Organisation du gène de mucine humaine *MUC5B* et bases moléculaires d'une nouvelle classification des gènes de mucines

Jean-Luc Desseyn et Anne Laine
Unité de "*Biologie & Physiopathologie des Cellules Mucipares*" INSERM U377 ;
Place de Verdun, 59045 Lille Cedex - FRANCE

Adresse :
Unité de "*Biologie & Physiopathologie des Cellules Mucipares*" INSERM U377 ; Place de Verdun, 59045 Lille Cedex – FRANCE
Tél. : +33 3 20 29 88 50 – Fax : +33 3 20 53 85 62 – e-mail : laine@lille.inserm.fr

Mots Clés : Mucines ; Facteur von Willebrand ; *MUC5B* ; TGF-β; EGF.

Abréviations
aa : acide aminé ; ASGP : Ascites sialomucin glycoproteins ; CK : Cystine knot ; EGF : Epidermal growth factor ; kb : kilobase ; ME : Matrice extracellulaire ; NDP : Norrie disease protein ; pb : paire de bases ; TGF : Transforming growth factor ; TR : Tandem repeat ; VNTR : Variable number of tandem repeat; vWF : facteur von Willebrand

Jean-Luc Desseyn a soutenu en octobre 1997 sa thèse de doctorat qu'il a preparée dans l'équipe de «*Génétique des Apomucines Humaines*» de l'unité de "*Biologie & Physiopathologie des Cellules Mucipares* " (INSERM U377 - Lille; Directeur. Pr. P. Degand). Il a travaillé sur l'organisation du gène de mucine humaine *MUC5B* sous la direction scientifique du Dr. Anne Laine. Il est actuellement en séjour postdoctoral à l'université de Washington dans le laboratoire de pharmacologie dirigé par le Pr. G. S. McKnight (Seattle, USA). Il developpe un modèle murin afin d'étudier le rôle biologique *in vivo* de la *N*-myristylation de la sous-unité catalytique Cα de la protéine kinase AMPc-dépendante (PKA).

Jean-Luc Desseyn s'est vu attribuer le prix Dina-Surdin 1998 par la *Société Française de Biochimie et Biologie Moléculaire* pour ses travaux de thèse.

Anne Laine est CR1 INSERM. Ses premiers travaux ont porté sur les constituants protéiques des sécrétions bronchiques humaines. Anne Laine a ensuite mis au point un procédé original de purification de l'α_1-antichymotrypsine humaine, inhibiteur de sérine-protéases alors peu connu. Elle en a déterminé les propriétés physico-chimiques et a étudié ses interactions avec les sérine-protéases impliquées dans l'inflammation. Il y a une dizaine d'années Anne Laine a rejoint le groupe '*Génétique des apomucines humaines'* dans lequel elle a pris la responsabilité des travaux de détermination de l'organisation génomique de certains gènes de mucines et de *MUC5B* en particulier. Elle se tourne maintenant vers l'étude des fonctions biologiques des mucines par l'obtention de protéines recombinantes correspondant aux domaines fonctionnels individualisés lors des études précédentes et isole les gènes murins de mucines pour aborder ensuite les fonctions *in vivo* de ces molécules complexes.

Regard sur la Biochimie 1999, 3 : 33–43

Le mucus est un terme générique désignant les sécrétions physiologiques visqueuses et viscoélastiques qui recouvrent les épithéliums des vertébrés (poissons, mammifères) et invertébrés (coelomates, mollusques). Chez les mammifères, le terme de mucus est restreint à la sécrétion qui recouvre et protège les épithéliums des tractus respiratoire, gastro-intestinal et urogénitaux ainsi que les muqueuses, sous-muqueuses et sécrétions des glandes exocrines. Ce mucus constitue l'interface entre l'environnement extérieur et l'épithélium. Les fonctions dévolues au mucus ne sont pas uniquement des fonctions de protection et/ou lubrification. En effet, le mucus a d'autres fonctions biologiques comme par exemple dans la locomotion des gastéropodes. Il joue aussi le rôle de barrière imperméable protégeant ainsi les organismes des changements soudains de la pression osmotique. Chez le ver par exemple, le mucus le pourvoit d'une paroi semi-perméable permettant le passage d'oxygène et de dioxyde de carbone. Le mucus est donc un système ancien de protection de l'organisme et ses rôles biologiques se sont diversifiés au cours de l'évolution. Le mucus joue, selon l'espèce, l'organe et le moment où il est produit, des rôles parfois contradictoires ; par exemple, si il est indispensable à la survie de la flore intestinale, il est aussi l'un des éléments majeurs responsables de l'épuration de l'épithélium bronchique et donc de l'élimination des bactéries. Cette adaptation évolutive doit donc être le fruit des variations des compositions biochimiques des différents mucus.

Chez les vertébrés, le mucus forme un tapis continu d'épaisseur variable à la surface des épithéliums, constituant ainsi une barrière de protection efficace entre l'environnement et la muqueuse. Ce film protecteur est formé de deux phases: une phase aqueuse au contact des cellules, de faible viscosité proche de celle de l'eau et une phase superficielle, ou phase gel, se caractérisant par une viscosité et une élasticité élevées. Le mucus est très hétérogène, il est constitué d'eau (95 %), d'électrolytes et de composés organiques, majoritairement les mucines, molécules qui confèrent au mucus ses propriétés rhéologiques.

Les mucines : généralités

Les mucines sont des *O*-glycoprotéines de haute taille moléculaire dont les chaînes glycanniques représentent plus de 50 % du poids sec. Les chaînes oligosaccharidiques sont liées aux nombreux résidus Ser et Thr de l'axe peptidique de la mucine (encore appelé apomucine) via des *N*-acétylgalactosamines. Ces chaînes oligosaccharidiques, composées de galactose, de glucose, de fucose, de *N*-acétylglucosamine, de *N*-acétylgalactosamine et d'acide sialique, maintiennent la mucine sous une forme linéaire [1,2].

Du fait de la structure même des mucines, la détermination de leurs séquences peptidiques par la dégradation de Edman s'est révélée inopérante. Les techniques de l'ADN recombinant ont permis de pallier les difficultés de l'étude des apomucines. Dès la fin des années 80, plusieurs laboratoires se sont lancés avec plus ou moins de succès dans le clonage d'ADNc partiels de mucines puis des gènes correspondants. Depuis, 8 gènes ont été identifiés chez l'Homme et dénommés *MUC1* à *MUC7*, incluant *MUC5B* et *MUC5AC* [3]. Des homologues chez l'animal (Tableau I) ont parfois été caractérisés (partiellement).

Tableau I : ***Les gènes de mucines humaines et leurs homologues chez l'animal.***

Gène humain	Homologue(s) chez l'animal
MUC1	Rat, souris, hamster etc…
MUC2	Rat, souris
MUC3	Rat, souris
MUC4	Rat
MUC5B	
MUC5AC	Rat, souris, porc
MUC6	
MUC7	Rat, souris

Les séquences déduites montrent que l'axe peptidique des mucines est constitué de motifs répétés riches en résidus Ser, Thr, Pro et Ala. Ces motifs répétés varient en longueur et dans leur séquence d'un gène à l'autre, d'une espèce à l'autre (Tableau II). La plupart des gènes de mucines présentent un polymorphisme de type VNTR, c'est-à-dire que le nombre de répétitions varie d'un individu à l'autre. Le fait que les gènes de mucines possèdent ces motifs répétés en tandem et codent de longs transcrits accroît à la fois

l'instabilité des clones et les difficultés de séquençage (répétition des séquences et des sites de restriction).

Tableau II : ***Séquences peptidiques consensus des domaines répétitifs des mucines humaines.***

Gène	pb	aa	Séquence consensus
MUC1	60	20	PDTRPAPGSTAPPAHGVTSA
MUC2	69	23	PTTTPITTTTTVTPTPTPTGTQT
MUC3	51	17	HSTPSFTSSITTTETTS
MUC4	48	16	TSSASTGHATPLPVTD
MUC5AC	24	8	TTSTTSAP
MUC5B	87	29	AT(G/S)STATPSS(T/S)PGT(T/A)(H/W)T (P/L)(P/T)VL(T/S)(T/S)T(A/T)TT(P/T)T
MUC6	507	169	SPFSSTGPMTATSFQTTTTYPTPSHP QTTLPTHVPPFSTSLVTPSTGTYITP THAQMATSASIHSTPTGTIPPPTTLK ATGSTHTAPPMTPTTSGTSQAHSSFS TAKTSTSLHSHTSSTHHPEVTPTSTT TITPNPTSTGTSTPVAHTTSATSSRL PTPFTTHSPPTGS
MUC7	69	23	TTAAPPTPSATTPAPPSSSAPPE

Il a été montré une spécificité tissulaire de l'expression des gènes de mucines, mais aussi une spécificité cellulaire. Par exemple, *MUC4* s'exprime indifféremment dans les cellules caliciformes et les cellules ciliées de la plupart des muqueuses ainsi que dans les entérocytes alors que, au contraire, *MUC5B* a une expression tissulaire limitée aux glandes sous-maxillaires et bronchiques, à la vésicule biliaire, au pancréas et à l'endocol. L'ensemble des résultats est résumé par le Tableau III. Il est aussi intéressant de noter que les mucines sont des molécules oncofœtales qui s'expriment au cours du développement embryonnaire et fœtal avant même que n'existent les phénomènes de la mucosécrétion [8, 9].

Outre des altérations de la fraction glycosylée, des sous- et/ou sur-expressions des mucines ont été trouvées dans de nombreuses pathologies touchant les épithéliums dont les cancers épithéliaux qui représentent plus de 90% des cancers chez l'Homme [10]. L'expression des mucines humaines en situation normale et pathologique est étudiée par hybridation *in situ*, Northern et Dot blots et en immunohistochimie depuis plusieurs années et les résultats confirment l'intérêt que peuvent avoir ces gènes comme marqueurs de diagnostic et/ou de pronostic.

