway, involved in

proliferation of tumour cells [49]. $MUC1_{CT}$ interacts with components of the IκB kinase (IKK) complex enhancing IKKα-IKKβ interaction, promoting IKKβ phosphorylation, IKKα degradation, NFκB-p65 targeting to the nucleus that leads to NF-κB p65 transcriptional activity [50, 51]. $MUC1_{CT}$ interacts with chaperone heat shock proteins (HSP) HSP70 and HSP90 leading to mitochondrial translocation of $MUC1_{CT}$ and therefore inhibiting apoptosis [52]. $MUC1_{CT}$ binds to Wnt signalling pathway components β-catenin and APC. $MUC1_{CT}$-β-catenin interaction enhances the levels of nuclear β-catenin during disruption of cadherin-mediated cell-cell adhesion and promotes expression of Wnt target genes [46, 53]. The $MUC1_{CT}$-β-catenin binding is regulated by GSK3β and HSP90 in a competitive inhibition manner [54]. $MUC1_{CT}$ is phosphorylated by several cell surface receptors such as fibroblast growth factor receptor 3 (FGFR3), platelet-derived growth factor receptor (PDGFR) and ErbB receptor family altering interaction with proteins mentioned above. FGF1 induces Tyr46 YEKV phosphorylation whereas PDGF catalyses phosphorylation of Tyr35 and enhances invasion *in vitro*, tumour growth and metastasis *in vivo* [55].

$MUC1_{CT}$ also regulates its own nucleo-cytoplasmic transport by binding importin-β and nucleoporin p62 (nup62) *via* a CQC motif and therefore alters transcriptional regulation of numerous target genes [56]. In the cytoplasm of breast and lung cancer cells, $MUC1_{CT}$ forms dimers that are necessary for its nuclear localization *via* the CQC motif [56, 57]. In the nucleus, $MUC1_{CT}$ associates with transcription factors such as β-catenin/TCF4, p53, CDKN1A (p21), nuclear factor-κB p65 or STATs [3]. The $MUC1_{CT}$ domain also stabilizes estrogen receptor-α (ERα) and is necessary for its nuclear localization [58].

$MUC1_{CT}$ has been extensively characterised whereas other mucin CT are scarcely described. $MUC3_{CT}$ contains an YVAL aa motif which is similar to motifs recognized by the SH2 domain [59]. $MUC13_{CT}$ contains a PKC phosphorylation site [60]. $MUC16_{CT}$ contains polybasic aa that are predicted to interact with ezrin/radixin/moesin (ERM) actin-binding

proteins [61] and is proposed to cross-link MUC16 with actin-cytoskeleton. $MUC16_{CT}$ is also proposed to interact with JAK2 *via* its ERM domain and mediate breast cancer cell proliferation [62]. PDZ-interacting domains are observed within $MUC3_{CT}$, $MUC12_{CT}$ and $MUC17_{CT}$ but only phosphorylation sites of $MUC17_{CT}$ exhibit a strong binding to PDZ domain of PDZK1 [63]. It was also shown using $Pdzk1^{-/-}$ mice that Pdzk1 plays a specific role in stabilizing mMuc3 (MUC17 mouse homologue) at the apical pole of polarized enterocytes. $MUC20_{CT}$ binds to a multifunctional docking site of hepatocyte growth factor (HGF) receptor Met and therefore suppresses the Grb2-Ras pathway [64].

The tremendous variability of MUC_{CT} is thus associated with a wide array of signalling pathways regulated by membrane-bound mucins and much is still to be discovered to fully characterize their biological activities and effects on epithelial cell behavior.

6. Mucin splice variants

6. 1. MUC1

Membrane-bound mucin isoforms or alternative splicing events have been reported for MUC1, MUC3A, MUC4 and MUC17 that lead to new apomucin polypeptides possibly lacking functional domains and presumably showing altered cellular/biological functions.

MUC1/X is an alternate isoform of MUC1 in which the extracellular domain is comprised of the SEA module in addition to thirty MUC1 N-terminal aa residues [37]. MUC1/Z and MUC1/Y are two isoforms lacking the PTS domain. MUC1/Z isoform is proteolytically cleaved after its synthesis and generates the two subunits MUC1α and MUC1β. On the contrary, MUC1/Y contains a truncated SEA domain and therefore lacks the cleavage site, resulting in a single uncleaved apomucin. MUC1/Y also contains the TM and cytoplasmic domains. Interestingly, MUC1/Y isoform expression increases the tumourigenicity of DA3 mouse mammary epithelial cells [65]. An alternate isoform MUC1/ZD results from a reading

frameshift caused by a splicing event that deletes the TR-PTS domain but contains the signal peptide and a subsequent stretch of 30 aa [66]. Variant MUC1/SEC lacks the TM and cytoplasmic sequences and has the potential to be directly secreted out of the cell. Interestingly, tumour cells expressing MUC1/SEC fail to develop tumours in immunocompetent mice. MUC1/SEC may inhibit tumour development and may support antitumour immune responses *via* the downregulation of Urokinase-Type Plasminogen Activator (uPA) and Signal Transducer and Activator of Transcription 1 (STAT1) [67]. Moreover, MUC1/SEC is capable of blocking expression of arginase 1 and production of reactive oxygen species (ROS) in myeloid-derived suppressor cells MDSCs that play a critical role in tumour-induced immunosuppression [68].

6.2. MUC4

Multiple splicing events have been described for *MUC4* [69]. Among them, MUC4/Y and MUC4/X lack the central TR/MUC4$_{PTS}$ domain, the main feature of membrane-bound mucins [69]. Some MUC4 isoforms are also lacking TM domain and might be secreted. One could hypothesize that the EGF-like domains of MUC4 act as growth factors when released in the extracellular environment. Globally, more than twenty cDNA isoforms have been described for *MUC4* that are generated by several mechanisms (alternative use of cassette exons, exon skipping or use of cryptic splice donor/acceptor sites) [70]. However, till now no splice variants for Muc4 have been observed in mouse tissues [9]. So far, the existence at the protein level and the biological roles of the different MUC4 isoforms remain to be determined.

6. 3. Other mucin variants

Transcription variants of *MUC3* encode truncated proteins, suggesting the possibility of expression of soluble forms [71]. Williams *et al.* confirmed the existence of two secreted

isoforms of MUC3 that lack the TM domain [72]. In a similar manner, an alternative MUC17 splice event occurs and lacks the second EGF domain, the TM domain, and the cytoplasmic tail and leads to a secreted form MUC17/SEC [73].

7. Using MUC domains as therapeutic targets/tools

Membrane-bound mucins display major biological activities in epithelium homeostasis and pathologies and hold promise as biological tools for therapy in cancer or inflammatory diseases [4].

