

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE I

Année 1999

N° d'ordre :

THESE

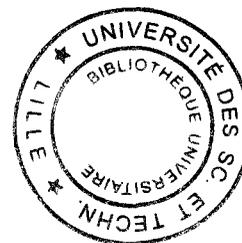
de DOCTORAT D'UNIVERSITE

pour l'obtention du grade de

DOCTEUR EN SCIENCES DE LA VIE ET DE LA SANTE

présentée par

Séverine NOLLET



ETUDE DU GÈNE D'APOMUCINE HUMAINE *MUC4*

Cartographie, clonage, polymorphisme et organisation de la région 5'

Présentée le 3 juin 1999 devant la commission d'examen

JURY

Président : Professeur Philippe Delannoy

Rapporteurs : Docteur Dallas Swallow

Docteur Alain Zweibaum

Membres : Docteur Jean-Pierre Aubert

Docteur Nicole Porchet, directeur de thèse

*A mes parents
pour leur confiance (et tout le reste...)*

*à mes sœurs, à mon frère
pour leur affection*

*à mes nièces
pour leur joie de vivre*

*à mes amis
pour leur soutien*

Je tiens à remercier très sincèrement :

Monsieur le Professeur Degand :

Vous m'avez accueillie au sein de votre laboratoire pour le DEA puis pour préparer cette thèse de Sciences. Je vous suis très reconnaissante pour la confiance que vous m'avez témoignée.

Madame le Docteur Nicole Porchet :

Tu as accepté de m'encadrer et m'as aidée tout au long de ce travail de thèse. Pendant la rédaction de ce mémoire, tu m'as fait profiter de tes connaissances scientifiques (assez impressionnantes) et soutenue dans les moments de découragement. Pour ta disponibilité et le temps passé à la correction de ce mémoire, je te suis sincèrement reconnaissante.

Monsieur le Docteur Jean-Pierre Aubert :

Pendant tout ce travail de thèse, tu m'as encouragée et apporté une aide scientifique (et technique) très appréciable, et ce, toujours avec l'humour qui te caractérise. Je t'en remercie très vivement.

Madame le Docteur Anne Laine :

Je tiens à t'exprimer ma profonde gratitude pour ton amitié, ton soutien et tes conseils avisés.

Nicolas :

Ce travail n'aurait pu être possible sans ta collaboration. Je te remercie pour ton enthousiasme communicatif et te prie de croire à toute ma sympathie.

Danièle, Pascal, Viviane, Annette, Christine et Michel :

Je vous remercie vivement pour votre aide technique si précieuse.

Marie-Pierre, Fabienne, Virginie, Isabelle, Jean-Luc, Pascal et toute l'équipe du "4^{ème} Ouest" :

Merci pour tous les bons moments passés ensemble. Je vous adresse toute mon amitié.

Je remercie Monsieur Claude Vandeperre pour la qualité de la documentation photographique.

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

Monsieur le Professeur Philippe Delannoy,

en acceptant de présider ce jury ;

Madame le Docteur Dallas Swallow,

Monsieur le Docteur Alain Zweibaum,

en acceptant d'être les rapporteurs de cette thèse ;

Monsieur le Docteur Jean-Pierre Aubert,

Madame le Docteur Nicole Porchet, mon tuteur de thèse,

en acceptant de juger ce travail.

Le travail présenté dans ce mémoire a été réalisé dans l'équipe de Recherche sur la Génétique des Apomucines Humaines, sous la direction scientifique de **Madame le Docteur PORCHET** et de **Monsieur le Docteur Jean-Pierre AUBERT**, au sein de l'Unité de Recherche N°377 "Biologie et Physiopathologie des Cellules Mucipares" de l'Institut National de la Santé et de la Recherche Médicale (INSERM) dirigée par **Monsieur le Professeur DEGAND**.

TABLE DES MATIERES

TABLE DES MATIERES	6
INDEX DES FIGURES	11
PUBLICATIONS ET COMMUNICATIONS	13
ABREVIATIONS	15
INTRODUCTION	17
I - Le mucus, les mucines : généralités	17
I.1 - Le mucus	17
I.1.1 - Définition	17
I.1.2 - Fonctions	18
I.2 - Propriétés biochimiques des mucines	19
I.2.1 - Définition	19
I.2.2 - Isolement et purification	20
I.2.3 - Structure des mucines	20
I.2.4 - Les glycanes	21
I.2.4.1 - La grande diversité des <i>O</i> -glycanes	21
I.2.4.2 - Les trois régions des chaînes <i>O</i> -glycanniques	23
I.2.4.3 - Fonctions des chaînes glycanniques	26
I.2.5 - L'axe peptidique	26
I.3 - Apport de la biologie moléculaire à la connaissance de l'axe peptidique des mucines	27
I.3.1 - Stratégie d'étude des mucines	27
I.3.2 - Les "mucin-like"	28
I.3.3 - Le domaine répétitif des mucines épithéliales	29
I.3.4 - Polymorphisme de type VNTR	29
I.3.5 - Expression des gènes de mucines épithéliales humaines	32
I.3.5.1 - Expression tissulaire analysée par la technique de Northern blot	32
I.3.5.2 - Expression cellulaire par la technique d'HIS	33
II - Connaissances actuelles sur les gènes d'apomucines épithéliales et sur leurs produits	35
II.1 - <i>MUC1</i> : gène codant une mucine transmembranaire	35
II.2 - Gènes codant les mucines sécrétées formant le gel de mucus	38
II.2.1 - <i>MUC2</i>	38
II.2.2 - <i>MUC5B</i>	39
II.2.3 - <i>MUC5AC</i>	42
II.2.4 - <i>MUC6</i>	44
II.2.5 - La famille des gènes <i>MUC</i> localisés en 11p15.5 : <i>MUC2</i> , <i>MUC5AC</i> , <i>MUC5B</i> , <i>MUC6</i>	46
II.2.5.1 Similarités	46
II.2.5.1 Phylogénèse	47
II.3 - <i>MUC3</i> : gène codant une mucine contenant un motif de type EGF	48
II.4 - <i>MUC7</i> : gène codant une mucine sécrétée soluble	50
II.5 - <i>MUC8</i>	52
II.6 - <i>MUC9</i>	52
II.7 - <i>MUC4</i>	52

II.8 - Les gènes de mucines animales.....	52
II.8.1 - Les gènes de mucines membranaires.....	54
II.8.1.1 - Le gène codant les sialomucines d'ascites de rat : ASGP-1 et ASGP-2	54
II.8.1.2 - L'homologue murin <i>Muc1</i>	57
II.8.1.3 - L'homologue murin <i>Muc3</i>	58
II.8.1.4 - Le gène <i>Muc3</i> chez le rat.....	59
II.8.2 - Les gènes de mucines animales sécrétées formant le gel de mucus.....	59
II.8.2.1 - L'homologue murin <i>Muc2</i>	59
II.8.2.2 - Le gène <i>Muc2</i> chez le rat.....	59
II.8.2.3 - L'homologue murin <i>Muc5ac</i>	61
II.8.2.4 - Le gène <i>Muc5ac</i> chez le rat.....	61
III - Biosynthèse des mucines.....	62
III.1 - Connaissances actuelles sur la régulation de la transcription des gènes de mucines.....	62
III.1.1 - Régulation de la transcription de <i>MUC1</i>	62
III.1.2 - Régulation de la transcription de <i>MUC2</i>	63
III.1.3 - Régulation de la transcription de <i>MUC5AC</i>	64
III.1.4 - Régulation de la transcription de <i>MUC5B</i>	64
III.1.5 - Régulation de la transcription de <i>MUC7</i>	64
III.2 - Maturation de l'apomucine	65
III.2.1 - Les cellules épithéliales	65
III.2.2 - Glycosylation	65
III.2.3 - Formation d'oligomères	66
III.2.4 - Transport intracellulaire et sécrétion.....	68
III.2.5 - Biosynthèse de MUC1.....	70
IV - Fonctions des mucines	72
IV.1 - Fonctions des mucines épithéliales formant le gel de mucus.....	72
IV.2 - Fonctions des mucines épithéliales membranaires	73
IV.3 - Fonctions des mucines contenant un motif de type EGF.....	75
IV.4 - Rôle des mucines dans les cellules tumorales	75
V - Ce que l'on connaissait du gène <i>MUC4</i> au début de ce travail.....	78
V.1 - Le clone JER64	78
V.2 - Localisation chromosomique.....	78
V.3 - Polymorphisme	80
V.4 - Expression.....	80
V.4.1 - Expression de l'ARN de <i>MUC4</i>	80
V.4.2 - Expression de la protéine MUC4	81
VI - Objectifs de la thèse.....	83
VII - Stratégie	85
RESULTATS.....	87
I - Caractérisation des clones génomiques	87
I.1 - Caractérisation des clones bactériophagiques contenant le gène <i>MUC4</i>	87
I.1.1 - ANT55.....	87
I.1.2 - ANT56.....	87
I.2 - Caractérisation des clones cosmidiqes.....	90
I.2.1 - LEA1 et LEA2.....	90
I.2.2 - LEA12	90
I.2.3 - LEA47	92
II - Le domaine répétitif	95
II.1 - Le domaine répétitif est homogène et de grande taille.....	95

II - Le domaine répétitif	95
II.1 - Le domaine répétitif est homogène et de grande taille	95
II.2 - Le domaine répétitif présente un polymorphisme de type VNTR	98
II.3 - Le domaine répétitif présente un polymorphisme de mutation pour l'enzyme de restriction <i>RsaI</i>	100
III - Identification d'un deuxième domaine répétitif non codant de <i>MUC4</i> en 3' du domaine répétitif (séquence de 15 pb présente dans la sonde HOR1)	105
III.1 - Caractérisation de la séquence répétitive de 15 pb	105
III.2 - Etablissement de la continuité entre les deux séquences répétitives de 48 pb et de 15 pb	105
III.3 - HOR1 : exon ou intron ?	105
III.4 - Polymorphisme de HOR1	112
III.4.1 - Polymorphisme de type VNTR	112
III.4.2 - Association allélique entre les deux domaines répétitifs de 48 pb et 15 pb	112
III.4.3 - Etude haplotypique du polymorphisme de type VNTR du gène <i>MUC4</i>	116
IV - Caractérisation de la séquence en 5' du domaine répétitif	118
IV.1 - Construction de deux banques d'ADNc (intestin grêle et côlon)	118
IV.2 - Caractérisation des clones d'ADNc obtenus par criblage de la banque d'expression colique	118
IV.3 - Obtention de clones d'ADNc par la technique de 5' RACE-PCR	120
IV.4 - Comparaison des séquences génomiques et des séquences ADNc en amont du domaine répétitif	120
IV.5 - Obtention de nouveaux clones génomiques contenant l'extrémité 5' terminale du gène de mucine humaine <i>MUC4</i>	123
V - Analyse des séquences nucléotidique et peptidique déduite de la partie 5' de <i>MUC4</i>	129
VI - Vérification du rôle potentiel du peptide signal	135
DISCUSSION	137
I - Le domaine répétitif de <i>MUC4</i>	137
I.1 - Clonage du domaine répétitif de <i>MUC4</i>	137
I.2 - Le domaine répétitif de <i>MUC4</i> est homogène et de grande taille	138
I.3 - Le domaine répétitif de <i>MUC4</i> est polymorphique	138
II - HOR1 : deuxième domaine répétitif de <i>MUC4</i>	140
III - La région 5' terminale de <i>MUC4</i>	142
III.1 - Obtention de la première séquence unique de <i>MUC4</i>	142
III.2 - Organisation génomique de la partie 5' de <i>MUC4</i>	143
III.3 - Premier anticorps dirigé contre une séquence unique de <i>MUC4</i>	144
III.4 - Peptide signal de <i>MUC4</i> et similarités avec celui de l'ASGP-1	144
IV - <i>MUC4</i> , ressemblances et différences avec les autres gènes d'apomucines humaines	145
IV.1 - <i>MUC4</i> et les gènes d'apomucines localisés en 11p15.5	145
IV.2 - <i>MUC4</i> et les gènes d'apomucines membranaires	146
V - La région promotrice de <i>MUC4</i>	147
VI - Perspectives	147
APPENDICE TECHNIQUE	149
I - Extraction et purification de l'ADN génomique à partir du sang	149
I.1 - Lyse cellulaire	149
I.2 - Protéolyse	149
I.3 - Précipitation de l'ADN	149

I.4 – Quantification et appréciation de la pureté.....	150
II - Marquage des sondes	150
II.1 - Sondes nucléiques : technique du multi-amorçage au hasard.....	150
II.2 - Sondes oligonucléotidiques : marquage en 5'OH.....	150
III – Analyse par la technique de Southern blot de l'ADN.....	150
III.1 - Transfert	151
III.2 - Hybridation.....	151
III.2.1 - Sonde nucléique	151
III.2.2 - Sonde oligonucléotidique	151
III.3 - Lavages et autoradiographie.....	151
IV – Extraction et purification des ARN totaux et des ARNm à partir de tissus	152
IV.1 - Préparation des ARN totaux par la technique dite "classique"	152
IV.2 - Protocole original adapté à la préparation de grands ARN.....	152
IV.3 - Sélection des ARN poly(A ⁺)	152
V – Analyse des ARN totaux ou messagers par la technique de Northern blot.....	153
V.1- Electrophorèse des ARN en milieu dénaturant.....	153
V.2 - Transfert	153
V.3 - Hybridation.....	153
V.4 - Lavage et autoradiographie.....	153
VI - Criblage des banques génomiques.....	154
VI.1 - Banque génomique construite en vecteur bactériophagique λ EMBL4	154
VI.1.1 - Criblage de la banque.....	154
VI.1.2 - Préparation des phages recombinants	154
VI.1.2.1 - Préministock.....	154
VI.1.2.2 - Ministock	155
VI.1.2.3 - Préparation rapide de l'ADN phagique	155
VI.1.2.4 - Amplification et extraction de l'ADN recombinant.....	155
VI.2 – Banque génomique construite en vecteur cosmique pWE15.....	156
VI.2.1 - Criblage de la banque.....	156
VI.2.2 - Purification de l'ADN des cosmides recombinants	157
VI.2.3 - Etablissement des cartes de restriction.....	157
VII - Criblage des banques d'expression.....	157
VIII – Sous clonage en vecteur plasmidique pKS	158
VIII.1 - Purification des fragments	158
VIII.1.1 - Méthode "GENECLEAN"	158
VIII.1.2 - Méthode "SeaPlaque"	158
VIII.2 - Hydrolyse du vecteur.....	159
VIII.3 – Déphosphorylation.....	159
VIII.4 – "Ligation" rapide	159
VIII.5 - Transformation.....	159
VIII.6 - Recherche des clones d'intérêt.....	160
VIII.6.1 - Empreintes plasmidiques	160
VIII.6.2 - Hybridation à une sonde nucléique	160
IX - Préparation de l'ADN plasmidique recombinant	161
IX.1 - Lyse bactérienne par un détergent	161
IX.2 - Purification de l'ADN plasmidique	161
IX.3 - Elution de l'ADN de la matrice.....	161
IX.4 - Détermination de la taille des inserts	161
X – Construction de mutants par délétion.....	162
XI - Séquençage nucléotidique.....	162

XII - Analyse des séquences.....	163
XIII - Synthèse d'ADNc.....	163
XIII.1 - Méthode d'amorçage aléatoire.....	163
XIII.2 - Méthode RACE-PCR.....	164
XIII.2.1 - Synthèse de l'ADNc.....	164
XIII.2.2 - Purification de l'ADNc.....	164
XIII.2.3 - Synthèse de la queue poly(A).....	164
XIV – Réactions d'amplification.....	166
XIV.1 - Principe.....	166
XIV.2 - Conditions techniques générales.....	166
XIV.3 - Réactions.....	167
XIV.3.1 - Amplification d'ADN génomique.....	167
XIV.3.2 - Amplification d'ADNc.....	167
XIV.3.2.1 - Minibanque d'ADNc.....	167
XIV.3.2.2 - RACE-PCR.....	167
XV - Clonage des fragments amplifiés.....	168
XV.1. – Clonage en vecteur plasmidique T/A.....	168
XV.2 - Ligation.....	168
XV.3 – Transformation bactérienne.....	168
XV.4 - Sélection des clones recombinants.....	169
XVI – Clonage des régions régulatrices.....	169
XVII – Transcription et traduction <i>in vitro</i>	170
COMPOSITION DES TAMPONS.....	171
BIBLIOGRAPHIE.....	176

INDEX DES FIGURES

Figure 1 : Représentation schématique de la structure linéaire des oligomères de mucines. . .	22
Figure 2 : Représentation schématique d'une région <i>O</i> -glycosylée de mucine [Roussel et Lamblin 1996].....	24
Tableau 1 : Séquences peptidiques consensus des régions répétitives des mucines humaines.....	30
Tableau 2 : Localisation chromosomique des gènes de mucines humaines.....	30
Tableau 3 : Profil d'expression des gènes de mucines humaines chez l'adulte analysé par hybridation <i>in situ</i>	34
Figure 3 : Représentation schématique de l'organisation génomique de <i>MUC1</i> et de l'organisation peptidique déduite [Gendler et Spicer 1995].....	37
Figure 4 : Organisation peptidique comparée du pro-vWF, de <i>MUC2</i> , de <i>MUC5B</i> , de <i>MUC5AC</i> et de <i>MUC6</i>	40
Figure 5 : Comparaison des organisations génomiques des régions 3' de <i>MUC5B</i> et <i>MUC5AC</i> [Buisine <i>et al.</i> 1998a] (A) et schéma hypothétique montrant l'évolution des gènes de mucines localisés en 11p15 à partir d'un ancêtre commun (<i>MUC11p15</i> originel) [Desseyn <i>et al.</i> 1997a ; 1998a] (B).	45
Figure 6 : Représentation schématique de l'organisation peptidique de l'extrémité C-terminale de <i>MUC3</i> [Gum <i>et al.</i> 1997b].	49
Figure 7 : Représentation schématique de l'organisation génomique de <i>MUC7</i> et de l'organisation peptidique déduite [Bobek <i>et al.</i> 1996].....	51
Tableau 4 : Séquences peptidiques consensus des régions répétitives de quelques mucines animales.	53
Figure 8 : Représentations schématiques de la biosynthèse et de l'organisation peptidique de SMC [McNeer <i>et al.</i> 1997].	55
Figure 9 : Séquence peptidique déduite du clone d'ADNc JER64 [Porchet <i>et al.</i> 1991].....	79
Figure 10 : Cartographie partielle de l'insert du clone bactériophagique ANT55.	88
Figure 11 : Cartographie partielle de l'insert du clone bactériophagique ANT56.	89
Figure 12 : Cartographie partielle de l'insert du clone cosmique LEA2.	91
Figure 13 : Cartographie partielle de l'insert du clone cosmique LEA12.	93
Figure 14 : Cartographie partielle de l'insert du clone cosmique LEA47.	94
Figure 15 : Détermination de la taille du domaine contenant les répétitions en tandem de 48 pb des cosmides LEA1 et LEA2.	96

Figure 16 : Caractérisation du domaine contenant les répétitions de 48 pb de l'insert du clone cosmique LEA2.....	97
Figure 17 : Le domaine répétitif de 48 pb présente un polymorphisme de type VNTR.....	99
Figure 18 : Le domaine répétitif de 48 pb présente un polymorphisme de mutation.	101
Figure 19 : HOR1, fragment <i>EcoRI</i> de 2,5 kb isolé du clone bactériophagique ANT55.	106
Figure 20 : Séquence nucléotidique de l'ADN du clone HOR1.	107
Figure 21 : Liaison entre les deux séquences répétitives de 48 pb et de 15 pb.....	108
Figure 22 : Analyse par Northern blot de l'expression du gène <i>MUC4</i> à partir d'un blot commercial poly(A ⁺) (Clontech laboratories).	110
Figure 23 : Séquences nucléotidique et peptidique déduite des clones obtenus par PCR et RT-PCR entre NAU174 et NAU99.....	111
Figure 24 : Le domaine répétitif de 15 pb présente un polymorphisme de type VNTR.....	113
Tableau 5 : Liaison entre les allèles observés avec la sonde JER64 (en majuscules) et ceux observés avec la sonde HOR1 (en minuscules).	115
Figure 25 : Le clone d'ADNc JER103.	119
Figure 26 : Organisation de la région 5' terminale de <i>MUC4</i>	121
Figure 27 : Séquence compilée des différents clones d'ADNc obtenus par des expériences de 5' RACE-PCR.	122
Figure 28 : La séquence unique riche en Thr/Ser en amont du domaine répétitif ne présente pas de polymorphisme de type VNTR.	124
Figure 29 : Cartographie partielle de l'insert du clone cosmique LEA51.	125
Figure 30 : Obtention de la région génomique en amont de la séquence connue de l'ARNm.	127
Figure 31 : Séquence des clones génomiques en amont de l'ATG potentiel.....	128
Figure 32 : Séquences nucléotidiques compilées et séquences peptidiques déduites des clones d'ADNc RAC1-5, JER103 et RT174-99.	131
Figure 33 : Courbe d'hydrophobie des 200 premiers résidus de <i>MUC4</i> [Kyte et Doolittle 1982].....	132
Figure 34 : Comparaison entre différents peptides signaux de mucines.....	133
Figure 35 : Marquage à la Met [³⁵ S] du produit d'expression du clone RT370-224 synthétisé à l'aide d'un système acellulaire en absence ou en présence de microsomes et d'endo H.	136
Figure 36 : Représentation schématique de l'organisation génomique de la région 5' ainsi que des deux domaines répétitifs du gène <i>MUC4</i>	141
Tableau 6 : Oligonucléotides utilisés.	165

Ce travail a fait l'objet de publications et de communications :

PUBLICATIONS

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

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

COMMUNICATIONS

Moniaux N., Nollet S., Laine A., Aubert J. P., Porchet N. (1996) The human mucin gene MUC4 contains at least two perfectly conserved sequences repeated in tandem. 4th international workshop on carcinoma-associated mucins, Cambridge.

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

Moniaux N., Nollet S., Desseyn J. L., Buisine M. P., Laine A., Porchet N., Aubert J. P. (1998) Genomic organization of the human mucin genes. MUC4 : an organization which differs from the 11p15 genes. 5th international workshop on carcinoma-associated mucins, Cambridge.

Nollet S., Moniaux N., Maury J., Petitprez D., Degand P., Laine A., Porchet N., Aubert J. P. (1998) Human mucin gene MUC4 : organisation of its 5'-region and polymorphism of its central tandem repeat array. 5th international workshop on carcinoma-associated mucins, Cambridge.