Mucines, «mucin-like» et «mucin-type»

On distingue généralement les mucines sécrétées (MUC2, MUC5B, MUC5AC, MUC6, MUC7 chez l'Homme et leurs homologues animaux) et les mucines membranaires, encore appelées "mucin-like" (MUC1, MUC3, MUC4 chez l'Homme et leurs homologues animaux). Cependant, ces mucin-like, au moins MUC1 et SMC (constituée de ASGP1 et ASGP2), l'homologue chez le rat de MUC4, sont dans certaines conditions protéolysées relarguant la fraction extracellulaire fortement *O*-glycosylée. De plus, il semble que certaines de ces mucines transmembranaires peuvent aussi être sécrétées lorsque l'exon codant le domaine transmembranaire est épissé. La distinction entre mucines sécrétées et mucines membranaires est donc insuffisante.

Des *O*-glycoprotéines de type mucine ont été trouvées associées à des cellules humaines non-épithéliales, comme par exemple la glycophorine, glycoprotéine intégrée dans la membrane des hématies ou encore la leucosialine (parfois appelée sialophorine ou encore CD34) qui s'exprime à la surface des cellules B, T, des macrophages, des plaquettes et des granulocytes. Ces mucines, parfois aussi appelées "mucin-type", jouent le rôle de ligand pour les sélectines [pour revue voir [11].

Des mucin-like sont aussi exprimées à la surface du protozoaire *Trypanosoma cruzi*. L'apomucine est essentiellement constituée d'un motif, riche en résidus Thr (motif trouvé dans la souche CA1/72 : TTTTTTTTKPP). Certains de ces motifs sont potentiellement *O*-glycosylables; ce motif, non conservé d'une souche à l'autre, est répété en tandem dans la région centrale de l'apomucine. Le nombre de motifs parfaits est variable selon les parasites d'une même souche; ce motif est donc de type VNTR [12]. Une famille de gènes similaires comprenant au moins 10 membres a été décrite chez *Leishmania major* [13]. Parce que la fonction de ces mucines ne correspond pas à la définition donnée plus haut quant à leur rôle, nous nous focaliserons désormais uniquement sur les mucines exprimées à la surface des épithéliums.

Proposition d'une nouvelle classification des mucines

Les études structurales des gènes de mucines chez l'Homme et l'animal montrent que les mucines sont des protéines mosaïques dont

Tableau III : ***Profil d'expression des gènes de mucines chez l'adulte par hybridation in situ [4-7].***

MUQUEUSES	SONDES						
HUMAINES	**MUC 2**	**MUC 3**	**MUC 4**	**MUC 5B**	**MUC 5AC**	**MUC 6**	**MUC 7**
Glandes salivaires	G : -	G : -	G : -	G : +/++	G : -	G : -	G : +++
Bronches	S : ++	S : -	S : ++	S : -	S : -/+++	S : -	S : -
	G : -	G : -	G : -	G : ++	G : -	G : -	G : +++
Fundus	S : -	S : -/+	S : -/+	S : -	S : ++++	S : - Ct : +	S : -
	G : -	G : -	G : -	G : -	G : -	G : -	G : -
Antre pylorique	S : -	S : ++	S : -/+	S : -	S : ++++	S : -	S : -
	G : -/+	G : +	G : -	G : -	G : -	G : ++/+++	G : -
Duodénum	S : ++++	S : +++	S : -	S : -	S : -	S : -	S : -
	C : ++++	C : -/+	C : -	C : -	C : -	C : -	C : -
	G : +	G : -	G : -	G : -	G : -	G : ++/+++	G : -
Jéjunum	S : ++++	S : +++	S : -	S : -	S : -	S : -	S : -
	C : ++++	C : -/+	C : -	C : -	C : -	C : -	C : -
Iléon	S : ++++	S : +++	S : -/+	S : -	S : -	S : -	S : -
	C : ++++	C : -/+	C : -/+	C : -	C : -	C : -	C : -
Côlon	S : ++++	S : ++	S : ++	S : -	S : -	S : -	S : -
	C : ++++	C : -	C : ++	C : -	C : -	C : -	C : -
Vésicule biliaire	S : +	S : +++	S : -	S : +	S : +	S : -/+	S : -
	I : +	I : +++	I : -	I : ++	I : +	I : +++	I : -
Pancréas	D : -/+	D : +++	D : -	D : ++	D : -	D : +	
Prostate	G : -	G : -	G : ++	G : -	G : -	G : -	G : -
Endocol	S : +	S : -	S : ++	S : +	S : ++	S : +	G : -
	G : +	G : -	G : ++	G : +	G : ++	G : +	S : -

S : épithélium de surface; G : épithélium glandulaire; C : cryptes; I : invaginations de l'épithélium; Ct : collet; D : cellules des canaux.
Marquage : ++++ : de très forte intensité; +++ : de forte intensité; ++ : d'intensité modérée; + : de faible intensité; -/+ : quelques cellules; - : absent,

certaines présentent des domaines de structure semblable. Ces données permettent à présent de proposer une nouvelle classification des mucines qui prend en compte la présence de certains domaines caractéristiques dans les séquences déduites des gènes correspondants. A côté de la petite mucine MUC1, on peut distinguer 3 familles de mucines : les mucines solubles, les mucines ayant 2 motifs EGF-like et les mucines ayant 1 motif TGF-β.

• MUC1

Le gène *MUC1* est bien caractérisé. Il est composé de 7 exons et s'étend sur 4 à 7 kb, selon le nombre variable de répétitions du motif de 60 pb. La protéine possède une séquence hydrophobe transmembranaire, mais une forme soluble de MUC1, engendrée par l'action d'une protéase, a aussi été décrite [14]. Deux autres formes dépourvues du domaine répétitif *O*-glycosylé ont été observées (pour revue [3]). Ces deux formes ne peuvent pas être réellement considérées comme des mucines et leur rôle est inconnu. Enfin, il a été suggéré que la queue cytoplasmique interagit avec des protéines de structure du cytosquelette [15].

• Les mucines solubles

L'organisation génomique de *MUC7* a été entièrement élucidée [16, 17]. Le gène s'étend sur 10 kb et comprend 3 exons. La protéine MG2 (377

aa), issue du gène *MUC7*, est sécrétée spécifiquement par les cellules séreuses des glandes salivaires et bronchiques et n'est pas impliquée, directement, dans la formation du gel de mucus car elle n'établit pas de ponts interchaînes. Ceci n'exclut cependant pas des interactions non covalentes avec des mucines qui forment le gel. De plus, cette molécule est fortement *O*-glycosylée et donc très hydrophile. Nous pouvons la qualifier de «mucine soluble». Il est probable que cette mucine joue essentiellement un rôle dans l'interaction avec les bactéries et d'autres molécules dont les IgAs (pour revue, voir [18]).

Chez le rat, un gène *Rmucsmg* constitué de 3 exons codant une petite mucine de 243 aa a été cloné [19]. Cette mucine soluble salivaire semble être l'homologue de MUC7. A partir des séquences codantes de *Rmucsmg* a été cloné l'ADNc de l'homologue murin *Mucsmg* [20].

- Les mucines ayant des domaines EGF-like

Les régions centrale et carboxy-terminale de MUC3 ont été partiellement caractérisées. La mucine possède dans sa région carboxy-terminale (617 aa) un motif riche en résidus Cys qui est de type EGF-like [21]. Très récemment, un deuxième motif EGF-like et une région transmembranaire ont été retrouvés (Crawley *et al.,* soumis).

Les ADNc des régions 3' des gènes MUC3 de rat (*RMuc3*) et de souris (*Muc3*) ont été clonés [22, 23]. Ces ADNc codent chacun 2 motifs EGF-like et un peptide potentiellement transmembranaire.

Un ADNc complet de *MUC4* codant une apomucine possédant entre 4468 et 8468 acides aminés, selon le nombre variable de répétitions du motif élémentaire de 16 aa, vient d'être publié [24, 25]. Comme MUC3 et ses homologues chez le rat et la souris, MUC4 possède dans sa région carboxy-terminale 2 motifs de type EGF et une séquence hydrophobe potentiellement transmembranaire. Les régions non répétitives de MUC4 présentent des similarités de séquence avec les régions non répétitives de SMC suggérant que SMC est probablement l'homologue de MUC4.

SMC est maintenant très bien caractérisée et est à ce titre un excellent modèle d'étude des mucines ayant des domaines EGF-like. Les deux composantes de SMC que sont ASGP-1 et ASGP–2 ("Ascites Sialomucin Glycoproteins") sont 2 *O*-glycoprotéines naturellement exprimées à la surface des cellules épithéliales. Elles ont été isolées à la surface de cellules d'un adénocarcinome mammaire chez la rate [26]. Les 2 glycoprotéines ASGP-1 et -2 sont issues d'un même précurseur polypeptidique. En s'appuyant sur des études biochimiques un modèle a été proposé dans lequel le précurseur peptidique appelé pSMC-1 (Figure 1; [27]) est synthétisé dans le réticulum endoplasmique et *N*-glycosylé [28].

