Université des Sciences et des Technologies de Lille

N° d'ordre

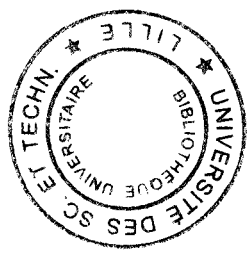
THESE

présentée pour l'obtention du grade de

Docteur en Sciences de la Vie et de la Santé

par

Nicolas MONIAUX



Caractérisation de l'extrémité 3' du gène *MUC4* (gène d'apomucine humaine) : isoformes solubles et membranaires, approche des relations structure-fonction

Soutenue le 04 juin 1999 devant la commission d'examen :

Président :	**Pr Steven BALL**
Rapporteurs :	**Dr Dallas SWALLOW**
	Dr Christian GESPACH
Examinateurs :	**Dr Jean Pierre AUBERT**
	Dr Nicole PORCHET

Ce travail a été réalisé au sein de l'Unité U377 INSERM intitulée "Biologie et Physiopathologie des Cellules Mucipares", dans l'équipe dirigée par Jean Pierre Aubert et Nicole Porchet, sous la direction scientifique de Jean Pierre Aubert.

Je tiens à remercier :

Monsieur le Professeur Pierre-Marie Degand, Directeur de l'Unité pour m'avoir accueilli au sein du laboratoire pour le DEA puis pour préparer cette thèse de sciences.

Monsieur le Docteur Jean Pierre Aubert, qui m'a confié un sujet de DEA et a accepté de m'encadrer tout au long de ce travail de thèse. Tu as toujours été disponible pour m'écouter et guider mon travail jusqu'à son accomplissement.

Madame le Docteur Nicole Porchet, qui m'a fait partager son enthousiasme et ses compétences pour la recherche.

Madame le Docteur Anne Laine, qui a su m'apporter conseils et encouragements depuis mon arrivée au laboratoire.

Madame le Docteur Isabelle Van Seuningen, qui a toujours été disponible pour répondre à mes questions.

Je tiens à remercier également Daniel Petitprez, Pascal Mathon, Annette Leclerc, Christine Mouton et Michel Crépin pour l'aide technique très précieuse qu'ils m'ont apportée.

Je tiens à remercier également tous les membres de l'équipe pour leur sympathie et leur bonne humeur.

Je remercie très sincérement pour l'honneur qu'ils me font :

Monsieur le Professeur Steven Ball,

en acceptant de présider ce jury ;

Madame le Docteur Dallas Swallow,
Monsieur le Docteur Christian Gespach,

en acceptant d'être les rapporteurs de cette thèse ;

Monsieur le Docteur Jean Pierre Aubert
Madame le Docteur Nicole Porchet

en acceptant de juger ce travail.

Résumé

Le gène de mucine humaine *MUC4,* localisé sur le chromosome 3 dans la région q29, est exprimé par les épithéliums des tractus respiratoire, digestif, uro-génital ainsi que par la surface oculaire. Il constitue un marqueur de carcinomes.

Nous avons étudié (par RACE-PCR, RT-PCR et criblage de banque d'ADNc) la structure de ce gène et montré que l'apomucine MUC4 est l'homologue humain de la sialomucine de rat SMC. Comme SMC, MUC4 a une structure modulaire composée de deux domaines, l'un de type mucine, MUC4α et l'autre contenant deux motifs de type EGF, un domaine hydrophobe potentiellement transmembranaire et une queue intracytoplasmique, MUC4β. MUC4 comme MUC1 et SMC appartiendrait donc au groupe des "mucin-like" dont les fonctions multiples dans la morphogénèse et le renouvellement des épithéliums ou dans la promotion des métastases seraient en rapport avec des propriétés d'adhérence.

Le gène *MUC4* code une famille d'isoformes comportant au moins 8 variants distincts, 5 formes solubles et 3 formes membranaires. Ces variants sont obtenus par un épissage alternatif complexe de l'extrémité 3'-terminale de l'ARNm de *MUC4*. Les formes solubles ne possèdent pas les domaines de type EGF. Un variant dénommé MUC4/Y semble, lui, être dépourvu du domaine de type mucine.

Comme SMC qui chez le rat est le ligand de $p185^{neu}$, MUC4 pourrait selon des travaux préliminaires se lier à ErbB2 (approche réalisé à l'aide de protéines de fusions) et induire la prolifération des cellules (après transfection transitoire) de la lignée de cancer mammaire MCF7.

Le gène *MUC4* présente un polymorphisme complexe (étude par Southern blot) de type VNTR en rapport avec la présence d'au moins 4 séquences répétitives distinctes exonique et introniques. Leur rôle sur l'expression qualitative ou quantitative du gène MUC4 est étudié au laboratoire en rapport avec la susceptibilité au développement des pathologies des muqueuses.

Publications et communications

I. Publications

Vandenhaute B., Buisine M. P., Debailleul V., Clement B., **Moniaux N.,** Dieu M. C., Degand P., Porchet N., Aubert J.P. Mucin gene expression in biliary epithelial cells. J. Hepatol. 1997 Dec;27(6):1057-66.

Nollet S., **Moniaux N.,** Maury J., Petitprez D., Degand P., Laine A., Porchet N., Aubert J.P. Human mucin gene MUC4: organization of its 5'-region and polymorphism of its central tandem repeat array. Biochem. J. 1998 Jun 15;332 (Pt 3):739-48.

Moniaux N., Nollet S. Porchet N., Degand P., Laine A., Aubert J. P. Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. Biochem J. 1999 Mar 1;338(Pt 2):325-333.

Porchet N., Buisine M. P., Desseyn J. L., **Moniaux N**., Nollet S., Degand P., Pigny P., Van Seuningen I., Laine A. and Aubert J.P. Gènes MUC : une superfamille de gènes ? Vers une clasification fonctionnelle des apomucines humaines. Société de Biologie de Lille 1999; 193(1): sous-presse.

II. Posters

Moniaux N., Nollet S., Laine A., Aubert J. P. and Porchet N. The human mucin gene *MUC*4 contains at least two perfectly conserved sequences repeated in tandem. 4th International Workshop on Carcinoma-Associated Mucins. Cambridge, 1996

Nollet S., Debailleul V., **Moniaux N**., Porchet N., Laine A. and Aubert J. P. Human mucin gene *MUC4*: organization and polymorphism of its central tandem repeat array at the DNA and RNA level. The 9th meeting of the Mucin Club, Bristol, Britain 1997

Moniaux N., Nollet S., Desseyn J. L., Buisine M. P., Laine A., Porchet N., and Aubert J. P. Genomic organization of the human mucin genes. *MUC4*: an organization which differs from the 11p15 genes. 5th International Workshop on Carcinoma-Associated Mucins. Cambridge, 1998

Nollet S., **Moniaux N**., Maury, J., Petitprez, D., Degand P., Laine A., Porchet N., and Aubert J. P. Human mucin gene *MUC4*: organization of its 5'-region and polymorphism of its central tandem repeat array. 5th International Workshop on Carcinoma-Associated Mucins. Cambridge, 1998

Moniaux N. MUC4 : une mucine humaine, glycoprotéine membranaire complexe et ligand du protooncogène ErbB2. Concours Alexandre Joël, Association pour la Recherche sur le Cancer, Paris, 1998

Moniaux N., Nollet S., Porchet N., Degand P., Laine A. and Aubert J. P. Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. The 10th meeting of the Mucin Club, Chantilly, France 1999

Degroote, S., Delmotte, P., **Moniaux N.**, Van Seuningen I., Davril, M., Roussel, P., Lamblin, G. and Perini, J. M. Modification de glycosylation des mucines sécrétées par une lignée cellulaire d'origine trachéobronchique (MM39) sous l'effet du TNFα. Concours des jeunes chercheurs sur la mucoviscidose, Paris, 1999.

Abréviations

aa : acide aminé

ADN : Acide DésoxyriboNucléique

ADNc : Acide DésoxyriboNucléique complémentaire

AMPc : Adénosine MonoPhosphate cyclique

APC : Adenomatous Polyposis Coli

ARN : Acide RiboNucléique

ARNm : Acide RiboNucléique messager

ASGP : Ascite SialoGlycoProtein

CK : Cystine Knot

CTL : Cytotoxic T Cell

Da : Dalton (unité de masse)

DO : Densité Optique

EGF : Epidermal Growth Factor

GSK3β : Glycogen Synthase Kinase 3β

M : unité de concentration Molaire

NDF : Neu Differentiation Factor

NDP : Norrie Disease Protein

NK : Natural Killer

nt : nucléotide

PCR : Polymerase Chain Reaction

RACE-PCR : Rapid Amplification of cDNA Ends-PCR

RER : Réticulum Endoplasmique Rugueux

RME : Responsive Mucin Element

RT-PCR : Reverse Transcriptase-PCR

SMC : SialoMucin Complex

TGF : Transforming Growth Factor

TR : Tandem Repeat

PMR : Perfect homopurine Mirror Repeats

VNTR : Variable Number of Tandem Repeat

vWF : von Willebrand Factor

Introduction

I. Le mucus

Les cellules des organismes vivants en contact avec le milieu extérieur sont agressées en permanence par des substances étrangères, telles que des micro-organismes, des toxines et des polluants. Au cours de l'Evolution, des mécanismes de protection se sont développés. Parmi ceux-ci on trouve la sécrétion de mucus. Le mucus recouvre les épithéliums des vertébrés ainsi que de certains invertébrés, ses rôles biologiques se sont diversifiés.

I. 1. Rôle de barrière.

Le mucus est une sécrétion dotée de propriétés rhéologiques (viscoélasticité et filance) qui protège des agressions extérieures la surface de l'épithélium des tractus respiratoire, gastro-intestinal et uro-génital. Il se compose de deux couches physiquement distinctes : un gel insoluble dans l'eau de 0,5 à 450 μm d'épaisseur, adhérant à la muqueuse et une couche visqueuse, hydrosoluble, qui recouvre ce gel. Le mucus constitue l'interface entre le milieu extérieur et la surface de l'épithélium.

L'épithélium respiratoire, qui tapisse l'ensemble des voies aériennes supérieures et inférieures depuis les fosses nasales jusqu'aux bronchioles terminales, est recouvert par un tapis muqueux (Figure 1). Le mucus respiratoire est principalement sécrété par les cellules glandulaires. Il forme un tapis continu (Basbaum et al., 1988) qui se situe à l'extrémité des cils vibratiles des cellules ciliées et constitue ainsi une barrière de protection efficace entre l'environnement et la muqueuse. Ce mince film de mucus est mobilisé en permanence par le battement ciliaire, et sa vitesse de transport croît depuis les bronchioles (vitesse de transport de l'ordre de 2 mm/min) jusqu'à la trachée et les voies aériennes supérieures (où la vitesse de transport atteint 10 à 15 mm/min). Ce mucus représente la première ligne de défense de la muqueuse respiratoire. La couche de mucus qui recouvre les cellules

épithéliales ciliées piège les particules exogènes. Les vibrations ciliaires font remonter ce tapis muqueux jusqu'au pharynx où il est dégluti.

Figure 1 : vue en microscopie électronique à balayage de l'épithélium respiratoire. Le mucus forme un tapis muqueux présent à l'extrémité des cils vibratiles (Gaillard et al., 1992).

Au niveau digestif, l'épaisseur et la composition de la couche de mucus qui tapisse la surface microvillositaire contribuent également à la défense de la muqueuse contre l'adhérence et la pénétration des toxines et des bactéries. Ainsi, la production de mucus par les cellules caliciformes est augmentée sous l'action de l'histamine produite par les mastocytes au cours d'une infection intestinale. Il a été expérimentalement démontré que *vibrio cholerae*, par l'intermédiaire de sa toxine cholérique, était susceptible d'épuiser la sécrétion des cellules caliciformes par effet sécrétagogue (Chadee et al., 1991). Le mucus est l'un des partenaires de l'écosystème digestif.

De nombreuses bactéries commensales du tube digestif peuvent se lier à des récepteurs glucidiques portés par les mucines (principales glycoprotéines du mucus) et ainsi inhiber la liaison des bactéries pathogènes (Chadee et al., 1991). Les diverses enzymes produites par la flore saprophyte peuvent endommager sélectivement les bactéries pathogènes et réduire leur viabilité par l'attaque de leur paroi cellulaire (Hoskins et al., 1985). Si la couche de mucus est affectée ou altérée par des changements de régime alimentaire ou des médicaments, la protection offerte par le mucus est réduite et conduit l'hôte à une plus grande susceptibilité aux surinfections bactériennes (Carlstedt-Duke et al., 1986). Toutes les études réalisées laissent supposer que le mucus joue un rôle protecteur contre les bactéries pathogènes en formant non seulement une barrière physique mais aussi en interférant avec l'adhérence de celles-ci sur leurs cibles.

Au niveau oesophagien, le mucus, notamment d'origine salivaire, est fortement impliqué dans la protection contre les reflux gastriques acides. Cette fonction de protection de la muqueuse se retrouve également au niveau gastrique et dans la vésicule biliaire. Dans l'estomac et la partie proximale du duodénum, le gel muqueux empêche les ions bicarbonates sécrétés par la muqueuse de se mélanger trop rapidement aux grandes quantités d'acide présentes dans la lumière. A la surface de l'épithélium, la neutralisation des sucs gastriques se fait par la diffusion des ions bicarbonates à travers le gel de mucus. Un gradient de pH se crée ainsi à travers cette couche de mucus, allant d'un pH 2 dans la lumière gastrique à un pH pratiquement neutre à la surface de l'épithélium (Allen et al., 1993).

La couche de mucus gastrique agit également comme une barrière physique, empêchant la pepsine de la lumière d'atteindre et de digérer la surface de l'épithélium (Pearson et al., 1986). La vésicule biliaire s'autoprotège contre les sels biliaires qui constituent de véritables détergents, naturellement présents à des concentrations élevées (jusqu'à 300 mM) dans la bile grâce au gel de mucus.

Au niveau génital, le mucus joue un rôle primordial dans le maintien de l'asepsie de l'utérus mais surtout dans la reproduction. Lors de l'ovulation, il facilite l'ascension des spermatozoïdes vers l'utérus et contribue au niveau tubaire à la survie et au transport des gamètes et de l'embryon. Les variations hormonales observées au cours du cycle menstruel semblent déterminer les changements de propriétés du mucus cervical.

I. 2. Composition chimique.

Le mucus est une sécrétion très hétérogène. Il est constitué de 95 % d'eau, 1 % d'électrolytes et de 4 % de macromolécules. Les principaux électrolytes rencontrés sont : Na^+, K^+, Mg^{2+}, Ca^{2+}. La composante organique est constituée de lipides (phospholipides, acides gras libres, cholestérol...) (Houdret et al., 1986) (Slomiany et al., 1983), d'acides nucléiques, de protéines et de glycoprotéines. Dans le cas de la muqueuse trachéo-bronchique, les protéines et les glycoprotéines présentes dans ces sécrétions peuvent être le résultat soit d'une synthèse et d'une sécrétion locales, soit d'une transsudation de protéines plasmatiques en particulier au cours de phénomènes inflammatoires (Kaliner et al., 1986) (Bhaskar et al., 1988). Les protéines et les glycoprotéines synthétisées localement par les cellules épithéliales sont de natures variées (Laine and Hayem, 1976). Outre les mucines, on y détecte :

- des inhibiteurs de protéases (Stockley and Afford, 1983), tels que l'α1-antiprotéase, l'antithrombine, l'α1-antichymotrypsine et l'antileucoprotéase, capables d'inhiber l'action des protéases libérées par les cellules inflammatoires et les bactéries au cours des phénomènes infectieux.

- le lysozyme, une protéine très basique de masse moléculaire 15 kDa douée de propriétés bactéricides (Konstan et al., 1982). Il s'agit d'une muramidase dont l'action bactéricide s'explique par son aptitude à dégrader les peptidoglycannes des parois bactériennes.

- la transferrine bronchique (analogue de la lactotransferrine) est une glycoprotéine possédant deux sites de fixation pour le fer. Cette capacité prive les bactéries du fer nécessaire à leur croissance (Spik and Montreuil, 1983).

- des immunoglobulines, dont les IgA. Elles sont synthétisées par les plasmocytes de la sous-muqueuse sous forme de dimères liés par une chaîne supplémentaire, la chaîne J. Pour traverser l'épithélium et être sécrétées, elles sont prises en charge par un récepteur situé à la partie baso-latérale de la membrane plasmique des cellules épithéliales. Elles sont ainsi internalisées puis sécrétées au pôle apical avec une partie du récepteur.

II. Les mucines épithéliales.

Les mucines sont les macromolécules majoritaires du mucus. Elles lui confèrent ses propriétés physico-chimiques. Ce sont des glycoprotéines de haute masse moléculaire, synthétisées et sécrétées par des cellules spécialisées de l'épithélium et, dans certains organes, par des cellules à mucus des glandes sous-muqueuses. Les mucines sont généralement définies comme étant des macromolécules dont la masse atteint plusieurs millions de Daltons. Leur composition en résidus d'acides aminés est riche en résidus hydroxylés (sérine et thréonine), sur lesquels se fixent de très nombreuses chaînes oligo-saccharidiques. La teneur en chaînes O-glycosidiques représente de 50 à 80 % de leur poids sec. De plus, certaines mucines peuvent contenir un petit nombre de chaînes N-glycanniques (Amerongen et al., 1983).

II. 1. Les cellules sécrétrices de mucines.

Les mucines humaines sont synthétisées et sécrétées par des cellules spécialisées de l'épithélium de surface (Smits and Kramer, 1981) et des glandes muqueuses. Ces cellules sont fortement polarisées et montrent une spécialisation pour la production massive des glycoprotéines du mucus. Le réticulum endoplasmique rugueux (RER) est généralement concentré au niveau du cytoplasme basal de ces cellules. L'appareil de Golgi y est très développé, ce qui est en rapport avec la forte activité de glycosylation requise lors de la synthèse de mucines. La partie apicale du cytoplasme comporte de nombreux granules de stockage des mucines, serrés les uns contre les autres. L'exocytose des mucines se produit par la fusion des membranes délimitant les granules avec la membrane apicale des cellules.

Ces cellules sont appelées cellules caliciformes au niveau de l'épithélium bronchique et colique ou cellules à pôle fermé au niveau de l'estomac. D'autres types cellulaires qui ne sont pas spécialisés dans la sécrétion de mucus expriment également des mucines. Ainsi *MUC4* est exprimé par les cellules ciliées de l'épithélium respiratoire et les entérocytes. *MUC3* est exprimé par les entérocytes et par les hépatocytes .

II. 2. Biosynthèse des mucines.

Comme pour toutes les protéines sécrétées, le squelette peptidique des mucines est synthétisé par les polysomes le long de la face cytoplasmique du RER. Les mucines ainsi traduites traversent la membrane pour se retrouver dans la lumière du RER. Le processus de N-glycosylation débute lors de cette internalisation. La N-glycosylation nécessite l'intervention d'un donneur de nature lipidique qui greffe sur une asparagine l'oligosaccharide suivant : $Glc_3Man_9GlcNAc_2$ (Kornfeld and Kornfeld, 1985). Pour être N-glycosylable, une asparagine doit appartenir à la séquence Asn-X-Ser ou Asn-X-Thr où X peut être n'importe lequel des acides aminés exception faite de la proline et de l'acide aspartique (Marshall, 1972). De plus, Aubert et al ont montré que la présence de la structure secondaire de type "β-turn" est souvent associée à la N-glycosylation d'un résidu d'asparagine (Aubert et al., 1976). La N-glycosylation est réalisée de manière précoce dans la biogénèse des glycoprotéines ce qui suggère son intervention pour le transport et la sécrétion de ces molécules.

Après la synthèse complète du squelette protéique, les mucines sont transportées par des vésicules dans l'appareil de Golgi. Pendant leur transit le long des citernes golgiennes, les mucines acquièrent leur glycosylation finale.

Pour la O-glycosylation, les mécanismes menant à la biosynthèse des chaînes oligosaccharidiques sont moins bien définis. Il n'y a aucune séquence précise connue qui favorise l'action des glycosyltransférases, sinon la présence de résidus de proline au voisinage de résidus de sérine ou thréonine qui, en rigidifiant la chaîne peptidique, favoriserait de manière stérique l'action des GalNac-transférases (Briand et al., 1981). La O-glycosylation est initialisée lors du passage des protéines néo-synthétisées, du RER aux citernes du cis Golgi. La structure et la composition des chaînes O-glycanniques dépendent de plusieurs éléments :

-de l'expression de manière spécifique de tissu d'un répertoire complexe de glycosyltransférases.

-de l'addition d'un monosaccharide qui est déterminée par la nature du dernier résidu ajouté.

-du degré de branchement des chaînes glycanniques.

-des différentes glycosyltransférases en compétition pour le même substrat (Hounsell and Feizi, 1982) (Feizi et al., 1984).

-de la quantité de monosaccharides présents.

L'étape finale de O-glycosylation est caractérisée par l'addition d'acide sialique et de résidus sulfate au niveau des citernes du trans Golgi. Ceci confère aux mucines une forte charge négative.

Les mucines sont alors concentrées. Il se forme au niveau des citernes du trans Golgi des vacuoles de concentration riche en calcium. Les vacuoles ou granules de sécrétion se retrouvent au pôle apical des cellules (Figure 2). Après leur sécrétion, la charge négative des mucines détermine la viscosité du gel de mucus. La présence de cations condense les molécules de mucines en excluant l'eau.

Certaines mucines comme MUC1 et l'asialoglycoprotéine de rat (SMC) montrent une biosynthèse en deux étapes. Quand MUC1 néo-synthétisée est exprimée à la surface, sa glycosylation est incomplète. MUC1 est alors réinternalisée de la membrane vers le trans Golgi. La O-glycosylation y est alors achevée. MUC1 est exprimée de nouveau à la surface cellulaire mais cette fois sous forme mature (Litvinov and Hilkens, 1993) (Pimental et al., 1996).

Figure 2 : Vue en microscopie électronique à transmission de l'épithélium respiratoire (Gaillard et al., 1992).

II. 3. Structure des chaînes glycanniques des mucines.

Les chaînes glycanniques représentent 50 à 80 % du poids sec des mucines. Elles sont liées de façon covalente à l'axe peptidique. Les liaisons sont de type O-glycosidique entre des résidus de N-acétyl galactosamine (GalNac) situés à l'extrémité réductrice des chaînes glycanniques et les groupements hydroxylés de résidus de sérine ou de thréonine de l'axe peptidique.

Quelques chaînes N-glycanniques sont potentiellement présentes parmi une majorité de chaînes O-glycosidiques (Aubert et al., 1991) (Gum et al., 1992).

Les O-glycannes, libérés par traitement alcalin ont été séparés selon leurs charges par chromatographie d'échange ionique en plusieurs fractions dont le nombre dépend du nombre de paliers du gradient discontinu (Roussel et al., 1975) (Lamblin et al., 1977). On distingue schématiquement :

- les mucines neutres (pauvres en résidus acides)
- les mucines acides ou sialomucines (riches en résidus d'acide sialique)
- les mucines très acides ou sulfomucines (riches en résidus sulfate).

Bien qu'il n'existe que 5 types de sucres dans la plupart des mucines (N-acétyl glucosamine, N-acétyl galactosamine, galactose, fucose, acide N-acétyl neuraminique), leur assemblage conduit à une très grande variété de structures (plus d'une centaine), qui diffèrent par leur composition, leur longueur (de 1 à 20 résidus) et leur acidité (Lamblin et al., 1991). Les chaînes O-glycanniques possèdent 3 domaines (Hounsell and Feizi, 1982):

- le core ou noyau qui comprend le résidu de GalNac lié à l'axe peptidique et le ou les 2 sucres fixés sur cette GalNac. Huit cores différents ont été décrits pour l'ensemble des mucines (Roussel et al., 1988). La GalNac et le Gal β1, 3 GalNac correspondent respectivement aux antigènes Tn et T, en général masqués par d'autres sucres.

- le squelette qui est formé d'une association linéaire ou branchée d'unités disaccharidiques de types Gal β1, 3 GalNac (type 1) ou Gal β1, 4 GlcNac (type 2). Ils constituent les antigènes i, I eux aussi cryptiques.

- la région périphérique qui est caractérisée par la présence de sucres tels que le fucose, le galactose, la GalNac, l'acide sialique auxquels peuvent s'ajouter sur les résidus de galactose des groupements sulfates. Les antigènes de groupes sanguins ABO sont exprimés à la périphérie des chaînes glycanniques des mucines (Lamblin et al., 1991). Les

antigènes Lewis a, b (chaîne de type 1) et X, Y (chaînes de type 2) sont localisés dans la région périphérique.

II. 4. Les apomucines, aspects biochimiques.

La fraction peptidique ne représente que 20 à 50 % de la masse moléculaire des mucines. Les résidus de sérine, thréonine, alanine, glycine et proline représentent de 66 à 75 % de la composition en acides aminés dont 18 à 48 % de résidus de sérine et de thréonine. La teneur en cystéine a fait l'objet de nombreuses controverses, son taux varie de 0,5 à 3 % selon les préparations. Les résidus de cystéine sont pourtant à la base du modèle structural proposé pour les mucines. La réduction des ponts disulfures inter ou intra-chaînes provoque une diminution de la viscosité du mucus.

Les mucines sont associées en oligomères (Figure 3) (Sheehan et al., 1986) (Carlstedt et al., 1983). 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. Il a été proposé l'hypothèse selon laquelle les mucines s'oligomériseraient par la formation de ponts disulfures inter-chaînes dans leurs extrémités N- et C-terminales (Dekker et al., 1991).

Les mucines composant le mucus apparaissent en microscopie électronique comme de longs filaments flexibles (Thornton et al., 1991a) (Thornton et al., 1991b) constitués de l'assemblage de monomères de mucines. A ce jour, 3 mucines distinctes ont été identifiées comme formant des oligomères. Il s'agit des produits des gènes *MUC5AC*, *MUC5B* (Thornton et al., 1996c) (Thornton et al., 1997) (Wickstrom et al., 1998) et *MUC2* (Asker et al., 1995a).

Jusqu'en 1989, peu d'informations étaient connues concernant l'axe peptidique des mucines. Les méthodes classiques d'étude des protéines (coupure par des protéases spécifiques puis détermination de la séquence des peptides par la technique de dégradation d'Edman) se sont avérées être inefficaces dans le cas des mucines. Ces problèmes sont liés à l'accessibilité limitée des enzymes à leur cible en raison de la densité des chaînes glycanniques et également à l'instabilité des dérivés engendrés par la sérine et la thréonine lors de la réaction de dégradation récurrente d'Edman.

Afin de pallier ces difficultés, des laboratoires se sont orientés vers les techniques de l'ADN recombinant. Les stratégies employées comportaient les étapes suivantes :

- préparation d'un anticorps polyclonal ou monoclonal dirigé contre des mucines déglycosylées chimiquement
- préparation d'une banque d'ADNc dans un vecteur d'expression construite à partir de l'ARN de muqueuses ou de cellules mucisécrétantes
- immuno-criblage de cette banque et purification des clones positifs.
- détermination de la séquence nucléotidique et de la localisation chromosomique.
- étude de l'organisation génomique.

Cette stratégie a permis l'obtention de séquences partielles d'ADNc de mucines animales et de mucines humaines (digestives et respiratoires).

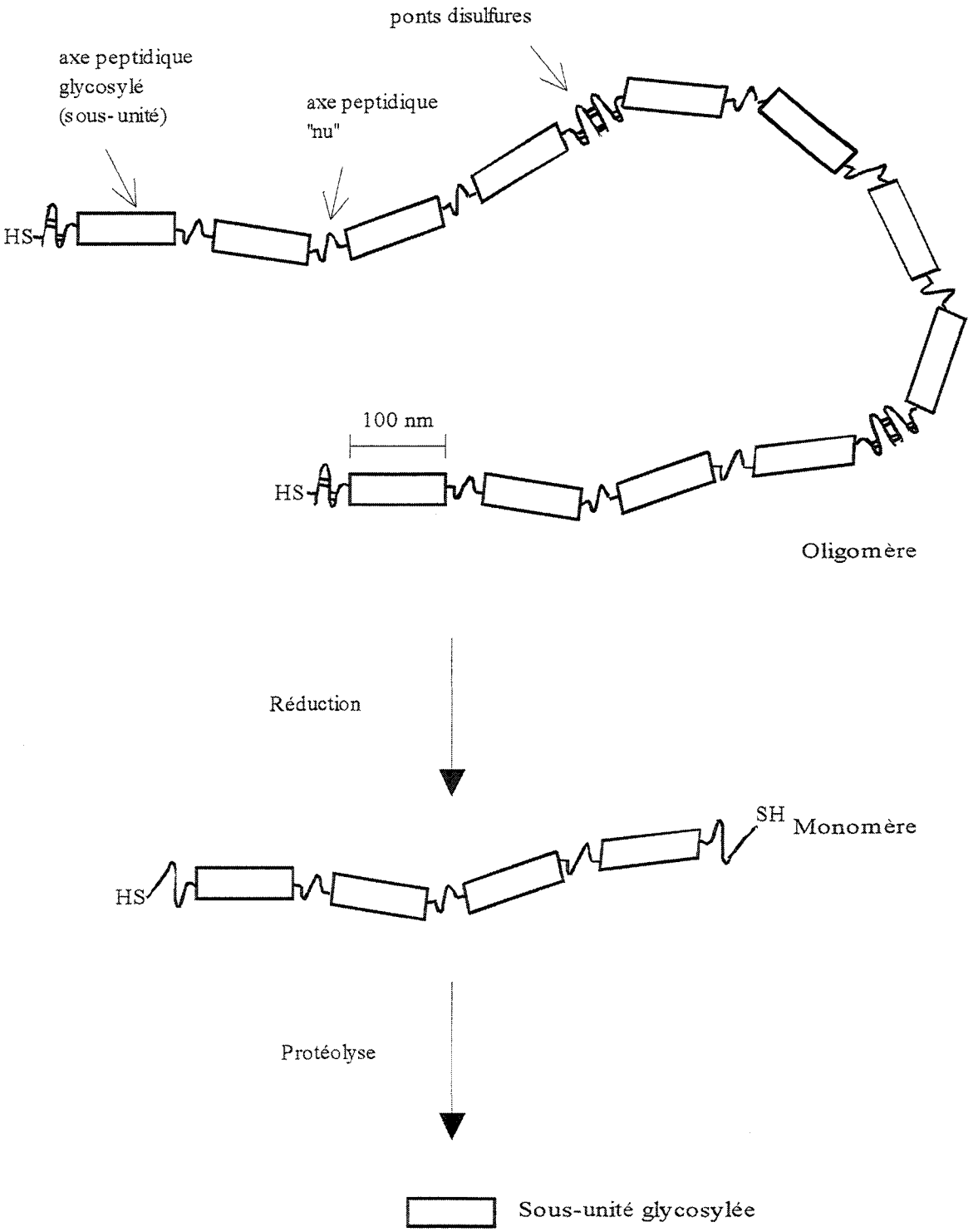


Figure 3 : Représentation schématique de la structure oligomérique des mucines d'après Carlstedt (Carlstedt et al., 1983).

II. 5. Les gènes d' apomucines.

Grâce à ces techniques de biologie moléculaire, des séquences partielles d'ADNc ont pu être isolées. Elles ont permis d'identifier et de localiser 8 gènes d'apomucines humaines, *MUC1, MUC2, MUC3, MUC4, MUC5B, MUC5AC, MUC6* et *MUC7* (Tableau 1). Leurs séquences déduites font apparaître des caractéristiques communes aux mucines.

Les apomucines sont des protéines modulaires, caractérisées par la présence d'un large domaine constitué de séquences répétées en tandem et riches en résidus de sérine, thréonine et proline (Tableau 2). Une séquence répétée différente est spécifique de chacune des mucines. Au niveau génomique, dans les gènes pour lesquels l'organisation est connue (c'est à dire *MUC1*, *MUC5B* et *MUC7*), ce domaine est codé par un seul exon situé en position centrale. En Southern blot, un polymorphisme interindividuel de type Variable Number of Tandem Repeat (VNTR) est détecté pour tous les gènes de mucines, exception faite du gène *MUC5B*.

Gène	**Localisation chromosomique**	**Référence**
MUC1	1q21-24	(Swallow et al., 1987)
MUC2	11p15.5	(Griffiths et al., 1990)
MUC3	7q22	(Fox et al., 1992)
MUC4	3q29	(Gross et al., 1992)
MUC5AC	11p15.5	(Nguyen et al., 1990)
MUC5B	11p15.5	(Nguyen et al., 1990)
MUC6	11p15.5	(Toribara et al., 1993)
MUC7	4q13-21	(Bobek et al., 1996)

Tableau 1 : Localisation chromosomique des gènes de mucines humaines.

Un long travail de criblage de banques génomiques a permis de cloner entièrement le gène *MUC4* grâce à la caractérisation et à l'étude essentiellement de 4 clones décrits dans l'article qui fait suite à cette partie.

En subtance les résultats sont les suivants :

- le domaine répétitif du gène *MUC4* est homogène et de grande taille

- il est localisé dans un seul exon

- le domaine répétitif présente un polymorphisme de type VNTR mais également un polymorphisme de mutation dans la population étudiée

- le gène *MUC4* contient un intron en aval du domaine répétitif lui-même organisé en répétition en tandem d'un motif de 15 pb

- l'organisation de la partie 5' du gène *MUC4* et du domaine répétitif codant a été déterminée et comporte 2 exons

- l'exon 1 est constitué d'une séquence 5' non traduite et d'un fragment de 82 pb codant les 27 premiers résidus N-terminaux. Ces derniers très hydrophobes correspondent au peptide signal de MUC4 qui présente de très fortes similarités avec celui de la mucine de rat SMC

- un intron d'au moins 15 kb est situé à la jonction entre les séquences codant le peptide signal et le peptide MUC4 mature

- un second exon code un peptide pouvant être divisé en 4 sous-domaines distincts, une région composée de répétitions imparfaites de 126 résidus d'acides aminés puis une région correspondant à une séquence unique, ces deux sous-domaines montrent des similarités avec la mucine de rat SMC. Un troisième domaine, le plus vaste, est composé des répétitions en tandem de 16 résidus. Enfin un quatrième domaine est constitué d'une séquence unique de 26 résidus.

Gène	aa	pb	Séquence consensus	Référence
MUC1	20	60	PDTRPAPGSTAPPAHGVTSA	(Gendler et al., 1988)
MUC2	23	69	PTTTPITTTTTVTPTPTPTGTQT	(Gum et al., 1989)
MUC3	17	51	HSTPSFTSSITTTETTS	(Gum et al., 1990)
MUC4	16	48	TSSASTGHATPLPVTD	(Porchet et al., 1991)
MUC5AC	8	24	TTSTTSAP	(Guyonnet-Dupérat, V et al., 1995)
MUC5B	29	87	Irrégulier ATG/SSTATPSST/SPGTT/AH/WTP/LP/TVLTTTATTPT	(Desseyn et al., 1997b)
MUC6	169	507	SPFSSTGPMTATSFOTTTTYPTPSHPOTTLPTHVPPFSTSLVTPSTGTYITPTHAOMATSASIHSTPTGTIPPPTTLKATGSTHTAPPMTPTTSGTSQAHSSFSTAKTSTSLHSHTSSTHHPEVTPTSTTTITPNPTSTGTSTPVAHTTSATSSRLPTPFTTHSPPTGS	(Toribara et al., 1993)
MUC7	23	69	TTAAPPTPSATTPAPPSSSAPPG	(Bobek et al., 1993)

Tableau 2 : Séquence peptidique consensus des domaines répétitifs des mucines humaines.

A l'origine, l'élément de classification le plus important des mucines reprenait le fait qu'elles soient des protéines sécrétées, constituants principaux du mucus. De ce fait, MUC1 qui peut être soit membranaire soit sécrétée était considérée comme une "mucin-like". Les données récentes, obtenues grâce à l'étude des séquences complètes des ADNc, permettent de classer les mucines épithéliales humaines en deux groupes :

- les mucines sécrétées
- les mucines membranaires ou associées à la membrane.

Notre étude du gène *MUC4* a participé à l'établissement de cette classification.

II. 5. 1. Les mucines sécrétées.

Les mucines appartenant à ce groupe sont exclusivement synthétisées et sécrétées par les cellules spécialisées de l'épithélium. Elles peuvent encore être subdivisées en deux groupes :

- les mucines formant le gel
- les mucines ne formant pas le gel.

II. 5. 1. 1. Les mucines sécrétées formant le gel.

Cette famille de mucines est exprimée par 4 gènes localisés sur le chromosome 11 dans la région p15.5. Ils sont rassemblés en un complexe de 400 kb très riche en îlots CpG situé entre *HRAS* et *IGF2* (Pigny et al., 1996) (Figure 4). Il s'agit de *MUC2, MUC5AC, MUC5B* et *MUC6*. Les séquences codantes déduites de ces 4 gènes sont organisées en modules selon un même schéma structural. Ils possèdent dans leurs parties distales de fortes homologies avec des modules du pré-profacteur de Von Willebrand (vWF) (Figure 5) (Mayadas and Wagner, 1991) (Mayadas and Wagner, 1992).

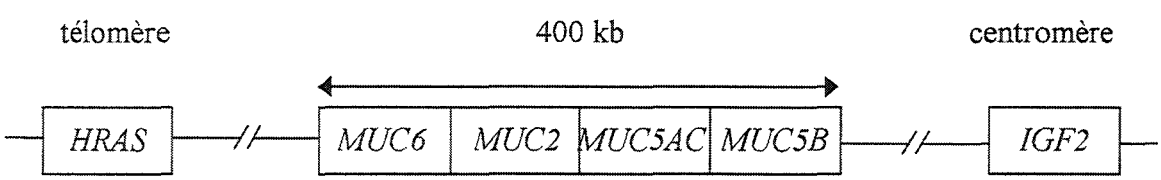


Figure 4 : organisation du cluster de gènes *MUC* localisés en 11p15.5.

Cette organisation en modules similaires n'est pas le seul trait commun des mucines dont les gènes sont localisés en 11p15.5. En effet, les 4 gènes sont rassemblés en un complexe de 400 kb où la distribution des sites de restriction, avec de nombreuses symétries et répétitions, est évocatrice de l'existence de nombreux événements de duplication (Pigny et al., 1996). Ces gènes seraient issus d'un gène ancestral commun (Desseyn et al., 1998a).

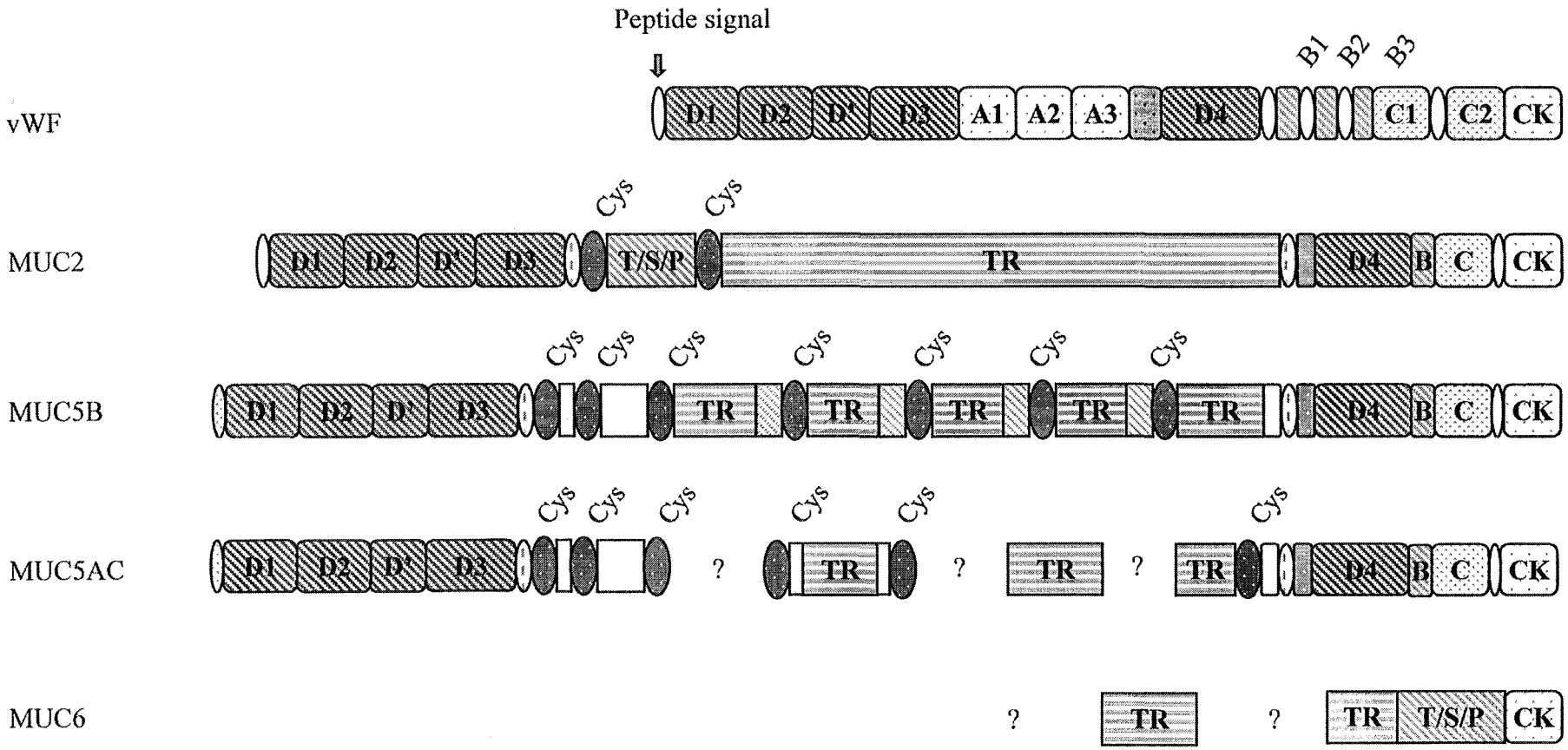


Figure 5 : Représentation schématique des séquences peptidiques déduites des ADNc du vWf et des gènes de mucines localisés sur le chromosome 11 dans la région p15.5.

MUC2

MUC2 a été identifié et décrit pour la première fois par Gum *et al* (Gum et al., 1989) après criblage d'une banque d'expression de jéjunum grâce à un immunsérum dirigé contre l'axe peptidique de mucines purifiées de la lignée tumorale colique LS174T. Tous les clones isolés sont constitués de séquences répétées en tandem de 69 nucléotides. De nouveaux criblages de banques d'ADNc et de gènes à partir de ces clones utilisés comme sondes ainsi que des expériences de RACE-PCR (Rapid Amplification of cDNA Ends) ont permis d'isoler un ADNc complet d'une taille de 15720 pb pour son plus grand allèle (Gum et al., 1994).

Le domaine répétitif de MUC2 est composé de deux domaines contenant des répétitions en tandem. Le premier, en position centrale, contient la séquence de 23 résidus d'acides aminés. Ce domaine montre un polymorphisme de type VNTR et le nombre de répétitions varie de 51 à 115 (Toribara et al., 1991). Le second est composé d'un motif irrégulier de 347 résidus. Ces deux domaines sont riches en résidus de sérine, thréonine et proline. L'étude des mucines purifiées par chromatographie à partir de la lignée cellulaire issue de cancer du côlon LS174T montrent que 78 % des résidus de thréonine sont O-glycosylés. MUC2 peut donc contenir plus de 1000 chaînes saccharidiques (Byrd et al., 1988). Le domaine contenant la séquence répétée de 347 résidus est encadré de part et d'autre par deux séquences peptidiques très similaires et contenant des résidus de cystéine. Ces motifs d'environ 110 résidus sont appelés motifs Cys.

MUC2 contient également 5 domaines D, appelés ainsi à cause de leur similarité avec les domaines D du pré-profacteur de Von Willebrand. Les domaines D1, D2, D' et D3 sont situés dans sa région N-terminale, le domaine D4 est lui en position C-terminale. En aval du domaine D4, deux autres modules montrent des similarités avec des domaines du vWF, un domaine C et un domaine CK (Cystine Knot). Ce domaine CK est retrouvé dans d'autres protéines sécrétées (Sun and Davies, 1995) dont la NDP (Norrie Disease Protein). La modélisation moléculaire de la NDP montre que son domaine CK possède une structure tertiaire similaire à celle du TGFβ (Transforming Growth Factor) (Meitinger et al., 1993).

MUC5B

Le premier clone d'ADNc isolé pour ce gène a été obtenu après criblage d'une banque d'expression de trachée humaine avec un immunsérum dirigé contre des glycopeptides déglycosylés de mucines trachéobronchiques humaines. Ce clone est constitué de la répétition imparfaite de 87 pb (Dufossé et al., 1993). Par des techniques

comparables à celles utilisées pour *MUC2* un ADNc complet ayant une séquence codante de 16986 pb a pu être isolé et caractérisé au laboratoire (Desseyn et al., 1997a), (Desseyn et al., 1997b), (Desseyn et al., 1998b).

Comme pour les autres mucines, son domaine répétitif est codé par un large exon en position centrale (Desseyn et al., 1997b). D'une taille de 3570 résidus d'acides aminés, le module central de MUC5B est composé d'un arrangement alterné de trois types de sous-domaines. 19 sous-domaines peuvent y être individualisés. 7 sous-domaines de 108 résidus, appelés Cys1 à Cys7 contiennent 10 résidus de cystéine chacun et montrent une organisation comparable aux motifs Cys rencontrés dans MUC2. 5 sous-domaines sont composés de la séquence imparfaite de 29 résidus répétée en tandem. 4 de ces 5 sous-domaines sont suivis d'une séquence identique mais différente du motif répétitif de 29 acides aminés. 3 sous-domaines sont composés de séquences uniques, riches en résidus de sérine, thréonine, proline et alanine. Les auteurs décrivent donc ce domaine central comme composé de 4 "super-repeats" de 528 résidus d'acides aminés.

L'extrémité 5'-terminale de son ADNc, d'une taille de 4023 pb, est constituée de 30 exons. La séquence déduite code 4 domaines D similaires à ceux du vWF, les domaines D1, D2, D', D3 (Desseyn et al., 1998b) (Offner et al., 1998). L'extrémité 3'-terminale de son ADNc a une taille de 2988 pb. Elle est constituée de 18 exons dont la séquence déduite code des domaines similaires aux domaines D4, C et CK du vWF. (Desseyn et al., 1997a) (Keates et al., 1997).

MUC5AC

Plusieurs séquences partielles d'ADNc sont connues pour *MUC5AC*. 2 clones ont été obtenus au laboratoire après criblage de la banque d'expression de trachée en parallèle du travail réalisé sur *MUC5B* (Guyonnet-Dupérat, V et al., 1995). Ces clones, JER47 et JER58, sont composés d'une répétition de 24 pb encadrée pour JER47 de séquences de 330 pb codant des domaines riches en cystéine comparables aux motifs Cys de MUC2 et MUC5B. Le domaine central de MUC5AC semble donc être similaire à celui de MUC5B.

Le clone NP3a a été obtenu par criblage d'une banque d'expression de polype nasal (Meerzaman et al., 1994) et le clone L31 a été isolé d'une banque d'expression construite à partir de la lignée cellulaire de cancer du côlon HT-29 MTX (clone résistant au méthotrexate 10^{-6} M) (Lesuffleur et al., 1995). Ces deux clones semblent coder l'extrémité C-terminale de MUC5AC. Ces clones contiennent les domaines similaires au vWF correspondant au domaine D4, C et CK. L'organisation génomique de l'extrémité 3'-

terminale de *MUC5AC* montre des similarités structurales avec l'organisation 3'-terminale de *MUC5B*. Elle est également composée de 18 exons qui ont une taille comparable à ceux composant l'extrémité 3'-terminale de *MUC5B* (Buisine et al., 1998a).

Deux autres clones, obtenus après criblage d'une banque d'expression construite à partir d'estomac humain, semblent coder l'extrémité N-terminale de MUC5AC. Le peptide déduit de HGM-1 (Klomp et al., 1995) est similaire au domaine D3 du vWF, et contient également 3 domaines riches en résidus de cystéine comparables aux motifs Cys de MUC2 et MUC5B. Le peptide déduit de HGM-2 (Li et al., 1998) (van de Bovenkamp et al., 1998) recouvre quant à lui les domaines D1, D2 et D'.

MUC6

MUC6 est la mucine la moins bien caractérisée de celles dont les gènes sont localisés en 11p15.5. Le criblage d'une banque d'expression construite à partir d'estomac a permis d'isoler le premier clone d'ADNc (Toribara et al., 1993). Sa séquence révèle la présence d'un motif répété en tandem de 507 pb.

Un second clone a été obtenu par la technique de RACE-PCR sur un ARN total d'estomac (Toribara et al., 1997). Ce clone de 1735 pb est composé d'une séquence codante de 1083 pb ainsi que d'une séquence non traduite de 652 pb. La séquence peptidique déduite code 2 domaines. Le premier, de 270 résidus d'acides aminés est riche en résidus de sérine, thréonine et proline. Il ne contient pas de cystéine. Le second module, de 91 résidus, montre une similarité avec le domaine CK du vWF, MUC2, MUC5B et MUC5AC.

Mis à part, MUC6 qui ne possède que le domaine CK, les mucines dont les gènes sont localisées en 11p15.5 montrent une organisation structurale semblable, comparable à celle du vWF. Les domaines A du vWF sont remplacés par le module répétitif central pour les mucines. Les domaines D et CK sont impliqués dans le processus d'oligomérisation du vWF (Voorberg et al., 1991). Le vWF est capable de former des dimères par l'intermédiaire de son domaine CK et de multimériser grâce à son domaine D3 (Dong et al., 1994). Des études récentes réalisées sur la lignée tumorale colique LS174T montrent que ces 4 mucines forment des homodimères (van Klinken et al., 1998) (Asker et al., 1998b) (Asker et al., 1995a). Ce processus d'homodimérisation est très précoce puisqu'il précède la O-glycosylation. Il s'effectue dans le réticulum endoplasmique rugueux par formation de ponts disulfures au niveau des domaines CK de deux monomères. Les

homodimères de mucines s'oligomérisent ensuite par l'intermédiaire de ponts disulfures grâce à leur domaine D3 (Offner et al., 1998). Les multimères forment alors le réseau tridimensionnel du gel de mucus.

De nombreux éléments liés à la structure et à l'expression de ces gènes laissent suspecter une régulation d'expression concertée pour ces 4 gènes. Peu de travaux sont à ce jour publiés sur la régulation de leur expression. Seules les régions promotrices de *MUC2* et *MUC5AC* sont connues.

Le promoteur de *MUC2* contient une boîte TATA en position -25. L'activité basale de transcription est assurée par une région comprise entre les nucléotides -91/-73 par rapport au point d'initiation de la transcription. Elle contient une boîte CACCC capable de lier les protéines de la famille Sp dont Sp1. La région comprise entre les bases -228/-171 confère la spécificité cellulaire d'expression (Gum et al., 1997a). L'activité transcriptionnelle maximale est détectée pour une région allant du nucléotide -848 à + 1. Deux éléments de réponse à *Pseudomonas aeruginosa* ont également été identifiés (Li et al., 1997).

Pour *MUC5AC*, la zone promotrice n'est identifiée que depuis peu. Elle comprend une boîte TATA et le même motif CACCC découvert pour *MUC2*. Elle contient également des sites potentiels de liaison à Sp1, NFkB et des éléments de réponse à *Pseudomonas aeruginosa* (Li et al., 1998).