The extended knowledge of MUC1 domains and their biological significance led to the development of peptide-based therapies that may have important clinical implications [1, 3, 4, 74]. As an example, MUC1 inhibitory peptide (MIP) was designed to block the intracellular interactions between MUC1/β-catenin and MUC1/EGFR. MIP was then fused to the protein transduction domain, PTD4 (PMIP) in order to increase cellular uptake. Exogenous treatment of PMIP led to a significant reduction in aggressiveness of metastatic breast cancer cells *in vitro*, and inhibition of growth and recurrence of breast tumour in an *in vivo* transgenic model [75]. Moreover, PMIP selectively inhibits lung adenocarcinoma proliferation, and inhibits ER responses *via* the blocking of the MUC1-ERα interaction [76]. $MUC1_{CT}$ dimerization may also be targeted by reducing agents. Similarly, cell-penetrating peptide containing the CQCRRKN sequence binds directly to the endogenous $MUC1_{CT}$ and blocks its ability to dimerize [57]. That blockage induces death of human breast cancer cells *in vitro* and in xenograft tumour models [77].

Since MUC1 displays ubiquitous expression in a wide variety of tumour types, numerous studies are targeting MUC1. The MUC1 peptide core or glycopeptide has been used in immunotherapy for immunization. The different strategies showed T cell- and antibody-responses and then were challenged for tumour protection [78, 79]. Recently, immunization

by a tripartite vaccine based on a synthetic 60-mer $MUC1_{PTS}$ derived peptide covalently linked to a Toll-like receptor (TLR) agonist could elicit IgG antibodies that efficiently lysed MUC1-expressing cancer cells, stimulated cytotoxicity of T lymphocytes, and activated innate immune responses [80].

Trastuzumab (also called herceptin®), a monoclonal antibody, targets ErbB2-expressing tumours such as breast cancers and leads to inhibition of intracellular signalling and induction of an immune system-mediated antitumour response [81]. Human MUC4 and ErbB2 do physically interact *via* $MUC4_{EGF3+1+2}$ region that contains the three EGF-like domains [28]. MUC4 overexpression decreases cancer cell sensitivity to trastuzumab, possibly by steric hindrance [82]. MUC4 is thought to impair access of trastuzumab to ErbB2 *via* its large PTS domain that protrudes outside the glycocalyx. More recently, we also showed an effect of MUC4 on gemcitabine sensitivity in pancreatic ductal adenocarcinoma [83] but via domains and partner still to identify.

$MUC17_{EGF}$ domains display a strong potential in the treatment of mucosal inflammatory diseases as healing is accelerated in acetic acid and dextran sodium sulfate (DSS)-induced colitis in mice [29]. MUC1 also has a protective effect against DSS-induced colitis in mice. Since Muc1 is lacking EGF-like domains, this effect is most likely mediated by another mechanism. MUC1 could act through its ability to increase the mucus barrier but also by decreasing T cell recruitment to the afflicted site [84].

7. Conclusions

Membrane-bound mucins are characterized by a multi-domain organisation with **a PTS domain,** conserved EGF or SEA domains and unique domains such as AMOP, NIDO or vWF-D that could mediate specific functions for some membrane-bound mucins. Understanding the structure and the biological functions of each domain might help explain

the major roles of mucins in inflammatory and cancerous diseases of the epithelium. In the future, the better structural and functional characterization of these domains should give rise to new mucin-based therapeutic targets/tools.

Acknowledgments

Dr Nicolas Jonckheere is a recipient of a postdoctoral fellowship from the Ligue Nationale Contre le Cancer (LNCC). Nicolas Skrypek is a recipient of a PhD fellowship from the Centre Hospitalier Régional et Universitaire (CHRU) de Lille/région Nord-Pas de Calais. Dr Frédéric Frénois is a recipient of a postdoctoral fellowship from the Fondation pour la Recherche Médicale (FRM). This work is supported by a grant from la Ligue Nationale Contre le Cancer (Equipe Labellisée Ligue 2010, IVS). Isabelle Van Seuningen is the recipient of a "Contrat Hospitalier de Recherche Translationnelle"/CHRT 2010, AVIESAN.

References

[1] Jonckheere N., Skrypek N., van Seuningen I., Mucins and pancreatic cancer, Cancers 2 (2010) 1794-1812.
[2] Jonckheere N., Van Seuningen I., The membrane-bound mucins: From cell signalling to transcriptional regulation and expression in epithelial cancers, Biochimie 92 (2010) 1-11.
[3] Kufe D.W., Mucins in cancer: function, prognosis and therapy, Nat Rev Cancer 9 (2009) 874-885.
[4] Jonckheere N., Van Seuningen I., The membrane-bound mucins: how large O-glycoproteins play key roles in epithelial cancers and hold promise as biological tools for gene-based and immunotherapies, Crit Rev Oncog 14 (2008) 177-196.
[5] Hollingsworth M.A., Swanson B.J., Mucins in cancer: protection and control of the cell surface, Nat Rev Cancer 4 (2004) 45-60.
[6] Van Seuningen I., Pigny P., Perrais M., Porchet N., Aubert J.P., Transcriptional regulation of the 11p15 mucin genes. Towards new biological tools in human therapy, in inflammatory diseases and cancer?, Front Biosci 6 (2001) D1216-1234.
[7] Linden S.K., Sutton P., Karlsson N.G., Korolik V., McGuckin M.A., Mucins in the mucosal barrier to infection, Mucosal Immunol 1 (2008) 183-197.
[8] Jonckheere N., Van Seuningen I., The ever growing family of membrane-bound mucins, in: Van Seuningen I. (Ed.), The Epithelial Mucins: Structure/Function. Roles in Cancer and Inflammatory Diseases, Research Signpost, Kerala, India, 2008, pp. 17-38.
[9] Desseyn J.L., Tetaert D., Gouyer V., Architecture of the large membrane-bound mucins, Gene 410 (2008) 215-222.
[10] Hijikata M., Matsushita I., Tanaka G., Tsuchiya T., Ito H., Tokunaga K., Ohashi J., Homma S., Kobashi Y., Taguchi Y., Azuma A., Kudoh S., Keicho N., Molecular cloning of two novel mucin-like genes in the disease-susceptibility locus for diffuse panbronchiolitis, Hum Genet 129 (2011) 117-128.
[11] Dekker J., Rossen J.W., Buller H.A., Einerhand A.W., The MUC family: an obituary, Trends Biochem Sci 27 (2002) 126-131.
[12] Moniaux N., Escande F., Porchet N., Aubert J.P., Batra S.K., Structural organization and classification of the human mucin genes, Front Biosci 6 (2001) D1192-1206.
[13] Duraisamy S., Ramasamy S., Kharbanda S., Kufe D., Distinct evolution of the human carcinoma-associated transmembrane mucins, MUC1, MUC4 AND MUC16, Gene 373 (2006) 28-34.
[14] Duraisamy S., Kufe T., Ramasamy S., Kufe D., Evolution of the human MUC1 oncoprotein, Int J Oncol 31 (2007) 671-677.
[15] Fowler J., Vinall L., Swallow D., Polymorphism of the human muc genes, Front Biosci 6 (2001) D1207-1215.
[16] Groux-Degroote S., Harduin-Lepers A., Delannoy P., Biosynthesis of mucin O-glycan chains in normal and pathological states, in: Van Seuningen I. (Ed.), The Epithelial Mucins: Structure/Function. Roles in Cancer and Inflammatory Diseases, Research Signpost, Kerala, India, 2008, pp. 39-54.
[17] Silverman H.S., Sutton-Smith M., McDermott K., Heal P., Leir S.H., Morris H.R., Hollingsworth M.A., Dell A., Harris A., The contribution of tandem repeat number to the O-glycosylation of mucins, Glycobiology 13 (2003) 265-277.
[18] Robbe-Masselot C., Michalski J.C., Capon C., Tumour-associated antigens of mucin O-glycans. Clinical relevance of glycobiology in cancer and inflammatory diseases of the epithelium, in: Van Seuningen I. (Ed.), The Epithelial Mucins: Structure/Function. Roles in Cancer and Inflammatory Diseases, Research Signpost, Kerala, India, 2008, pp. 55-74.
[19] Brockhausen I., Pathways of O-glycan biosynthesis in cancer cells, Biochim Biophys Acta 1473 (1999) 67-95.
[20] Brockhausen I., Sulphotransferases acting on mucin-type oligosaccharides, Biochem Soc Trans 31 (2003) 318-325.
[21] Rahn J.J., Chow J.W., Horne G.J., Mah B.K., Emerman J.T., Hoffman P., Hugh J.C., MUC1 mediates transendothelial migration in vitro by ligating endothelial cell ICAM-1, Clin Exp Metastasis 22 (2005) 475-483.
[22] Senapati S., Das S., Batra S.K., Mucin-interacting proteins: from function to therapeutics, Trends Biochem Sci 35 (2010) 236-245.
[23] Senapati S., Chaturvedi P., Chaney W.G., Chakraborty S., Gnanapragassam V.S., Sasson A.R., Batra S.K., Novel INTeraction of MUC4 and galectin: potential pathobiological implications for metastasis in lethal pancreatic cancer, Clin Cancer Res 17 (2011) 267-274.