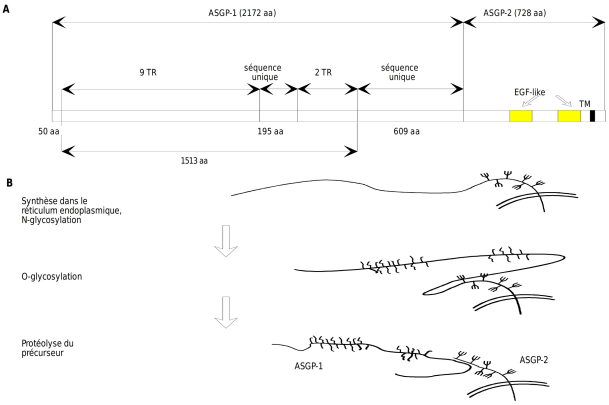


Figure 1. ***Schéma d'organisation peptidique de SMC (A) et modèle de biosynthèse avant la réinternalisation (B).***

Ce précurseur est ensuite *O*-glycosylé et est clivé pour donner les 2 entités ASGP-1 et -2, et ce probablement avant d'arriver dans le compartiment médian du Golgi. Ces 2 molécules restent associées par des liaisons non covalentes. Il a ensuite été montré que l'ASGP-1, une fois à la surface de la cellule épithéliale sécrétrice, peut être internalisée rapidement dans un compartiment intracellulaire de la cellule où il y a addition et élongation de nouvelles chaînes oligosaccharidiques [29]. Ces glycoprotéines jouent un rôle dans la protection de ces cellules cancéreuses. L'augmentation de l'expression des ASGP et leur mauvais adressage font qu'elles sont impliquées dans le processus métastatique et leur expression est associée à la résistance de la cellule cancéreuse aux cellules NK [30]. L'ADNc complet a été séquencé [31]. Le transcrit a une taille de 9270 pb. Le complexe ASGP-1/ASGP-2 existe sous une forme sécrétée et une forme membranaire dans les glandes mammaires et le lait, mais uniquement sous une forme soluble dans le côlon [32]. Cette dernière forme résulte de l'absence du

domaine carboxy-terminal de l'ASGP-2, c'est-à-dire de la région transmembranaire. La sous-unité ASGP-2 contenant le domaine transmembranaire possède 2 motifs EGF-like [33] dont l'un est similaire aux motifs EGF ayant une activité facteur de croissance. La sous-unité ASGP-2 peut moduler la phosphorylation et l'activité du récepteur tyrosine-kinase $p185^{neu}$/ErbB2 par interaction directe entre l'ASGP-2 et ErbB2 via un motif EGF de l'apomucine [34].

• Les mucines ayant un domaine TGF-β

Des 8 gènes de mucines humaines, 4 ont été localisés sur un fragment génomique de 400 kb en 11p15.5. Une étude en électrophorèse en champs pulsés a permis de déterminer l'ordre de ces 4 gènes et de les localiser par rapport aux autres gènes de la région [35] : Télomère-*HRAS-MUC6-MUC2-MUC5AC-MUC5B-IGF2*-Centromère. Ces mucines de masses moléculaires élevées sont sécrétées. Elles sont synthétisées par des cellules glandulaires de la sous-muqueuse et des cellules épithéliales spécialisées, appelées cellules sécrétrices au niveau des épithéliums bronchiques, des cellules caliciformes dans l'épithélium colique ou cellules à pôle fermé au niveau de l'estomac. Ces mucines sont sécrétées sous forme de granules, qui, libérés dans la lumière, forment au contact de l'eau un gel viscoélastique à la surface de l'épithélium. Une caractéristique essentielle des mucines sécrétées est leur structure oligomérique. La réduction des ponts disulfures provoque une diminution de la viscosité du mucus ; l'intégrité de ces ponts disulfures est indispensable à la fonction de ces mucines. Une séquence de 15720 pb d'un transcrit "virtuel" (puisque la séquence de la partie centrale n'a été qu'extrapolée) du gène *MUC2* codant une apomucine de 5177 aa a permis pour la première fois de schématiser sur des bases moléculaires une mucine sécrétée humaine. La partie centrale de MUC2 est constituée d'une énorme région purement répétitive de 2 à 3000 aa de type VNTR précédée d'un domaine répétitif plus petit, irrégulier et non VNTR. Ce domaine est bordé de part et d'autre par 2 motifs riches en résidus Cys (108 aa, 10 Cys) retrouvés dans MUC5AC, MUC5B et les homologues animaux. La séquence de la région répétitive a essentiellement été déduite d'après des expériences de Southern blot utilisant des coupures enzymatiques totales et partielles et une sonde reconnaissant la région répétitive. La partie carboxy-terminale possède un petit domaine de type mucine de 139 aa suivi d'un long domaine de 845 aa riche en résidus Cys et possédant des sites potentiels de *N*-glycosylation. Le domaine amino-terminal est lui aussi très complexe, riche en résidus Cys et présente, comme le domaine carboxy-terminal, des homologies avec certains domaines du pro-facteur de vWF [36-40].

Des ADNc des régions qui bordent la partie centrale de RMuc2 (MUC2 de rat) ont été publiés [41-45] ainsi que l'ADNc tout à fait 5' de MUC2 murin (Muc2) [46]. Les séquences montrent que ces ADNc sont très similaires à l'ADNc de *MUC2* et codent tous des domaines retrouvés dans le vWF ainsi que 2 motifs CYS pour RMuc2.

Quelques éléments d'organisation génomique et un ADNc 3' de *MUC6* ont été publiés [47, 48]. La partie en aval de la région centrale portant le motif répété en tandem caractéristique de MUC6 contient 2 exons codant une région riche en résidus Ser, Thr et Pro mais pauvre en résidus Cys. Cette région est suivie d'un domaine d'une centaine d'acides aminés contenant 10 résidus Cys aux positions conservées par rapport aux séquences carboxy-terminales de MUC2, MUC5B, MUC5AC et du vWF (voir plus loin).

Plusieurs ADNc de MUC5AC chez l'Homme, la souris et le rat ont été clonés ces dernières années [3] et suggèrent que MUC5AC a une construction très semblable à celle de MUC5B . Nous reviendrons dans la suite de cet article sur les connaissances acquises au laboratoire sur MUC5AC.

• MUC5B : modèle d'étude des mucines sécrétées

Nous avons déterminé la séquence complète du gène *MUC5B*. C'est, après *MUC1* et *MUC7* qui sont des gènes relativement petits, la seule séquence entièrement connue pour un très grand gène de mucine puisqu'il s'étend sur une quarantaine de kilobases. MUC5B est donc un excellent modèle d'étude de mucine sécrétée. Deux cosmides d'environ 45 kb chevauchants sur 10 kb permettent de disposer du gène entier. Le transcrit de *MUC5B* a une taille de 17,6 kb et ceci est en bonne corrélation avec les études en Northern blot

[49]. La région centrale de MUC5B [50] code un peptide déduit de 3571 aa composé majoritairement (3254 aa sur les 3571 aa) de l'agencement en alternance de 3 sous-domaines (Figure 2) :

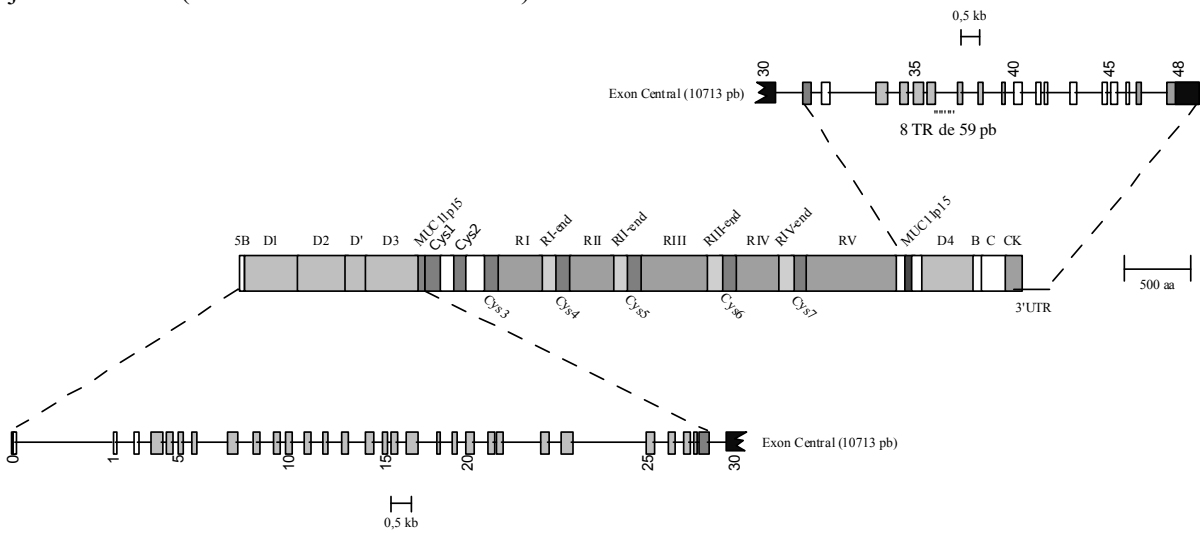


Figure 2 : ***Organisations génomique et peptidique déduite de MUC5B.***

- un domaine de 108 aa, appelé sous-domaine CYS («Cys subdomain»), riche en résidus Cys (10 Cys soit 9,3%) et trouvé 7 fois dans MUC5B, 2 fois dans MUC2, au moins 6 fois dans MUC5AC (Figure 3). Ce domaine est également retrouvé dans des mucines animales *Muc5ac*, *RMuc2*, *RMuc5ac* et PGM-2A (Pig Gastric Mucin). Ce domaine est pauvre en résidus Thr et Ser.
- un domaine de 111 aa, appelé R-end, retrouvé 4 fois dans MUC5B et très conservé. Ce domaine est riche en résidus Thr et Ser (Thr+Ser=48,6%) et en résidus Pro (14,2%).
- un domaine trouvé 5 fois et appelé R. Ce domaine est constitué de la répétition en tandem (11, 17 ou 23 fois) du motif élémentaire et irrégulier de 29 aa (87 pb). Ces sous-domaines R sont riches en résidus Ser et Thr, ainsi qu'en résidus Pro.