II. 5. 1. 2. Les mucines sécrétées ne formant pas le gel.

MUC7

Cette définition ne concerne à ce jour qu'une seule mucine : MUC7. La mucine MUC7 a été caractérisée dans un premier temps par des techniques biochimiques comme étant la fraction MG2 issue de glandes salivaires humaines. L'ADNc complet a été isolé. Il code une protéine de 377 résidus qui peut être divisé en 3 domaines. Le domaine central est composé de la répétition de 23 résidus d'acides aminés. Le gène correspondant ne montre que 2 allèles, l'allèle majeur contient 6 fois la répétition, le second ne la contient que 5 fois. Les domaines N- et C-terminaux ne possèdent aucun module cystéine mais montrent une richesse en sérine, thréonine et proline (Bobek et al., 1993) (Bobek et al., 1996). En comparaison avec celles des mucines dont les gènes sont localisés en 11p15.5, la structure de MUC7 est très simple (Figure 6). L'expression des messagers de *MUC7* est observée spécifiquement dans les cellules séreuses salivaires et bronchiques.

Quelques éléments se rapportant à la régulation de son expression ont pu être identifiés. Ainsi, en plus d'une boîte TATA et d'une boîte CAAT, sa séquence régulatrice comporte un élément de réponse aux glucocorticoïdes et à l'AMPc.

MUC7 est décrit comme une molécule d'adhérence privilégiée aux souches de streptocoques. Le rôle qui lui est attribué serait de moduler la clairance bactérienne dans la cavité orale (Levine et al., 1978).

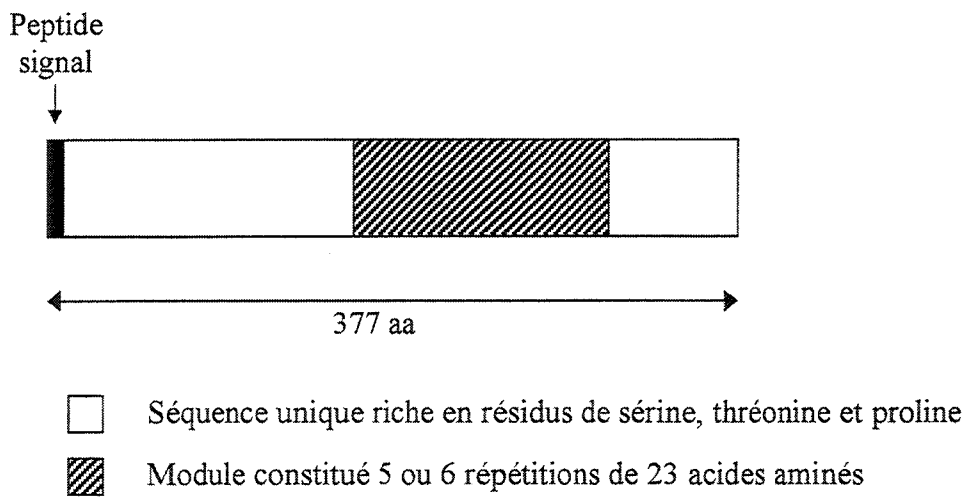


Figure 6 : Représentation schématique de la mucine MUC7.

II. 5. 2. Les mucines membranaires ou associées à la membrane.

Cette famille de mucines a très longtemps été constituée d'un seul membre, MUC1. MUC1 peut être à la fois soluble et membranaire. Elle a très longtemps été considérée comme une "mucin-like". Les données récentes concernant MUC3, ainsi que nos résultats concernant MUC4 permettent de définir une nouvelle famille de mucines, les mucines membranaires ou associées à la membrane.

MUC1

Comme son expression et sa glycosylation sont perturbées dans de nombreux carcinomes, MUC1 suscite de nombreux travaux. Un grand nombre d'anticorps dirigés spécifiquement contre des cellules cancéreuses se fixent en réalité sur des épitopes de MUC1. MUC1 a été identifiée successivement sous plusieurs dénominations dont PEM (polymorphic epithelial mucin), épisialine, épitectine, EMA (epithelial membrane antigen) PAS-O, DUPAN-2, PUM (peanut-lectin binding urinary mucin), Ca1, NPGP (non

penetrating glycoprotein), NCRC11, MAM-6, DF3 antigène, SGA (sebaceous gland antigen), H23 antigène et HMFG (human milk fat globul) antigène jusqu'à ce que la découverte du gène clarifie la nomenclature.

MUC1 a été isolée à partir de tissus variés comme la glande mammaire, le pancréas et l'ovaire (Gendler et al., 1987) (Gendler et al., 1990) (Stern et al., 1992). Bien que dans ces tissus l'apomucine MUC1 soit identique, la mucine mature glycosylée est différente. En effet, il existe des glycoformes distinctes pour MUC1. MUC1, synthétisée dans la glande mammaire, a une masse moléculaire qui varie entre 250 à 500 kDa, ce qui correspond à un taux de glycosylation de 50 % (Shimizu and Yamauchi, 1982). Au niveau du pancréas, elle a une masse qui dépasse 1000 kDa, donc un taux de glycosylation de 80 % (Lan et al., 1987).

MUC1 est la première mucine dont le gène ait été caractérisé et étudié. Les séquences complètes de son ADNc et de son gène sont publiées, aussi bien chez l'homme (Gendler et al., 1987) (Gendler et al., 1990) (Lan et al.,1990) que chez la souris (Spicer et al., 1991a). Comme les autres mucines, MUC1 est une protéine modulaire. Son domaine central est composé de la répétition d'un motif de 20 résidus. Ce domaine présente un polymorphisme de type VNTR, il varie en taille de 400 à 2400 résidus. Les séquences en position N- et C-terminales du domaine central se composent du même motif répétitif mais dégénéré. La dégénérescence s'accroît en s'éloignant du centre de la protéine. L'extrémité C-terminale contient également une séquence transmembranaire ainsi qu'une queue cytoplasmique. L'ADNc est composé de 7 exons. L'exon 1 code le peptide signal. L'exon 2 est très grand. Il code le domaine central répétitif ainsi que les séquences N- et C-terminales dégénérées de ce domaine. La séquence transmembranaire est codée par l'exon 6 et le domaine cytoplasmique par l'exon 7 (Figure 7).

Deux isoformes sont identifiées pour MUC1. Elles résultent de l'utilisation alternative de 2 sites accepteurs d'épissage pour l'exon 2 (Figure 8). Cet épissage alternatif qui délète l'un des deux variants de 27 pb se traduit au niveau de l'apomucine par une modification du peptide signal. La résultante pour MUC1 mature divise les deux groupes de recherche qui ont identifié ce mécanisme d'épissage alternatif. En effet pour Ligtenberg et *al* (Ligtenberg et al., 1990) la modification du peptide signal est le seul point qui différencie les deux variants. Cette modification serait cependant suffisante pour induire un "processing" post traductionnel spécifique à chacune des isoformes. Ceci expliquerait la différence de localisation membranaire détectée dans les cellules cancéreuses. Pour Wreschner et *al* (Wreschner et al., 1990) la modification de l'exon 2 entraîne un défaut

d'épissage en aval du domaine répétitif de 60 pb. Le transcrit correspondant code une protéine tronquée de ses domaines C-terminaux dont le domaine transmembranaire, ce qui se traduirait par la synthèse de MUC1 sécrétée.

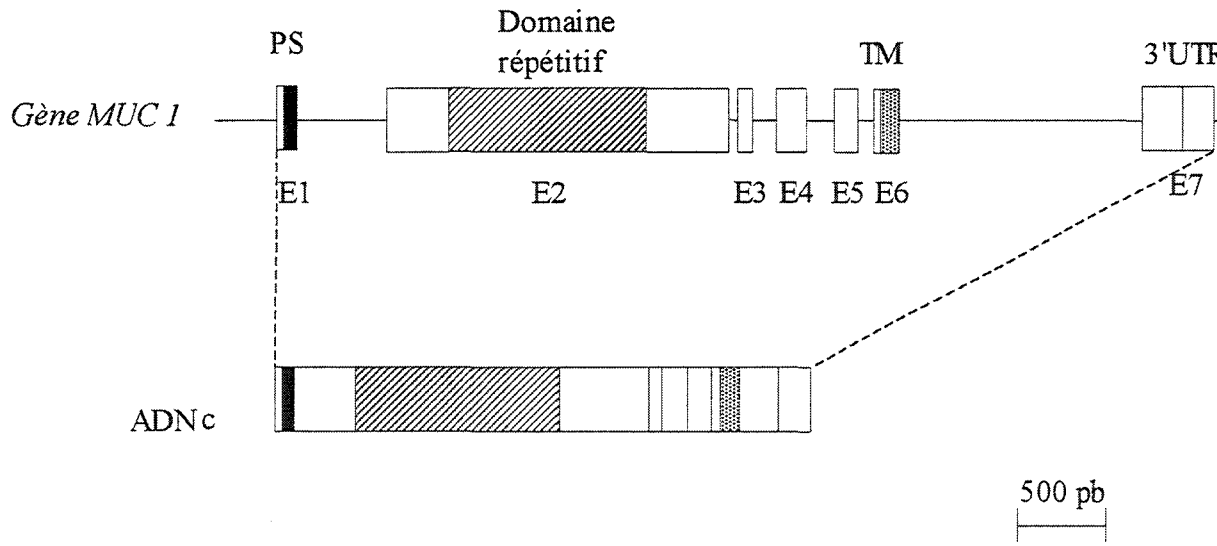


Figure 7 : Représentation schématique de l'organisation génomique de *MUC1*.

Bien que MUC1 soit une mucine membranaire, une forme soluble ou sécrétée existe dans le surnageant de cellules en culture (Levine et al., 1978) comme dans les liquides biologiques (Burchell et al., 1984). Plus récemment, il a été montré que la forme soluble pouvait être détectée sans qu'il y ait d'épissage alternatif. La forme soluble, caractérisée biochimiquement, est juste dépourvue de sa séquence transmembranaire et de son domaine cytoplasmique. Boshell et *al* suggèrent que la présence de MUC1 soluble puisse résulter de l'action d'une protéase (Boshell et al., 1992). Il est proposé que l'action de la protéase suive la seconde étape de O-glycosylation qui a lieu dans le trans Golgi lors du processus de réinternalisation de MUC1 (Ligtenberg et al., 1992) (Litvinov and Hilkens, 1993).

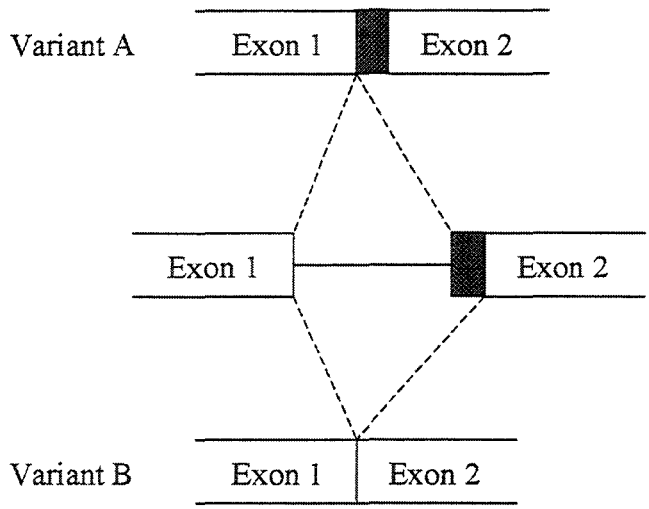


Figure 8 : Représentation schématique du mécanisme d'épissage alternatif de *MUC1* (organinastion genommique partielle).

L'étude du promoteur du gène *MUC1* montre une organisation typique d'un gène spécifique de tissu avec la présence de nombreuses séquences GC. La région promotrice en amont du site d'initiation comprend une boîte TATA et une boîte CAAT, 3 boîtes GC, 2 sites Sp1, un site de fixation pour le "milk binding protein factor" et une boite E. Cette boite E détermine l'expression spécifique de tissu de *MUC1* (Kovarik et al., 1993). La région –505/-485 pb fixe une protéine non identifiée de 45 kDa (Abe and Kufe, 1993). La région en aval de la boîte E contient une boîte PMR (perfect homopurine mirror repeats) (Hollingsworth et al., 1994a). Cette boîte fixe une protéine de 27 kDa. Une autre boîte dénommée RME (responsive mucin element) a été identifiée. Elle se lie à une protéine de 70 kDa (Shirotani et al., 1994).

Muc1

L'analogue murin de MUC1 (Muc1) est entièrement connu (Spicer et al., 1991a). Comme son homologue humain, elle est codée à partir d'un ARNm constitué de 7 exons. Les séquences déduites de Muc1 sont similaires à MUC1 pour les domaines C-terminaux. En position N-terminale, seuls les peptides signaux des 2 molécules sont similaires (Figure 9).

Bien que la séquence du peptide signal soit conservée entre MUC1 humain et Muc1 de souris, aucun site accepteur cryptique n'est retrouvé au voisinage du site d'épissage

entre l'exon 1 et l'exon 2 chez la souris. Les auteurs concluent que l'épissage alternatif décrit pour MUC1 entre ces 2 exons n'existe pas pour Muc1 de souris.

Le module central de Muc1 est constitué de séquences répétées en tandem riches en sérine, thréonine et proline. Le motif répétitif est imparfait et varie de 20 à 21 résidus. Le domaine central ne montre pas de polymorphisme interindividuel de type VNTR.

Les similarités les plus fortes entre les 2 apomucines sont retrouvées dans le domaine transmembranaire et la queue cytoplasmique. Pour ces auteurs, le fait que la queue cytoplasmique soit le domaine le plus conservé montre bien qu'elle est responsable des fonctions biologiques les plus importantes de MUC1.

Une autre différence importante existe entre les deux molécules : la structure de leur promoteur n'est pas similaire. En effet, tous les éléments de réponse potentiels de *MUC1* humain ne sont pas retrouvés pour *Muc1* de souris. Seul l'élément de réponse à Sp1 est présent dans les 2 molécules. Par contre, une région comprise entre les nucléotides –92 et –34 montre 100 % de similarité. L'analyse des banques de données informatiques n'a pas permis d'identifier d'élément de réponse potentiel dans cette zone. Pour les auteurs, cette petite région serait responsable de l'expression tissu spécifique de *MUC1* et *Muc1*.

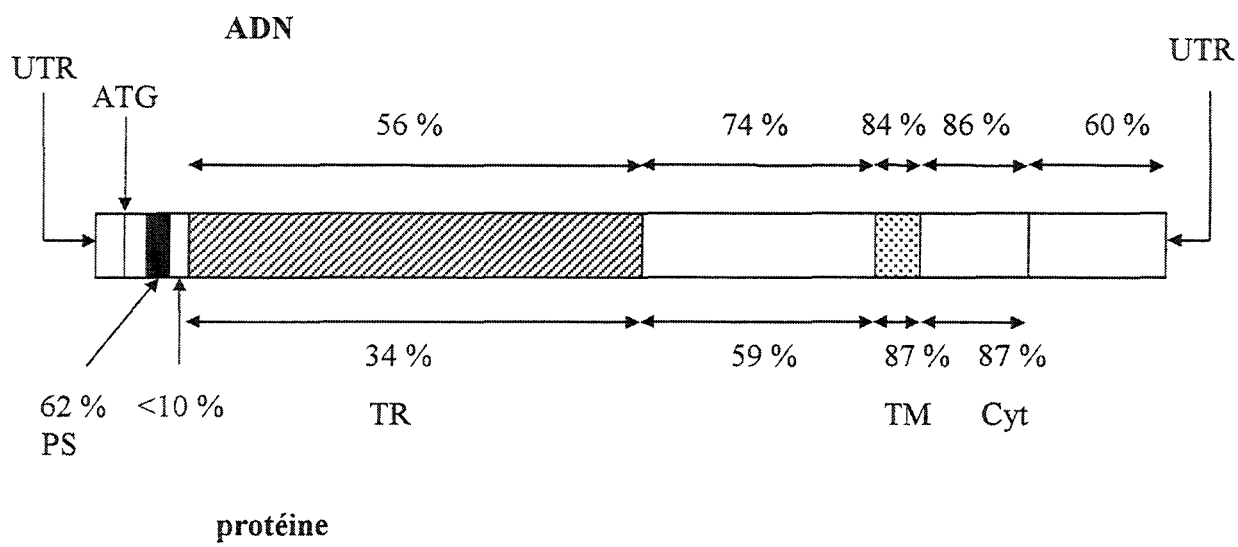


Figure 9 : Représentation schématique de Muc1 de souris. Les pourcentages de similarité de chacun des domaines par rapport à MUC1 humain sont donnés pour l'ADN (au-dessus du schéma) et pour la protéine (en dessous du schéma).
PS : peptide signal ; TR : tandem repeat ; TM : séquence transmembranaire ; Cyt : queue cytoplasmique.

MUC3

Les informations concernant l'organisation structurale de MUC3 sont encore incomplètes. Il apparaît cependant que MUC3 est une protéine modulaire comme les autres mucines. Les modules N-terminaux sont inconnus. Son domaine central est encore aujourd'hui sujet à controverse. Il serait au moins composé de 2 séquences répétées en tandem, l'une ayant un motif de 59 résidus d'acides aminés pour van Klinken et al. (van Klinken et al., 1997) ou un motif de 375 résidus d'acide aminé pour Gum et al. (Gum et al., 1997b) et l'autre ayant un motif de 17 résidus (Gum et al., 1990). Comme il n'existe aucune connaissance quant à l'organisation génomique de *MUC3*, il n'est pas possible de vérifier si ce domaine est codé par un seul et même exon.

L'extrémité C-terminale contient un grand domaine riche en résidus de sérine, thréonine et proline, suivi de 2 domaines de type EGF, un domaine transmembranaire et une queue cytoplasmique (Figure 10) (Gum et al., 1997b) (Crawley et al., 1999). Comme aucun clone codant l'extrémité N-terminale n'a pu être isolé, aucune information sur la régulation de son expression n'est connue.

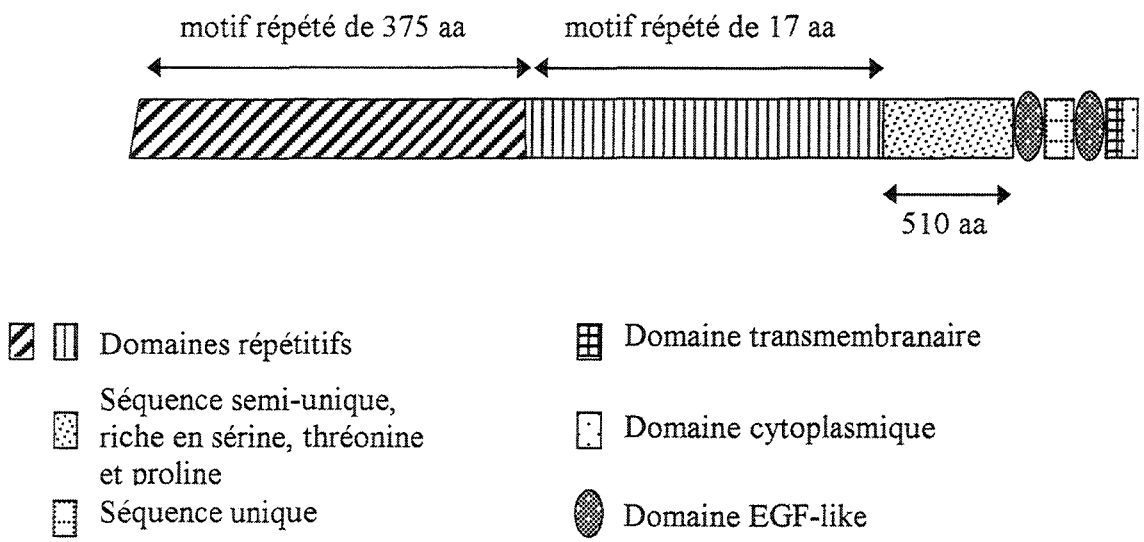


Figure 10 : Représentation schématique des séquences déduites pour MUC3.

Après criblage de banques et RT-PCR, différents clones d'ADNc résultant de mécanismes d'épissage alternatif ont été isolés (Crawley et al., 1999). Ces clones codent soit un MUC3 tronqué de ses modules C-terminaux, soit un MUC3 délété de l'un de ses modules. Ces résultats sont très récents et aucune fonction précise n'a encore été associée à ces différents variants.

Muc3

Seul les domaines C-terminaux de MUC3 ont pu être isolés pour le rat (Khatri et al., 1997) et la souris (Shekels et al., 1998). Chez la souris, le domaine central constitué de la répétition d'un motif de 18 pb est suivi d'une séquence unique de 1137 pb qui code différents modules présents dans MUC3 humain. Muc3 de souris possède 2 domaines de type EGF, un domaine transmembranaire et un domaine cytoplasmique. La même structure est rencontrée pour Muc3 de rat. Aucun variant n'a été décrit à ce jour concernant Muc3 de souris ou de rat. Ces résultats étant récents, aucune étude fonctionnelle n'est publiée.

MUC4

Ce gène a été isolé au laboratoire par criblage d'une banque d'ADNc d'origine trachéobronchique (Porchet et al., 1991). Il a été localisé en 3q29 (Gross et al., 1992). Le clone obtenu était caractérisé par une répétition en tandem d'une séquence élémentaire de 48 pb. Le criblage d'une banque génomique à l'aide de cette sonde était en cours au laboratoire lorsque nous avons abordé notre travail de DEA et de thèse. Nos résultats ont permis de rattacher MUC4 à cette classe de mucines membranaires.

SMC

Découverte dans les années 1980, cette mucine de rat dénommée "sialomucin complex" (SMC) est une glycoprotéine hétérodimérique de haute masse moléculaire mise en évidence à la surface des cellules d'ascites de la lignée 13762 issue d'adénocarcinome mammaire de rat.

SMC est composée de 2 sous-unités : l'"ascite sialoglycoprotein-1" (ASGP-1) et l'"ascite sialoglycoprotein-2" (ASGP-2) (Sherblom and Carraway, 1980). L'ASGP-1 est une glycoprotéine sécrétée de 600 kDa. Il s'agit d'un peptide de 220 kDa O-glycosylé à 70 % (Hull et al., 1984).

En 1994, la caractérisation d'un ADNc partiel codant l'ASGP-1 a permis de définir la structure de cette glycoprotéine (Wu et al., 1994). Sa région N-terminale est composée d'une séquence unique de 80 résidus d'acides aminés dont les 30 premiers forment le peptide signal. Son module central est constitué de 1513 résidus riches en sérine, thréonine et proline. Il correspond à 12 fois la répétition d'un motif variant de 117 à 124 résidus d'acides aminés. Sa région C-terminale est composée d'une séquence unique de 609 résidus. L'ASGP-1 est une glycoprotéine sécrétée qui se lie d'une manière stable grâce à

des liaisons non covalentes à une glycoprotéine membranaire, l'ASGP-2 (Sherblom and Carraway, 1980) (Helm and Carraway, 1981).

Un ADNc partiel codant l'ASGP-2 a également été isolé (Sheng et al., 1992). La séquence déduite code une protéine de 80 kDa organisée en module. L'ASGP-2 comprend 2 domaines riches en sites potentiels de N-glycosylation, 2 domaines de type EGF, un domaine riche en résidus de cystéine, un domaine hydrophobe transmembranaire et un domaine cytoplasmique. L'ASGP-2 contient 24 sites potentiels de N-glycosylation dont 17 sont glycosylés (Hull et al., 1990). L'ASGP-1 et l'ASGP-2 sont issues d'un même précurseur protéique, pSMC-1 (Figure 11) et donc du même ARNm (Sheng et al., 1990).

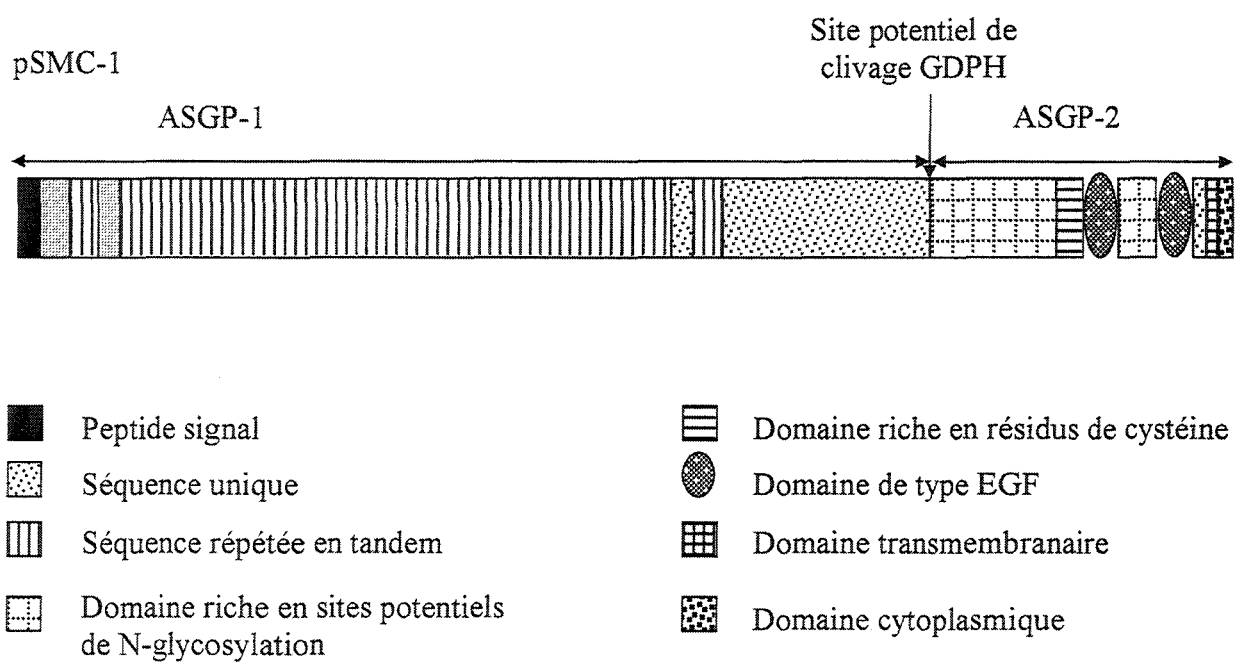


Figure 11 : Représentation schématique du précurseur protéique pSMC-1 (Mc Neer et al., 1997).

Un site potentiel de clivage protéolytique GDPH (GlyAspProHis) est rencontré dans la séquence déduite de l'ADNc complet de SMC entre l'ASGP-1 et l'ASGP-2.

Les auteurs proposent un modèle de biosynthèse de SMC. pSMC-1 est synthétisé et N-glycosylé dans le réticulum endoplasmique rugueux. Le précurseur est alors clivé en même temps que débute la O-glycosylation dans le Golgi (Figure 12). Notons qu'au cours de ces expériences, un autre précurseur du nom de pSMC-2 a pu être détecté. pSMC-2 ne contient qu'une partie des domaines constitutifs de l'ASGP-2. L'origine de ce second précurseur est encore inconnue. Il semble cependant qu'il ne soit pas issu de pSMC-1 après maturation post-traductionnelle. La recherche par RT-PCR de variant dû à un mécanisme d'épissage alternatif est restée également infructueuse.

La régulation de l'expression de SMC n'est pas encore connue. Aucune séquence régulatrice n'est publiée. Cependant, certains facteurs ont une action sur cette régulation. Le tableau 3 résume les connaissances actuelles dans ce domaine.

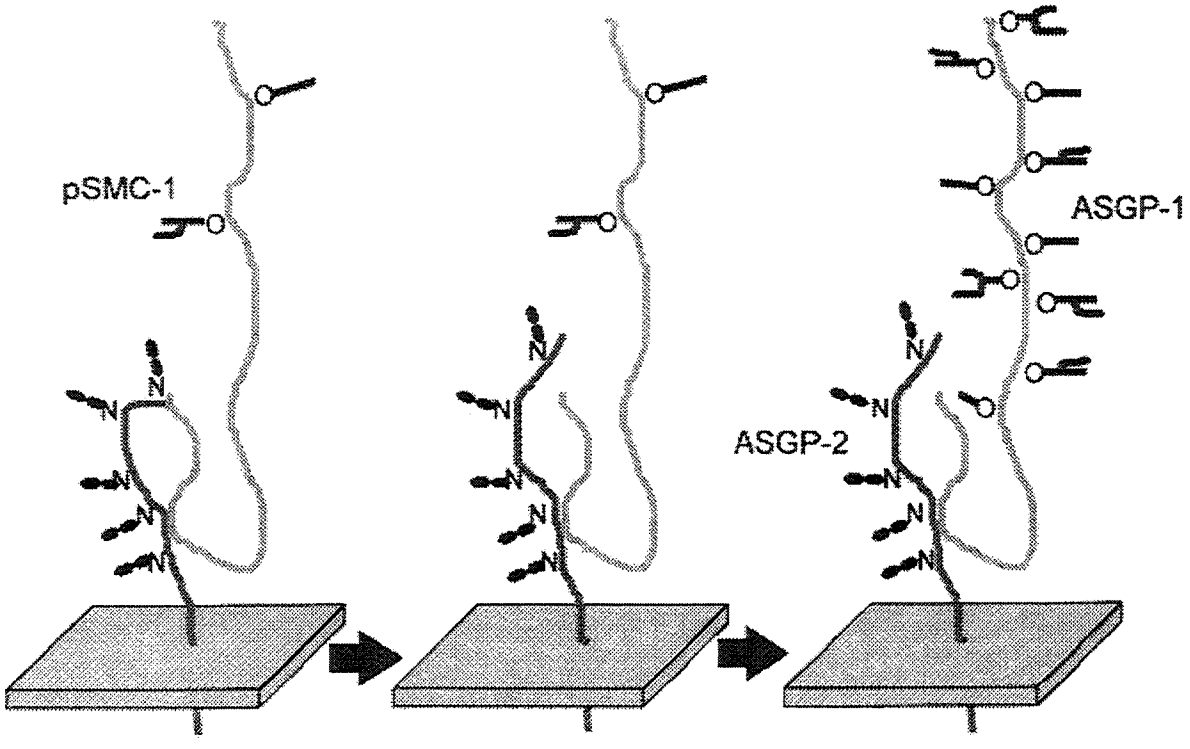


Figure 12 : Représentation schématique de la biosynthèse de l'ASGP-1 et l'ASGP-2 à partir de pSMC-1.

	niveau d'expression chez la rate			facteur		
tissu	non gestante	implantation	fin de gestation	positif	negatif	sans effet
glande mammaire	-/+	+	++	prolactine		oestradiol progestérone
utérus	++	-	++	oestradiol	progestérone	

Tableau 3 : Expression et régulation de SMC dans différents types tissulaires.

III. Les mucines endothéliales et leucocytaires

Ces molécules ne sont pas le propos de cette thèse mais elles sont parfois dénommées sous le terme de mucines dans la littérature (Shimizu and Shaw, 1993). Les mucines endothéliales regroupent de nombreuses molécules qui ne se retrouvent pas sous forme sécrétée dans le mucus. Ainsi de nombreuses O-glycoprotéines membranaires, impliquées dans la reconnaissance cellule-cellule ou cellule-matrice extracellulaire, sont également dénommées mucines ou "mucin-like". Ces protéines sont impliquées dans les mécanismes d'adhérence, les mécanismes initiant les réactions inflammatoires et le "homing" des lymphocytes T. Ces molécules sont associées à l'endothélium ou aux leucocytes et agissent comme ligand envers les membres de la famille des sélectines (van Klinken et al., 1995) (Figure 13).

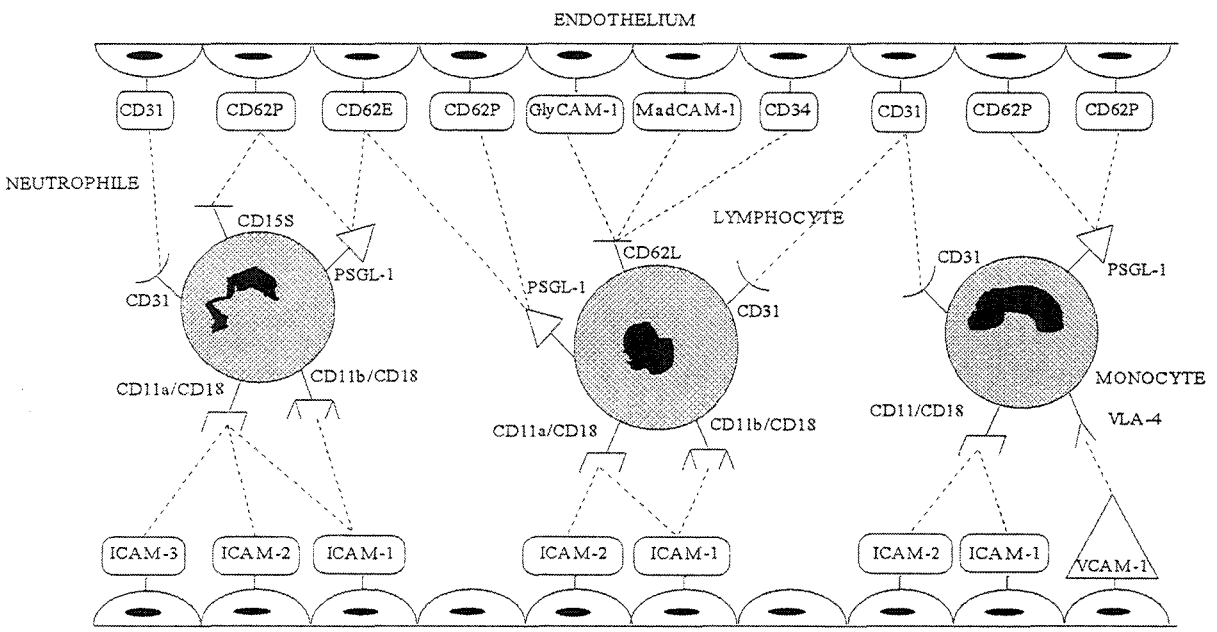


Figure 13 : molécules adhésives et leurs ligands à la surface de l'endothélium et des cellules immunitaires : leukocyte function antigen-1 (CD11a/CD18) ; Mac-1 (CD11b/CD18) ;p150.95 (CD11a/CD18) ; intercellular adhesion molecule ICAM-1 (CD54) ; ICAM-2 (CD102) ; ICAM-3 (CD50) ; very late activation VLA-4 (CD49d) ; vascular adhesion molecule VCAM-1 (CD106) ; platelet-endothelial cell adhesion molecule PE-CAM-1 (CD31) ; E-selectin (CD62E) ; P-selectin (CD62P) ; tetrasaccharide sialyl Lewis^x (CD15s) ; P-selectin glycoprotein ligand-1 (PSGL-1) ; L-selectin (CD62L) ; GlyCAM-1 ; mucosal addressin cell adhesion molecule MadCAM-1 ; CD34.

Shimizu et Shaw proposent une classification pour ces molécules. Ils parlent de mucines endothéliales et de mucines leucocytaires (Figure 14). Ces molécules sont également classées dans d'autres familles de protéines, comme les intégrines, la superfamille des immunoglobulines, les sélectines et les cadhérines (Elangbam et al., 1997).

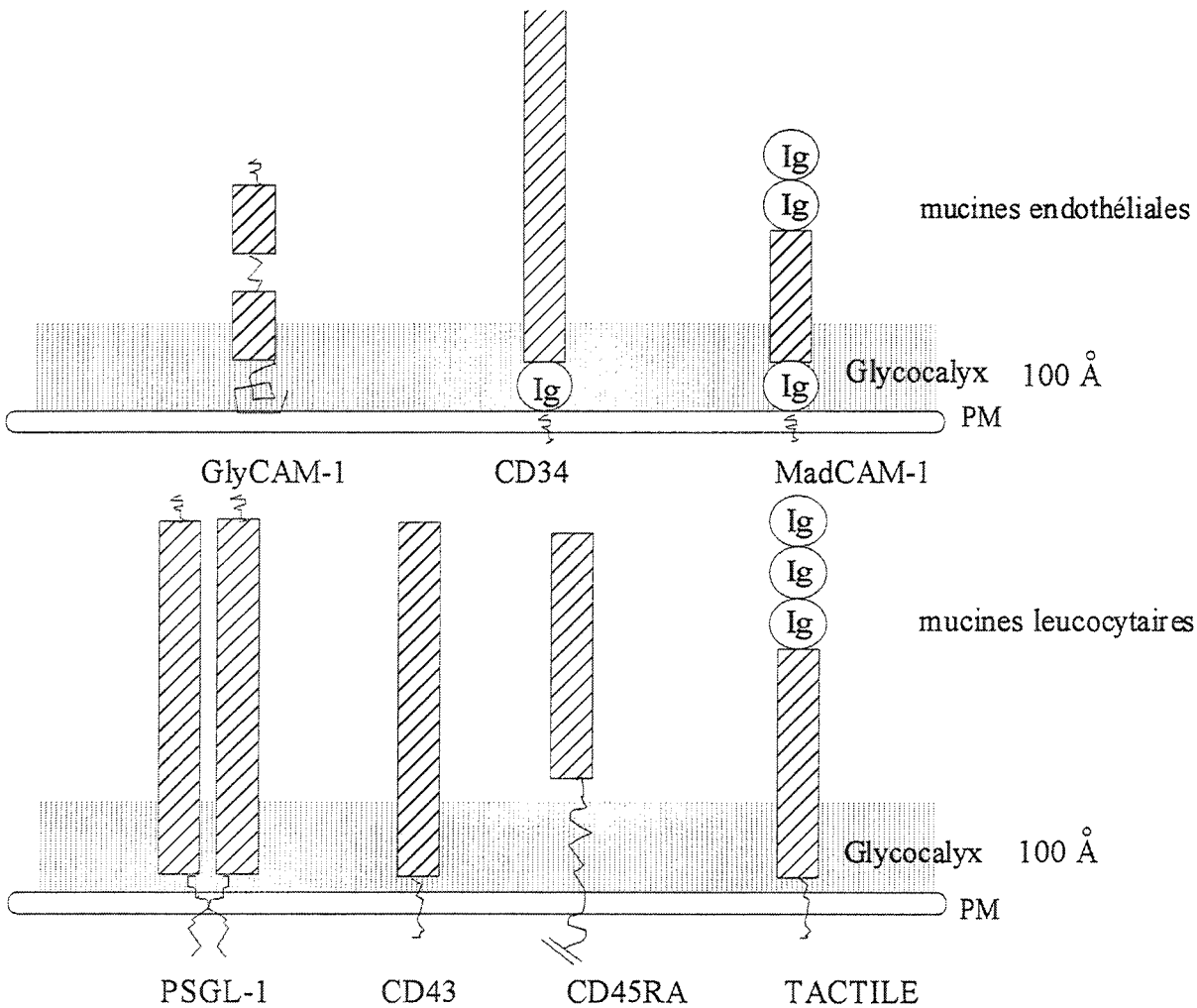


Figure 14 : Représentation schématique des mucines endothéliales et leucocytaires. Les boites hachurées représentent les domaines riches en sérine et thréonine O-glycosylables, Ⓘg : domaine immunoglobuline, PM : membrane cytoplasmique.

IV Expression des gènes de mucines humaines et pathologies.

Le profil d'expression des gènes de mucines humaines est assez bien caractérisé, que ce soit dans les tissus sains adultes (Tableau 3) (Audié et al., 1993) (Porchet et al., 1995) (Carrato et al., 1994) et fœtaux (Buisine et al., 1998b) (Buisine et al., 1999b) (Reid et al., 1997) ou dans certaines situations pathologiques où les profils d'expression peuvent varier :

- maladies infectieuses : infection à *Helicobacter pylori* et expression perturbée de *MUC5AC* et *MUC6* (Byrd et al., 1997)

mucoviscidose, infection à *Pseudomonas aeruginosa* et expression perturbée de *MUC2* (Li et al., 1997)

- maladies inflammatoires : MICI et expression de *MUC1*, *MUC3*, *MUC4* (Buisine et al., 1999a)

- situations tumorales (Lesuffleur et al., 1994) (Ho et al., 1993).

IV. 1. Expression physiologique.

Le profil d'expression reflète des particularités spécifiques à chacune des familles de mucines humaines. Ainsi les mucines sécrétées montrent une expression cellulaire spécifique alors que les mucines membranaires ont un profil plus large. En effet, les mucines membranaires sont exprimées non seulement par les cellules spécialisées dans la production de mucus (comme les cellules caliciformes) mais aussi par les cellules ciliées, les entérocytes et les hépatocytes.

MUC1 est détecté à la surface apicale des cellules de l'épithélium bronchique, mammaire, salivaire, pancréatique, prostatique, utérin et à un degré moindre au niveau des cellules épithéliales du côlon, de l'intestin grêle, de l'estomac et de la vésicule biliaire (Ho et al., 1993).

MUC4 est lui exprimé aussi bien par les tractus respiratoire, digestif et uro-génital. Seules les annexes du tube digestif ainsi que le duodénum et le jéjunum ne l'expriment pas. *MUC4* est synthétisé par les cellules spécifiques mucisécrétantes mais également par des

types cellulaires qui ne sont pas spécialisés dans la synthèse de mucines comme les entérocytes.

Cette particularité est également vraie pour *MUC3* qui est exprimé par les hépatocytes et les entérocytes. *MUC3* est largement exprimé au niveau du tube digestif et de ses annexes.

MUC7 est exprimé par les glandes séreuses salivaires et les glandes séreuses de l'épithélium trachéobronchique.

MUC2 est fortement exprimé au niveau intestinal (grêle et côlon) et plus faiblement au niveau respiratoire.

MUC5AC est détecté au niveau trachéobronchique, gastrique et cervical.

MUC6 est également retrouvé au niveau gastrique, mais aussi au niveau de l'endocol, de la vésicule biliaire et du pancréas.

MUC5B est quant à lui détecté dans les glandes muqueuses salivaires, au niveau bronchique dans les glandes muqueuses de la sous-muqueuses, dans le pancréas, dans la vésicule biliaire et dans l'endocol.

Les gènes correspondant aux mucines formant le gel offrent un profil d'expression fœtal cellulaire et tissulaire complexe, différent de celui de l'adulte (Buisine et al., 1998b) (Reid et al., 1997) (Buisine et al., 1999b).

MUC2 est exprimé dans l'intestin fœtal dès 9,5 semaines d'aménorrhée, et est associé au phénomène d'organisation spatiale de la muqueuse. Il apparaît également exprimé par de nombreux types cellulaires peu ou pas différenciés au niveau de l'appareil respiratoire fœtal.

MUC5AC apparaît quant à lui concentré dans les extrémités des bourgeons bronchiques qui se ramifient dans le mésenchyme. Ce gène pourrait donc jouer un rôle dans les phénomènes d'élongation et de dichotomie qui caractérisent l'arborisation bronchique. Il apparaît également être impliqué dans la différenciation des cellules superficielles destinées à former les glandes et les canaux glandulaires. Une coexpression dans ces cellules de *MUC5B* et *MUC5AC* évoque une concertation de fonction alors qu'ensuite, un gradient d'expression croissant de *MUC5B* accompagne la différenciation et l'enfouissement des glandes bronchiques jusque dans la sous-muqueuse. Au niveau de l'intestin fœtal, *MUC5AC* présente une expression entre 8 et 12 semaines d'aménorrhée

puis n'est plus exprimé durant la vie fœtale et adulte. Il s'exprime à nouveau dans les adénomes coliques (Buisine et al., 1996)

MUC4 a une expression cellulaire précoce au cours du développement de l'intestin primitif. Les transcrits de *MUC4* sont détectés dans l'appareil respiratoire et l'intestin dans tous les échantillons embryonnaires et fœtaux étudiés, le plus jeune étant daté à 6,5 semaines d'aménorrhée (Buisine et al., 1998b) (Buisine et al., 1999b). Tout au long du développement embryonnaire et fœtal de l'appareil respiratoire, les transcrits de *MUC4* sont constamment présents dans l'épithélium de surface de la trachée, des grosses bronches, des bronches lobaires et des bronchioles selon un gradient d'expression décroissant.

Au niveau de l'intestin grêle *MUC4* est faiblement présent entre 9 et 15 semaines d'aménorrhée dans certaines cryptes et villosités alors qu'au niveau du côlon, l'expression de *MUC4* accompagne sans discontinuité la structuration de la muqueuse en cryptes ainsi que la différenciation cellulaire de l'épithélium.

En ce qui concerne *MUC3*, il apparaît dans l'intestin primitif indifférencié en même temps que *MUC4* dès 6,5 semaines d'aménorrhée. Après 9 semaines, son expression est observée dans l'intestin grêle et le côlon, dans les cellules caliciformes et les entérocytes. Le profil d'expression persiste tout au long de la vie fœtale et chez l'adulte.

IV. 2. Expression des apomucines dans les cellules tumorales.

L'expression de *MUC1* est fortement augmentée dans de nombreux carcinomes (comme dans le carcinome de la glande mammaire, du poumon, de l'ovaire, des canaux biliaires, de l'estomac, du pancréas et de la prostate). Dans les cellules malignes, la distribution apicale de MUC1 est perdue, elle est alors exprimée non seulement par le pôle apical mais aussi par les pôles baso-latéraux (Hilkens et al., 1984).

Tout comme *MUC1*, l'expression de *MUC4* est puissamment activée dans de nombreux carcinomes. Citons les exemples du cancer du pancréas (Balague et al., 1994) (Balague et al., 1995) (Hollingsworth et al., 1994b), des voies biliaires (Vandenhaute et al., 1997), de l'endocol, de la prostate, du côlon (Lesuffleur et al., 1994), de l'estomac (Ho et al., 1995) et du poumon non à petites cellules (Nguyen et al., 1996).

Muqueuses humaines	*MUC2*	*MUC3*	*MUC4*	*MUC5B*	*MUC5AC*	*MUC6*	*MUC7*
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 : -
	G : +	G : -	G : -	G : -	G : -	G : ++/+++	G : -
	C : ++++	C : -/+	C : -	C : -	C : -	C : -	C : -
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 : +	D : -
prostate	G : -	G : -	G : ++	G : -	G : -	G : -	G : -
endocol	S : +	S : -	S : ++	S : +	S : ++	S : +	S : -
	G : +	G : -	G : ++	G : +	G : ++	G : +	G : -

Tableau 3 : Profil d'expression des gènes de mucines chez l'adulte par hybridation in situ (résultats obtenus au laboratoire).
S : épithélium de surface ; G :épithélium glandulaire ; C : cryptes ; I : invagination ; Ct : collet ; D : cellules des canaux.
Marquage : ++++ : de très forte intensité ; +++ : de forte intensité ; ++ : d'intensité modérée ; + : de faible intensité ; -/+ : quelques cellules ; - : absent.

Dans le cancer du côlon, l'expression de *MUC2* et de *MUC3* est fortement diminuée et ceci quel que soit son type histologique à l'exception du cancer mucineux (Ho et al., 1993) (Weiss et al., 1996). Une surexpression de *MUC2* ainsi qu'une expression aberrante de *MUC5AC* est détectée dans les tumeurs villeuses recto-sigmoïdiennes (Buisine et al., 1996). L'expression de *MUC5AC* semble être corrélée négativement avec le degré de dysplasie de la tumeur. *MUC5AC* est parfois également exprimé par les zones histologiquement saines à distance de la tumeur, faisant de ce gène un marqueur potentiel de récidive tumorale (Buisine et al., 1996).

IV. 3. Isoformes, glycoformes d'apomucines et tumorigénèse.

En situation pathologique, la dérégulation de l'expression des gènes de mucines n'est pas la seule modification observée. En effet, une modification au niveau post-transcriptionnelle ainsi que des modifications post-traductionnelles sont également décrites pour la synthèse de mucines en situation tumorale.

Une seule modification post-transcriptionnelle est connue à ce jour. Il s'agit de la synthèse d'une nouvelle isoforme, issue du gène codant MUC1 par épissage alternatif. Dénommée MUC1/Y, cette isoforme ne contient pas le trait essentiel qui caractérise les mucines. En effet, un épissage alternatif délète MUC1/Y de son domaine central constitué de la répétition de 20 résidus d'acides aminés (Figure 15) (Zrihan-Licht et al., 1994). Cette isoforme est présente dans les tissus tumoraux de patients atteints de cancer mammaire mais absente dans les tissus sains adjacents.

Les modifications post-traductionnelles en situations pathologiques touchent la synthèse des chaînes glycanniques. Ces modifications peuvent être observées par la technique de séparation des O-glycannes libérés par un traitement alcalin. Podolsky *et al* ont montré la disparition d'une fraction glycannique de mucines purifiées de muqueuse colique de patients atteints de rectocolite hémorragique (Podolsky and Isselbacher, 1984) (Podolsky and Isselbacher, 1983).

Durant la carcinogénèse la structure des oligosaccharides des régions périphériques change et ainsi des antigènes présents durant le développement fœtal mais absents chez l'adulte réapparaissent (Strous and Dekker, 1992).

Toutes ces modifications rendent difficile l'étude du rôle des mucines en situation pathologique.

La mucine la plus étudiée actuellement dans ce domaine est MUC1. De nombreuses fonctions lui sont attribuées. Son expression élevée dans les carcinomes et dans les lésions métastatiques suggère qu'elle ait un rôle dans la progression métastatique.

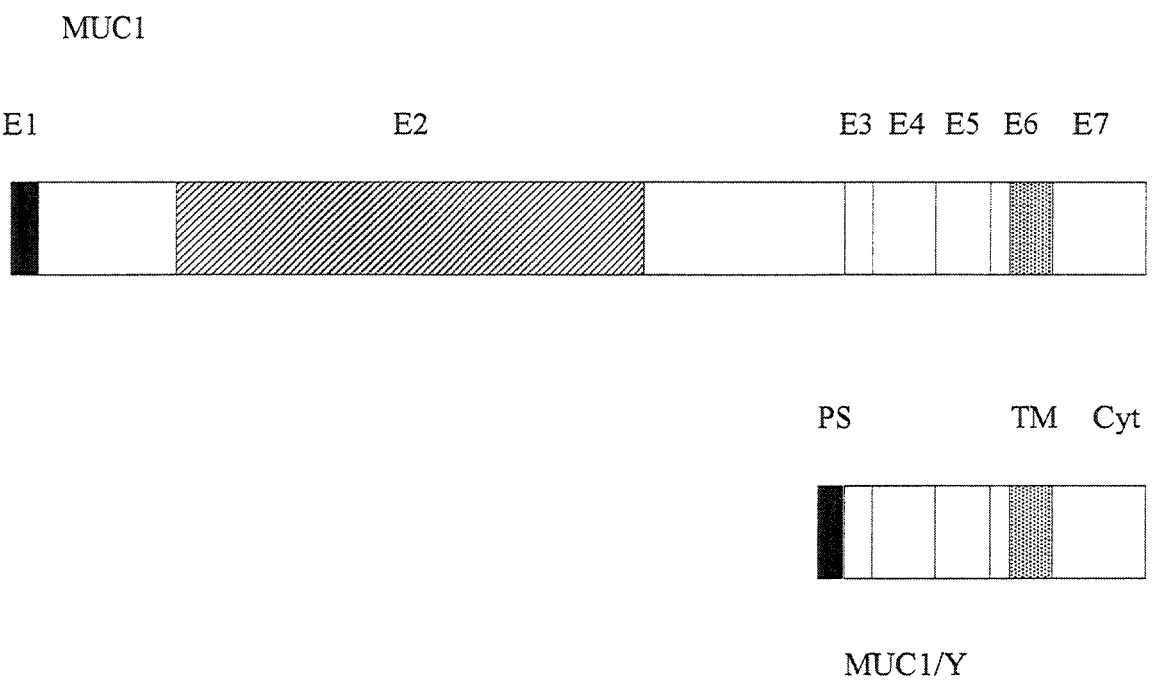


Figure 15 : Représentation schématique de MUC1 et MUC1/Y. PS : peptide signal, TM : séquence transmembranaire, Cyt : queue cytoplasmique, la boite hachurée : séquence répétitive.