[24] Zhao Q., Guo X., Nash G.B., Stone P.C., Hilkens J., Rhodes J.M., Yu L.G., Circulating galectin-3 promotes metastasis by modifying MUC1 localization on cancer cell surface, Cancer Res 69 (2009) 6799-6806.
[25] Zhao Q., Barclay M., Hilkens J., Guo X., Barrow H., Rhodes J.M., Yu L.G., Interaction between circulating galectin-3 and cancer-associated MUC1 enhances tumour cell homotypic aggregation and prevents anoikis, Mol Cancer 9 (2010) 154.
[26] Merlin J., Stechly L., de Beauce S., Monte D., Leteurtre E., van Seuningen I., Huet G., Pigny P., Galectin-3 regulates MUC1 and EGFR cellular distribution and EGFR downstream pathways in pancreatic cancer cells, Oncogene 30 (2011) 2514-2525.
[27] Carraway K.L., Ramsauer V.P., Haq B., Carothers Carraway C.A., Cell signaling through membrane mucins, Bioessays 25 (2003) 66-71.
[28] Jonckheere N., Skrypek N., Merlin J., Dessein A.F., Dumont P., Leteurtre E., Harris A., Desseyn J.L., Susini C., Frenois F., Van Seuningen I., The Mucin MUC4 and Its Membrane Partner ErbB2 Regulate Biological Properties of Human CAPAN-2 Pancreatic Cancer Cells via Different Signalling Pathways, PLoS One 7 (2012) e32232.
[29] Luu Y., Junker W., Rachagani S., Das S., Batra S.K., Heinrikson R.L., Shekels L.L., Ho S.B., Human intestinal MUC17 mucin augments intestinal cell restitution and enhances healing of experimental colitis, Int J Biochem Cell Biol 42 (2010) 996-1006.
[30] Ho S.B., Luu Y., Shekels L.L., Batra S.K., Kandarian B., Evans D.B., Zaworski P.G., Wolfe C.L., Heinrikson R.L., Activity of recombinant cysteine-rich domain proteins derived from the membrane-bound MUC17/Muc3 family mucins, Biochim Biophys Acta 1800 (2010) 629-638.
[31] Kelley L.A., Sternberg M.J., Protein structure prediction on the Web: a case study using the Phyre server, Nat Protoc 4 (2009) 363-371.
[32] Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A., Harris A., Identification of MUC1 proteolytic cleavage sites in vivo, Biochem Biophys Res Commun 283 (2001) 715-720.
[33] Palmai-Pallag T., Khodabukus N., Kinarsky L., Leir S.H., Sherman S., Hollingsworth M.A., Harris A., The role of the SEA (sea urchin sperm protein, enterokinase and agrin) module in cleavage of membrane-tethered mucins, Febs J 272 (2005) 2901-2911.
[34] Bork P., Patthy L., The SEA module: a new extracellular domain associated with O-glycosylation, Protein Sci 4 (1995) 1421-1425.
[35] Macao B., Johansson D.G., Hansson G.C., Hard T., Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin, Nat Struct Mol Biol 13 (2006) 71-76.
[36] Maeda T., Inoue M., Koshiba S., Yabuki T., Aoki M., Nunokawa E., Seki E., Matsuda T., Motoda Y., Kobayashi A., Hiroyasu F., Shirouzu M., Terada T., Hayami N., Ishizuka Y., Shinya N., Tatsuguchi A., Yoshida M., Hirota H., Matsuo Y., Tani K., Arakawa T., Carninci P., Kawai J., Hayashizaki Y., Kigawa T., Yokoyama S., Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16), J Biol Chem 279 (2004) 13174-13182.
[37] Levitin F., Stern O., Weiss M., Gil-Henn C., Ziv R., Prokocimer Z., Smorodinsky N.I., Rubinstein D.B., Wreschner D.H., The MUC1 SEA module is a self-cleaving domain, J Biol Chem 280 (2005) 33374-33386.
[38] Senapati S., Gnanapragassam V.S., Moniaux N., Momi N., Batra S.K., Role of MUC4-NIDO domain in the MUC4-mediated metastasis of pancreatic cancer cells, Oncogene 31 (2012) 3346-3356.
[39] Piscaglia F., Dudas J., Knittel T., Di Rocco P., Kobold D., Saile B., Zocco M.A., Timpl R., Ramadori G., Expression of ECM proteins fibulin-1 and -2 in acute and chronic liver disease and in cultured rat liver cells, Cell Tissue Res 337 (2009) 449-462.
[40] Xiang W., Ke Z., Zhang Y., Cheng G.H., Irwan I.D., Sulochana K.N., Potturi P., Wang Z., Yang H., Wang J., Zhuo L., Kini R.M., Ge R., Isthmin is a novel secreted angiogenesis inhibitor that inhibits tumour growth in mice, J Cell Mol Med 15 (2011) 359-374.
[41] Lang T., Hansson G.C., Samuelsson T., Gel-forming mucins appeared early in metazoan evolution, Proc Natl Acad Sci U S A 104 (2007) 16209-16214.
[42] Desseyn J.L., Gouyer V., Tetaert D., Architecture of the gel-forming mucins, in: Van Seuningen I. (Ed.), The Epithelial Mucins: Structure/Function. Roles in Cancer and Inflammatory Diseases, Research Signpost, Kerala, India, 2008, pp. 1-16.
[43] Desseyn J.L., Clavereau I., Laine A., Cloning, chromosomal localization and characterization of the murine mucin gene orthologous to human MUC4, Eur J Biochem 269 (2002) 3150-3159.
[44] Hassan M.I., Saxena A., Ahmad F., Structure and function of von Willebrand factor, Blood Coagul Fibrinolysis 23 (2012) 11-22.
[45] Hattrup C.L., Gendler S.J., Structure and function of the cell surface (tethered) mucins, Annu Rev Physiol 70 (2008) 431-457.
[46] Carson D.D., The cytoplasmic tail of MUC1: a very busy place, Sci Signal 1 (2008) pe35.