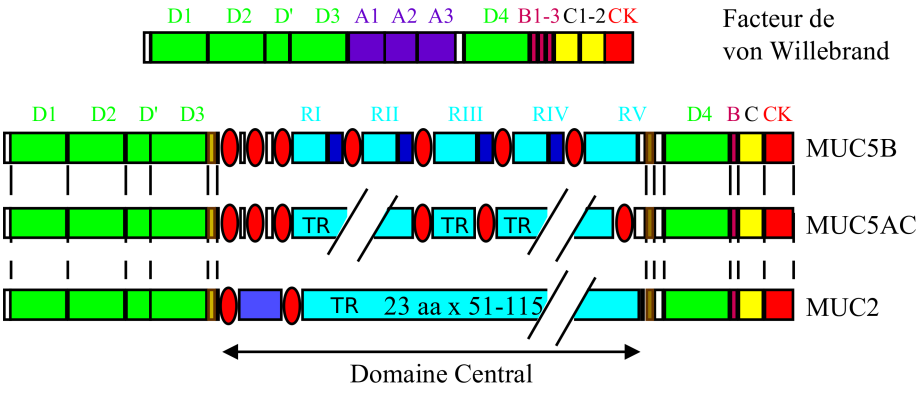


Figure 3 : ***Organisations peptidiques comparées du facteur de von Willebrand, de MUC5B, MUC5AC et MUC2.Les régions centrales de MUC2 et MUC5AC n'ont pas été clonées. Les ovales représentent les sous-domaines CYS (108 aa; 10 résidus Cys).***

Puisque MUC5B possède une alternance de plusieurs domaines distincts et qu'un motif donné répété en tandem est rarement 100% homologue à un autre motif (motifs imparfaits) nous avons pu sous-cloner et séquencer entièrement cette région. Ces caractéristiques nous ont permis de plus de montrer que la région centrale de MUC5B est codée par une énorme région exonique (pas d'intron) et que cette région exonique de 10,7 kb n'a pas de site d'épissage

interne. Cet exon, le plus grand jamais caractérisé chez l'Homme (10713 pb), est équidistant d'une dizaine de kilobases d'un îlot CpG en 5' et d'un signal de polyadénylation en 3'. Dans la littérature, bon nombre d'auteurs qui ont étudié de grands transcrits pensent souvent avoir caractérisé en même temps un énorme exon. Cependant, pour la plupart de ces gènes, aucun élément de preuve indiscutable n'a pu être rapporté car ces gènes codent presque toujours des motifs répétés en tandem. La taille des fragments génomiques portant ces motifs est alors estimée par Southern blot avec des coupures de restriction partielles et totales et le nombre de motifs et la séquence sont extrapolés. Cette méthode a l'inconvénient de ne pas pouvoir détecter d'éventuels petits introns ni d'exclure qu'une partie du cadre ouvert de lecture (extrapolé le plus souvent), même contenant des motifs répétés en tandem, ne soit en fait intronique. Grâce aux technologies actuelles, il semble possible de montrer l'existence d'un énorme exon, si et seulement si ce putatif exon possède des motifs répétés mais non parfaitement identiques entre eux. En effet, ceci permet alors d'effectuer des RT-PCR chevauchantes et parallèlement des PCR sur de l'ADN génomique. On peut alors avoir la quasi-certitude qu'une grande région est totalement exonique et qu'elle ne possède ni site d'épissage alternatif 5' donneur interne, ni site d'épissage alternatif 3' accepteur interne.

La région 3' du gène *MUC5B* (10,6 kb) (voir Figure 2) comprend 18 introns (nommés tout d'abord de A à R, puis de 30 à 47) qui ont des tailles comprises entre 114 et 1118 pb [51]. L'intron 36 (7^{ième} intron en aval de l'exon central) a une taille de 538 pb. Il est majoritairement constitué d'un motif parfait de 59 pb répété 8 fois en tandem. Le premier motif est en partie sur l'exon précédant l'intron 36. Ces motifs, très riches en G+C, contiennent tous une boîte GC. Une protéine nucléaire de 42 kDa [52] reconnaît cette boîte et pourrait être impliquée dans la régulation de *MUC5B* ou d'un gène proche de *MUC5B*. Par ailleurs, le nombre de répétitions de ce motif varie d'un individu à l'autre [53] ce qui peut se révéler important dans le cadre de l'étude de la régulation de l'expression. La partie 3' de *MUC5B* code un peptide de 808 aa qui présente des similarités avec la région carboxy-terminale du vWF. Ainsi, MUC2, MUC5AC et MUC5B possèdent des régions carboxy-terminales très homologues entre elles, tout particulièrement en ce qui concerne les positions des résidus Cys. L'analyse des séquences peptidiques permet de subdiviser en 6 domaines la région carboxy-terminale de MUC5B comparativement aux régions carboxy-terminales de MUC5AC et MUC2 [51](Figure 3) :

- un premier domaine appelé «MUC11p15-type» car retrouvé uniquement dans ces 3 apomucines
- un petit domaine commun aux 3 mucines et au vWF, situé dans ce dernier entre le domaine A3 et le domaine D4 et nommé A3uD4.
- un domaine D4-like semblable à celui du vWF
- un domaine B-like unique au lieu des 3 domaines B du vWF
- un domaine C-like unique au lieu des 2 domaines C1 et C2 du vWF
- un domaine CK ("pour Cystine Knot") qui fait de MUC5B, comme de MUC5AC, MUC2 et MUC6, des membres de la mégafamille des protéines à CK (famille TGF-β). Ce domaine est impliqué dans la dimérisation des protéines.

Le travail de détermination de l'organisation génomique de la région 3' de *MUC5AC* mené en parallèle au laboratoire a bénéficié des connaissances acquises sur *MUC5B* puisque les positions des introns sont conservées entre les 2 gènes [54]. *MUC5AC* présente une structure plus proche de celle de MUC5B que de celle des 2 autres gènes de mucines situés dans le même cluster [55].

La majeure partie de l'ADNc 5' de *MUC5B* a été clonée grâce à des expériences de RT-PCR utilisant des oligonucléotides dégénérés. L'ADNc total a été obtenu par des expériences complémentaires de RACE-PCR et RT-PCR. Parallèlement, l'organisation génomique complète de cette région (15 kb) qui comprend 29 exons a été déterminée [56]. Récemment, un ADNc 5' de *MUC5B* plus long a été publié [57], suggérant que le gène possède en amont de l'exon 1 un second site d'initiation de la transcription. Un nouvel exon de 128 pb (exon 0) est séparé de l'exon 1, par un intron de 2,4 kb (résultat non publié). Le peptide déduit, pour toute la région amino-terminale, est constitué de 1322 aa et est riche en résidus Cys . Il est homologue aux régions amino-terminales de MUC2, de MUC5AC et du pro-vWF (Figure 3).

Le polymorphisme VNTR n'est plus une caractéristique de toutes les mucines

En utilisant des oligonucléotides choisis dans l'exon central de *MUC5B* et des oligonucléotides qui bornent cet exon central, des RT-PCR ont été effectuées sur de l'ARN provenant de différents individus. Par cette méthode, aucun polymorphisme de type VNTR n'a pu être visualisé. Le nombre de répétitions en tandem du motif de 87 pb est donc très peu ou pas variable. Ceci a par ailleurs été confirmé par Southern blot en utilisant la sonde répétitive de *MUC5B* [58]. Il est souvent admis que le nombre de répétitions nucléotidiques en tandem de la région codante des mucines varie d'un individu à l'autre. Certains auteurs font de ce polymorphisme VNTR un critère de définition des mucines ([11] par exemple). Ceci amenait à penser que la taille de l'axe peptidique n'est pas cruciale pour la fonction de la molécule [3], or il semble bien que le critère de polymorphisme ne convienne ni à *MUC7* [59], ni à *MUC5B*. L'absence de polymorphisme VNTR se conçoit plus aisément pour un petit gène comme *MUC7*. Il est, par contre, beaucoup plus difficile d'admettre l'absence d'un tel polymorphisme sur de très grandes régions construites de motifs répétés en tandem. Il doit donc probablement exister un "phénomène" génomique qui préserve *MUC5B* d'une augmentation ou au contraire d'une diminution du nombre de répétitions. La taille de la région répétitive de *MUC5B* est certainement un facteur important dans son rôle biologique. Il n'en est pas exactement de même pour les 3 autres gènes du cluster MUC en 11p15.5. En effet, *MUC6* est, de ces 4 gènes, le plus polymorphique (VNTR), puis vient *MUC2*. *MUC5AC* n'est pas très polymorphique et enfin *MUC5B* ne présente pas de polymorphisme VNTR dans l'ADNc. En comparant les 4 gènes de mucines localisés en 11p15 on remarque que cette différence entre ces 4 gènes peut être reliée à la présence ou non des régions codant les motifs dénommés sous-domaines CYS («Cys-subdomains»). Ces domaines ont été retrouvés 2 fois dans MUC2, au moins 6 fois dans MUC5AC et exactement 7 fois dans MUC5B. Il n'en existe probablement pas dans MUC6. Ainsi, moins il y a de ces sous-domaines, plus le gène correspondant montre un polymorphisme important de type VNTR dans sa région centrale [58].

Caractéristiques structurales et phylogénèse des mucines du chromosome humain 11

MUC2, MUC5AC et MUC5B possèdent des domaines D (D1, D2, D', D3 et D4) retrouvés dans le pro-vWF. Sur le modèle du pro-vWF [60], ces molécules polymériseraient par ces domaines [61, 62]. Les organisations peptidiques du pro-vWF, de MUC2, MUC5AC et de MUC5B sont schématisées sur la Figure 3. Ces 3 mucines ont par ailleurs, dans leur domaine central, un domaine appelé sous-domaine CYS (108 aa; 10 Cys ; Figure 3). Ce domaine a été retrouvé jusqu'à présent dans ces mucines humaines mais aussi dans des mucines animales. Nos analyses phylogénétiques des séquences de ces motifs, l'ordonnancement de ces motifs (Figure 4) et des gènes sur le chromosome 11 montrent que *MUC5AC* et *MUC5B* sont plus proches l'un de l'autre dans l'évolution que de *MUC2*. Ceci amène à penser que *MUC5AC* et *MUC5B* ont un gène progéniteur commun [55]. Ce gène provient d'un gène ancestral commun également à *MUC2*.