V Les fonctions des mucines membranaires.

Les fonctions attribuées aux mucines membranaires s'inscrivent dans le cadre de l'adhérence cellule-cellule ou cellule-matrice extracellulaire.

V. 1. MUC1 et dissémination métastatique.

Sa structure en filament rigide dont la longueur est estimée entre 200 à 500 nm provoquerait la rupture des interactions cellule-cellule et cellule-matrice extracellulaire (Ligtenberg et al., 1992). Sa structure tridimensionnelle associée à sa distribution aux pôles baso-latéraux par les cellules cancéreuses se traduisent par la suppression de l'agrégation cellulaire. C'est pourquoi, MUC1 est souvent décrite comme une molécule d'anti-adhérence (Bramwell et al., 1986) (Fontenot et al., 1993) (Jentoft, 1990).

Cependant, MUC1 a également des fonctions d'adhésivité. En effet, dans les cellules tumorales où elle s'exprime, MUC1 présente les déterminants sialyl Lewis^x et sialyl Lewis^a (Baeckstrom et al., 1991) (Hanski et al., 1993). Ces déterminants sont les ligands pour la P- et la E-sélectine. La P- et la E-sélectine interviennent dans la dissémination métastatique (Majuri et al., 1992) (Rice and Bevilacqua, 1989). Les cellules épithéliales cancéreuses circulantes ont la capacité de se fixer aux cellules endothéliales grâce aux déterminants sialyl Lewis^x et sialyl Lewis^a. Ces cellules sont alors capables de proliférer en métastases cancéreuses.

Une étude récente montre que des sites d'interactions cryptiques, présents au niveau de l'apomucine comme au niveau des chaînes oligosaccharidiques, sont révélés par la modification de la glycosylation des cellules cancéreuses. MUC1 est alors capable de se lier aux molécules de la famille ICAM dont ICAM-1 (Regimbald et al., 1996). Les auteurs montrent que cette interaction MUC1 membranaire/ICAM-1 est inhibée par MUC1 soluble. MUC1 soluble pourrait ainsi inhiber l'interaction des cellules métastatiques avec l'endothélium ce qui aurait pour effet de diminuer le recrutement des cellules du système immunitaire vis à vis des cellules cancéreuses en migration.

De plus, il est montré que les cellules exprimant MUC1 sont résistantes aux cellules NK (natural killer) et CTL (cytotoxic T cell) (van de Wiel-van Kemenade et al., 1993). MUC1 exprime les antigènes T et Tn pouvant être reconnus par les CTL (Henningson et al., 1987). La surexpression de MUC1 au niveau des cellules cancéreuses devrait donc faciliter le recrutement des cellules CTL. Cependant, la présence de MUC1 soluble dans le sang circulant de patients présentant un cancer du sein (Hayes et al., 1985) ou du pancréas (Metzgar et al., 1984) apparaît inhiber la lyse des cellules cibles par les CTL (Barnd et al., 1989). Il semble que MUC1 soluble soit un immunosuppresseur (Fung and Longenecker, 1991).

De plus, les cellules cancéreuses exprimant MUC1 sont capables d'induire l'apoptose des CTL activées (Gimmi et al., 1996).

V. 2. MUC1 et la morphogénèse des organes épithéliaux.

Le profil d'expression spatial et temporel de *Muc1* dans l'embryon de souris coïncide avec la morphogenèse des organes épithéliaux (Braga et al., 1992) (Hilkens et al., 1992). Braga *et al* proposent que Muc1 puisse participer à la différenciation des tissus épithéliaux. Les auteurs s'appuient sur 4 observations pour émettre leur hypothèse :

- l'expression précoce de *Muc1* au cours de l'organogénèse
- la localisation apicale de Muc1 dans toutes les cellules où elle s'exprime
- la diminution des interactions cellule-cellule quand Muc1 est surexprimée
- l'interaction de son domaine cytoplasmique avec le cytosquelette (Parry et al., 1990).

Le fait qu'aucun phénotype ne soit associé à la déficience de *Muc1* chez la souris, peut s'expliquer par une redondance des fonctions par une autre mucine membranaire qui reste à identifier.

V. 3. MUC1 et le maintien/renouvellement des épithéliums.

L'un des mécanismes possibles de l'action de MUC1 pourrait se rattacher à la β caténine et la E-cadhérine. MUC1 possède un site de fixation à la β caténine (Yamamoto et

al., 1997). La β caténine est une protéine intervenant dans la formation des jonctions entre les cellules épithéliales grâce à une interaction avec la E-cadhérine (Hulsken et al., 1994).

La β caténine peut interagir également avec le produit du gène *APC* (adenomatous polyposis coli), gène suppresseur de tumeur (Hulsken et al., 1994). La protéine APC constitue un effecteur important de la voie de transmission du signal Wingless/Wnt-1 (pour revue (Peifer, 1996)). Cette voie de signalisation intracellulaire est impliquée dans le développement du système nerveux central des mammifères (Bhat et al., 1994). L'activation de cette voie de transmission aboutit à l'élévation du taux de β caténine cytoplasmique libre par inhibition de la GSK3β (Hinck et al., 1994).

Quel que soit le partenaire qui se lie à la β caténine, les complexes sont mutuellement exclusifs (Rubinfeld et al., 1995b). L'augmentation de l'expression de la protéine APC diminue le taux de β caténine cytoplasmique libre, ce qui réduit le taux de complexes β caténine/E-cadhérine et par voie de conséquence le niveau d'adhérence intercellulaire (Peifer, 1993) (Burchill, 1994).

La formation de ces complexes est régulée par la phosphorylation des domaines cytoplasmiques de chacun des partenaires par la GSK3β (glycogen synthase kinase 3β) (Rubinfeld et al., 1993a). Après phosphorylation, la β caténine est dégradée. La GSK3β est capable de phosphoryler le domaine d'interaction de MUC1 avec la β caténine (Li et al., 1998). Plus le domaine cytoplasmique de MUC1 est phosphorylé, moins elle interagit avec la β caténine (Quin et al., 1998). La concentration relative de MUC1, de la E-cadhérine, de la β caténine, de la GSK3β, et de APC semble donc importante dans le maintien de l'intégrité de l'épithélium.

Un autre aspect complique la compréhension des fonctions de MUC1. Cet aspect est la phosphorylation de son domaine cytoplasmique (Zrihan-Licht et al., 1994). MUC1, qui est capable d'interagir avec le cytosquelette (Parry et al., 1990) par l'intermédiaire de son domaine cytoplasmique, pourrait avoir des fonctions de facteur de croissance et agir dans la chaîne de cascade de transduction du signal via Grb2 et Sos (Pandey et al., 1995). L'activation de Grb2 et de Sos par MUC1 pourrait activer la voie de transduction liée à Ras (van der Geer and Hunter, 1991).

Il semble donc qu'outre les fonctions généralement dévolues aux mucines, comme la lubrification et la protection des épithéliums contre les agressions exogènes,

MUC1 soit impliquée dans la morphogenèse et l'organogenèse ainsi que dans le maintien de l'intégrité des épithéliums.

Récemment, une autre protéine membranaire phosphorylée, d'une masse de 180 kDa, a pu être co-immunoprécipitée avec MUC1 (Mockenstrum-Gardner et al., 1998). La nature de cette protéine membranaire est inconnue ainsi que les conséquences de cette interaction.

V. 4. MUC1 et le modèle de souris knock-out.

La compréhension des fonctions des mucines membranaires nécessite l'établissement de modèles d'étude. L'identification d'homologues murins permet d'aborder les fonctions des mucines dans un modèle animal, la souris. La caractérisation de Muc1 a permis d'amorcer ce travail.

En vue d'approcher ces fonctionnalités, des souris déficientes pour *Muc1* ont été générées (Spicer et al., 1995b). Malgré l'expression de *Muc1* durant l'organogenèse des tissus épithéliaux, des souris recombinantes déficientes pour *Muc1* sont obtenues. Leur fréquence et leur développement apparaissent tout à fait normaux. Aucun phénotype n'est associé à la déficience de *Muc1*. Les auteurs suggèrent que la déficience de *Muc1* puisse être compensée par la surexpression d'un autre gène de mucines. Récemment, une diminution de l'obstruction intestinale liée au mucus a pu être mis en évidence chez des souris doublement déficientes pour *Muc1* et *CF* (Parmley and Gendler, 1998). Les auteurs suggèrent que Muc1 puisse intervenir dans le largage des mucines sécrétées coliques par un mécanisme inconnu ou intervenir dans la formation du réseau qui structure le mucus colique.

V.5. Un autre modèle de mucine membranaire : la mucine de rat SMC.

Les fonctions d'adhérence et d'antiadhérence découvertes pour MUC1 sont également mises en évidence pour SMC (Carraway et al., 1992). En effet par sa sous-unité ASGP-1, SMC forme une glycoprotéine étirée implantée dans la membrane. SMC est impliquée dans le processus métastatique (Steck and Nicolson, 1983) et dans la résistance

aux cellules NK (natural killer) (Moriarty et al., 1990). Les auteurs suggèrent que sa surexpression par les cellules cancéreuses puisse masquer les antigènes ou les protéines tumorales néoformées (Figure 16).

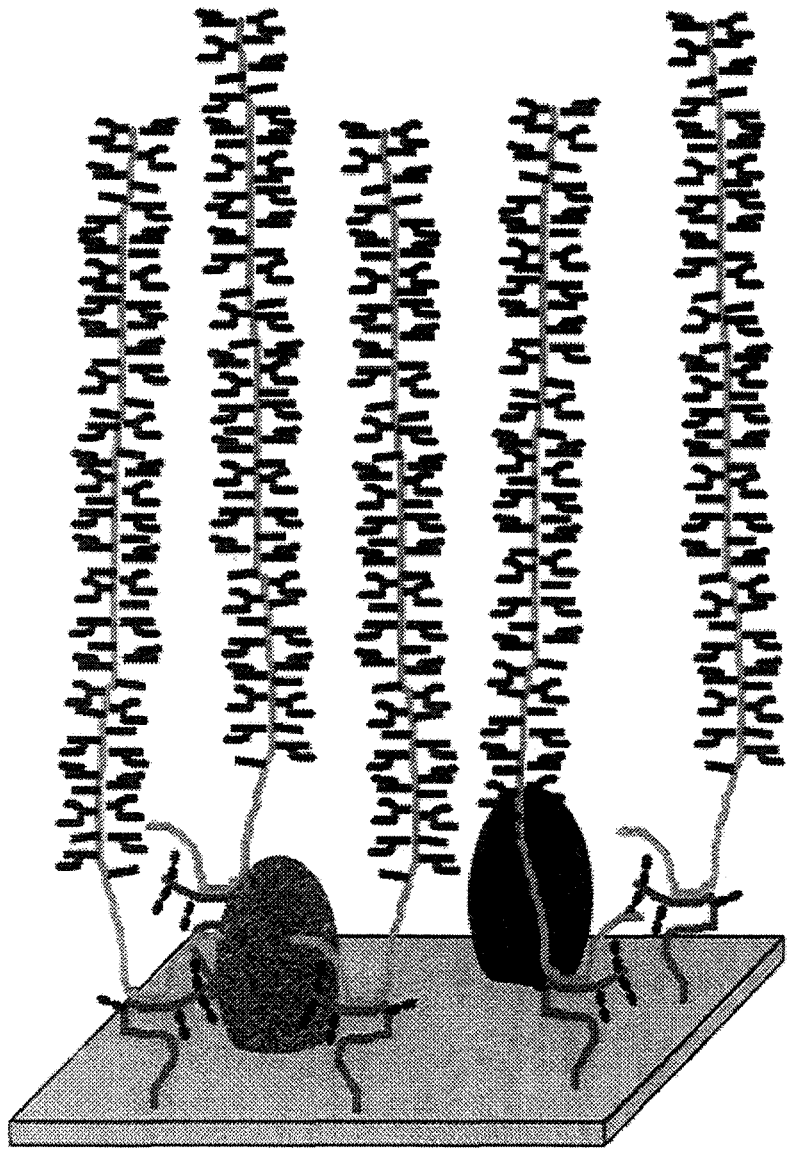


Figure 16 : Modèle de protection de la surface membranaire cellulaire résultant de surexpression de SMC par les cellules tumorales (Mc Neer et al., 1997).

La mucine de rat SMC pourrait avoir des fonctions de facteur de croissance par le biais de ses domaines de type EGF. Deux observations importantes laissent supposer que SMC forme un complexe avec une protéine membranaire $p185^{neu}$ (Figure 17) (Carraway et

al., 1997) (Carraway and Cantley, 1994). L'ASGP-2 et p185^{neu} sont co-immunoprécipitées à partir d'un extrait de protéines membranaires issu de la lignée tumorale 13762. Après transfection de cellules d'insecte par l'ASGP-2 et p185^{neu}, un complexe des deux molécules se forme et est sécrété. Le complexe ne se forme que lorsque les 2 protéines sont exprimées dans la même cellule.

Des expériences à l'aide de mutants de délétion de l'ASGP-2 montrent que le domaine intervenant dans la formation de ce complexe est le domaine EGF1. Les auteurs proposent que l'ASGP-2 puisse intervenir dans le processus de perte du contrôle de croissance des cellules malignes.

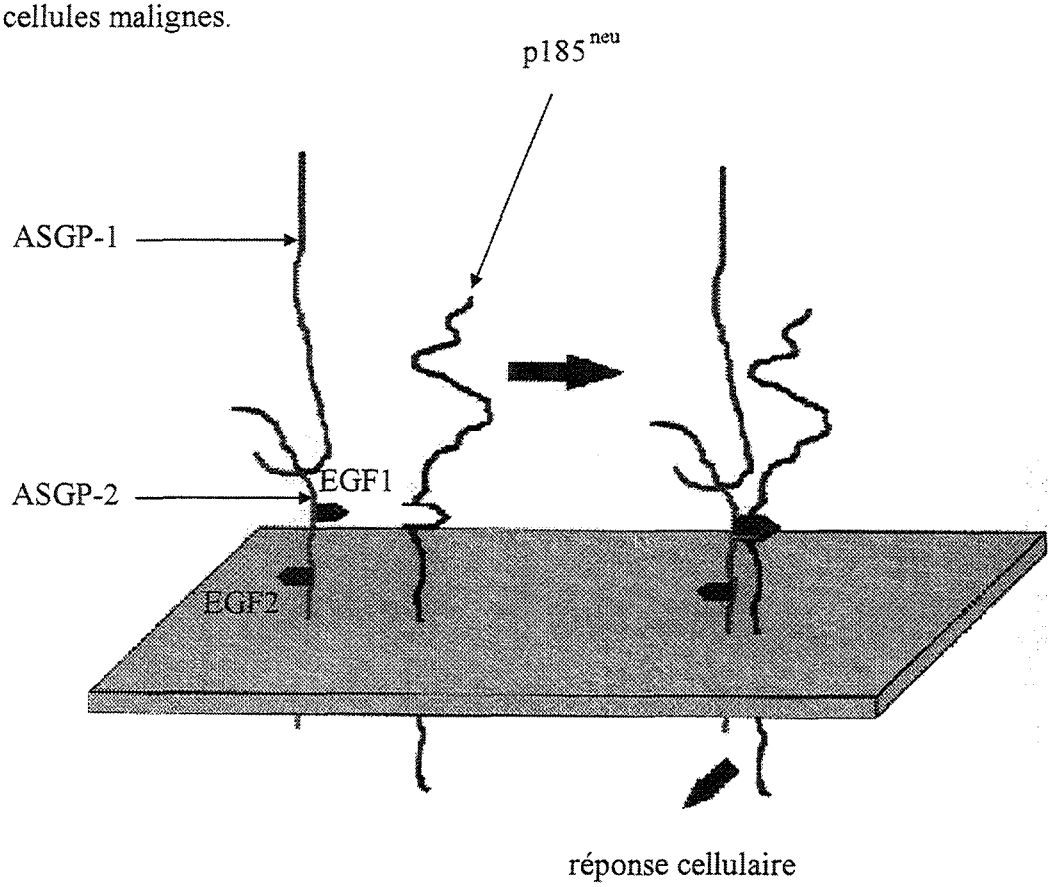


Figure 17 : Modèle d'interaction de SMC avec p185^{neu} (Mc Neer et al., 1997).

p185^{neu} est l'homologue de rat du récepteur de type tyrosine kinase ErbB2 humain. SMC est le premier ligand spécifique décrit pour p185^{neu}. Leur interaction se traduit par la phosphorylation et donc l'activation du dimère. Les effets biologiques qui peuvent en découler ne sont pas encore connus.

Les domaines N-terminaux et le domaine central de MUC4.

I. Situation du projet de recherche au début de notre travail.

Quand a débuté notre travail de thèse sur la caractérisation des domaines composant l'extrémité 3' de *MUC4* et l'approche des relations structure-fonction, une autre étude avait débuté en parallèle dans le laboratoire. Le travail de thèse de Séverine Nollet consistait en effet en l'obtention de clones génomiques de *MUC4* et en l'étude de l'extrémité 5' de *MUC4*, en l'étude des variations polymorphiques liées à la séquence répétée en tandem de 48 pb, ainsi qu'en la caractérisation des séquences impliquées dans la régulation de l'expression du gène *MUC4*. Nous disposions pour mener à bien ces deux sujets de thèse d'un clone d'ADNc, JER64, isolé par criblage d'une banque d'expression construite à partir de l'ARNm extrait d'une muqueuse de trachée humaine (Porchet et al., 1991). JER64 est constituée de 39 fois la répétition du motif de 48 pb. Le gène reconnu par le clone JER64 utilisé comme sonde est localisé sur le chromosome 3 dans la région q29 (Gross et al., 1992). Il a reçu le nom de *MUC4* dans la nomenclature internationale.

Les deux sujets de thèse étant liés, le début des travaux s'est fait dans une collaboration étroite. Les deux études ont nécessité d'une part le criblage avec la sonde JER64 de deux banques de gènes, l'une construite en vecteur cosmidique pWE15 et l'autre construite en vecteur phagique, et d'autre part la construction et le criblage avec la sonde JER64 d'une banque d'ADNc construite à partir de l'ARNm extrait d'une muqueuse colique.

II. Résultats acquis par Séverine Nollet.

Les résultats de l'étude des domaines 5' et du domaine central de *MUC4* ont fait l'objet de la publication suivante :

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

Human mucin gene *MUC4*: organization of its 5′-region and polymorphism of its central tandem repeat array

Séverine NOLLET*[1], Nicolas MONIAUX*[1], Jacques MAURY*, Danièle PETITPREZ*, Pierre DEGAND*†, Anne LAINE*, Nicole PORCHET*† and Jean-Pierre AUBERT*†[2]

*INSERM Unité 377, Place de Verdun, 59045 Lille Cedex, France, and †Laboratoire de Biochimie et de Biologie Moléculaire, l'Hôpital C. Huriez, Centre Hospitalier Régional et Universitaire, 59037 Lille Cedex, France

In a previous study we isolated a partial cDNA with a tandem repeat of 48 bp, which allowed us to map a novel human mucin gene named *MUC4* to chromosome 3q29. Here we report the organization and sequence of the 5′-region and its junction with the tandem repeat array of *MUC4*. Analysis of three overlapping genomic clones allowed us to obtain a partial restriction map of *MUC4* and to locate the complete 48 bp tandem repeat domain on a *PstI*/*EcoRI* genomic fragment that exhibits a very large variation in number of tandem repeats (7–19 kb). cDNA clonal extension allowed us to obtain the entire 5′ coding region of *MUC4*. Exon 1 consists of a 5′ untranslated region and an 82 bp fragment encoding the signal peptide. This latter shows a high degree of similarity to the signal peptide of another apomucin, ASGP-1. Exon 2 is extremely large and contains a unique sequence that is followed by the whole tandem repeat domain. It encodes only one cysteine residue, making *MUC4* different from mucin genes belonging to the 11p15.5 family. Moreover, an intron downstream from the tandem repeat array consists mainly of a 15 bp tandem repeat that exhibits a polymorphism in having a variable number of tandem repeats.

INTRODUCTION

Mucins constitute a complex family of glycoconjugates, characterized by a large amount of O-glycans linked to the peptide backbone (reviewed in [1–3]). These O-linked glycans usually account for more than 50 % of the molecular mass of the mucin. The potential O-glycosylation sites (threonine and serine residues) are located mainly in tandemly repeated amino acid sequences. Mucins can be subdivided into secreted mucins and 'mucin-like' membrane-bound O-glycoproteins. So far nine human mucin genes, designated *MUC1*, *2*, *3*, *4*, *5AC*, *5B*, *6*, *7* and *8* in the international nomenclature, have been identified (reviewed in [1]; *MUC8* reviewed in [4]). *MUC1* is located on 1q21-24, *MUC3* on 7q22, *MUC4* on 3q29, *MUC7* on 4q13-21 and *MUC8* on 12q24.3. *MUC2*, *MUC5AC*, *MUC5B* and *MUC6* are clustered on 11p15.5. Several partial cDNA species containing tandem repeats (TRs) have been isolated in our laboratory from a human tracheobronchial cDNA library and correspond to the three distinct genes *MUC4*, *MUC5AC* and *MUC5B* [5–7]. The first partial cDNA from *MUC4*, called JER64, contains 39 repeats of 48 nearly identical nucleotides that encode a Thr/Ser-rich (consensus sequence ATPLPVTDTSSASTGH) potentially glycosylated peptide mucin domain [6]. *MUC4* displays genetic polymorphism, which suggests that it is related to a variable number of tandem repeats (VNTR) specific to each individual [8].

The gene-expression pattern of *MUC4* has been determined by using Northern blotting and/or hybridization *in situ*. *MUC4* mRNA species are expressed in colon, stomach, cervix and lung [9,10] but are not detected in normal pancreas, gall bladder and breast tissues. The expression of *MUC4* mRNA species is ubiquitous in epithelial cells, where it is equally expressed in goblet and ciliated cells of the trachea and bronchi. Similarly, in the intestinal mucosa *MUC4* is expressed in both goblet and absorptive cells.

Abnormal expression of *MUC4* has been reported in various carcinomas. For example, increased expression of *MUC4* mRNA species is observed in pancreatic carcinoma and cancer cell lines [11,12] and mammary carcinomas [13]. Moreover, the frequent occurrence of increased *MUC4* transcripts in a variety of non-small-cell lung cancers [14] and in colon carcinomas [15] indicates that the overexpression of this gene might have an important, albeit undefined, role in tumour biology. Therefore it is of considerable interest to define more fully the genomic structure of *MUC4*.

Here we report the genomic organization of the 5′-region and the whole TR array of the human mucin gene *MUC4*. The nature of the VNTR polymorphism of the 48 bp TR region is also described.

EXPERIMENTAL

Screening of genomic libraries

The probe designated JER64 used in this study corresponds to the *MUC4* cDNA probe (1.83 kb) and contains 39 identical 48 bp TRs [6]. Human genomic DNA fragments (of approx.

Abbreviations used: endo H, endoglycosidase H; RACE, rapid amplification of cDNA ends; RT–PCR, reverse transcription PCR; TR, tandem repeat; VNTR, variable number of tandem repeats.

[1] The first two authors contributed equally to this work and should therefore be considered as equal first authors.

[2] To whom correspondence should be addressed at INSERM U-377 (e-mail jpa@lille.inserm.fr).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers AJ000281 and AJ000282.

20 kb) were obtained by partial digestion with *Sau*3A and were cloned into *Bam*HI sites of a λEMBL4 phage vector. Screening of the library was performed with the JER64 probe. One positive clone (ANT55) with an insert of approx. 15.5 kb was obtained.

To isolate larger genomic clones of *MUC4*, a human placenta genomic cosmid DNA library in pWE15 (Stratagene) was screened with the JER64 probe. Two positive clones (LEA2 and LEA47), each containing inserts of approx. 45 kb, were obtained and analysed. The same cosmid genomic library was also screened with the insert of the RAC3 clone described below. One positive clone with an insert of approx. 40 kb (LEA51) was isolated and studied.

Restriction mapping of cosmids

The restriction mapping strategy of Wahl et al. [16] was modified slightly, as described previously [17].

Study of the 48 bp tandem repeat region

To evaluate the number of 48 bp repeats, we first cut the LEA2 cosmid with *Pst*I and *Eco*RI at positions that flank the region containing these repeats and isolated the fragment. Complete or partial digestion with *Dde*I was achieved with 10 units/μg of DNA for 4 h or 0.14 unit/μg of DNA for 1 h respectively. Southern blot analysis was conducted with the JER64 probe, as described previously [6].

RNA extraction

Total RNA was extracted from normal human colon mucosa with the guanidine isothiocyanate/CsCl method [18].

cDNA library preparation and screening

Total RNA from human colon mucosa was prepared and used as a template for cDNA synthesis. All details of double-stranded cDNA synthesis and cloning into λgt11 vector were as described by the commercial supplier (Amersham).

Screening of the cDNA library was performed with the JER64 probe. One single positive clone, JER103, was isolated and sequenced.

Cloning into pKS

The fragments of interest from phage or cosmid clones were subcloned into pBluescript KS(+) vector from Stratagene.

5′ Rapid amplification of cDNA ends (RACE) procedures

The 5′/3′ RACE kit (Boehringer Mannheim) was used to synthesize first-strand cDNA species from total human colon RNA (2 μg) with specific primers for *MUC4* (NAU124, NAU155, NAU168 and NAU287; their locations are given below). Terminal transferase was then used to add a poly(dA) tail to the 3′ end of the cDNA. RACE–PCR experiments were performed in 50 μl reaction volumes containing 5 μl of 10 × buffer (100 mM Tris/HCl/15 mM $MgCl_2$/500 mM KCl, pH 8.3), 5 μl of 10 mM deoxynucleoside triphosphates, 5 μl of poly(dA)-tailed cDNA, 12.5 pmol of each primer [a set of four specific primers, NAU138, NAU156, NAU169 and NAU 288, and an oligo(dT) anchor primer were used], and 2 units of *Taq* DNA polymerase (Boehringer Mannheim). After being overlaid with 60 μl of mineral oil (Sigma), the mixture was denatured at 94 °C for 2 min followed by 30 cycles at 94 °C for 1 min, 60 °C for 1 min and finally 72 °C for 2 min. The elongation step was extended for an additional 15 min period. A 1 μl sample of the primary amplification product was further amplified by a second PCR reaction with a nested specific primer of *MUC4* (NAU138, NAU156, NAU204, NAU289 and NAU327) and the PCR anchor primer. The thermal cycling protocol used was the same as for the primary RACE amplification step, except that the annealing was performed at 62 °C. PCR experiments were performed with a Perkin–Elmer thermal cycler model 480.

Various antisense primers were used to extend the 5′ end sequence: NAU124 (nt 565–588) (see Figure 4), NAU138 (nt 538–564), NAU155 (nt 308–325), NAU156 (nt 275–301), NAU168 (nt 105–124), NAU169 (nt 81–100), NAU204 (nt 51–77), NAU287 (nt −15 to 6), NAU288 (nt −41 to −21), NAU289 (nt −76 to −56) and NAU327 (nt −130 to −109).

Reverse transcription and amplification

Human tracheal poly(A)$^+$ RNA (0.5 μg) (Clontech) and 1 μg of total RNA extracted from human colon mucosa were reverse transcribed with the 1st-STRAND® cDNA synthesis kit (Clontech) with random primers in accordance with the manufacturer's instructions. First-strand cDNA (8 μl) was amplified by PCR with various primers: NAU162 (sense) (nt 2–19), NAU138 (antisense) (nt 538–564), NAU370 (sense) (nt −71 to −48), NAU224 (antisense) (nt 955–974), NAU174 (sense) (nt 3015–3035) and NAU99 (antisense) (nt 3093–3109). The thermal cycling protocol was the same as that described above.

Cloning of amplification products

RACE–PCR and reverse transcription PCR (RT–PCR) products were separated by electrophoresis, excised and purified with Preps DNA purification resin (Promega), and finally cloned into pGEMT (Promega) or pCR2.1 (Invitrogen) vector.

Plasmid DNA purification

The Wizard® minipreps DNA purification system (Promega) was used in accordance with the manufacturer's instructions.

DNA sequencing

Clones were sequenced on both strands by the dideoxynucleotide chain termination method, by using [α-^{35}S]dATP with Sequenase version 2.0 (U.S. Biochemical Corp.) and synthetic oligonucleotides corresponding to the T7 and T3 primers of the pKS plasmid, and to the T7 and −40 primers of the pGEMT or pCR2.1 vector. Part of the sequence was determined by primer walking with primers specific to *MUC4*. The locations of these specific primers on the cDNA are indicated above (in the section on 5′ RACE and reverse transcription procedures) except that of NAU103 (5′-GTAATGCGAATGCACCAGTG-3′, antisense) located within an intron. We performed DNA sequencing directly on cosmids, as described previously [17]. The inner regions of some clones were sequenced with exonuclease III-deleted clones. Sequencing reaction mixtures were subjected to electrophoresis on 6% (w/v) polyacrylamide gel (Sequagel-6®; National Diagnostics). Nucleic acid and protein sequence results were analysed with PC/GENE Software.

Transcription and translation assays *in vitro*

RT370-224 is an RT–PCR product (NAU370/NAU224) containing the cDNA coding a peptide with a predicted size of 34 kDa. This peptide comprises the 27-residue N-terminal signal sequence and two potential *N*-glycosylation sites. The corre-

sponding cDNA was cloned into pCR2.1 vector under the control of the T7 promoter. Transcription and translation experiments *in vitro* were performed with the TNT Coupled Reticulocyte Lysate System (Promega) in accordance with the manufacturer's instructions. The amino acid mixture lacking methionine, supplemented with [^{35}S]methionine (approx. 1000 Ci/mmol; Amersham), was used. To detect a signal peptide, canine pancreatic microsomal membranes (Promega) were added to the TNT Coupled Reticulocyte Lysate System for co-translational processing of translation products. To determine the extent of N-glycosylation of the translation products, ^{35}S-labelled peptide was incubated overnight at 37 °C with 10 m-units of endoglycosidase H (endo H; Boehringer Mannheim) in a buffer containing 0.02% SDS and 0.1 M sodium citrate, pH 5.5. Translation products were analysed by SDS/PAGE in slab gels with a 10–30% (w/v) polyacrylamide gradient. Rainbow™ coloured protein molecular mass markers (Amersham) varying in size from 14.3 to 220 kDa were used to determine the molecular mass.

Southern blot analysis

Human genomic DNA from 18 healthy volunteers was digested with *Pst*I and *Eco*RI restriction endonucleases; 12 out of the 18 DNA species were digested with *Dde*I (complete digestion). Fragments were separated by electrophoresis and transferred to nylon N$^+$ membrane (Amersham). They were hybridized with three different probes and washed as described previously [6].

RESULTS

MUC4 genomic clones

We isolated three genomic clones from *MUC4* with the JER64 probe. One of these, designated ANT55, was isolated from a phage library and the other two, LEA2 and LEA47, from a cosmid library. Alignment and partial restriction maps of these three clones are shown in Figure 1. The two cosmid clones overlap ANT55, which is 15.5 kb in length. A 2.5 kb *Eco*RI fragment located at the 3′ end of ANT55 was subcloned into pKS vector. This fragment, designated HOR1, was completely sequenced. It consists of 109 repetitions of very similar 15 bp units (consensus sequence GGTGTGGAAGGTATG) and two unique flanking sequences of 795 and 62 bp at the 5′ and 3′ ends respectively. The 15 bp TR domain of HOR1 has been submitted to the EMBL Data Bank with accession number AJ000282. The oligonucleotide primer NAU103 situated on the 5′ end of HOR1 (its position is shown in Figure 1) was chosen and used to perform direct sequencing on ANT55 clone. The 48 bp TR domain of JER64 was located 85 bp upstream of the *Eco*RI restriction site. The direction of the transcription could then be determined as shown in Figure 1. The primer NAU103 was also used to perform direct sequencing on LEA2 and LEA47 cosmid clones. The same sequence was determined. The region of HOR1 containing only the 15 bp repeats did not hybridize to RNA species from human colon and bronchus mucosae by Northern blot or *in situ* hybridization analyses (results not shown).

The 48 bp tandem repeat region

*Pst*I/*Eco*RI digestion allowed us to isolate the smallest fragment containing the whole 48 bp repetitive domain. In the ANT55 insert this fragment was 11 kb in length; it was approx. 20 kb in length in the LEA2 and LEA47 inserts. Only a few usual restriction enzymes were able to cut within this 48 bp repetitive sequence. Among these, the *Rsa*I restriction endonuclease cleaved the LEA2 insert into four distinct bands (approx. 8, 5, 3.2 and 2 kb), which hybridized with the JER64 probe. Moreover, a combined enzyme digestion *Pst*I/*Eco*RI/*Rsa*I gave the same bands. So the sum of the sizes of the subfragments included into the *Pst*I/*Eco*RI fragment of the LEA2 insert allowed us to calculate a more precise size for the whole *Pst*I/*Eco*RI fragment, which was predicted to be approx. 18.2 kb. This region was completely devoid of all conventional endonuclease restriction sites and could thus not be cloned into a plasmid vector, thereby making direct sequence determination impossible. Some restriction sites that are not present in the multicloning site of the vector exist in this region (*Dde*I, *Fok*I, *Hph*I, *Ksp*632I, *Mae*III, *Mbo*II, *Mnl*I, *Sfa*NI) but they cut it into a range of very small fragments (results not shown). However, because one *Dde*I restriction site (CTNAG) exists in most of the JER64 48 bp repeats (in 36 out of 39 units), *Dde*I was used to cut this region. Figure 2 shows the 18.2 kb *Pst*I/*Eco*RI fragment of LEA2 digested to completion with the restriction enzyme *Dde*I. Only two ethidium bromide-stained fragments (48 and 96 bp) were observed, as expected from the JER64 nucleotide sequence. These two bands hybridized with the JER64 probe. The presence of a 96 bp fragment is due to two adjacent units lacking one *Dde*I site. The fact that no additional band was observed strongly suggests that the TR region of the LEA2 insert is not interrupted by an additional unique sequence. Partial *Dde*I digest of the 18.2 kb *Pst*I/*Eco*RI fragment of LEA2 (Figure 2) showed a ladder where individual bands consisted of 48 bp multiples. Beyond approx. 20 bands, a smear was observed and precise estimation of the number of units was difficult. The same observation was made with ANT55 and LEA47. Thus the TR arrays in the phage clone and in the cosmid clones appear to contain approx. 208 and 380 uninterrupted individual TR units respectively. Southern blot analysis of the *Dde*I fragments from DNA species isolated from lymphocytes of 12 individuals gave the same result: only fragments of 48 and 96 bp hybridized to the JER64 probe (results not shown).

JER103 cDNA clone

The JER64 cDNA probe was also used to screen a human colon mucosa cDNA library. One positive clone was obtained and designated JER103 (Figure 3). The JER103 insert consists of 2653 bp, of which 200 bp belong to the 48 bp TR recognized by JER64. A *Pst*I restriction site was found 263 bp upstream of the TR (labelled 3 in Figure 1). The choice of the reading frame was directed by that of the JER64 cDNA clone. The nucleotide sequence of JER103 revealed the presence of a unique fragment encoding a threonine/serine-rich peptide. This sequence was found immediately upstream of the 48 bp repeat array on the cosmid inserts, indicating that both of these unique and repetitive Thr/Ser-rich regions are encoded by a single exon. The JER103 insert sequence is identical with that found in LEA2 and LEA47 cosmid and ANT55 phage clones, this last being 439 bp shorter.

Extension of the JER103 5′ end sequence

Two antisense primers, NAU124 and NAU138, were used to extend the sequence by 5′ RACE on human colon mucosa total RNA. One 358 bp cDNA fragment was obtained, cloned and designated RAC1 (Figure 3). Its 3′-end sequence overlaps the 5′ end sequence of JER103, indicating that it contains the JER103 5′-extended end. This sequence is identical with that found in LEA2 and LEA47 cosmid clones immediately upstream of the sequence corresponding to that of the JER103 clone. This experiment was followed by a second 5′ RACE procedure with two novel primers deduced from the sequence of the RAC1 clone

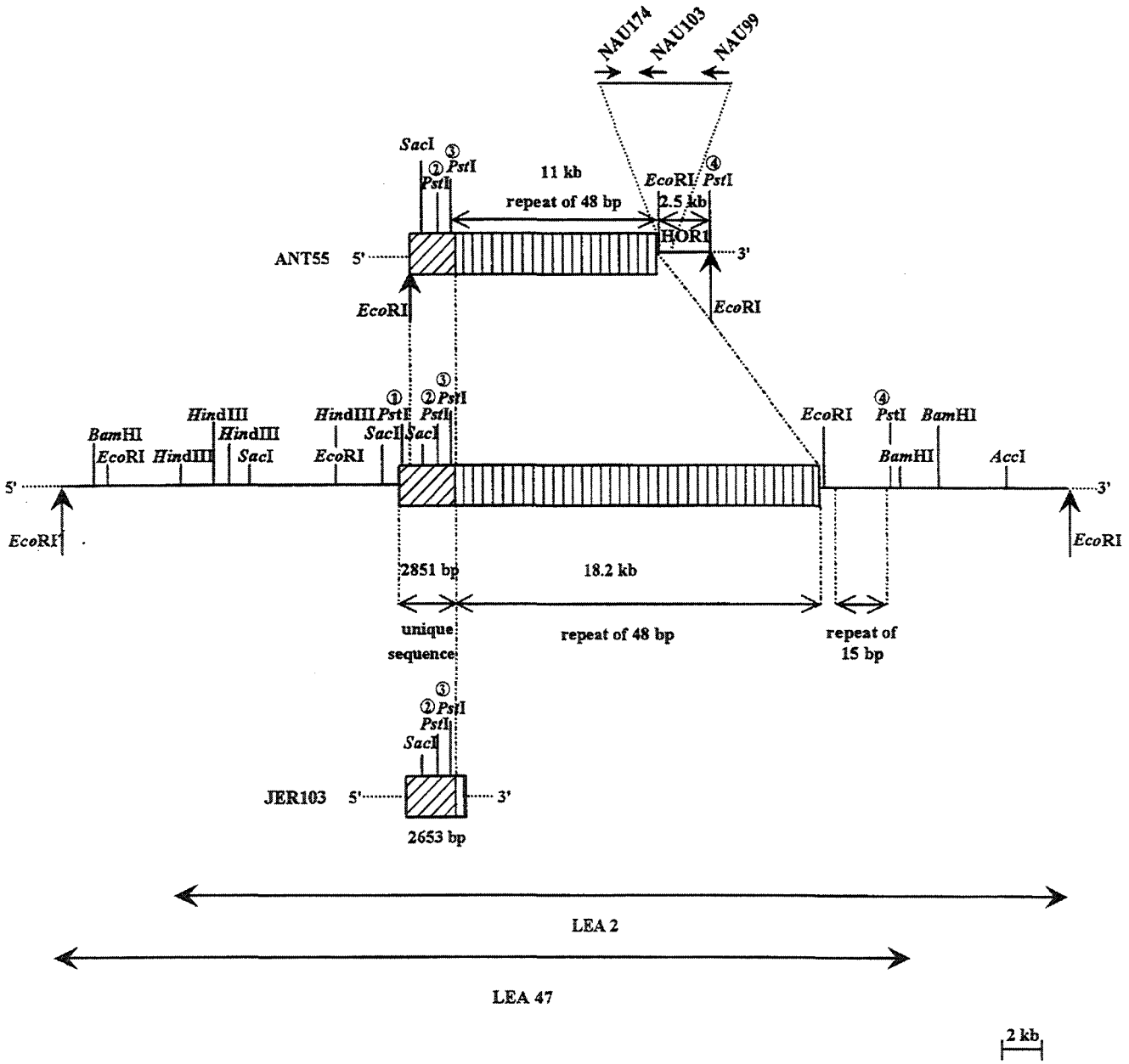


Figure 1 Partial restriction map of three genomic clones (ANT55, LEA2 and LEA47) and one cDNA clone (JER103) from *MUC4*

The vertically hatched box represents the 48 bp TR array. The diagonally hatched box represents the unique sequence encoding a threonine/serine-rich region. The *PstI*/*EcoRI* fragment that delimited the 48 bp TR region is 11 kb in length for the phage clone and approx. 18.2 kb in length for the cosmid clones. Primers and their directions are indicated (not to scale) by horizontal arrows and their NAU numbers (their locations are given in the Experimental section). All genomic inserts are flanked by *EcoRI* sites of the vectors indicated by vertical arrows. *PstI* restriction sites used in polymorphism analysis are numbered in a circle.

(NAU155 and NAU156). This produced a 300 bp fragment (RAC2). The sequence of the 3′ region of the RAC2 clone is the same as that of a 219 bp stretch found in the three genomic clones ANT55, LEA2 and LEA47. However, the first 81 bp of the RAC2 clone are not found within these genomic clones. Thus an intronic region was suspected. Another 5′ RACE procedure with the oligonucleotides NAU168, NAU169 and NAU204 (chosen in the sequence of the RAC2 clone) produced a fragment of 183 bp (RAC3). Two other RAC clones, RAC4 (140 bp) and RAC5 (352 bp), were then obtained by using four novel oligonucleotides (NAU287, NAU288, NAU289 and NAU327). These five RAC clones allowed us to extend the 5′ end sequence of JER103 over 942 bp.

The RAC3 probe does not hybridize to Southern blots from LEA2 and LEA47 DNA species, showing that the 5′ sequence of 542 bp obtained by compiling the sequences of clones RAC3, RAC4 and RAC5 is not located on these cosmids. This indicates that this coding sequence is situated at least 15 kb upstream of the RAC1–2 sequence and that the 15 kb fragment constituting the 5′ end of LEA47 corresponds to a large intron (Figure 3).

To obtain the 5′-region of *MUC4*, the pWE15 cosmid genomic library was screened with probe RAC3. One novel clone

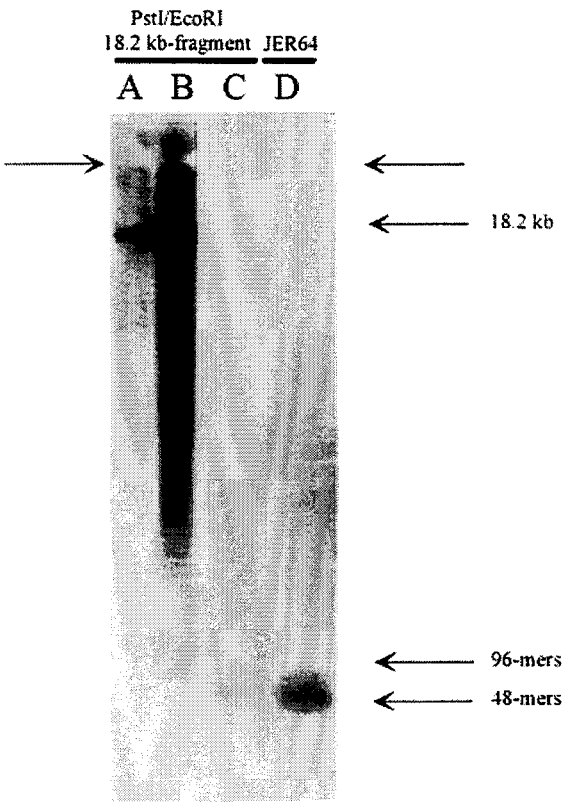


Figure 2 Characterization of the 48 bp repeat in the LEA2 clone

After electrophoretic separation in an agarose gel, blot analysis was conducted with the JER64 probe. Lanes A–C, LEA2 was cut at *PstI* and *EcoRI* sites flanking the 48 bp TR region: lane A, *PstI/EcoRI* fragment; lane B, *PstI/EcoRI* fragment partly digested for 1 h with *DdeI* (0.14 unit/μg of DNA); lane C, *PstI/EcoRI* fragment totally digested with *DdeI* (10 units/μg of DNA). Lane D, JER64 insert totally digested with *DdeI*.

designated LEA51 was isolated and its partial restriction map revealed one 2.2 kb *EcoRI* fragment situated at the 3′ end of this clone. This fragment was subcloned in pKS vector and partly sequenced. Its 3′ part consists of an intronic region of 400 bp. The 542 bp upstream of this intronic region showed 100% sequence similarity to the 542 bp compiled cDNA sequences of RAC3–5.

The exon–intron boundaries were determined by comparing the genomic and cDNA nucleotide sequences. Splice acceptor and donor sequences agree with the 'GT–AG' rule proposed by Mount [19].This intron is class 1 because it interrupts the coding sequence between the first and second bases of the codon [20].

RT–PCRs with the primers NAU162 and NAU138 were performed starting from human RNA species from tracheobronchial and colon mucosae; 563 bp amplification products were generated and cloned. Their nucleotide sequence showed 100% similarity to the corresponding sequences obtained by compiling RAC2 and RAC1 sequences.

Analysis of the nucleotide and deduced amino acid sequences of the MUC4 5′ end cDNA

The compiled nucleotide sequences of the different cDNA clones obtained allowed us to establish the whole coding sequence of the MUC4 cDNA 5′ part and its junction with the expansive 48 bp TR region. The 460 bp 5′ untranslated region is followed by a region encoding an open reading frame of 978 residues (Figure 4). This open reading frame is continuous with that of the 16-residue TR.

A methionine residue starting at nt 1 is contained within the optimal context for initiation of translation, GCCGCAGCCATGA, as described by Kozak [21], except for the −4 and +4 positions, which are A instead of C and G respectively. The Kyte–Doolittle [22] hydropathy plot of the first 200 residues of the deduced sequence showed that the initial 27 residues encoded by exon 1 are very hydrophobic, suggesting that these comprise the putative signal peptide. Three potential cleavage sites exist between amino acids 23 and 24, 27 and 28, and 28 and 29. To confirm that this peptide can act as a signal sequence, the RT370-224 clone (1045 bp) [see its position in Figures 3(C) and 4] was analysed by transcription and translation *in vitro* in the presence or absence of dog pancreatic microsomes (Figure 5). Lane 1 (in the presence of microsomes) shows three bands, where the upper is stronger and corresponds to a translated product with an apparent molecular mass of 43 kDa. Lane 2 (after treatment with endo H) shows two bands, where the smaller (corresponding to a 39 kDa product) migrates with the same mobility as the third band before treatment with endo H and the weaker like the second. The treatment with endo H in lane 2 was probably incomplete. The upper band in lane 1 is probably due to two *N*-glycans, the second to one *N*-glycan and the very weak third to the non-glycosylated one. In the absence of microsomes (lane 3), the translated product (324 residues) had an apparent molecular mass of 40 kDa. Thus the translated product of lane 2 is smaller than the 40 kDa peptide formed in the absence of microsomes, indicating that it has been cleaved as a result of translation in the presence of the membranes containing signal peptidase. Searching the GenBank database, we noticed that a high degree of similarity exists between signal peptides of MUC4 and rat ASGP-1 [23], as well as between MUC1 [24] and mouse Muc1 [25], and MUC2 [26] and rat Muc2 [27] (Figure 6A). A 62% similarity between rat *ASGP-1* and *MUC4* was observed at the nucleotide level, whereas a 59% similarity was seen at the protein level. Similarity between the two signal peptides is particularly striking when considering their respective C-terminal regions, where 12 residues out of 15 are perfectly conserved (80% similarity) (Figure 6B).

The region from nt 83 to nt 2934, together with the TR array, forms a single large exon. This region is found to encode a unique 951-residue sequence typical of apomucins, comprising 21.5% threonine and 19% serine. The sequence TXXP, considered to be a major O-glycosylation site [28], is repeated on 21 occasions. This region is also proline-rich (7.4%). Three potential N-glycosylation sites are also found in this sequence at positions 235, 260 and 622 (it is likely that at least two were glycosylated *in vitro*) and only one cysteine residue at position 256. The sequence for this unique Thr/Ser-rich region bears no significant similarity to the 48 bp TR of JER64 or to any other sequences from the GenBank database. Closer examination reveals that this sequence contains three subregions that share a high degree of similarity. These three repeated units begin at residue 43. The first and the second repeats are both 126 residues in length, whereas the third is 130 residues long. The amino acid composition of each subregion contains 24% threonine and 20% serine and is typical of apomucins. A tetrabasic amino acid RKRR site is also found at position 822.

Determination of intron–exon boundaries on the 3′ end of the 48 bp TR region

To determine the 3′ end of the central exon, RT–PCR experiments were performed with NAU174 and NAU99 as primers (see their positions in Figure 1), starting from several sources of poly(A)$^+$

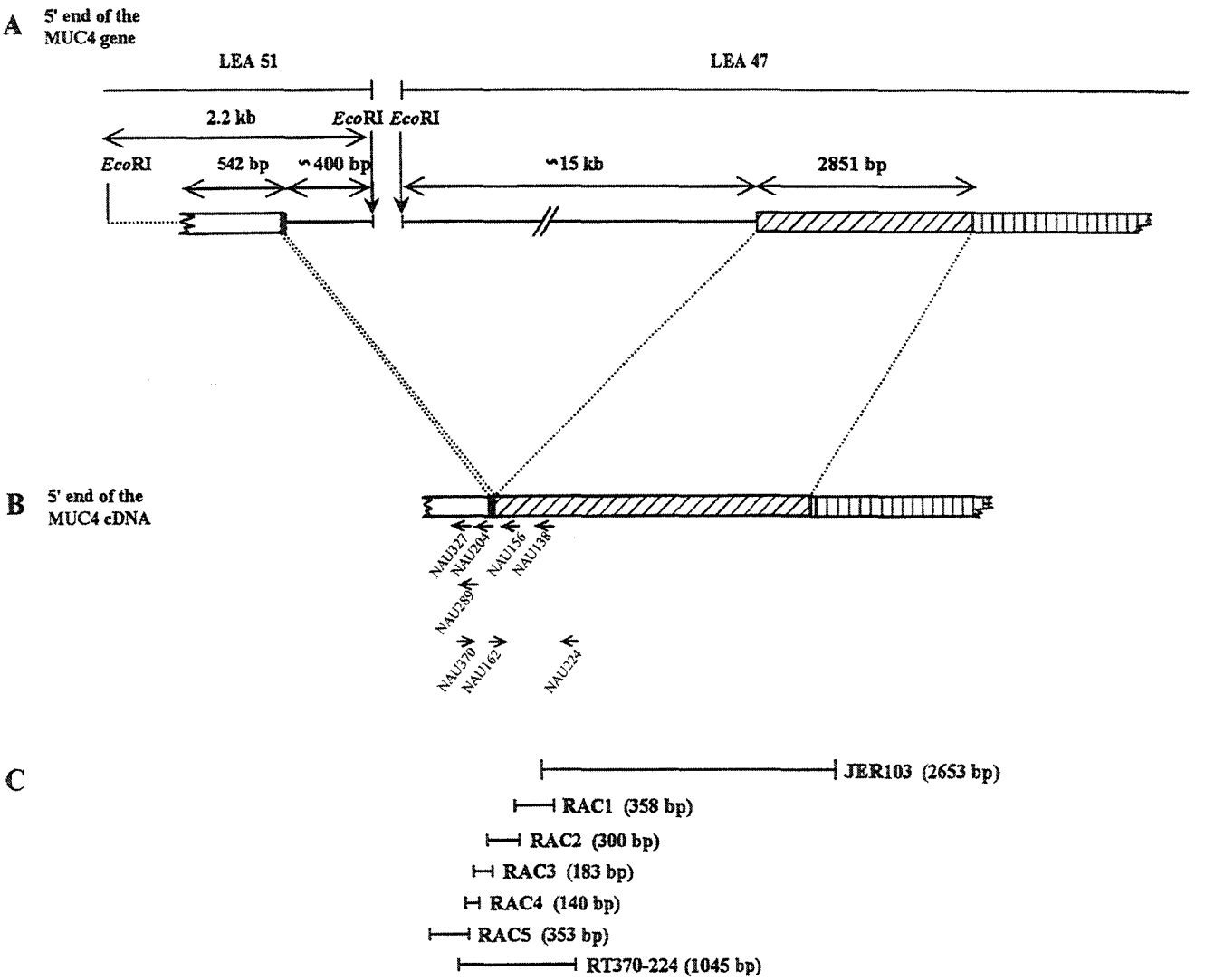


Figure 3 Organization of the 5′-terminal region of *MUC4*

(**A**) Schematic representation of the exon/intron structure of the 3′ end of clone LEA51 and the 5′ end of clone LEA2. The open box on the left represents the 5′ untranslated region, the black box the region encoding the signal peptide, the diagonally hatched box the unique sequence encoding a threonine/serine-rich region and the vertically hatched box on the right the 5′-terminal region of the *MUC4* 48 bp TR array. The broken line represents the intronic region. (**B**) Illustration of the compiled sequence of the different cDNA clones obtained. The primers and their directions are indicated by horizontal arrows and their NAU numbers (their locations are given in the Experimental section). (**C**) Location and length of the JER103 cDNA clone and of cDNA clones obtained by 5′ RACE and RT–PCR experiments.