[47] Kufe D.W., Functional targeting of the MUC1 oncogene in human cancers, Cancer Biol Ther 8 (2009) 1197-1203.
[48] Raina D., Ahmad R., Kumar S., Ren J., Yoshida K., Kharbanda S., Kufe D., MUC1 oncoprotein blocks nuclear targeting of c-Abl in the apoptotic response to DNA damage, Embo J 25 (2006) 3774-3783.
[49] Pandey P., Kharbanda S., Kufe D., Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein, Cancer Res 55 (1995) 4000-4003.
[50] Ahmad R., Raina D., Trivedi V., Ren J., Rajabi H., Kharbanda S., Kufe D., MUC1 oncoprotein activates the IkappaB kinase beta complex and constitutive NF-kappaB signalling, Nat Cell Biol 9 (2007) 1419-1427.
[51] Ahmad R., Raina D., Joshi M.D., Kawano T., Ren J., Kharbanda S., Kufe D., MUC1-C oncoprotein functions as a direct activator of the nuclear factor-kappaB p65 transcription factor, Cancer Res 69 (2009) 7013-7021.
[52] Ren J., Bharti A., Raina D., Chen W., Ahmad R., Kufe D., MUC1 oncoprotein is targeted to mitochondria by heregulin-induced activation of c-Src and the molecular chaperone HSP90, Oncogene 25 (2006) 20-31.
[53] Wen Y., Caffrey T.C., Wheelock M.J., Johnson K.R., Hollingsworth M.A., Nuclear association of the cytoplasmic tail of MUC1 and beta-catenin, J Biol Chem 278 (2003) 38029-38039.
[54] Ren J., Raina D., Chen W., Li G., Huang L., Kufe D., MUC1 oncoprotein functions in activation of fibroblast growth factor receptor signaling, Mol Cancer Res 4 (2006) 873-883.
[55] Singh P.K., Wen Y., Swanson B.J., Shanmugam K., Kazlauskas A., Cerny R.L., Gendler S.J., Hollingsworth M.A., Platelet-derived growth factor receptor beta-mediated phosphorylation of MUC1 enhances invasiveness in pancreatic adenocarcinoma cells, Cancer Res 67 (2007) 5201-5210.
[56] Leng Y., Cao C., Ren J., Huang L., Chen D., Ito M., Kufe D., Nuclear import of the MUC1-C oncoprotein is mediated by nucleoporin Nup62, J Biol Chem 282 (2007) 19321-19330.
[57] Raina D., Ahmad R., Rajabi H., Panchamoorthy G., Kharbanda S., Kufe D., Targeting cysteine-mediated dimerization of the MUC1-C oncoprotein in human cancer cells, Int J Oncol 40 (2012) 1643-1649.
[58] Wei X., Xu H., Kufe D., MUC1 oncoprotein stabilizes and activates estrogen receptor alpha, Mol Cell 21 (2006) 295-305.
[59] Songyang Z., Shoelson S.E., McGlade J., Olivier P., Pawson T., Bustelo X.R., Barbacid M., Sabe H., Hanafusa H., Yi T., et al., Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav, Mol Cell Biol 14 (1994) 2777-2785.
[60] Maher D.M., Gupta B.K., Nagata S., Jaggi M., Chauhan S.C., Mucin 13: structure, function, and potential roles in cancer pathogenesis, Mol Cancer Res 9 (2011) 531-537.
[61] Blalock T.D., Spurr-Michaud S.J., Tisdale A.S., Heimer S.R., Gilmore M.S., Ramesh V., Gipson I.K., Functions of MUC16 in corneal epithelial cells, Invest Ophthalmol Vis Sci 48 (2007) 4509-4518.
[62] Lakshmanan I., Ponnusamy M.P., Das S., Chakraborty S., Haridas D., Mukhopadhyay P., Lele S.M., Batra S.K., MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells, Oncogene 31 (2012) 805-817.
[63] Malmberg E.K., Pelaseyed T., Petersson A.C., Seidler U.E., De Jonge H., Riordan J.R., Hansson G.C., The C-terminus of the transmembrane mucin MUC17 binds to the scaffold protein PDZK1 that stably localizes it to the enterocyte apical membrane in the small intestine, Biochem J 410 (2008) 283-289.
[64] Higuchi T., Orita T., Katsuya K., Yamasaki Y., Akiyama K., Li H., Yamamoto T., Saito Y., Nakamura M., MUC20 suppresses the hepatocyte growth factor-induced Grb2-Ras pathway by binding to a multifunctional docking site of met, Mol Cell Biol 24 (2004) 7456-7468.
[65] Baruch A., Hartmann M., Zrihan-Licht S., Greenstein S., Burstein M., Keydar I., Weiss M., Smorodinsky N., Wreschner D.H., Preferential expression of novel MUC1 tumor antigen isoforms in human epithelial tumors and their tumor-potentiating function, Int J Cancer 71 (1997) 741-749.
[66] Levitin F., Baruch A., Weiss M., Stiegman K., Hartmann M.L., Yoeli-Lerner M., Ziv R., Zrihan-Licht S., Shina S., Gat A., Lifschitz B., Simha M., Stadler Y., Cholostoy A., Gil B., Greaves D., Keydar I., Zaretsky J., Smorodinsky N., Wreschner D.H., A novel protein derived from the MUC1 gene by alternative splicing and frameshifting, J Biol Chem 280 (2005) 10655-10663.
[67] Ilkovitch D., Handel-Fernandez M.E., Herbert L.M., Lopez D.M., Antitumor effects of Mucin 1/sec involves the modulation of urokinase-type plasminogen activator and signal transducer and activator of transcription 1 expression in tumor cells, Cancer Res 68 (2008) 2427-2435.
[68] Ilkovitch D., Lopez D.M., Urokinase-mediated recruitment of myeloid-derived suppressor cells and their suppressive mechanisms are blocked by MUC1/sec, Blood 113 (2009) 4729-4739.