Les 4 mucines dont les gènes sont localisés sur le chromosome 11 sont très similaires et possèdent toutes dans leur région tout à fait carboxy-terminale un domaine d'une centaine d'acides aminés comportant 11 résidus Cys aux positions très conservées (Figure 5). Ce motif, appelé CK pour «cystine knot», a été retrouvé dans des mucines animales (mucine tégumentaire FIM-B.1 de xénope, mucines sous-maxillaires de porc et de bovin, mucines RMuc2 et RMuc5ac) et est très similaire au NDP ("Norrie Disease Protein"). L'alignement des motifs CK conforte l'arbre phylogénétique proposé ci-dessus et montre que *MUC6* est, des 4 gènes de mucines du chromosome 11, le plus éloigné [63]. La structure tridimensionnelle du NDP déterminée par modélisation moléculaire est superposable à la structure 3-D expérimentale du TGF-β obtenue par cristallographie [64, 65]. Ainsi, la structure et le rôle biologique du NDP suggèrent que cette protéine pourrait jouer le rôle de facteur de croissance. Des études préliminaires utilisant un protocole de modélisation par homologie montrent que le domaine TGF-β de MUC5B est superposable à la structure 3D du TGF-β obtenue par cristallographie ([63] et R. Effremov, résultats non publiés).

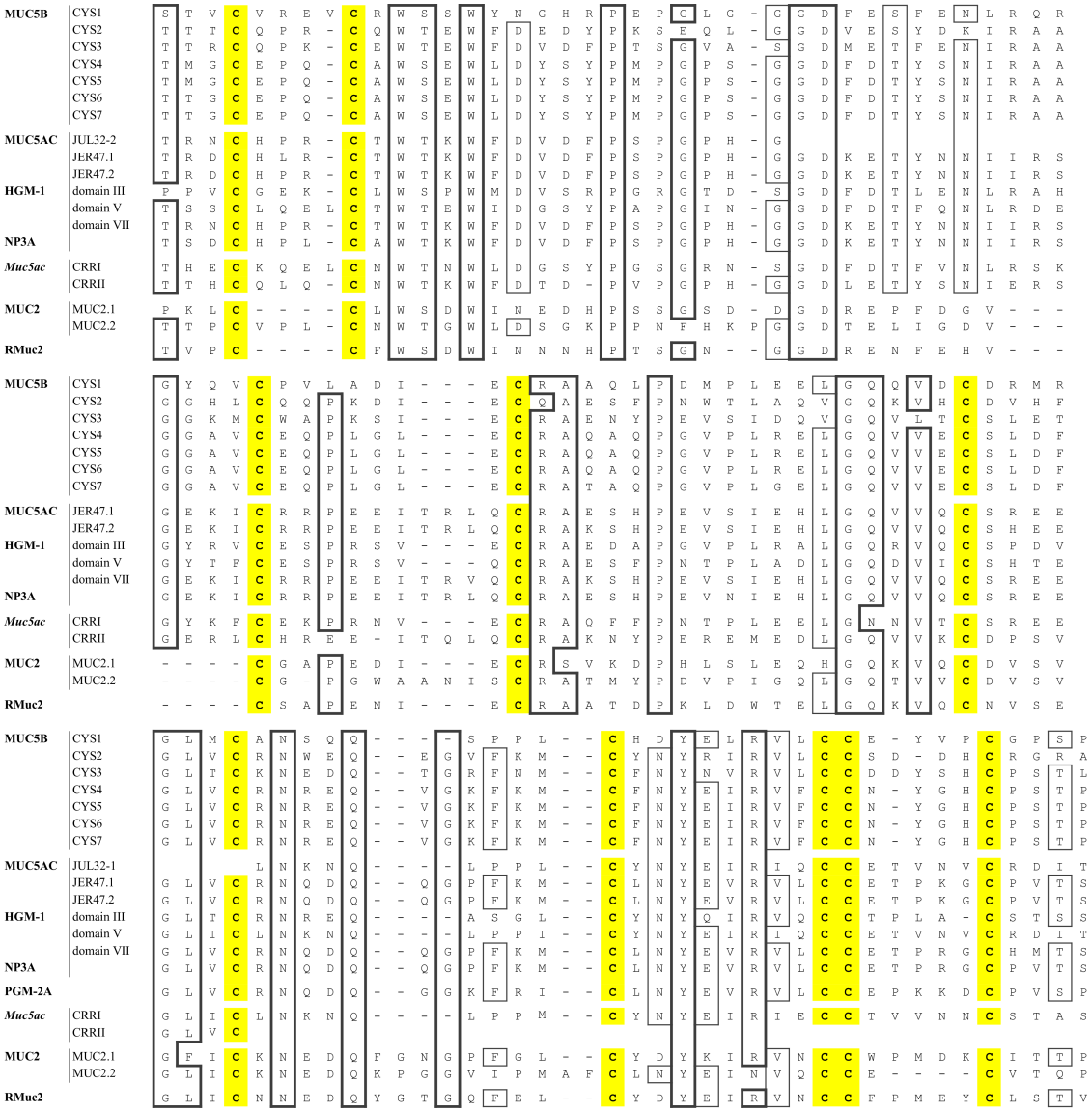


Figure 4 : ***Alignement des séquences peptidiques des sous-domaines CYS des mucines humaines et des mucines animales.***

L'étude de l'organisation de *MUC5B* valide les études biochimiques antérieures sur les mucines

• Identification de MG1 et MG2 comme produits de *MUC5B* et *MUC7*

La cavité orale est la porte d'entrée du système digestif et constitue donc le premier segment du tube digestif. Parce que les mucines salivaires peuvent être facilement récoltées, et en particulier chez des sujets qui sont sains, la bouche est un modèle privilégié d'étude du mucus et des mucines (pour revues voir [18, 66, 67]). Deux catégories de mucines ont été caractérisées biochimiquement depuis plus d'une dizaine d'années: une mucine de haut poids moléculaire, appelée MG1, et une mucine de faible poids moléculaire, appelée MG2 [68-70]. Ces 2 mucines sont structurellement et fonctionnellement distinctes [71]. En effet, MG2 est trouvée sous forme de monomères alors que MG1 est trouvée sous forme de polymères reliés par des ponts disulfures dans les parties amino- et carboxy-terminales. Il a été montré plus récemment que MG2 est codée par le gène *MUC7* [16] et que le constituant majeur de MG1 est codé par le gène *MUC5B* [51, 72]. MG1 est associée de façon non-covalente à des IgA sécrétoires et possède des régions nues et accessibles. Ces régions apparaissent hydrophobes, plus riches en acides aminés chargés négativement et en acides aminés aromatiques. Ces régions correspondent

probablement aux sous-domaines CYS de la partie centrale de MUC5B.

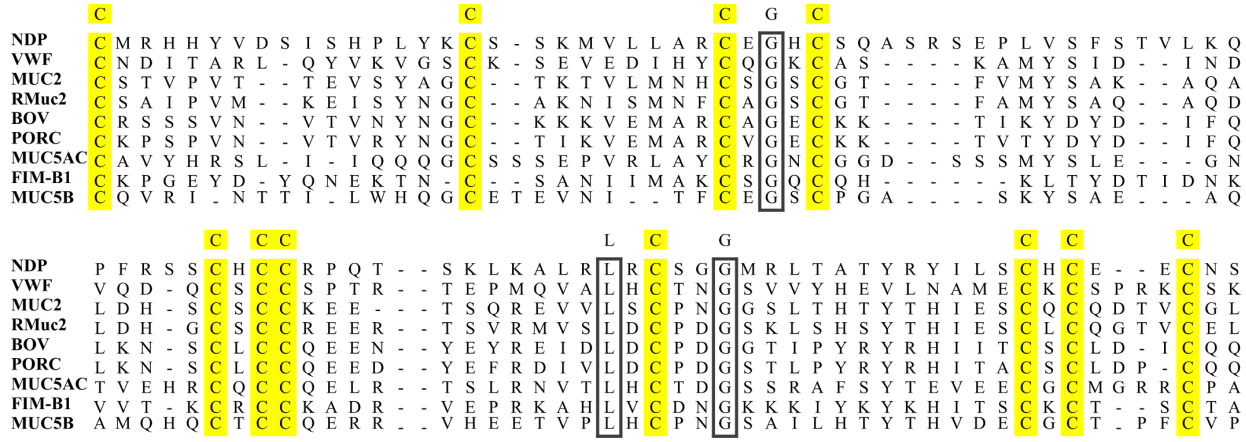


Figure 5 : ***Comparaison des motifs peptidiques CK du NDP, du vWF, des mucines humaines dont les gènes sont localisés en 11p15 et de certaines mucines animales.*** *BOV, mucine sous-maxillaire de bœuf ; PORC, mucine sous-maxillaire de porc ; FIM-B.1, mucine tégumentaire de xénope ; RMuc2, MUC2 de rat.*

• Le peptide de liaison

De nombreux groupes ont décrit un peptide de liaison qui relierait plusieurs grandes sous-unités de mucine. Ce peptide de liaison a une M_r de 70 000 dans l'estomac de porc [73]. Des travaux effectués en collaboration entre 4 laboratoires ont montré en 1989 [74] qu'une sous-unité de M_r 118 000 purifiée par différentes techniques biochimiques à partir de nombreux échantillons de différents organes et de différentes espèces, est présente dans les granules de sécrétion des cellules caliciformes de l'intestin, et dérive d'une réduction de ponts disulfures d'une plus grande sous-unité d'une mucine. Dans la salive un peptide ayant une M_r de 150 000 a été trouvé [74, 75]. Cette sous-unité a été isolée après réduction des ponts disulfures à partir de la mucine MG1. Elle contient des chaînes *N*-glycosidiques tout comme le peptide de 118 000 [76, 77]. Compte tenu que l'antisérum dirigé contre la protéine de liaison de 118 000 isolée à partir d'intestin de rat reconnaît aussi le peptide salivaire de 150 000, il a été suggéré que la différence de masse observée entre ces deux peptides pourrait être due à une glycosylation différente d'un même peptide ou de peptides très similaires [77]. Ainsi, le peptide de liaison, plus hydrophobe que la région *O*-glycosylable de l'apomucine, semble être un peptide libéré par protéolyse d'une mucine et lié par un ou des ponts disulfures à cette même mucine. Dans la salive, ce peptide de liaison correspond sans doute à la partie carboxy-terminale de MUC5B.