Moniaux N., Nollet S., Porchet N., Degand P., Laine A., Aubert J. P. (1999) Complete sequence of the human mucin MUC4 : a putative cell membrane-associated mucin. *Meeting of the Mucin Club, Chantilly.*

Par ailleurs, nous avons participé aux travaux du laboratoire dans d'autres cadres :

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

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

ABREVIATIONS

aa :	acide aminé
ADN :	Acide Déoxyribonucléique
ADNc :	ADN complémentaire à l'ARNm
AMV :	"Avian Myeloblastosis Virus"
ARN :	Acide Ribonucléique
ARNm :	ARN messenger
ASGP :	"Ascite Sialomucin Glycoprotein"
BEt :	Bromure d'Ethidium
CK :	"Cystine Knot"
cpm :	coup par minute
CsCl :	Chlorure de Césium
Da :	Dalton
DEPC :	Diéthylpyrocarbonate
DNase :	Déoxyribonucléase
dNTP :	Déoxynucléotide Triphosphate
DO :	Densité Optique
EDTA :	Acide Ethylène Diamine Tétracétique
EGF :	"Epidermal Growth Factor"
endo H :	endoglycosidase H
Fuc :	Fucose
G3PDH :	Glycéraldéhyde-3 Phosphate Déshydrogénase
Gal :	Galactose
GalNAc :	N-Acétyl Galactosamine
GlcNAc :	N-Acétyl Glucosamine
HIS :	Hybridation <i>In Situ</i>
IPTG :	Isopropyl- β -D-Thiogalactopyranoside
kb :	kilobase
KDa :	Kilodalton
LB :	milieu Luria Bertani
Man :	mannose

MMLV :	"Moloney-Murine Leukemia Virus"
M_r :	Masse moléculaire relative
MTX :	méthotrexate
NeuNAc :	acide N-acétylneuraminique
nt :	nucléotide
pb :	paire(s) de bases
PCR :	"Polymerase Chain Reaction"
PEG :	Polyéthylène Glycol
pfu :	"plaque forming units"
PSM :	"Porcin Submaxillary Mucin"
qsp :	quantité suffisante pour
RACE-PCR :	"Rapid Amplification of cDNA Ends"
RE :	Réticulum Endoplasmique
RER :	Réticulum Endoplasmique Rugueux
RNase :	Ribonucléase
RT-PCR :	"Reverse Transcription PCR"
SDS :	Sodium Dodécyl Sulfate
SMC :	"Sialo-Mucin Complex"
SSC :	Standard Sodium Citrate
ssDNA :	ADN de sperme de saumon
SSPE :	Standard Sodium Phosphate EDTA
STE :	Sodium Tris EDTA
TE :	Tris EDTA
TEA :	Tris EDTA Acétate
T_m :	Température de fusion
TMN :	Tris – Magnésium – NaCl
UV :	Ultraviolet
VNTR :	"Variable Number of Tandem Repeats"
vWF :	Facteur de von Willebrand
X-gal :	5-bromo-4-chloro-3-indolyl- β -galactoside

INTRODUCTION

I - LE MUCUS, LES MUCINES : GENERALITES

I.1 - Le mucus

I.1.1 - Définition

Dans le cas de la plupart des types cellulaires, la barrière physique que constitue la membrane plasmique est suffisante au maintien de l'intégrité cellulaire. Cependant, les épithéliums des **tractus respiratoire, digestif et reproducteur** ainsi que ceux de la surface **oculaire**, en contact direct avec l'environnement, requièrent une protection supplémentaire. Ils sont ainsi recouverts d'une **sécrétion protectrice**, servant également de **barrière sélective**. Cette sécrétion, appelée mucus, est produite par des cellules épithéliales spécialisées dans les fonctions de stockage et de sécrétion : les **cellules caliciformes** ou cellules en gobelet des épithéliums de surface et les **cellules muqueuses** des glandes sous-muqueuses. Le mucus est formé de **deux phases** : une phase aqueuse au contact des cellules (de faible viscosité, proche de celle de l'eau) et une phase superficielle ou phase gel (caractérisée par une viscosité et une élasticité élevées).

Le mucus forme un **gel fortement hydraté** et composé de 95 % d'eau et de 5 % de substances minérales (électrolytes : Na^+ , K^+ , Ca^{2+} , Mg^{2+} ...) et organiques : protéines cellulaires, protéines sériques (protéases, anti-protéases, albumine, lysozyme, lactoferrine), immunoglobulines sécrétoires (IgA), lipides et acides nucléiques [Strous et Dekker 1992]. Les macromolécules principales de ce gel sont les **mucines**, responsables de l'hydrophilie et des **propriétés rhéologiques** (viscosité, fluidité, adhésivité, filance et élasticité) du mucus. En effet, ces propriétés sont dues à la structure oligomérique de ces mucines dont les monomères sont reliés entre eux par des ponts disulfures. La formation du gel est également favorisée par les interactions intermoléculaires non covalentes entre les différents carbohydrates présents sur les molécules de mucines [Bansil *et al.* 1995].

I.1.2 - Fonctions

Le mucus possède plusieurs fonctions plus ou moins spécifiques de la muqueuse qui le produit [Allen 1983] :

- Le mucus recouvre l'ensemble des voies aériennes supérieures depuis les fosses nasales jusqu'aux bronchioles terminales, formant un tapis continu d'une épaisseur de 0,5 à 2 μm . Dans le **tractus respiratoire**, le mucus joue un rôle vital dans l'**humidification** des voies aériennes et l'**hydratation** des muqueuses trachéobronchiques qui sont au contact de l'air. Il représente également la **première défense** de la muqueuse respiratoire. Celle-ci est en effet exposée en permanence à de multiples agents agresseurs (gaz, particules toxiques, bactéries, virus et allergènes) inhalés au cours de la respiration. La fine couche de mucus qui recouvre les cellules épithéliales ciliées piège alors ces particules exogènes. La vibration des cils fait remonter ce tapis muqueux jusqu'au pharynx où il est dégluti (**épuration mucociliaire**) [Gaillard *et al.* 1992].

- Le mucus tapissant l'épithélium du **tractus gastro-intestinal** contribue à la protection de la muqueuse œsophagienne contre les reflux gastriques très acides. Il **protège** l'estomac et le duodénum contre les enzymes digestives et le pH acide. Il **lubrifie** les parois du tractus digestif pour les protéger des forces mécaniques lors du passage du bol alimentaire et au cours de la digestion [Allen *et al.* 1986]. Le mucus est également la **première barrière défensive** de l'épithélium : il prévient l'adhésion et la pénétration des toxines ainsi que l'invasion bactérienne. Le gel muqueux régule également le mouvement des molécules vers les cellules absorbantes. Le gel muqueux du tractus digestif est dégradé par la bile et la pepsine. L'établissement d'un équilibre entre sécrétion du mucus et altération de ce mucus est donc nécessaire au maintien de l'intégrité de l'épithélium sous-jacent. Bien que le mucus soit produit par l'hôte pour protéger la surface de l'épithélium gastro-intestinal, il peut aussi **favoriser la colonisation par des micro-organismes** commensaux ou pathogènes en fonctionnant comme source de nutriments et comme matrice dans laquelle ils peuvent proliférer [Neutra et Forstner 1987]. La microflore abondante dans l'intestin des mammifères, particulièrement au niveau du côlon, semble être principalement concentrée dans la couche de mucus. Il peut y avoir accumulation dans le gel lorsque la vitesse à laquelle les organismes se multiplient dépasse la vitesse à laquelle ils sont expulsés via le turnover et l'érosion de la couche de mucus. Ce dernier peut ainsi favoriser la prolifération de nombreuses espèces de bactéries, y compris certaines souches de bactéries virulentes.

- Dans **l'appareil urogénital féminin**, le mucus joue un rôle primordial dans la reproduction [Elstein 1978]. Il est principalement sécrété au niveau du col utérin (endocol) mais également au niveau de l'endomètre. Les propriétés du mucus cervical (quantité, filance et viscosité) **varient au cours du cycle menstruel**. Ces changements de propriétés du mucus semblent être déterminés par les variations endocriniennes observées au cours du cycle menstruel. Au moment de l'ovulation, le mucus est abondant et peu visqueux, **protégeant ainsi les spermatozoïdes** de l'environnement vaginal hostile et de la phagocytose et **permettant leur ascension** vers l'utérus. Par contre, en dehors de cette période, le mucus est épais et imperméable aux spermatozoïdes. Le mucus contribue également à **maintenir l'aseptie** lors des différentes situations physiologiques. Au niveau tubaire, il contribue à la survie et au transport des gamètes et de l'embryon.

- La **surface oculaire**, qui est l'un des tissus les plus accessibles et vulnérables chez les mammifères, est protégée par un liquide lacrymal complexe contenant des mucines. Ce liquide lacrymal assure la clarté optique, lubrifie la surface oculaire et protège des agents pathogènes et nocifs [Gipson et Inatomi 1997b].

I.2 - Propriétés biochimiques des mucines

I.2.1 - Définition

Différentes caractéristiques physicochimiques ou biochimiques des mucines ont été mises en évidence à partir du mucus [pour revue, voir Devine et McKenzie 1992 ; Gum 1992a ; Rose 1992 ; Strous et Dekker 1992] :

- Les mucines du mucus sont capables de former des **gels viscoélastiques insolubles** dans l'eau.

- Elles ont une **masse moléculaire élevée** variant de 1.10^6 à 44.10^6 Da selon les méthodes d'isolement et de purification employées [Mantle et Allen 1989].

- Les mucines sont constituées de **sous-unités** reliées entre elles par des **ponts disulfures**.

- Chaque sous-unité glycoprotéique est constituée d'un axe peptidique appelé **apomucine**, hérissé de chaînes glycaniques reliées au peptide par une liaison covalente.

- Les mucines sont des **O-glycoprotéines** : les chaînes glycosidiques sont liées au squelette protéique par une liaison O-glycosidique entre un résidu de N-acétylgalactosamine et le groupement hydroxyle d'un résidu Thr ou Ser.

- Les chaînes glycaniques représentent plus de 50 % du poids sec.
- Ces chaînes oligosaccharidiques sont composées de **cinq sucres** différents : la *N*-acétylgalactosamine (GalNAc), la *N*-acétylglucosamine (GlcNAc), le galactose (Gal), le fucose (Fuc) et l'acide *N*-acétylneuraminique (NeuNAc) ou acide sialique. Elles peuvent également contenir des groupements sulfates ainsi que de faibles quantités de mannose (Man) supposant l'existence de quelques chaînes *N*-glycaniques. Les mucines ne contiennent pas d'acides uroniques.
- L'assemblage de ces sucres conduit à une **très grande hétérogénéité** des chaînes différentes dans leur composition (plus d'une centaine), leur longueur (un à vingt sucres) et leur acidité (neutres, acides ou très acides) [Roussel *et al.* 1988].
- Les mucines ont une **charge globale négative** due à la présence de résidus **d'acide sialique et de sulfate**.
- Le squelette protéique contient un nombre important de résidus **Thr** et/ou **Ser**, et de **Pro**. A eux seuls, ces trois résidus constituent 20 à 55 % de la totalité des résidus d'acides aminés.

1.2.2 - Isolement et purification

Trois principales techniques biochimiques sont utilisées pour isoler et purifier les mucines [Gum 1992a ; Strous et Dekker 1992] :

- la **chromatographie de gel-filtration** : isole les mucines de haut poids moléculaire des protéines de faible poids moléculaire ; sépare les sous-unités des polymères de mucines.
- l'**ultracentrifugation sur gradient de densité** : sépare les mucines des protéines et des lipides ayant une densité plus faible, ainsi que des acides nucléiques (densité plus élevée).
- la **chromatographie par échange d'ions** : permet de sous-fractionner les mucines (chargées négativement).

La méthode la plus utilisée car donnant les meilleurs résultats (rendement et pureté) est la centrifugation sur gradient de densité en chlorure de césium.

Les parties C- et N-terminales non-glycosylées des mucines étant très sensibles à la protéolyse, des **inhibiteurs de protéases** sont ajoutés pour prévenir leur dégradation.

1.2.3 - Structure des mucines

Les mucines sont les principales molécules responsables de la formation du gel de mucus et de ses propriétés rhéologiques.

Une des caractéristiques des mucines est leur **structure oligomérique**. L'utilisation d'agents thiols réduisant les ponts disulfures provoque la réduction de la structure oligomérique en sous-unités et une diminution de la viscosité du gel de mucus [Allen 1986]. Les sous-unités glycoprotéiques ainsi obtenues ont une viscosité plus faible et sont incapables de reformer le gel de mucus. L'intégrité des ponts disulfures est indispensable à la fonction des mucines.

La viscosité du gel de mucus peut également être diminuée sous l'action d'une large variété de protéases telles que la trypsine ou la pronase. La protéolyse a permis de mettre en évidence que les mucines sont constituées d'une **alternance de domaines fortement glycosylés**, hydrophiles (insensibles aux agents protéolytiques) **et de domaines "nus"**, moins glycosylés et donc plus hydrophobes (sensibles à la protéolyse).

Parmi les modèles moléculaires proposés pour schématiser les molécules de mucines [Allen *et al.* 1976 ; Carlstedt *et al.* 1983], le modèle structural couramment retenu est celui décrit par Carlstedt *et al.* à propos de la mucine cervicale humaine. Les mucines selon ce modèle auraient une **forme filamenteuse et linéaire** où les monomères seraient reliés bout à bout par des ponts disulfures pour former un long filament (Figure 1). Chaque monomère comprendrait des zones glycosylées et non glycosylées. Cette structure linéaire a été confirmée par les données de la microscopie électronique. En effet, l'analyse en microscopie électronique montre les mucines de différentes origines tissulaires sous forme de longs filaments de taille variable (de 100 à 5 000 nm) [Slayter *et al.* 1984 ; Sheehan *et al.* 1986 ; Thornton *et al.* 1991].

1.2.4 - Les glycannes

1.2.4.1 - La grande diversité des O-glycannes

La grande majorité des chaînes oligosaccharidiques des mucines sont reliées à l'apomucine par une **liaison α -O-glycosidique** entre un **résidu de GalNAc** et le **groupement hydroxyle** d'un résidu d'aa **Thr** ou **Ser** [Carlson 1968]. Cette liaison est labile à pH alcalin. Les mucines peuvent également contenir quelques chaînes *N*-glycanniques. Les chaînes oligosaccharidiques représentent de 50 à 80 % de la masse moléculaire des mucines.

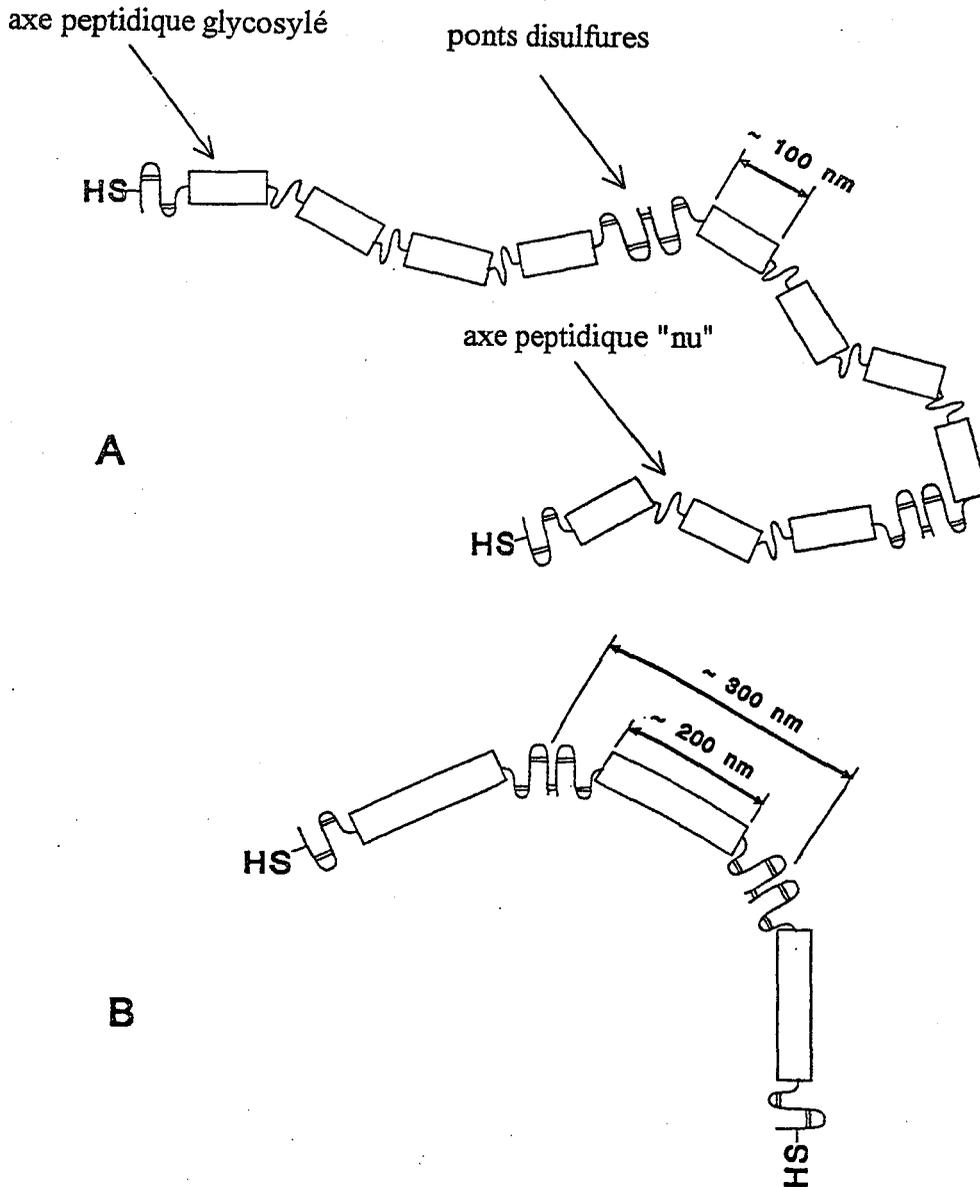


Figure 1 : Représentation schématique de la structure linéaire des oligomères de mucines.

A, Modèle structural de la mucine cervicale humaine d'après Carlstedt [Carlstedt *et al.* 1983]. Ce modèle est basé sur des données de physicochimie et de microscopie électronique. Les monomères, d'une taille uniforme de 500 nm, sont reliés bout à bout pour former des oligomères de différentes tailles (longueur maximale : environ 3 000 nm). Le monomère de mucine est composé de plusieurs parties : une région centrale *O*-glycosylée interrompue par de petites séquences non-glycosylées et sensibles aux protéases. La digestion protéolytique produit 3 à 4 glycopeptides d'environ 100 nm par monomère.

B, Modèle structural de la mucine gastrique de rat d'après Strous et Dekker [Strous et Dekker 1992]. Contrairement à la mucine cervicale humaine, la région centrale *O*-glycosylée de la mucine gastrique de rat ne contient pas de domaines sensibles aux protéases. La longueur des oligomères est très hétérogène et peut atteindre 3 000 nm alors que les monomères ont une taille uniforme d'environ 300 nm. La région glycosylée correspond à environ 72 % de la longueur du monomère.

Il existe une **hétérogénéité considérable** dans la composition, la longueur (1 à 20 sucres), la séquence et l'acidité (neutres, acides, très acides) des chaînes glycaniques des mucines (plus d'une centaine de chaînes distinctes). Cette diversité a été un obstacle majeur à leur étude structurale.

I.2.4.2 - Les trois régions des chaînes *O*-glycaniques

Les chaînes glycaniques des mucines peuvent être décomposées en trois parties (Figure 2) [pour revue, voir Roussel et Lamblin 1996] :

a) Le cœur ou noyau :

La synthèse des chaînes *O*-glycosidiques est initiée par l'action sur l'apomucine d'enzymes très spécifiques, les **UDP-GalNAc-polypeptide- α -*N*-acétylgalactosaminyl transférases**. Ces enzymes sont probablement localisées dans le *cis*-Golgi. Le noyau est constitué du résidu de GalNAc lié à l'axe peptidique par une liaison *O*-glycosidique ainsi que du (ou des deux) sucre(s) fixé(s) sur cette GalNAc. Actuellement, **huit noyaux différents** ont été identifiés bien que le noyau 7 n'ait pas été détecté dans les mucines humaines. Différents types de noyaux, résultant de l'action de nombreuses glycosyl transférases, peuvent être trouvés dans les mucines sécrétées d'un même individu.

b) Le squelette :

La formation du squelette résulte de l'action de différentes glycosyl transférases catalysant le transfert de résidus Gal ou GlcNAc. Le squelette est formé par l'**addition linéaire ou branchée** (sur un résidu interne de Gal du squelette) de **deux types principaux** d'unités disaccharidiques: Gal β (1-3)-GlcNAc (type 1) ou Gal β (1-4)-GlcNAc (type 2).

c) La périphérie :

La périphérie est représentée par les sucres qui viennent se greffer sur le squelette sous l'action de différentes glycosidases. Elle est caractérisée par la présence de Fuc, Gal, GalNAc, acide sialique, le plus souvent en liaison α . Des résidus de sulfate peuvent aussi être additionnés. Ce domaine est porteur des **activités antigéniques** reliées aux **groupes sanguins** (A, B, H, Lewis). Les mucines des sujets de groupe O se terminent par Fuc α (1-2)Gal. Les mucines des sujets sécréteurs de groupe A ou B ont en plus un résidu GalNAc (A) ou un résidu Gal (B). Chez les sujets non sécréteurs (environ 20 % de la population), les oligosaccharides se terminent par Gal seul. Des résidus de Fuc ajoutés à des disaccharides

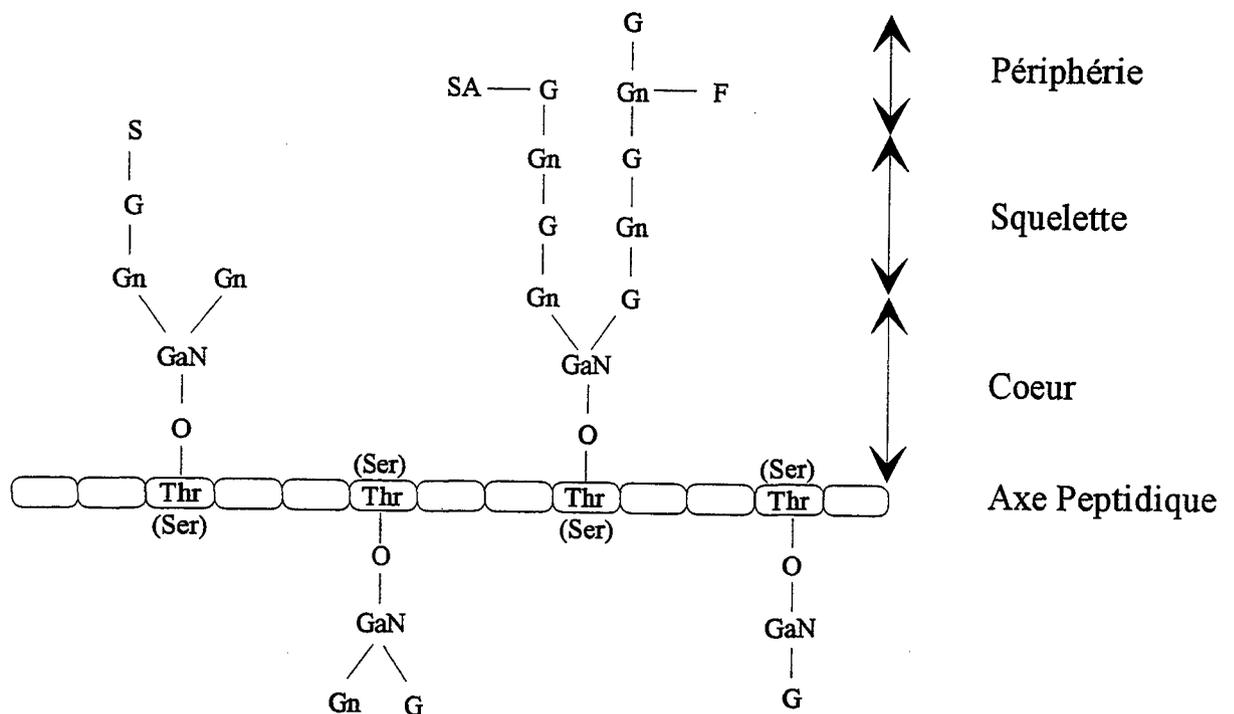


Figure 2 : Représentation schématique d'une région O-glycosylée de mucine [Roussel et Lamblin 1996].

La liaison O-glycosidique s'établit entre un acide aminé hydroxylé [sérine (Ser) ou thréonine (Thr)] et la N-acétylgalactosamine (GaN). F, fucose ; G, galactose ; Gn, N-acétylglucosamine ; M, mannose ; SA, acide sialique ; S, sulfate.

terminaux de type 1 ou de type 2 génèrent différentes structures (H1, Lewis a, Lewis b, H2, X et Y).