RNA (tracheal mucosa or colon mucosa). The different amplification products were subcloned into pCR2.1 vector and sequenced. The clones were designated RT174-99. The sequences obtained (Figure 4) were analysed and compared with those obtained from HOR1 (the 2.5 kb *Eco*RI fragment of ANT55) (Figure 1) and its 5′ extension with NAU103 as primer with direct sequencing. A short intron of 333 bp whose location is marked in Figure 4 is 37 bp downstream of the 48 bp repeat, as evidenced when comparing cDNA and genomic sequences. Splice acceptor and donor sequences conform to the 'GT–AG' rule. This intron might also be categorized as class 1.

Polymorphism studies

Lymphocyte-derived DNA isolated from 18 volunteers (unrelated Caucasian individuals from the north of France) was digested with *Pst*I and *Eco*RI. As shown previously, the 48 bp TR array is flanked by a *Pst*I site (labelled 3 on Figure 1) and an *Eco*RI site, and the 15 bp TR array is flanked by the same *Eco*RI site and a *Pst*I site (labelled 4 on Figure 1). Three additional *Pst*I sites permitted the cleavage of the unique 5′-exonic sequence close to the 48 bp TR domain into two fragments. This double digestion was therefore useful in investigating contiguous fragments for polymorphism studies of *MUC4* (Figure 7). The two *Pst*I fragments located on the 5′ part of *MUC4*, which hybridized with the JER103 probe, were constantly 2 and 0.5 kb in length in all individuals tested. In contrast, the two TR arrays showed high degrees of polymorphism. With each probe, Southern blot analysis demonstrated two bands with the same intensity in 15 out of 18 individuals. These two bands were considered to be allelic forms of each TR domain; no additional band indicated the presence of *Pst*I or *Eco*RI sites within the TR domains. In the

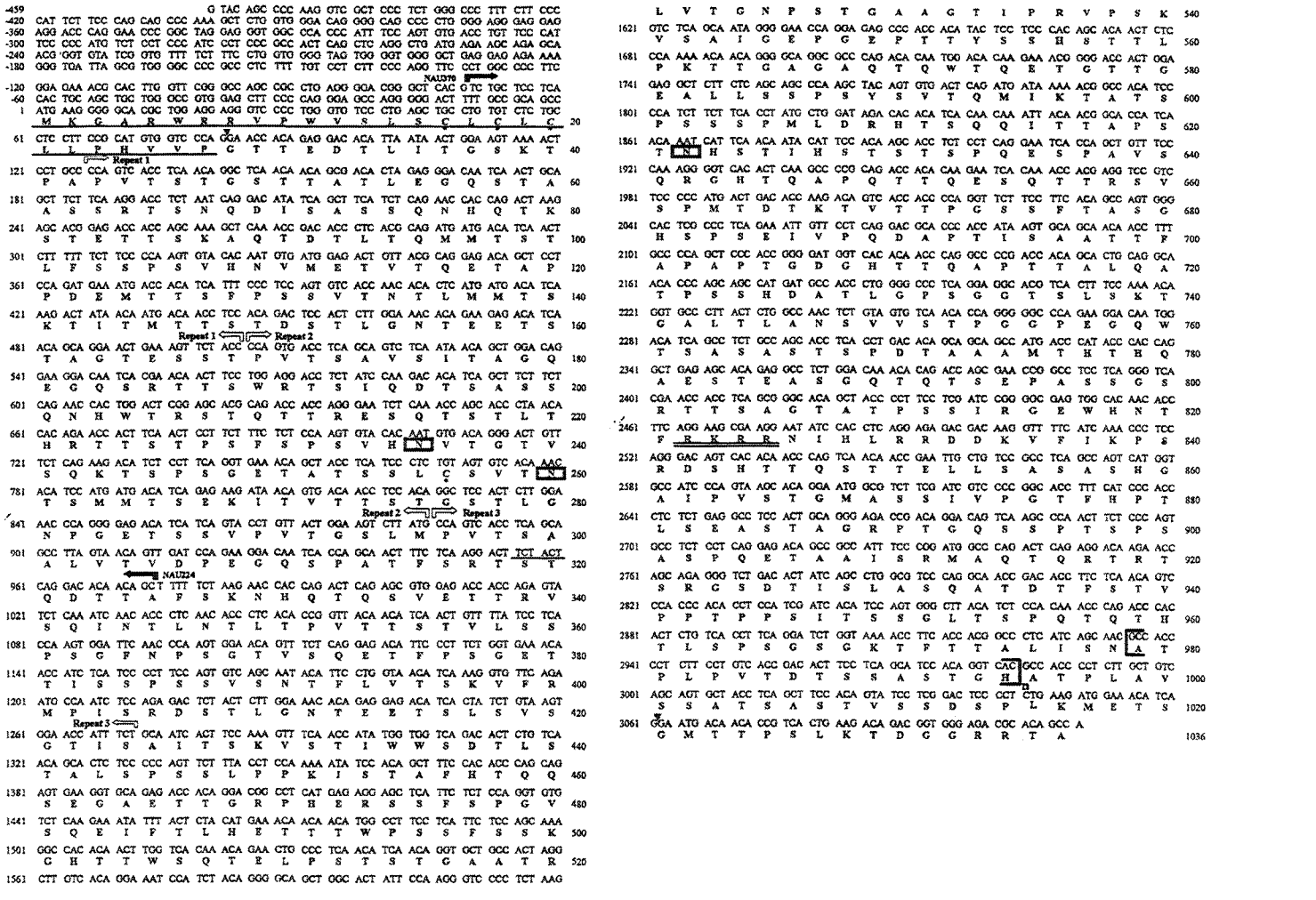


Figure 4 Compiled nucleotide sequences and deduced amino acid sequences of the RAC1-5, JER103 and RT174-99 cDNA clones

Nucleotide positions are indicated by the numbers at the left, and amino acid positions at the right. The locations of primers NAU370 and NAU224 are indicated by black arrows. The secretory protein signal sequence is underlined. The asterisks indicate the positions of cysteine residues (lines 20 and 260). Black downwards-pointing arrowheads indicate the positions of the introns. Open arrows indicate the three repeat units found at the beginning of the Thr/Ser-rich region. Three putative N-glycosylation sites are boxed. A tetrabasic amino acid sequence RKRR is doubly underlined. The sequence is presented with only one complete 48 bp tandem repeat. The number of repeats (*n*) varies from approx. 145 to approx. 395 in the present study. The cDNA sequence reported in this figure has been submitted to the EMBL Data Bank with accession number AJ000281.

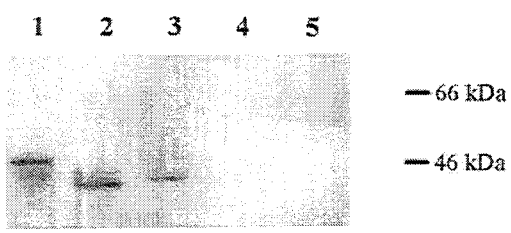


Figure 5 [^{35}S]Methionine labelling of the expression product of RT370-224 synthesized in a cell-free system in the absence and in the presence of microsomes and of endo H

Transcription and translation reactions with the TNT reticulocyte lysate system were performed and the reaction products were processed as described in the Experimental section. Lane 1, incubation with 1 μg of RT370-224 and microsomes; lane 2, incubation with 1 μg of RT370-224, microsomes and endo H; lane 3, incubation with 1 μg of RT370-224; lane 4, incubation with 1 μg of pCR2.1 vector; lane 5, incubation with 1 μg of pCR2.1 vector and microsomes. The sizes of the Rainbow$^{(R)}$ coloured protein molecular mass markers are indicated at the right.

48 bp TR domain, 13 distinct alleles were observed from 19 to 7 kb (A, 19 kb; B, 17 kb; C, 14 kb; E, 12 kb; F, 11.5 kb; G, 11 kb; H, 10.5 kb; I, 10 kb; J, 9.2 kb; K, 8.9 kb; L, 8.5 kb; M, 7.5 kb; N, 7 kb). The largest allele (A) was the most common. However, most individuals exhibiting the largest allele (12 out of 18) were heterozygous (10 out of 18), as only two displayed homozygosity. In these individuals, the size of the second allele varied from 12 kb (allele E) to 7 kb (allele N). Six individuals lacking the common allele A exhibited various other alleles (B, C, E, G, H, I and L); one of these was homozygous for E. The length of the *PstI*/*Eco*RI fragment containing the 48 bp TR domain in LEA2 and LEA47 corresponds to allele B. This in ANT55 was like allele G. In the intronic 15 bp TR domain of *MUC4*, 12 distinct alleles were observed from 6.5 to 1.5 kb (a, 6.5 kb; b, 6.2 kb; c, 5.5 kb; d, 5.2 kb; e, 5.0 kb; f, 4.0 kb; h, 3.6 kb; i, 3.3 kb; j, 2.8 kb; k, 2.4 kb; l, 1.8 kb; m, 1.5 kb). The most common allele found in 12 out of 18 individuals was h. Ten individuals were heterozygous and two homozygous; these two

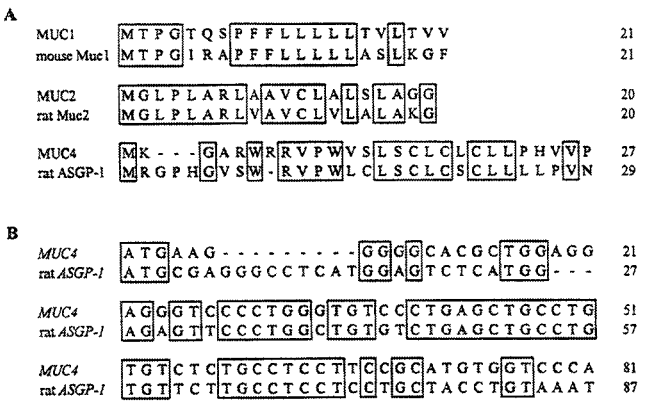


Figure 6 Comparison of several mucin signal peptides

(**A**) Alignment of the deduced amino acid sequences of the signal peptide of MUC1 with that of its mouse homologue, the signal peptide of MUC2 with that of its rat homologue, and the signal peptide of MUC4 with that of rat ASGP-1. Dashes indicate gaps introduced into the sequence for alignment purposes. Identical amino acids are boxed. The numbers indicate amino acid positions. (**B**) Nucleotide sequence similarity between the signal sequence of MUC4 and that of rat ASGP-1. Identical nucleotides are boxed. The numbers indicate nucleotide positions.

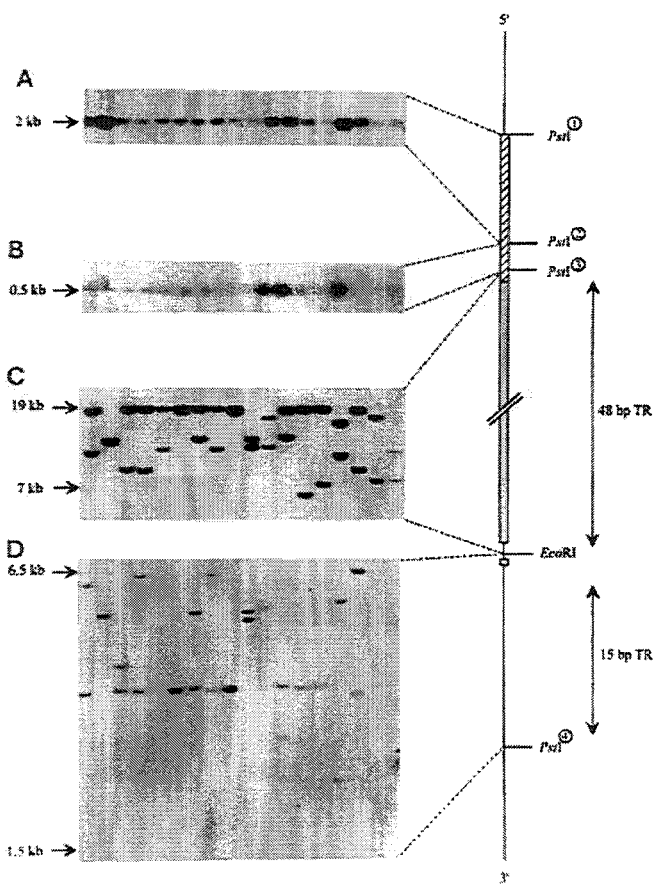


Figure 7 Location of sequence polymorphisms

Genomic DNA prepared from 18 random unrelated individuals was digested to completion with *PstI* and *EcoRI*. The same blot of *PstI*/*EcoRI*-digested DNA samples was hybridized sequentially with: *PstI*①–*PstI*② 5′ probe (**A**); *PstI*②–*PstI*③ 5′ probe (**B**); *MUC4* TR probe (JER64) (**C**); and 3′ probe (HOR1) (**D**).

individuals were also homozygous for A. The *Eco*RI/*Pst*I fragment containing the 15 bp TR domain in the two cosmids (LEA2 and LEA47) corresponds to allele h and that of the phage clone (ANT55) to allele j.

DISCUSSION

A human tracheobronchial mucin cDNA from a mucin gene named *MUC4* has previously been isolated in our laboratory [6]. The 1.83 kb JER64 clone insert consists entirely of nearly identical 48 bp repeats, encoding a Thr/Ser-rich stretch. This cDNA was mapped to chromosome 3q29 [8]. The whole genomic TR domain hybridizing to the JER64 probe is defined by a *Pst*I/*Eco*RI band of approx. 18.2 kb in cosmid clones and 11 kb in phage clones. Digestion with *Dde*I allowed us to establish that the *Pst*I/*Eco*RI fragment consists entirely of 48 bp repeats because no fragments larger than 96 bp (corresponding to two units) were observed. Thus the TR array in cosmid clones seems to contain approx. 380 uninterrupted individual units of 48 bp (approx. 18.2 kb) that are almost identical. Moreover, Southern blot analysis confirmed that the *MUC4* TR array of 12 individuals consists only of 48 bp units because alleles cut to completion by restriction enzyme *Dde*I showed only 48 or 96 bp units. Therefore the TR region of *MUC4* is uninterrupted and located within a single exon.

Gross et al. [8] suggested that the polymorphism observed between individuals with the JER64 probe was VNTR, owing to the repetitive structure of this cDNA. However, the restriction sites for the enzymes used in this study did not accurately flank the 48 bp TR. In the present study we chose to use *Pst*I and *Eco*RI, which just flank the 48 bp TR array, thereby enabling the isolation of the whole TR domain. Eighteen unrelated individuals were examined by Southern blot analysis; 13 different alleles were detected. In contrast, the unique 5′ sequences flanking the 48 bp TRs are of the same length in all genomic or cDNA clones analysed. Thus *MUC4* exhibits a length polymorphism in its TR array that can be characterized as the VNTR type. Alleles observed vary between 7 and 19 kb and correspond to a variation in the number of 48 bp TRs, ranging from approx. 145 to approx. 395 units. Nevertheless, all the individuals studied had at least one large allele. A similar expansive variation in the number of repeats has also been demonstrated for *MUC1*, whose alleles seem to contain between 20 and 125 TRs [24]. Carvalho et al. [29] have reported that individuals with small *MUC1* genotypes are more susceptible to developing gastric carcinoma, suggesting that these mucin gene VNTR variations have a possible functional significance.

The human mucin genes are characterized by a large TR array coding for a peptide that is typically rich in hydroxylated amino acid residues. The largest 60 bp TR region of *MUC1* is 7.5 kb [24]; that of *MUC2* is approx. 8 kb, corresponding to approx. 115 individual TR units of 69 bp [30]. The central region of *MUC7* consists of six highly similar TRs of 69 bp [31]. As with *MUC1*, *MUC2* and *MUC7*, the TR region of *MUC4* is not interrupted by a unique segment. In contrast, Desseyn et al. [17] have indicated that the irregular repeat of 87 bp of *MUC5B* is interrupted by seven unique conserved subdomains that are cysteine-rich. The central exon of *MUC5B* has been completely sequenced and encompasses 10713 bp. The sequence of the TR domain of *MUC5AC* is still incomplete, but in this gene, as in *MUC5B*, this domain is interrupted several times by cysteine-rich unique subdomains [7]. Only partial information is available about the size of the repeat region of *MUC3*, *MUC5AC* and *MUC6*. Hence, with regard to human mucin genes, the TR array of *MUC4* (as far as the longest alleles are concerned) is the

largest described so far. Eckhardt et al. [32] have recently reported that the TR domain of pig submaxillary mucin is encoded by an unusually long exon (34–32.8 kb).

The JER64 probe allowed us to isolate a *MUC4* cDNA clone from a human colon mucosa cDNA library. This clone contained a unique sequence located on the 5′ side of the 48 bp TR domain. This sequence was extended by a 5′ RACE procedure. Exon 1 consists of a 5′ untranslated sequence and an 82 bp fragment encoding the first 27 N-terminal residues, which are very hydrophobic and might comprise the MUC4 signal peptide as demonstrated with transcription and translation assays *in vitro* in the presence of dog pancreatic microsomes. An intron spanning at least 15 kb is located close to the boundary between the putative signal peptide and the mature peptide. Signal peptides are short N-terminal extension sequences required for the translocation of growing polypeptide chains through membranes of the endoplasmic reticulum [33]. In spite of the fact that these short cleaved N-terminal sequences constantly contain clustered hydrophobic residues flanked by hydrophilic residues, the primary sequences themselves vary considerably, in both length and structure [34–36]. A comparison of human and other mammalian preproteins therefore usually shows very little sequence similarity [37]. Nevertheless some proteins belonging to family groups show a high level of conservation through evolution of the signal peptide and occasionally of the 5′ untranslated region sequence. For example, milk proteins (alpha S1, alpha S2 and beta caseins) [38], saliva proteins (proline-rich proteins) [39], statherin and histidine-rich peptides [40] are striking examples of conserved signal peptides. The high degrees of similarity (70–99%) of the signal peptide of the different rat and mouse proteins belonging to the proline-rich protein family is particularly striking, especially when considering that the tandem sequence repeats of these proteins highly diverge. Therefore, as the putative signal peptide of MUC4 shows a high degree of similarity to that of ASGP-1, we can postulate that MUC4 and rat ASGP-1 apomucins might have derived from a common ancestor. After the signal sequence, no similarities were found between the two proteins; however, the 3′ cDNA sequence of *MUC4* remains to be studied to define whether other regions of MUC4 demonstrate similarity to rat ASGP-1. However, for human statherin and basic histidine-rich peptide, two submandibular-gland proteins considered to be derived from a common ancestral gene, the majority of the coding sequence shows no significant similarity, except the first 14 N-terminal residues and the signal peptide. The unusual conservation of the sequence surrounding the AUG initiation codon has been proposed to contribute to specific regulation and/or to mRNA stability and conformation [39].

The second exon of *MUC4* encodes a peptide that can be divided into four distinct subdomains: (1) a first region with three imperfect repeats of 126 residues containing a high proportion of threonine and serine residues, (2) a second region of 554 residues corresponding to a unique sequence typically enriched for threonine and serine, (3) a third major domain, the 16-residue TR region, followed by (4) a fourth domain with a unique 26-residue sequence. In this regard *MUC4* resembles *MUC1* and *MUC2* genes. The N-terminal region of *MUC1* is also rich in serine, threonine and proline residues [24]. Similarly, the region upstream of the major TR domain of *MUC2* is rich in hydroxylated amino acid residues [30]. It consists mostly of 16-residue repeat units that are often non-continuous. As seen for *MUC4*, this region together with the TR array forms a single large exon.

Exon 2 also contains a tetrabasic amino acid RKRR site. The consensus sequence RXK/RR is conserved at the cleavage sites of a wide variety of secretory and membrane protein precursors (e.g. von Willebrand factor, growth factors, serum proteases, receptors, and viral envelope glycoproteins) and is proposed to be processed by furin [41].

The four human mucin genes *MUC2*, *MUC5AC*, *MUC5B* and *MUC6* are clustered on chromosome 11p15.5 [42], whereas the five other genes are dispersed randomly throughout the human genome. A typical cysteine-rich domain, termed the Cys-subdomain, seems to be a characteristic of at least three (*MUC2*, *MUC5AC* and *MUC5B*) of the four human mucin genes located on 11p15.5 [17]. Desseyn et al. [43] suggested that these three genes might be a part of a highly conserved protein family and proposed an evolutionary tree.

No consensus cysteine-rich region was observed either in the N-terminus or in the TR array of MUC4, in contrast with the gel-forming mucins encoded by genes belonging to the 11p15.5 gene family [26,44–47]. Moreover, cell expression of *MUC4* distinguishes it from mucins located on 11p15.5. Thus *MUC4* is expressed in several epithelial cells other than typical mucin-secreting cells such as goblet and mucous cells; *MUC4* is expressed both in respiratory ciliated cells and in intestine absorptive cells [9].

Further investigations to elucidate the entire genomic organization of *MUC4* will be of great interest in shedding light on the dysregulation of this gene in tumorigenesis. The production of novel antibodies specifically defining MUC4 apomucin is currently being performed and will greatly facilitate studies of MUC4 protein/carbohydrate expression in epithelial disease.

We thank Annette Leclercq and Christine Mouton for performing polymorphism analysis, Marie-José Dejonghe for performing the SDS/PAGE analysis, Pascal Mathon for help in the preparation of the Figures, and Dr. Tor Savidge for his help in improving the style of this paper. This work was supported by le Comité du Nord de La Ligue contre le Cancer and l'Association de Recherche contre le Cancer. J. M. is a recipient of an Association Française de Lutte contre la Mucoviscidose fellowship.

REFERENCES

1 Gendler, S. J. and Spicer, A. P. (1995) Annu. Rev. Physiol. **57**, 607–634
2 Bansil, R., Stanley, E. and LaMont, J. T. (1995) Annu. Rev. Physiol. **57**, 635–657
3 Forstner, G. (1995) Annu. Rev. Physiol. **57**, 585–605
4 Shankar, V., Pichan, P., Eddy, Jr., R. L., Tonk, V., Nowak, N., Sait, S. N. J., Shows, T. B., Schultz, R. E., Gotway, G., Elkins, R. C., Gilmore, M. S. and Sachdev, G. P. (1997) Am. J. Respir. Cell Mol. Biol. **16**, 232–241
5 Crépin, M., Porchet, N., Aubert, J. P. and Degand, P. (1990) Biorheology **27**, 471–484
6 Porchet, N., Nguyen, V. C., Dufossé, J., Audié, J. P., Guyonnet Dupérat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P.(1991) Biochem. Biophys. Res. Commun. **175**, 414–422
7 Guyonnet Dupérat, V., Audié, J. P., Debailleul, V., Laine, A., Buisine, M. P., Galiegue-Zouitina, S., Pigny, P., Degand, P., Aubert, J. P. and Porchet, N. (1995) Biochem. J. **305**, 211–219
8 Gross, M. S., Guyonnet Dupérat, V., Porchet, N., Bernheim, A., Aubert, J. P. and Nguyen, V. C. (1992) Ann. Genet. **35**, 21–26
9 Audié, J. P., Janin, A., Porchet, N., Copin, M. C., Gosselin, B. and Aubert, J. P. (1993) J. Histochem. Cytochem. **41**, 1479–1485
10 Audié, J. P., Tetaert, D., Pigny, P., Buisine, M. P., Janin, A., Aubert, J. P., Porchet, N. and Boersma, A. (1995) Hum. Reprod. **10**, 98–102
11 Balagué, C., Gambús, G., Carrato, C., Porchet, N., Aubert, J. P., Kim, Y. S. and Real, F. X. (1994) Gastroenterology **106**, 1054–1061
12 Hollingsworth, M. A., Strawhecker, J. M., Caffrey, T. C. and Mack, D. R. (1994) Int. J. Cancer **57**, 198–203
13 Walsh, M. D., McGuckin, M. A., Devine, P. L., Hohn, B. G. and Wright, R. G. (1993) J. Clin. Pathol. **46**, 922–925
14 Nguyen, P. L., Niehans, G. A., Cherwitz, D. L., Kim, Y. S. and Ho, S. B. (1996) Tumor Biol. **17**, 176–192
15 Ogata, S., Uehara, H., Chen, A. and Itzkowitz, S. H. (1992) Cancer Res. **52**, 5971–5978
16 Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B., Zhao, J. and Evans, G. A. (1987) Proc. Natl. Acad. Sci. U.S.A. **4**, 2160–2164

17 Desseyn, J. L., Guyonnet Dupérat, V., Porchet, N., Aubert, J. P. and Laine, A. (1997) J. Biol. Chem. **272**, 3168–3178
18 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
19 Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
20 Patthy, L. (1987) FEBS Lett. **214**, 1–7
21 Kozak, M. (1987) Nucleic Acids Res. **15**, 8125–8148
22 Kyte, J. and Doolittle, R.F (1982) J. Mol. Biol. **157**, 105–132
23 Wu, K., Fregien, N. and Carraway, K. L. (1994) J. Biol. Chem. **269**, 11950–11955
24 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N. and Wilson, D. (1990) J. Biol. Chem. **265**, 15286–15293
25 Spicer, A. P., Parry, G., Patton, S. and Gendler, S. J. (1991) J. Biol. Chem. **266**, 15099–15109
26 Gum, Jr., J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) J. Biol. Chem. **269**, 2440–2446
27 Ohmori, H., Dohrman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, Jr., J. R., Kim, Y. S. and Basbaum, C. B. (1994) J. Biol. Chem. **269**, 17833–17840
28 Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G. and Kezdy, F. J. (1993) J. Biol. Chem. **268**, 10029–10038
29 Carvalho, F., Seruca, R., David, L., Amorim, A., Seixas, M., Bennett, E., Clausen, H. and Sobrinho-Simoes, M. (1997) Glycoconj. J. **14**, 107–111
30 Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M. and Kim, Y. S. (1991) J. Clin. Invest. **88**, 1005–1013
31 Bobek, L. A., Tsai, H., Biesbrock, A. R. and Levine, M. J. (1993) J. Biol. Chem. **268**, 20563–20569
32 Eckhardt, A. E., Timpte, C. S., DeLuca, A. W. and Hill, R. L. (1997) J. Biol. Chem. **272**, 33204–33210
33 Kreil, G. (1981) Annu. Rev. Biochem. **50**, 317–348
34 Haeuptle, M. T., Flint, N., Gough, N. M. and Dobberstein, B. (1989) J. Cell Biol. **108**, 1227–1236
35 Bird, P., Gething, M. J. and Sambrook, J. (1990) J. Biol. Chem. **265**, 8420–8425
36 Sakaguchi, M., Tomiyoshi, R., Kuroiwa, T., Mihara, K. and Omura, T. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 16–19
37 Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H. and Lingappa, V. R. (1979) Symp. Soc. Exp. Biol. **33**, 9–36
38 Gaye, P. and Mercier, J. C. (1981) Reprod. Nutr. Dev. **21**, 199–208
39 Clements, S., Mehansho, H. and Carlson, D. M. (1985) J. Biol. Chem. **260**, 13471–13477
40 Dickinson, D. P., Ridall, A. L. and Levine, M. J. (1987) Biochem. Biophys. Res. Commun. **149**, 784–790
41 Rehemtulla, A. and Kaufman, R. J. (1992) Curr. Opin. Biotechnol. **3**, 560–565
42 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. et al. (1996) Genomics **38**, 340–352
43 Desseyn, J. L., Buisine, M. P., Porchet, N., Aubert, J. P., Degand, P. and Laine, A. (1998) J. Mol. Evol. **46**, 102–106
44 Gum, Jr., J. R., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E. and Kim, Y. S. (1992) J. Biol. Chem. **267**, 21375–21383
45 Lesuffleur, T., Roche, F., Hill, A. S., Lacasa, M., Fox, M., Swallow, D. M., Zweibaum, A. and Real, F. X. (1995) J. Biol. Chem. **270**, 13665–13673
46 Desseyn, J. L., Aubert, J. P., Van Seuningen, I., Porchet, N. and Laine, A. (1997) J. Biol. Chem. **272**, 16873–16883
47 Toribara, N. W., Ho, S. B., Gum, E., Gum, Jr., J. R., Lau, P. and Kim, Y. S. (1997) J. Biol. Chem. **272**, 16398–16403

Received 19 December 1997/6 March 1998; accepted 23 March 1998

Stratégie

I. Situation de notre sujet.

Le criblage de la banque génomique construite en vecteur phagique avait permis d'isoler deux clones chevauchants, ANT55 et ANT56. Ces clones avaient une taille d'environ 15 kb. Leur étude avait permis d'identifier une nouvelle séquence répétitive de motif élémentaire de 15 pb.

Comme nous suspections que le gène *MUC4* soit de grande taille, nous avons criblé, en collaboration avec Séverine Nollet, une banque de gènes construite en vecteur cosmidique. Ce type de vecteur permet d'isoler des fragments d'ADN d'une taille allant de 40 à 50 kb. Environ 50 clones positifs ont pu être isolés de cette nouvelle banque génomique. Les deux clones, dénommés LEA2 et LEA47, ont été retenus pour établir, par la technique dite d'hydrolyse partielle, la carte de restriction de la partie du gène *MUC4* isolé. Ces deux clones offrent le plus long fragment du gène *MUC4* identifié d'une taille d'environ 50 kb.

Grâce à ces deux clones, Séverine Nollet a débuté l'étude du domaine répétitif de motif élémentaire de 48 pb ainsi que la caractérisation de la région 5' du gène. Afin d'identifier les séquences codantes de *MUC4*, Séverine Nollet a construit et criblé une banque d'ADNc à partir de l'ARNm extrait de muqueuse colique.

II. Notre travail de thèse.

II. 1. Caratérisation de l'extrémité 3' de *MUC4*.

Le but de notre travail était d'isoler et d'identifier l'extrémité 3' du gène *MUC4*. Pour ce faire, nous avons commencé par cribler avec la sonde correspondant au motif élémentaire de 48 pb, la banque d'ADNc. Après le criblage de plus de 50000 clones, aucun des clones positifs isolés n'avait permis d'identifier de séquence d'ADNc de la région 3' de *MUC4*.

Nous avons donc décidé de changer de technique d'approche et de sous-cloner les séquences d'ADN génomique en aval du domaine répétitif des clones LEA2 et LEA47. Tous les sous-clones obtenus ont été séquencés à leurs extrémités et les séquences comparées aux banques de données internationales. Les sous-clones, utilisés comme sonde, ont également été hybridés à un Northern blot réalisé à partir d'ARN total extrait de plusieurs tissus épithéliaux commerciaux.

L'un des sous-clones, S610, a montré un profil d'expression après hybridation sur Northern blot, comparable à celui détecté par le motif élémentaire de 48 pb. Une séquence correspondant à l'une des extrémités de S610, a également montré 78 % de similarité avec une séquence en aval du domaine répétitif de SMC.

Les travaux de recherche de Séverine Nollet avaient permis, depuis peu, d'isoler l'extrémité 5' de *MUC4*. Elle avait mis en évidence que la séquence codant le peptide signal de SMC montrait 60 % de similarité avec la séquence codant le peptide signal de *MUC4*. A ce stade nous avons donc émis l'hypothèse selon laquelle MUC4 pouvait appartenir à la même famille de mucines que SMC et peut-être même représenter l'homologue chez l'homme de cette mucine membranaire de rat. Un certain nombre d'arguments tirés de la littérature et montrant pour ces deux mucines des territoires d'expression similaires et des dérégulations associées aux carcinomes de mêmes organes autorisaient de fonder notre stratégie sur cette hypothèse. Nous avons donc supposé que le fragment de S610 montrant 78 % de similarité avec *SMC* pouvait être exonique. Nous avons donc choisi un oligonucléotide antisens dans cette séquence et réalisé une RT-PCR. Comme oligonucléotide sens, nous avons choisi un oligonucléotide juste en aval de la séquence répétitive de motif élémentaire de 48 pb. Un produit d'amplification, S1217, a été obtenu et cloné. La comparaison de la séquence de ce produit d'amplification avec les séquences génomiques a montré que ce clone était composé de 3 exons.

La découverte d'un degré de similarité important entre une séquence d'ADN génomique de *MUC4* et une séquence de l'ADNc de *SMC* a donc permis d'amorcer la marche vers l'extrémité 3' de l'ADNc de *MUC4*. Des expériences de criblages de banque d'ADNc avec S1217 utilisé comme sonde, et une expérience de RACE-PCR ont permis d'obtenir l'extrémité 3' complète de l'ADNc de *MUC4*. La figure 18 rassemble tous les sous-clones obtenus par RT-PCR, criblage de banque et RACE-PCR.

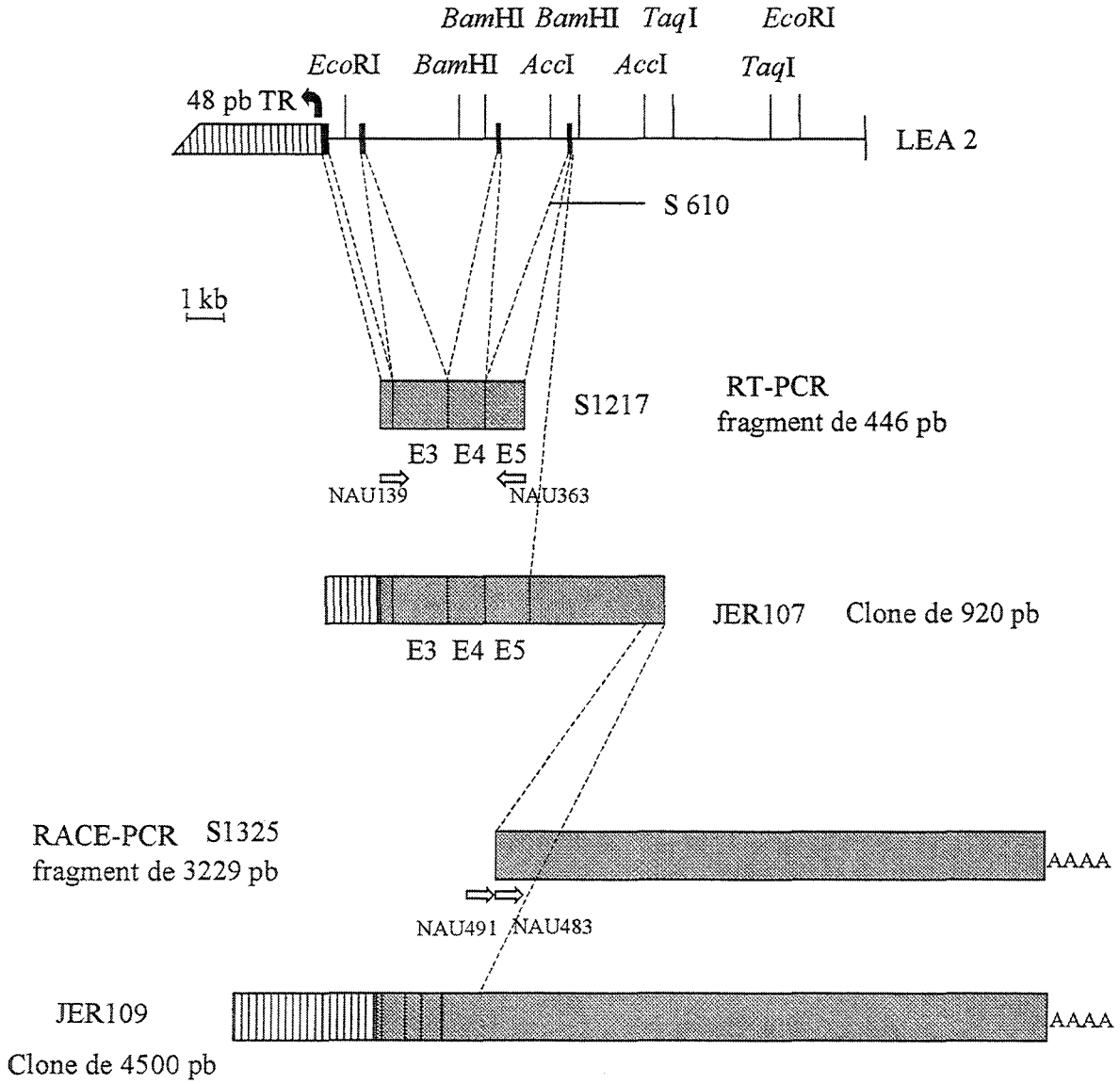


Figure 18 : Représentation schématique des clones d'ADNc de MUC4 obtenus par PCR, criblage de banque et RACE-PCR (boite hachurée : séquence répétitive, boites grises : séquences uniques).

La réalisation de la RACE-PCR a nécessité des modifications par rapport à la technique de base. Le degré de similarité rencontré entre SMC et les fragments isolés de l'ADNc de MUC4 nous ont conforté dans notre hypothèse selon laquelle MUC4 puisse être l'homologue humain de SMC. De ce fait, nous avons supposé que l'extrémité 3' de MUC4 était d'une taille telle que les techniques classiques de PCR ne permettent de l'amplifier. Nous savions que la technique de synthèse des ADNc par la technique "Advantage for RT-PCR kit" (Clontech) permettait l'obtention d'ADNc de grandes tailles. Nous avons donc réalisé l'ADNc à partir de l'ARN total extrait de muqueuse colique avec ce kit mais en

utilisant comme oligonucléotide, l'oligonucléotide spécifique pour la réalisation de RACE-PCR du kit 5'/3'-RACE kit (Boerhinger Mannheim). L'amplification de l'ADNc complémentaire synthétisé lors de cette expérience a été réalisée avec comme enzyme la polymérase du kit "Expand Long Template PCR system". La figure 19 résume la technique de RACE-PCR utilisée.

II. 2. Caractérisation des isoformes de MUC4.

Au cours de la caractérisation de l'extrémité 3' de l'ADNc de *MUC*4, nous avons isolé par criblage de banque d'ADNc et par RT-PCR des clones de séquences différentes. Ces clones se différenciaient par l'ajout ou la perte de domaines évoquant un mécanisme d'épissage alternatif. Nous avons étudié la présence d'événements d'épissage alternatif par RT-PCR avec des oligonucléotides choisis tout au long de la séquence 3'-terminale de *MUC4*. Les oligonucléotides ont été choisis afin d'amplifier des fragments d'une taille allant de 400 à 600 pb. Nous avons pu identifier par cette technique des événements complexes d'épissage alternatif tout au long de l'extrémité 3' de l'ADNc de *MUC4* et nous avons pu identifier l'isoforme prépondérante dans l'échantillon biologique étudié c'est à dire le côlon.

Cette technique de RT-PCR a été menée sur plusieurs ADNc préparés à partir d'ARN commerciaux extraits de différents tissus. Nous avons pu mettre en évidence que la qualité et la quantité des isoformes de *MUC4* étaient spécifiques de tissu. Le niveau d'expression des isoformes additionnelles de *MUC4* semblaient être maximum pour l'ARN extrait du testicule.

Afin de différencier les isoformes de *MUC4*, l'amplification de toute l'extrémité 3'-terminale était nécessaire. Nous avons utilisé à nouveau la technique "d'expand long RT-PCR". Le clonage des produits d'amplification a été réalisé sans purification sur gel afin de pouvoir cloner toutes les formes amplifiées.

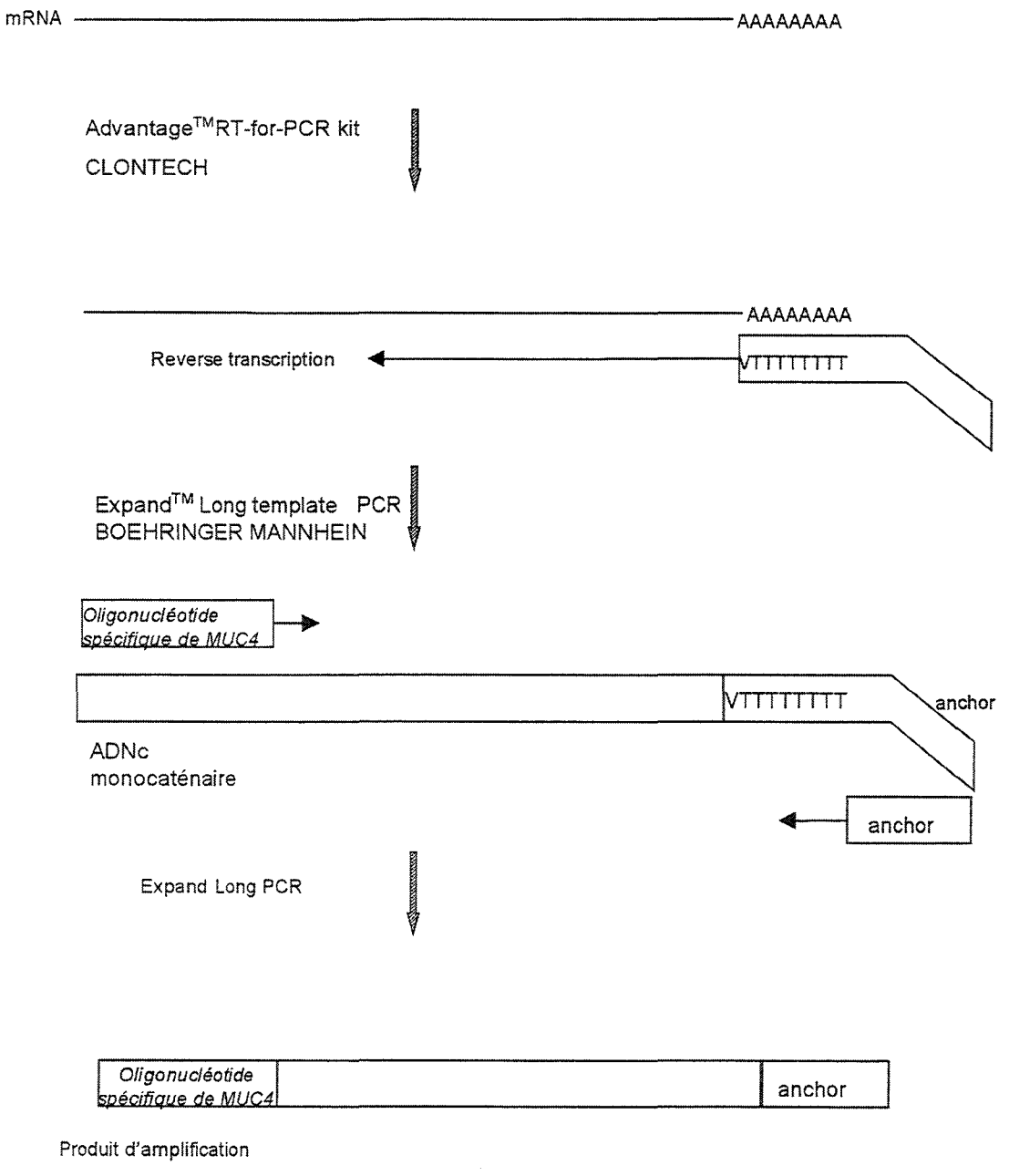


Figure 19 : Représentation schématique de l'expérience de RACE-PCR.

II. 3. Relation structure-fonction.

L'analyse des séquences de l'extrémité 3' de *MUC4* a montré un haut degré de similarité avec les séquences de *SMC*.

SMC est composée de deux sous-unités, une sous-unité de type mucine et une sous-unité de type facteur de croissance. Cette dernière possède deux domaines de type EGF, un

domaine transmembranaire et une queue cytoplasmique. SMC est capable d'interagir directement avec l'oncogène $p185^{neu}$. Le degré de similarité entre les séquences 3' des ADNc codant MUC4 et SMC nous a fait supposer que MUC4 puisse avoir la même organisation que SMC et puisse donc interagir avec ErbB2 qui est l'homologue humain de la $p185^{neu}$.

Nous ne disposions d'aucun anticorps permettant d'immunoprécipiter MUC4. Nous avons opté pour l'étude de l'interaction potentielle de MUC4 avec ErbB2 grâce à la technique de production d'une protéine de fusion GST-MUC4. La taille de MUC4 ne permettant pas la production d'une protéine de fusion ayant tous ses domaines peptidiques, nous avons décidé de limiter notre étude aux domaines composant la sous-unité de type facteur de croissance.

Nous avons pu, lors d'une étude préliminaire, montrer que MUC4 pouvait interagir avec ErbB2. Nous avons ensuite étudié l'hypothèse selon laquelle la surexpression de MUC4 pouvait avoir un effet mitogène. Nous avons transfecté la sous-unité de type facteur de croissance de MUC4 dans la lignée de cellules issue de cancer mammaire MCF7 par la technique de transfection instable sous contrôle d'un promoteur constitutif. Cette technique nous a permis d'obtenir des résultats préliminaires sur la fonction potentielle de MUC4 dans la croissance.

II. 4. *MUC4* et polymorphisme.

La comparaison des séquences d'ADNc et des séquences génomiques de l'extrémité 3' de MUC4 nous ont permis d'identifier et de localiser deux nouvelles séquences répétitives introniques.

Au cours de son travail de thèse, Séverine Nollet a étudié le polymorphisme associé à la séquence répétitive exonique de motif élémentaire de 48 pb ainsi que celui associé à une séquence intronique de motif élémentaire de 15 pb. Elle a remarqué que les variations haplotypiques de ces deux séquences répétitives ne semblaient pas être aléatoires.

Nous avons remarqué que ces séquences répétitives introniques représentent presque toutes les séquences introniques situées au voisinage des sites d'épissage alternatif et sont donc potentiellement en relation avec le mécanisme d'épissage. Nous avons donc décidé d'étudier le polymorphisme associé aux deux nouvelles séquences répétitives introniques et de vérifier si leurs variations montraient une relation.

Pour réaliser ce travail, et afin de pouvoir le comparer aux résultats de Séverine Nollet, nous avons choisi la même stratégie d'étude par Southern blot après hydrolyse par *Eco*RI/*Pst*I.

Résultats et discussion

I. Caractérisation de l'extrémité 3' du gène *MUC4*.

I. 1. Stratégie.

Nous disposions pour débuter ce travail de 2 clones d'ADN génomique obtenus par le criblage de la banque de gènes construite en vecteur cosmidique. Les inserts contenus dans ces 2 clones chevauchants, LEA2 et LEA47, représentent notre fragment le plus long d'ADN isolé pour le gène *MUC4*. Nous disposions également d'un clone d'ADNc JER64 et d'une banque d'ADNc construite à partir de l'ARNm extrait de muqueuse colique. Les aspects techniques d'obtention et de caractérisation de ces outils sont décrits dans le mémoire de thèse de Séverine Nollet ainsi que dans l'article précédement présenté.

Nous avons commencé par cribler la banque d'expression avec la sonde JER64 afin d'isoler des clones chevauchant la région répétitive et pouvant étendre l'extrémité 3'. Après criblage de plus de 50000 clones, aucun des clones positifs isolés n'a permis d'identifier de séquence d'ADNc de la région 3' de *MUC4*. Nous avons donc décidé de changer de stratégie.

Nous avons sous-cloné les séquences d'ADN génomique en aval du domaine répétitif des clones cosmidiques LEA2 et LEA47. Les sous-clones obtenus ont été séquencés à chacune de leurs extrémités. Les séquences ont alors été comparées aux banques de données internationales. Une séquence correspondant à l'une des extrémités d'un fragment *Acc*I/*Acc*I de 2,8 kb (appelé S610) montre 78 % de similarité avec une séquence en aval du domaine répétitif de SMC. Comme la séquence codant le peptide signal de MUC4 est également similaire à la séquence codant le peptide signal de SMC, nous avons suspecté que la séquence du clone S610 puisse être exonique. De plus, S610 utilisé comme sonde montre un profil d'expression comparable à celui détecté par la sonde JER64 en Northern blot.

Nous avons donc choisi un oligonucléotide antisens dans cette séquence de S610 (NAU363) et réalisé une expérience de RT-PCR sur un ARNm extrait d'un épithélium

colique. Nous avons choisi comme oligonucléotide sens (NAU139) les 21 premiers nucléotides de la séquence unique juste en aval du domaine répétitif. Un produit d'amplification d'une taille de 446 pb a été obtenu et cloné (S1217). La comparaison des séquences de S1217 avec les séquences génomiques montre que S1217 est composé de 3 exons.

Nous avons criblé à nouveau la banque d'expression avec S1217, un clone positif a été isolé, JER107. D'une taille de 920 pb, il est composé de 70 pb correspondant au motif répétitif de 48 pb ainsi que de la séquence de S1217 qu'il prolonge de 404 pb.

Une expérience de 3'RACE-PCR a été réalisée avec un oligonucléotide sens choisi à l'extrémité 3' de JER107. Un produit d'amplification de 3229 pb a pu être obtenu, S1325.

Parallèlement à cette expérience, nous avons criblé la banque d'expression avec comme sonde le clone JER107. Un clone d'une taille de 4500 pb a pu être obtenu, JER109. L'analyse des séquences de tous les clones d'ADNc obtenus nous permet d'établir la séquence codante complète de l'extrémité 3'-terminale de *MUC4*.

Cette séquence est maintenant disponible dans la banque internationale de données sous le numéro d'accession AJ010901.