• MUC5B et le modèle alterné

La digestion protéolytique des mucines cervicales et bronchiques met en évidence une alternance de zones hautement glycosylées et résistantes à la protéolyse et de zones peu glycosylées, nues, et sensibles aux protéases. Ces données biochimiques ont été confortées par des observations en microscopie électronique : des anticorps dirigés contre les régions nues se fixent régulièrement le long des filaments de mucines [78]. D'après ces études, il a été proposé que les mucines s'oligomérisent par la formation de ponts disulfures interchaînes dans les extrémités amino- et carboxy-terminales des mucines. Les zones nues, plus hydrophobes, pourraient former des liaisons non covalentes avec d'autres composants du mucus, dont des cations, des enzymes, des protéines et d'autres mucines [79].

Les régions nues correspondent probablement aux sous-domaines CYS (108 aa, 10 résidus Cys). En effet, la distance entre ces régions nues observées en microscopie électronique est d'un peu plus de 100 nm, ce qui correspond, si on considère qu'un aa d'une région étirée a une taille d'environ 2,5 Å [11], à environ 400 aa. Ce nombre est en très bonne corrélation avec les tailles des fragments séparant les sous-domaines CYS de MUC5B. Il est possible que les mucines possédant les sous-domaines CYS forment des liaisons entre elles, directement ou indirectement, par ces sous-

domaines. Ainsi, les mucines MUC2, MUC5AC et MUC5B auraient la capacité de former des mailles où la taille de chaque maille est fonction à la fois de la distance entre les sous-domaines CYS et de la quantité de mucines sécrétées. Les propriétés rhéologiques du mucus seraient donc influencées par deux paramètres, l'un régulé (la sécrétion), et l'autre non régulé (la distance entre les sous-domaines CYS). Comme nous l'avons vu plus haut, ce dernier paramètre est invariable d'un individu à l'autre, tout au moins pour MUC2 et MUC5B car, à l'inverse de la région purement répétitive, la région riche en résidus Thr, Ser et Pro située entre les deux sous-domaines CYS de MUC2 et MUC5B ne présente pas de polymorphisme VNTR.

Les mucines comme modulateurs de l'activité facteur de croissance ?

• La famille des mucines ayant 2 domaines EGF-like

Un des 2 motifs EGF-like de SMC (dans ASGP-2) possède les acides aminés consensus des motifs EGF ayant une activité facteur de croissance [33]. Comme nous l'avons vu plus haut, la sous-unité transmembranaire ASGP-2 se lie au récepteur tyrosine-kinase $p185^{neu}$ (dont l'homologue humain est ErbB2) via ce motif EGF modulant ainsi la signalisation cellulaire médiée par la famille des récepteurs de croissance EGF [34]. SMC est donc une molécule régulée lors du développement [80] qui agit comme facteur de croissance ou/et modulateur de l'activité facteur de croissance.

• La famille des mucines ayant 1 domaine TGFβ-like

Le motif CK est, comme pour la protéine NDP et le vWF, un motif à 11 résidus Cys aux positions très conservées. Les protéines ayant ce domaine constituent une sous-famille qui fait partie de la famille des protéines TGF-β contenant [81]:

- la sous-famille CK à 7 résidus Cys dont un résidu Cys est impliqué dans une liaison intermoléculaire et conservé dans cette sous-famille CK. C'est la famille des TGF-β.
- la sous-famille CK à 6 résidus Cys où la dimérisation est non covalente. Le prototype est le NGF (ou Nerve Growth Factor).

Les mucines humaines dont les gènes sont localisés sur le chromosome 11 et les homologues chez l'animal possèdent ce motif structural CK et peuvent donc être classées dans la mégafamille des facteurs de croissance. Ce domaine CK a-t-il pour autant une activité facteur de croissance ou/et modulateur de l'activité facteur de croissance ? Bien que ceci reste encore à démontrer, plusieurs arguments vont dans ce sens. De nombreuses observations relient les mucines et certaines molécules de la matrice extracellulaire (ME) (comme le Perlecan et la Décorine par exemple) à la différenciation cellulaire et à la mise en place des tissus [8, 9, 82] suggérant fortement que les produits des gènes correspondants jouent un rôle crucial dans l'embryogenèse. Ces protéines de la ME et les mucines ont par ailleurs une très grande région centrale répétitive, parfois fortement glycosylée (l'Aggrecan par exemple), bordée par des régions plus riches en résidus Cys. On retrouve parfois des motifs EGF-like dans ces mêmes protéines de la ME. Enfin, il a été montré que le Perlecan intervient directement dans la croissance et l'invasion des cellules tumorales grâce à sa capacité à lier et à stocker les facteurs de croissance. De même, il a été montré que la décorine lie avec une forte affinité le TGF-β1. La décorine joue ainsi un rôle de réservoir de facteur de croissance dans le milieu extracellulaire [83-85]. Il est donc tout à fait probable que les très grandes mucines sécrétées ont aussi une fonction de réservoir des facteurs de croissance. Il est aussi envisageable, à l'image de SMC, que ces mucines agissent directement comme facteur de croissance, d'autant plus qu'il a été montré récemment, en utilisant des anticorps dirigés contre la séquence peptidique déduite de l'ADNc de *MUC5B*, qu'une fraction carboxy-terminale de MUC5B est protéolysée libérant alors le domaine CK de l'apomucine [86].

Perspectives

Les fonctions biologiques des mucines ont surtout été abordées par des études portant sur la copule glycannique. La molécule MUC5B étirée, linéaire, est constituée d'une alternance de régions fortement *O*-glycosylables (les domaines R et R-end) et de domaines nus, c'est-à-dire peu ou pas glycosylables (sous-domaines CYS, régions amino- et carboxy-terminales). Ces régions dites nues représentent en fait pour MUC5B 50% de

l'apomucine et sont organisées en domaines. Aux domaines nus retrouvés dans les mucines doivent correspondre des fonctions biologiques importantes car ces domaines sont très conservés dans l'évolution. Les travaux de clonage des ADNc puis des gènes de mucines n'ont pu voir le jour que très récemment. Ils confortent les observations biochimiques antérieures et vont permettre d'aborder la régulation de l'expression des gènes afin de comprendre les dysrégulations observées dans les nombreuses pathologies qui touchent les muqueuses. Nul doute qu'aux ères successives de la biochimie et de la génétique moléculaire des mucines vont succéder les études de biologie cellulaire et les études *in vivo* grâce au développement des souris transgéniques et knock-out. Ces technologies permettront d'appréhender le rôle physiologique de chacune de ces mucines et de chacun de leurs domaines.

Remerciements : Les auteurs remercient les membres du groupe de *«Génétique des Apomucines Humaines*» de l'Unité INSERM n° 377 et particulièrement Marie-Pierre Buisine pour les résultats d'hybridation *in situ* non encore publiés. J.-L. D. remercie la Fondation Dina-Surdin pour le prix qu'elle vient de lui attribuer.

Bibliographie

1. **Roussel P., Lamblin, G.** (1996) Human mucosal mucins in diseases. Editions J. Montreuil, J.F.G. Vliegenthart et H. Schachter, *Glycoproteins and disease* pp. 351-393
2. **Rose, M.C., Voter, W.A., Sage, H., Brown, C.F., Kaufman, B.** (1984) Effects of deglycosylation on the architecture of ovine submaxillary mucin glycoprotein. *J. Biol. Chem.* **259**, 3167-3172
3. **Gendler, S.J., Spicer A.P.** (1995) Epithelial mucin genes. *Ann. Rev. Physiol.* **57**, 607-634
4. **Audié, J.P., Janin, A., Porchet, N., Copin, M.C., Gosselin, B., Aubert, J.P.** (1993) Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by in situ hybridization. *J. Histochem. Cytochem.* **41**, 1479-1485
5. **Audié, J.P., Tetaert, D., Pigny, P., Buisine, M.P., Janin, A., Aubert, J.P., Porchet, N., Boersma, A.** (1995) Mucin gene expression in the human endocervix. *Hum. Reprod.***10**, 98-102
6. **Balagué, C., Audié, J.P., Porchet, N., Real, F.X.** (1995) In situ hybridization shows distinct patterns of mucin gene expression in normal, benign, and malignant pancreas tissues. *Gastroenterology* **109**, 953-964
7. **Vandenhaute, B., Buisine, M.P., Debailleul, V., Clément, B., Moniaux, N., Dieu, M.C., Degand, P., Porchet, N., Aubert, J.P.** (1997) Mucin gene expression in biliary epithelial cells. *J. Hepatol.* **27**, 1057-1066
8. **Buisine, M.P., Devisme, L., Savidge, T.C., Gespach, C., Gosselin, B., Porchet, N., Aubert, J.P.** (1998) Mucin gene expression in human embryonic and fetal intestine. *Gut* **43**, 519-24
9. **Buisine, M.P., Devisme, L., Copin, M.C., Durand-Réville, M., Gosselin, B., Aubert, J.P., Porchet, N.** (1999) Developmental mucin gene expression in the human respiratory tract. *Am. J. Respir. Cell Mol. Biol.* **20**, 209-218
10. **Ho, J.J.L, Kim, Y.S.** (1995) Do mucins promote tumor cell metastasis ? (Review) *Int. J. Oncol.* **7**, 913-926
11. **Van Klinken, B.J.W., Dekker, J., Büller, H.A., Einerhand, A.W.C.** (1995) Mucin gene structure and expression : protection vs. adhesion. *Am. J. Physiol.* **269**, G613-G627
12. **Noia, J.M.D., Sanchez, D.O., Frasch, A.C.C.** (1995) The protozoan *Trypanosoma cruzi* has a family of genes resembling the mucin genes of mammalian cells. *J. Biol. Chem.* **270**, 24146-24149
13. **Murray, P.J., Spithill, T.W.** (1991) Variants of a *Leishmania* surface antigen derived from a multigenic family. *J. Biol. Chem.* **266**, 24477-24484
14. **Litvinov, S.V., Hilkens, J.** (1993) The epithelial sialomucin, episialin, is sialylated during recycling. *J. Biol. Chem.* **268**, 21364-21371
15. **Parry, G., Beck, J.C., Moss, L., Bartley, J., Ojakian, G.K.** (1990) Determination of apical membrane polarity in mammary epithelial cell cultures : the role of cell-cell, cell-substratum, and membrane-cytoskeleton interactions. *Exp. Cell Res.* **188**, 302-311
16. **Bobek, L.A., Tsai, H., Biesbrock, A.R., Levine, M.J.** (1993) Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (*MUC7*). *J. Biol. Chem.* **268**, 20563-20569
17. **Bobek, L.A., Liu, J., Sait, S.N.J., Shows, T.B., Bobek, Y.A., Levine, M.J.** (1996) Structure and chromosomal localization of the human salivary mucin gene, MUC7. *Genomics* **31**, 277-282
18. **Tabak, L.A.** (1995) In defense of the oral cavity : structure, biosynthesis, and function of salivary mucins. *Annu. Rev. Physiol.* **57**, 547-564
19. **Albone, E.F., Hagen, F.K., Szpirer, C., Tabak, L.A.** (1996) Molecular cloning and characterization of the gene encoding rat submandibular gland apomucin, *Mucsmg*. *Glycoconj. J.* **13**, 709-716