La présence de résidus de **sulfate** et d'**acide sialique** terminaux donnent aux mucines leurs propriétés **polyanioniques**. Par chromatographie d'échange d'ions, on peut fractionner les mucines en trois fractions principales selon leur charge :

- 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 de sulfate).

Chacun des trois domaines peut porter des structures reconnaissables par des lectines, des antigènes de groupes sanguins, ainsi que des "déterminants antigéniques associés aux tumeurs". Les structures de groupes sanguins les plus simples sont, au niveau du noyau, GalNAc et Gal β (1-3)GalNAc connus comme les antigènes Tn et T (Thomsen-Friedenreich). Chez l'adulte sain, dans la plupart des tissus, les déterminants antigéniques associés au noyau et au squelette sont masqués par les sucres périphériques.

Des motifs peptidiques consensuels pour la *O*-glycosylation n'ont pu être clairement mis en évidence contrairement à ce qui a été mis en évidence à propos de la *N*-glycosylation (séquence consensus : Asn-Xaa-Ser/Thr avec Xaa \neq Pro). Cependant, Yoshida *et al.* proposent un motif effectif pour la *O*-glycosylation de type mucine : **Xaa-Thr-Pro-Xaa-Pro** [Yoshida *et al.* 1997].

L'étude de MUC1 a montré que l'apomucine est exprimée par les glandes mammaires, le pancréas et l'ovaire. Or, les **formes glycosylées matures** apparaissent **différentes** :

- MUC1 exprimée par les glandes mammaires a une M_r variant entre $250 \cdot 10^3$ et $500 \cdot 10^3$ et contient 50 % de carbohydrates [Shimizu et Yamauchi 1982].
- MUC1 provenant du pancréas a une M_r supérieure à 10^6 et contient ~ 80 % de carbohydrates [Lan *et al.* 1987].
- MUC1 exprimée par les cellules de carcinome du côlon a une M_r de $600 \cdot 10^3$ à $800 \cdot 10^3$ et contient $\sim 80-90$ % de carbohydrates [Baeckstrom *et al.* 1991].

A propos de l'exemple de MUC1, il apparaît donc vraisemblable qu'une même apomucine puisse être glycosylée de façon différente selon les individus, mais également chez

un même sujet selon l'organe, la cellule et l'état physiologique. Il est fort probable que dans une même sécrétion coexistent des molécules processées différemment à partir d'une même apomucine. La séquence primaire de l'apomucine ne semble donc pas dicter à elle seule le **profil de glycosylation**, celui-ci dépendant sûrement de l'**expression d'un registre complexe de glycosyl transférases** et/ou de la **disponibilité des substrats**.

I.2.4.3 - Fonctions des chaînes glycaniques

Les chaînes glycaniques **protègent le squelette protéique** d'une dégradation protéolytique et jouent un rôle important dans les caractéristiques biochimiques et fonctionnelles des mucines : grâce à cette diversité glycanique, il y a **interaction entre les mucines et les micro-organismes** au niveau respiratoire, ce qui est primordial pour leur élimination et la défense de la muqueuse. Au niveau gastrointestinal, l'interaction mucines/souche symbiotique protège la muqueuse des infections par des agents pathogènes. En ce qui concerne la formation du gel de mucus, la charge négative globale attribuée à la mucine par les résidus de sulfate et d'acide sialique produit des forces répulsives **permettant l'hydratation** de la mucine et contribue à la formation d'une structure gélifiée par **captation d'eau**. Les chaînes glycaniques contribuent à la viscosité du gel de mucus non seulement par leur taille, mais peuvent aussi participer à des interactions intermoléculaires intensifiant la formation du gel.

I.2.5 - L'axe peptidique

Les mucines se caractérisent par un pourcentage élevé en Thr, Ser, Pro, Ala, et Gly. Les résidus Thr, Ser et Pro représentent 20 à 50 % de la composition totale en acides aminés [Van Klinken *et al.* 1995]. Les résidus **Ser et Thr** sont les sites d'attache des chaînes O-glycaniques. Le contenu élevé en **Pro** permet la **conformation nécessaire** à une glycosylation importante et au maintien de l'apomucine sous forme étirée.

Les régions glycosylées contiennent la majorité des résidus Thr, Ser et Pro présents dans les mucines. Les régions non glycosylées sont caractérisées par leur richesse en résidus Cys, essentiels au maintien de l'intégrité de la structure macromoléculaire des mucines. C'est dans ces régions que se situent les sites potentiels de **N-glycosylation**.

Les techniques classiques de biochimie utilisées pour déterminer les séquences protéiques : coupures spécifiques par des **protéases** et détermination de la séquence des

peptides ainsi obtenus par la **dégradation d'Edman**, n'ont apporté que peu d'éléments sur la composition peptidique des mucines et n'ont permis d'obtenir que de très courtes séquences en acides aminés (inférieures à 30 acides aminés).

Divers problèmes directement liés à la structure des mucines ont été rencontrés par de nombreux auteurs :

- les très nombreuses chaînes glycaniques masquent l'axe peptidique et limitent l'accès des protéases ;
- l'étape de déglycosylation préalable à l'étude est délicate : une déglycosylation enzymatique est incomplète alors qu'une déglycosylation chimique plus efficace entraîne une dégradation de la partie protéique (coupure des résidus Met, hydrolyse acide) ;
- la richesse en acides aminés hydroxylés (Ser et Thr) donne naissance, lors de la dégradation séquentielle d'Edman, à des composés instables (dérivés phénylthiohydantoïnes) qui rendent l'interprétation impossible au-delà de vingt cycles.

I.3 - Apport de la biologie moléculaire à la connaissance de l'axe peptidique des mucines

I.3.1 - Stratégie d'étude des mucines

L'étude biochimique des mucines a évolué très lentement en raison des difficultés rencontrées lors de la purification des mucines mais aussi du fait de l'hétérogénéité et de la grande taille de ces molécules. Jusqu'à récemment, pour caractériser les mucines, on les purifiait à partir de mucus provenant de tissus ou de lignées cellulaires et on analysait leur composition en acides aminés et en carbohydrates. Quelques résultats très fragmentaires de séquence peptidique ont été obtenus mais leur interprétation restait souvent difficile. Plusieurs laboratoires ont donc développé des **techniques de biologie moléculaire** appliquées aux mucines. **Depuis 1987**, quelques dizaines de séquences partielles ou complètes d'ADNc de mucines animales et humaines ont été publiées. La **stratégie** généralement utilisée comportait les étapes suivantes :

- **purification** de mucines à partir de mucus provenant d'un tissu sain, tumoral ou d'une lignée cellulaire,
- **déglycosylation** chimique en milieu anhydre par l'acide fluorhydrique ou l'acide trifluorométhane sulfonique de mucines entières purifiées ou de fragments glycopeptidiques,
- obtention d'**anticorps** polyclonaux ou monoclonaux anti-apomucines,

- **immunocriblage** de banques d'ADNc construites en vecteur d'expression (λ gt11) à partir des ARNm isolés de la muqueuse ou des cellules étudiées,
- authentification des clones positifs, **séquençage** et déduction de la séquence peptidique,
- **localisation chromosomique**,
- **expression** tissulaire (Northern blots) et cellulaire (Hybridation *in situ* ou HIS) sur tissus sains et pathologiques,
- **clonage** et étude des gènes.

Dix gènes de mucines humaines ont ainsi pu être caractérisés. Selon les conventions internationales de l'HGM (Human Genome Mapping), le gène de mucine est désigné avec les lettres *MUC*. Les gènes de mucines humaines ont donc été désignés : ***MUC1-MUC4***, ***MUC5AC***, ***MUC5B***, ***MUC6-MUC9***. Les lettres AC, B ont permis de distinguer 2 gènes distincts initialement désignés par le symbole *MUC5* attribué à tort à plusieurs ADNc partiels. Par convention, le gène murin est appelé *Muc* suivi du nombre de son homologue humain. L'homologue de rat est désigné *rMuc*.

1.3.2 - Les "mucin-like"

Les études fonctionnelles ont permis de distinguer progressivement deux types de mucines : les mucines **sécrétées** et les mucines **membranaires**. En effet, à l'origine, le mot "mucine" était employé pour désigner les glycoprotéines constituant le gel de mucus. Le terme "mucine" était donc associé à la **notion de sécrétion**. Or, le premier gène de mucine a été cloné, ***MUC1***, code une mucine comportant un domaine transmembranaire hydrophobe et ne formant pas de complexes oligomériques [Gendler *et al.* 1990 ; Lan *et al.* 1990 ; Ligtenberg *et al.* 1990]. Elle fut donc appelée "**mucin-like**".

D'autres *O*-glycoprotéines membranaires trouvées dans des tissus non-épithéliaux sont également parfois appelées "mucin-like" car plus de 50 % de leur masse est due à la présence de chaînes *O*-glycanniques. Ce sont des **protéines membranaires endothéliales** (GlyCAM-1, CD34 ou leucosialine, MadCAM-1) et **leucocytaires** (PSGL-1, CD45RA, CD43, TACTILE ou CD96) [Shimizu et Shaw 1993 ; Van Klinken *et al.* 1995]. Elles contiennent un **domaine de type mucine**, c'est à dire un domaine riche en résidus Ser et Thr et donc potentiellement *O*-glycosylable. Leurs chaînes glycanniques présentent des structures sialyl-Le^a et sialyl-Le^x, déterminants essentiels pour les interactions avec les E- et P-sélectines qui sont des

glycoprotéines membranaires exprimées par les cellules endothéliales. Ces molécules sont impliquées dans la **cascade moléculaire d'interactions (propriétés d'adhérence)** par laquelle les leucocytes migrent du sang dans les tissus, initiant le processus inflammatoire, et jouent un rôle dans la **circulation des lymphocytes T**. Elles se distinguent biochimiquement des mucines épithéliales car elles sont généralement beaucoup plus petites. De plus, contrairement à la plupart des mucines épithéliales, elles ne contiennent pas de répétitions de séquences, ou peu dont le nombre ne varie pas selon les individus.

I.3.3 - Le domaine répétitif des mucines épithéliales

Une grande partie de la région codante de tous les gènes de mucines épithéliales est constituée de la **répétition en tandem d'un certain nombre d'acides aminés** (Tableau 1). Ces unités répétitives, **riches en résidus Thr, Ser, Pro et Ala**, contiennent de **8 à 169 aa** et sont répétées de **5 à plus de 100 fois** suivant la mucine considérée. Il n'existe pas de similarités, ni en séquence nucléotidique ou peptidique, ni en nombre de nucléotides par unité répétée, entre les différents gènes de mucines. Cependant, dans tous les cas, cette région répétitive code un domaine riche en résidus Thr et Ser et donc **potentiellement fortement O-glycosylé**.

I.3.4 - Polymorphisme de type VNTR

Pour la plupart des gènes de mucines, le nombre de répétitions en tandem varie d'un individu à l'autre, donnant lieu à un **polymorphisme génomique de type VNTR** pour "Variable Number of Tandem Repeats". Le domaine répétitif a donc une taille variable d'un allèle à l'autre. Cette variation dans le nombre d'unités répétitives serait due à des "crossing-overs" inégaux mais homologues lors de la méiose.

Concernant *MUC1*, le nombre de répétitions de 60 nucléotides peut varier de 20 à 125 par allèle [Gendler *et al.* 1990], d'où une variation de **la taille du gène de 4 à 7 kb**. Les tailles les plus communes du domaine répétitif varient de 820 à 1700 aa, soit de 41 à 85 répétitions.

Pour *MUC2*, le nombre d'unités répétitives par allèle varie de **51 à 115** (soit une variation de la région O-glycosylable de 3,5 à 8,7 kb), la majorité des allèles contenant environ 100-115 répétitions en tandem [Griffiths *et al.* 1990 ; Toribara *et al.* 1991].

Gène	Longueur du motif répétitif (aa)	Séquence peptidique consensus du motif répétitif	Référence
<i>MUC1</i>	20	G S T A P P A H G V T S A P D T R P A P	Gendler <i>et al.</i> 1988
<i>MUC2</i>	23	P T T T P I T T T T T V T P T P T P T G T Q T	Gum <i>et al.</i> 1989
<i>MUC3</i>	17	H S T P S F T S S I T T T E T T S	Gum <i>et al.</i> 1990
<i>MUC4</i>	16	T S S A S T G H A T P L P V T D	Porchet <i>et al.</i> 1991
<i>MUC5AC</i>	8	T T S T T S A P	Aubert <i>et al.</i> 1991 Guyonnet Dupérat <i>et al.</i> 1995
<i>MUC5B</i>	29	A T G S T A T P S S T P G T T H T P P V L T T T A T T P T	Aubert <i>et al.</i> 1991 Desseyn <i>et al.</i> 1997c
<i>MUC6</i>	169	S P F S S T G P M T A T S F Q T T T T Y P T P S H P Q T T L P T H V P P F S T S L V T P S T G T V I T P T H A Q M A T S A S I H S T P T G T I P P P T T L K A T G S T H T A P P M T P T T S G T S Q A H S S F S T A K T S T S L H S H T S S T H H P E V T P T S T T T I T P N P T S T G T S T P V A H T T S A T S S R L P T P F T T H S P P T G S	Toribara <i>et al.</i> 1993
<i>MUC7</i>	23	T T A A P P T P S A T T P A P P S S S A P P E	Bobek <i>et al.</i> 1993
<i>MUC8</i>	13 et 41	T S C P R P L Q E G T R V T S C P R P L Q E G T P G S R A A H A L S R R G H R V H E L P T S S P G G D T G F	Shankar <i>et al.</i> 1994
<i>MUC9</i>	15	G E K T L T P V G H Q S V T P	Lapensée <i>et al.</i> 1997

Tableau 1 : Séquences peptidiques consensus des motifs répétitifs élémentaires des mucines humaines.

Gène	Localisation chromosomique	Référence
<i>MUC1</i>	1q21-24	Swallow <i>et al.</i> 1987b
<i>MUC2</i>	11p15.5	Griffiths <i>et al.</i> 1990
<i>MUC3</i>	7q22	Fox <i>et al.</i> 1992
<i>MUC4</i>	3q29	Gross <i>et al.</i> 1992
<i>MUC5AC</i>	11p15.5	Nguyen <i>et al.</i> 1990
<i>MUC5B</i>	11p15.5	Nguyen <i>et al.</i> 1990
<i>MUC6</i>	11p15.5	Toribara <i>et al.</i> 1993
<i>MUC7</i>	4q13-q21	Bobek <i>et al.</i> 1996
<i>MUC8</i>	12q24.3	Shankar <i>et al.</i> 1997
<i>MUC9</i>	13	Lapensée <i>et al.</i> 1997

Tableau 2 : Localisation chromosomique des gènes de mucines humaines.

Récemment, Vinall *et al.* ont étudié le polymorphisme de *MUC2*, *MUC5AC*, *MUC5B* et *MUC6* [Vinall *et al.* 1998]. Le plus petit allèle observé pour *MUC2* par l'équipe de San Francisco est d'environ 3,5 kb et le plus grand d'environ 14 kb, soit **une différence de plus de 150 unités répétitives** de 23 aa. Ces nombres indiquent que les différents allèles de *MUC2* codent un polypeptide naissant dont la M_r varierait de 350.10^3 à 680.10^3 [Gum *et al.* 1994]. La différence de taille entre le plus petit et le plus grand allèle observés jusqu'à présent pour *MUC6* est de 5 kb, ce qui correspond à une M_r de 170.10^3 . ***MUC5AC* ne présente que très peu de variations** en ce qui concerne la longueur du domaine répétitif [Pigny *et al.* 1995]. Quant à ***MUC5B*, aucun polymorphisme de type VNTR** n'a pu être mis en évidence dans le domaine répétitif.

En ce qui concerne *MUC7*, une étude préliminaire a montré un **très faible polymorphisme** : la majorité des individus étudiés ont un gène dont le domaine répétitif contient **six répétitions**, seuls quelques individus ont un gène avec cinq répétitions [Bobek *et al.* 1996].

La région répétitive polymorphique de *MUC1* est **transcrite et traduite**, et donc ce polymorphisme s'observe également au niveau de l'apomucine [Swallow *et al.* 1987a ; Gendler *et al.* 1987 ; Ligtenberg *et al.* 1990]. De plus, les produits peptidiques sont **codominants**. Debailleul *et al.* ont également observé des variations alléliques de la longueur des transcrits des différents gènes de mucines (*MUC2* à *MUC6*) [Debailleul *et al.* 1998]. Ils ont démontré que ces variations étaient directement reliées à un polymorphisme de type VNTR observé au niveau de l'ADN. Le fait que le **domaine VNTR soit en région codante** est une caractéristique importante des gènes de mucines. En effet, généralement, les régions hypervariables de l'ADN humain sont observées dans des zones non codantes (ex : région VNTR en 5' du gène de l'insuline [Bell *et al.* 1982] ou en 5' de *H-RAS1* [Krontiris *et al.* 1985]).

Les mucines ayant un rôle dans la protection de la surface des cellules épithéliales, on peut se demander quelles sont les conséquences fonctionnelles et la signification biologique de ces variations dans la longueur des apomucines. Au vu d'une telle variation, il semblerait que la longueur précise des mucines ne soit pas primordiale pour leur fonction [Gendler et Spicer 1995 ; Lancaster *et al.* 1990]. Cependant, **cette variation de taille du domaine glycosylé pourrait, au delà d'un certain seuil, influencer les propriétés physicochimiques**

du mucus, conférant aussi au sujet une protection plus ou moins efficace des muqueuses et donc une susceptibilité plus ou moins grande à développer des pathologies inflammatoires ou tumorales des muqueuses. Carvalho *et al.* ont observé une différence significative dans la distribution des allèles de *MUC1* entre un groupe témoin et un groupe de patients atteints de carcinome gastrique [Carvalho *et al.* 1997]. En effet, de petits allèles de *MUC1* sont plus fréquents dans ce dernier groupe. L'hypothèse des auteurs de ce travail est que des individus ayant de petits allèles *MUC1* sont plus prédisposés à développer un carcinome gastrique. Ces individus auraient une couche de mucus plus mince et seraient donc plus sensibles aux agressions provenant de l'environnement telles que enzymes et sucs digestifs, toxines ou micro-organismes. De même, Kyo *et al.* suggèrent que des individus possédant de rares allèles de *MUC3* seraient prédisposés à développer une rectocolite ulcéro-hémorragique [Kyo *et al.* 1999].

I.3.5 - Expression des gènes de mucines épithéliales humaines

I.3.5.1 - Expression tissulaire analysée par la technique de Northern blot

Jusqu'à présent, sur les 10 gènes de mucines caractérisés, seules les séquences complètes de *MUC1* et de *MUC7* ont été publiées rapidement. Cela est surtout dû à leur petite taille par rapport aux autres gènes de mucines. Leur transcrit peut être détecté par analyse en Northern blot. Jusqu'à récemment, on observait pour tous les autres gènes *MUC* un profil d'expression par Northern blot présentant un **haut degré d'hétérogénéité**. Cette hétérogénéité était même considérée comme une caractéristique typique des mucines, bien qu'on ne sache pas l'expliquer. Debailleul *et al.* ont démontré que, en améliorant les méthodes de purification et de transfert de l'ARN, quel que soit le tissu étudié, **les gènes *MUC2-6* expriment des ARNm stables et d'une taille inhabituellement grande pour des ARN d'eucaryotes (14-24 kb)** [Debailleul *et al.* 1998].

L'étude de l'expression tissulaire par la technique de Northern blot a permis de dégager une notion importante : il n'y a pas d'expression exclusive d'un seul gène de mucine dans un type de muqueuse donné, mais un profil complexe d'expression sur le plan qualitatif et quantitatif. L'expression des gènes de mucines humaines est spécifique du tissu, chaque *MUC* possédant un territoire majeur d'expression :

- **MUC1** est exprimé dans la plupart des tissus épithéliaux et en particulier dans les canaux et les glandes exocrines, avec une expression majoritaire dans le **sein** et le **pancréas** [Gendler et Spicer 1995 ; Lan *et al.* 1990],

- **MUC2** est dominant dans le **côlon**, mais il est aussi fortement exprimé dans l'intestin **grêle** [Gum *et al.* 1989],

- **MUC3** est détecté majoritairement au niveau de l'**intestin grêle**, mais également au niveau du **côlon** et de la **vésicule biliaire** [Gum *et al.* 1990 ; Vandenhautte *et al.* 1997],

- **MUC4** semble beaucoup plus **ubiquitaire**, bien qu'il ne soit pas exprimé dans les canaux des annexes du tube digestif (pancréas et vésicule biliaire) [Porchet *et al.* 1991],

- **MUC5AC** montre une expression importante au niveau **respiratoire** et **gastrique** [Guyonnet Dupérat *et al.* 1995],

- **MUC5B** s'exprime spécifiquement et intensément dans la **bronche** et les **glandes sous-maxillaires** ; il est aussi retrouvé au niveau de la **vésicule biliaire** [Dufossé *et al.* 1993 ; Vandenhautte *et al.* 1997],

- **MUC6** est présent au niveau des glandes sous-muqueuses de l'**antre pylorique** et dans les canaux de la **vésicule biliaire** et du **pancréas** [Toribara *et al.* 1993 ; Vandenhautte *et al.* 1997],

- **MUC7** offre une expression restreinte exclusivement aux **glandes salivaires** et **bronchiques (cellules séreuses)** [Bobek *et al.* 1993],

- **MUC8** semble être exprimé dans la **trachée** et l'**estomac** [Shankar *et al.* 1997], mais peu d'études d'expression ont concerné ce gène jusqu'à présent.

I.3.5.2 - Expression cellulaire par la technique d'HIS

A cette spécificité tissulaire s'ajoute une **spécificité cellulaire**. Le tableau 3 résume les résultats obtenus dans notre laboratoire. Par exemple, dans la bronche, **MUC5B** est exprimé spécifiquement dans les **glandes sous-muqueuses**. Dans l'intestin, **MUC2** est exprimé dans les **cellules caliciformes** alors que **MUC3** est exprimé dans les cellules caliciformes et les **entérocytes** [Chang *et al.* 1994 ; Audié *et al.* 1993]. De même, dans l'estomac, **MUC5AC** s'exprime uniquement dans les **cellules caliciformes** de l'antre et du fundus, alors que les transcrits de **MUC6** sont présents essentiellement dans les **glandes sous-muqueuses** de l'antre.

Muqueuses	Gènes						
	<i>MUC2</i>	<i>MUC3</i>	<i>MUC4</i>	<i>MUC5B</i>	<i>MUC5AC</i>	<i>MUC6</i>	<i>MUC7</i>
Glandes salivaires	G: -	G: -	G: -	G: ++	G: -	G: -	G: +++
Bronches	S: ++ G: +	S: - G: -	S: ++ G: -	S: -/+ G: ++	S: -/++++ G: -	S: - G: -	S: - G: +++
Fundus	S: - G: -	S: -/+ G: -	S: -/+ G: -	S: - G: -	S/Ct: ++++ G: -	S: - Ct: + G: -	S: - G: -
Antre	S: - G: +	S: ++ G: -	S: + G: -	S: - G: -	S: ++++ G: -	S: - G: +++	S: - G: -
Duodénum	S: ++++ C: ++++ G: +	S: +++ C: -/+	S: - C: -	S: - C: -	S: - C: -	S: - C: -	S: - C: -
Jéjunum	S: ++++ C: ++++ G: -	S: +++ C: -/+	S: - C: -	S: - C: -	S: - C: -	S: - C: -	S: - C: -
Iléon	S: ++++ C: ++++ G: -	S: +++ C: -/+	S: -/+ C: -/+	S: - C: -	S: - C: -	S: - C: -	S: - C: -
Côlon	S: ++++ C: ++++ G: -	S: ++ C: -	S: ++ C: ++	S: - C: -	S: - C: -	S: - C: -	S: - C: -
Pancréas	Ca: -/+	Ca: +++	Ca: -	Ca: ++	Ca: -	Ca: +	Ca: -
Vésicule biliaire	S: + I: +	S: +++ I: +++	S: - I: -	S: ++ I: ++	S: + I: +	S: + I: +++	S: - I: -
Prostate	G: -	G: -	G: ++	G: -	G: -	G: -	G: -
Endocol	S: + G: +	S: - G: -	S: ++ G: ++	S: + G: +	S: ++ G: ++	S: + G: +	G: - S: -

Tableau 3 : Profil d'expression des gènes de mucines humaines chez l'adulte analysé par hybridation *in situ*.