I. 2. Résultats

Les résultats de l'étude des domaines C-terminaux de MUC4 ont fait l'objet de la publication suivante :

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

La séquence en aval du domaine répétitif de 48 pb a une taille de 3468 pb et code un peptide de 1156 résidus d'acides aminés. Elle est suivie d'une séquence non traduite de 405 pb. Comme les autres mucines, MUC4 est une protéine modulaire. Son extrémité C-terminale est constituée de 12 modules différents, de CT1 à CT12 (Tableau 4). Les 4 premiers domaines sont séparés des 8 suivants par la présence d'un site potentiel de clivage protéolytique GlyAspProHis (GDPH). A l'exception du domaine CT1, tous les autres

domaines C-terminaux montrent un fort degré de similarité avec les séquences C-terminales de SMC. Comme SMC, MUC4 contient deux domaines riches en sites potentiels de N-glycosylation, deux domaines de type EGF, un domaine transmembranaire et un domaine cytoplasmique.

nom	position	caractéristique	similarité avec SMC
CT1	1 to 168	domaine de type mucine	
CT2	169 to 912	séquence unique	ASGP1
CT3	913 to 1251	domaine riche en résidus de cystéine	ASGP1
CT4	1252 to 1293	séquence unique	ASGP1
	1288 to 1299	site de clivage GDPH	site de clivage GDPH
CT5	1294 to 2331	domaine riche en sites de N-glycosylation	ASGP2
CT6	2332 to 2580	domaine riche en résidus de cystéine	ASGP2
CT7	2581 to 2700	domaine de type EGF	ASGP2
CT8	2701 to 3135	domaine riche en sites de N-glycosylation	ASGP2
CT9	3136 to 3270	domaine de type EGF	ASGP2
CT10	3271 to 3327	séquence unique	ASGP2
CT11	3328 to 3401	domaine transmembranaire	ASGP2
CT12	3402 to 3468	domaine cytoplasmique	ASGP2

Tableau 4 : Position et nature des différents domaines qui composent l'extrémité 3'-terminale de *MUC4*.

I. 3. Discussion.

La comparaison des séquences C-terminales de MUC4 et de SMC nous permet d'établir que MUC4 est l'homologue humain de la sialomucine de rat SMC. Toutes ces données ajoutées aux résultats concernant l'extrémité N-terminale ainsi que ceux du domaine central nous permet d'établir l'organisation structurale complète de MUC4 (Figure 20). Pour son allèle le plus grand, le gène *MUC4* transcrit un ARNm de 26,5 kb qui pourrait coder un hétérodimère complexe implanté dans la membrane d'une taille maximale estimée à 2,12 µm.

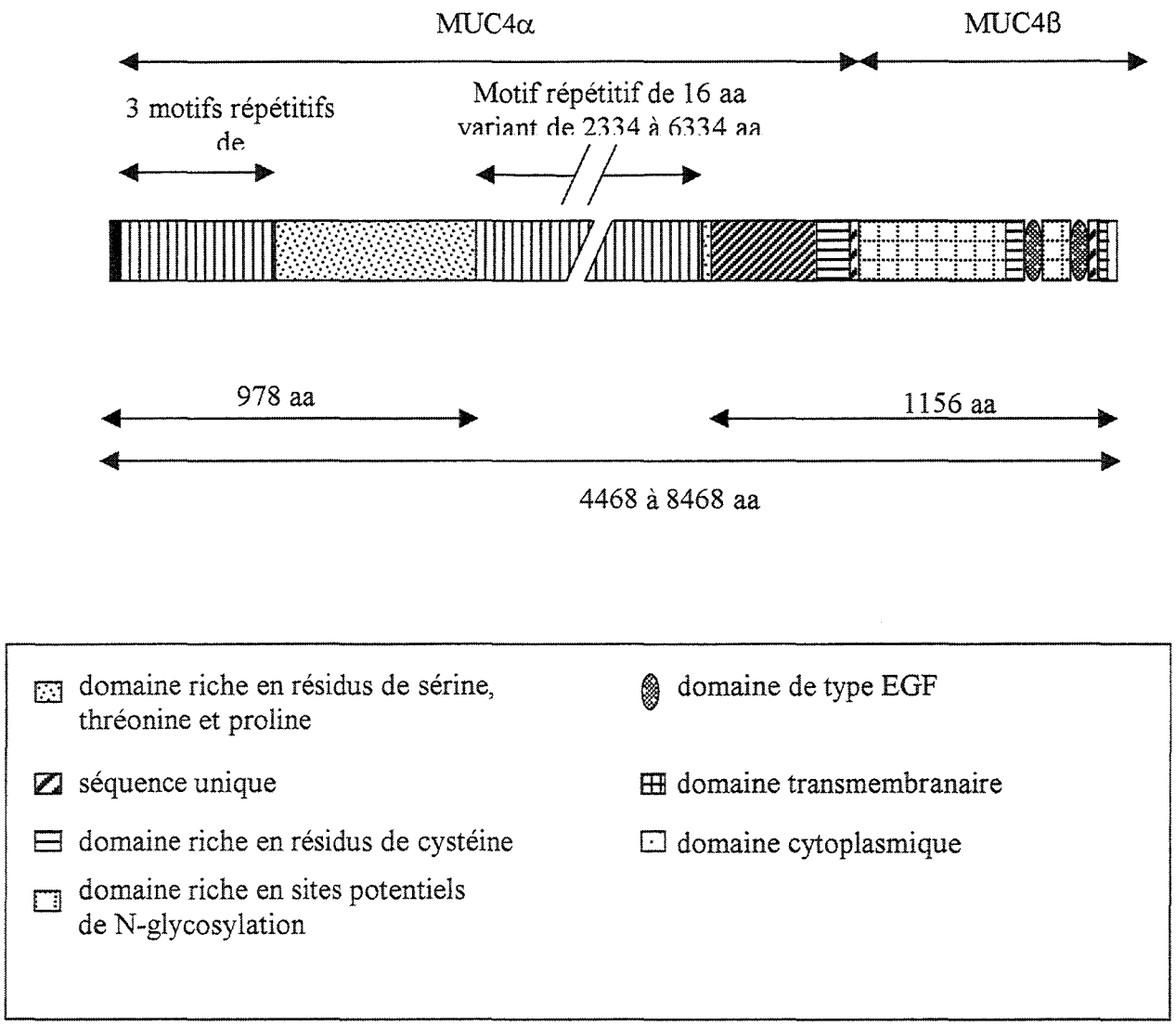


Figure 20 : Représentation schématique de l'organisation structurale complète de MUC4.

Tout comme son homologue de rat, SMC, MUC4 pourrait donc être une glycoprotéine bifonctionnelle constituée de 2 sous-unités, une sous unité sécrétée de type mucine de 850 kDa, MUC4α, reliée de façon non covalente à une sous-unité membranaire de type facteur de croissance de 80 kDa, MUC4β.

Ce résultat nous amène évidemment à poser la question de l'existence d'une interaction moléculaire entre MUC4 et la protéine ErbB2 tout comme chez le rat où Muc4 (SMC) interagit avec la protéine $p185^{neu}$.

SMC est le premier ligand spécifique décrit chez le rat pour $p185^{neu}$. Leur interaction se traduit par la phosphorylation et donc l'activation du dimère. Les effets biologiques qui peuvent en découler ne sont pas encore connus.

ErbB2 appartient à la même famille de molécules que le récepteur à l'EGF. Aucun ligand naturel de ErbB2 chez l'homme n'est décrit dans la littérature. La famille des récepteurs à tyrosine kinase comprend 4 membres : le récepteur de l'EGF ou ErbB1, ErbB2, ErbB3 et ErbB4. Les récepteurs de type ErbB sont exprimés par les épithéliums, le mésenchyme et les tissus nerveux. ErbB1 et ErbB2 sont impliqués dans le développement de nombreux types de cancers (Hynes and Stern, 1994) (Salomon et al., 1995).

Les mécanismes qui permettent la transduction d'un signal extracellulaire à travers la membrane sont relativement bien connus (Ullrich and Schlessinger, 1990) (van der Geer et al., 1994). L'interaction spécifique d'un ligand à un récepteur membranaire est suivie rapidement par la phosphorylation du récepteur. Le recrutement de molécules cytoplasmiques qui reconnaissent spécifiquement la tyrosine phosphorylée du récepteur initie les événements de la cascade de transduction du signal (Ullrich and Schlessinger, 1990) (Fantl et al., 1993)). Cette voie de transduction du signal est dite verticale.

La voie latérale de transduction du signal est beaucoup moins bien connue. Cette voie latérale représente la somme des mécanismes moléculaires qui favorisent la propagation latérale du signal et permettent la transactivation des récepteurs membranaires aussi bien que leur activité tyrosine kinase. La transduction latérale semble être régulée par la dimérisation des récepteurs (Heldin, 1995). La dimérisation ne concerne que les membres d'une même famille de récepteurs.

L'un des meilleurs exemples de la voie latérale du signal est fourni par les membres de la famille ErbB. Un grand nombre de facteurs de croissance sont des ligands pour les récepteurs de la famille ErbB (Riese and Stern, 1998) (Salomon et al., 1995). Ces ligands peuvent être classés en 3 groupes :

- l'EGF, l'amphiréguline et le TGFα (Transforming Growth Factor) se lient avec ErbB1

- la bétacelluline, l'épiréguline et le facteur de croissance de type EGF se liant à l'héparine (heparin binding EGF-like growth factor) se lient aussi bien à ErbB1 qu'à ErbB4

- toutes les isoformes du NDF (Neu Differentiation Factor ou héréguline) se lient à ErbB3 et ErbB4.

Aucun ligand n'est décrit chez l'homme comme pouvant permettre l'homodimérisation de ErbB2. Cependant, l'homodimérisation de ErbB2 peut être obtenue après mutation d'un seul résidu d'acide aminé dans son domaine transmembranaire (Bargmann et al., 1986) ou par interaction avec un anticorps qui reconnaît son domaine extracellulaire (Harwerth et al., 1993). Cette homodimérisation permet d'activer ErbB2. L'activation de ErbB2 peut être inhibée par l'action de la phosphatase (Weiss et al., 1997), par phosphorylation de résidus de sérine ou thréonine (Davis and Czech, 1984) (Decker, 1984) ou après réinternalisation du récepteur (Sorkin and Waters, 1993).

La réponse biologique liée à l'activation de ErbB2 peut être une transformation (Bargmann et al., 1986), une inhibition de la croissance (Harwerth et al., 1993), (Hudziak et al., 1989), une stimulation de la croissance (Beerli et al., 1996) (Graus-Porta et al., 1997) ou l'apoptose (Daly et al., 1997). La cascade verticale de transduction du signal semble donc être spécifique selon le type de mécanisme qui permet l'activation.

Bien qu'aucun ligand spécifique ne permette son homodimérisation, ErbB2 apparaît être le partenaire préférentiel pour l'hétérodimérisation des membres de la famille ErbB (Graus-Porta et al., 1997) (Karunagaran et al., 1996) (Tzahar et al., 1996). L'homo ou l'hétérodimérisation des récepteurs de la famille ErbB se traduit par la phosphorylation d'un résidu de tyrosine spécifique de leur domaine cytoplasmique. La tyrosine phosphorylée forme un site d'interaction protéine-protéine de type SH2 (Src homology 2). Ce site permet l'interaction avec des protéines possédant un site PTB (phosphotyrosine binding) telles que Shc, Grb2 et la sous unité p85 de la phosphatidylinositol kinase (Cohen et al., 1995) (Kavanaugh and Williams, 1994), ce qui permet de produire un signal mitogène ou de différenciation.

De nombreux types de cancers sont connus pour surexprimer ErbB2 comme le cancer du côlon (Kapitanovic et al., 1997) du poumon (non à petites cellules) (Yu et al., 1997), de l'ovaire (Meden and Kuhn, 1997), du sein (Slamon et al., 1989) et de l'utérus (Costa et al., 1995). Ces mêmes cancers sont connus pour surexprimer MUC4.

La transfection de la lignée A375 issue de mélanome ou de la lignée MCF7 issue de cancer du sein par l'ADNc codant SMC placé sous le contrôle d'un promoteur inductible montre que suite à la surexpression de SMC, les cellules s'arrondissent avec perte de l'adhérence cellule-cellule et cellule-matrice extracellulaire (Mcneer et al., 1997). Ce phénomène est réversible par inhibition de l'expression de SMC.

Récemment, une étude a permis d'établir le profil d'expression de *SMC* dans différents tissus d'origine épithéliale (Tableau 5) (Rossi et al., 1996). Une forme soluble de SMC est détectée dans certains tissus.

tissu	niveau d'expression	isoforme	régulation	fonction potentielle
glande mammaire/lait	++++	membranaire soluble	post-transcriptionnelle	ligand, protection
tractus digestif	+++	soluble	constitutive ?	ligand, lubrification, protection
tractus respiratoire	++	membranaire soluble	constitutive ?	protection
utérus	+++	membranaire soluble	transcriptionnelle	protection, anti-implantation

Tableau 5 : profil d'expression de *SMC*.

SMC s'exprime dans la glande mammaire à un niveau 100 fois moins élevé que celui détecté dans la lignée 13762 (Rossi et al., 1996). *SMC* n'est pas exprimé dans les glandes mammaires de rates vierges. Faiblement exprimé dans la glande mammaire de la rate non gestante et en début de gestation, son expression augmente fortement au milieu de la phase de gestation et atteint son maximum en fin de gestation et en phase *post partum*. Ce profil d'expression est comparable à celui de la β caséine et de $p185^{neu}$.

SMC est présent dans la glande mammaire aussi bien sous forme membranaire (75 %) que sous forme soluble (25 %). Grâce à sa structure en filament rigide, l'isoforme membranaire de SMC permettrait l'ouverture des canaux impliqués dans la sécrétion du lait.

Son isoforme soluble serait quant à elle impliquée dans l'aseptie du tractus digestif du nouveau-né. En effet, les chaînes glycanniques des mucines du lait permettent la liaison de bactéries pathogènes (Patton et al., 1995) (Schroten et al., 1992). Ces mucines, résistantes à la protéolyse, sont retrouvées dans les féces des nouveau-nés.

Le lait joue également un rôle majeur dans le développement intestinal des nouveau-nés. L'EGF est un des facteurs importants de ce développement. Comme la concentration de SMC dans le lait est supérieure à celle de l'EGF (Koldovsky, 1989) Rossi et al (Rossi et al., 1996) suggèrent que l'interaction de SMC avec $p185^{neu}$ puisse jouer un rôle sur le développement intestinal du nouveau-né.

Cette fonction de SMC au cours du développement intestinal pourrait débuter dès le développement du système digestif embryonnaire. L'expression de *SMC* dans le système digestif embryonnaire coïncide avec la différenciation cellulaire. *SMC* est détecté dans le pharynx dès le $14^{ème}$ jour du développement embryonnaire. Après 16,5 jours de gestation, il est détecté dans les glandes salivaires et à la surface apicale de l'épithélium stratifié non différencié de la jonction gastro-intestinale. Son expression suit alors jusqu'à la naissance la différenciation du tube digestif vers le côlon. Le niveau d'expression de SMC dans l'intestin grêle reste inchangé de l'embryon à l'adulte. SMC est localisée au niveau des cellules de Paneth présentes à la base des cryptes. Les cellules de Paneth sont décrites comme sécrétant des agents anti-microbiens (Mallow et al., 1996) et l'EGF (Raaberg et al., 1988).

Au niveau du côlon, le profil d'expression de SMC ne correspond au profil d'expression adulte qu'après la période de sevrage. Il est alors exprimé exclusivement sous forme soluble par les cellules en gobelet. Les auteurs supposent qu'après le sevrage, la seule fonction de SMC dans le côlon est de lubrifier et de protéger l'épithélium. SMC est présentée comme pouvant être un des facteurs impliqués dans la différenciation et la morphogenèse du tube digestif embryonnaire.

SMC est faiblement détectée au niveau de l'utérus avec une expression à 60 % membranaire. Chez la rate, l'expression de SMC diminue fortement 2 jours après la fécondation et disparaît complètement quand débute la phase d'implantation du blastocyte (Mcneer et al., 1998). Cette expression qui semble être hormono-dépendante semble jouer un rôle important dans l'implantation blastocytaire. Les fonctions précises de SMC sous sa forme soluble et membranaire sont encore mal connues.

Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin

Nicolas MONIAUX*, Séverine NOLLET*, Nicole PORCHET*†, Pierre DEGAND*†, Anne LAINE* and Jean-Pierre AUBERT*†[1]

*Unité 377 INSERM, Place de Verdun, 59045 Lille Cedex, France, and †Laboratoire de Biochimie et de Biologie Moléculaire de l'Hôpital C. Huriez, 59037 Lille Cedex, France

The *MUC4* gene, which encodes a human epithelial mucin, is expressed in various epithelial tissues, just as well in adult as in poorly differentiated cells in the embryo and fetus. Its N-terminus and central sequences have previously been reported as comprising a 27-residue peptide signal, followed by a large domain varying in length from 3285 to 7285 amino acid residues. The present study establishes the whole coding sequence of MUC4 in which the C-terminus is 1156 amino acid residues long and shares a high degree of similarity with the rat sialomucin complex (SMC). SMC is a heterodimeric glycoprotein complex composed of mucin (ascites sialoglycoprotein 1, ASGP-1) and transmembrane (ASGP-2) subunits. The same organization is found in MUC4, where the presence of a GlyAspProHis proteolytic site may cleave the large precursor into two subunits, MUC4α and MUC4β. Like ASGP-2, which binds the receptor tyrosine kinase p185^{neu}, MUC4β possesses two epidermal growth factor-like domains, a transmembrane sequence and a potential phosphorylated site. MUC4, the human homologue of rat SMC, may be a heterodimeric bifunctional cell-surface glycoprotein of 2.12 μm. These results confer a new biological role for MUC4 as a ligand for ErbB2 in cell signalling.

Key words: ascites sialoglycoprotein, epidermal growth factor-like domain, epithelial tissue, membrane glycoprotein.

INTRODUCTION

Originally, mucins were defined as the major highly glycosylated glycoproteins that composed the slimy and viscous secretion that covers epithelial surfaces, the mucus. These mucins, now called epithelial mucins, are thought to play an important role in the protection of the epithelial cells, and they have been implicated in epithelial renewal and differentiation [1,2]. Currently, nine human mucin genes have been identified, designated *MUC1–4*, *MUC5B*, *MUC5AC* and *MUC6–8* [3–11]. Four of these *MUC* genes, *MUC6*, *MUC2*, *MUC5AC* and *MUC5B*, are clustered between *HRAS* and *IGF2* on chromosome 11 in p15.5 [12]. These genes, which exhibit some sequence similarities with the cysteine-rich domains of the pro-von Willebrand factor, have been proposed to be derived from a common ancestral gene [13], and are believed to be the gel-forming mucins. MUC7, which does not show any cysteine-rich domain, is a soluble secreted mucin. The human epithelial mucin MUC3 is expressed in the small intestine, in both goblet cells and villus columnar cells [14,15]. MUC3 exhibits one epidermal growth factor (EGF)-like domain of which a precise role is still unknown [16]. Until now, no transmembrane region has been identified in human MUC3, although its mouse and rat homologues have one [17,18]. MUC1, which is expressed on the apical surface of most secretory epithelia [19], was the first human epithelial membrane-bound mucin identified.

The first partial cDNA from *MUC4* was isolated in our laboratory from a human tracheobronchial cDNA library [6]. MUC4 is expressed in a variety of tissues, including trachea and the bronchial area, cervix, stomach, small intestine and colon [15–20]. Like MUC3, MUC4 is not restricted to goblet cells. It is also expressed in the ciliated cells of trachea and bronchi and in absorptive cells of intestinal mucosa. Recently, the genomic organization of the 5′ region and the central part of the *MUC4* gene was determined [21]. The first exon codes the signal peptide of 27 residues and shares a high degree of similarity with that of ascites sialoglycoprotein 1 (ASGP-1), part of the rat sialomucin complex (SMC). The second exon is a large one and contains a unique sequence (951 residues) that is followed by a long tandem-repeat (TR) domain. This TR domain varies in length from 2334 to 6334 amino acid residues. This variation is due to variable number of tandem repeats (VNTR) polymorphism. The rat SMC is a well-characterized membrane-associated mucin [22]. SMC was originally isolated and characterized as a heterodimeric glycoprotein complex from highly metastatic 13762 rat mammary adenocarcinoma ascites cells, in which the mucin subunit ASGP-1 is the major detectable glycoprotein [23]. The other subunit of SMC, ASGP-2, which contains two EGF-like domains [24], has been shown to act as a ligand for the tyrosine kinase p185^{neu} [25]. The present study establishes the whole deduced coding sequence of the MUC4 C-terminus, which is a 1156-residue peptide. The MUC4 C-terminus, which shares a high degree of similarity with SMC, also possesses two EGF-like domains, a potential transmembrane sequence, a putative GlyAspProHis (GDPH) proteolytic cleavage site, two domains rich in potential N-glycosylation sites and two cysteine-rich domains. Our results allow us to conclude that MUC4 is the human homologue of rat SMC.

EXPERIMENTAL

Library screening

Total RNA was extracted from a human colon mucosa using the guanidinium isothiocyanate/CsCl method [26] and used as a template for cDNA synthesis. All details concerning double-stranded cDNA synthesis and cloning into λgt11 vector were as described by the commercial supplier, Amersham (Saclay,

Abbreviations used: SMC, sialomucin complex; EGF, epidermal growth factor; ASGP, ascites sialoglycoprotein; TR, tandem repeat; RACE, rapid amplification of cDNA ends; RT-PCR, reverse-transcriptase PCR.

[1] To whom correspondence should be addressed (e-mail jpa@lille.inserm.fr).

The nucleotide sequence data reported is in the EMBL Nucleotide Sequence Database under the accession number AJ010901.

France). Nitrocellulose membranes (Schleicher and Schüll, Ceralabo, Ecquevilly, France) were used to obtain plaque lifts. These membranes were prehybridized and hybridization was performed with 2.5×10^6 c.p.m./membrane at 42 °C overnight. Inserts of positive phages were subcloned into pBluescript KS$^+$ vector.

Cloning in pBluescript KS$^+$

Restriction enzyme digestions (*Bam*HI, *Acc*I, *Pst*I) were performed under standard conditions with the appropriate buffer on the cosmid genomic clone, LEA2, isolated previously [21]. The different fragments obtained were subcloned into pBluescript KS$^+$ vector from Stratagene (Ozyme, Saint Quentin en Yvelines, France). Subclones were sequenced using the T3 and T7 vector primers, and sequences were analysed with the GenBank® database.

Plasmid DNA purification

Qiaprep Spin Plasmid Kit (Qiagen, Courtaboeuf, France) was used according to the manufacturer's instructions.

3′-Rapid amplification of cDNA ends (RACE)-PCR procedure

Total RNA from human colon mucosa was extracted as described previously [26]. Advantage® RT-for-PCR kit (Clontech, Heidelberg, Germany) was used to synthesize first-strand cDNA from 1 μg of RNA using the oligo (dT)-anchor primer of the 5′/3′-RACE kit (Boehringer Mannheim, Roche Diagnostics, Meylan, France). Expand long PCR was performed using Expand® Long Template PCR System (Boehringer Mannheim) with the sense primer NAU 491 (5′-AGCAGGCCGAGTCTTGGATTA-3′), and as antisense primer the PCR anchor primer of the 5′/3′-RACE kit was used. The PCR amplification reaction mixture (50 μl) contained 5 μl of cDNA, 10 mM sodium dNTPs, 0.4 μM of each primer, 5 μl of 10 × Expand® Long Template PCR buffer 3, 0.75 mM $MgCl_2$ and 2.5 units of enzyme mixture. The PCR was performed using a Perkin-Elmer Thermal Cycler Gene Amp® PCR System 9700. PCR parameters were 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, annealing at 60 °C for 45 s and elongation at 71 °C for 4 min, of which the 20 last cycles had their elongation time extended by 40 s for each new cycle, followed by a final elongation at 71 °C for 15 min. Nested PCR was carried out using NAU 483 (5′-CTGTTTCTCTACCAGAGCGGT-3′) and the PCR anchor primer. The amplified product was electrophoresed on 1% TBE (1 × TBE = 45 mM Tris/borate/1 mM EDTA) agarose gel and stained with ethidium bromide. The band was cut out, purified using QIAquick Gel Extraction Kit (Qiagen), and subcloned into the Original TA Cloning® Kit (Invitrogen, Leek, The Netherlands).

Oligonucleotide primers

Oligonucleotide primers used in PCR, RACE-PCR, reverse-transcriptase PCR (RT-PCR) and sequencing experiments were synthesized by MWG-Biotech (Ebersberg, Germany). These primers were: sense NAU 139 (nt 1–22), antisense NAU 363 (nt 425–445), sense NAU 491 (nt 515–535), sense NAU 483 (nt 682–702), antisense NAU 577 (nt 1273–1293), sense NAU 576 (nt 1294–1314), sense NAU 590 (nt 1660–1680), sense NAU 591 (nt 1976–1996), sense NAU 585 (nt 2339–2359), antisense NAU 584 (nt 2582–2602), antisense NAU 555 (nt 2728–2748), sense NAU 586 (nt 2728–2748), antisense NAU 589 (nt 2910–2930), sense NAU 511 (nt 2994–3014), sense NAU 587 (nt 3213–3233), antisense NAU 535 (nt 3302–3322) and antisense NAU 533 (nt 3569–3589).

Sequencing and sequence analyses

Clones were sequenced on both strands by the dideoxy chain-termination method using [α-^{35}S]dATP with Sequenase version 2.0 (Amersham). Sequences were also determined by automatic sequencing, using internal primers with an ABI Prism model 377 XL automatic sequencer and the ABI PRISM dRhodamine terminator cycle sequencing ready reaction kit (Perkin-Elmer, Inc., Courtaboeuf, France) or using the standard vector primers, with a DNA Sequencer model 4000L LI-COR and the SequiTherm Excel® II long-read Premix DNA Sequencing Kit-LC (TEBU, Le Perray en Yvelines, France).

Analyses of nucleic acid and protein sequence data were performed using PC/GENE Software. The nucleotide sequence reported in this paper has been submitted to the EMBL Databank with accession number AJ010901.

RT-PCR amplification

RNA from human colon mucosa (1 μg) was used to perform single-strand cDNA using the Advantage® RT-for-PCR kit (Clontech) with a poly(T) primer. PCR was performed with sense NAU 139 and antisense NAU 363 as primer using a Perkin-Elmer Thermal Cycler Gene Amp® PCR System 9700. PCR parameters were 94 °C for 2 min, followed by 30 cycles at 94 °C for 30 s, annealing at 60 °C for 45 s and extension at 72 °C for 1 min, followed by a final elongation at 72 °C for 15 min. PCR were performed using 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim). The amplified product was electrophoresed on 1% TBE agarose gel and stained with ethidium bromide.

Northern-blot analysis

RNA from human colon prepared with the improved method for isolation of large RNA was used to perform Northern-blot analysis as described previously [27].

RESULTS

Isolation of the first exons downstream of the 48 bp repetitive sequence

Fragments of the previously isolated genomic clone LEA2 [21] corresponding to the region downstream of the 48 bp TR were subcloned and partially sequenced (Figure 1A). The different sequences obtained were compared with the GenBank® data base. The 3′ end of the *Acc*I–*Acc*I 2.8 kb fragment called S610 showed 78% of similarity with a region situated downstream of the rat ASGP-1 tandem repeat. S610 used as a probe and hybridized with a multiple-tissue Northern blot exhibited the same pattern of expression as that obtained with the JER64 probe (results not shown). An antisense oligonucleotide, NAU 363, was chosen from the end of S610 to perform RT-PCR on mRNA extracted from a normal human colon mucosa with the sense oligonucleotide, NAU 139, chosen in the first 21 nucleotides downstream of the 48 bp repetitive sequence. The RT-PCR procedure produced a 446 bp fragment, S1217 (Figure 1B). The sequence determined was analysed and compared with that of the genomic clone LEA2. S1217 showed 100% identity with the sequence of LEA2, in three exons dispersed along 7 kb of the cosmid clone. The first exon (E3) of 175 bp encodes a domain rich in serine, threonine and proline, the second (E4) is 134 bp long and the third (E5) is at least 137 bp long.

S1217 was used to screen a human colon mucosa cDNA library. One positive clone was isolated and named JER107 (Figure 1C). The JER107 insert consists of 920 bp, of which the

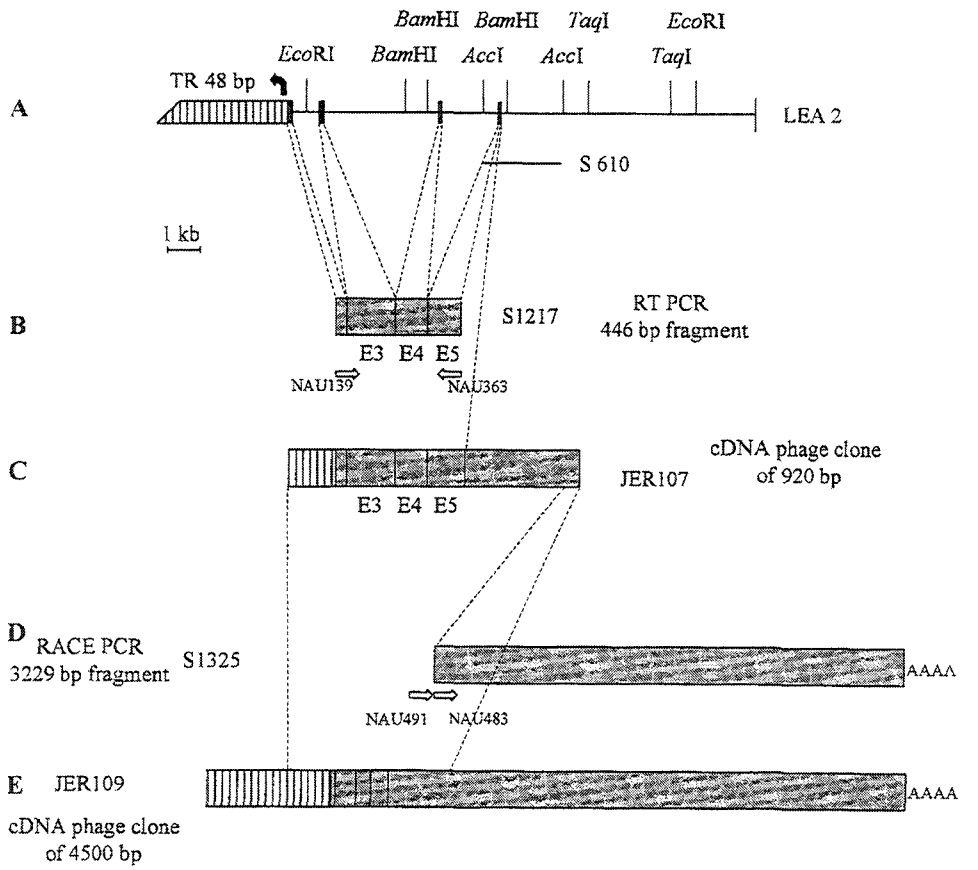


Figure 1 Map of *MUC4* 3′-terminal clones

(**A**) Partial restriction map of LEA2 pWE15 cosmid clone. The fragment called S610 was subcloned into pBluescript KS$^+$ vector. The positions of the three exons, E3, E4 and E5, are indicated by black boxes. Some primers and their directions are indicated (not to scale) by horizontal arrows. (**B**, **C**, **D**, **E**) Different cDNA clones isolated by RT-PCR, library screening and RACE-PCR.

first 70 bp is 48 bp repetitive sequence; the following 446 bp show 100% identity with the S1217 sequence and extend the exon E5 by 28 bp. Comparison between JER107 and the cosmid LEA2 sequences reveals the presence of at least four introns, I2–I5, of which three contain sequences repeated in tandem. The first is in I3 and consists of the 15 bp TR isolated previously [21]. The second in I4 is a novel 26–32 bp imperfect TR. The third TR in I5 is a 32 bp nearly perfect TR.

Extension of the 3′-terminus cDNA of *MUC4*

One sense primer, NAU 491, was chosen in the 3′ end of the phage clone JER107 to extend the sequence by 3′-RACE-PCR on human colonic mucosa total RNA. One 3396 bp cDNA fragment was obtained. Another sense primer, NAU 483, chosen in JER107, was used to perform a nested PCR on the 3396 bp fragment. One cDNA of 3229 bp was obtained and named S1325 (Figure 1D). The first 169 bp of S1325 show 100% identity with the 3′ end of JER107. JER107 was also used as a probe to screen the colon cDNA library. One positive clone, called JER109, was isolated (Figure 1E), which overlaps with S1325. It was sequenced in its entirety.

Analysis of the nucleotide and deduced amino acid sequences of the *MUC4* 3′-end cDNA

The compiled nucleotide sequences of the different cDNA clones isolated allowed us to establish the whole coding sequence of the *MUC4* 3′-terminus and its junction with the large 48 bp TR region. The unique sequence downstream of the 48 bp TR consists of a 3468 bp sequence that encodes a 1156-residue peptide (Figure 2) followed by a 3′-untranslated region of 405 bp. This compiled nucleotide sequence shows a high degree of similarity with the C-terminal sequence of the well-characterized rat membrane mucin called SMC, and can be subdivided in 13 regions, which encode 12 distinct domains (Table 1). Structural organizations of the C-termini of both peptides (MUC4 and SMC) are very similar (Figure 3).

The first four domains (CT1–CT4) are separated from the others by a putative GDPH proteolytic cleavage site. An identical proteolytic cleavage site exists in SMC between ASGP-1 and ASGP-2. Except CT1, all the C-terminal MUC4 domains exhibit sequence similarity to the corresponding domains of ASGP-1 or ASGP-2.

CT1 encodes a mucin-like domain comprising 12.5% serine, 23% threonine and 16% proline. This sequence is different from the unique domain rich in serine, threonine and proline of the MUC4 N-terminus described previously [21], or from the 16-amino-acid TR domain.

CT2 encodes a unique non-mucin type sequence, which shows a high degree of similarity with a 3′ region of ASGP-1. The degree of identity with ASGP-1 is higher at the nucleotide level. The similarity between the two molecules is particularly striking, about 60%, if we consider amino acids 195–284 of MUC4 and amino acids 1972–2061 of ASGP-1 [28].

```
1    CCT CTG AAG ATG GAA ACA TCA GGA ATG ACA ACA CCG TCA CTG AAG ACA GAC GGT GGG AGA
       P   L   K   M   E   T   S   G   M   T   T   P   S   L   K   T   D   G   G   R   20
61   CGC ACA GCC ACA TCA CCA CCC CCC ACA ACC TCC CAG ACC ATC ATT TCC ACC ATT CCC AGC
       R   T   A   T   S   P   P   P   T   T   S   Q   T   I   I   S   T   I   P   S   40
121  ACT GCC ATG CAC ACC CGC TCC ACA GCT GCC CCC ATC CCC ATC CTG CCT GAG AGA GGA GTT
       T   A   M   H   T   R   S   T   A   A   P   I   P   I   L   P   E   R   G   V   60
181  TCC CTC TTC CCC TAT GGG GCA GAC GCC GGG GAC CTG GAG TTC GTC AGG AGG ACC GTG GAC
       S   L   F   P   Y   G   A   D   A   G   D   L   E   F   V   R   R   T   V   D   80
241  TTC ACC TCC CCA CTC TTC AAG CCG GCG ACT GGC TTC CCC CTT GGC TCC TCT CTC CGT GAT
       F   T   S   P   L   F   K   P   A   T   G   F   P   L   G   S   S   L   R   D   100
301  TCC CTC TAC TTC ACA GAC AAT GGC CAG ATC ATC TTC CCA GAG TCA GAC TAC CAG ATT TTC
       S   L   Y   F   T   D   N   G   Q   I   I   F   P   E   S   D   Y   Q   I   F   120
361  TCC TAC CCC AAC CCA CTC CCA ACA GGC TTC ACA GGC CGG GAC CCT GTG GCC CTG GTG GCT
       S   Y   P   N   P   L   P   T   G   F   T   G   R   D   P   V   A   L   V   A   140
421  CCG TTC TGG GAC GAT GCT GAC TTC TCC ACT GGT CGG GGG ACC ACA TTT TAT CAG GAA TAC
       P   F   W   D   D   A   D   F   S   T   G   R   G   T   T   F   Y   Q   E   Y   160
481  GAG ACG TTC TAT GGT GAA CAC AGC CTG CTA GTC CAG CAG GCC GAG TCT TGG ATT AGA AAG
       E   T   F   Y   G   E   H   S   L   L   V   Q   Q   A   E   S   W   I   R   K   180
541  ATC ACA AAC AAC GGG GGC TAC AAG GCC AGG TGG GCC CTA AAG GTC ACG TGG GTC AAT GCC
       I   T   N   N   G   G   Y   K   A   R   W   A   L   K   V   T   W   V   N   A   200
601  CAC GCC TAT CCT GCC CAG TGG ACC CTC GGG AGC AAC ACC TAC CAA GCC ATC CTC TCC ACG
       H   A   Y   P   A   Q   W   T   L   G   S   N   T   Y   Q   A   I   L   S   T   220
661  GAC GGG AGC AGG TCC TAT GCC CTG TTT CTC TAC CAG AGC GGT GGG ATG CAG TGG GAC GTG
       D   G   S   R   S   Y   A   L   F   L   Y   Q   S   G   G   M   Q   W   D   V   240
721  GCC CAG CGC TCA GGC AAC CCG GTG CTC ATG GGC TTC TCT AGT GGA GAT GGC TAT TTC GAA
       A   Q   R   S   G   N   P   V   L   M   G   F   S   S   G   D   G   Y   F   E   260
781  AAC AGC CCA CTG ATG TCC CAG CCA GTG TGG GAG AGG TAT CGC CCT GAT AGA TTC CTG AAT
       N   S   P   L   M   S   Q   P   V   W   E   R   Y   R   P   D   R   F   L   N   280
841  TCC AAC TCA GGC CTC CAA GGG CTG CAG TTC TAC AGG CTA CAC CGG GAA GAA AGG CCC AAC
       S   N   S   G   L   Q   G   L   Q   F   Y   R   L   H   R   E   E   R   P   N   300
901  TAC CGT CTC GAG TGC CTG CAG TGG CTG AAG AGC CAG CCT CGG TGG CCC AGC TGG GGC TGG
       Y   R   L   E   C   L   Q   W   L   K   S   Q   P   R   W   P   S   W   G   W   320
961  AAC CAG GTC TCC TGC CCT TGT TCC TGG CAG CAG GGA CGA CGG GAC TTA CGA TTC CAA CCC
       N   Q   V   S   C   P   C   S   W   Q   Q   G   R   R   D   L   R   F   Q   P   340
1021 GTC AGC ATA GGT CGC TGG GGC CTC GGC AGT AGG CAG CTG TGC AGC TTC ACC TCT TGG CGA
       V   S   I   G   R   W   G   L   G   S   R   Q   L   C   S   F   T   S   W   R   360
1081 GGA GGC GTG TGC TGC AGC TAC GGG CCC TGG GGA GAG TTT CGT GAA GGC TGG CAC GTG CAG
       G   G   V   C   C   S   Y   G   P   W   G   E   F   R   E   G   W   H   V   Q   380
1141 CGT CCT TGG CAG TTG GCC CAG GAA CTG GAG CCA CAG AGC TGG TGC TGC CGC TGG AAT GAC
       R   P   W   Q   L   A   Q   E   L   E   P   Q   S   W   C   C   R   W   N   D   400
1201 AAG CCC TAC CTC TGT GCC CTG TAC CAG CAG AGG CGG CCC CAC GTG GGC TGT GCT ACA TAC
       K   P   Y   L   C   A   L   Y   Q   Q   R   R   P   H   V   G   C   A   T   Y   420
1261 AGG CCC CCA CAG CCC GCC TGG ATG TTC GGG GAC CCC CAC ATC ACC ACC TTG GAT GGT GTC
       R   P   P   Q   P   A   W   M   F   G   D   P   H   I   T   T   L   D   G   V   440
1321 AGT TAC ACC TTC AAT GGG CTG GGG GAC TTC CTG CTG GTC GGG GCC CAA GAC GGG AAC TCC
       S   Y   T   F   N   G   L   G   D   F   L   L   V   G   A   Q   D   G   N   S   460
1381 TCC TTC CTG CTT CAG GGC CGC ACC GCC CAG ACT GGC TCA GCC CAG GCC ACC AAC TTC ATC
       S   F   L   L   Q   G   R   T   A   Q   T   G   S   A   Q   A   T   N   F   I   480
1441 GCC TTT GCG GCT CAG TAC CGC TCC AGC AGC CTG GGC CCC GTC ACG GTC CAA TGG CTC CTT
       A   F   A   A   Q   Y   R   S   S   S   L   G   P   V   T   V   Q   W   L   L   500
1501 GAG CCT CAC GAC GCA ATC CGT GTC CTG CTG GAT AAC CAG ACT GTG ACA TTT CAG CCT GAC
       E   P   H   D   A   I   R   V   L   L   D   N   Q   T   V   T   F   Q   P   D   520
1561 CAT GAA GAC GGC GGA GGC CAG GAG ACG TTC AAC GCC ACC GGA GTC CTC CTG AGC CGC AAC
       H   E   D   G   G   G   Q   E   T   F   N   A   T   G   V   L   L   S   R   N   540
1621 GGC TCT GAG GTC TCG GCC AGC TTC GAC GGC TGG GCC ACC GTC TCG GTG ATC GCG CTC TCC
       G   S   E   V   S   A   S   F   D   G   W   A   T   V   S   V   I   A   L   S   560
1681 AAC ATC CTC CAC GCC TCC GCC AGC CTC CCG CCC GAG TAC CAG AAC CGC ACG GAG GGG CTC
       N   I   L   H   A   S   A   S   L   P   P   E   Y   Q   N   R   T   E   G   L   580
1741 CTG GGG GTC TGG AAT AAC AAT CCA GAG GAC GAC TTC AGG ATG CCC AAT GGC TCC ACC ATT
       L   G   V   W   N   N   N   P   E   D   D   F   R   M   P   N   G   S   T   I   600
1801 CCC CCA GGG AGC CCT GAG GAG ATG CTT TTC CAC TTT GGA ATG ACC TGG CAG ATC AAC GGG
       P   P   G   S   P   E   E   M   L   F   H   F   G   M   T   W   Q   I   N   G   620
1861 ACA GGC CTC CTT GGC AAG AGG AAT GAC CAG CTG CCT TCC AAC TTC ACC CCT GTT TTC TAC
```

Figure 2 For legend see facing page

CT3 encodes a cysteine-rich domain comprising 11.3% cysteine. The nucleotide sequence of this domain shows 78% similarity with the ASGP-1 sequence, but the two deduced peptides are different. As in CT2, there are several changes in the reading frame. The analysis of this sequence with the GenBank® data base does not exhibit evidence of similarity with any other cysteine-rich domain and particularly with the cysteine-rich domains found in the 11p15.5 mucin gene family.

CT4 encodes a peptide which shows 64% similarity with ASGP-1 in amino acids 2189–2202. CT5 encodes a large domain that shows 60% similarity with the first subdomain of ASGP-2, which contains 16 putative N-glycosylation sites. CT5 contains

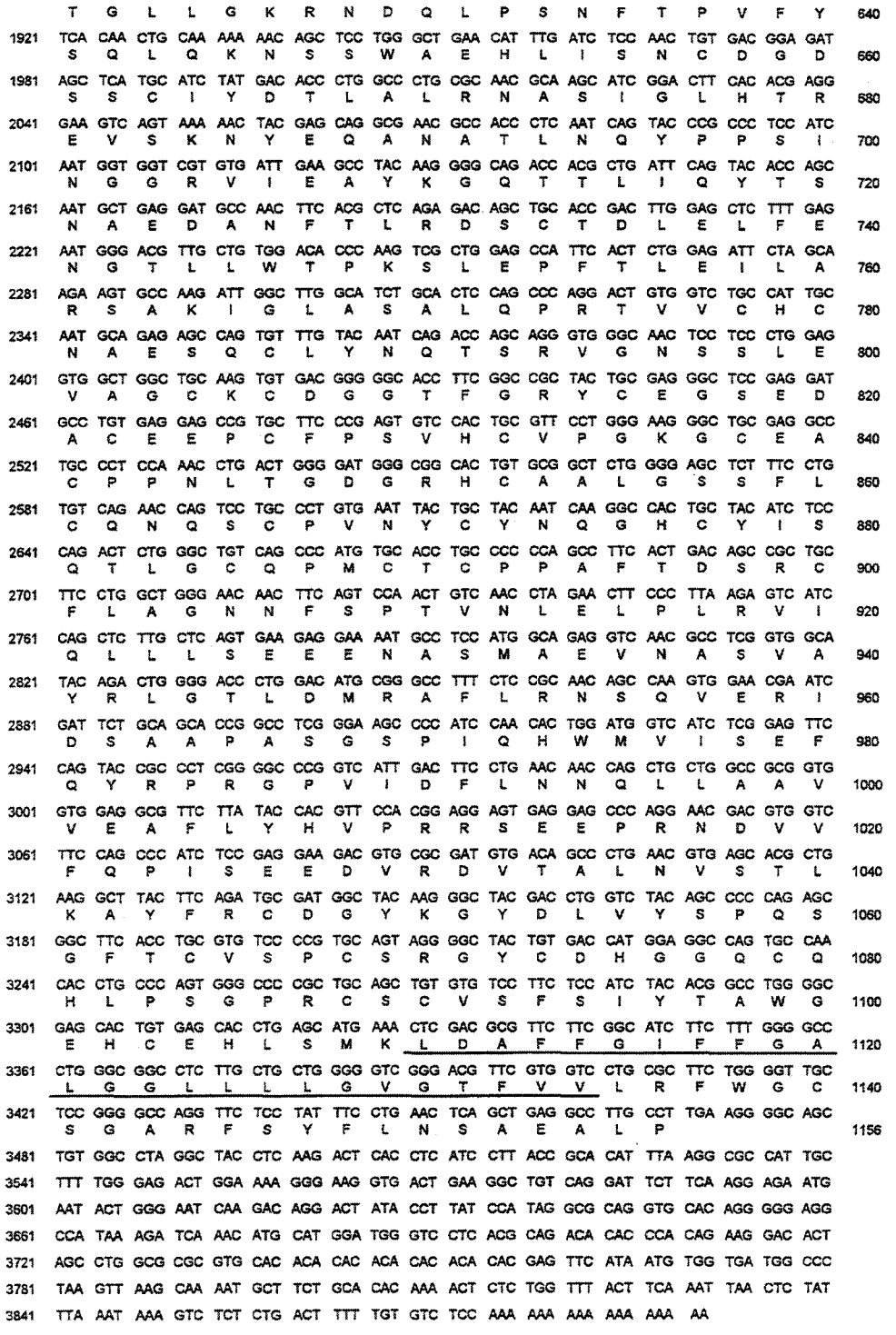


Figure 2 Compiled nucleotide sequences and deduced amino acid sequence of the 3′ terminus of *MUC4*

Nucleotide 1 corresponds to the first nucleotide just downstream of the 48 bp TR sequence. The hydrophobic stretch of amino acid residues is underlined. The nucleotides are numbered on the left and the amino acid residues are numbered on the right side of the Figure.

13 potential N-glycosylated sites, of which 10 are conserved in both peptides.

CT6 encodes another cysteine-rich domain. This domain, which contains 14.5% cysteine residues, shows 65% similarity with a cysteine-rich domain in ASGP-2. In ASGP-2, this region follows the N-glycosylation-rich domain as well. Like the N-glycosylation sites, the cysteines are conserved in both peptides. This domain contains three potential N-glycosylation sites that are also found in ASGP-2. As is the case for the cysteine-rich domain found upstream of the GDPH cleavage site, no similarity

Table 1 Position and characterization of the different domains of the MUC4 C-terminal region and their similarity with the different subunits of rat SMC

Name	Position in nucleotide	Characteristic	Similarity with SMC
CT1	1–168	Mucin-like domain	
CT2	169–912	Unique sequence	ASGP1
CT3	913–1251	Cysteine-rich domain	ASGP1
CT4	1252–1293	Unique sequence	ASGP1
	1288–1299	GDPH cleavage site	GDPH cleavage site
CT5	1294–2331	N-Glycosylated rich domain	ASGP2
CT6	2332–2580	Cysteine-rich domain	ASGP2
CT7	2581–2700	EGF1 domain	ASGP2
CT8	2701–3135	N-Glycosylated rich domain	ASGP2
CT9	3136–3270	EGF2 domain	ASGP2
CT10	3271–3327	Unique sequence	ASGP2
CT11	3328–3401	Transmembrane domain	ASGP2
CT12	3402–3468	Cytoplasmic tail	ASGP2
CT13	3469–3873	3′ Untranslated sequence	

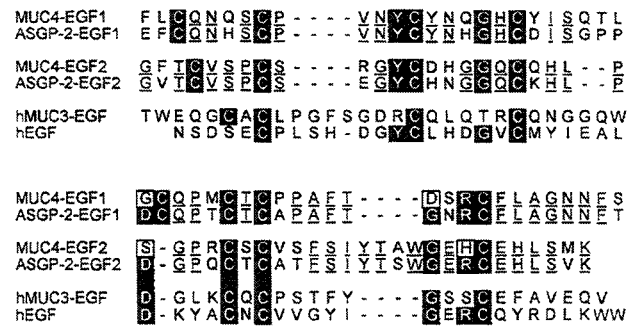


Figure 4 Comparison of EGF-like domains of MUC4β with those of rat ASGP-2, human MUC3 and human EGF

Highly conserved cysteine residues and other essential residues are in white lettering on a dark background and conserved residues between SMC and MUC4 are underlined. Non-conserved essential amino acids are boxed.

exists with the cysteine-rich domains in the *MUC* genes in the 11p15.5 mucin family.

CT7 and CT9 are two EGF-like domains. Comparison between different EGF-like domains is shown in Figure 4. The similarity between both peptides (MUC4 and ASGP-2) is, respectively, 68% for EGF1 and 67% for EGF2. The positions of all the cysteine residues are identical in both molecules, with a $CX_4CX_4CX_5CX_7CX_3CXCX_8C$ motif for EGF1 and a $CX_{17}CX_3CX_4CX_5CX_8CXCX_{12}C$ motif for EGF2 (where X denotes any other residue). Moreover, there is a putative N-glycosylation site in position 863, which is also found in ASGP-2. It is important to note that one aspartic acid and one glycine found in most of the EGF-like domains are replaced, respectively, by one glycine in position 884 and by one aspartic acid in position 887 in MUC4 EGF1. In MUC4 EGF2, one aspartic acid and one arginine are replaced, respectively, by a serine in position 1084 and by a histidine in position 1102. Moreover, MUC4 EGF1 possesses the supplementary cysteine residue found in ASGP-2 EGF1 (second block in Figure 4). The motif found in MUC4 and ASGP-2 EGF-like domains is CX_7CX_3C instead of the $CX_{10}C$ motif that is usually found in the other EGF-like domains. As in ASGP-2, a domain of 147 amino acid residues (CT8) separates the two EGF-like domains. This domain contains four potential N-glycosylation sites that are conserved in ASGP-2.