[69] Moniaux N., Escande F., Batra S.K., Porchet N., Laine A., Aubert J.P., Alternative splicing generates a family of putative secreted and membrane-associated MUC4 mucins, Eur J Biochem 267 (2000) 4536-4544.
[70] Escande F., Lemaitre L., Moniaux N., Batra S.K., Aubert J.P., Buisine M.P., Genomic organization of MUC4 mucin gene. Towards the characterization of splice variants, Eur J Biochem 269 (2002) 3637-3644.
[71] Crawley S.C., Gum J.R., Jr., Hicks J.W., Pratt W.S., Aubert J.P., Swallow D.M., Kim Y.S., Genomic organization and structure of the 3' region of human MUC3: alternative splicing predicts membrane-bound and soluble forms of the mucin, Biochem Biophys Res Commun 263 (1999) 728-736.
[72] Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A., The MUC3 gene encodes a transmembrane mucin and is alternatively spliced, Biochem Biophys Res Commun 261 (1999) 83-89.
[73] Moniaux N., Junker W.M., Singh A.P., Jones A.M., Batra S.K., Characterization of human mucin MUC17. Complete coding sequence and organization, J Biol Chem 281 (2006) 23676-23685.
[74] Kufe D.W., MUC1-C oncoprotein as a target in breast cancer: activation of signaling pathways and therapeutic approaches, Oncogene doi: 10.1038/onc.2012.158. (2012).
[75] Bitler B.G., Menzl I., Huerta C.L., Sands B., Knowlton W., Chang A., Schroeder J.A., Intracellular MUC1 peptides inhibit cancer progression, Clin Cancer Res 15 (2009) 100-109.
[76] Klinge C.M., Radde B.N., Imbert-Fernandez Y., Teng Y., Ivanova M.M., Abner S.M., Martin A.L., Targeting the intracellular MUC1 C-terminal domain inhibits proliferation and estrogen receptor transcriptional activity in lung adenocarcinoma cells, Mol Cancer Ther 10 (2011) 2062-2071.
[77] Raina D., Ahmad R., Joshi M.D., Yin L., Wu Z., Kawano T., Vasir B., Avigan D., Kharbanda S., Kufe D., Direct targeting of the mucin 1 oncoprotein blocks survival and tumorigenicity of human breast carcinoma cells, Cancer Res 69 (2009) 5133-5141.
[78] Tang C.K., Katsara M., Apostolopoulos V., Strategies used for MUC1 immunotherapy: human clinical studies, Expert Rev Vaccines 7 (2008) 963-975.
[79] Tang C.K., Apostolopoulos V., Strategies used for MUC1 immunotherapy: preclinical studies, Expert Rev Vaccines 7 (2008) 951-962.
[80] Lakshminarayanan V., Thompson P., Wolfert M.A., Buskas T., Bradley J.M., Pathangey L.B., Madsen C.S., Cohen P.A., Gendler S.J., Boons G.J., Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine, Proc Natl Acad Sci U S A 109 (2011) 261-266.
[81] Mukohara T., Mechanisms of resistance to anti-human epidermal growth factor receptor 2 agents in breast cancer, Cancer Sci 102 (2011) 1-8.
[82] Nagy P., Friedlander E., Tanner M., Kapanen A.I., Carraway K.L., Isola J., Jovin T.M., Decreased accessibility and lack of activation of ErbB2 in JIMT-1, a herceptin-resistant, MUC4-expressing breast cancer cell line, Cancer Res 65 (2005) 473-482.
[83] Skrypek N., Duchêne B., Hebbar M., Leteurtre E., van Seuningen I., Jonckheere N., The MUC4 mucin mediates gemcitabine resistance of human pancreatic cancer cells via the Concentrative Nucleoside Transporter family, Oncogene doi: 10.1038/onc.2012.179 (2012).
[84] Petersson J., Schreiber O., Hansson G.C., Gendler S.J., Velcich A., Lundberg J.O., Roos S., Holm L., Phillipson M., Importance and regulation of the colonic mucus barrier in a mouse model of colitis, Am J Physiol Gastrointest Liver Physiol 300 (2011) G327-333.

Figure legends

Figure 1: Membrane-bound mucin prototype. (A) Membrane-bound mucins are modular proteins sharing conserved domains such as epidermal growth factor-like (EGF), Sea urchin sperm protein Enterokinase and Agrin (SEA) or Pro/Thr/Ser (PTS) domains or MUC4-specific domains (AMOP, Nido and vWF-D). (B) Alternative splicing events have been reported and lead to new secreted or membrane-bound apomucin polypeptides possibly lacking functional domains and presumably showing altered cellular/biological functions.

Figure 2: Evolution of the membrane-bound mucins. Phylogenetic trees of membrane-bound mucins were built after aligning of amino acid sequences of various domains conserved in mucin families. Three independent trees were deducted clustering (A) MUC1, MUC3, MUC12, MUC13, MUC17 evolving from SEA-containing protein HSGP2. $MUC3_{EGF}$ domains display the highest identity with EGF. MUC1 cytoplasmic tail (CT) peptide sequence is related to MUC5B tandem repeat (TR) (B) MUC4 arises from two ancestor proteins common to Nidogen (Nido and EGF-like domains) and Susd2 (AMOP and vWF-D domains) (C) MUC16 arises from multiplication of SEA domain related to Agrin proteins.

Figure 3: EGF-like and SEA domains of membrane-bound mucins. (A) Alignment of EGF domains of membrane-bound mucins and EGF. Peptidic sequences of the domains were identified by Protein Knowledgebase (UniprotKB, http://www.uniprot.org) (B) Phylogenetic tree of identity of EGF domains. (C) Alignment of SEA domains of membrane-bound mucins. Peptidic sequences of the domains were identified by Protein Knowledgebase. * MUC16 contains up to 56 SEA domains. For clarity purpose, we decided to use the SEA1 domain as a template for alignment analysis. (D) Phylogenetic tree of identity of SEA domains.