C : cryptes ; Ca : cellules canalaire ; Ct : collet ; G : épithélium glandulaire ; I : invaginations de l'épithélium ; S : épithélium de surface

Marquage : ++++ : de très forte intensité ; +++ : de forte intensité ; ++ : d'intensité modérée ; + : de faible intensité ; - : absent

II - CONNAISSANCES ACTUELLES SUR LES GENES D'APOMUCINES EPITHELIALES ET SUR LEURS PRODUITS

La séquence complète de l'ADNc n'est connue jusqu'à présent que pour quatre gènes de mucines épithéliales humaines : celui de la mucine membranaire MUC1 et ceux de trois mucines sécrétées, MUC2, MUC5B et MUC7. En ce qui concerne les autres mucines épithéliales, seules des séquences partielles contenant les répétitions en tandem et quelques séquences uniques ont été publiées.

II.1 - MUC1 : gène codant une mucine transmembranaire

MUC1 est une mucine transmembranaire. La mucine MUC1 avait été caractérisée bien avant le gène. La découverte du gène a permis de clarifier les très nombreuses dénominations différentes qu'elle avait pu recevoir, selon les études et selon les milieux biologiques étudiés. Elle est exprimée à la **surface apicale** de la plupart des épithéliums sécrétoires simples. Elle est surexprimée de façon considérable dans les carcinomes, avec une localisation cellulaire anormale, tout autour de la cellule. Elle représente donc un **marqueur important de malignité** [Taylor-Papadimitriou *et al.* 1981 ; Burchell *et al.* 1987 ; Zotter *et al.* 1988 ; Girling *et al.* 1989]. Des anticorps dirigés contre des carcinomes provenant de différents tissus réagissent avec différents épitopes de MUC1 (70 recensés actuellement !). Comme plusieurs épitopes correspondant à cette mucine épithéliale ont été décrits indépendamment, l'apomucine MUC1 a été désignée de nombreuses façons : HMFG, EMA, PAS-O, DUPAN-2, PUM, Ca1, NPGP, NCRC11, épitectine, MAM-6, DF3, SGA, H23, PEM et épisialine [Gendler et Spicer 1995], avant que la découverte du gène n'éclaircisse la nomenclature.

MUC1 a été le premier gène de mucine à être cloné. Les clones d'ADNc de *MUC1* ont été obtenus à partir d'ARN de glandes mammaires, de pancréas et d'ovaires. L'ADNc complet de *MUC1* [Gendler *et al.* 1990 ; Lan *et al.* 1990] et celui de son homologue murin [Spicer *et al.* 1991] ont été publiés ainsi que l'organisation des clones génomiques [Lancaster *et al.* 1990 ; Spicer *et al.* 1991]. *MUC1* a été localisé sur le bras long du **chromosome 1** humain dans une région dense en gènes **en q21-24** [Swallow *et al.* 1987b] (Tableau 2).

La séquence codante complète de *MUC1* est contenue dans **7 exons**. Le gène humain *MUC1*, dont la taille varie **entre 4 et 7 kb** selon le nombre de répétitions, code **une protéine transmembranaire** comportant différents domaines (Figure 3) :

- un **domaine amino-terminal** constitué du peptide signal et de répétitions en tandem dégénérées s'étendant sur 127 aa,

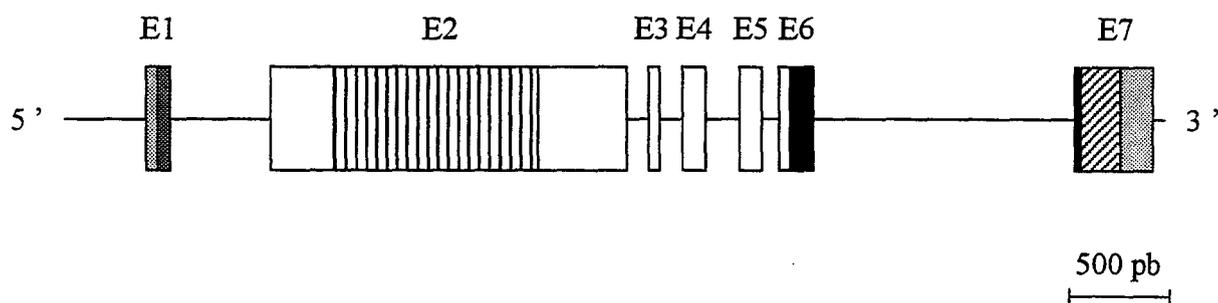
- un **grand domaine central** comprenant un nombre variable de répétitions presque parfaites de 20 aa, ce domaine répétitif ayant la caractéristique d'être porté par un seul exon, qui est aussi le plus grand,

- un **domaine carboxy-terminal** constitué d'une région de 227 aa (riche en Ser, Thr, Pro) comprenant des répétitions en tandem dégénérées et une séquence unique de type mucine, d'un **domaine transmembranaire** hydrophobe de 31 aa, et d'un domaine cytoplasmique de 69 aa.

Deux variants de transcription ont été décrits [Ligtenberg *et al.* 1990 ; Wreschner *et al.* 1990]. Ils résulteraient de l'utilisation alternative de deux sites accepteurs lors de l'épissage de l'exon 2. Ces variants **diffèrent de 27 pb** et codent des protéines immatures dont le **peptide signal est différent** et donc des protéines matures dont l'extrémité amino-terminale est différente. Ces variants ne résultent pas d'un épissage alternatif mais sont deux variants alléliques ne différant que par un seul nucléotide (A/G) dans l'exon 2, situé 8 nucléotides en aval du second site accepteur d'épissage [Ligtenberg *et al.* 1991].

Une **forme sécrétée** privée du domaine transmembranaire est produite semble-t-il par épissage alternatif ; une **forme soluble** vraisemblablement obtenue par protéolyse de la forme membranaire est retrouvée dans les liquides biologiques de l'organisme. Enfin, plusieurs formes dépourvues du domaine répétitif *O*-glycosylé ont été observées ; ces formes, dénommées **MUC1/Y** et **MUC1/Z** [Zrihan-Licht *et al.* 1994 ; Oosterkamp *et al.* 1997], ne sont plus à proprement parler des mucines. Le rôle de ces différents variants demeure encore inconnu.

La masse moléculaire calculée du squelette peptidique de MUC1 est d'environ 120 à 225 kDa, soit une masse supérieure à 300 kDa pour la protéine glycosylée. Le domaine répétitif central est riche en résidus Pro (25 %), Thr (15 %) et Ser (10 %). Cette région est donc fortement *O*-glycosylable. La séquence déduite en acides aminés contient aussi 5 sites potentiels de *N*-glycosylation. Des études ont montré que MUC1 était réellement



-  Régions 5' et 3' non traduites
-  Région codant le peptide signal
-  Domaine répétitif (20 à 125 répétitions de 60 pb)
-  Région codant le domaine transmembranaire
-  Région codant le domaine cytoplasmique

Figure 3 : Représentation schématique de l'organisation génomique de *MUC1* et de l'organisation peptidique déduite [Gendler et Spicer 1995].
Les rectangles représentent les 7 exons (E1-E7).

N-glycosylée [Hilkens et Buijs 1988 ; Linsley *et al.* 1988]. La région répétitive et les régions adjacentes également riches en résidus Ser/Thr constituent le domaine de type mucine. Il représente la majeure partie de la protéine et a une structure filamenteuse. Ce domaine de type mucine s'élève de 200 à 500 nm au-dessus de la membrane cellulaire, ce qui est beaucoup plus que les autres protéines membranaires [Jentoft 1990].

II.2 - Gènes codant les mucines sécrétées formant le gel de mucus

Quatre gènes de mucines humaines sont regroupés dans la même région du chromosome 11 en p15.5. : *MUC2*, *MUC5AC*, *MUC5B* et *MUC6*. Il existe donc un complexe de gènes *MUC* dans cette région du chromosome. Une étude par électrophorèse en champs pulsés réalisée au laboratoire a permis de déterminer l'ordre de ces 4 gènes sur le chromosome 11 et de les orienter par rapport aux autres gènes de la région [Pigny *et al.* 1996a] :

Télomère-*HRAS-MUC6-MUC2-MUC5AC-MUC5B-IGF2*-Centromère

II.2.1 - MUC2

Plusieurs clones d'ADNc ont été isolés par criblage d'une banque d'expression de muqueuse intestinale humaine, à l'aide d'un anticorps anti-peptide [Gum *et al.* 1989] mais également à partir d'une banque d'ADNc de trachée humaine [Gerard *et al.* 1990 ; Jany *et al.* 1991]. Le gène *MUC2* a été localisé en 11p15.5 [Griffiths *et al.* 1990]. La séquence complète de l'ADNc de *MUC2* s'étend sur 15 720 pb pour la forme allélique la plus commune, c'est à dire contenant 100 répétitions soit environ 6 900 pb (2 300 aa) [Gum *et al.* 1994]. La mucine *MUC2* est présente dans l'intestin grêle, le côlon et l'appareil respiratoire [Gum *et al.* 1989 ; Gerard *et al.* 1990 ; Jany *et al.* 1991 ; Shankar *et al.* 1995]. *MUC2* contient une séquence répétée en tandem de 23 aa, riche en résidus Thr et Pro et relativement pauvre en résidus Ser. Ce domaine répétitif correspond à la fraction acide isolée par Wesley *et al.* [Wesley *et al.* 1985] dans le mucus intestinal et non à la fraction neutre plus abondante.

Le segment répétitif central consiste en deux régions répétitives de tailles inégales, séparées d'environ 600 pb et situées toutes les deux dans le même exon [Toribara *et al.* 1991]. Ces deux régions répétitives constituent plus de 50 % de l'ADNc de *MUC2*. La région 1 la plus en 5' et la plus petite, est constituée de répétitions de 48 pb interrompues par des segments de 21-24 pb et codant un domaine peptidique également riche en résidus Thr et Pro, d'une longueur de 347 aa. Ce domaine est encadré par deux séquences peptidiques très

similaires et contenant des résidus Cys [Guyonnet Dupérat *et al.* 1995]. Ce motif de 130 aa est appelé "**motif Cys**". Cette région 1 ne présente pas de polymorphisme. La région 2 est constituée de répétitions en tandem de 69 pb non interrompues.

La **partie C-terminale** (984 aa) est constituée [Gum *et al.* 1992b] :

- d'un petit domaine de type mucine (139 aa),
- d'un long domaine (845 aa) riche en résidus Cys (>10 %) et possédant 18 sites potentiels de *N*-glycosylation.

La **partie N-terminale** est elle aussi riche en résidus Cys.

Les domaines riches en résidus Cys contiennent 4 domaines répétés de 350 acides aminés, présentant des **similarités de séquence avec les domaines D du pro-facteur de von Willebrand (vWF)** (Figure 4) [Gum *et al.* 1994]. La **partie N-terminale** comporte quatre domaines D homologues aux domaines **D1, D2, D' et D3 du vWF**. La **partie C-terminale** possède un domaine fortement homologue au domaine **D4 du vWF**. La position de ces domaines est similaire dans les deux protéines. Ils constituent plus de 25 % du polypeptide MUC2.

II.2.2 - MUC5B

Les premiers clones d'ADNc correspondants à *MUC5B* ont été isolés dans le laboratoire à partir d'une banque d'ADNc d'origine **trachéobronchique** [Aubert *et al.* 1991]. Ils sont caractérisés par des **répétitions en tandem dégénérées de 87 nucléotides (29 aa)** [Dufossé *et al.* 1993].

MUC5B est le seul gène de mucine sécrétée de grande taille dont la structure génomique est entièrement connue. Le gène *MUC5B* est constitué de **48 exons** répartis sur **36,5 kb**. Les **exons** ont des tailles qui varient de **32 pb à 10 713 pb**. Les **47 introns** ont des tailles comprises entre **87 pb et environ 1 700 pb** [Desseyn *et al.* 1998b].

Le transcrit de *MUC5B* a une taille de **17 592 pb** (plus la queue polyA⁺) et code une apomucine de 5 664 aa. Desseyn *et al.* ont subdivisé cette molécule en trois parties :

- la **région N-terminale** : 1 283 aa ; 10 sites potentiels de *N*-glycosylation ; 29 exons répartis sur environ 15,2 kb [Desseyn *et al.* 1998b].

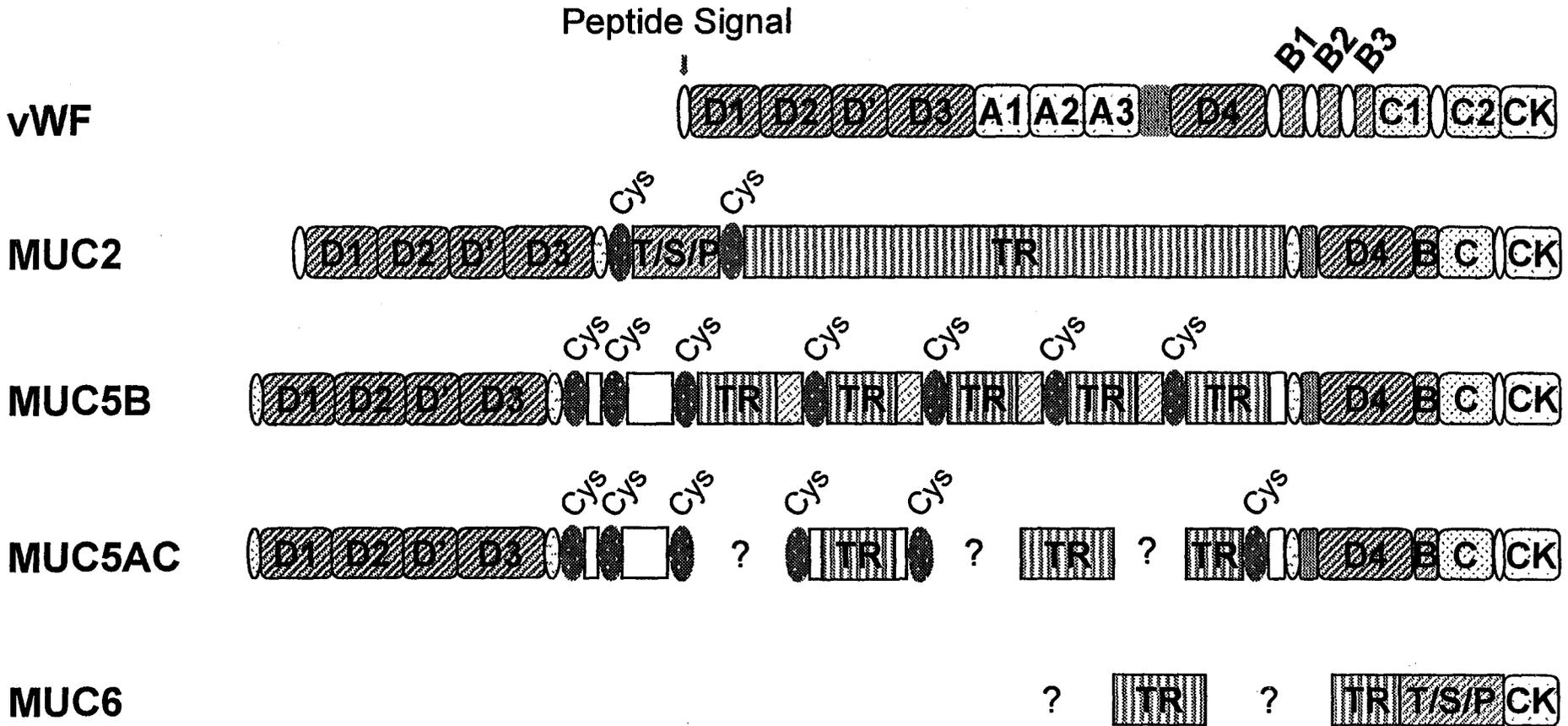


Figure 4 : Organisation peptidique comparée du pro-vWF, de MUC2, de MUC5B, de MUC5AC et de MUC6.

- la **région centrale** : 7 sites potentiels de *N*-glycosylation ; 3 570 aa codés par un exon unique de 10 713 pb [Desseyn *et al.* 1997c].

- la **région C-terminale** : 808 aa ; 15 sites potentiels de *N*-glycosylation ; 18 exons répartis sur 10,6 kb [Desseyn *et al.* 1997b].

L'exon central est l'un des plus grands décrits jusqu'à présent pour un gène de mammifère. Le peptide déduit comprend **19 sous-domaines. 4 super-répétitions de 528 aa (UpA à UpD)** peuvent également être observées. Chaque super-répétition est composée de :

- un sous-domaine constitué de répétitions irrégulières en tandem (11, 17 ou 23 fois) de 29 aa riches en Thr (37 %), Ser (15 %) et Pro (11 %) (ce sous-domaine est retrouvé 5 fois dans MUC5B). Chacun de ces sous-domaines possède un site potentiel de *N*-glycosylation,

- une séquence unique de 111 aa également riche en Ser/Thr (Thr + Ser = 48,6 %) et en Pro (14,2 %), appelée R-end,

- un sous-domaine de 108 aa riche en résidus Cys (10 Cys soit 9,3 %). Ce dernier sous-domaine (trouvé 7 fois dans MUC5B) présente des similarités avec les motifs Cys retrouvés 2 fois dans MUC2 (Figure 4).

Les régions N- et C-terminales de MUC5B sont riches en résidus Cys et sont similaires, comme pour MUC2, aux domaines D qui flanquent les domaines A(A1-A2-A3) du pro-vWF.

Des séquences répétitives ont été trouvées **dans 4 des 18 introns** de la région 3' de *MUC5B*. La séquence peptidique de la région C-terminale de *MUC5B* comporte une région 100 % homologue à la séquence peptidique déduite du clone **pSM2-1**. Cet ADNc code 196 aa de la région carboxy-terminale de la mucine salivaire de haut poids moléculaire **MG1** [Troxler *et al.* 1995]. *MUC5B* s'exprimant dans les glandes salivaires, Desseyn *et al.* en ont conclu que *MUC5B* code la mucine salivaire MG1. Deux autres articles concernant la région 3' de *MUC5B* ont été publiés [Keates *et al.* 1997 ; Nielsen *et al.* 1997]. Offner *et al.* ont également cloné la région 5' terminale de *MUC5B* [Offner *et al.* 1998]. Pour ces auteurs, la région N-terminale de *MUC5B* est plus longue que celle décrite précédemment [Desseyn *et al.* 1998 b]. Elle contiendrait 1 321 aa organisés en un peptide signal, une courte région riche en Ser/Thr et trois domaines D.

II.2.3 - *MUC5AC*

Les premiers clones d'ADNc partiels de *MUC5A* et *MUC5C* ont été isolés dans le laboratoire à l'aide d'un antisérum dirigé contre des mucines déglycosylées **trachéobronchiques** [Crépin *et al.* 1990 ; Aubert *et al.* 1991]. Guyonnet Dupérat *et al.* ont démontré, par une étude électrophorétique en champs pulsés, que les clones de *MUC5A* et *MUC5C* provenaient du même gène, dénommé ensuite *MUC5AC* [Guyonnet Dupérat *et al.* 1995]. *MUC5AC* a été localisé en 11p15.5 [Nguyen *et al.* 1990].

Un des clones étudiés, **JER47** (1,4 kb), contient des répétitions en tandem de 24 pb codant la séquence peptidique consensus de 8 aa : **TTSTTSAP**. Ces répétitions sont interrompues par deux domaines de 130 aa riches en Cys. Ces domaines contiennent les **séquences peptidiques TR-3A et TR-3B** précédemment décrites et isolées d'une mucine humaine trachéobronchique par Rose *et al.* [Rose *et al.* 1989]. **JER58** (0,8 kb), un autre clone séquencé, ne contient que des répétitions de 24 pb. La séquence en aa déduite de ces différents clones suggère que le domaine répétitif central est composé de répétitions en tandem alternant avec des domaines riches en Cys. Ces derniers contiennent dix Cys et sont similaires aux **motifs Cys** retrouvés deux fois dans la séquence de MUC2 et sept fois dans celle de MUC5B.

En dehors de notre laboratoire, d'autres équipes ont publié des séquences partielles d'ADNc de *MUC5AC* :

- Meerzaman *et al.* ont isolé un clone d'ADNc (**NP3a** ; 3,6 kb) à partir d'une banque de polype nasal [Meerzaman *et al.* 1994]. La séquence peptidique déduite de ce clone contient les peptides **TR-3A** et **TR-3B** et a donc été rapportée à *MUC5AC*, bien qu'elle ne contienne pas de séquence consensus de 8aa. Ce clone NP3a contient un signal de polyadénylation (ATAAATA) et représenterait donc l'**extrémité C-terminale** de *MUC5AC*.

- **L31** (3 310 pb) est un clone d'ADNc isolé par Lesuffleur *et al.* à partir d'une banque d'ADNc de cellules HT-29 MTX, lignée cellulaire de cancer du côlon résistante au méthotrexate (MTX) et exprimant des mucines de type gastrique [Lesuffleur *et al.* 1995]. L31 présente une forte homologie avec NP3a au niveau nucléotidique (98,6 %). L31 code une protéine de 1 042 aa riche en Cys (9,2 %) et comprenant un domaine homologue au **domaine**

D4 du vWF (Figure 4). L31 correspondrait également à l'**extrémité C-terminale** de *MUC5AC*.

- Le clone **HGM-1** (2 551 pb) a été obtenu à partir d'une banque d'expression d'estomac humain [Klomp *et al.* 1995]. Il code un peptide constitué de 4 domaines riches en Cys dont le premier est homologue au **domaine D3 du vWF**. Les trois autres domaines correspondent à trois motifs Cys identiques au motif retrouvé dans la région C-terminale de *MUC5AC*, trouvé au moins deux fois dans la région centrale de *MUC5AC*, deux fois dans *MUC2* et sept fois dans *MUC5B*. Ces domaines sont séparés par trois courtes séquences non répétitives riches en Ser/Thr. HGM-1 correspond à l'**extrémité N-terminale** de *MUC5AC*.

- Récemment, Li et al. ont publié une séquence nucléotidique de 3,3 kb [Li *et al.* 1998]. Cette séquence a été obtenue par différentes expériences de 5' RACE-PCR à partir de l'extrémité 5' de l'ADNc HGM-1. Elle contient un site potentiel d'initiation de la traduction suivi d'une **séquence signal potentielle de 45 nucléotides**. La séquence nucléotidique ne contient pas de répétitions en tandem mais présente des similarités (> 60 %) avec la séquence en 5' de *MUC2*. La position des Cys est conservée à 100 % entre les deux gènes. *MUC5AC* contient, comme *MUC2*, trois domaines homologues aux domaines D1, D2 et D3 du vWF dans sa partie N-terminale (Figure 4).