CT10, encodes a domain that shows 83% similarity with ASGP-2. This domain separates MUC4 EGF2 from a very

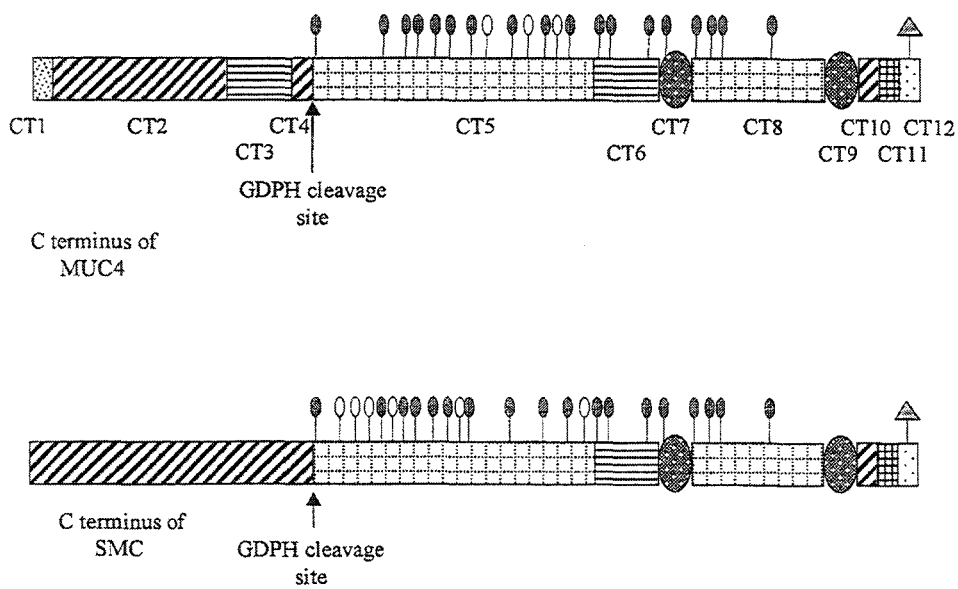


Figure 3 Schematic representation of human MUC4 and rat SMC C-termini

Dense dots, serine/threonine-rich non-repetitive sequence domain; diagonal lines, unique sequence; horizontal lines, cysteine-rich domain; dotted grid, domain rich in potential N-glycosylation sites; hatched ovals, EGF-like domain; solid grid, potential transmembrane sequence; spaced dots, potential cytoplasmic tail; and (on stalks above sequence) hatched ovals, conserved potential N-glycosylation sites; open ovals, non-conserved potential N-glycosylation sites; hatched triangles, potential phosphorylated sites.

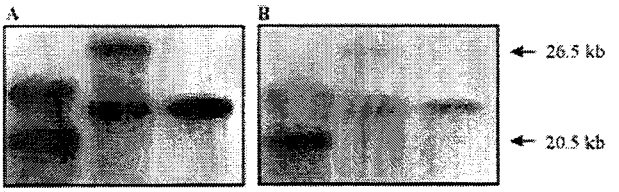


Figure 5 Comparison between Northern-blot patterns obtained with the JER64 and S1325 probes

Total RNA prepared from three individual colons was hybridized with the JER64 (**A**) and S1325 (**B**) probes.

hydrophobic domain (CT11) [29]. CT11 shows 63% similarity with the transmembrane sequence of ASGP-2.

The last coding region, CT12, shows 55% similarity with the cytoplasmic tail of ASGP-2. It does not possess the palmitoylation site CXC found in ASGP-2 [24] and MUC1 [30]. This domain contains one tyrosine residue at position 1147. A tyrosine at the same position in the cytoplasmic tail of ASGP-2 is suspected to be a putative phosphorylation site.

CT13, a 3′-untranslated region of 405 bp, contains two potential polyadenylation signals (AAATTAA and AAATAAA). This region does not share any similarity with the 3′-untranslated region of *ASGP-2* except for a 10 CA motif repeated in tandem in both cDNAs.

RNA analysis

The whole 3′-end cDNA fragment (S1325) was used as a probe to hybridize a Northern blot of three individuals' colonic mucosae (prepared with improved method for isolation of large RNA [27]). This fragment revealed the same double bands that were revealed with the JER64 probe (Figure 5).

DISCUSSION

MUC4, located on chromosome 3 in the q29 region [31], encodes a human epithelial mucin that is detected in various epithelial tissues in adult but also in poorly differentiated cells in embryo and fetus [32,33]. Thus MUC4 is expressed early in the primitive gut, before respiratory and digestive epithelial cells have acquired their tissue and cell specificity. Moreover, abnormal expression of MUC4 has been reported in various cancers, such as in pancreatic [34,35] and colon carcinomas [36]. These observations suggest that several distinct functions might be fulfilled by this mucin. To approach these functions, we have determined the complete sequence of *MUC4* cDNA and deduced the peptide organization (Figure 6). Its N-terminus and central sequences have previously been reported [21] as a 27-residue peptide signal, followed by a large domain varying in length from 3285 to 7285 amino acid residues. The C-terminal region of MUC4 shows a very high degree of similarity with the rat heterodimeric glycoprotein complex [22] called SMC. SMC consists of a cell-surface sialomucin (ASGP-1) of 600 kDa associated in a non-covalent manner with an 80 kDa cell-membrane-bound peptide (ASGP-2). Both subunits are translated by a unique cDNA. A GDPH proteolytic cleavage site is present in both peptides in the same position. This suggests that the MUC4 precursor could be cleaved into two subunits and form, as with SMC, a heterodimeric complex. The subunit upstream of the GDPH cleavage site is now called MUC4α and the unit downstream is called MUC4β. Another well-characterized mucin, MUC1, is synthesized on the cell surface as a heterodimeric complex, both subunits originating from a single apomucin precursor [37]. MUC1 is a transmembrane protein [38] for which a soluble form has been reported to be present in cell-culture media and body fluids [39,40]. Although the nucleotide sequence downstream of the 48 bp repeat of MUC4α exhibits similarity with ASGP-1, both peptide domains are different except for regions from amino acid 195 to 284 and from 1252 to 1293. The differences are due to several changes in the reading frame between both apomucins. Thus, a cysteine-rich domain, present in MUC4α, is absent in ASGP-1.

MUC4β subunit is closely related to ASGP-2. Indeed the structural organization of both apomucins is identical and both peptide sequences show more than 60% similarity. These results suggest that SMC could be considered as the rat homologue of human MUC4. MUC4β is rich in potential N-glycosylation sites, of which 18 out of 21 are conserved within ASGP-2. As in SMC, the hydropathy profile of MUC4β reveals a hydrophobic region of about 24 amino acid residues, which represents a potential membrane-spanning domain. Thus the MUC4 complex, like SMC, is probably also a heterodimeric membrane-associated

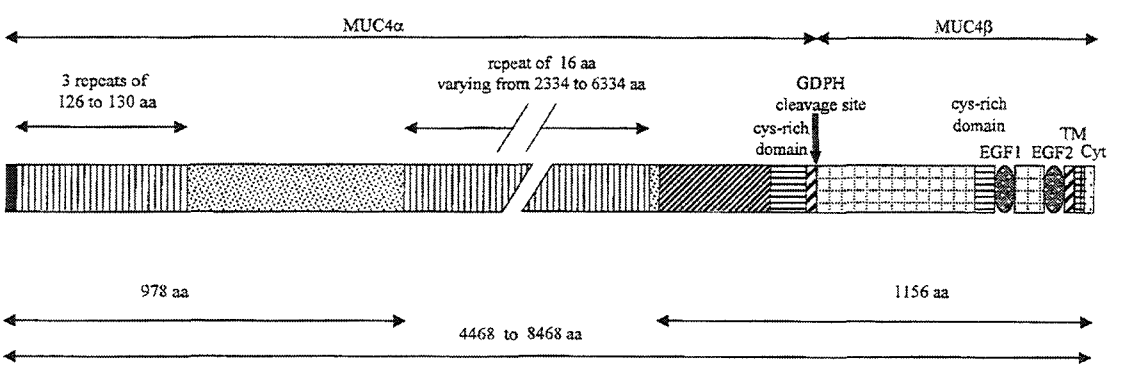


Figure 6 Schematic representation of the structure of MUC4

Black, peptide signal; vertical lines, TR; dense dots, serine, threonine-rich non-repetitive sequence domain; diagonal lines, unique sequence; horizontal lines, cysteine-rich domain; dotted grid, domain rich in potential N-glycosylation sites; hatched ovals, EGF-like domain; TM, potential transmembrane sequence; Cyt, potential cytoplasmic tail.

mucin with its N-terminus orientated extracellularly. SMC is present in different rat tissues as both soluble and membrane-associated forms. For instance, SMC is expressed as two isoforms in the mammary gland and milk; a membrane-associated form (75%) and a soluble or secreted form (25%) [41]. Since no evidence of alternative splicing had been observed, a proteolytic cleavage event was suggested to be responsible for the generation of the soluble form. Our knowledge of MUC4 expression in absorptive and ciliated cells as well as goblet cells suggests that MUC4 could exist as both membrane-associated and secreted forms.

Similar EGF-like domains are found in MUC4β and in ASGP-2. The ASGP-2 EGF1 is considered to interact with the tyrosine kinase p185^{neu}, which is the rat homologue of the proto-oncogene c-ErbB2. ASGP-2 and p185^{neu} are co-immunoprecipitable from cell-surface fractions, and a complex of ASGP-2 and p185^{neu} extracellular domains is formed and secreted from insect cells when the two are co-infected [25]. p185^{neu} shows similarity with the EGF receptor [42], but does not bind EGF. No other p185^{neu} real ligand has been reported. ErbB2 is a member of the class-I EGF receptor tyrosine kinase family, a family of four members, ErbB1–ErbB4. Lupu and colleagues reported a putative ligand, the gp30, that presumably interacts directly with ErbB2 [43]. However, it was not proven that the activity corresponds to a direct ErbB2 ligand. Even without a ligand of its own, ErbB2 can undergo activation by heterologous ligands. EGF and Neu differentiation factor (NDF or heregulin) have been shown to activate the phosphorylation of ErbB2 through the formation of heterodimers, respectively, between ErbB1 and ErbB3 or ErbB4 [44,45]. The ErbB2 gene product is overexpressed in many human cancers, including colorectal [46], non-small-cell lung [47], ovarian [48], breast [49] and uterine cervix carcinoma [50]. It is also expressed in a tissue- and developmental-stage-specific manner [51]. It turned out that the expression pattern of MUC4 is very similar to that of ErbB2 [32,33,47]. In non-small-cell lung cancer, sialomucin expression is associated with ErbB2 overexpression [47]. The heterodimeric membrane-associated isoform of MUC4 could be (as is the case with SMC for p185^{neu}) the natural ligand of the proto-oncogene c-ErbB2. Regulation of ErbB2 receptor activity appears to be very complex. The formation of a MUC4/ErbB2 complex or ErbB2/ErbB1, ErbB2/ErbB3 and ErbB2/ErbB4 may serve to diversify the nature of the intracellular signal elicited by ErbB2. Thus, MUC4 may be a heterodimeric bifunctional cell-surface glycoprotein complex. Recently, MUC1 has been described as a bifunctional cell-surface glycoprotein too [52]. The MUC4 complex, which is very rich in potential O- and N- glycosylation sites, has an extended structure. According to Jentoft [53], the glycosylated polypeptide of 20 amino acid residues is approximately 5 nm long. The MUC4 TR domain varies from 2334 to 6334 residues, so the size of the extended apomucin MUC4 complex varies from 4468 to 8468 residues. This means that MUC4 extends at least 1.12–2.12 μm above the cell membrane, far above all other membrane-associated proteins. For instance, MUC1, which is considered as the largest membrane-associated glycoprotein, extends from 200 to 500 nm [53]. With such size and its putative bifunctionality, MUC4 could be considered as an essential cell membrane-associated glycoprotein, involved in cell–cell communication and the adhesion cascade. Like SMC for p185^{neu}, the MUC4 complex could be involved in a signalling pathway that is required for proliferation and differentiation of epithelial cells.

This work was supported by l'Association de Recherche contre le Cancer and by le Comité du Nord de la Ligue contre le Cancer. N.M. is a recipient of l'Association de Recherche contre le Cancer. We gratefully acknowledge P. Mathon, M. Crépin and C. Mouton for performing automatic sequences, and A. Leclercq and C. Mouton for performing polymorphism analysis. We thank the members of our E.U. consortium CEEBMH4-CT98-3222 for stimulating discussion.

REFERENCES

1 Guzman, K., Bader, T. and Nettesheim, P. (1996) Am. J. Physiol. **270**, L846–L853
2 Braga, V. M. M., Pemberton, L. F., Duhig, T. and Gendler, S. J. (1992) Development **115**, 427–437
3 Lan, M. S., Batra, S. K., Qi, W. N., Metzgar, R. S. and Hollingworth, M. A. (1990) J. Biol. Chem. **265**, 15294–15299
4 Gum, Jr., J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) J. Biol. Chem. **269**, 2440–2446
5 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. E., Byrd, J. C., Lamport, D. T. A., Siddiki, B. and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. **171**, 407–415
6 Porchet, N., Nguyen, V. C., Dufossé, J., Audié, J. P., Guyonnet Dupérat, V., Gross, M. S., Denis, C., Degand, P., Berheim, A. and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. **175**, 414–422
7 Dufossé, J., Porchet, N., Audié, J. P., Guyonnet Dupérat, V., Laine, A., Van Seuningen, I., Marrakchi, S., Degand, P. and Aubert, J. P. (1993) Biochem. J. **293**, 329–337
8 Aubert, J. P., Porchet, N., Crépin, M., Duterque-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. (1991) Am. J. Respir. Cell. Mol. Biol. **5**, 178–185
9 Toribara, N. W., Roberton, A. M., Ho, S. B., Kuo, W. M., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) J. Biol. Chem. **268**, 5879–5885
10 Bobek, L. A., Liu, J., Sait, S. N. J., Shows, T. B., Bobek, Y. A. and Levine, M. J. (1996) Genomics **31**, 277–282
11 Shankar, V., Pichan, P., Eddy, Jr., R. L., Tonk, V., Nowak, N., Sait, S. N. J., Shows, T. B., Schultz, R. E., Gotway, G., Elkins, R. C., Gilmore, M. S. and Sachdev, G. P. (1997) Am. J. Respir. Cell. Mol. Biol. **16**, 232–241
12 Pigny, P., Guyonnet Dupérat, V., Hill, A., Pratt, W. S., Galiègue-Zouitinia, S., Collyn d'Hooge, M., Laine, A., Van Seuningen, I., Gum, J. R., Kim, Y. S., Swallow, D. M., Aubert, J. P. and Porchet, N. (1996) Genomics **38**, 340–352
13 Desseyn, J. L., Buisine, M. P., Porchet, N., Aubert, J. P., Degand, P. and Laine, A. (1998) J. Mol. Evol. **46**, 102–106
14 Ho, S. B., Niehans, G. A., Lyftogt, C., Yan, P. S., Cherwitz, D. L., Gum, E. T., Dahira, R. and Kim, Y. S. (1993) Cancer Res. **53**, 641–651
15 Audié, J. P., Janin, A., Porchet, N., Copin, M. C., Gosselin, B. and Aubert, J. P. (1993) J. Histochem. Cytochem. **43**, 1479–1485
16 Gum, Jr., J. R., Ho, J. J. L., Pratt, W. S., Hicks, J. W., Hill, A. S., Vinall, L. E., Roberton, A. M., Swallow, D. M. and Kim, Y. S. (1997) J. Biol. Chem. **272**, 26678–26686
17 Shekels, L. L., Hunninghake, D. A., Tisdales, A. S., Gipson, I. K., Kieliszewski, M., Kozak, C. A. and Ho, S. B. (1998) Biochem. J. **330**, 1301–1308
18 Khatri, I. A., Forstner, G. G. and Forstner, J. F. (1997) Biochim. Biophys. Acta **1326**, 7–11
19 Gendler, S. J., Lancaster, C. A., Taylor Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalami, E. N. and Wilson, D. (1990) J. Biol. Chem. **265**, 15286–15293
20 Audié, J. P., Tétaert, D., Pigny, P., Buisine, M. P., Janin, A., Aubert, J. P., Porchet, N. and Boersma, A. (1995) Hum. Reprod. **10**, 98–102
21 Nollet, S., Moniaux, N., Maury, J., Petitprez, D., Degand, P., Laine, A., Porchet, N. and Aubert, J. P. (1998) Biochem. J. **332**, 739–748
22 Sherblom, A. P. and Carraway, K. L. (1980) J. Biol. Chem. **255**, 12051–12059
23 Sherblom, A. P., Buck, R. L. and Carraway, K. L. (1980) J. Biol. Chem. **255**, 783–790
24 Sheng, Z., Wu, K., Carraway, K. L. and Fregien, N. (1992) J. Biol. Chem. **267**, 16341–16346
25 Carraway, K. L., Carraway, C. A. C. and Carraway, III, K. L. (1997) J. Mammary Gland Biol. Neoplasia **2**, 187–198
26 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
27 Debailleul, V., Laine, A., Huet, G., Mathon, P., Collyn d'Hooghe, M., Aubert, J. P. and Porchet, N. (1998) J. Biol. Chem. **273**, 881–890
28 Wu, K., Fregen, N. and Carraway, K. L. (1994) J. Biol. Chem. **269**, 11950–11955
29 Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
30 Vos, H. L., de Vries, Y. and Hilkens, J. (1991) Biochem. Biophys. Res. Commun. **181**, 121–130
31 Gross, M. S., Guyonnet Dupérat, V., Porchet, N., Bernheim, A., Aubert, J. P. and Van Cong, N. (1992) Ann. Hum. Genet. **35**, 21–26
32 Buisine, M. P., Devisme, L., Savidge, T. C., Gespach, C., Gosselin, B., Porchet, N. and Aubert, J. P. (1998) Gut **43**, 519–524
33 Buisine, M. P., Devisme, L., Copin, M. C., Durand-Réville, M., Gosselin, B., Aubert, J. P. and Porchet, N. (1999) Am. J. Respir. Cell. Mol. Biol. **19**, in the press

34 Balagué, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J. P., Kim, Y. S. and Real, F. X. (1994) Gastroenterology **106**, 1054–1061
35 Balagué, C., Audié, J. P., Porchet, N. and Real, F. X. (1995) Gastroenterology **109**, 953–964
36 Ogata, S., Uehara, H., Chen, A. and Itzkowitz, S. H. (1992) Cancer Res. **52**, 5971–5978
37 Ligtenberg, M. J. L., Kruijshaar, L., Buijs, F., van Meijer, M., Litvinov, S. V. and Hilkens, J. (1992) J. Biol. Chem. **267**, 6171–6177
38 Pemberton, L., Taylor-Papadimitriou, J. and Gendler, S. J. (1992) Biochem. Biophys. Res. Commun. **185**, 167–175
39 Boshell, M., Lalani, E.-N., Pemberton, L., Burchell, J., Gendler, S. J. and Taylor-Papadimitriou, J. (1992) Biochem. Biophys. Res. Commun. **185**, 1–8
40 Burchell, J., Wang, D. and Taylor-Papadimitriou, J. (1984) Int. J. Cancer **34**, 763–768
41 Rossi, E. A., McNeer, R. R., Price-Schiavi, S. A., Van den Brande, J. M. H., Komatsu, M., Thompson, J. F., Carraway, C. A. C., Friegien, N. L. and Carraway, III, K. L. (1996) J. Biol. Chem. **271**, 33476–33485
42 Gullick, W. J. (1990) Int. J. Cancer suppl. **5**, 55–61
43 Lupu, R., Colomer, R., Zugmaier, G., Sarup, J., Slamon, D. and Lippman, M. E. (1990) Science **249**, 1552–1555
44 Wada, T., Qian, X. and Greene, M. I. (1990) Cell **61**, 1339–1347
45 Carraway, III, K. L. and Cantley, L. C. (1994) Cell **78**, 5–8
46 Kapitanovic, S., Radosevic, S., Kapitanovic, M., Andelinovic, S., Frerencic, Z., Tavassoli, M., Primorac, D., Sonicki, Z., Spaventi, S., Pavelic, K. and Spaventi, R. (1997) Gastroenterology **112**, 1103–1113
47 Yu, C.-J., Shun, C.-T., Yang, P.-C., Lee, Y.-C., Shew, J.-Y., Kuo, S.-H. and Luh, K.-T. (1997) Am. J. Respir. Crit. Care Med. **155**, 1419–1427
48 Meden, H. and Kuhn, W. (1997) Eur. J. Obstet. Gynecol. Reprod. Biol. **71**, 173–179
49 Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., Ullrich, A. and Press, M. F. (1989) Science **244**, 707–712
50 Costa, M. J., Walls, J. and Trelford, J. D. (1995) Am. J. Clin. Pathol. **104**, 634–642
51 Kokai, Y., Cohen, J. A., Drebin, J. A. and Greene, M. I. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8498–8501
52 Mockensturm-Gardner, M., Rowles, J. and Gendler, S. J. (1998) 5th International Workshop on Carcinoma-Associated Mucin, Abstract, D6
53 Jentoft, N. (1990) Trends Biochem. Sci. **15**, 291–294

Received 14 September 1998/11 November 1998; accepted 10 December 1998

II. Le gène *MUC4* génère par épissage alternatif une famille de mucines solubles et membranaires.

II.1. Détection des isoformes.

Afin de compléter le profil d'expression de *MUC4*, des expériences de RT-PCR ont été réalisées sur de l'ARN polyA+ isolé de nombreux tissus épithéliaux. Après de nombreux essais, nous avons choisi pour réaliser cette étude d'utiliser l'oligonucléotide sens NAU591 (nt 1976/1996) et l'oligonucléotide antisens NAU555 (nt 2728/2748). Ce couple d'oligonucléotides donne les meilleurs résultats d'amplification sur l'ARN de tous les tissus testés. Deux produits d'amplification sont détectés. Un produit d'amplification attendu de 772 pb est détecté à partir d'ARN commerciaux ou préparés au laboratoire, extraits de thymus, thyroïde, glandes salivaires et mammaires, trachée, poumon, œsophage, estomac, d'intestin grêle, côlon, testicule, prostate, ovaire, utérus et placenta (Figure 21).

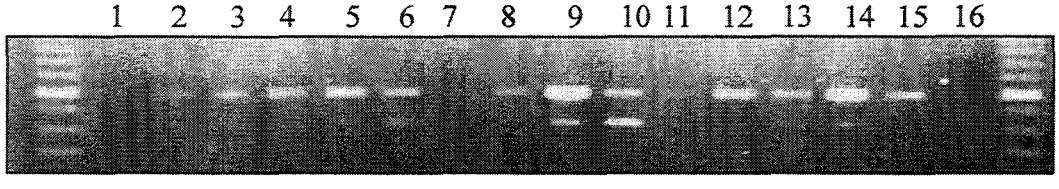


Figure 21 : RT-PCR réalisée sur l'ARN polyA+ de : 1, thymus ; 2, thyroïde ; 3, glandes mammaires ; 4, glandes salivaires ; 5, trachée ; 6, estomac ; 7, intestin grêle ; 8, utérus ; 9, prostate ; 10, testicule ; 11, côlon ; 12, œsophage ; 13, ovaire ; 14, poumon ; 15, placenta ; 16, témoin eau.

Le clonage et le séquençage du produit d'amplification de 772 pb confirme qu'il correspond au fragment attendu. Un fragment supplémentaire de 613 pb est également détecté dans la majorité des tissus testés. Cette bande supplémentaire de 613 pb a été clonée puis séquencée. La comparaison de cette séquence avec celle du fragment de 772 pb révèle la présence d'un épissage alternatif. Le fragment de 613 pb est délété des nucléotides de la position 2401 à 2560. Ces nucléotides codent le domaine riche en résidus de cystéine situé en amont du domaine EGF1 de MUC4. Bien que la technique de

RT-PCR utilisée ne soit pas quantitative, il apparaît que la forme préalablement publiée et qui correspond au fragment de 772 pb, est la plus abondante dans tous les tissus testés. Le niveau d'expression de la forme additionnelle est variable selon la source de l'ARN étudié. Son niveau d'expression maximum est détecté pour l'ARN extrait de testicules.

Pour isoler l'ADNc de l'isoforme détectée, l'idéal aurait été de choisir des oligonucléotides dans chacun des UTR (5' et 3'). Cette expérience était impossible à réaliser étant donnée la taille de l'exon 2. Comme nous n'avons jamais détecté d'événement d'épissage alternatif entre l'exon 2 et l'exon 3, un oligonucléotide sens NAU412 (nt 117/137) a été choisi dans l'exon 3 afin d'obtenir d'éventuels autres produits d'épissage alternatif par la technique d'"expand long RT-PCR". L'oligonucléotide antisens NAU533 (nt 3569/3589) est choisi dans la séquence 3' non traduite. Cette expérience a été réalisée sur l'ARN polyA+ extrait de testicules. 2 bandes principales sont détectées, d'une taille d'environ 3,5 kb (figure 22). Afin d'isoler toutes les formes présentes, la totalité du produit d'amplification est cloné sans purification préalable sur gel. 9 clones de taille différente sont isolés et séquencés (Figure 23). La figure 23 montre les différents sous-clones obtenus après hydrolyse par *Eco*RI. La bande de 4 kb correspond au vecteur linéarisé. Un site *Eco*RI est présent dans la séquence de certaines des isoformes.

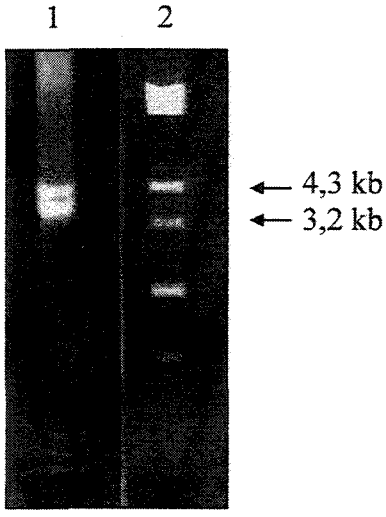


Figure 22 : Ligne 1 : expand long RT-PCR réalisée sur l'ARN polyA+ de testicule avec comme oligonucléotide NAU412 et NAU533. Ligne 2 : témoin de masse moléculaire.

Un des clones isolés (Figure 23 Ligne 1) correspond à la séquence de *MUC4* préalablement publiée (Moniaux et al., 1999). Cette isoforme est maintenant dénommée sv0-*MUC4*. Les 8 autres clones (Figure 23, ligne 2 à 9), qui résultent de la combinaison de 8 événements d'épissage alternatif (Tableau 6), codent des formes distinctes de MUC4 (Figure 24). Ces nouveaux variants sont dénommés sv1-MUC4 à sv8-MUC4.

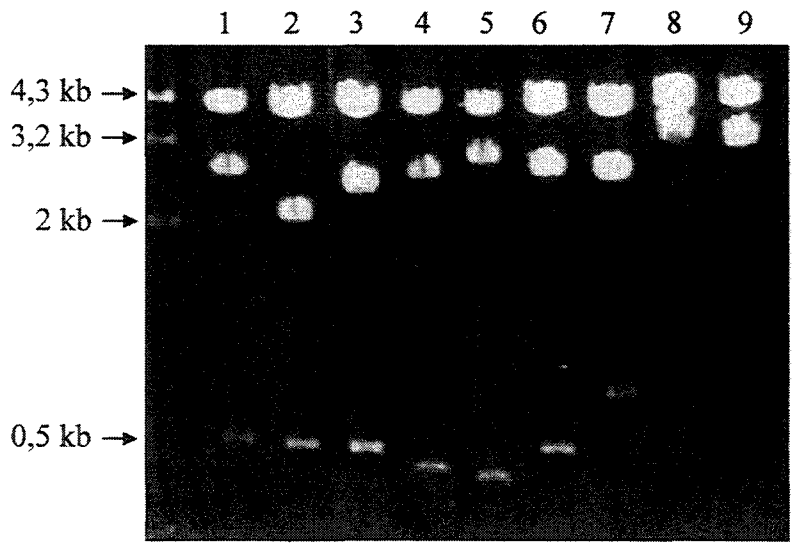


Figure 23 : Hydrolyse *Eco*RI et migration en gel d'agarose des 9 clones isolés par expand long RT-PCR. Les variants sv0-MUC4 à sv8-MUC4 sont respectivement déposés de la ligne 1 à 9.

L'extrémité 3' de l'ADNc de sv1-*MUC4* a une taille de 3938 pb, sv2-*MUC4* de 3798 pb, sv3-*MUC4* de 3637 bp, sv4-*MUC4* de 3769 pb, sv5-*MUC4* de 3708 pb, sv6-*MUC4* de 3282 pb sv7-*MUC4* de 3713 pb et sv8-*MUC4* de 3195 pb. L'extrémité 3' de l'ADNc de sv0-*MUC4*, préalablement publiée et accessible sous le numéro AJ010901, a une taille de 3873 pb. Comme les tailles des différentes isoformes sont proches, nous ne pouvons pas savoir à laquelle des isoformes correspond les deux bandes majoritaires détectées par la technique d'"expand long RT-PCR" (Figure 22, ligne 1). L'étude avec des oligonucléotides choisis tout au long de l'extrémité 3' de l'ADNc de *MUC4* et permettant d'amplifier des fragments dont la taille varie entre 400 et 600 kb nous permet de savoir que l'isoforme prépondérante dans tous les tissus testés est sv0-*MUC4*. Pour les autres isoformes, nous ne pouvons pas définir leur niveau d'expression.

II.2. Les événements d'épissage alternatif.

Les événements d'épissage alternatif observés sont résumés dans le tableau 6 et sont au nombre de 8. Les mécanismes précis de l'épissage alternatif ne peuvent être décrits que pour les événements A à C car le clone génomique LEA2 que nous possédons ne contient pas toute la séquence codante.

événement d'épissage alternatif	type	taille de l'insertion ou de la délétion en pb	cadre de lecture
A	insertion	14	modifié
B	délétion	75	modifié
C	insertion	211	modifié
D	délétion	28	modifié
E	délétion	88	modifié
F	délétion	428	non modifié
G	délétion	677	non modifié
H	délétion	159	modifié

Tableau 6 : Tableau résumant les 8 événements d'épissage alternatif ainsi que leurs caractéristiques.

L'événement A résulte de l'utilisation d'un site donneur d'épissage cryptique, et l'événement B de l'utilisation d'un site accepteur d'épissage cryptique. Avec l'événement A, l'exon 3 est plus grand de 14 pb et avec l'événement B, l'exon 5 est plus court de 75 pb. L'événement C résulte quant à lui de l'insertion d'un nouvel exon (cassette) de 211 pb. Tous les autres événements d'épissage alternatif conduisent à une délétion de la position 732 à 760 pour l'événement D, de la position 760 à 849 pour l'événement E, de la position 967 à 1395 pour l'événement F, de la position 1022 à 1699 pour l'événement G et de la position 2401 à 2560 pour l'événement H. A l'exception des événements F et G, tous les autres génèrent un changement du cadre de lecture.

Des expériences de PCR réalisées sur l'ADN génomique extrait de lymphocytes et le séquençage direct des produits d'amplification ont montré que les délétions générées par les événements F et G résultent de l'exclusion de plusieurs exons en cassette. Les

expériences de PCR et les séquences sont réalisées à partir d'oligonucléotides localisés tout au long de l'extrémité 3'-terminale de l'ADNc de MUC4. Plusieurs jonctions d'épissage ont ainsi pu être identifiées dans les larges domaines délétés suite aux événements F et G. Même si tous les sites d'épissages n'ont pu être identifiés, ces expériences montrent que les domaines délétés par les événements F et G sont composés d'au moins 3 exons.

II. 3. La famille des mucines solubles et membranaires issus du gène *MUC4.*

Tous les variants de *MUC4* identifiés sont générés par un ou plusieurs événements associés d'épissage alternatif décrits précédemment et sont présentés dans le tableau 7. sv1-*MUC4* résulte des événements A, C et H. Comme l'événement A crée un changement du cadre de lecture, la séquence déduite de sv1-*MUC4* est différente de celle de sv0-*MUC4.* La nouvelle séquence est riche en résidus de sérine, thréonine et proline. L'événement C génère un codon stop. La séquence C-terminale en aval du domaine central répétitif de sv1-MUC4 est maintenant de 164 résidus au lieu de 1156 résidus pour sv0-MUC4. sv1-*MUC4* code une forme tronquée de MUC4α.

variant	type d'évenement
sv1-*MUC4*	A, C, H
sv2-*MUC4*	A, E
sv3-*MUC4*	B, H
sv4-*MUC4*	B, D
sv5-*MUC4*	B, E
sv6-*MUC4*	F, H
sv7-*MUC4*	H
sv8-*MUC4*	G

Tableau 7 : Tableau représentant les événements d'épissage alternatif qui produisent les différentes isoformes de *MUC4.*

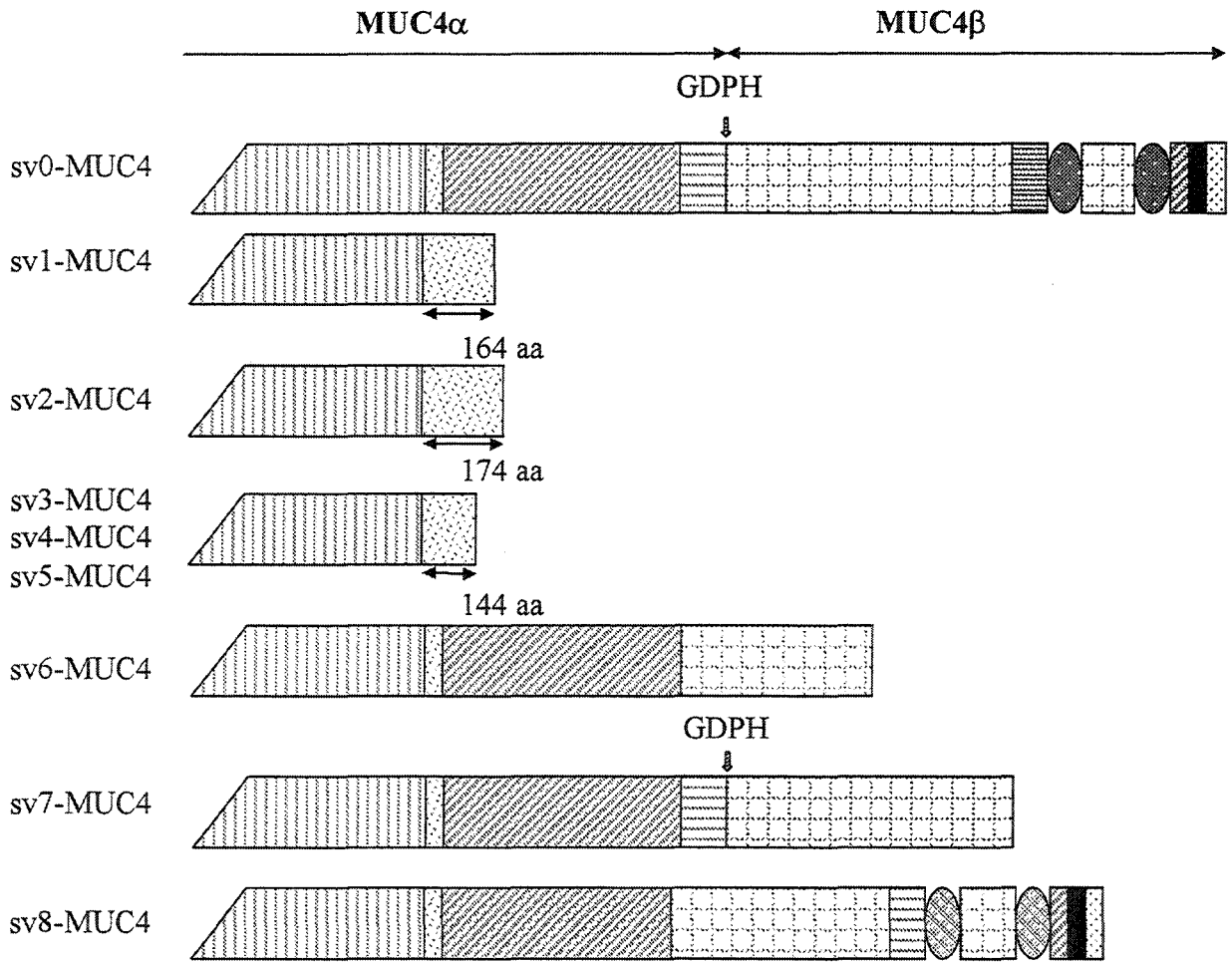


Figure 24 : Représentation schématique des séquences C-terminales déduites des différentes isoformes de MUC4.

domaine riche en résidus de sérine, thréonine et proline	domaine de type EGF
séquence unique	domaine transmembranaire
domaine riche en résidus de cystéine	domaine cytoplasmique
domaine riche en sites potentiels de N-glycosylation	séquence répétitive

sv2-*MUC4* résulte des événements A et E. Comme sv1-*MUC4*, sv2-*MUC4* code une forme tronquée de MUC4α avec une extrémité C-terminale composée de 174

résidus. Le changement du cadre de lecture induit par l'événement A génère un codon stop en amont de l'évenement E.

sv3-*MUC4* résulte des événements B et H, sv4-*MUC4* des événements B et D et sv5-*MUC4* des événements B et E. Si les séquences d'ADNc de ces 3 isoformes sont différentes, elles codent la même forme tronquée de MUC4α avec une extrémité C-terminale composée de 144 résidus. L'événement B induit un changement du cadre de lecture et crée un codon stop.

sv6-*MUC4* résulte des événements F et H. L'événement F ne change pas le cadre de lecture mais délète MUC4 du domaine riche en résidus de cystéine situé à l'extrémité C-terminale de MUC4α ainsi que du site potentiel de clivage protéolytique GDPH. L'événement H génère un codon stop. L'extrémité C-terminale de sv6-MUC4 n'est composée que des domaines CT1, CT2 et CT5.

sv7-*MUC4* résulte de l'événement H. sv7-*MUC4* code MUC4α avec une forme tronquée de MUC4β, composée uniquement du domaine riche en sites potentiels de N-glycosylation, le domaine CT5.

sv8-*MUC4* résulte de l'événement G. Bien que le peptide déduit soit toujours une forme membranaire de MUC4 avec 2 domaines de type EGF, sv8-*MUC4* code un peptide unique délété des domaines CT3, CT4, d'une partie du CT5 et du site potentiel de clivage protéolytique.

Seules les extrémités 3'-terminales des isoformes sv1 à sv8-*MUC4* ont pu être isolées et séquencées. La taille de l'exon 2 ne permet pas d'amplifier et de cloner les séquences complétes de chacune des isoformes. Comme aucun événement d'épissage alternatif n'a pu être détecté entre l'exon 1 et l'exon 2 ni entre l'exon 2 et l'exon 3, nous considérons que les isoformes sv1 à sv8-*MUC4* contiennent également l'extrémité 5'-terminale et la séquence répétitive centrale de sv0-*MUC4*.

II. 4. MUC4/Y

Comme MUC4 appartient à la même famille fonctionnelle de mucines que MUC1 à savoir le groupe des mucines membranaires, nous avons recherché par RT-PCR la présence d'une isoforme de MUC4 délétée de son domaine central répétitif. Cette isoforme avait été décrite pour MUC1 sous la dénomination MUC1/Y.

Une expérience de RT-PCR a été réalisée sur l'ARNm extrait de cellules de la lignée de cancer pancréatique, CAPAN1. Le choix de cette lignée cellulaire a été guidé par les résultats d'une étude réalisée en collaboration avec le Dr Surinder Batra (Omaha, Nebraska, USA) sur l'expression des gènes de mucines dans différentes lignées cellulaires de cancer du pancréas. De toutes les lignées testées, la lignée CAPAN1 est la lignée cellulaire de cancer pancréatique qui exprime le plus *MUC4*.

Un oligonucléotide sens est choisi dans la séquence 5' non traduite (nt -72/nt -52 [AJ000281]) et un oligonucléotide antisens est choisi en aval du domaine répétitif, NAU413 (nt 289/nt 309 [AJ010901], exon 4). Un produit d'amplification de 440 pb est détecté, cloné et séquencé. La séquence obtenue montre la présence d'un nouvel événement d'épissage alternatif qui génère une délétion de l'exon codant le domaine central répétitif de MUC4. Par analogie avec MUC1, cette isoforme est dénommée MUC4/Y.

Des expériences de RT-PCR sont alors réalisées à partir de l'ARN total extrait de tissus de poumons sains et tumoraux de 3 patients. Le choix de ce tissu pour notre étude ce justifie par le fait que MUC4 est exprimé par l'épithélium respiratoire, et que nous possédions au laboratoire des piéces opératoires prélévées pour un même patient en zone saine et tumorale. Ces expériences montrent la présence de MUC4/Y dans les 3 échantillons tumoraux et pour un des cas, en zone saine. L'expression en zone saine est très faible par rapport à celle détectée dans l'échantillon tumoral de ce patient.

Afin d'isoler la séquence compléte codant le variant MUC4/Y, une expérience d'"expand long RT-PCR" est réalisée sur l'ARN total extrait du poumon tumoral de l'un des patients précédents. Un oligonucléotide sens est choisi dans la séquence 5' non traduite (NAU744 nt -39/nt -19 [AJ000281]) et un oligonucléotide antisens est choisi dans la séquence 3' non traduite (NAU533 nt 3569/nt 3589 [AJ010901]). Aucun produit d'amplification n'a pu être détecté. Une expérience de "nested expand long RT-PCR" est alors réalisée sur un aliquot de la première "expand long RT-PCR". Les oligonucléotides sont NAU743 (nt -14/nt 7 [AJ000281]) et NAU555 (nt 2728/nt 2748 [AJ010901]). 2 bandes principales sont détectées d'une taille d'environ 3 kb (figure 25). Le clonage et le séquençage des produits d'amplifications sont en cours actuellement au laboratoire.

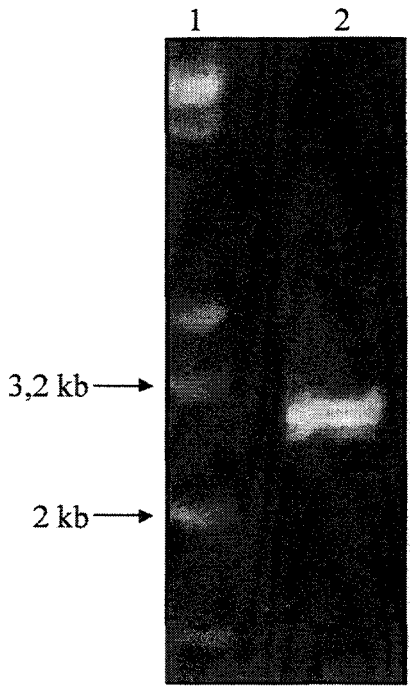


Figure 25 : Ligne 1 : témoin de masse moléculaire. Ligne 2 : nested expand long RT-PCR réalisée sur l'ARN total de poumon tumoral avec les oligonucléotides NAU743 et NAU555.

II. 5. Discussion.

MUC4 coderait un hétérodimère associé à la membrane et composé de deux sous-unités, MUC4α et MUC4β (Moniaux et al., 1999). Les deux sous-unités seraient issues d'un même précurseur, clivé au niveau d'un site de clivage protéolytique GDPH. Ce précurseur correspond à l'isoforme sv0-MUC4. sv0-MUC4 apparaît être la forme la plus abondante de MUC4 dans tous les tissus testés, comme le thymus, la thyroïde, les glandes salivaires et mammaires, la trachée, le poumon, l'œsophage, l'estomac, l'intestin grêle, le côlon, le testicule, la prostate, l'ovaire, l'utérus et le placenta.

Grâce aux techniques d'hybridation *in situ* et de Northern blot, *MUC4*, tout comme *MUC1*, est détecté dans la trachée et l'épithélium digestif (Audié et al., 1993), dans l'épithélium du tractus génital (Gipson et al., 1999) (Audié et al., 1995), dans l'épithélium oculaire (Inatomi et al., 1996), nasal, maxillaire et dans l'épithélium des glandes sous-maxillaires (Liu et al., 1998). MUC4 appartient à la famille des mucines

associées à la membrane, famille qui regroupe également MUC1. Il faut souligner que contrairement à MUC1 MUC4 ne s'exprime pas dans la vésicule biliaire et les canaux pancréatiques. MUC4 pourrait donc avoir des fonctions essentiellement tournées vers la protection des épithéliums de surface.

L'homologue de rat de MUC4, SMC est également exprimée dans l'épithélium du tractus respiratoire (Mcneer et al., 1997), digestif (Rossi et al., 1996), génital (Mcneer et al., 1998), oculaire (Price-Schiavi et al., 1998) et dans le lait (Rossi et al., 1996). SMC est présent dans ces différents tissus sous deux formes, une forme soluble et une forme membranaire. Comme la forme soluble de SMC contient les domaines de type EGF, les auteurs ont recherché la présence d'épissage alternatif entre la séquence codant pour les domaines de type EGF et la séquence 3' non traduite (Rossi et al., 1996). Aucun épissage alternatif n'ayant pu être détecté, les auteurs concluent que la forme soluble est le résultat d'un clivage protéolytique de la forme membranaire.

A l'exception de sv8-*MUC4* qui code une forme membranaire non hétérodimérique de MUC4, tous les autres variants codent des formes solubles. sv1 à sv6-*MUC4* codent des formes tronquées de MUC4α. La raison pour laquelle coexistent autant de formes solubles différentes de MUC4α est inconnue. sv7-*MUC4* code la même sous-unité de type mucine que sv0-*MUC4*, mais avec une forme tronquée et soluble de MUC4β. Pour ce variant, MUC4β n'est composée que du domaine CT5. Aucune fonction spécifique n'est connue pour ce domaine riche en sites potentiels de N-glycosylation. sv8-MUC4 est le seul variant qui pourrait avoir une double fonctionnalité, comme sv0-MUC4. Bien que sv8-MUC4 soit délété de plusieurs des domaines de sv0-MUC4, il possède toujours les domaines de type EGF, la séquence transmembranaire et le domaine cytoplasmique.

Nous suspectons la présence de fonctions de facteur de croissance associé à sv0-MUC4 via une interaction à ErbB2. La présence de sv0-MUC4 ou sv8-MUC4 pourrait induire des cascades de transduction de signal spécifiques. L'existence de l'isoforme MUC4/Y ajoute un niveau de complexité supérieure aux fonctions de MUC4. Les deux produits d'amplification obtenus après "nested expand long RT-PCR" pour *MUC4/Y* (Fig 25) nous rappellent le profil d'amplification obtenu pour *MUC4* sur l'ARN extrait de testicules (Fig 22). Nous suspectons la présence d'événements d'épissage alternatif tout au long de l'extrémité 3'-terminale de *MUC4/Y* comme c'est le cas pour l'extrémité 3'-terminale de *MUC4*. Ces isoformes sont en cours de caractérisation au laboratoire. Nous

pouvons penser qu'il existe une famille de variants MUC4/Y avec des formes solubles et des formes membranaires.

Un article est en cours de rédaction et sera soumis pour publication dans Biochemical Journal.

III. MUC4, approche des relations structure-fonction.

III. 1. Interaction MUC4 et ErbB2.

Comme nous l'avons démontré, MUC4 est l'homologue humain de la sialomucine de rat SMC. SMC forme un complexe hétérodimérique avec l'oncogène $p185^{neu}$ (Carraway et Cantley, 1994). La formation de ce complexe est réalisée grâce au domaine EGF1 de la sous-unité membranaire de SMC, l'ASGP-2. Ce complexe est mis en évidence par co-immunoprécipitation.

Nous ne disposions d'aucun anticorps permettant d'immunoprécipiter MUC4. Pour étudier l'interaction potentielle de MUC4 avec l'homologue humain de la $p185^{neu}$, ErbB2, nous avons donc utilisé une stratégie différente.

L'ADNc complet codant la sous-unité MUC4β a été cloné dans le vecteur d'expression pGEX-KT. Ce vecteur a permis la production d'une protéine de fusion GST-MUC4β. L'ADNc codant les domaines de MUC4β de l'isoforme sv7-*MUC4* a été également cloné dans le vecteur d'expression pGEX-KT. L'isoforme sv7-*MUC4* résulte de l'événement d'épissage alternatif H et code une protéine tronquée dont le segment MUC4β n'est composé que du domaine CT5. La protéine de fusion obtenue, dénommée GST-MUC4βΔH, est délétée des domaines de types EGF ainsi que du domaine transmembranaire et de la queue cytoplasmique de MUC4β.

Un extrait de protéines totales préparé à partir de la lignée cellulaire de cellules respiratoires glandulaires transformées MM39 est incubé avec les protéines de fusion GST-MUC4β et GST-MUC4βΔH immobilisées sur une résine glutathion. Les protéines retenues sur la résine sont ensuite étudiées par Western blot, la membrane est incubée avec un anticorps anti-ErbB2 (Figure 26). La figure 26 montre que ErbB2 a été retenu par la protéine de fusion GST-MUC4β immobilisée sur la résine (couloir 5) mais ne l'a pas été par la protéine de fusion GST-MUC4βΔH (couloir 3).

Nous avons pu monter que la protéine ErbB2 contenue dans l'extrait cellulaire interagit avec la protéine de fusion GST-MUC4β. La reconnaissance par ErbB2 de la protéine de fusion GST-MUC4β est faible. Elle nécessite une forte concentration d'extrait de protéines totales de la lignée MM39 pour être détectée. Ceci peut s'expliquer par le fait que la conformation de la GST-MUC4β n'est pas la même que celle de MUC4β native. De plus, l'étude par RT-PCR de l'ARN extrait des cellules MM39 montre que *MUC4* est

exprimé par les cellules de cette lignée cellulaire. MUC4β native contenue dans l'extrait de protéines totales de la lignée MM39 peut donc entrer en compétition avec la protéine de fusion GST-MUC4β pour l'interaction au récepteur ErbB2. Nous aurions préféré travailler à partir d'une lignée cellulaire n'exprimant pas *MUC4*. Cependant, toutes les lignées cellulaires que nous possédons au laboratoire et qui exprime *ErbB2*, exprime également *MUC4*.

Grâce à la protéine GST-MUC4βΔH, nous pouvons dire qu'un ou plusieurs des domaines de MUC4β en aval du domaine CT5 sont indispensables à la formation du complexe. Par analogie avec SMC, nous pouvons supposer que l'interaction est réalisée grâce au domaine EGF1 (CT7) de MUC4β.

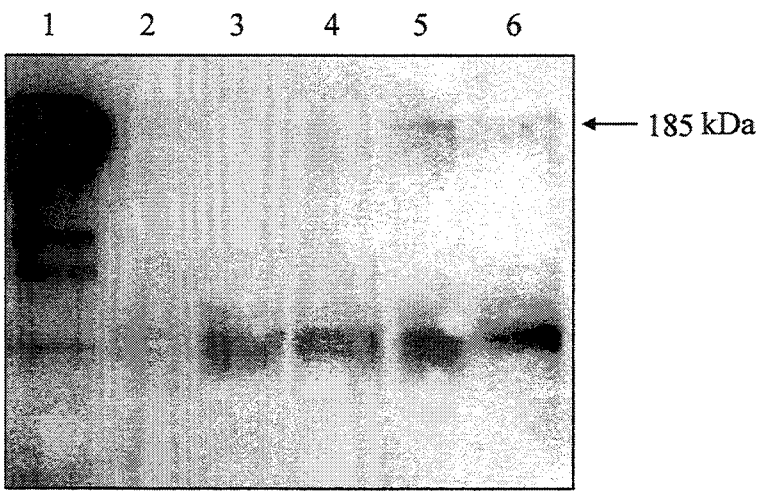


Figure 26 : Western blot réalisé après interaction des protéines de fusion de MUC4β avec un extrait de protéines totales de la lignée cellulaire MM39. Le Western blot est révélé avec un anticorps anti ErbB2. 1, protéines de MM39 ; 2, résine glutathion et protéines de MM39 ; 3, résine glutathion et GST-MUC4βΔH et protéines de MM39; 4, résine glutathion et GST-MUC4βΔH; 5, résine glutathion et GST-MUC4β et protéines de MM39 ; 6, résine glutathion et GST-MUC4β.