Figure 4: Three dimensional structure of EGF and SEA domains of membrane-bound mucins. (A) Ribbon diagram showing the superimposition of the 3D structure of the EGF1and EGF2 domains of MUC4 and the human EGF. (The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC). (B) Ribbon diagram showing the superimposition of the 3D structure of $MUC1_{SEA}$ with the 3D predicted structure of $MUC3A_{SEA}$, $MUC12_{SEA}$ and $MUC17_{SEA}$ (The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC).

Table 1. RMSD values of mucin EGF domains superimposition on alpha carbons. The 3D structure predictions of the mucins EGF domains was carried out with the Phyre 2 server (Protein Homology/analogY Recognition Engine V 2.0) by using the deduced amino-acids sequences of each domain. The predicted 3D structure of each EGF domain was compared with the resolved 3D structure of the human EGF (PDB code : 1jl9) and with the other EGF domains predicted structures using the molecular visualization system ,PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC). To determine whether the EGF domains were structurally homologous, PyMOL was used to measure the root-mean-square-deviation (*RMSD): the average distance between the alpha carbons atoms (the backbone atoms) of superimposed proteins. n.s:no 3D structure prediction. *RMSD : Root Mean Square Deviation is the square root of the mean of the square of the distances between the matched atoms. RMSD = SQRT[(SUM(dii)2)/N]. dii is the distance between the ith atom of structure 1 and the ith atom of structure 2 and N is the number of atoms matched in each structure

Table 2: RMSD values of mucin SEA domains superimposition on alpha carbons. The 3D structure predictions of the mucin SEA domains was carried out with the Phyre 2 server (Protein Homology/analogY Recognition Engine V 2.0) by using the amino-acids sequence of each domain. The predicted 3D structure of each SEA domain was compared with the resolved 3D structure of the human SEA domain of MUC1 (PDB code : 2ACM) and with the others SEA domains predicted structures using the molecular visualization system ,PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC). To determine whether the SEA domains were structurally homologous, PyMOL was used to measure the root-mean-square-deviation (*RMSD).

Table 3: **Sequence and functions of the cytoplasmic tails of membrane-bound mucins.**

Table 1

Table 1. RMSD values of mucin EGF domains superimposition on alpha carbons

EGF	Human EGF	MUC4_EGF1	MUC4_EGF2	MUC4_EGF3	MUC3A_EGF	MUC17_EGF	MUC12_EGF	MUC13_EGF1	MUC13_EGF2	MUC13_EGF3	MUC16_EGF
MUC4_EGF1	1.272 (25 atoms)		3.708 (21 atoms)	n.s	0.502 (15 atoms)	0.876 (13 atoms)	1.947 (14 atoms)	3.698 (12 atoms)	n.s	5.958 (30 atoms)	1.128 (27 atoms)
MUC4_EGF2	3.929 (24 atoms)	3.708 (21 atoms)		n.s	4.664 (24 atoms)	5.175 (35 atoms)	8.013 (16 atoms)	5.163 (26 atoms)	n.s	3.188 (14 atoms)	3.082 (30 atoms)
MUC4_EGF3	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	ns
MUC3A_EGF	0.502 (15 atoms)	2.177 (17 atoms)	4.664 (24 atoms)	n.s		4.776 (19 atoms)	7.122 (23 atoms)	0.148 (7 atoms)	n.s	8.884 (19 atoms)	2.559 (16 atoms)
MUC17_EGF	2.235 (22 atoms)	0.876 (13 atoms)	5.175 (35 atoms)	n.s	4.776 (19 atoms)		3.195 (23 atoms)	0.226 (4 atoms)	n.s	3.490 (18 atoms)	1.747 (21 atoms)
MUC12_EGF	2.093 (24 atoms)	1.947 (14 atoms)	8.013 (16 atoms)	n.s	7.122 (23 atoms)	3.195 (23 atoms)		1.455 (12 atoms)	n.s	1.812 (8 atoms)	2.643 (25 atoms)
MUC13_EGF1	0.793 (22 atoms)	3.698 (12 atoms)	5.163 (26 atoms)	n.s	0.148 (7 atoms)	0.226 (4 atoms)	1.455 (12 atoms)		n.s	3.272 (27 atoms)	2.087 (24 atoms)
MUC13_EGF2	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	n.s	ns
MUC13_EGF3	1.178 (7 atoms)	5.958 (30 atoms)	3.188 (14 atoms)	n.s	8.884 (19 atoms)	3.490 (18 atoms)	1.812 (8 atoms)	3.272 (27 atoms)	n.s		2.867 (28 atoms)
MUC16_EGF	2.803 (41 atoms)	1.128 (27 atoms)	3.082 (30 atoms)	n.s	2.559 (16 atoms)	1.747 (21 atoms)	2.643 (25 atoms)	2.087 (24 atoms)	n.s	2.867 (28 atoms)	

Table 2

Table 2. RMSD values of mucin SEA domains superimposition on alpha carbons

SEA	MUC1_SEA (2ACM)	MUC3A_SEA	MUC12_SEA	MUC13_SEA	MUC16-SEA	MUC17_SEA
MUC3A_SEA	3.812 (61 atoms)		4.615 (87 atomes)	11.740 (20 atoms)	1.759 (66 atoms)	0.004 (48 atoms)
MUC12_SEA	5.363 (66 atoms)	4.615 (87 atomes)		10.773 (45 atoms)	8.417 (82 atoms)	0.004 (56 atoms)
MUC13_SEA	11.740 (20 atoms)	13.996 (75 atoms)	10.773 (45 atoms)		3.361 (40 atoms)	6.417 (30 atoms)
MUC16_SEA	0.912 (44 atoms)	1.759 (66 atoms)	8.417 (82 atoms)	3.361 (40 atoms)		7.400 (48 atoms)
MUC17_SEA	1.204 (52 atoms)	0.004 (48 atoms)	0.004 (56 atoms)	6.417 (30 atoms)	7.400 (48 atoms)	