- La région 5' de *MUC5AC* a également été caractérisée par Van de Bovenkamp *et al.* [Van de Bovenkamp *et al.* 1998]. Ces auteurs ont observé l'existence d'un **motif à répétition de leucines** de type "**leucine zipper**" ou "**fermeture Eclair**" à leucines, dans le domaine D1. Un tel motif est retrouvé dans différentes protéines dimériques telles que les facteurs de transcription. En formant une structure dite hélice α -super-enroulée, le motif de "**fermeture Eclair**" à leucines pourrait faciliter la dimérisation par des interactions non covalentes : deux hélices α , une par monomère, se regrouperaient par des interactions entre les chaînes latérales des leucines (aa hydrophobes). La même équipe a également estimé la taille du domaine répétitif par une technique de PCR réalisée sur de l'ADN génomique. Le domaine répétitif de *MUC5AC* aurait une taille d'environ 2 500 aa (7,5 kb). Ces auteurs suggèrent que l'ADNc total de *MUC5AC* a une longueur d'environ **16,6 kb** et code un peptide de 5 525 aa.

Au laboratoire, Marie-Pierre Buisine et Fabienne Escande ont pu, grâce à des expériences utilisant la PCR, établir la continuité physique entre les séquences uniques distales et le domaine répétitif de *MUC5AC* [Buisine *et al.* 1998a ; Escande 1998].

En ce qui concerne l'organisation génomique de *MUC5AC*, seule la partie 3' est connue [Buisine *et al.* 1998a]. Buisine *et al.* ont montré que l'organisation génomique des exons et des introns de *MUC5AC* est très proche de celle de *MUC5B* et que ces deux gènes ont la même orientation transcriptionnelle (figure 5A).

II.2.4 - MUC6

Le premier clone d'ADNc de *MUC6* a été obtenu par immunocriblage d'une banque commerciale d'ADNc gastrique préparée en vecteur d'expression λ gt11, à l'aide d'un antisérum dirigé contre des mucines gastriques humaines déglycosylées [Toribara *et al.* 1993]. Sa séquence révèle un **motif de 507 pb (169 aa) répété en tandem**. Ce motif représente la plus grande unité répétitive de mucine publiée jusqu'à présent (Tableau 1). Ces répétitions ont une composition en aa typique des mucines puisque les résidus Thr (30 %), Ser (18 %) et Pro (15 %) représentent 63 % du nombre total d'aa.

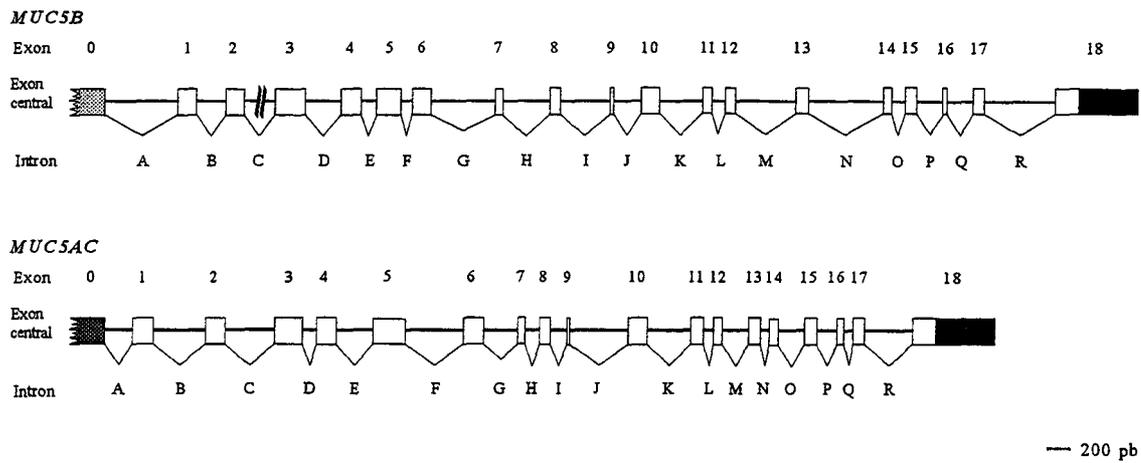
Toribara *et al.* ont publié en 1997 la séquence de la région 3' de *MUC6* [Toribara *et al.* 1997]. La **région 3' unique** de *MUC6* contient **1 083 pb** codant un peptide de **361 aa** ainsi qu'une **région 3' non traduite de 632 pb**. La séquence codante est constituée de deux régions distinctes :

- **région 1** ; 270 aa ; riche en Thr/Ser/Pro (62 %) ; aucune Cys ; un site potentiel de *N*-glycosylation

- **région 2** ; 91 aa terminaux ; Thr/Ser/Pro : 22 % ; 11 Cys (12 %) aux positions conservées par rapport aux C-terminaux de *MUC2* ; un site potentiel de *N*-glycosylation.

La séquence peptidique déduite de la région 2 présente 25 % de similarités avec les régions C-terminales des trois autres gènes de mucines localisés en 11p15.5 ainsi qu'avec le vWF, bien qu'elle soit plus courte (Figure 4). Les auteurs argumentent ainsi cette différence de structure de *MUC6* par rapport aux trois autres gènes de la famille : les régions non glycosylées flanquant le domaine répétitif central de *MUC2*, *MUC5AC* et *MUC5B* sont longues. Elles sont donc relativement peu protégées de l'action des protéases. Au contraire, la **région 2 de *MUC6* étant plus petite**, il apparaît qu'une plus petite partie de l'apomucine soit

A



B

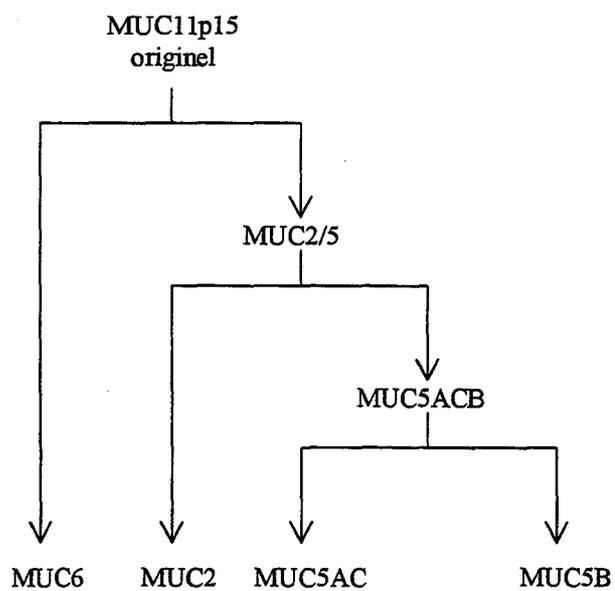


Figure 5 : Comparaison des organisations génomiques des régions 3' de *MUC5B* et *MUC5AC* [Buisine *et al.* 1998a] (A) et schéma hypothétique montrant l'évolution des gènes de mucines localisés en 11p15 à partir d'un ancêtre commun (*MUC11p15* originel) [Desseyne *et al.* 1997a ; 1998a] (B).

potentiellement exposée à un environnement hostile. Or, les tissus exprimant *MUC6* : vésicule biliaire, pancréas, vésicules séminales, appareil reproducteur féminin, seraient exposés de façon chronique à l'agression chimique ou enzymatique de bile ou protéases... Selon les auteurs, *MUC6* devrait donc avoir des caractéristiques uniques lui assurant une résistance supplémentaire.

En ce qui concerne l'organisation génomique de *MUC6*, la **région 1** est codée par **un seul exon** contenant aussi le **domaine répétitif central**. Les régions 1 et 2 sont séparées par un intron. La région 2 est codée par les exons 2 et 3.

II.2.5 - La famille des gènes MUC localisés en 11p15.5 : MUC2, MUC5AC, MUC5B, MUC6

II.2.5.1 Similarités

De part et d'autre de la région répétitive de chacun des 4 gènes de mucines localisés en 11p15.5, on trouve des séquences codant des **domaines N- et C-terminaux** dont la particularité est d'être riches en **résidus Cys**. Ces parties N- et C-terminales peuvent être divisées en plusieurs domaines par **analogie au pro-vWF** (Figure 4). Les résidus Cys ont des **positions bien conservées**.

L'organisation de la région N-terminale est connue pour *MUC2*, *MUC5AC* et *MUC5B*. Cette région comprend **trois domaines D homologues (D1, D2 et D3)** et le **domaine D tronqué (domaine D')** situé entre les domaines D2 et D3. *MUC2*, *MUC5AC*, *MUC5B* et le pro-vWF contiennent un peptide signal N-terminal hydrophobe. Entre les domaines D3 et central (domaine répétitif) des trois apomucines, il existe un domaine typique des MUC localisés en 11p15.5 et ne présentant pas de similarité avec le pro-vWF. Ce domaine a été appelé **MUC11p15** [Desseyn *et al.* 1998b].

Le domaine central constitué de la répétition en tandem de séquences spécifiques à chaque gène et riches en Ser/Thr, contient également un **motif Cys** (10 résidus Cys dont les positions sont conservées) retrouvé 2 fois dans *MUC2*, au moins 6 fois dans *MUC5AC* et 7 fois dans *MUC5B*.

La **région C-terminale** de MUC2, MUC5AC et MUC5B peut être subdivisée en plusieurs domaines :

- un premier domaine appelé "**MUC11p15-type**",
- un petit domaine (nommé **A3uD4**) commun aux 3 mucines et situé dans le vWF entre le domaine A3 et le domaine D4,
- un domaine de type **D4** semblable à celui du vWF,
- un domaine de type **B** unique au lieu des 3 domaines B du vWF,
- un domaine de type **C** unique au lieu des 2 domaines C du vWF,
- un domaine **CK** pour "Cystine knot" ou nœud cystine (11 résidus Cys).

MUC6 est différent des 3 autres gènes localisés en 11p15.5 dans la mesure où l'apomucine MUC6 ne possède pas derrière sa région répétitive de domaines "MUC11p15-type", A3uD4, D4, B ou C. Elle possède **uniquement un domaine CK terminal**.

Ces 4 apomucines sont donc des membres de la mégafamille des protéines à CK. Ce domaine est probablement impliqué dans la dimérisation des protéines possédant ce domaine.

II.2.5.1 Phylogénèse

MUC2, MUC5AC et MUC5B contiennent un motif Cys retrouvé jusqu'à présent uniquement dans ces mucines. Ce sous-domaine définit ainsi une famille de protéines. L'alignement des séquences nucléotidiques codant les sous-domaines Cys et le calcul des similarités entre ces séquences, ont permis à Desseyn *et al.* de bâtir un modèle cohérent de l'évolution des gènes humains MUC2, MUC5AC et MUC5B [Desseyn *et al.* 1998a]. Il est plus que probable que ces gènes soient issus d'un **unique gène ancestral** à la suite de phénomènes successifs de **duplications de régions géniques** (Figure 5B). Les analyses phylogénétiques des séquences des motifs Cys, l'ordonnancement et la séquence de ces motifs de même que la disposition des gènes sur le chromosome 11 montrent que MUC5AC et MUC5B seraient plus proches l'un de l'autre au cours de l'évolution que de MUC2. Ceci amène les auteurs à penser que **MUC5AC et MUC5B ont un gène progéniteur commun baptisé MUC5ACB**. Ce gène provient d'un gène ancestral commun également à MUC2.

MUC6 est, des 4 gènes de mucines localisés sur le chromosome 11, le plus éloigné des trois autres dans la phylogénèse [Desseyn 1997a]. Ce gène a probablement perdu un fragment

par délétion d'une région comprise entre deux introns de même classe. Ainsi, la délétion génomique ne modifie pas le cadre de lecture et le motif peptidique CK reste conservé. Une première duplication a donc probablement eu lieu donnant naissance tout d'abord à l'**ancêtre de MUC6** et à l'**ancêtre commun de MUC2, MUC5AC et MUC5B**. Ce dernier s'est ensuite dupliqué en *MUC2* et *MUC5ACB*. L'ancêtre de *MUC6* a perdu une partie 3' pour donner naissance au gène que nous connaissons actuellement.

II.3 - MUC3 : gène codant une mucine contenant un motif de type EGF

Les premiers clones d'ADNc partiels caractérisant *MUC3* ont été isolés à partir de la banque d'ADNc d'**intestin grêle** utilisée précédemment pour isoler *MUC2* à l'aide d'un anticorps polyclonal dirigé contre des mucines intestinales humaines déglycosylées [Gum *et al.* 1990]. Ces clones (SIB124 et SIB139) contiennent des **répétitions en tandem d'une séquence de 51 pb, soit 17 aa** (riche en Thr/Ser : 71 % et en Pro : 6 %). *MUC3* a été localisé sur le **chromosome 7 en q22** [Fox *et al.* 1992].

Van Klinken *et al.* ont criblé une banque d'ADNc à l'aide d'une sonde représentant la région répétitive [Van Klinken *et al.* 1997]. En amont du domaine répétitif central se trouvent :

- une région comportant un **autre domaine répétitif**, celui-ci de **59 aa** (177 pb), l'ADNc codant ce domaine s'étendant sur au moins 5 kb,
- suivie d'une **région semi-unique de 270 aa de type mucine** (riche en Ser/Thr/Pro) et contenant un seul résidu Cys.

Ces résultats sont **en désaccord** avec ceux de Gum *et al.* [Gum *et al.* 1997b]. Ces derniers ont également décrit, devant le domaine répétitif de 17 aa, **une seconde** structure répétitive de type mucine mais avec une périodicité imparfaite de 375 aa (1125 pb) (Figure 6).

Le **domaine C-terminal** de 617 aa serait composé de deux parties [Gum *et al.* 1997b] :

- un **domaine de type mucine non répétitif de 511 aa**, riche en Thr (27 %), Ser (22 %) et Pro (11 %),
- et un **domaine de 106 aa riche en Cys** (10 %) et présentant des similarités avec des **motifs structuraux de type EGF** ("Epidermal Growth Factor") (similarités dans le nombre et l'espacement des Cys).

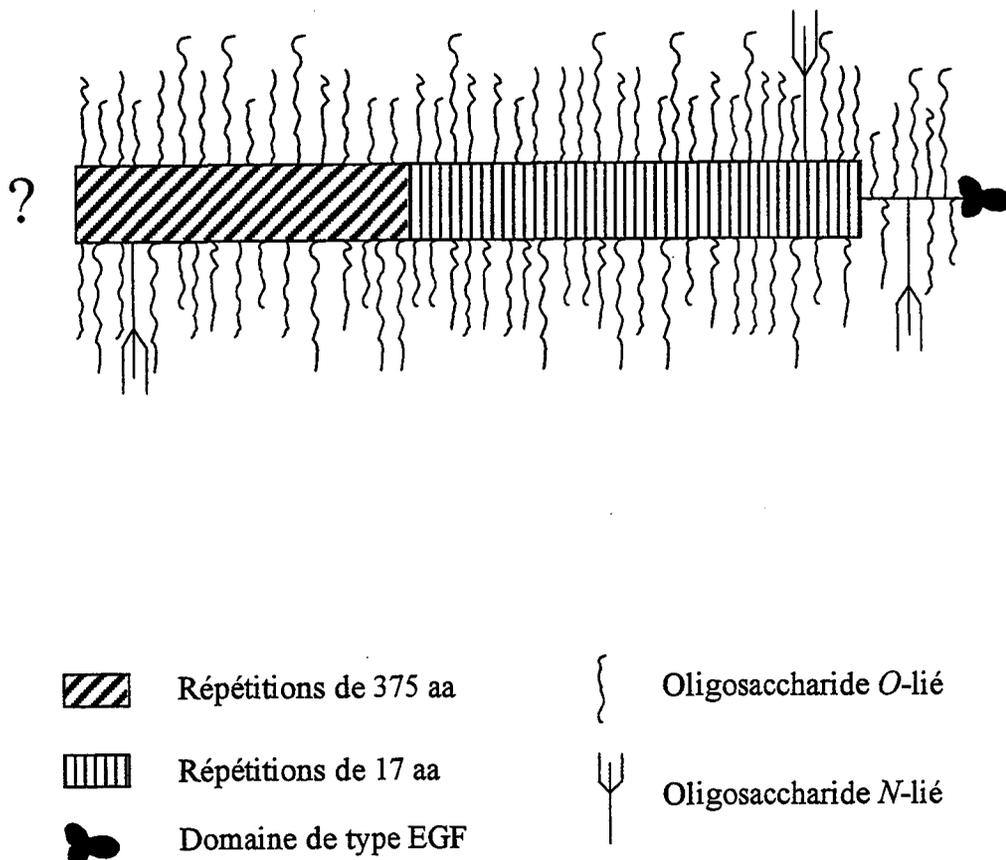


Figure 6 : Représentation schématique de l'organisation peptidique de l'extrémité C-terminale de MUC3 [Gum et al. 1997b].

La région contenant des répétitions en tandem de 17 aa présente un polymorphisme de type VNTR. Les séquences en amont des répétitions en tandem de 375 aa n'ont pas été déterminées jusqu'à présent.

Les motifs de type EGF, retrouvés dans de nombreuses protéines possédant une activité de facteur de croissance, ont un rôle dans la présentation des sites de liaison aux ligands.

II.4 - *MUC7* : gène codant une mucine sécrétée soluble

Dans la **salive** humaine ont été caractérisées deux mucines : **MG1**, de haut poids moléculaire codée par *MUC5B* et **MG2**, de faible poids moléculaire codée par *MUC7* [Tabak *et al.* 1982]. MG2 est une protéine de 125 kDa composée d'environ 30 % d'acides aminés et de 68 % de carbohydrates. MG2 est **trouvée sous forme de monomères**, contrairement à MG1/*MUC5B*, trouvée sous forme de polymères reliés par des ponts disulfures dans les parties N- et C-terminales.

En utilisant un antisérum dirigé contre la mucine MG2 déglycosylée pour cribler une banque d'ADNc de glande sous-maxillaire humaine, Bobek *et al.* ont isolé des clones d'ADNc représentant le gène *MUC7* [Bobek *et al.* 1993]. La même équipe a également décrit l'organisation génomique de *MUC7* [Bobek *et al.* 1996]. *MUC7* a été localisé sur le **chromosome 4 en q13-q21**. L'ADNc complet de *MUC7* comprend **2 368 pb**. La séquence déduite contient un peptide correspondant à la séquence de MG2, confirmant ainsi l'identité de l'ADNc [Reddy *et al.* 1992]. La région traduite code une protéine de 377 aa et contient une **région centrale constituée de six répétitions en tandem de 23 aa riche en Thr/Ser/Pro** (Figure 7). Cette région répétitive est **encadrée par deux séquences uniques de type mucine** (riches en Thr et Ser). La région N-terminale (164 aa) contient une région fortement hydrophobe de 20 aa correspondant au **peptide signal**. Elle contient également deux résidus Cys probablement impliqués dans la formation de ponts disulfures intramoléculaires ainsi que 4 sites potentiels de *N*-glycosylation. La région C-terminale (75 aa) contient un site potentiel de *N*-glycosylation.

MUC7 s'étend sur une distance d'environ **10 kb**. Il comporte **3 exons et 2 introns**. Les deux premiers exons, de respectivement 100 et 68 pb, codent la région 5' non traduite et le peptide signal. Le troisième exon de 2,2 kb comporte le reste de la séquence codante de MG2 ainsi que la partie 3' non traduite s'étendant, sans compter la queue polyA, sur 1,1 kb. Les deux introns ont une taille d'environ 1,7 et 6 kb.

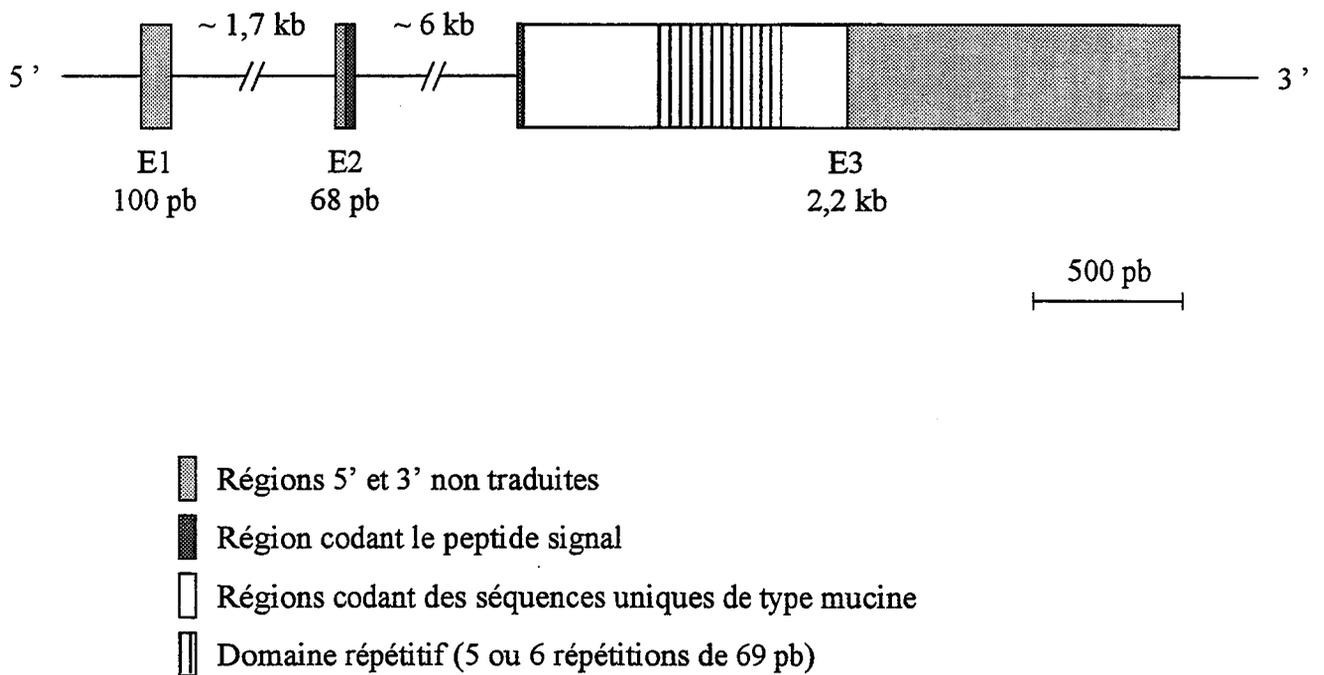


Figure 7 : Représentation schématique de l'organisation génomique de *MUC7* et de l'organisation peptidique déduite [Bobek *et al.* 1996].

II.5 - MUC8

Un anticorps polyclonal dirigé contre des mucines **trachéobronchiques** humaines déglycosylées chimiquement a permis à Shankar *et al.* d'isoler en 1994 un clone d'ADNc appelé pAM1 [Shankar *et al.* 1994]. Ce clone, correspondant au gène *MUC8*, contient **941 pb** codant un peptide de 313 aa. La région située entre les bases 3 et 892 consiste en séquences répétitives de 41 pb codant un peptide composé de **2 types de séquences répétitives consensuelles : une de 13 aa et une de 41 aa**. Le motif répété est riche en Thr, Ser et Pro et comprend un résidu Cys par répétition. *MUC8* a été localisé sur le **chromosome 12 en q24.3** [Shankar *et al.* 1997]. La même équipe a également isolé un clone d'ADNc correspondant à la région C-terminale de *MUC8* [Shankar *et al.* 1997]. Ce clone contient une région non traduite de 458 pb et une queue polyA.

II.6 - MUC9

Le gène de l'**oviductine** humaine code une glycoprotéine présentant les caractéristiques des mucines épithéliales. Lapensée *et al.* proposent donc d'appeler ce gène *MUC9* [Lapensée *et al.* 1997]. *MUC9* a été localisé sur le **chromosome 13**. L'oviductine est riche en Ser, Thr et Pro et est fortement glycosylée [Arias *et al.* 1994]. De plus, elle contient une région constituée de **répétitions en tandem de 15 aa**, le nombre de ces répétitions variant selon les individus.