20. Denny, P.C., Mirels, L., Denny, P.A. (1996) Mouse submandibular gland salivary apomucin contains repeated *N*-glycosylation sites. *Glycobiology* **6**, 43-50

21. Gum, J.R., Ho, J.J., Pratt, W.S., Hicks, J.W., Hill, A.S., Vinall, L.E., Roberton, A.M., Swallow, D.M., Kim, Y.S. (1997) MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl terminus, and a novel upstream repetitive region. *J. Biol. Chem.* **272**, 26678-26686

22. Khatri, I.A., Forstner, G.G., Forstner, J.F. (1997) The carboxyl-terminal sequence of rat intestinal mucin RMuc3 contains a putative transmembrane region and two EGF-like motifs. *Biochim. Biophys. Acta* **1326**, 7-11

23. Shekels, L.L., Hunninghake, D.A., Tisdale, A.S., Gipson, I.K., Kieliszewski, M., Kozak, C.A., Ho, S.B. (1998) Cloning and characterization of mouse intestinal MUC3 mucin: 3' sequence contains epidermal-growth-factor-like domains. *Biochem. J.* **330**, 1301-1308

24. Nollet, S., Moniaux, N., Maury, J., Petitprez, D., Degand, P., Laine, A., Porchet, N., Aubert, J.P. (1998) Human mucin gene MUC4: organization of its 5'-region and polymorphism of its central tandem repeat array. *Biochem. J.* **332**, 739-748

25. Moniaux, N., Nollet, S., Porchet, N., Degand, P., Laine, A., Aubert, J.P. (1999) Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. *Biochem. J.* **338**, 325-333

26. Hull, S.R., Sheng, Z., Vanderpuye, O., David, C., Carraway, K.L. (1990) Isolation and partial characterization of ascites sialoglycoprotein-2 of the cell surface sialomucin complex of 13762 rat mammary adenoma cells. *Biochem. J.* **265**, 121-129

27. Sheng, Z., Hull, S.R., Carraway, K.L. (1990) Biosynthesis of the cell surface sialomucin complex of ascites 13762 rat mammary *adenocarcinoma* cells from a high molecular weight precursor. *J. Biol. Chem.* **265**, 8505-8510

28. Bharathan, S., Moriarty, J., Moody, C.E., Sherblom, A.P. (1990) Effect of tunicamycin on sialomucin and natural killer susceptibility of rat mammary tumor ascites cells. *Cancer Res.* **50**, 5250-5256

29. Hull, S.R., Sugarman, E.D., Spielman, J., Carraway, K.L. (1991) Biosynthetic maturation of an ascites tumor cell surface sialomucin. Evidence for *O*-glycosylation of cell surface glycoprotein by the addition of new olisaccharides during recycling. *J. Biol. Chem.* **266**, 13580-13586

30. Moriarty, J., Skelly, C.M., Bharathan, S., Moody, C.E., Sherblom, A.P. (1990) Sialomucin and lytic susceptibility of rat mammary tumor ascites cells. *Cancer Res.* **50**, 6800-6805

31. Wu, K., Fregien, N.L., Carraway, K.L. (1994) Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifunctional cell surface glycoprotein complex of ascites rat mammary adenocarcinoma cells. *J. Biol. Chem.* **269**, 11950-11955

32. Rossi, E.A., McNeer, R.R., Price-Schiavi, S.A., Van den Brande J.M.H., Komatsu, M., Thompson, J.F., Carothers Carraway, C.A., Fregien, N.L., Carraway, K.L. (1996) Sialomucin complex, a heterodimeric glycoprotein complex. Expression as a soluble, secretable form in lactating mammary gland and colon. *J. Biol. Chem.* **271**, 33476-33485

33. Sheng, Z., Wu, K., Carraway, K.L., Fregien, N.L. (1992) Molecular cloning of the transmembrane component of the 13762 mammary adenocarcinoma sialomucin complex. A new member of the epidermal growth factor superfamily. *J. Biol. Chem.* **267**, 16341-16346

34. Carraway, K.L., Rossi, E.A., Komatsu, M., Price-Schiavi, S.A., Huang, D., Guy, P.M., Carvajal, M.E., Fregien, N., Carraway, C.A., Carraway, K.L. (1999) An intramembrane modulator of the ErbB2 receptor tyrosine kinase that potentiates neuregulin signaling. *J. Biol. Chem.* **274**, 5263-5266

35. Pigny, P., Guyonnet Dupérat, V., Hill, A.S., Pratt, W.S., Galiegue-Zouitina, S., Collyn d'Hooge, M., Laine, A., Van Seuningen, I., Degand, P., Gum, J.R., Kim, Y.S., Swallow, D.M., Aubert, J.P., Porchet, N. (1996) Human mucin genes assigned to 11p15: identification and organization of a cluster of genes. *Genomics* **38**, 340-352

36. Gum, J.R., Byrd, J.C., Hicks, J.W., Toribara, N.W., Lamport, D.T.A., Kim, Y.S. (1989) Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for genetic polymorphism. *J. Biol. Chem.* **264**, 6480-6487

37. Gum, J.R., Hicks, J.W., Swallow, D.M., Lagace, R.E., Byrd, J.C., Lamport, D.T.A., Siddiki, B., Kim, Y.S. (1990) Molecular cloning of cDNAs derived from a novel human intestinal mucin gene. *Biochem. Biophys. Res. Commun.* **171**, 407-415

38. Toribara, N.W., Gum, J.R., Culhane, P.J., Lagace, R.E., Hicks, J.W., Petersen, G.M., Kim, Y.S. (1991) MUC-2 human small intestinal mucin gene structure (Repeated arrays and polymorphism). *J. Clin. Invest.* **88**, 1005-1013

39. Gum, J.R., Hicks, J.W., Toribara, N.W., Rothe, E.M., Lagace, R.E., Kim, Y.S. (1992) The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. *J. Biol. Chem.* **267**, 21375-21383

40. Gum, J.R., Hicks, J.W., Toribara, N.W., Siddiki, B., Kim, Y.S. (1994) Molecular cloning of human intestinal mucin (*MUC2*) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J. Biol. Chem.* **269**, 2440-2446

41. Ohmori, H., Dohrman, A.F., Gallup, M., Tsuda, T., Kai, H., Gum, J.R., Kim, Y.S., Basbaum, C.B. (1994) Molecular cloning of the amino-terminal region of a rat MUC 2 mucin gene homologue. Evidence for expression in both intestine and airway. *J. Biol. Chem.* **269**, 17833-17840

42. Xu, G., Huan, L., Khatri, I., Sajjan, U.S., McCool, D., Wang, D., Jones, C., Forstner, G., Forstner, J. (1992) Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminal region, and is encoded by a gene on chromosome 11 p 15.5. *Biochem. Biophys. Res. Commun.* **183**, 821-828

43. Xu, G., Huan, L.J., Khatri, I.A., Wang, D., Bennick, A., Fahim, R.E., Forstner, G.G., Forstner, J.F. (1992) cDNA for the carboxyl-terminal region of a rat intestinal mucin-like peptide. *J. Biol. Chem.* **267**, 5401-5407

44. Hansson, G.C., Baeckstrom, D., Carlstedt, I., Klinga-Levan, K. (1994) Molecular cloning of a cDNA coding for a region of an apoprotein from the 'insoluble' mucin complex of rat small intestine. *Biochem. Biophys. Res. Commun.* **198**, 181-190

45. Gum, J.R., Hicks, J.W., Lagace, R.E., Byrd, J.C, Toribara, N.W., Siddiki, B., Fearney, F.J., Lamport, D.T.A., Kim, Y.S. (1991) Molecular cloning of rat intestinal mucin. Lack of conservation between mammalian species. *J. Biol. Chem.* **266**, 22733-22738

46. Van Klinken, B.J., Einerhand, A.W., Duits, L.A., Makkink, M.K., Tytgat, K.M., Renes, I.B., Verburg, M., Buller, H.A., Dekker, J. (1999) Gastrointestinal expression and partial cDNA cloning of murine Muc2. *Am. J. Physiol.* **276**, G115-124