Table 3

apomucin	Cytoplasmic tail sequence	Protein binding-domains and phosphorylated tyrosines
MUC1	CQCRRKN**Y**GQLDIFPARDT**Y**HPMSE**Y**PT **Y**HTHGR**Y**VPPSSTDRSP**Y**EKVSAGNGGS SLS**Y**TNPAVAATSANL	7 Tyr SH2 domain c-Src, Lyn, Lck, ZAP70, c-Abl, Grb2, IKK, HSP70, HSP90, b-catenin, GSK3, FGFR3
MUC3	AVRSGWWGGQRRGRSWDQDRKWFET WDEEVVGTFSNWGFEDDGTDKDTNFH VALENVDTTMKVHIKRPEMTSSSV	SH2 domain
MUC4	LRFWGCSGARFS**Y**FLNSAEALP	1 Tyr
MUC12	SQRKRHREQ**Y**DVPQEWRKEGTPGIFQK TAIWEDQNLRESRFGLENA**Y**NNFRPTLE TVDSGTELHIQRPEMVASTV	2 Tyr
MUC13	VTARSNNKTKHIEEENLIDEDFQNLKLRS TGFTNLGAEGSVFPKVRITASRDSQMQN P**Y**SRHSSMPRPD**Y**	2 Tyr PKC
MUC15	GKRKTDSFSHRRL**Y**DDRNEPVLRLDNAP EP**Y**DVSFGNSS**YY**NPTLNDSAMPESEEN ARDGIPMDDIPPLRTSV	4 Tyr
MUC16	VTTRRRKKEGE**Y**NVQQQCPG**YY**QSHLD LEDLQ	3 Tyr moesin
MUC17	SI**Y**SNFQPSLRHIDPETKIRIQRPQVMTTS F	1 Tyr PDZK1
MUC20	ENGGFLLLRLSVASPEDLTDPRVAERLM QQLHRELHAHAPHFQVSLLRVRRG	c-Met
MUC21	RNSLSLRNTFNTAV**Y**HPHGLNHGLGPGP GGNHGAPHRPRWSPNWFWRRPVSSIAM EMSGRNSGP	1 Tyr
MUC22	SFCLRNLFFPLR**Y**CGI**YY**PHGHSHSLGLD LNLGLGSGTFHSLGNALVHGGELEMGH GGTHGFG**Y**GVGHGLSHIHGDG**Y**GVNHG GH**Y**GHGGGH	6 Tyr

Table 3: Sequence and functions of the cytoplasmic tails of membrane-bound mucins. Tyrosines that can be phosphorylated are bold.