II.7 - MUC4

Ce gène a été isolé au laboratoire par criblage d'une banque d'ADNc **trachéobronchique** humaine [Porchet *et al.* 1991]. Il a été localisé en **3q29** [Gross *et al.* 1992]. Le clone obtenu est caractérisé par une **répétition en tandem d'une séquence de 48 pb** codant un peptide de 16 aa riche en Ser/Thr. Tel était l'état de nos connaissances à propos de *MUC4* lorsque nous avons abordé notre travail de DEA puis de thèse.

II.8 - Les gènes de mucines animales

On observe une homologie importante entre une mucine humaine et son homologue chez l'animal au niveau des séquences uniques N- et C-terminales (jusqu'à 90 %), alors que l'homologie est faible au niveau des répétitions en tandem d'une espèce à une autre (stricte spécificité d'espèce des sondes oligonucléotidiques correspondant aux domaines répétitifs élémentaires) (Tableau 4).

Gène	Espèce	Longueur de la région répétitive (aa)	Séquence peptidique consensus de la région répétitive	Référence
<i>ASGP-1</i>	rat	117-125	L T T L P Q S Q H T G G M K T T R N P Q T T T T E V T T T L S A S S S D Q V Q V E T T S Q T T L S P D T T T T S H A P R E S S S P P S T S V I L T T T A S T E G T S G D T G H T T A V T T Q G S T P A T T Q I S V T P S S Q K M S T V S T L S T S T Q E	[Wu <i>et al.</i> 1994]
<i>Muc1</i>	souris	20-21	D S T S S P V H S G T S S P A T S A P X	[Spicer <i>et al.</i> 1991]
<i>Muc3</i>	souris	6	T T T A D V ou T T T V V V	[Shekels <i>et al.</i> 1998]
<i>rMuc3</i>	rat	6	T T T P D V	[Gum <i>et al.</i> 1991]
<i>rMuc2 (MLP)</i>	rat	11-12	P S T P S T T S S K S T	[Xu <i>et al.</i> 1992]
<i>Muc5ac (MGM)</i>	souris	16	Q T S S P N T G K T S T I S T T	[Shekels <i>et al.</i> 1995]
<i>rMuc5ac</i>	rat	16	H T S S P P T G G T S P T S T T	[Inatomi <i>et al.</i> 1997]

Tableau 4 : Séquences peptidiques consensus des régions répétitives de quelques mucines animales.

II.8.1 - Les gènes de mucines membranaires

II.8.1.1 - Le gène codant les sialomucines d'ascite de rat : ASGP-1 et ASGP-2

Chez le rat, il existe une mucine membranaire différente de Mucl, appelée "sialomucin complex" ou **SMC** et composée de deux sous-unités (Figure 8) :

- une **sialomucine θ -glycosylée, ASGP-1** (abréviation de l'anglais "Ascite Sialomucin Glycoprotein")
- et une ***N*-glycoprotéine transmembranaire, ASGP-2** [Sherblom et Carraway 1980 ; Hull *et al.* 1990].

La sous-unité ASGP-1 est fortement liée, mais de façon non covalente, à la sous-unité transmembranaire ASGP-2. SMC a été à l'origine isolée sous la forme d'un complexe **hétérodimérique ancré dans la bicouche lipidique** de la surface de **cellules métastatiques** provenant d'un **adénocarcinome mammaire de rat (cellules 13762)** [Sherblom et Carraway 1980]. Mais *SMC* est également exprimé chez le rat adulte normal dans différents épithéliums : intestin grêle, côlon, trachée, utérus, glandes mammaires et surface oculaire [Rossi *et al.* 1996 ; McNeer *et al.* 1997 ; Price-Schiavi *et al.* 1998]. SMC provient d'un gène unique [Sheng *et al.* 1992 ; Wu *et al.* 1994]. Le précurseur peptidique est **clivé par protéolyse**, lors du passage dans le réticulum endoplasmique, en ses 2 sous-unités, ASGP-1 et ASGP-2 [Sheng *et al.* 1990].

L'ARNm de pSMC-1 a une longueur de **9 kb** [Sheng *et al.* 1992]. La séquence en acides aminés d'ASGP-2 (~80 kDa) consiste en :

- un **grand domaine extracellulaire** de 684 acides aminés contenant 24 sites potentiels de *N*-glycosylation. Ce domaine peut être subdivisé en 5 sous-domaines : 2 domaines hydrophiles *N*-glycosylés, **2 séquences de type EGF** (EGF-1 et EGF-2) et un domaine riche en Cys (n'étant pas de type EGF),
- un **domaine transmembranaire** de 25 acides aminés,
- une **queue cytoplasmique** de 20 acides aminés.

Les **répétitions de type EGF contiennent 6 résidus Cys** à des positions constantes (CX₄CX₅CX₃CXCX₈CX) et seraient impliquées dans des interactions protéines/protéines.

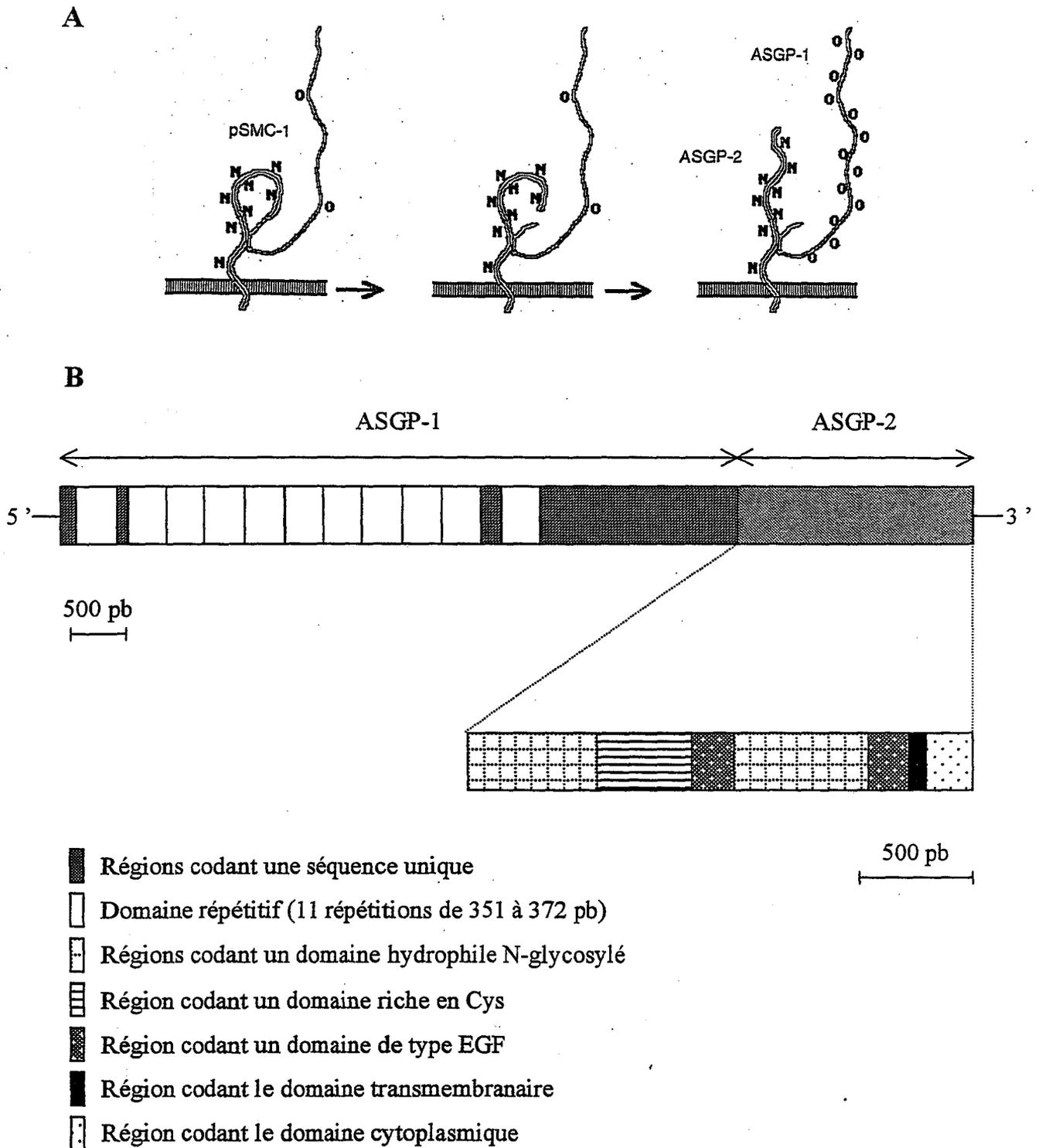


Figure 8 : Représentations schématiques de la biosynthèse et de l'organisation peptidique de SMC [McNeer *et al.* 1997].

A, Le complexe est synthétisé sous la forme d'un précurseur pSMC-1 N-glycosylé mais peu O-glycosylé. Le clivage en ses deux sous-unités, ASGP-1 et ASGP-2, a lieu lors du passage dans le réticulum endoplasmique. La plus grande partie de la O-glycosylation se déroule immédiatement après ce clivage.

B, Schéma d'organisation du précurseur peptidique SMC.

L'équipe de Carraway a montré que ASGP-2 agit en tant que **ligand** du récepteur à activité tyrosine kinase de l'EGF, codé par le proto-oncogène *p185^{neu}* (homologue murin de *c-ErbB-2* ou *HER-2/Neu*), lorsque les deux protéines sont exprimées dans la même cellule [Carraway *et al.* 1997 et 1999]. Le ligand susceptible de se fixer au récepteur ErbB-2 induit non pas la croissance mais la différenciation cellulaire, d'où son nom NDF ("neu differentiation factor").

La séquence peptidique de ASGP-1 contient [Wu *et al.* 1994] :

- une courte région N-terminale unique de 80 acides aminés, contenant le **peptide signal potentiel de 30 acides aminés**,
- un **domaine répété en tandem contenant 11 répétitions de 117 à 124 acides aminés** (9 répétitions le sont en tandem et sont 70-90 % identiques à la répétition consensus de 124 acides aminés. Les 2 répétitions aux extrémités sont séparées des autres par de courtes séquences uniques.),
- une **séquence unique du côté C-terminal de 609 acides aminés** et constituée de 1,6 % de résidus Cys.

La mucine ASGP-1 (~ 220 kDa) contient de nombreux O-glycannes riches en acide sialique. ASGP-1 est impliquée dans les **capacités métastatiques** [Steck et Nicolson 1983] et dans la **résistance à l'action des cellules NK** ("Natural Killer" ou cellules tueuses) [Moriarty *et al.* 1990] des cellules 13762. La présence d'une sous-unité de type mucine (ASGP-1) ainsi que la capacité d'ASGP-2 d'agir en tant que ligand de *p185^{neu}*, indiquent que **pSMC-1 est une molécule hétérodimérique possédant des capacités bifonctionnelles** [Wu *et al.* 1994].

Le complexe ASGP-1/ASGP-2 peut **exister sous deux formes** : une **forme soluble** ne contenant ni la région cytoplasmique C-terminale ni le domaine transmembranaire et une **forme associée à la membrane** contenant ces deux domaines C-terminaux [Rossi *et al.* 1996]. Le complexe ASGP-1/ASGP-2 est présent sous les **deux formes, sécrétée et membranaire**, dans les **glandes mammaires** et le **lait** mais sous forme uniquement **soluble** dans le **côlon** [Rossi *et al.* 1996]. Il existe sous les deux formes dans la **trachée** et les **bronches** [McNeer *et al.* 1998].

Il existe des **similarités entre le complexe ASGP-1/ASGP-2 et MUC1** :

- Tous les deux sont exprimés dans les mêmes tissus épithéliaux sécrétoires tels que les glandes mammaires, la trachée et l'utérus [Gendler et Spicer 1995 ; Rossi *et al.* 1996 ;

McNeer *et al.* 1997]. Cependant, SMC est exprimée dans des tissus où MUC1 n'est pas ou peu détectée, tels que le côlon et l'intestin grêle.

- Les deux protéines forment un complexe hétérodimérique contenant une sous-unité transmembranaire [Sherblom et Carraway 1980 ; Ligtenberg *et al.* 1992b].

- De plus, SMC et MUC1 sont retrouvées sous forme membranaire et sous forme soluble, la forme soluble ne contenant pas les domaines C-terminaux et transmembranaires [Gendler et Spicer 1995 ; Rossi *et al.* 1996]. La forme soluble de MUC1 pourrait être générée par un clivage protéolytique de la forme membranaire [Boshell *et al.* 1992]. La forme soluble de SMC est également produite par clivage protéolytique.

- SMC comme MUC1 inhibe les interactions cellule/cellule et cellule/matrice et assure la résistance des cellules à l'action des cellules NK [Komatsu *et al.* 1997].

- L'ASGP-1, une fois à la surface de la cellule épithéliale sécrétrice, peut être **internalisée** rapidement comme MUC1 dans un compartiment intracellulaire où il y a addition de résidus d'acide sialique [Hull *et al.* 1991].

II.8.1.2 - L'homologue murin *Muc1*

Le gène de souris *Muc1* a été cloné à l'aide d'une sonde d'ADNc correspondant à une partie 3' du gène humain. L'organisation du gène de souris homologue à *MUC1* a été publiée en 1991 [Spicer *et al.* 1991 ; Vos *et al.* 1991]. Il code également une région répétitive constituée de **16 répétitions de 20 à 21 aa**. Comparées au gène humain, ces répétitions sont moins similaires entre elles et elles ne présentent pas de polymorphisme génétique comme c'est le cas pour *MUC1*. Les homologies entre la souris et l'homme sont très faibles en ce qui concerne le domaine de type mucine (34 %) mais d'autres régions sont très conservées (87 %). C'est le cas notamment de la région extracellulaire située juste au-dessus de la membrane, du domaine transmembranaire et du domaine cytoplasmique [Pemberton *et al.* 1996]. La **région extracellulaire fortement conservée** est impliquée dans le clivage de la molécule et dans l'association des parties C- et N-terminales. Ces régions sont également retrouvées chez le gibbon, le boeuf, le lapin, le hamster et le cochon d'Inde [Gendler et Spicer 1995] avec un haut degré d'homologie. Il semblerait donc que ces régions aient une grande importance fonctionnelle. Ainsi, le **domaine cytoplasmique jouerait un rôle important** dans l'adressage de la molécule au pôle apical de la cellule, dans le recyclage de cette molécule et dans l'interaction avec les molécules du cytosquelette.

La **structure génomique de l'homologue murin est similaire à *MUC1* humain**. Les jonctions exon/intron sont conservées. Les introns montrent 50 à 62 % de similarités entre les deux espèces. Les régions promotrices sont similaires à 72 % ce qui n'est pas étonnant puisque les gènes humain et murin ont le même profil d'expression tissulaire et cellulaire [Pemberton *et al.* 1992 ; Spicer *et al.* 1991]. Le gène murin a été localisé sur le **chromosome 3** dans une région synténique au chromosome 1 humain. Récemment, la mucine membranaire Muc1 a été caractérisée chez le lapin [Hewetson et Chilton 1997].

II.8.1.3 - L'homologue murin *Muc3*

Récemment, Shekels *et al.* ont identifié l'homologue murin de la mucine intestinale humaine MUC3 [Shekels *et al.* 1998]. En criblant une banque d'ADNc d'intestin de souris à l'aide d'anticorps dirigés contre des mucines déglycosylées de **caecum** de souris (MCM), ces auteurs ont isolé des clones contenant des **répétitions en tandem de 18 pb** codant deux séquences peptidiques prédominantes : TTTADV et TTTVVV. Le domaine répétitif est suivi par une séquence unique de 1 137 pb et par une séquence 3' non traduite de 521 pb précédant la queue poly(A). Le domaine unique en 3' code **deux régions riches en Cys**. La position des Cys dans ces régions correspond à celle des **domaines de type EGF**. Derrière la seconde région riche en Cys se trouve un **domaine composé de 19 aa hydrophobes**. Ce domaine pourrait servir de **domaine transmembranaire** ou permettrait des **interactions avec des molécules hydrophobes**.

L'ARNm de *Muc3* est d'une taille d'environ **13,5 kb**. Cet ARNm est exprimé dans le côlon et l'intestin grêle. *Muc3* a été localisé sur le **chromosome 5** murin dans une région synténique de la région du chromosome 7 humain où a été localisé le gène *MUC3*. La séquence de l'ADNc de *Muc3* suggère que Muc3 soit une mucine qui ne forme pas de gel de mucus et qui puisse interagir avec la membrane ou avec d'autres molécules via sa région C-terminale.

Muc3 présente des homologies (41 %) avec la région C-terminale unique de MUC3 humain [Gum *et al.* 1997b]. Pourtant, les mucines intestinales humaine et murine différeraient dans leur organisation générale. Immédiatement en aval du domaine répétitif de la mucine de souris se trouve une région riche en Cys, alors que la mucine humaine possède un domaine riche en Ser/Thr entre son domaine répétitif et sa région riche en Cys. La mucine Muc3 de souris a un deuxième domaine riche en Cys séparé du premier par 119 aa. Ce second domaine

serait absent dans la mucine humaine qui se termine (en l'état actuel des connaissances) 30 aa derrière son unique domaine riche en Cys. L'existence de possibilités complexes d'épissage est suspectée à propos de *MUC3*.

II.8.1.4 - Le gène *Muc3* chez le rat

Deux groupes ont isolé des clones d'ADNc à partir de banques d'ADNc d'intestin de rat [Gum *et al.* 1991 ; Khatri *et al.* 1993]. Ces clones contiennent **des répétitions en tandem de 18 pb** codant un peptide répétitif (séquence consensus : TTPDV). Le clone **rMUC176** isolé par Gum *et al.* contient également une séquence unique en aval du domaine répétitif. Cette séquence code un peptide de 92 aa riche en résidus Cys (9 %) et montre des similarités de séquence avec la mucine humaine MUC3 décrite par Gum *et al.* [Gum *et al.* 1997b] ainsi qu'avec l'homologue murin *Muc3* [Shekels *et al.* 1998]. La position des Cys dans cette région est conservée et similaire aux **motifs structuraux de type EGF**. Contrairement à rMUC176, le clone **M2-798** isolé par Khatri *et al.* code 30 répétitions en tandem suivies, non pas d'un domaine de type EGF mais d'une **séquence hydrophobe unique de 82 aa**. Ces auteurs suggèrent que cette région soit un domaine transmembranaire ou un domaine associé à la membrane, et/ou établisse des interactions hydrophobes avec d'autres molécules.

Khatri *et al.* ont également publié une séquence d'ADNc obtenue par une technique de RACE-PCR indiquant que *Muc3* de rat, comme *Muc3* de souris, **contient un deuxième domaine riche en Cys** de type EGF [Khatri *et al.* 1997].

II.8.2 - Les gènes de mucines animales sécrétées formant le gel de mucus

II.8.2.1 – L'homologue murin *Muc2*

Van Klinken *et al.* ont cloné un fragment d'ADNc de 911 pb codant l'extrémité N-terminale de *Muc2* [Van Klinken *et al.* 1999]. La séquence peptidique déduite de 302 aa est similaire aux séquences de MUC2 et de son homologue chez le rat (75 % et 86 % respectivement). Les 17 résidus Cys et le seul site potentiel de *N*-glycosylation sont parfaitement conservés entre les trois espèces.

II.8.2.2 - Le gène *Muc2* chez le rat

Plusieurs groupes ont cloné différentes régions de l'homologue de MUC2 chez le rat. Les régions N- et C-terminales uniques de rMuc2 présentent des similarités avec les domaines équivalents de MUC2 humain.

- Le clone **MLP 2677** ("mucin-like peptide") :

Les mucines intestinales de l'homme et de différentes espèces animales donnent, après réductions des thiols, des composants de haut poids moléculaire ainsi qu'une petite glycoprotéine de **118 kDa** appelée **peptide de liaison** potentiel [Mantle *et al.* 1984 ; Fahim *et al.* 1987 ; Robertson *et al.* 1989]. Le peptide de liaison semble être un peptide libéré par protéolyse d'une mucine et lié par un ou des ponts disulfures à cette même mucine. Xu *et al.* ont isolé un clone d'ADNc d'une banque d'ADNc d'intestin de rat, appelé **MLP**, dont la région 3' code ce peptide de liaison [Xu *et al.* 1992]. Ce clone de **2,6 kb** code un peptide de 837 aa. La séquence peptidique déduite de ce clone montre que le peptide de liaison potentiel est équivalent aux 689 derniers aa du domaine C-terminal. Ce domaine est très similaire aux régions C-terminales des mucines humaines dont le gène a été localisé en 11p15.5. La région 5' du clone MLP code un peptide de 148 aa contenant un domaine de 75 aa riche en Cys et Pro et une région constituée de 4,5 répétitions de 11-12 aa et riche en Ser/Thr/Pro. La présence d'un **domaine de type mucine** suggère que le peptide de liaison représente la région C-terminale d'un peptide de type mucine (MLP pour "mucin-like peptide"). Dans le domaine C-terminal riche en Cys, il existe un point de clivage entre un résidu Asp et un résidu Pro situés à environ 690 résidus de l'extrémité C-terminale, libérant le peptide de liaison *in vitro*.

- Le clone **VR1-A** :

Ce clone d'ADNc code un peptide de **235 aa** composé d'un domaine riche en Cys de 53 aa dans sa région N-terminale suivi d'un domaine de type mucine de 182 aa [Hansson *et al.* 1994]. Le domaine riche en Cys est similaire aux **motifs de 10 Cys** retrouvés dans les mucines humaines. Des études par Northern blot et par Southern blot par électrophorèse en champs pulsés ont démontré que les clones MLP et VR1-A étaient issus du même gène **rMuc2**. Ce gène a été localisé sur le **chromosome 1** de rat [Hansson *et al.* 1994 ; Klings-Levan *et al.* 1996].

Le peptide de liaison trouvé dans l'intestin de rat est donc le produit de la protéolyse de la région C-terminale de rMuc2. Khatri *et al.* ont démontré que les régions N- et C-terminales de rMuc2 sont très **sensibles** à la dégradation protéolytique lors de l'homogénéisation tissulaire, même en présence de plusieurs inhibiteurs de protéases. Cette sensibilité est également retrouvée lors de la purification des mucines [Khatri *et al.* 1998]. Deux autres sites de clivage Asp-Pro existent en aval du premier libérant le peptide de liaison de 118 kDa. Un

fragment de 200 kDa est également libéré par protéolyse. Ce fragment a été identifié par séquençage peptidique comme étant l'extrémité N-terminale de rMuc2.

- La région 5' de *rMuc2* :

La région 5' de *rMuc2* a été obtenue à l'aide d'une sonde 5' de *MUC2* humain [Ohmori *et al.* 1994]. Le peptide déduit du cadre de lecture ouvert de 4 546 pb contient un domaine unique riche en Cys de 1 391 aa et une région répétitive de 122 aa composée de répétitions en tandem irrégulières riche en Thr/Ser/Pro (92 %). Ce peptide est très **similaire à l'extrémité 5' de *MUC2***. Outre les domaines D1-D2-D'-D3 du pro-vWF, l'ADNc code un motif Cys de 10 résidus, motif retrouvé 2 fois dans *MUC2*, au moins 6 fois dans *MUC5AC* et 7 fois dans *MUC5B*.

II.8.2.3 - L'homologue murin *Muc5ac*

Shekels *et al.* ont criblé une banque d'ADNc d'estomac de souris à l'aide d'un anticorps dirigé contre des mucines gastriques déglycosylées de souris (**MGM**) [Shekels *et al.* 1995]. Les clones obtenus contiennent des répétitions en tandem de **48 pb** et une séquence unique de 768 pb. La séquence peptidique déduite du domaine répétitif est composée de 25 % de Ser et 38 % de Thr. Ce domaine est suivi de deux motifs en tandem d'une centaine d'aa riches en Cys et similaires aux **motifs Cys** retrouvés dans *MUC2*, *MUC5AC* et *MUC5B*. Le gène correspondant, *Muc5ac*, a été localisé sur le **chromosome 7** de la souris qui regroupe des gènes synténiques de la région humaine 11p15.5.