L'étude plus précise de la formation du complexe MUC4/ErbB2 nécessite l'obtention et l'utilisation d'anticorps spécifiques de MUC4.

III. 2. Etude de MUC4β comme facteur de croissance.

III. 2. 1. Stratégie.

La structure de la sous-unité MUC4β et le fait que *MUC4* s'exprime précocement dans la majorité des muqueuses épithéliales fœtales nous font suspecter que MUC4 puisse avoir des fonctions de facteur de croissance et être impliquée dans la prolifération cellulaire.

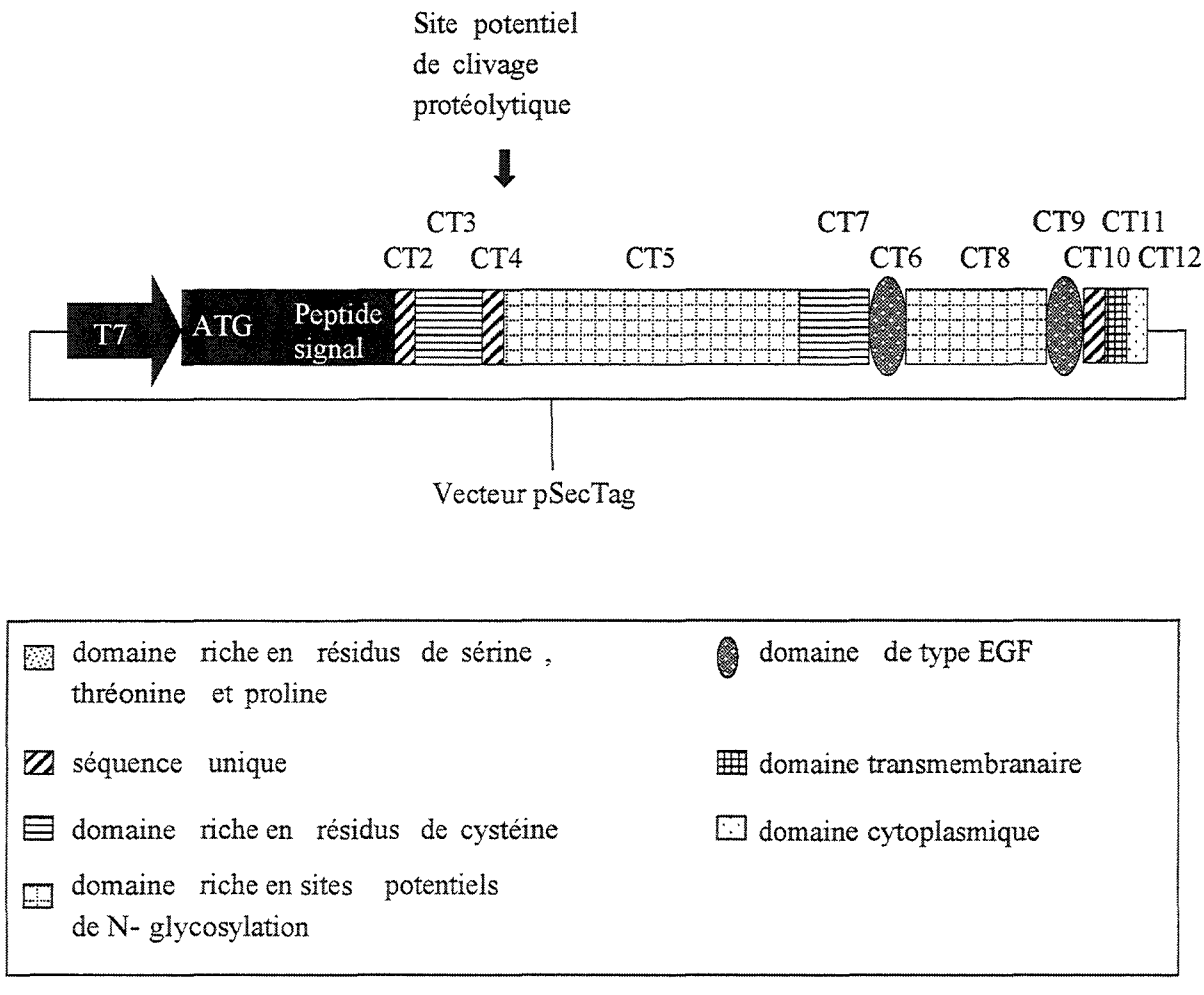


Figure 27 : Représentation schématique de la construction clonée dans le vecteur pSecTag.

Afin de vérifier cette hypothèse, nous avons cloné l'ADNc codant les domaines CT2 à CT12 de MUC4 dans un vecteur d'expression, sous le contrôle d'un promoteur constitutif qui permet de lier une séquence codant un peptide signal dans le même cadre de lecture que l'ADNc codant les domaines C-terminaux de MUC4 (Figure 27). Ce vecteur, pSecTag, permet la transfection de cellules eucaryotes.

Les cellules de la lignée de cancer mammaire MCF7 sont transfectées par différentes concentrations de la construction pSecTag-MUC4 et du vecteur vide pSecTag comme témoin. Cette lignée cellulaire est choisie pour la transfection car elle est stable, bien caractérisée et exprimant ErbB2 et MUC4. La prolifération est testée après 70 h de culture par l'utilisation du kit CellTiter 96® AQ_{ueous} One solution Cell Proliferation Assay. Les résultats sont exprimés en unités de DO.

III. 2. 2. Résultats et discussion.

La figure 28 A montre les résultats obtenus après la transfection de quantités croissantes de la construction pSecTag-MUC4 et du vecteur vide. Au temps 0, la même quantité de cellules transfectées est mise en culture pour 70 h. Nous pouvons remarquer que le nombre de cellules vivantes après transfection par le vecteur vide est pratiquement constant, quel que soit la quantité de vecteur transfecté. Par contre, le nombre de cellules vivantes après transfection par la construction pSecTag-MUC4 est croissant de 1 à 5 µg de DNA transfecté par 10^6 cellules, puis décroît de 5 à 20 µg de DNA transfecté par 10^6 cellules. La différence maximale observée se situe pour 5 µg de DNA transfecté par 10^6 cellules. En effet le nombre de cellules vivantes est alors de 26 % plus élevé après transfection par pSecTag-MUC4 qu'après transfection par le vecteur vide pSecTag. La figure 28 B représente le profil de croissance observé.

Il semble donc que MUC4 puisse avoir des fonctions de facteur de croissance en favorisant la prolifération cellulaire. Cette fonction semble être "dose-dépendante". En effet, jusqu'à une concentration de 5 µg de pSecTag-MUC4 transfecté par 10^6 cellules, la prolifération des cellules de la lignée cellulaire MCF7 est augmentée pour atteindre une valeur maximale de plus 26 %. Après la valeur de 5 µg de pSecTag-MUC4 transfecté par 10^6 cellules, la baisse du taux de prolifération observée peut s'expliquer par un effet

toxique de la protéine à forte concentration, par une saturation des partenaires potentiels de MUC4 ou par un effet inhibiteur de la croissance dose dépendant. Une des fonctions attribuées à ErbB2 est de produire un effet mitogène. L'action de MUC4 sur la prolifération cellulaire est donc en adéquation avec une association au protooncogène ErbB2.

A

	µg DNA/10^6 cellules	1	3	5	6	10	12	20
DO	pSecTag	1,13	1,10	1,17	1,16	1,11	1,13	1,10
	pSecTag-MUC4β	1,12	1,33	1,57	1,34	1,30	1,26	1,07

B

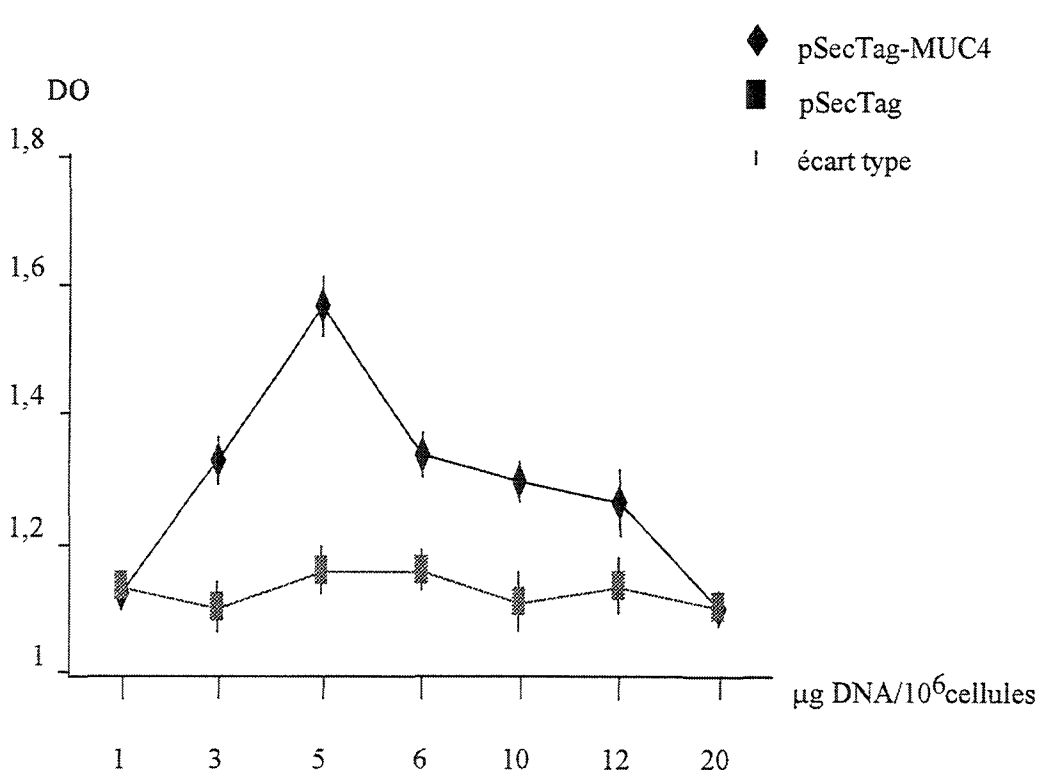


Figure 28 : A, valeurs des DO après transfection par la construction pSecTag-MUC4 et pSecTag après 70h de culture. B, profil de croissance en unités de DO par µg de DNA transfecté par 10^6 cellules (moyenne de 6 expériences).

Ces résultats doivent être vérifiés par transfection stable avec l'expression de MUC4 placée sous contrôle d'un promoteur inductible. La technique utilisée de transfection instable nécessite un comptage précis des cellules transfectées avant la culture de 70 h. Des erreurs de comptage avant le repiquage des cellules transfectées peuvent induire un biais au niveau des résultats.

IV. Polymorphisme VNTR du gène *MUC4.*

IV. 1. Analyse du polymorphisme associé à 2 séquences introniques répétées en tandem.

La comparaison de la séquence d'ADNc qui code les domaines C-terminaux de MUC4 avec celles des séquences d'ADN génomique du clone cosmidique LEA2 a permis d'établir pour une part l'organisation exon-intron de *MUC4* mais également de découvrir 2 nouvelles séquences introniques répétées en tandem (Figure 29). La première séquence répétée en tandem a un motif imparfait de 26 à 32 pb, la seconde a un motif parfait de 32 bp.

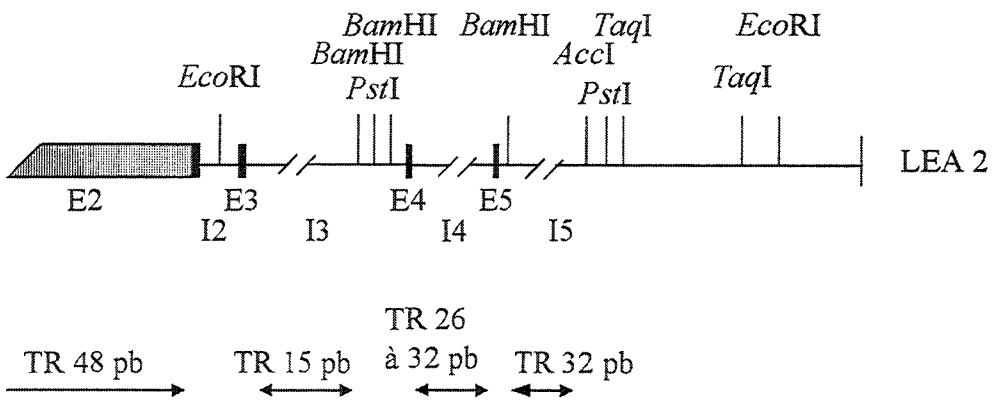


Figure 29 : Organisation génomique partielle de l'extrémité 3' de *MUC4.*

L'étude menée en parallèle de la nôtre par Séverine Nollet sur le polymorphisme de la séquence exonique répétée en tandem de 48 pb et de la séquence intronique répétée en tandem de 15 pb a révélé dans les deux cas un polymorphisme de type VNTR. De plus cette étude permet de montrer que l'association haplotypique des deux séquences répétées en tandem n'est pas aléatoire. A un allèle donné pour la séquence répétitive de 48 pb, correspond le plus souvent un allèle donné pour la séquence répétitive de 15 pb. Cette étude a été menée par Southern blot sur de l'ADN extrait de lymphocytes après une double hydrolyse enzymatique *Eco*RI/*Pst*I. Comme au moins un site *Eco*RI ou *Pst*I encadre les 2 nouvelles séquences répétées en tandem, nous avons décidé d'étendre l'étude du polymorphisme en Southern blot à l'aide des même coupures enzymatiques pour les 2 nouvelles séquences.

La séquence répétée ayant le motif imparfait de 26 à 32 pb ne montre que peu de polymorphisme de type VNTR. Une étude menée à partir de l'ADN extrait de lymphocytes de 18 volontaires caucasiens ne révèle que 3 allèles différents pour cette séquence (Figure 30 C). L'allèle le plus grand a une taille de 4 kb (*a*), il n'est représenté qu'une seule fois dans cette étude. Les 2 autres allèles ont une taille de 3,8 (*b*) et de 2,8 kb (*c*). 16 individus sur 18 présentent l'allèle de 3,8 kb et 7 individus sur 18 sont homozygotes.

La séquence répétée ayant le motif parfait de 32 pb montre quant à elle un haut niveau de polymorphisme de type VNTR (Figure 30 D). Au moins 7 allèles distincts sont présents, variant en taille de 0,98 à 2,9 kb (**a**, 2,9 kb ; **b**, 1,32 kb ; **c**, 1,3 kb ; **d**, 1.25 kb ; **e**, 1,25kb ; **f**, 1,05 kb ; **g**, 0,98 kb). La majorité des allèles ont une taille comprise entre 1 et 1,5 kb. 7 individus sur 18 sont homozygotes. 5 de ces individus sont également homozygotes pour le motif répétitif imparfait.

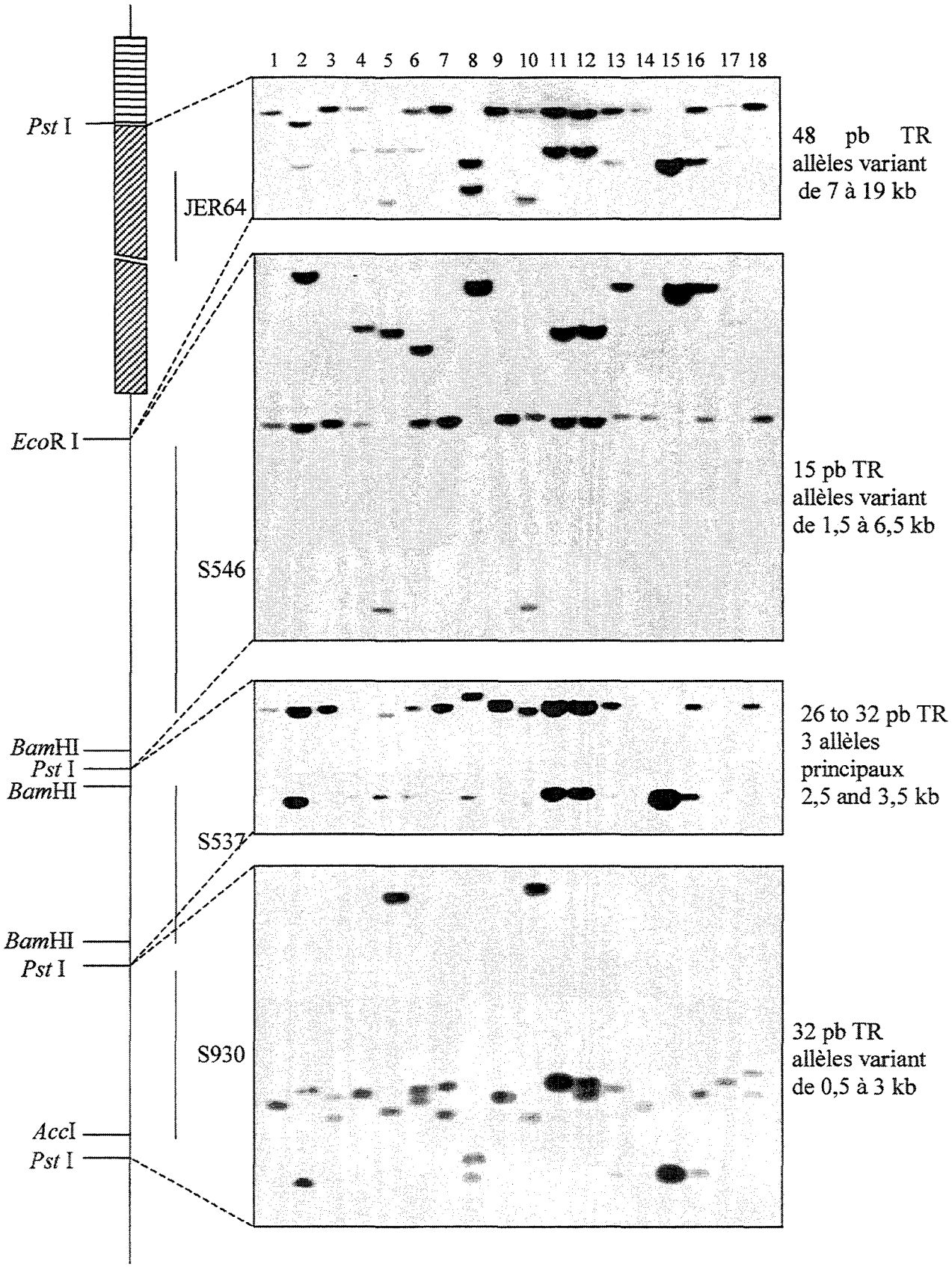


Figure 30 : Southern blot réalisé avec les sondes correspondant aux 4 motifs répétitifs ; JER64 pour le motif de 48 bp, S546, pour le motif de 15 pb, S537 pour le motif de 26 à 32 pb et S930 pour le motif de 32 bp. Chacun des motifs répétitifs est localisé sur la carte de restriction génomique. (chaque numéro au-dessus de la figure correspond à un individu)

IV. 2. Comparaison des variations polymorphiques associées aux 4 séquences répétées en tandem.

Afin d'étudier les variations polymorphiques des 4 séquences répétées en tandem, l'ADN de 18 individus caucasiens non apparentés est analysé par Southern blot (Figure 30).

4 individus sont homozygotes pour les 4 séquences répétées en tandem et 3 individus le sont pour 3 des motifs. Dans ces derniers cas, les 3 individus ne sont pas homozygotes pour le motif de 32 pb. Les allèles de ce motif sont de petite taille, et donc facilement séparables par la technique électrophorétique utilisée. Nous pensons que ces 3 individus ne sont réellement homozygotes pour aucune de leurs séquences répétées en tandem, mais que la différence de taille de chaque allèle des motifs de 48 pb, de 15 pb et du motif imparfait est trop petite pour être détectée. Cette supposition a été vérifiée dans d'autres exemples à propos du motif de 48 pb par le travail de thèse mené par Séverine Nollet. Cette étude a permis de mettre en évidence la présence d'un nouvel allèle dénommé A^+ pour le motif de 48 pb. Cet allèle ne peut être différencié de l'allèle A par hydrolyse *Eco*RI/*Pst*I mais est détecté après hydrolyse par *Rsa*I. Le fragment *Eco*RI/*Pst*I de 19 kb correspondant à l'allèle A donne 3 fragments de 10 kb, 5,5 kb et 3,6 kb après hydrolyse par *Rsa*I. Le fragment *Eco*RI/*Pst*I de 19 kb correspondant à l'allèle A^+ donne lui 3 fragments de 10 kb, 5,4 kb et 3,6 kb après hydrolyse par *Rsa*I.

Les résultats de thèse de Séverine Nollet montrent également que les variations alléliques du motif de 48 pb sont associées aux variations du motif de 15 pb. Le tableau 8 montre les associations haplotypiques des 4 séquences répétées en tandem. Nous pouvons remarquer que la distribution haplotypique des allèles des 4 motifs répétitifs ne semble pas être aléatoire. Dans notre échantillonnage, l'allèle A pour le motif de 48 pb est toujours associé à l'allèle h pour le motif de 15 pb et l'allèle *b* pour le motif imparfait. L'allèle correspondant au motif de 32 pb associé à cette formule haplotypique est plus variable. Cependant deux allèles sont prépondérants, l'allèle **c** et l'allèle **d.** Comme précédemment à propos de l'homozygotie apparente, cette variabilité peut s'expliquer par une meilleure séparation des allèles due à leur taille. Les formules haplotypiques Ah*b***d** et Ah*b***c** sont les plus fréquentes.

Nous pouvons remarquer également que lorsque l'allèle d'un motif n'est pas conservé, il est remplacé par un allèle dont la taille est très proche. Ainsi, nous trouvons

la formule haplotypique Bh*b***c** au lieu de Ah*b***c**, Ah*b***b** au lieu de Ah*b***c** ou Ee*c***c** au lieu de Ed*c***c**.

individus	48 pb	15 pb	26 à 32 pb	32 pb	haplotype
1	AA	hh	*bb*	**dd**	Ah*h***d** Ah*b***d**
2	BH	ah	*bc*	**cg**	Bh*b***c** Ha*c***g**
3	AA	hh	*bb*	**ce**	Ah*b***c** Ah*b***e**
4	AE	dh	*bc*	**cc**	Ah*b***c** Ed*c***c**
5	EL	dl	*bc*	**ea**	Ed*c***c** Ll *b***a**
6	AE	eh	*bc*	**cd**	Ah*b***d** Ee*c***c**
7	AA	hh	*bb*	**ce**	Ah*b***e** Ah*b***c**
8	GK	bb	*ac*	**fg**	Gb*a* **?** Kb*c* **?**
9	AA	hh	*bb*	**dd**	Ah*b***d** Ah*d***d**
10	AL	hl	*bc*	**ea**	Ah*b***e** L l*c***a**
11	AE	dh	*bc*	**cc**	Ah*b***c** Ed*c***c**
12	AE	dh	*bc*	**cd**	Ah*b***d** Ed*c***c**
13	AH	bh	*bc*	**cg**	Ah*b***c** Hb*b***g**
14	AA	hh	*bb*	**ee**	Ah*b***e** Ah*b***e**
15	HH	bb	*cc*	**gg**	Hb*c***g** Hb*c***g**
16	AH	bh	*bc*	**dg**	Ah*b***d** Hb*c***g**
17	AE	dh	*bc*	**cc**	Ah*b***c** Ed*c***c**
18	AA	hh	*bb*	**bd**	Ah*b***d** Ah*b***b**

Tableau 8 : Distribution haplotypique (déduite des individus homozygotes) des 4 séquences répétées en tandem qui caractérisent *MUC4*.

IV. 3. Discussion.

Le gène *MUC4* est caractérisé par la présence d'au moins 4 séquences répétitives ayant un motif élémentaire différent. Ces séquences répétées en tandem, de type minisatellite, montrent un polymorphisme interindividuel de type VNTR. L'une des séquences ayant un motif répétitif de 48 pb est exonique. Comme son allèle le plus fréquent est l'allèle le plus grand, la taille du domaine qu'il code apparaît être importante quant à sa fonctionnalité. En effet, cette séquence code un domaine riche en résidus de sérine et thréonine qui sont autant de sites potentiels de O-glycosylation. Ce domaine est indispensable aux fonctions de type mucine de protection et de lubrification des épithéliums.

Les variations polymorphiques associées aux séquences répétitives introniques semblent être en relation (tout au moins dans le faible échantillonnage de notre étude) entre elles et avec la séquence répétitive exonique. La raison pour laquelle les 4 séquences répétitives distinctes montrent des variations polymorphiques avec une répartition haplotypique conservée est encore inconnue. A ce jour, aucune fonction n'est associée à la présence de séquences répétitives situées dans les introns des gènes de mucines. Cependant dans le cas du gène *MUC4*, les 3 séquences répétitives introniques représentent presque la totalité de l'intron correspondant et sont donc situées au voisinage des sites d'épissage. Les séquences répétitives pourraient donc être en relation avec le mécanisme d'épissage alternatif. En effet, des études récentes montrent que des séquences répétées en tandem introniques sont impliquées dans la régulation, la transcription ou la traduction des gènes qui les contiennent [pour revue (Nakaruma et al., 1998)].

Conclusion

Notre travail a permis de caractériser les domaines C-terminaux de la mucine humaine MUC4. Ces résultats compilés aux résultats concernant l'extrémité N-terminale et le domaine central nous permettent d'établir l'organisation structurale complète de MUC4 et de démontrer son homologie avec la sialomucine de rat SMC.

MUC4 appartient au groupe des mucines associées à la membrane. Comme pour les autres membres de ce groupe, MUC4 est exprimée à la fois sous forme soluble et sous forme membranaire. En effet, 10 ARNm différents codent une famille de 8 variants distincts. 5 de ces variants sont des formes solubles de MUC4 et 3, des formes membranaires.

La forme la plus représentée, sv0-MUC4, a une taille maximale théorique de 2,12 µm lorsqu'elle est implantée dans la membrane. Elle est codée par un ARN de 26, 5 kb pour son allèle le plus grand. sv0-MUC4 a une structure parfaitement similaire à celle de SMC. Le degré de similarité ainsi que leur profil d'expression respectif démontrent que SMC est l'homologue de rat de MUC4. Comme SMC, sv0-MUC4 est une protéine modulaire. Elle pourrait être composée de deux sous-unités, une sous-unité de type mucine, MUC4α et une sous-unité membranaire, MUC4β. Les deux sous-unités pourraient, comme c'est le cas pour SMC, être issues d'un même précurseur clivé au niveau d'un site de clivage protéolytique GlyAspProHis (GDPH) présent aussi dans la séquence de MUC4.

Selon le nombre de répétitions du motif de 16 résidus d'acides aminés, la taille de MUC4α varie de 530 à 930 kDa. Exception faite de son peptide signal et de son domaine en position C-terminale, tous les autres montrent une richesse en résidus de sérine, thréonine et proline. Le domaine en position C-terminale de MUC4α est, lui, riche en résidus de cystéine. Il ne montre aucune similarité avec d'autres domaines décrits pour les mucines. Ce domaine est peut-être impliqué dans l'interaction des deux sous-unités qui composent MUC4. Une seule forme soluble contient ce domaine, il s'agit de sv7-MUC4. sv7-MUC4, comme tous les autres variants solubles de MUC4, a une structure de type mucine. Ce domaine permet peut-être à sv7-MUC4 d'interagir avec les mucines sécrétées composant le mucus et ainsi de modifier ses propriétés rhéologiques.

MUC4β est composée de 8 domaines, dont 2 domaines riches en sites potentiels de N-glycosylation, 2 domaines de type EGF, un domaine riche en résidus de cystéine, un

domaine transmembranaire et une queue cytoplasmique. La queue cytoplasmique contient un site potentiel de phosphorylation.

Comme son homologue de rat, MUC4β peut former un complexe avec le protooncogène ErbB2. MUC4β est également capable d'induire la prolifération des cellules de la lignée de cancer mammaire MCF7. Cette fonction de MUC4β sur la prolifération cellulaire apparaît être "dose-dépendante". Bien que nos résultats ne permettent pas de savoir si l'action de MUC4β sur la prolifération cellulaire nécessite la formation du complexe avec ErbB2, nous pouvons le supposer.

Le variant dénommé MUC4/Y dépourvu du domaine de type mucine, principalement exprimé au niveau de cellules ou de tissus tumoraux, nous semble pouvoir jouer une fonction nouvelle, importante dans la prolifération cellulaire. Il ne peut donc pas intervenir dans les fonctions généralement dévolues aux mucines comme la protection des epithéliums. Bien que nos résultats n'aient pas permis de caractériser les domaines C-terminaux de MUC4/Y, nous pouvons penser qu'il conserve les domaines EGF. La conformation spatiale de MUC4/Y pourrait alors favoriser l'interaction avec ErbB2 ainsi que la prolifération des cellules cancéreuses.

Tous les variants de MUC4 identifiés à ce jour résultent d'un mécanisme d'épissage alternatif. Comme c'est le cas pour les autres mucines membranaires, nous pouvons penser que de nouvelles formes solubles, issues d'un clivage protéolytique des formes membranaires, peuvent exister.

L'étude par RT-PCR a permis de montrer l'expression de *MUC4* dans presque tous les tissus. Dans tous les tissus testés, sv0-MUC4 apparaît être l'isoforme prépondérante. Bien que nous ne puissions pas étudier l'expression spécifique de chacune des isoformes, nous pouvons tout de même conclure que selon le tissu étudié, le niveau d'expression des isoformes est variable par rapport à celui de sv0-MUC4. La régulation de l'expression de *MUC4* apparaît donc complexe. La quantité et la qualité de MUC4 synthétisée sont régulées d'une manière spécifique de tissu.

Le gène *MUC4* est composé d'au moins 4 séquences répétitives distinctes de type minisatellite qui montrent un polymorphisme interindividuel de type VNTR. L'une des séquences est exonique et code le domaine central de MUC4α. Les 3 autres sont introniques et pourraient être impliquées dans les mécanismes d'épissage alternatif de *MUC4*. Sur le faible échantillonnage de notre étude, les variations polymorphiques des séquences répétées en tandem semblent être en relation les unes avec les autres. Nous ne

savons pas à ce jour quelle pourrait être la raison d'une telle relation entre ces 4 séquences répétitives. Elles pourraient faire l'objet d'analyses plus précises pour rechercher un rôle possible de ce polymorphisme dans le fonctionnement qualitatif ou quantitatif de ce gène en particulier dans les maladies inflammatoires, infectieuses et tumorales affectant les muqueuses.

Appendice technique.

I. Tampons

↳ TE 10X

Tris-HCl 1 M, pH 8,0	10 ml
EDTA 0,5 M	2 ml
H_2O	qsp 100 ml
pH 7,0	

↳ SSC 20X

NaCl	175,6 g
Citrate trisodique	88,2 g
H_2O	qsp 1 l
pH 7,0	

↳ SSPE 20X

NaCl	174 g
NaH_2PO4, H_2O	27,6 g
EDTA	7,4 g
H_2O	qsp 1 l
pH 7,4	

↳ Denhart's 50X

Ficoll 400	1 g
Polyvinylpyrrolidone	1 g
Sérum albumine bovine	1 g
H_2O	100 ml

↳ LB

Bactotryptone	10 g
Yeast extract	5 g
NaCl	5 g

H_2O	qsp 1 l
pH 7,2	

↳ LB-Agar

LB contenant 15 g d'agar pour 1 l

↳ SOC

Bactotryptone (w\v)	2 %
Yeast extract (w\v)	0,5 %
NaCl	10 mM
KCl	25 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

↳ Tampon borate

Acide borique	55 g
Tris	108 g
EDTA	9,3 g
H_2O	qsp 1 l
pH 8,3	

↳ Tampon GT

Isothiocyanate de guanidium	23,6 g
Citrate trisodique	73,5 mg
Sarcosyl	250 mg
H_2O	qsp 50 ml
DEPC	50 µl
β mercaptoéthanol (extemporanément)	375 µl

↳ MOPS 10X

MOPS	16 g
Acétate de sodium 3 M	6,72 ml
EDTA 0,5 M, pH 8,0	8 ml

H_2O	400 ml
pH 7,0	

↪ Tampon de dénaturation pour ARN

Formamide désionisée	500 µl
Formaldéhyde désionisé	178 µl
(sur résine AG501X)	
MOPS 10X	100 µl
H_2O	qsp 1 ml

↪ Hybridation ARN

SSPE 20X	25 ml
Denhardt's 50X	20 ml
ADN de sperme de hareng	10 mg
Formamide	50 ml
SDS	2 g
H_2O	qsp 100 ml

↪ Tampon de dénaturation de l'ADN phagique

NaOH	0,2 M
NaCl	1,5 M

↪ Tampon de neutralisation de l'ADN phagique

Tris-HCl	0,5 M
NaCl	3 M
pH 7,5	

↪ TMN

Tris-HCl	20 mM
NaCl	0,5 M
$MgSO_4$	10 mM
pH 7,5	

↳ Tampon de lyse pour préparation de l'ADN humain

Tris-HCl, pH 7,4	10 mM
EDTA	10 mM

↳ Tampon de préhybridation Southern blot

SSC 20X	30 ml
Denhardt's 50X	10 ml
SDS 10 %	5 ml
H_2O	qsp 100 ml

↳ Tampon d'hybridation Southern blot

SSC 20X	30 ml
Denhardt's 50X	10 ml
ADN de sperme de hareng	25 mg
SDS 10 %	5 ml
Sulfate de dextran 50 %	10 g
H_2O	qsp 100 ml

↳ PBS 10X

KH_2PO_4	15 mM
Na_2HPO_4	81 mM
KCl	27 mM
NaCl	1,37 mM
H_2O	qsp 1 l
pH 7,4	

↳ TBST

Tris-HCl, pH 7,4	0,5 mM
NaCl	150 mM
Tween 20	0,5 ml
H_2O	qsp 1 l

↳ Tampon de binding 5X

NaCl	400 mM

Tris-HCl, pH 7,4	80 mM
EDTA	4 mM
NP 40 10 %	12,5 ml
H_2O	qsp 500 ml
Extemporanément pour 10 ml de tampon binding 1X	
Leupeptine 10 mg/ml	10 µl
Aprotinine 10 mg/ml	10 µl
PMSF 0,1 M	60 µl
DTT 1 M	10 µl

↪ **Tampon de migration des protéines**

Tris-HCl, pH 8,3	25 mM
Glycine	192 mM
SDS	0,1 %

↪ **Tampon d'électrotransfert**

Tris-HCl, pH 8,2	25 mM
Glycine	192 mM
Méthanol	20 %
SDS	0,005 %

II. Matériels.

↪ Les ARN polyA+ extraits de thymus, thyroïde, trachée, poumon, estomac, prostate, ovaire, utérus, testicule, placenta, glandes mammaires et glandes salivaires sont d'origine commerciale (Clontech).

↪ Les ARNm extraits de poumon sain et de cancer du poumon, d'œsophage, de côlon, des lignées cellulaires MM39 et CAPAN1, sont préparés par la méthode d'extraction des ARN de grande taille (Debailleul et al., 1998).

↪ Prélèvements sanguins de volontaires caucasiens.

↪ Une banque d'expression construite en vecteur phagique à partir de l'ARNm extrait d'un épithélium colique selon le protocole donné par le fournisseur (Amersham).

↪ Les vecteurs de clonage :

- pBluescript KS(+), vecteur de clonage des fragments d'ADN (Stratagene)
- pCR®2.1, vecteur de clonage des fragments obtenus par PCR (Invitrogen)
- pGEX-KT, vecteur de clonage pour les protéines de fusion (don de David J. Hakes, Department of Biological Chemistry, University of Michigan Medical School, Michigan, USA)
- pSecTag B, vecteur d'expression en cellules eucaryotes.

↪ Les cellules bactériennes :

- La souche JM109 d'*Escherichia coli* (Promega)
- La souche INVαF' d'*Escherichia coli* (Invitrogen)

↪ Les cellules eucaryotes :

- Les cellules MM39 : ce sont des cellules qui dérivent de cellules respiratoires glandulaires transformées par SV40. Elles sont maintenues en culture en milieu DMEN F12 complémenté à 10 % en sérum de veau fœtal décomplémenté. La lignée cellulaire nous a été donnée par D^r Marc Merten et le D^r Catherine Figarella (Merten, et al. 1996).
- Les cellules MCF7 : ce sont des cellules qui dérivent de cancer mammaire. Elles sont maintenues en culture en milieu MEN complémenté en glutamine et à 10 % en sérum de veau fœtal décomplémenté.

↪ Des enzymes de restriction (Boehringer Mannheim).

↪ Taq polymérase:

- Taq polymérase (Boerhinger Mannheim)
- Expand Long ^{TM} Template PCR system (Boerhinger Mannheim)

↪ Kit de marquage des sondes d'ADN, Kit Random Primed DNA Labeling (Boehringer Mannheim).

↪ Kit de synthèse d'ADNc, Advantage^{TM} RT-for-PCR Kit (Clontech)

↪ 5'/3' RACE Kit (Boehringer Mannheim)

↪ Purification de l'ADN après migration en gel d'agarose, QIAquick Gel Extraction Kit (Qiagen)

↪ Extraction de l'ADN plasmidique :

- mini-préparation, Qiaprep Spin Plasmid Kit (Qiagen)
- maxi-préparation, Qiaprep Maxi Plasmid Kit (Qiagen)

↪ Séquence de l'ADN

- SequiTherm Exce® II Long-Read Premix DNA Sequencing Kit-LC (TEBU) pour l'automate DNA Sequencer model 4000L LI-COR
- ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Kit (Perkin-Elmer) pour l'automate ABI PRISM model 377 XL automatic sequencer (Perkin-Elmer)

↪ Test de prolifération cellulaire

- Celltiter 96® AQ_{ueous} One Solution Cell Proliferation Assay

↪ anticorps monoclonal de souris CB11 anti c-ErbB2 (Novo Castra)

↪ révélation immunologique par électro-chimioluminescence

- SuperSignal® Substrate Western Blotting Kit (Pierce)

III. Méthodologie.

III.1. Caractérisation de l'extrémité 3'-terminale de *MUC4*.

III.1.1. Criblage de la banque d'expression construite en vecteur phagique à partir d'ARNm extrait d'un épithélium colique.

La banque est mise en contact avec une culture de la souche bactérienne Y1088 puis étalée sur boîtes LB-agar ampicilline (50 µg/ml). Après incubation la nuit à 37°C, les clones phagiques sont répliqués sur membrane de nitrocellulose. Les membranes sont dénaturées, neutralisées, préhybridées et hybridées à une sonde d'ADN radio-marquée à $\alpha^{32}P$ (les sondes utilisées sont : JER64, S1217 et JER107). Les clones positifs sont isolés et leur titre est augmenté par les étapes de préministock et de ministock.

Pour le préministock, un clone recombinant positif est mis en contact 15 min dans 200 µl de suspension bactérienne en présence de $MgSO_4$ à 10 mM. 5 ml de LB-ampicilline (50 µg/ml) -$MgSO_4$ sont alors ajoutés à la suspension, le tout est incubé à 37 °C jusqu'à obtention de la lyse. Lorsque la lyse est terminée, 1/100ème de chloroforme est ajouté à la suspension, laissé en contact 15 min puis on centrifuge à 7500g pendant 10 min. Le surnageant est conservé avec quelques gouttes de chloroforme à 4 °C.

Pour le ministock, 50 µl de préministock sont mis en contact 10 min à 37 °C sans agitation avec 2 ml de LB-ampicilline (50 µg/ml) -$MgSO_4$ et 2 ml de culture bactérienne Y1088. 16 ml de LB-ampicilline (50 µg/ml) -$MgSO_4$ sont alors ajoutés, le mélange est incubé à 37 °C jusqu'à la lyse. Lorsque la lyse est terminée, 1/100ème de chloroforme est ajouté à la suspension, laissé en contact 15 min puis centrifugé à 7500g pendant 10 min. Le surnageant est conservé avec quelques gouttes de chloroforme à 4 °C.

Pour l'extraction de l'ADN phagique, 2 ml du ministock sont incubés avec de ml d'une suspension bactérienne Y1088 jusqu'à la lyse dans 500 ml LB-ampicilline (50 µg/ml) -$MgSO_4$. Après la lyse, les phages sont précipités la nuit à 4 °C avec 25 g de NaCl et 50 g de PEG. Le précipité est alors centrifugé à 10000g pendant 15 min. Le culot est repris dans 8 ml de TMN avec 8 ml de chloroforme. Après 10 min, le tout est centrifugé 10 min à 3800g. A 4,7 g de surnageant, 3,5 g de CsCl sont ajoutés, puis centrifugés 2 h à 4 °C à 300000g. L'ADN phagique est alors prélevé à la seringue, le CsCl est éliminé par dialyse. L'ADN phagique est alors incubé 2 h à 37 °C avec 20 mg/ml de protéinase K et 63 µl de

SDS. L'ADN est ensuite purifié par une extraction phénol/chloroforme, puis précipité à l'alcool. L'ADN phagique est alors dissous en TE 1X et dosé.

III.1. 2. La technique dite de 3'-RACE-PCR.

Les ADNc sont synthétisés à partir de 1 µg d'ARNm extrait d'un épithélium colique grâce au kit Advantage™ RT-for-PCR Kit (Clontech) avec comme oligonucléotide l'oligo dT-anchor du kit 5'/3' RACE Kit (Boehringer Mannheim). Une PCR est alors réalisée grâce au kit Expand Long ™ Template PCR system (Boerhinger Mannheim) avec comme oligonucléotide sens NAU491 (5'AGCAGGCCGAGTCTTGGATTA3', position : nt 515/535 dans la séquence AJ010901). NAU491 est choisi à l'extrémité 3'-terminale du clone d'ADNc JER107. L'oligonucléotide antisens est l'oligonucléotide correspondant à l'anchor du kit 5'/3' RACE Kit (Boehringer Mannheim). Le mélange réactionnel est réalisé à partir de :

- 5 µl d'ADNc
- 10 mM de dNTP en sel de sodium
- 0,4 µM de chaque oligonucléotide
- 5 µl du tampon 3 10X
- 0,75 mM de $MgCl_2$
- 2,5 unités de Taq polymérase.
- QSP 50 µl

La PCR est réalisée avec le Perkin-Elmer Thermal Cycler Gene Amp® PCR System 9700. Les paramètres de PCR sont les suivants :

- dénaturation : 94 °C 2 min
- 1ère élongation : 94 °C 30 s
  60 °C 45 s
  71 °C 4 min

cette opération est répétée 10 fois

- 2ème élongation : 94 °C 30 s
  60 °C 45 s
  71 °C 4 min (avec une extension de 40 s à chaque cycle)

cette opération est répétée 20 fois

- terminaison : 71 °C 30 min.

Une seconde étape de PCR est ensuite réalisée à partir de 1µl de la PCR précédente avec comme oligonucléotide sens NAU483 (5'CTGTTTCTCTACCAGAGCGGT3', position dans la séquence AJ010901 : nt 682/702) choisi juste en aval de NAU 491 et comme oligonucléotide antisens l'oligo dT-anchor du kit 5'/3' RACE Kit (Boehringer Mannheim). La seconde PCR est réalisée dans les mêmes conditions que la première PCR. Les produits d'amplification sont déposés en gel d'agarose, purifiés à l'aide du kit QIAquick Gel Extraction Kit (Qiagen), clonés en vecteur pCR®2.1 et séquencés. Les séquences sont réalisées avec les oligonucléotides du vecteur (T7 et "reverse" M13) grâce au kit SequiTherm Exce® II Long-Read Premix DNA Sequencing Kit-LC (TEBU) sur l'automate DNA Sequencer model 4000L LI-COR ou avec des oligonucléotides internes grâce au kit ABI PRISM dRhodmanine Terminator Cycle Sequencing Ready Kit (Perkin-Elmer) pour l'automate ABI PRISM model 377 KL automatic sequencer (Perkin-Elmer)
Les séquences sont analysées grâce à PC/GENE Software et grâce à la banque de données internationales : http://www.ncbi.nlm.nih.gov.

III.2. Etude de l'expression de *MUC4* et caractérisation de ces variants.

III. 2. 1. Etude de l'expression de *MUC4*.

1 µg d'ARN polyA+ extrait de thymus, thyroïde, trachée, poumon, estomac, prostate, ovaire, utérus, testicule, placenta, glandes mammaires et glandes salivaires (Clontech) ainsi que 1 µg d'ARNm extrait de poumon sain et de cancer du poumon, d'œsophage, de côlon sont utilisés pour réaliser des ADNc grâce au kit Advantage™ RT-for-PCR Kit (Clontech) avec un oligonucléotide oligo dT. Des PCR sont réalisées avec la Taq polymérase (Boehringer Mannheim), avec un oligonucléotide sens choisi juste en amont de la séquence codant pour le domaine EGF1 NAU591 (5'GTCTCGGTGATCGCGCTCTCC3', position dans la séquence AJ010901 nt 1976/1996), et comme oligonucléotide antisens NAU533 (5'AAGAATCCTGACAGCCTTCAG3', position dans la séquence AJ010901 nt 3569/3589) choisi dans la séquence 3' non traduite. Le mélange réactionnel est réalisé à partir de :

- 5 µl d'ADNc
- 10 mM de dNTP en sel de lithium

- 0,4 µM de chaque oligonucléotide
- 5 µl du tampon 10X
- 1 unité de Taq polymérase.
- QSP 50 µl

La PCR est réalisée avec le Perkin-Elmer Thermal Cycler Gene Amp® PCR System 9700.
Les paramètres de PCR sont les suivants :

- dénaturation : 94 °C 3 min
- élongation : 94 °C 30 s
  60 °C 30 s
  72 °C 1 min

cette opération est répétée 30 fois
Les produits d'amplification sont déposés en gel d'agarose à 1,5 %.

III. 2. 2 Caractérisation des variants de MUC4.

Les ADNc sont synthétisés à partir de 1 µg d'ARN polyA+ extrait de testicules (Clontech) grâce au kit Advantage™ RT-for-PCR Kit (Clontech) avec comme oligonucléotide l'oligo dT. Une PCR est alors réalisée grâce au kit Expand Long ™ Template PCR system (Boerhinger Mannheim) avec comme oligonucléotide sens NAU412 (5'CCAGCACTGCCATGCACACCC3', position dans la séquence AJ010901 nt 117/127). NAU412 est choisi dans l'exon 3 (aucun événement d'épissage alternatif n'est détecté entre l'exon 2 et l'exon 3). L'oligonucléotide antisens est l'oligonucléotide NAU533 choisi dans la séquence 3' non traduite afin d'isoler la séquence complète des différentes isoformes. Le mélange réactionnel est réalisé à partir de :

- 5 µl d'ADNc
- 10 mM de dNTP en sel de sodium
- 0,4 µM de chaque oligonucléotide
- 5 µl du tampon 3 10X
- 0,75 mM de $MgCl_2$
- 2,5 unité de Taq polymérase.
- QSP 50 µl

La PCR est réalisée avec le Perkin-Elmer Thermal Cycler Gene Amp® PCR System 9700.
Les paramètres de PCR sont les suivants :

- dénaturation : 94 °C 2 min

-1ère élongation : 94 °C 30 s

60 °C 45 s

71 °C 4 min

cette opération est répétée 10 fois

- 2ème élongation : 94 °C 30 s

60 °C 45 s

71 °C 4 min (avec une extension de 40 s à chaque cycle)

cette opération est répétée 20 fois

- terminaison : 71 °C 30 min.

Les produits d'amplification sont clonés en vecteur pCR®2.1 sans purification préalable sur gel. Les clones obtenus sont séquencés comme vu précédemment et les séquences sont analysées grâce à PC/GENE Software.

III. 3. Approche des relations structure-fonction.

III. 3. 1. Interaction MUC4 et ErbB2.

L'ADNc codant pour la totalité de MUC4β est cloné dans le vecteur pGEX-KT. Ce vecteur permet la production d'une protéine de fusion GST-MUC4β. L'ADNc codant la sous unité MUC4β de l'isoforme sv7-*MUC4* est également cloné dans le vecteur d'expression pGEX-KT. La protéine de fusion obtenue, dénommé GST-MUC4βΔH, est délétée des domaines de type EGF, du domaine transmembranaire et de la queue cytoplasmique. Les protéines de fusion sont produites dans la souche bactérienne JM109 après une induction par l'IPTG (1 mM). Les protéines bactériennes totales sont extraites en triton X100 à 0,5 %. Les protéines sont purifiées par centrifugation 10 min à 17500g. Les protéines totales de la lignée cellulaire MM39 sont préparées de la même manière. L'extrait de protéines bactériennes totales contenant la protéine de fusion GST-MUC4β et celui contenant la GST-MUC4βΔH sont incubés 1 h à 4 °C en présence de résine glutathion en tampon de "binding". Après 4 lavages successifs en tampon de "binding", la résine glutathion-GST-MUC4β et la résine glutathion- GST-MUC4βΔH sont incubées 1 h à 4 °C en présence de l'extrait de protéines totales de la lignée cellulaire MM39. Le rapport de

protéines bactériennes (quantité d'extrait total utilisé pour l'interaction de la résine avec les protéines de fusion)/protéines de MM39 est de 1/1000. Les résines sont ensuite lavées 4 fois en tampon "binding". Les résines sont ensuite portées à ébullition 5 min en tampon de dénaturation de protéines. Après centrifugation 1 min à 14000g, les surnageants sont déposés en gel d'acrylamide SDS-PAGE à 8 %. Après transfert sur membrane de nitrocellulose, la membrane est mise en contact avec l'anticorps anti-ErbB2. Après lavage en TBST, un anticorps secondaire anti-souris fourni dans le kit SuperSignal® Substrate Western Blotting Kit (Pierce) permet de révéler la présence de l'anticorps ErbB2 par électro-chimioluminescence.

III. 3. 2. Etude de MUC4 comme facteur de croissance.

La séquence qui code les domaines C-terminaux de MUC4 est clonée dans le vecteur d'expression pSecTag. Ce vecteur qui contient une séquence codant un peptide signal, permet la production de protéines en cellules eucaryotes sous contrôle d'un promoteur constitutif. 1, 3, 5, 6, 10, 12 et 20 µg du vecteur pSecTag-MUC4 sont transfectés par électroporation (300 volts, 975 µfarads) à 10^6 cellules de la lignée MCF7. Les mêmes quantités de vecteur vide sont transfectées dans les mêmes conditions comme témoin. Après comptage, 5000 cellules transfectées sont mises en culture pour 70 h (10 % de CO_2) en plaque 96 puits. Un changement de milieu est réalisé après 48 h de culture. Afin de contrôler la prolifération cellulaire, un test colorimétrique est réalisé avec le kit Celltiter 96® AQ_{ueous} One Solution Cell Proliferation Assay (Promega).