Figure 1

Click here to download high resolution image

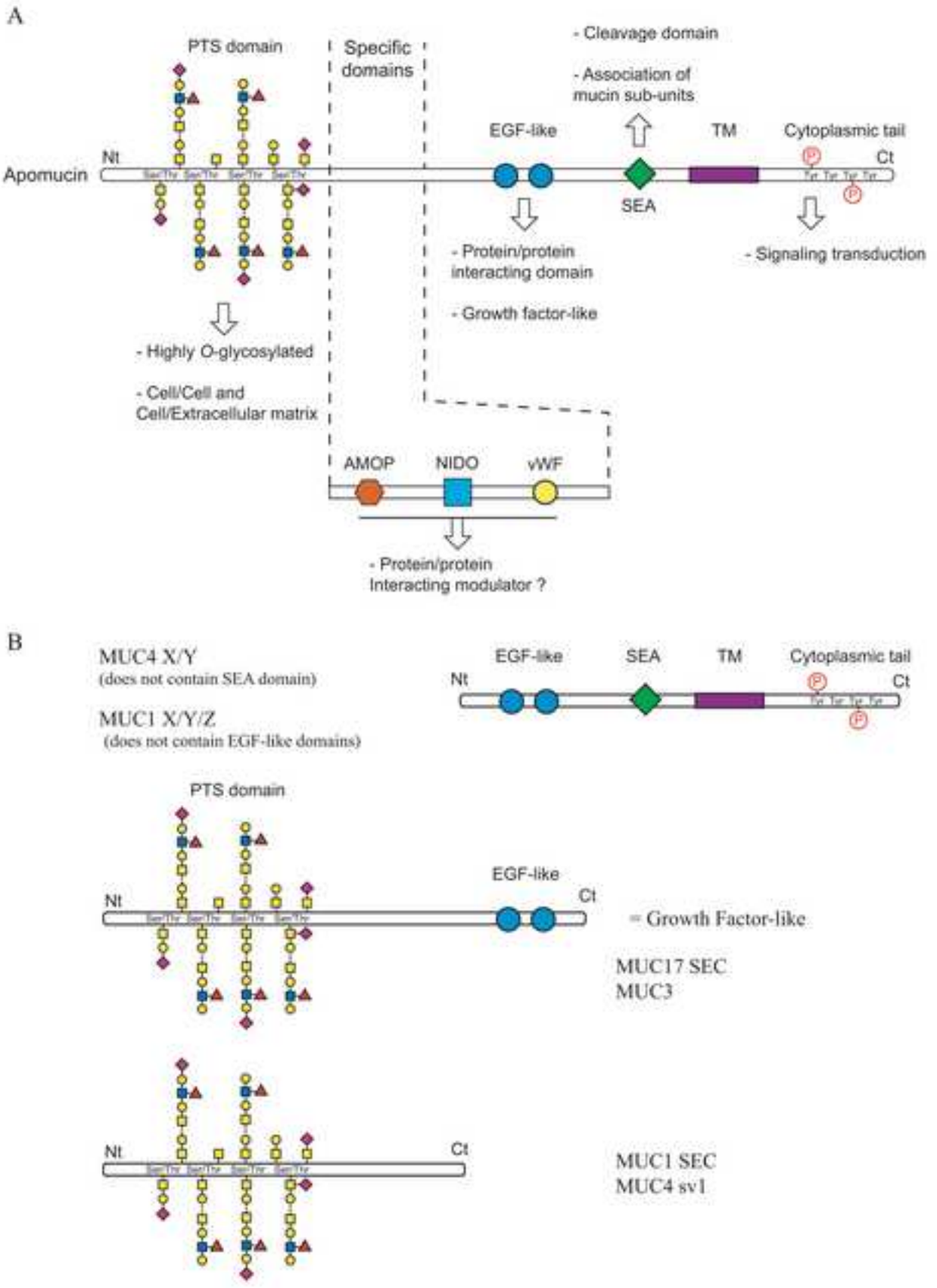

Figure 2

Click here to download high resolution image

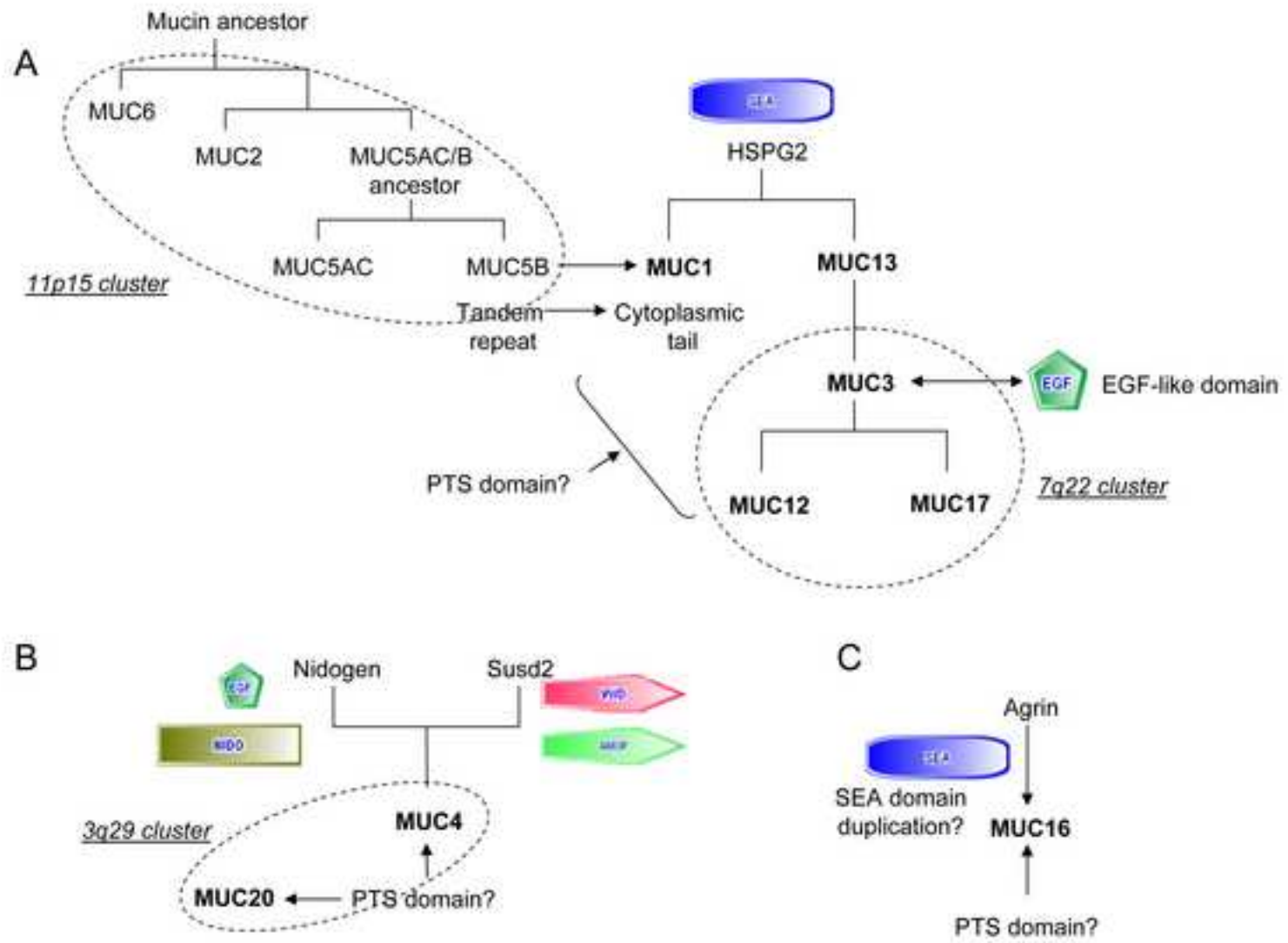

Figure 3

Click here to download high resolution image

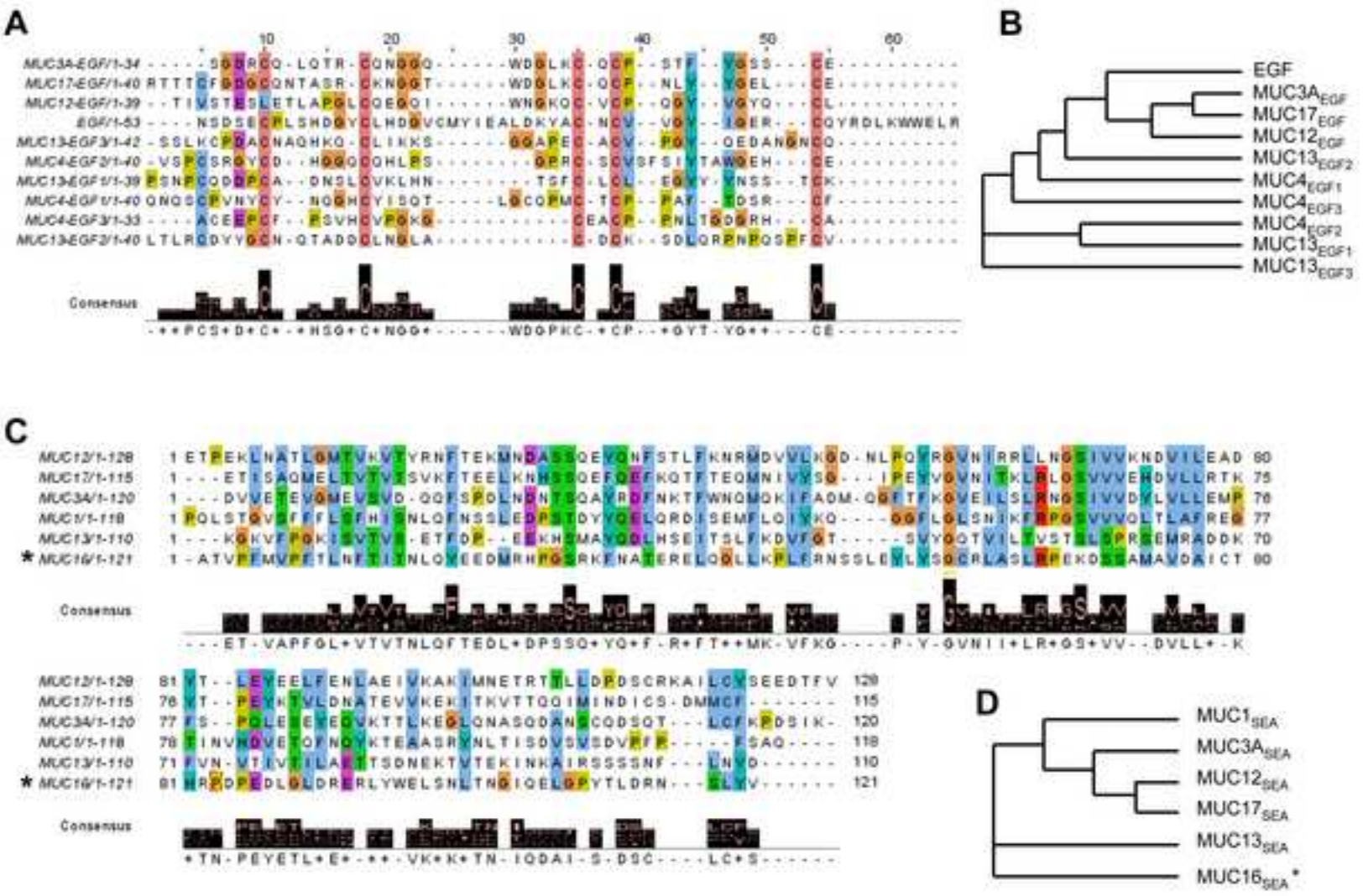

Figure 4

Click here to download high resolution image