II.8.2.4 - Le gène *Muc5ac* chez le rat

Inatomi *et al.* ont publié une séquence partielle d'ADNc de 3,8 kb obtenue à partir du criblage d'une banque gastrique de rat [Inatomi *et al.* 1997]. Ce clone du gène rMuc5ac contient 6 unités répétitives de **16 aa** présentant 69 % de similarités avec l'homologue murin *Muc5ac*. Contrairement au domaine répétitif, l'extrémité C-terminale de 1 142 aa riche en résidus Cys de *MUC5AC* est conservée, non seulement chez les deux espèces de rongeurs, mais également chez l'homme.

III - BIOSYNTHESE DES MUCINES

III.1 - Connaissances actuelles sur la régulation de la transcription des gènes de mucines

III.1.1 - Régulation de la transcription de MUC1

MUC1 a été localisé en **1q21** dans une région riche en gènes [Swallow *et al.* 1987b ; Middleton-Price *et al.* 1988]. A environ 2,6 kb en amont du site d'initiation de la transcription de *MUC1* se trouve la séquence consensuelle de polyadénylation du gène de la **thrombospondine 3** [Vos *et al.* 1992]. En aval, à 1 kb du site de polyadénylation de *MUC1*, un gène non identifié a été localisé [Gendler et Spicer 1995]. Ces deux gènes ont un profil d'expression différent de *MUC1*. Leurs promoteurs respectifs sont donc sûrement indépendants. L'expression spécifique de tissu de *MUC1* peut être reproduite expérimentalement chez une souris transgénique portant deux copies d'un fragment d'ADN contenant le gène *MUC1* ainsi que les séquences flanquantes (**1,6 kb en 5'** et **2,3 kb en 3'**) mais il semble que les séquences 5' soient suffisantes pour maintenir le niveau de transcription *in vivo* [Peat *et al.* 1992].

Différentes régions incluses dans les 743 pb en amont du site de transcription apparaissent être impliquées dans le contrôle de l'expression de *MUC1* :

- Une boîte **TATA** est présente en -24/-19 (par rapport au site d'initiation de la transcription) [Lancaster *et al.* 1990].

- Une séquence appelée **Ebox** ou **E-*MUC1***, localisée à -84/-72 pb semble déterminer la spécificité tissulaire de l'expression de *MUC1* [Kovarik *et al.* 1993].

- **Deux boîtes GC** (site de liaison de la protéine **Sp1**) ont été localisées en -576/-568 et -99/-90 et jouent un rôle dans l'**activation** de la transcription [Kovarik *et al.* 1996]. Le site Sp1 proximal peut également lier un autre facteur dénommé **SpA**. Ce facteur exerce un effet inhibiteur sur la transcription.

- Hollingsworth *et al.* ont identifié dans le promoteur de MUC1 trois régions appelées M-PMR1 (-641/-615), M-PMR2 (-253/-237) et M-PMR3(-133/-102) (**M-PMR** pour Mucin-Purine pyrimidine Mirror Repeat) [Hollingsworth *et al.* 1994a]. Ces régions riches en purines (G+A) sur un brin et en pyrimidines (T+C) sur l'autre brin sont également retrouvées dans le promoteur du gène CFTR (Cystic Fibrosis Transmembrane conductance Regulator). Ces régions sont caractérisées par une symétrie en miroir presque parfaite et ont tendance à former une triple hélice intramoléculaire d'ADN (conformation ADN de type H) qui pourrait être impliquée dans la régulation de la transcription de certains gènes. La même équipe a identifié une protéine nucléaire d'environ **27 kDa** se liant au brin riche en purines de ces régions. La spécificité du site SpA semble influencée par l'élément M-PMR3 qui le jouxte. M-PMR3 serait doté d'une fonction régulatrice négative.

- Une région de régulation localisée à **-505/-485** apparaît se lier à une **protéine de 45 kDa** [Abe et Kufe 1993]. Cette région contient un motif également trouvé dans le promoteur des protéines du lait [Watson *et al.* 1991]. Or *MUC1* est lui aussi surexprimé lors de la lactation.

- Un **facteur soluble d'environ 70 kDa**, présent dans le côlon humain et dénommé mucomoduline, est capable de stimuler la production de MUC1 par des cellules de carcinome colique [Irimura *et al.* 1990]. La région impliquée dans cette régulation est localisée à -531/-520 et a été appelée **RME** pour "Responsive Mucin Element" [Shirotani *et al.* 1994].

III.1.2 - Régulation de la transcription de MUC2

Les premiers éléments de la régulation du gène *MUC2* ont été publiés en 1997 :

La région comprise entre les nucléotides **-848/+1** confère une activité transcriptionnelle maximale dans plusieurs lignées cellulaires intestinales [Velcich *et al.* 1997]. Des éléments localisés plus en amont exercent un effet négatif sur l'expression de gènes rapporteurs.

Gum *et al.* ont par ailleurs montré qu'un court segment compris entre les bases **-91/-73** est important pour l'**activité basale** de transcription [Gum *et al.* 1997a]. Ce segment contient une boîte **CACCC** liant la protéine **Sp1** et d'autres protéines de la famille Sp. Enfin, la région comprise entre les bases **-228/-171** semble conférer la **spécificité cellulaire** d'expression.

Le promoteur de *MUC2* possède **deux éléments de réponse spécifique à *Pseudomonas aeruginosa*** compris entre les nucléotides -2 864/-1308 et -343/-73 [Li *et al.* 1997].

III.1.3 - Régulation de la transcription de MUC5AC

Li *et al.* ont cloné environ 4 kb d'ADN génomique en amont du site d'initiation de la transcription de MUC5AC [Li *et al.* 1998]. Le promoteur de MUC5AC contient une boîte **TATA** (-29/-23), une boîte **CACCC** (-68/-64) ainsi que des **sites de liaison potentiels pour Sp1 et NFκB**. Les 4 kb en amont du site d'initiation de la transcription contiennent également des **éléments de réponse spécifique à *Pseudomonas aeruginosa***.

III.1.4 - Régulation de la transcription de MUC5B

L'étude de la régulation du gène *MUC5B* a été entreprise dans notre laboratoire. Une séquence potentielle de liaison de la protéine Sp1 a été localisée dans l'**intron G** (7^{ème} intron en aval de l'exon central répétitif). Cet intron, dénommé maintenant intron 36, est majoritairement constitué d'un motif parfait de 59 pb répété en tandem. Ces motifs contiennent tous une boîte **GC** [Desseyn 1997a]. Des études de retard sur gel ont montré que cette séquence se lie spécifiquement à un facteur nucléaire extrait de cellules HT-29 MTX. Ce facteur, qui n'est pas la protéine Sp1, est une **protéine de 42 kDa nommée NF1-MUC5B** [Pigny *et al.* 1996b] exerçant une activité inhibitrice sur le promoteur de MUC5B. Deux sites d'initiation de la transcription ont été déterminés [Van Seuningen *et al.* 1999]. La région en amont du site d'initiation proximal est dépourvue de boîtes TATA ou CAAT et contient de nombreux sites potentiels de fixation du facteur nucléaire Sp1. Le second site d'initiation est localisé 2,64 kb en amont du premier. Une boîte TATA potentielle ainsi que deux sites de fixation de Sp1 sont contenus dans un fragment de 222 pb flanquant le site d'initiation. Le promoteur distal est le plus actif.

III.1.5 - Régulation de la transcription de MUC7

Une **TATA box** est située en -24/-19 du site d'initiation de la transcription de MUC7, ainsi qu'une séquence consensus **CAAT** en -83/-79 [Bobek *et al.* 1996]. D'autres séquences régulatrices ont été trouvées : un élément **AP-1**, un élément de réponse aux **glucocorticoïdes** et un à l'**AMPc** (AMP cyclique). Le rôle de ces séquences dans la régulation du gène *MUC7* reste à déterminer.

III.2 - Maturation de l'apomucine

III.2.1 - Les cellules épithéliales

Les cellules épithéliales forment des feuillets cellulaires cohérents appelés épithéliums qui tapissent les surfaces internes et externes du corps. Les cellules épithéliales adjacentes sont liées entre elles par des jonctions donnant au feuillet sa résistance mécanique et le rendant imperméable aux petites molécules. Le feuillet repose sur une lame basale. Les cellules épithéliales sont de plusieurs types et possèdent des morphologies en rapport avec leur spécialisation fonctionnelle. Par exemple :

- Les **cellules absorbantes** possèdent de nombreuses microvillosités qui augmentent la surface d'absorption.
- Les cils des **cellules ciliées** battent de façon synchronisée afin de déplacer le mucus sur le feuillet épithélial.
- La plupart des couches épithéliales possèdent des **cellules sécrétrices** sécrétant des substances à la surface du feuillet cellulaire. Les cellules épithéliales sécrétrices peuvent se regrouper pour former une glande spécialisée dans la sécrétion d'une substance particulière telle que le mucus. Ces **glandes exocrines** sécrètent leur produit dans des canaux.

Les cellules exocrines sécrétant le mucus correspondent aux cellules dites "muqueuses" (cellules caliciformes, cellules à pôle muqueux fermé de l'épithélium gastrique, très nombreuses glandes du tube digestif, de l'arbre trachéo-bronchique et du tractus urogénital). Habituellement, l'abondance des grains de sécrétion de mucus fait qu'en microscopie optique la cellule muqueuse a un aspect "clair", contrairement à l'aspect "sombre" des cellules séreuses.

Ces cellules ont un réticulum endoplasmique rugueux (RER) concentré dans la partie basale. L'appareil de Golgi est bien développé ce qui correspond à une activité importante de glycosylation. Ces cellules sont **polarisées** et caractérisées par la présence de **granules de sécrétion au pôle apical**.

III.2.2 - Glycosylation

L'apomucine est synthétisée par les ribosomes. Lorsque le peptide signal sort du ribosome, il le dirige vers un récepteur protéique sur la membrane du RE. Pendant sa synthèse, le polypeptide est vraisemblablement transloqué à travers la membrane du RER, par

l'intermédiaire d'un pore protéique associé au récepteur. Le peptide signal est excisé par une signal peptidase au cours de la traduction, et la protéine mature est libérée dans la lumière du RE immédiatement après avoir été synthétisée.

C'est dans le RER que l'apomucine est *N*-glycosylée. Puisque l'importation de la majorité des protéines dans le RE est co-translationnelle, les oligosaccharides *N*-liés sont presque toujours ajoutés au cours de la synthèse protéique. La transformation de l'oligosaccharide se poursuit dans l'appareil de Golgi.

Les protéines exportées du RE pénètrent dans le premier des compartiments de transformation du Golgi (le **compartiment *cis***), que l'on pense être en continuité avec le réseau *cis*-golgien ; elles se déplacent ensuite vers le compartiment suivant (le **compartiment *médian***, qui correspond au saccule central de l'empilement), et enfin vers le **compartiment *trans***, où la glycosylation est achevée. On pense que la lumière de ce compartiment *trans* est en continuité avec le réseau *trans*-golgien, où les protéines sont isolées dans différentes vésicules de transport et expédiées vers leur destination finale : la membrane, les lysosomes ou les vésicules de sécrétion. Les voies de modification des oligosaccharides se produisent dans une séquence correspondante organisée dans l'empilement golgien, chaque saccule contenant son propre lot d'enzymes modificatrices. Les protéines sont modifiées par étapes successives, lorsqu'elles passent d'un saccule à l'autre à travers l'empilement. La modification se produit dans une séquence respectant des contraintes aussi bien spatiales que biochimiques: les enzymes catalysant des étapes précoces sont localisées dans des saccules de la face *cis* de l'empilement golgien, alors que les enzymes catalysant des étapes tardives sont localisées dans des saccules de la face *trans*.

La ***O*-glycosylation**, comme l'extension des chaînes d'oligosaccharides *N*-liées, est catalysée par toute une série de **glycosyl transférase**s qui utilisent les nucléotides glucidiques dans la lumière de l'appareil de Golgi pour ajouter un résidu glucidique à la fois à l'apomucine. Une GalNAc est ajoutée en premier, suivie d'un nombre variable de résidus glucidiques supplémentaires, allant de quelques-uns à 10 ou plus.

III.2.3 - Formation d'oligomères

Les mucines sécrétées dont les gènes sont localisés en 11p15 forment des oligomères. De fortes similarités ont été détectées entre les **domaines D** du vWF et les régions N- et

C-terminales de ces mucines. Les domaines D riches en résidus Cys du pro-vWF sont impliqués dans l'oligomérisation de la protéine par des ponts disulfures ainsi que dans l'emballage de cette protéine dans des granules spécifiques de stockage. Le vWF dimérise par les domaines C-terminaux dans le RER puis oligomérisé par les domaines N-terminaux dans le compartiment *trans* du Golgi [Voorberg *et al.* 1990 et 1991].

Des mécanismes d'oligomérisation analogues sont supposés exister pour ces mucines : Van Klinken *et al.* ont étudié l'oligomérisation des précurseurs de mucines gastrointestinales humaines dans la lignée cellulaire LS174T (carcinome de côlon humain) [Van Klinken *et al.* 1998]. Dans cette lignée cellulaire, les précurseurs oligomériques de MUC2, MUC5AC, MUC5B et MUC6 sont probablement sous forme de dimères. Seuls des **homo-dimères** sont formés. Ces auteurs ont démontré que les apomucines **MUC2 dimérisaient dans le RER**. En se basant sur les similarités de séquence avec le vWF, ils postulent que MUC2, MUC5AC, MUC5B et MUC6 dimérisent certainement par leurs domaines C-terminaux. **Cette dimérisation est suivie de la O-glycosylation** de MUC2 [Asker *et al.* 1995]. Récemment, Asker *et al.* ont démontré que MUC5AC dimérisait également dans le RER [Asker *et al.* 1998]

L'inhibition de la *N-glycosylation* par la **tunicamycine** retarde mais n'inhibe pas la dimérisation de MUC2 humaine. De même, Perez-Vilar *et al.* ont démontré que la dimérisation par les domaines C-terminaux de PSM (mucine sous-maxillaire de porc) pouvait avoir lieu en absence de *N-glycannes* [Perez-Vilar *et al.* 1996]. Par contre, il semblerait que la *N-glycosylation* soit nécessaire à la dimérisation et à la sécrétion de Muc2 de rat [Bell *et al.* 1998]. Il est probable que les *N-glycannes* soient impliqués dans la configuration et la stabilité du domaine C-terminal bien qu'il existe des résultats contradictoires à propos des différentes mucines.

La dimérisation du vWF requiert les 151 aa C-terminaux contenant le **motif "Cystine knot"** [Voorberg *et al.* 1991]. Perez-Vilar *et al.* ont démontré que l'information nécessaire à la formation des dimères de PSM est contenue dans les 90 aa C-terminaux, comprenant 11 résidus Cys [Perez-Vilar et Hill 1998]. Ces auteurs suggèrent également que trois de ces résidus Cys soient impliqués dans la formation de ponts disulfures intercaténaux dans les dimères de mucines.

Sheehan *et al.* ont suggéré que la multimérisation de MUC2 dans la lignée cellulaire de côlon humain PC/AA s'effectuait après la glycosylation et la dimérisation [Sheehan *et al.* 1996]. La formation et la sécrétion de MUC2, MUC5AC et MUC5B matures dans les cellules LS174T est **lente**, puisqu'il faut jusqu'à 20 heures pour que les mucines matures soient formées et sécrétées [Van Klinken *et al.* 1996]. La sécrétion des dimères de mucines est précédée d'une **sialylation** [Bell *et al.* 1998].

Les domaines D1, D2, D3 et D' situés dans la région N-terminale du pro-vWF sont nécessaires à la multimérisation du pro-vWF [Wise *et al.* 1988]. Cette étape de multimérisation est indépendante de la dimérisation. De plus, la multimérisation par les domaines D1 et D2 joue un rôle important dans le stockage de la molécule [Wagner *et al.* 1991]. Le pro-vWF a la capacité de **catalyser sa propre oligomérisation**. Les séquences peptidiques impliquées : TCGLCG (domaine D1) et VCGLCGN (domaine D3), sont parfaitement retrouvées dans MUC2, MUC5AC et MUC5B. Les Cys vicinales de ces 2 motifs ont été retrouvées dans le site actif de PDI ("protéine disulfide isomerase"). Il a été suggéré que cette séquence joue un rôle primordial dans la multimérisation du pro-vWF [Mayadas et Wagner 1992]. Il est probable que les mucines qui possèdent ces domaines D multimérisent de la même manière que le pro-vWF.

Il a été montré que le **clivage du pro-vWF** (séquence RSKR↓S entre les domaines D2 et D') n'est pas essentiel à la formation des multimères [Verweij *et al.* 1988]. Les deux premiers domaines D de MUC2 correspondent à la pro-protéine de 741 résidus du pro-vWF (domaines D1 et D2). Ce site de clivage n'a pas été retrouvée dans MUC2, MUC5AC ou MUC5B.

III.2.4 - Transport intracellulaire et sécrétion

Les vésicules de transport destinées à la membrane plasmique bourgeonnent à partir de la face *trans* de l'appareil de Golgi. Les protéines associées à la membrane sont destinées à la membrane plasmique alors que les protéines solubles contenues dans ces vésicules sont sécrétées dans l'espace extracellulaire par un phénomène d'exocytose.

Lors d'une **sécrétion** continue dite **constitutive**, il existe un flux continu de vésicules d'origine golgienne venant fusionner avec la membrane plasmique. Par contre, dans les cellules spécialisées dans la sécrétion, les produits de sécrétion sont stockés dans des

vésicules de sécrétion (ou granules de sécrétion ou vésicules à noyau dense), avant d'être libérés à la demande (**sécrétion régulée**). McCool *et al.* ont démontré que deux types de sécrétion se produisaient dans les **cellules LS180** (cellules de carcinome de côlon) : une **sécrétion basale** (non régulée) et une **sécrétion stimulée** (régulée), les deux sécrétions faisant intervenir des granules dans lesquels les mucines sont **condensées** [McColl *et al.* 1995]. La sécrétion non régulée dépend du transport continu de granules de mucines des vésicules de Golgi à la surface cellulaire et n'utilise pas de mucines stockées. Les microtubules sont nécessaires à la sécrétion basale des mucines. Au contraire, la sécrétion régulée nécessite la libération des mucines des granules de stockage et ne dépend pas des microtubules. Les vésicules sont stockées du côté apical de la cellule.

Dans les vésicules de sécrétion, les protéines sécrétées subissent divers types de modifications. Elles sont tout d'abord **concentrées**, environ 200 fois par rapport à leur concentration dans le Golgi. La condensation se produit brusquement et elle est sans doute provoquée par une acidification de la lumière de la vésicule. La charge négative des mucines joue un rôle important dans ce stockage, les mucines formant des complexes avec les ions Ca^{2+} présents en grande quantité dans les granules. C'est parce que les vésicules matures sont remplies de façon aussi dense que les cellules pourront sécréter rapidement de grandes quantités de mucus par exocytose. D'autres modifications interviennent dans les vésicules de sécrétion : beaucoup de polypeptides sécrétés ont, par exemple, une **pro-pièce N-terminale** qui est clivée par protéolyse peu de temps avant la sécrétion pour donner la protéine mature.

Les cellules épithéliales sont des cellules polarisées, possédant un **domaine apical**, qui fait face à la lumière et a souvent des caractéristiques particulières tels que des cils ou une bordure en brosse de microvillosités, et un **domaine basolatéral**, qui recouvre le reste de la cellule. Les cellules épithéliales co-sécrètent souvent un groupe de produits (enzymes digestives, mucus, dans le cas des cellules intestinales) à leur surface apicale et un autre groupe de produits (laminine et autres composants de la lame basale) à leur surface basolatérale. La signalisation membranaire ajoutée dans l'étape golgienne sur la face *trans* assure cette régulation.

La dernière étape dans la sécrétion régulée est celle de la **libération par exocytose** déclenchée par un signal. Le **signal de sécrétion** est souvent un messenger chimique, tel qu'une hormone, s'associant à son récepteur à la surface de la cellule sécrétrice, déclenchant une

cascade d'événements intracellulaires dont une **augmentation du Ca^{2+} cytosolique**. Ce signal induit la mobilisation de la vésicule sur le lieu d'exocytose, sa fusion à la membrane plasmique, la libération du contenu et le recyclage de la membrane de la vésicule par endocytose.

Dès l'exocytose, les mucines sont instantanément **hydratées**. Leur volume augmente de 400 à 600 fois par rapport au volume condensé dans le granule en quelques millisecondes. Cette augmentation de volume est due à des répulsions électrostatiques entre les charges négatives des mucines qui vont former un gel de mucus hydraté.

III.2.5 - Biosynthèse de MUC1

Contrairement aux mucines sécrétées formant le gel de mucus, **MUC1 n'est pas sécrétée** par des **cellules hautement spécialisées**.

La biosynthèse de MUC1 a été beaucoup étudiée dans les cellules de cancer du sein [Hilkens et Buijs 1988 ; Ligtenberg *et al.* 1992b]. La molécule est synthétisée sous forme de **précurseur N-glycosylé** et est **immédiatement clivée** dans le RE. Le domaine de type mucine est alors séparé du reste de la molécule comprenant le domaine cytoplasmique, le domaine transmembranaire et environ 65 acides aminés du domaine extracellulaire. Le produit de clivage contenant le domaine de type mucine reste néanmoins **associé à la membrane** car il forme un **complexe stable** avec le domaine C-terminal du précurseur de MUC1 ancré dans la membrane. Ce complexe ne met pas en jeu de ponts disulfures.

Il semblerait que **deux motifs**, l'un situé à la jonction du domaine transmembranaire et du domaine cytoplasmique et l'autre dans le domaine extracellulaire non répétitif, soient **impliqués dans la localisation apicale de MUC1** [Pemberton *et al.* 1996]. Parry *et al.* ont démontré que le domaine cytoplasmique de MUC1 se lie aux microtubules du cytosquelette, ce qui permettrait son transport intracellulaire [Parry *et al.* 1990]. MUC1 pourrait également servir de point d'attache de ces filaments dans la membrane apicale.

Les étapes post-traductionnelles de **glycosylation** se déroulent lors du cheminement vers la surface cellulaire mais la protéine n'est pas totalement glycosylée quand elle arrive à la surface. Ce n'est qu'après un **processus d'internalisation et de recyclage** qu'elle arrive à un stade pleinement mature. L'addition des résidus d'**acide sialique** se déroule sûrement dans le

compartiment trans du Golgi [Litvinov et Hilkens 1993]. Ce processus a également été observé pour une "mucin-like" de rat fortement sialylée : ASGP-1 [Hull *et al.* 1991]. Le recyclage de ces glycoprotéines est nécessaire au maintien de leur état totalement sialylé, état conférant au domaine de type mucine une structure fortement étendue.

Plusieurs heures après sa synthèse, **MUC1 est libérée de la cellule** et détectée dans le milieu de cellules en culture [Boshell *et al.* 1992]. Une **forme soluble** existe également dans les liquides biologiques (sérum, lait) [Burchell *et al.* 1984]. Une **forme sécrétée de MUC1**, présumée être produite par épissage alternatif, a aussi été décrite [Williams *et al.* 1990]. Cependant, l'existence de ce produit d'épissage n'a jamais été confirmée. Boshell *et al.* ont démontré la présence d'une forme soluble ne contenant pas de domaine cytoplasmique et suggèrent que MUC1 est libérée de la membrane plasmique sous l'action d'une protéase [Boshell *et al.* 1992]. La présence de MUC1 dans le milieu extracellulaire est donc due, soit à une dissociation du complexe, soit à un clivage protéolytique supplémentaire.