III. 4. Etude du polymorphisme VNTR du gène *MUC4*.

III. 4. 1. Préparation de l'ADN génomique humain.

Le sang prélevé sur anticoagulant est congelé puis décongelé rapidement à 37°C et réparti en fractions de 10 ml dans des tubes de 50 ml. La lyse des cellules est réalisée par adjonction de 40 ml d'un tampon de lyse pendant 10 min. Une centrifugation de 10 min à 3000g permet d'éliminer dans le surnageant, l'hémoglobine et les autres constituants solubles. L'opération de lavage est répétée plusieurs fois (5 à 6 fois) jusqu'à l'obtention d'un culot de débris cellulaires et de noyaux pratiquement blanc. 14 ml d'un dénaturant protéique, l'hydrochlorure de guanidinium 6 M filtré, sont ajoutés au culot qui est alors

homogénéisé par une agitation ménagée durant une trentaine de minutes. Puis 2 ml d'un détergent anionique, le sarcosyl à 10 %, et 150 µl de protéinase K à 10 mg/ml (préparée extemporanément) sont ajoutés. La solution est incubée une heure à 60°C ou une nuit à 37°C. L'ADN est précipité par addition (qsq 50 ml) d'éthanol absolu froid auquel a préalablement été ajouté 1 ml d'acétate d'ammonium 7,5 M. Il est recueilli par enroulement autour d'une baguette de verre, et rincé 4 fois à l'éthanol à 70 % (v/v). L'ADN ainsi purifié est redissous dans du TE X (4 ml pour un tube de 10 ml de sang). Afin d'obtenir une bonne homogénéisation, l'ADN est laissé 48 h sous agitation douce.

La concentration d'ADN extrait est calculée par la mesure spectrophotométrique de la densité optique (DO) à 260 nm. Une concentration de 50 µg/ml correspond à 1 de DO. La pureté de l'ADN est appréciée par le rapport de la DO à 260 nm sur la DO à 280 nm, qui doit être voisin de 2. De plus la qualité de l'ADN peut être contrôlée par migration électrophorétique dans un gel d'agarose à 0,8 %. La migration a lieu à voltage élevé (2 V/cm) pendant une heure et demi environ. L'ADN de bonne qualité doit migrer sous forme d'une bande de 50 kb ou plus, la présence d'une traînée signifie que l'ADN est dégradé. Les solutions d'ADN sont stockées à 4°C.

III. 4. 2. Réalisation des Southern blots.

L'hydrolyse est réalisée sur 10 à 40 µg d'ADN génomique à raison de 5 à 10 U d'enzyme par µg d'ADN dans les conditions optimales recommandées par le fabriquant. Le temps d'incubation est de l'ordre de 24 heures. Une recharge d'enzyme peut être effectuée au bout de 6 heures si nécessaire. Les enzymes utilisées sont *Eco*RI, *Pst*I et *Rsa*I. Les fragments d'ADN hydrolysés sont séparés par une électrophorèse en gel d'agarose (0,8 %, en tampon phosphate) de 18 cm de long contenant du BEt (0,5 µg/ml), une nuit à 20 volts. Sur chaque gel est déposé un témoin de masse moléculaire. Ce marqueur "13i" est un phage recombinant qui hydrolysé par l'endonucléase *Eco*RI libère 8 fragments dont les tailles sont 31, 22; 11; 4,3; 3,2; 2; 1,3 et 0,5 kb. Avant le transfert sur membrane, l'ADN doit être dénaturé. Pour cela, le gel est trempé 2 fois 30 mn dans un tampon de dénaturation puis neutralisé par deux bains de 30 mn dans le tampon de neutralisation. L'ADN est alors transféré sur membrane de Nylon Hybond N+ (Amersham) sous vide à 50 mbar pendant 1h30 à l'aide du "vaccum blotter" (Appligene). Les acides nucléiques sont ensuite fixés sur la membrane par une exposition de 3 mn aux U.V. (312 nm). La

membrane avant d'être hybridée, doit passer par une étape de préhybridation qui consiste à saturer ses sites afin d'obtenir une fixation spécifique de la sonde sur l'ADN. Pour cela, la membrane est rincée dans du SSC 3X puis incubée dans du tampon de préhybridation à raison de 50 µl/cm², au moins une heure à 65°C. Ce tampon est remplacé par du tampon d'hybridation auquel est ajoutée la sonde marquée au α^{32}PdCTP. L'hybridation se déroule une nuit à 65°C. L'excès de sonde est éliminé par plusieurs lavages réalisés comme suit :

- rinçage dans du SSC 3X
- 2 fois 15 mn en tampon SSC 0,1X - 0,1% SDS à 65°C
- rinçage dans du SSC 3X.

La membrane ainsi lavée est mise en autoradiographie à -80°C pendant des temps variables pouvant aller d'une journée à 8 jours. Avant de tester une nouvelle sonde, le blot est déshybridé par une solution de SDS 0,1% bouillant et un contrôle de déshybridation est effectué par autoradiographie.

Bibliographie

Abe, M. and Kufe, D. Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene. *Proc. Natl. Acad. Sci. USA* 90:282-286, 1993.

Allen, A., Flemstrom, G., Garner, A. and Kivilaakso, E. Gastroduodenal mucosal protection. *Physiol .Rev.* 73:823-857, 1993.

Amerongen, A.V., Oderkerk, C.H., Roukema, P.A., Wolf, J.H., Lisman, J.J. and Overdijk, B. Murine submandibular mucin (MSM): a mucin carrying N-and O- glycosylically bound carbohydrate-chains. *Carbohydr. Res.* 115:C1-C51983.

Asker, N., Axelsson, M.A.B., Olofsson, S.O. and Hansson, G.C. Human MUC5AC mucin dimerizes in the rough endoplasmic reticulum, similarly to the MUC2 mucin. *Biochem .J.* 335:381-387, 1998b.

Asker, N., Baeckstrom, D., Axelsson, M.A., Carlstedt, I. and Hansson, G.C. The human MUC2 mucin apoprotein appears to dimerize before O- glycosylation and shares epitopes with the 'insoluble' mucin of rat small intestine. *Biochem. J.* 308:873-880, 1995a.

Aubert, J.P., Biserte, G. and Loucheux-Lefebvre, M. H. Carbohydrate-peptide linkage in glycoproteins. *Arch. Biochem. Biophys.* 175: 410-418, 1976.

Aubert, J.P., Porchet, N., Crepin, M., Duterque-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. Evidence for different human tracheobronchial mucin peptides deduced from nucleotide cDNA sequences. *Am .J. Respir. Cell Mol. Biol.* 5:178-185, 1991.

Audié, J.P., Janin, A., Porchet, N., Copin, M.C., Gosselin, B. and Aubert, J.P. Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by in situ hybridization. *J. Histochem. Cytochem.* 41:1479-1485, 1993.

Baeckstrom, D., Hansson, G.C., Nilsson, O., Johansson, C., Gendler, S.J. and Lindholm, L. Purification and characterization of a membrane-bound and a secreted mucin-type

glycoprotein carrying the carcinoma-associated sialyl-Lea epitope on distinct core proteins. *J. Biol. Chem.* 266:21537-21547, 1991.

Balagué, C., Audié, J.P., Porchet, N. and Real, F.X. In situ hybridization shows distinct patterns of mucin gene expression in normal, benign, and malignant pancreas tissues. *Gastroenterology* 109:953-964, 1995.

Balagué, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J.P., Kim, Y.S., and Real, F.X. Altered expression of MUC2, MUC4, and MUC5 mucin genes in pancreas tissues and cancer cell lines. *Gastroenterology* 106:1054-1061, 1994.

Bargmann, C.I., Hung, M.C. and Weinberg, R.A. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell* 45:649-657, 1986.

Barnd, D.L., Lan, M.S., Metzgar, R.S. and Finn, O.J. Specific, major histocompatibility complex-unrestricted recognition of tumor-associated mucins by human cytotoxic T cells. *Proc. Natl. Acad. Sci. USA* 86:7159-7163, 1989.

Basbaum, C., Carlson, D., Davidson, E., Verdugo, P. and Gail, D.B. NHLBI Workshop summary. Cellular mechanisms of airway secretion. *Am. Rev. Respir. Dis.* 137:479-485, 1988.

Beerli, R.R., Hynes, N.E. and Graus-Porta, D. Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities Single-chain antibody-mediated intracellular retention of ErbB-2 impairs Neu differentiation factor and epidermal growth factor signaling. *J. Biol. Chem.* 271:6071-6076, 1996.

Bhaskar, K.R., Drazen, J.M., O'Sullivan, D.D., Scanlon, P.M. and Reid, L.M. Transition from normal to hypersecretory bronchial mucus in a canine model of bronchitis: changes in yield and composition. *Exp. Lung Res.* 14:101-120, 1988.

Bhat, R.V., Baraban, J.M., Johnson, R.C., Eipper, B.A. and Mains, R.E. High levels of expression of the tumor suppressor gene APC during development of the rat central nervous system. *J. Neurosci.* 14:3059-3071, 1994.

Bobek, L.A., Liu, J., Sait, S.N., Shows, T.B., Bobek, Y.A. and Levine, M.J. Structure and chromosomal localization of the human salivary mucin gene, MUC7. *Genomics* 31:277-282, 1996.

Bobek, L.A., Tsai, H., Biesbrock, A.R. and Levine, M.J. 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, 1993.

Boshell, M., Lalani, E.N., Pemberton, L., Burchell, J., Gendler, S. and Taylor-Papadimitriou, J. The product of the human MUC1 gene when secreted by mouse cells transfected with the full-length cDNA lacks the cytoplasmic tail. *Biochem. Biophys. Res. Commun.* 185:1-8, 1992.

Braga, V.M., Pemberton, L.F., Duhig, T. and Gendler, S.J. Spatial and temporal expression of an epithelial mucin, *Muc1*, during mouse development. *Development* 115:427-437, 1992.

Bramwell, M.E., Wiseman, G. and Shotton, D.M. Electron-microscopic studies of the CA antigen, epitectin. *J. Cell Sci.* 86:249-261, 1986.

Briand, J.P., Andrews, S.P.J., Cahill, E., Conway, N.A. and Young, J.D. Investigation of the requirements for O-glycosylation by bovine submaxillary gland UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosamine transferase using synthetic peptide substrates. *J. Biol. Chem.* 256:12205-12207, 1981.

Buisine, M.P., Janin, A., Maunoury, V., Audié, J.P., Delescaut, M.P., Copin, M.C., Colombel, J.F., Degand, P., Aubert, J.P., and Porchet, N. Aberrant expression of a human mucin gene (MUC5AC) in rectosigmoid villous adenoma. *Gastroenterology* 110:84-91, 1996.

Buisine, M.P., Desseyn, J.L., Porchet, N., Degand, P., Laine, A. and Aubert, J.P.,Genomic organization of the 3'-region of the human *MUC5*AC mucin gene: additional evidence for common ancestral gene for the 11p15.5 mucin gene family. *Biochem. J.* 332:29-738, 1998a.

Buisine, M.P., Devisme, L., Savidge, T.C., Gespach, C., Gosselin, B., Porchet, N., and Aubert, J.P. Mucin gene expression in human embryonic and fetal intestine. *Gut* 43:519-524, 1998b.

Buisine, M.P., Desreumaux, P., Debailleul, V., Gambiez, L., Geboes, K., Ectors, N., Delescaut, M.P., Degand, P.,Aubert J.P., Colombel, J.F., Porchet, N. Abnormalities in mucin gene expression in Crohn's disease. *Inflamm. Bowel Dis.* 5:24-32, 1999a

Buisine, M.P., Devisme, L., Copin, M.C., Durand-Reville, M., Gosselin, B., Aubert, J.P., and Porchet, N. Developmental mucin gene expression in the human respiratory tract. *Am.J.Respir.Cell Mol.Biol.* 20:209-218, 1999b.

Burchell, J., Wang, D. and Taylor-Papadimitriou, J. Detection of the tumour-associated antigens recognized by the monoclonal antibodies HMFG-1 and 2 in serum from patients with breast cancer. *Int .J. Cancer* 34:763-768, 1984.

Burchill, S.A. The tumour suppressor APC gene product is associated with cell adhesion. *Bioessays* 16:225-227, 1994.

Byrd, J.C., Nardelli, J., Siddiqui, B. and Kim, Y.S. Isolation and characterization of colon cancer mucin from xenografts of LS174T cells. *Cancer Res.* 48:6678-6685, 1988.

Carlstedt-Duke, B., Hoverstad, T., Lingaas, E., Norin, K.E., Saxerholt, H., Steinbakk, M., and Midtvedt, T. Influence of antibiotics on intestinal mucin in healthy subjects. *Eur. J. Clin. Microbiol.* 5:634-638, 1986.

Carlstedt, I., Lindgren, H. and Sheehan, J.K. The macromolecular structure of human cervical-mucus glycoproteins. Studies on fragments obtained after reduction of disulphide bridges and after subsequent trypsin digestion. *Biochem. J.* 213:427-435, 1983.

Carrato, C., Balagué, C., de Bolos, C., Gonzalez, E., Gambus, G., Planas, J., Perini, J.M., Andreu, D., and Real, F.X. Differential apomucin expression in normal and neoplastic human gastrointestinal tissues. *Gastroenterology* 107:160-172, 1994.

Carraway, K.L. and Cantley, L.C. A neu acquaintance for erbB3 and erbB4: a role for receptor heterodimerization in growth signaling. *Cell* 78:5-8, 1994.

Carraway, K.L., Carraway, C.A.C. and Carraway, K.L.I. Role of ErbB-3 and ErbB-4 in the physiology and pathology of mammary gland. *J. Mammary Gland Biol. Neoplasia* 2:187-198, 1997.

Carraway, K.L., Fregien, N. and Carraway, C.A. Tumor sialomucin complexes as tumor antigens and modulators of cellular interactions and proliferation. *J. Cell. Sci.* 103:299-307, 1992.

Chadee, K., Keller, K., Forstner, J., Innes, D.J. and Ravdin, J.I. Mucin and nonmucin secretagogue activity of Entamoeba histolytica and cholera toxin in rat colon. *Gastroenterology* 100:986-997, 1991.

Cohen, G.B., Ren, R. and Baltimore, D. Modular binding domains in signal transduction proteins. *Cell* 80:237-248, 1995.

Costa, M.J., Walls, J. and Trelford, J.D. c-erbB-2 oncoprotein overexpression in uterine cervix carcinoma with glandular differentiation. A frequent event but not an independent prognostic marker because it occurs late in the disease. *Am. J. Clin. Pathol.* 104:634-642, 1995.

Crawley, S. C., Gum, J. R., Hicks, J. W., Pratt, W., Aubert, J. P., Swallow, D. M., and Kim, Y. S. Genomic organization and structure of the 3' region of human *MUC3*. evidence of alternative splicing. soumis . 1999.

Daly, J.M., Jannot, C.B., Beerli, R.R., Graus-Porta, D., Maurer, F.G. and Hynes, N.E. Neu differentiation factor induces ErbB2 down-regulation and apoptosis of ErbB2-overexpressing breast tumor cells. *Cancer Res.* 57:3804-3811, 1997.

Davis, R.J. and Czech, M.P. Tumor-promoting phorbol diesters mediate phosphorylation of the epidermal growth factor receptor. *J. Biol. Chem.* 259:8545-8549, 1984.

Debailleul, V., Laine, A., Huet, G., Mathon, P., d'Hooghe, M.C., Aubert, J.P. and Porchet, N. 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, 1998.

Decker, S.J. Effects of epidermal growth factor and 12-O-tetradecanoylphorbol-13- acetate on metabolism of the epidermal growth factor receptor in normal human fibroblasts. *Mol. Cell Biol.* 4:1718-1724, 1984.

Dekker, J., van der Ende, A., Aelmans, P.H. and Strous, G.J. Rat gastric mucin is synthesized and secreted exclusively as filamentous oligomers. *Biochem .J.* 279:251-256, 1991.

Desseyn, J.L., Aubert, J.P., Van, Seuningen, I, Porchet, N. and Laine, A. Genomic organization of the 3' region of the human mucin gene MUC5B. *J. Biol. Chem.* 272:16873-16883, 1997a.

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

Desseyn, J.L., Buisine, M.P., Porchet, N., Aubert, J.P., Degand, P. and Laine, A. Evolutionary history of the 11p15 human mucin gene family. *J. Mol. Evol.* 46:102-106, 1998a.

Desseyn, J.L., Buisine, M.P., Porchet, N., Aubert, J.P. and Laine, A. Genomic organization of the human mucin gene MUC5B. cDNA and genomic sequences upstream of the large central exon. *J. Biol. Chem.* 273:30157-30164, 1998b.

Dong, Z., Thoma, R.S., Crimmins, D.L., McCourt, D.W., Tuley, E.A. and Sadler, J.E. Disulfide bonds required to assemble functional von Willebrand factor multimers. *J. Biol. Chem.* 269:6753-6758, 1994.

Dufossé, J., Porchet, N., Audié, J.P., Guyonnet, D., V, Laine, A., Van-Seuningen, I., Marrakchi, S., Degand, P., and Aubert, J.P. Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15. *Biochem. J.* 293:329-337, 1993.

Elangbam, C.S., Qualls, C.W.J. and Dahlgren, R.R. Cell adhesion molecules--update. *Vet. Pathol.* 34:61-73, 1997.

Fantl, W.J., Johnson, D.E. and Williams, L.T. Signalling by receptor tyrosine kinases. *Annu. Rev. Biochem.* 62:453-481, 1993.

Feizi, T., Gooi, H.C., Childs, R.A., Picard, J.K., Uemura, K., Loomes, L.M., Thorpe, S.J., and Hounsell, E.F. Tumour-associated and differentiation antigens on the carbohydrate moieties of mucin-type glycoproteins. *Biochem. Soc. Trans.* 12:591-596, 1984.

Fontenot, J.D., Tjandra, N., Bu, D., Ho, C., Montelaro, R.C. and Finn, O.J. Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (MUC-1) protein core. *Cancer Res.* 53:5386-5394, 1993.

Fox, M.F., Lahbib, F., Pratt, W., Attwood, J., Gum, J., Kim, Y., and Swallow, D.M. Regional localization of the intestinal mucin gene *MUC3* to chromosome 7q22. *Ann. Hum. Genet.* 56:281-287, 1992.

Freedman, V.H. and Shin, S.I. Cellular tumorigenicity in nude mice: correlation with cell growth in semi-solid medium. *Cell* 3:355-359, 1974.

Fung, P.Y. and Longenecker, B.M. Specific immunosuppressive activity of epiglycanin, a mucin-like glycoprotein secreted by a murine mammary adenocarcinoma (TA3-HA). *Cancer Res.* 51:1170-1176, 1991.

Gaillard, D., Plotkowski, C. and Puchelle, E. Mucus et protection de la muqueuse respiratoire. In: Anonymous1992,

Gendler, S.J., Burchell, J.M., Duhig, T., Lamport, D., White, R., Parker, M., and Taylor-Papadimitriou, J. Cloning of partial cDNA encoding differentiation and tumor-associated mucin glycoproteins expressed by human mammary epithelium. *Proc. Natl. Acad. Sci. U S A* 84:6060-6064, 1987.

Gendler, S., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J. and Burchell, J. A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. *J. Biol. Chem.* 263:12820-12823, 1988.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.N., and Wilson, D. Molecular cloning and expression of

human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* 265:15286-15293, 1990.

Gimmi, C.D., Morrison, B.W., Mainprice, B.A., Gribben, J.G., Boussiotis, V.A., Freeman, G.J., Park, S.Y., Watanabe, M., Gong, J., Hayes, D.F., Kufe, D.W., and Nadler, L.M. Breast cancer-associated antigen, DF3/MUC1, induces apoptosis of activated human T cells. *Nat. Med.* 2:1367-1370, 1996.

Gipson, I.K., Spurr-Michaud, S., Moccia, R., Zhan, Q., Toribara, N., Ho, S.B., Gargiulo, A.R. and Hill, J.A. *MUC4* and *MUC5B* Transcripts are the prevalent mucin messenger ribonucleic acids of the human endocervix. *Biol. Reprod.* 60:58-64, 1999.

Graus-Porta, D., Beerli, R.R., Daly, J.M. and Hynes, N.E. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J.* 16:1647-1655, 1997.

Griffiths, B., Matthews, D.J., West, L., Attwood, J., Povey, S., Swallow, D.M., Gum, J.R., and Kim, Y.S. Assignment of the polymorphic intestinal mucin gene (*MUC2*) to chromosome 11p15. *Ann. Hum. Genet.* 54:277-285, 1990.

Gross, M.S., Guyonnet-Dupérat, V., Porchet, N., Bernheim, A., Aubert, J.P. and Nguyen, V.C. Mucin 4 (*MUC4*) gene: regional assignment (3q29) and RFLP analysis. *Ann. Genet.* 35:21-26, 1992.

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

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

Gum, J.R.J., Hicks, J.W., Toribara, N.W., Rothe, E.M., Lagace, R.E. and Kim, Y.S. 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, 1992.

Gum, J.R.J., Hicks, J.W., Toribara, N.W., Siddiki, B. and Kim, Y.S. 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, 1994.

Gum, J.R., Hicks, J.W. and Kim, Y.S. Identification and characterization of the *MUC2* (human intestinal mucin) gene 5'-flanking region: promoter activity in cultured cells. *Biochem. J.* 325:259-267, 1997a.

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

Guyonnet-Dupérat, V., Audié, J.P., Debailleul, V., Laine, A., Buisine, M.P., Galiègue-Zouitina, S., Pigny, P., Degand, P., Aubert, J.P., and Porchet, N. Characterization of the human mucin gene *MUC5AC*: a consensus cysteine- rich domain for 11p15 mucin genes? *Biochem. J.* 305:211-219, 1995.

Hanski, C., Drechsler, K., Hanisch, F.G., Sheehan, J., Manske, M., Ogorek, D., Klussmann, E., Hanski, M.L., Blank, M., and Xing, P.X. Altered glycosylation of the MUC-1 protein core contributes to the colon carcinoma-associated increase of mucin-bound sialyl-Lewis(x) expression. *Cancer Res.* 53:4082-4088, 1993.

Harwerth, I.M., Wels, W., Schlegel, J., Muller, M. and Hynes, N.E. Monoclonal antibodies directed to the erbB-2 receptor inhibit in vivo tumour cell growth. *Br. J. Cancer* 68:1140-1145, 1993.

Hayes, D.F., Sekine, H., Ohno, T., Abe, M., Keefe, K. and Kufe, D.W. Use of a murine monoclonal antibody for detection of circulating plasma DF3 antigen levels in breast cancer patients. *J. Clin. Invest.* 75:1671-1678, 1985.

Heldin, C.H. Dimerization of cell surface receptors in signal transduction. *Cell* 80:213-223, 1995.

Helm, R.M. and Carraway, K.L. Evidence for the association of two cell surface glycoproteins of 13762 mammary ascites tumor cells. Concanavalin A-induced

redistribution of peanut agglutinin-binding proteins. *Exp. Cell Res.* 135:418-424, 1981.

Henningson, C, Selvaraj, S., MacLean, G., Luresh, M., Noujaim, A., and Longenecker, B. M. T cell recognition of a tumor-associated glycoprotein and its synthetic carbohydrate epitopes: stimulation of anticancer T cell immunity in vivo. *Cancer Immunol. Immunother.* 25, 231-241. 1987.

Hilkens, J., Buijs, F., Hilgers, J., Hageman, P., Calafat, J., Sonnenberg, A., and van der Valk, M. Monoclonal antibodies against human milk-fat globule membranes detecting differentiation antigens of the mammary gland and its tumors. *Int. J. Cancer* 34:197-206, 1984.

Hilkens, J., Ligtenberg, M.J., Vos, H.L. and Litvinov, S.V. Cell membrane-associated mucins and their adhesion-modulating property. *Trends Biochem. Sci.* 17:359-363, 1992.

Hinck, L., Nelson, W.J. and Papkoff, J. Wnt-1 modulates cell-cell adhesion in mammalian cells by stabilizing beta-catenin binding to the cell adhesion protein cadherin. *J. Cell Biol.* 124:729-741, 1994.

Ho, S.B., Niehans, G.A., Lyftogt, C., Yan, P.S., Cherwitz, D.L., Gum, E.T., Dahiya, R., and Kim, Y.S. Heterogeneity of mucin gene expression in normal and neoplastic tissues. *Cancer Res.* 53:641-651, 1993.

Ho, S.B., Shekels, L.L., Toribara, N.W., Kim, Y.S., Lyftogt, C., Cherwitz, D.L., and Niehans, G.A. Mucin gene expression in normal, preneoplastic, and neoplastic human gastric epithelium. *Cancer Res.* 55:2681-2690, 1995.

Hollingsworth, M.A., Closken, C., Harris, A., McDonald, C.D., Pahwa, G.S. and Maher, L.J. A nuclear factor that binds purine-rich, single-stranded oligonucleotides derived from S1-sensitive elements upstream of the *CFTR* gene and the *MUC1* gene. *Nucleic Acids Res.* 22:1138-1146, 1994a.

Hollingsworth, M.A., Strawhecker, J.M., Caffrey, T.C. and Mack, D.R. Expression of *MUC1*, *MUC2*, *MUC3* and *MUC4* mucin mRNAs in human pancreatic and intestinal tumor cell lines. *Int. J. Cancer* 57:198-203, 1994b.

Hoskins, L.C., Agustines, M., Mckee, W.B., Boulding, E.T., Kriaris, M. and Niedermeyer, G. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J. Clin. Invest.* 75:944-953, 1985.

Houdret, N., Périni, J.M., Galabert, C., Scharfman, A., Humbert, P., Lamblin, G., and Roussel, P. The high lipid content of respiratory mucins in cystic fibrosis is related to infection. *Biochim. Biophys. Acta* 880:54-61, 1986.

Hounsell, E.F. and Feizi, T. Gastrointestinal mucins. Structures and antigenicities of their carbohydrate chains in health and disease. *Med. Biol.* 60:227-236, 1982.

Hudziak, R.M., Lewis, G.D., Winget, M., Fendly, B.M., Shepard, H.M. and Ullrich, A. p185HER2 monoclonal antibody has antiproliferative effects in vitro and sensitizes human breast tumor cells to tumor necrosis factor. *Mol. Cell Biol.* 9:1165-1172, 1989.

Hull, S.R., Laine, R.A., Kaizu, T., Rodriguez, I. and Carraway, K.L. Structures of the O-linked oligosaccharides of the major cell surface sialoglycoprotein of MAT-B1 and MAT-C1 ascites sublines of the 13762 rat mammary adenocarcinoma. *J. Biol. Chem.* 259:4866-4877, 1984.

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

Hulsken, J., Birchmeier, W. and Behrens, J. E-cadherin and APC compete for the interaction with beta-catenin and the cytoskeleton. *J. Cell Biol.* 127:t-91994.

Hynes, N.E. and Stern, D.F. The biology of erbB-2/neu/HER-2 and its role in cancer. *Biochim. Biophys. Acta* 1198:165-184, 1994.

Inatomi, T., Spurr-Michaud, S., Tisdale, A.S., Zhan, Q., Feldman, S.T. and Gipson, I.K. Expression of secretory mucin genes by human conjunctival epithelia. *Invest. Ophthalmol. Vis. Sci.* 37:1684-1692, 1996.

Jentoft, N. Why are proteins O-glycosylated? *Trends Biochem. Sci.* 15:291-294, 1990.

Kaliner, M., Shelhamer, J.H., Borson, B., Nadel, J., Patow, C. and Marom, Z. Human respiratory mucus. *Am. Rev. Respir. Dis.* 134:612-621, 1986.

Kapitanovic, S., Radosevic, S., Kapitanovic, M., Andelinovic, S., Ferencic, Z., Tavassoli, M., Primorac, D., Sonicki, Z., Spaventi, S., Pavelic, K., and Spaventi, R. The expression of p185(HER-2/neu) correlates with the stage of disease and survival in colorectal cancer. *Gastroenterology* 112:1103-1113, 1997.

Karunagaran, D., Tzahar, E., Beerli, R.R., Chen, X., Graus-Porta, D., Ratzkin, B.J., Seger, R., Hynes, N.E., and Yarden, Y. ErbB-2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. *Embo J.* 15:254-264, 1996.

Kavanaugh, W.M. and Williams, L.T. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* 266:1862-1865, 1994.

Keates, A.C., Nunes, D.P., Afdhal, N.H., Troxler, R.F. and Offner, G.D. Molecular cloning of a major human gall bladder mucin: complete C- terminal sequence and genomic organization of MUC5B. *Biochem. J.* 324:295-303, 1997.

Khatri, I.A., Forstner, G.G. and Forstner, J.F. 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, 1997.

Klomp, L.W., Van Rens, L. and Strous, G.J. Cloning and analysis of human gastric mucin cDNA reveals two types of conserved cysteine-rich domains. *Biochem. J.* 308:831-838, 1995.

Koldovsky, O. Is breast-milk epidermal growth factor biologically active in the suckling? *Nutrition* 5:223-225, 1989.

Konstan, M.W., Cheng, P.W. and Boat, T.F. A comparative study of lysozyme and its secretion by tracheal epithelium. *Exp. Lung Res.* 3:175-181, 1982.

Kornfeld, R. and Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 1985.

Kovarik, A., Peat, N., Wilson, D., Gendler, S.J. and Taylor-Papadimitriou, J. Analysis of the tissue-specific promoter of the MUC1 gene. *J. Biol. Chem.* 268:9917-9926, 1993.

Laine, A. and Hayem, A. Identification and characterization of proteins from human bronchial secretion. *Clin. Chim. Acta* 67:159-167, 1976.

Lamblin, G., Humbert, P., Degand, P. and Roussel, P. Heterogeneity of carbohydrate chains of acidic bronchial mucin isolated from the spatum of two subjects with chronic bronchitis. *Clin. Chim. Acta* 79:425-436, 1977.

Lamblin, G., Lhermitte, M., Klein, A., Périni, J. M., and Roussel, P. Diversité des chaînes glycanniques des mucines bronchiques humaines et défense antimicrobienne de la muqueuse bronchique. *Med. Sci.* 7, 1031-1040. 1991

Lan, M.S., Bast, R.C.J., Colnaghi, M.I., Knapp, R.C., Colcher, D., Schlom, J., and Metzgar, R.S. Co-expression of human cancer-associated epitopes on mucin molecules. *Int. J. Cancer* 39:68-72, 1987.

Lan, M.S., Batra, S.K., Qi, W.N., Metzgar, R.S., and Hollingsworth, M.A. Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J. Biol. Chem.* 265:15294-15299, 1990.

Lesuffleur, T., Zweibaum, A. and Real, F.X. Mucins in normal and neoplastic human gastrointestinal tissues. *Crit. Rev. Oncol. Hematol.* 17:153-180, 1994.

Lesuffleur, T., Roche, F., Hill, A.S., Lacasa, M., Fox, M., Swallow, D.M., Zweibaum, A., and Real, F.X. Characterization of a mucin cDNA clone isolated from HT-29 mucus- secreting cells. The 3' end of MUC5AC? *J. Biol. Chem.* 270:13665-13673, 1995.

Levine, M.J., Herzberg, M.C., Levine, M.S., Ellison, S.A., Stinson, M.W., Li, H.C., and van Dyke, T. Specificity of salivary-bacterial interactions: role of terminal sialic acid residues in the interaction of salivary glycoproteins with Streptococcus sanguis and Streptococcus mutans. *Infect. Immun.* 19:107-115, 1978.

Li, D., Dohrman, A. F., Gallup, M., Miyata, S., Gum, J. R., Kim, Y. S., Nadel, J. A., Prince, A. and Basbaum, C. B. Transcriptional activation of mucin by *Pseudomonas aeruginosa* lipopolysaccharide in the pathogenesis of cystic fibrosis lung disease. *Proc. Natl. Acad. Sci. USA* 94:436-441, 1997.

Li, D., Gallup, M., Fan, N., Szymkowski, D.E. and Basbaum, C.B. Cloning of the amino-terminal and 5'-flanking region of the human *MUC5AC* mucin gene and transcriptional up-regulation by bacterial exoproducts. *J. Biol. Chem.* 273:6812-6820, 1998.

Li, Y., Bharti, A., Chen, D., Gong, J. and Kufe, D. Interaction of glycogen synthase kinase 3beta with the DF3/MUC1 carcinoma-associated antigen and beta-catenin. *Mol. Cell Biol.* 18:7216-7224, 1998.

Ligtenberg, M.J., Vos, H.L., Gennissen, A.M. and Hilkens, J. Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. *J. Biol. Chem.* 265:5573-5578, 1990.

Ligtenberg, M.J., Buijs, F., Vos, H.L. and Hilkens, J. Suppression of cellular aggregation by high levels of episialin. *Cancer Res.* 52:2318-2324, 1992.

Litvinov, S.V. and Hilkens, J. The epithelial sialomucin, episialin, is sialylated during recycling. *J. Biol. Chem.* 268:21364-21371, 1993.

Liu, B., Offner, G.D., Nunes, D.P., Oppenheim, F.G. and Troxler, R.F. MUC4 is a major component of salivary mucin MG1 secreted by the human submandibular gland. *Biochem. Biophys. Res. Commun.* 250:757-761, 1998.

Majuri, M.L., Mattila, P. and Renkonen, R. Recombinant E-selectin-protein mediates tumor cell adhesion via sialyl- Le(a) and sialyl-Le(x). *Biochem. Biophys. Res. Commun.* 182:1376-1382, 1992.

Mallow, E.B., Harris, A., Salzman, N., Russell, J.P., DeBerardinis, R.J., Ruchelli, E., and Bevins, C.L. Human enteric defensins. Gene structure and developmental expression. *J. Biol. Chem.* 271:4038-4045, 1996.

Marshall, R.D. Glycoproteins. *Annu. Rev. Biochem.* 41:673-702, 1972.

Mayadas, T.N. and Wagner, D.D. von Willebrand factor biosynthesis and processing. *Ann. N. Y. Acad. Sci.* 614:153-166, 1991.

Mayadas, T.N. and Wagner, D.D. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc. Natl. Acad. Sci. U S A* 89:3531-3535, 1992.

Mc Neer, R.R., Price-Schiavi, S., Komatsu, M., Fregien, N.R. and Carraway, K.L. Sialomucin complex in tumors and tissues. *Front. Biosci.* 2:d449-d459, 1997.

Mc Neer, R.R., Carraway, C.A., Fregien, N.L. and Carraway, K.L. Characterization of the expression and steroid hormone control of sialomucin complex in the rat uterus: implications for uterine receptivity. *J. Cell Physiol.* 176:110-119, 1998.

Meden, H. and Kuhn, W. Overexpression of the oncogene c-erbB-2 (HER2/neu) in ovarian cancer: a new prognostic factor. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 71:173-179, 1997.

Meerzaman, D., Charles, P., Daskal, E., Polymeropoulos, M.H., Martin, B.M. and Rose, M.C. Cloning and analysis of cDNA encoding a major airway glycoprotein, human tracheobronchial mucin (MUC5). *J. Biol. Chem.* 269:12932-12939, 1994.

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

Merten, M. D., Kammouni, W., Renaud, W., Birg, F., Mattéi, M. G. and Figarella, C. A transformed human tracheal gland cell line, MM39, that retains serous secretory functions. *Am . J. Cell Mol. Biol.* 15 : 520-528, 1996.

Metzgar, R.S., Rodriguez, N., Finn, O.J., Lan, M.S., Daasch, V.N., Fernsten, P.D., Meyers, W.C., Sindelar, W.F., Sandler, R.S. and Seigler, H.F. Detection of a pancreatic cancer-associated antigen (DU-PAN-2 antigen) in serum and ascites of patients with adenocarcinoma. *Proc. Natl. Acad. Sci. U S A* 81:5242-5246, 1984.

Mockenstrum-Gardner, M., Rowles, J. and Gendler, S.J. MUC1 is phosphorylated and co-immunoprecipitates pp180 upon EGF-like ligand induction. *5th International Workshop on Carinoma-Associated Mucins* 1998.(Abstract)

Moniaux, N., Nollet, S., Porchet, N., Degand, P., Laine, A. and Aubert, J.P. Complete sequence of the human mucin MUC4: a putative cell membrane- associated mucin. *Biochem. J.* 338:2-333, 1999.

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

Nakamura, Y., Koyama, K. and Matsushima, M. VNTR (Variable Number of Tandem Repeat) sequences as transcriptional, translational, or functional regulators. *J. Hum. Genet.* 43:149-152, 1998.

Nguyen, P.L., Niehans, G.A., Cherwitz, D.L., Kim, Y.S. and Ho, S.B. Membrane-bound (*MUC1*) and secretory (*MUC2*, *MUC3*, and *MUC4*) mucin gene expression in human lung cancer. *Tumour Biol.* 17:176-192, 1996.

Nguyen, V.C., Aubert, J.P., Gross, M.S., Porchet, N., Degand, P. and Frezal, J. Assignment of human tracheobronchial mucin gene(s) to 11p15 and a tracheobronchial mucin-related sequence to chromosome 13. *Hum. Genet.* 86:167-172, 1990.

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

Pandey, P., Kharbanda, S. and Kufe, D. Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein. *Cancer Res.* 55:4000-4003, 1995.

Parmley, R.R. and Gendler, S.J. Cystic fibrosis mice lacking MUC1 have reduced amounts of intestinal mucus. *J. Clin. Invest.* 102:1798-1806, 1998.

Parry, G., Beck, J.C., Moss, L., Bartley, J. and Ojakian, G.K. 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, 1990.

Patton, S., Gendler, S.J. and Spicer, A.P. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim. Biophys. Acta* 1241:407-423, 1995.

Pearson, J.P., Ward, R., Allen, A., Roberts, N.B. and Taylor, W.H. Mucus degradation by pepsin: comparison of mucolytic activity of human pepsin 1 and pepsin 3: implications in peptic ulceration. *Gut* 27:243-248, 1986.

Peifer, M. Cancer, catenins, and cuticle pattern: a complex connection. *Science* 262:1667-1668, 1993.

Peifer, M. Regulating cell proliferation: as easy as APC. *Science* 272:974-975, 1996.

Pigny, P., Guyonnet-Duperat, V., Hill, A.S., Pratt, W.S., Galiegue-Zouitina, S., d'Hooge, M.C., Laine, A., Van-Seuningen, I., Degand, P., Gum, J.R., Kim, Y.S., Swallow, D.M., Aubert, J.P., and Porchet, N. Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. *Genomics* 38:340-352, 1996.

Pimental, R.A., Julian, J., Gendler, S.J. and Carson, D.D. Synthesis and intracellular trafficking of Muc-1 and mucins by polarized mouse uterine epithelial cells. *J. Biol. Chem.* 271:28128-28137, 1996.

Podolsky, D.K. and Isselbacher, K.J. Composition of human colonic mucin. Selective alteration in inflammatory bowel disease. *J. Clin. Invest.* 72:142-153, 1983.

Podolsky, D.K. and Isselbacher, K.J. Glycoprotein composition of colonic mucosa. Specific alterations in ulcerative colitis. *Gastroenterology* 87:991-998, 1984.

Porchet, N., Nguyen, V.C., Dufossé, J., Audié, J.P., Guyonnet-Dupérat, V., Gross, M.S., Denis, C., Degand, P., Bernheim, A., and Aubert, J.P. Molecular cloning and chromosomal localization of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs. *Biochem. Biophys. Res. Commun.* 175:414-422, 1991.

Porchet, N., Pigny, P., Buisine, M.P., Debailleul, V., Degand, P., Laine, A., and Aubert, J.P. Human mucin genes: genomic organization and expression of *MUC4*, *MUC5AC* and *MUC5B*. *Biochem. Soc. Trans.* 23:800-805, 1995.

Price-Schiavi, S.A., Meller, D., Jing, X., Merritt, J., Carvajal, M.E., Tseng, S.C. and Carraway, K.L. Sialomucin complex at the rat ocular surface: a new model for ocular surface protection. *Biochem. J.* 335:457-463, 1998.

Quin, R.J., Ward, B.G. and McGuckin, M.A. Phosphorylation of the MUC1 cytoplasmic tail correlates with changes in cell-cell adhesion. *5th International Workshop on Carinoma-Associated Mucins* 1998.(Abstract)

Raaberg, L., Nexo, E., Damsgaard, M.J. and Seier, P.S. Immunohistochemical localisation and developmental aspects of epidermal growth factor in the rat. *Histochemistry* 89:351-356, 1988.

Regimbald, L.H., Pilarski, L.M., Longenecker, B.M., Reddish, M.A., Zimmermann, G. and Hugh, J.C. The breast mucin MUC1 as a novel adhesion ligand for endothelial intercellular adhesion molecule 1 in breast cancer. *Cancer Res.* 56:4244-4249, 1996.

Reid, C.J., Gould, S. and Harris, A. Developmental expression of mucin genes in the human respiratory tract. *Am. J. Respir. Cell. Mol. Biol.* 17:592-598, 1997.

Rice, G.E. and Bevilacqua, M.P. An inducible endothelial cell surface glycoprotein mediates melanoma adhesion. *Science* 246:1303-1306, 1989.

Riese D.J., Stern D.F. Specificity within the EGF family/ErbB receptor family signaling network. *Bioessays* 20:41-48, 1998

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

Roussel, P., Lamblin, G. and Degand, P. Heterogeneity of the carbohydrate chains of sulfated bronchial glycoproteins isolated from a patient suffering from cystic fibrosis. *J. Biol. Chem.* 250:2114-2122, 1975.

Roussel, P., Lamblin, G., Lhermitte, M., Houdret, N., Lafitte, J.J., Périni, J.M., Klein, A., and Scharfman, A. The complexity of mucins. *Biochimie* 70:1471-1482, 1988.

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., and Polakis, P. Association of the APC gene product with beta-catenin. *Science* 262:1731-1734, 1993a.

Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis, P. The APC protein and E-cadherin form similar but independent complexes with alpha-catenin, beta-catenin, and plakoglobin. *J. Biol. Chem.* 270:5549-5555, 1995b.

Salomon, D.S., Brandt, R., Ciardiello, F. and Normanno, N. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit. Rev. Oncol. Hematol.* 19:183-232, 1995.

Schroten, H., Lethen, A., Hanisch, F.G., Plogmann, R., Hacker, J., Nobis-Bosch, R., and Wahn, V. Inhibition of adhesion of S-fimbriated Escherichia coli to epithelial cells by meconium and feces of breast-fed and formula-fed newborns: mucins are the major inhibitory component. *J. Pediatr. Gastroenterol. Nutr.* 15:150-158, 1992.

Sheehan, J.K., Oates, K. and Carlstedt, I. Electron microscopy of cervical, gastric and bronchial mucus glycoproteins. *Biochem. J.* 239:147-153, 1986.

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

Sheng, Z.Q., Hull, S.R. and Carraway, K.L. 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, 1990.

Sheng, Z., Wu, K., Carraway, K.L. and Fregien, N. 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, 1992.

Sherblom, A.P. and Carraway, K.L. A complex of two cell surface glycoproteins from ascites mammary adenocarcinoma cells. *J. Biol. Chem.* 255:12051-12059, 1980.

Shimizu, M. and Yamauchi, K. Isolation and characterization of mucin-like glycoprotein in human milk fat globule membrane. *J. Biochem.(Tokyo)* 91:515-524, 1982.

Shimizu, Y. and Shaw, S. Cell adhesion. Mucins in the mainstream. *Nature* 366:630-631, 1993.

Shirotani, K., Taylor-Papadimitriou, J., Gendler, S.J. and Irimura, T. Transcriptional regulation of the MUC1 mucin gene in colon carcinoma cells by a soluble factor. Identification of a regulatory element. *J. Biol. Chem.* 269:15030-15035, 1994.

Slamon, D.J., Godolphin, W., Jones, L.A., Holt, J.A., Wong, S.G., Keith, D.E., Levin, W.J., Stuart, S.G., Udove, J., and Ullrich, A. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244:707-712, 1989.

Slomiany, A., Witas, H., Aono, M. and Slomiany, B.L. Covalently linked fatty acids in gastric mucus glycoprotein of cystic fibrosis patients. *J. Biol. Chem.* 258:8535-8538, 1983.

Smits, H.L. and Kramer, M.F. Glycoprotein synthesis in the mucous cells of the vascularly perfused rat stomach. III. Mucous cells of the antrum and the duodenal glands. *Am. J. Anat.* 161:365-374, 1981.

Sorkin, A. and Waters, C.M. Endocytosis of growth factor receptors. *Bioessays* 15:375-382, 1993.

Spicer, A.P., Parry, G., Patton, S. and Gendler, S.J. Molecular cloning and analysis of the mouse homologue of the tumor- associated mucin, MUC1, reveals conservation of potential O- glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism. *J. Biol. Chem* 266:15099-15109, 1991a.

Spicer, A.P., Rowse, G.J., Lidner, T.K. and Gendler, S.J. Delayed mammary tumor progression in Muc-1 null mice. *J. Biol. Chem.* 270:30093-30101, 1995b.

Spik, G. and Montreuil, J. The role of lactotransferrin in the molecular mechanisms of antibacterial defense. *Bull. Eur. Physiopathol. Respir.* 19:123-130, 1983.

Steck, P.A. and Nicolson, G.L. Cell surface glycoproteins of 13762NF mammary adenocarcinoma clones of differing metastatic potentials. *Exp. Cell Res.* 147:255-267, 1983.

Stern, L., Palatsides, M., de Kretser, T. and Ford, M. Expression of the tumor-associated mucin MUC1 in an ovarian tumor cell line. *Int. J. Cancer* 50:783-790, 1992.

Stockley, R.A. and Afford, S.C. The immunological assessment of alpha 1-antitrypsin with reference to its function in bronchial secretions. *Clin. Sci.* 65:373-381, 1983.

Strous, G.J. and Dekker, J. Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* 27:57-92, 1992.

Sun, P. D. and Davies, D. R. The cystine-knot growth-factor superfamily. *Ann. Rev. Biophys. Biomol. Struct.* 24, 269-291. 1995.

Swallow, D.M., Gendler, S., Griffiths, B., Kearney, A., Povey, S., Sheer, D., Palmer, R.W., and Taylor-Papadimitriou, J. The hypervariable gene locus PUM, which codes for the tumour associated epithelial mucins, is located on chromosome 1, within the region 1q21- 24. *Ann. Hum. Genet.* 51:289-294, 1987.

Thornton, D.J., Sheehan, J.K., Lindgren, H. and Carlstedt, I. Mucus glycoproteins from cystic fibrotic sputum. Macromolecular properties and structural 'architecture'. *Biochem. J.* 276:667-675, 1991a.

Thornton, D.J., Sheehan, J.K. and Carlstedt, I. Heterogeneity of mucus glycoproteins from cystic fibrotic sputum. Are there different families of mucins? *Biochem. J.* 276:677-682, 1991b.

Thornton, D.J., Carlstedt, I., Howard, M., Devine, P.L., Price, M.R. and Sheehan, J.K. Respiratory mucins: identification of core proteins and glycoforms. *Biochem. J.* 316:967-975, 1996.

Thornton, D.J., Howard, M., Khan, N. and Sheehan, J.K. Identification of two glycoforms of the MUC5B mucin in human respiratory mucus. Evidence for a cysteine-rich sequence repeated within the molecule. *J. Biol. Chem.* 272:9561-9566, 1997.

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

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

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

Tzahar, E., Waterman, H., Chen, X., Levkowitz, G., Karunagaran, D., Lavi, S., Ratzkin, B.J., and Yarden, Y. A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol. Cell Biol.* 16:5276-5287, 1996.

Ullrich, A. and Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell* 61:203-212, 1990.

van de Bovenkamp, J.H., Hau, C.M., Strous, G.J., Buller, H.A., Dekker, J. and Einerhand, A.W. Molecular cloning of human gastric mucin MUC5AC reveals conserved cysteine-rich D-domains and a putative leucine zipper motif. *Biochem. Biophys. Res. Commun.* 245:853-859, 1998.

van de Wiel-van Kemenade, E., Ligtenberg, M.J., de Boer, A.J., Buijs, F., Vos, H.L., Melief, C.J., Hilkens, J., and Figdor, C.G. Episialin (MUC1) inhibits cytotoxic lymphocyte-target cell interaction. *J. Immunol.* 151:767-776, 1993.

van der Geer, P. and Hunter, T. Tyrosine 706 and 807 phosphorylation site mutants in the murine colony- stimulating factor-1 receptor are unaffected in their ability to bind or phosphorylate phosphatidylinositol-3 kinase but show differential defects in their

ability to induce early response gene transcription. *Mol. Cell Biol.* 11:4698-4709, 1991.

van der Geer, P., Hunter, T. and Lindberg, R.A. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* 10:251-337, 1994.

van Klinken, B.J., Dekker, J., Buller, H.A. and Einerhand, A.W. Mucin gene structure and expression: protection vs. adhesion. *Am. J. Physiol.* 269:G613-G627, 1995.

van Klinken, B.J., Van Dijken, T.C., Oussoren, E., Buller, H.A., Dekker, J. and Einerhand, A.W. Molecular cloning of human *MUC3* cDNA reveals a novel 59 amino acid tandem repeat region. *Biochem. Biophys. Res. Commun.* 238:143-148, 1997.

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

Vandenhaute, B., Buisine, M.P., Debailleul, V., Clément, B., Moniaux, N., Dieu, M.C., Degand, P., Porchet, N., and Aubert, J.P. Mucin gene expression in biliary epithelial cells. *J. Hepatol.* 27:1057-1066, 1997.

Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J.A. and Pannekoek, H. Assembly and routing of von Willebrand factor variants: the requirements for disulfide-linked dimerization reside within the carboxy-terminal 151 amino acids. *J. Cell. Biol.* 113:195-205, 1991.

Weiss, A.A., Babyatsky, M.W., Ogata, S., Chen, A. and Itzkowitz, S.H. Expression of MUC2 and MUC3 mRNA in human normal, malignant, and inflammatory intestinal tissues. *J. Histochem. Cytochem.* 44:1161-1166, 1996.

Weiss, F.U., Daub, H. and Ullrich, A. Novel mechanisms of RTK signal generation. *Curr. Opin. Genet. Dev.* 7:80-86, 1997.

Wickstrom, C., Davies, J.R., Eriksen, G.V., Veerman, E.C. and Carlstedt, I. 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, 1998.

Wreschner, D.H., Hareuveni, M., Tsarfaty, I., Smorodinsky, N., Horev, J., Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., and Dion, A. Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein forms. *Eur. J. Biochem.* 189:463-473, 1990.

Wu, K., Fregien, N. and Carraway, K.L. 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, 1994.

Yamamoto, M., Bharti, A., Li, Y. and Kufe, D. Interaction of the DF3/MUC1 breast carcinoma-associated antigen and beta-catenin in cell adhesion. *J. Biol. Chem.* 272:12492-12494, 1997.

Yu, C.J., Shun, C.T., Yang, P.C., Lee, Y.C., Shew, J.Y., Kuo, S.H., and Luh, K.T. Sialomucin expression is associated with erbB-2 oncoprotein overexpression, early recurrence, and cancer death in non-small-cell lung cancer. *Am. J. Respir. Crit. Care Med.* 155:1419-1427, 1997.

Zrihan-Licht, S., Baruch, A., Elroy-Stein, O., Keydar, I. and Wreschner, D.H. Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins. Cytokine receptor-like molecules. *Febs Lett.* 356:130-136, 1994.

Zrihan-Licht, S., Vos, H.L., Baruch, A., Elroy-Stein, O., Sagiv, D., Keydar, I., Hilkens, J., and Wreschner, D.H. Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue. *Eur. J. Biochem.* 224:787-795, 1994.