47. Toribara, N.W., Roberton, A.M., Ho, S.B., Kuo, W.L., Gum, E., Hicks, J.W., Gum, J.R., Byrd, J.C., Siddiki, B., Kim, Y.S. (1993) Human gastric mucin. Identification of a unique sequence by expression cloning. *J. Biol. Chem.* **268**, 5879-5885

48. Toribara, N.W., Ho, S.B., Gum, E., Gum, J.R , Lau, P., Kim, Y.S. (1997) The carboxyl-terminal sequence of the human secretory mucin, MUC6. Analysis of the primary amino acid sequence. *J. Biol. Chem.* **272**, 16398-16403

49. Debailleul, V., Laine, A., Huet, G., Mathon, P., d'Hooghe, M.C., Aubert, J.P., Porchet, N. (1998) Human mucin genes MUC2, MUC3, MUC4, MUC5AC, MUC5B, and MUC6 express stable and extremely large mRNAs and exhibit a variable length polymorphism. An improved method to analyze large mRNAs. *J. Biol. Chem.* **273**, 881-890

50. Desseyn, J.L., Guyonnet Dupérat, V., Porchet, N., Aubert, J.P., Laine, A. (1997) Human mucin gene MUC5B: the 10.7 kb large central exon encodes various subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family. *J. Biol. Chem.* **272**, 3168-3178

51. Desseyn J.L., Aubert J.P., Van Seuningen, I., Porchet N., Laine A. (1997) Genomic Organization of the 3' Region of the Human Mucin Gene *MUC5B*. *J. Biol. Chem*. **272**, 16873-16883

52. Pigny, P., Van Seuningen, I., Desseyn, J.L., Nollet, S., Porchet, N., Laine, A., Aubert, J.P. (1996) Identification of a 42 kDa nuclear factor (NF1-MUC5B) from HT-29 MTX cells that binds to the 3' region of human mucin gene *MUC5B*. *Biochem. Biophys. Res. Commun.* **220**, 186-191

53. Desseyn, J.L., Rousseau, K., Laine, A. (1999) Fifty-nine-bp repeat polymorphism in the uncommon intron 36 of the human mucin gene *MUC5B*. *Electrophoresis* **20**, 493-496.

54. Buisine, M.P., Desseyn, J.L., Porchet, N., Degand, P., Laine, A., Aubert, J.P. (1998) Genomic organization of the 3'-region of the human MUC5AC mucin gene: additional evidence for a common ancestral gene for the 11p15.5 mucin gene family. *Biochem. J.* **332**, 729-738

55. Desseyn J.L., Buisine, M.P., Porchet N., Aubert, J.P., Degand, P., Laine A. (1998) Evolutionary History of the 11p15 Human Mucin Genes. *J. Mol. Evol.* **46,** 102-106

56. Desseyn JL, Buisine MP, Porchet N, Aubert JP, Laine A (1998) Genomic organization of the human mucin gene MUC5B. cDNA and genomic sequences upstream of the large central exon. *J. Biol. Chem.* **273**,:30157-30164

57. Offner, G.D., Nunes, D.P., Keates, A.C., Afdhal, N.H., Troxler, R.F. (1998) The amino-terminal sequence of MUC5B contains conserved multifunctional D domains: implications for tissue-specific mucin functions. *Biochem. Biophys. Res. Commun.* **251**, 350-355

58. Vinall, L.E., Hill, A.S., Pigny, P., Pratt, W.S., Toribara, N., Gum, J.R., Kim, Y.S., Porchet, N., Aubert, J.P., Swallow, D.M. (1998) Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15. *Hum. Genet.* **102**, 357-366

59. Biesbrock, A.R., Bobek, L.A., Levine, M.J. (1997) *MUC7* gene expression and genetic polymorphism. *Glycoconj. J.* **14**, 415-422

60. Sadler, J.E. (1998) Biochemistry and genetics of von Willebrand factor. *Annu. Rev. Biochem.* **67**, 395-424

61. Perez-Vilar, J., Eckhardt, A.E., DeLuca, A., Hill, R.L. (1998) Porcine submaxillary mucin forms disulfide-linked multimers through its amino-terminal D-domains. *J. Biol. Chem.* **273**, 14442-14449

62. Van Klinken, B.J., Einerhand, A.W., Buller, H.A., Dekker, J. (1998) The oligomerization of a family of four genetically clustered human gastrointestinal mucins. *Glycobiology* **8**, 67-75

63. Desseyn J.L. (1997) Organisation du gène de mucine humaine *MUC5B*. Bases moléculaire d'une nouvelle classification des mucines. Thèse de doctorat d'université, Lille, France

64. Meindl, A., Berger, W., Meitinger, T., van de Pol, D., Achatz, H., Dörner, C., Haasemann, M. Hellebrand, H., Gal, A., Cremers, F., Ropers, H.H. (1992) Norrie disease is caused by mutations in an extracellular protein resembling C-terminal globular domain of mucins. *Nat. Genet.* **2**, 139-143

65. Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M., Murken, J. (1993) Molecular modeling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. *Nat. Genet.* **5**, 376-380

66. Nieuw Amerongen, A.V.N., Bolscher, J.G.M., Veerman, E.C.I. (1995) Salivary mucins : protective functions in relation to their diversity. *Glycobiology* **5**, 733-740

67. Levine, M.J., Reddy, M.S., Tabak, L.A., Loomis, R.E., Bergey, E.J., Jones, P.C., Cohen, R.E., Stinson, M.W., Al-Hashimi, I. (1987) Structural aspects of salivary glycoproteins. *J. Dent. Res.* **66**, 436-441

68. Tabak, L.A., Levine, M.J., Mandel, I.D., Ellison, S.A. (1982) Role of salivary mucins in the protection of the oral cavity. *J. Oral Pathol.* **11**, 1-17

69. Prakobphol, A., Levine, M.J., Tabak, L.A., Reddy, M.S. (1982) Purification of low-molecular-weight, mucin-type glycoprotein from human submandibular-sublingual saliva. *Carbohydr. Res.* **108**, 111-122

70. Loomis, R.E., Prakobphol, A., Levine, M.J., Reddy, M.S., Jones, P.C. (1987) Biochemical and biophysical comparison of two mucins from human submandibular-sublingual saliva. *Arch. Biochem. Biophys.* **258**, 452-464

71. Al-Hashimi, I., Levine, M.J. (1989) Characterization of in vivo salivary-derived enamel pellicle. *Arch. Oral Biol.* **34**, 289-295

72. Nielsen, P.A., Bennett, E.P., Wandall, H.H., Therkildsen, M.H., Hannibal, J., Clausen, H. (1997) Identification of a major human high molecular weight salivary mucin (MG1) as tracheobronchial mucin MUC5B. *Glycobiology* **7**, 413-419

73. Pearson, J.P., Allen, A., Parry, S. (1981) A 70000-molecular-weight protein isolated from purified pig gastric mucus glycoprotein by reduction of disulphide bridges and its implication in the polymeric structure. *Biochem. J.* **197**, 155-162

74. Roberton, A.M., Mantle, M., Fahim, R.E.F., Specian, R.D., Bennick, A., Kawagishi, S., Sherman, P., Forstner, J.F. (1989) The putative 'link' glycopeptide associated with mucus glycoproteins. Composition and properties of preparations from the gastrointestinal tracts of several mammals. *Biochem. J.* **261**, 637-647

75. Slomiany, A., Tamura, S., Grzelinska, E., Piotrowski, J., Slomiany, B.L. (1992) Mucin complexes: characterization of the "link" component of submandibular mucus glycoprotein. *Int. J. Biochem.* **24**, 1003-1015

76. Mantle, M., Forstner, G.G., Forstner, J.F. (1984) Biochemical characterization of the component parts of intestinal mucin from patients with cystic fibrosis. *Biochem. J.* **224**, 345-354

77. Kawagishi, S., Fahim, R.E.F., Wong, K.H., Bennick, A. (1990) Purification and characterization of subnits of high molecular weight human salivary mucin. *Archs. Oral Biol.* **35**, 265-272

78. Sheehan, J.K., Boot-Handford, R.P., Chantler, E., Carlstedt, I., Thornton, D.J. (1991) Evidence for shared epitopes within the 'naked' protein domains of human mucus glycoproteins. A study performed by using polyclonal antibodies and electron microscopy. *Biochem. J.* **274**, 293-296

79 Shankar, V., Naziruddin, B., Reyes de la Rocha, S., Sachdev, G.P. (1990) Evidence of hydrophobic domains in human respiratory mucins. Effect of sodium chloride on hydrophobic binding properties. *Biochemistry* **29**, 5856-5864

80. Price-Schiavi, S.A., Carraway, C.A., Fregien, N., Carraway, K.L. (1998) Post-transcriptional regulation of a milk membrane protein, the sialomucin complex (Ascites sialoglycoprotein (ASGP)-1/ASGP-2, rat muc4), by transforming growth factor beta. *J. Biol. Chem.* **273**, 35228-35237

81. Sun, P.D., Davies, D.R. (1995) The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269-91

82. Iozzo, R.V. Matrix proteoglycans: from molecular design to cellular function.(1998) *Annu. Rev. Biochem.*, **67**, 609-652

83. Hildebrand, A., Romaris, M., Rasmussen, L.M., Heinegard, D., Twardzik, D.R., Border, W.A., Ruoslahti, E. (1994) Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem. J.* **302**, 527-534

84. Ruoslahti, E., Yamaguchi, Y. (1991) Proteoglycans as modulators of growth factor activities. *Cell* **64**, 867-869

85. Iozzo, R.V., Moscatello, D.K., McQuillan, D.J., Eichstetter, I. (1999) Decorin is a biological ligand for the epidermal growth factor receptor. *J. Biol. Chem.* **274**, 4489-4892

86. Wickstrom, C., Davies, J.R., Eriksen, G.V., Veerman, E.C., Carlstedt, I. (1998) 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**, 685-693