IV - FONCTIONS DES MUCINES

IV.1 - Fonctions des mucines épithéliales formant le gel de mucus

Ce sont les oligomères de mucines qui forment le gel de mucus et sont responsables de ses caractéristiques de viscosité et d'élasticité. La configuration oligomérique est essentielle : la réduction des ponts disulfures existant entre les oligomères ainsi que la dégradation protéolytique en glycopeptides conduisent à la perte de capacité du mucus à former le gel.

Les fonctions du mucus et des mucines sont donc étroitement liées. On attribue aux mucines un certain nombre de fonctions biologiques du mucus [Devine et McKenzie 1992] :

- Très **hydrophiles** du fait de leur haute teneur en résidus glycaniques, elles contribuent à maintenir une **hydratation optimale** du mucus respiratoire et le protègent contre l'évaporation au contact de l'air.

- Plusieurs microorganismes pathogènes utilisent des structures carbohydrates comme récepteurs afin de reconnaître et de pénétrer spécifiquement certains tissus ou cellules mucipares. Les glycanes des mucines sont considérés comme formant une **mosaïque de récepteurs** permettant, par des mécanismes d'adhérence, de **piéger les micro-organismes** qui seront éliminés vers le pharynx grâce à l'**activité mucociliaire**. Ces molécules représentent donc un élément important de la **défense de la muqueuse** et du **maintien de la stérilité** de l'arbre trachéobronchique. Les mucines protègent également l'intestin contre la **colonisation** par des parasites et autres agents pathogènes grâce à certains résidus glycaniques (Gal et GalNAc) qui favorisent le **maintien de la flore naturelle**.

- Les mucines permettent la **lubrification** des muqueuses, assurant ainsi une **protection mécanique** lors de la progression du bol alimentaire et au cours de la digestion.

- Elles facilitent l'accès des nutriments aux entérocytes au niveau intestinal et constituent donc une **barrière sélective d'échange** [Strous et Dekker 1992] ; les nutriments tels que les petits peptides, le sucrose et le lactose doivent d'abord traverser la couche de mucus avant d'être hydrolysés par les hydrolases liées à la membrane intestinale telles que les dipeptidases, la sucrose-isomaltase et la lactase.

- Dans l'estomac, où le pH est extrêmement acide et où il existe une forte concentration de pepsines, les mucines protègent la muqueuse en exerçant un **pouvoir tampon** [Allen 1983]. Les ions H^+ peuvent diffuser à travers la couche de mucus, de la surface épithéliale à la lumière gastrique, mais le retour vers l'épithélium est rendu impossible par la formation d'un gradient de pH (pH 1-2 dans la lumière gastrique à pH 6-7 à la surface de l'épithélium gastrique). L'acidité est neutralisée dans cette couche stable par des bicarbonates sécrétés par les cellules épithéliales [Williams et Turnberg 1980].

IV.2 - Fonctions des mucines épithéliales membranaires

Les mucines membranaires jouent un rôle de lubrification et de protection cellulaire contre les substances toxiques. Mais elles ont également une fonction d'adhérence. En effet, la densité des chaînes O-glycanniques entraîne deux caractéristiques importantes de la structure des mucines :

- les mucines ont une structure étendue, avec une extension moyenne par acide aminé de 2,5 Å. Les mucines membranaires sont donc assez longues pour dépasser le glycocalyx c'est-à-dire l'ensemble des glycoprotéines et des polysaccharides recouvrant les cellules et dont l'épaisseur est d'environ 100 Å. Ainsi, MUC1 s'élève de 200 à 500 nm au-dessus de la membrane cellulaire.

- les nombreux sucres terminaux des chaînes oligosaccharidiques sont ainsi exposés de façon optimale.

La fonction d'adhérence a pour conséquence la protection cellulaire : du fait de leurs charges négatives et de leur structure étendue, les mucines forment une **barrière répulsive** autour de la cellule et jouent donc un rôle de protection ; par contre, si une cellule a des récepteurs spécifiques pour des sucres attachés aux mucines, l'adhérence est alors plus importante que la répulsion. Comme les sucres sont nombreux et bien exposés, de nombreuses interactions potentielles avec des leucocytes, des bactéries ou des virus contenant des récepteurs pour ces résidus glycanniques sont possibles. Ces fonctions d'**adhérence** et d'**anti-adhérence** ont surtout été étudiées pour MUC1 mais elles peuvent être extrapolées aux autres mucines. Ligtenberg *et al.* ont montré que des transfectants exprimant des taux élevés de MUC1 ont une capacité d'agrégation fortement réduite [Ligtenberg *et al.* 1992a]. Cette diminution des propriétés d'adhérence cellulaire en relation avec la surexpression de MUC1 est observée sur de nombreux types cellulaires : cellules épithéliales de sein immortalisées, cellules de mélanomes ou fibroblastes de souris transfectés avec de l'ADNc de MUC1. La

présence de MUC1 ne serait-ce que sur une seule cellule est suffisante pour perturber les propriétés d'**adhérence de type cellule/cellule**.

L'adhérence cellule/substrat vis à vis de différents composants de la matrice extracellulaire est également **diminuée**. Il a été montré que ces **fonctions d'anti-adhérence** étaient surtout dues à l'**encombrement stérique** provoqué par la forte densité en chaînes glycaniques et non à la répulsion des charges négatives portées par les résidus d'acide sialique. En effet, le traitement de MUC1 par la neuraminidase (éliminant les résidus d'acide sialique) n'a pas d'influence sur les fonctions de répulsion [Ligtenberg *et al.* 1992].

Une forte densité de MUC1 peut donc masquer les molécules d'adhérence cellulaire et les empêcher d'agir. Cette hypothèse de répulsion cellulaire a déjà été postulée pour d'autres molécules : par exemple, la forme polysialylée embryonnaire de N-CAM module l'adhérence de certaines cellules malignes et de cellules neurales lors de l'embryogenèse par répulsion des charges électriques [Yang *et al.* 1992]. De même, la podocalyxine, une molécule fortement sialylée présente à la surface luminale des podocytes (cellules épithéliales recouvrant les capillaires des reins), est probablement impliquée dans l'organisation et la maintenance de ces cavités par un mécanisme identique [Charest et Roth 1985].

Cet effet anti-adhésif *in vitro* doit avoir une fonction dans les cellules normales *in vivo*. Dans les cellules épithéliales glandulaires différenciées normales, MUC1 n'est exprimée que du côté apical. Elle ne peut donc interférer avec les molécules d'adhérence situées du côté basolatéral des cellules. Mais elle peut prévenir les interactions entre les glycoprotéines membranaires de cellules opposées et donc l'adhérence des membranes apicales. MUC1 pourrait donc **faciliter la formation et le maintien de la lumière**, présente non seulement dans les structures glandulaires, mais également dans d'autres tissus tels que les veinules endothéliales [Braga *et al.* 1992 ; Hilkens *et al.* 1992].

La présence de MUC1 sur l'endothélium peut prévenir l'adhérence des lymphocytes. Cependant, des lymphocytes spécifiques peuvent encore adhérer aux cellules endothéliales par les **sélectines**. En effet, ces molécules de type lectine peuvent se lier aux carbohydrates de molécules contenant un domaine de type mucine et s'étendant largement au-dessus de la membrane plasmique (ex : récepteurs de la L-sélectine se liant peut-être aux carbohydrates de MUC1 [Lasky *et al.* 1992]).

IV.3 - Fonctions des mucines contenant un motif de type EGF

Les **résidus Cys** trouvés dans les domaines de type **EGF** (6 à 8 résidus) de certaines apomucines pourraient former trois ponts disulfures intramoléculaires créant ainsi un domaine structural important dans le maintien des **interactions protéine-protéine** voire même **protéine-membrane** [Baron *et al.* 1991]. Des domaines de type EGF sont actifs dans plusieurs **facteurs de croissance** ainsi que dans diverses protéines extracellulaires impliquées dans la formation de la **matrice extracellulaire**, l'**adhérence cellulaire**, le **chimiotactisme** et la **réparation tissulaire**. Du fait de l'analogie structurale, on peut proposer pour MUC3 une fonction similaire. Il a été suggéré que les mucines puissent jouer un rôle dans la **réparation des lésions**. Les cellules en gobelet sécrètent, en même temps que les mucines, des **peptides en trèfle** qui sont de petites protéines facilitant la réparation de la muqueuse. Les mucines peuvent, seules ou associées à des peptides en trèfle, promouvoir la réparation de l'épithélium endommagé [Kindon *et al.* 1995].

IV.4 - Rôle des mucines dans les cellules tumorales

La dérégulation de l'expression des mucines dans les carcinomes peut résulter d'une **glycosylation aberrante** ou **incomplète** ou d'une **expression anormale des apomucines** [Devine et McKenzie 1992]. Des expressions aberrantes de mucines et d'antigènes apparentés aux mucines dans des carcinomes atteignant différents organes tels que cancers du côlon et du sein, sont en **relation avec un faible taux de survie** [Nakamori *et al.* 1994 ; McGuckin *et al.* 1995]. Les adénocarcinomes mucisécrétants semblent biologiquement plus agressifs, avec un pronostic plus défavorable, que les tumeurs non mucisécrétantes [Bresalier et Kim 1989]. L'expression de mucines dans les cellules cancéreuses est corrélée à l'**agrégation** des cellules tumorales, à l'**invasion** des cellules et aux **métastases**. Les mucines modulent également les phénomènes de **reconnaissance immunitaire** des cellules cancéreuses [Jentoft 1990 ; Berg *et al.* 1991 ; Fung et Longenecker 1991 ; Handa *et al.* 1991].

Gum suggère des fonctions supplémentaires pour **MUC2** en se basant sur les similarités avec le vWF : **liaison au collagène** ou à d'autres **composants du tissu conjonctif** lorsque l'architecture tissulaire est désorganisée au décours d'un carcinome ou d'une maladie inflammatoire [Gum *et al.* 1994]. Ces liaisons moléculaires pourraient entraîner une augmentation de la dysplasie ou de l'inflammation, des interactions avec des cellules impliquées dans l'inflammation ou la réponse immunitaire voire des métastases.

Le rôle des mucines dans les cellules tumorales a surtout été étudié à propos de la mucine transmembranaire **MUC1**. Des études immunohistochimiques ont été réalisées avec différents anticorps monoclonaux. MUC1 est surtout présente à la surface des cellules épithéliales glandulaires [Zotter *et al.* 1988]. On la retrouve également à la surface luminale des cellules endothéliales des veinules postcapillaires dans les nodules lymphatiques, à la surface de certains types cellulaires recouvrant d'autres cavités telles que le mésothélium, et enfin dans différents types de carcinomes et autres cancers. Dans certains carcinomes tels que ceux du sein, de l'ovaire ou du pancréas, l'expression de MUC1 est augmentée plus de 10 fois par rapport à celle observée dans le tissu épithélial adjacent (ces résultats ont été observés au niveau nucléaire : dot blot ou HIS comme au niveau protéique : immunohistochimie). Les cellules épithéliales de carcinomes n'étant plus polarisées, MUC1 est également ancrée sur la membrane plasmique **basolatérale**. La surexpression de MUC1 entraîne une diminution des contacts cellule/cellule [Ligtenberg *et al.* 1992a] et cellule/matrice [Wesseling *et al.* 1995].

Les cellules exprimant MUC1 peuvent moduler les fonctions immunitaires. En effet, ces cellules sont **résistantes à l'action des lymphocytes T cytotoxiques et des cellules NK** [Van de Wiel-Van Kemenade *et al.* 1993 ; Zhang *et al.* 1997]. MUC1 exprimé dans les carcinomes contient un épitope reconnu par les lymphocytes T cytotoxiques [Barnd *et al.* 1989 ; Ioannides *et al.* 1993 ; Jerome *et al.* 1993]. Il semblerait donc qu'un niveau d'expression élevé de MUC1 au sein d'une tumeur puisse être défavorable aux cellules de cette tumeur, à certains égards mais pas à d'autres [Lalani *et al.* 1991]. Chez des patients atteints du cancer du sein ou du pancréas, MUC1 est détectée dans la circulation sanguine [Burchell *et al.* 1984 ; Metzgar *et al.* 1984 ; Hayes *et al.* 1985 ; Hilkens *et al.* 1986]. La glycoprotéine MUC1 libre produite par les tumeurs inhibe la lyse des cellules cancéreuses par les lymphocytes T cytotoxiques [Barnd *et al.* 1989] et pourrait avoir une **activité immunosuppressive** [Fung et Longenecker 1991]. Un taux élevé de MUC1 circulant pourrait donc bloquer l'activité spécifique des lymphocytes T et ainsi aider les cellules cancéreuses à échapper à la lyse médiée par ces cellules T. La surexpression de MUC1 dans les cancers du sein joue donc un rôle dans la promotion des métastases en inhibant l'adhérence des cellules tumorales et leur reconnaissance par le système immunitaire.

De plus, MUC1 présente à sa surface des motifs carbohydrates, ligands spécifiques des molécules de type **P** et **E sélectine** [Rice et Bevilacqua 1989 ; Aruffo *et al.* 1992 ; Majuri *et*

al. 1992] : sialyl Lewis^X et sialyl Lewis^Y [Baeckstrom *et al.* 1991 ; Hanski *et al.* 1993]. Par ce type d'interactions, MUC1 faciliterait ainsi la dissémination métastatique.

Une molécule d'apomucine est glycosylée différemment suivant qu'elle est exprimée par un tissu normal ou tumoral. On observe ainsi l'expression d'épitopes tumoraux spécifiques [Devine *et al.* 1990]. Cela s'explique par **l'expression aberrante d'apomucines** mais surtout par une **glycosylation différente**. En effet, les glycanes observés dans les cellules de carcinomes sont plus courts que ceux observés dans les cellules normales, augmentant ainsi l'accessibilité de l'apomucine à des ligands ou à des anticorps. En conséquence, la glycosylation différente observée dans les cellules provenant de différentes lignées permettrait d'expliquer la réactivité limitée de certains anticorps anti-MUC1 à l'égard de certains tissus. Comme MUC1 est libérée dans le sérum en grande quantité, il est utilisé comme **marqueur sérique tumoral** [Hilkens *et al.* 1984]. En effet, lorsque la protéine transmembranaire est clivée, la fraction soluble devient antigène circulant et **marqueur tumoral (CA15.3)** détectable par de nombreux anticorps. Ces anticorps permettent le dépistage précoce des récidives et le suivi du traitement des formes métastasées.

Des études comparables réalisées à propos d'autres mucines fourniront de nouveaux outils de diagnostic et/ou de pronostic du cancer ainsi que de nouvelles cibles potentielles d'agents anticancéreux.

V - CE QUE L'ON CONNAISSAIT DU GÈNE *MUC4* AU DEBUT DE CE TRAVAIL

Lorsque le travail d'étude de l'organisation génomique de *MUC4* m'a été confié dans le cadre du DEA puis de la thèse des Sciences de la Vie et de la Santé, le contexte de la littérature était le suivant :

Une vingtaine d'ADNc partiels d'apomucines humaines avaient été caractérisés dont l'expression était activement étudiée. Quant aux gènes eux-mêmes, peu de choses étaient connues si ce n'est l'organisation du gène *MUC1*. Au laboratoire, *MUC5B* était cloné, cartographié, non encore complètement séquencé. Il a fallu encore cinq années pour que l'organisation génomique complète du gène soit élucidée.

Quant à *MUC4*, les informations relatives à ce gène d'apomucine humaine étaient les suivantes :

V.1 - Le clone JER64

Une banque d'ADNc en vecteur λ gt11 construite à partir d'un fragment de muqueuse trachéobronchique humaine avait été criblée avec un antisérum polyclonal dirigé contre des mucines humaines bronchiques déglycosylées chimiquement [Porchet *et al.* 1991]. Sur les 20 clones positifs sélectionnés, un seul clone d'ADNc partiel de *MUC4* d'une taille de **1.83 kb**, et désigné **JER64**, avait été obtenu. Sa séquence montre **39 répétitions de 48 nucléotides** codant des **répétitions de 16 acides aminés** (séquence consensus : Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Asp) (Figure 9). La séquence protéique déduite est constituée d'environ **50 % d'acides aminés hydroxylés** (Ser/Thr : 1/1). Sur les 16 aa, 7 sont parfaitement conservés. Aucun site potentiel de *N*-glycosylation n'a été trouvé dans la séquence. Le clone JER64 caractérise la région répétitive du gène *MUC4*.

V.2 - Localisation chromosomique

Par HIS, le gène *MUC4* avait été localisé sur le **chromosome 3 en q29** [Gross *et al.* 1992].

V.3 - Polymorphisme

La sonde JER64 était capable de détecter un **polymorphisme interindividuel avec BamHI** [Porchet *et al.* 1991]. Deux phénotypes : 1 et 1,2 étaient observés. Dans le phénotype 1,2 les bandes 1 (30 kb) et 2 (26 kb) sont équivalentes en intensité. JER64 détecte également un **polymorphisme de restriction avec toutes les différentes enzymes utilisées : BamHI, HindIII, PstI, EcoRI et TaqI** [Gross *et al.* 1992]. Le polymorphisme observé était supposé de type VNTR c'est à dire en rapport avec la présence de séquences nucléotidiques répétées en tandem dans le gène *MUC4*.

V.4 - Expression

V.4.1 - Expression de l'ARN de MUC4

L'ARNm de *MUC4* a été détecté dans les muqueuses du **tractus respiratoire**, des **glandes mammaires**, de **l'estomac**, de **l'intestin grêle**, du **côlon** et du **col utérin** [Audié *et al.* 1993 ; Lesuffleur *et al.* 1994 ; Audié *et al.* 1995 ; Gendler et Spicer 1995]. Des ARNm de *MUC4* sont également présents dans l'épithélium stratifié de la **conjonctive** [Inatomi *et al.* 1996]. L'expression de *MUC4* apparaissait assez ubiquitaire, intéressant des cellules mucipares mais également des cellules des épithéliums non spécialisées dans les fonctions de sécrétion : entérocytes, cellules ciliées bronchiques.

Au niveau de l'épithélium de l'endocol, *MUC4* est exprimé à un niveau significativement plus élevé pendant la phase lutéale que pendant la phase folliculaire [Audié *et al.* 1995]. L'étude en Northern blot démontre que l'expression de *MUC4* n'est pas reliée à l'imprégnation en progestérone du col utérin [Gollub *et al.* 1993]. Récemment, Gipson *et al.* n'ont détecté aucune variation dans le niveau d'expression de *MUC4* lors du cycle menstruel [Gipson *et al.* 1997a].

MUC4 n'est **pas exprimé dans les cellules canalaire du pancréas normal** [Balagué *et al.* 1994 et 1995] **ni dans celles de la vésicule biliaire normale** [Vandehaute *et al.* 1997]. Cependant, des transcrits de *MUC4* ont été **détectés dans des adénocarcinomes pancréatiques** [Balagué *et al.* 1994 et 1995]. Au laboratoire, il a été mis en évidence une **expression aberrante du gène *MUC4* dans l'adénocarcinome de la vésicule biliaire** [Vandehaute *et al.* 1997]. *MUC4* a également été détecté dans la métastase hépatique d'un

adénocarcinome pancréatique [Vandehaute *et al.* 1997]. Ceci confirme des observations préalables relatant une expression élevée de *MUC4* dans un grand nombre de lignées tumorales pancréatiques [Hollingsworth *et al.* 1994b]. Toutes ces données suggèrent que l'expression du gène *MUC4* puisse constituer un **marqueur de tumorigenèse** au cours de la transformation des cellules épithéliales biliaires et pancréatiques. Cette étude confirme les premiers résultats obtenus sur des cellules de vésicule biliaire en culture [Campion *et al.* 1995].

Les ARNm de *MUC4* sont présents dans les **cancers du côlon** avec une expression **plus importante ou comparable** à celle de la muqueuse colique normale [Ogata *et al.* 1992]. Les **cancers gastriques** présentent une **augmentation** du niveau d'expression de *MUC4* par rapport à celui observé dans l'estomac normal [Ho *et al.* 1995]. Les modifications d'expression semblent d'autant plus importantes que le stade du cancer est avancé. L'expression des transcrits de *MUC4* est également **augmentée** dans une large variété de **cancers du poumon** (sauf dans les cancers à petites cellules), en particulier dans les cancers bien différenciés [Nguyen *et al.* 1996]. **MUC4 semble donc être un marqueur tumoral intéressant.**

V.4.2 - Expression de la protéine MUC4

Très récemment, Gipson *et al.* ont produit un **anticorps polyclonal**, désigné **M4P**, dirigé contre la séquence peptidique déduite du domaine répétitif [Gipson *et al.* 1997a]. La spécificité de cet anticorps à l'égard de l'apomucine MUC4 et l'absence de réactions croisées vis à vis d'autres apomucines n'ont pas été vérifiées. Un signal positif de M4P a été observé dans les couches basale et suprabasale du tissu ectocervical ainsi que d'une façon diffuse dans les glandes endocervicales.

Xing *et al.* [Xing *et al.* 1997] ont synthétisé un peptide de 22 acides aminés (M4.22) à partir de la séquence de JER64. Ils ont obtenu **deux anticorps monoclonaux** : **M4.171** et **M4.275**. Ces anticorps leur ont permis d'étudier l'expression de MUC4 dans des tissus normaux ou cancéreux. L'apomucine MUC4 native serait fortement exprimée dans le côlon : les deux anticorps marquent les cellules en gobelet, les cellules glandulaires et les sécrétions contenues dans la lumière. Le marquage des cellules est surtout cytoplasmique. MUC4 serait également présente dans **l'oesophage** (glandes sous-muqueuses), **l'estomac** (cellules glandulaires), **l'intestin grêle** (enterocytes et cellules en gobelet), les **glandes salivaires**

(acinis muqueux et séreux), la **prostate** (cellules épithéliales des glandes). M4.171 marque faiblement les **neurones de la substance grise** et M4.275 marque faiblement les **hépatocytes**. Contrairement à ce qu'indique l'HIS, aucun des anticorps ne marque le poumon témoin. Par contre, les deux anticorps **marquent fortement les pièces histopathologiques de cancers du poumon** (~ 70 %) où le marquage est cytoplasmique mais aussi présent à la surface apicale. L'apomucine MUC4 semble surtout exprimée dans les groupes histologiques de cancers du poumon à petites cellules puisqu'elle est présente dans 73 % des adénocarcinomes, dans 86 % des carcinomes **épidermoïdes** et pas du tout dans les carcinomes non à petites cellules. MUC4 est exprimée **très faiblement dans les cancers du côlon** (~ 30 %) ainsi que dans les cancers de l'ovaire, de l'estomac et du sein.

Pourquoi MUC4 n'est-elle pas retrouvée dans le poumon alors que son ADNc provient d'une banque d'ADNc d'origine trachéobronchique ? On peut supposer que les épitopes reconnus par ces deux anticorps sont masqués dans le poumon normal et qu'ils sont exposés dans le cas des cancers. Le même cas a été décrit pour MUC1 où l'apparition d'un épitope est due à une glycosylation défectueuse [Burchell *et al.* 1987].

La mucine MUC4 semble donc d'un grand intérêt en tant que marqueur diagnostique dans les carcinomes notamment ceux du poumon (expression augmentée).

VI - OBJECTIFS DE LA THESE

Lorsque nous avons abordé ce travail de thèse, les relations de proximité et les parentés structurales des 4 gènes MUC localisés en 11p15.5 n'étaient pas connues.

Les études d'expression réalisées par Northern blot montraient une **hétérogénéité de signal** sous la forme d'un **continuum** détectable dès les zones de migration correspondant à des masses moléculaires très élevées de l'ordre de **20 kb**. La signification de ce continuum ou "smear" n'était pas connue mais les gènes capables de les produire, c'est à dire les 4 gènes localisés en 11p15.5 : *MUC2*, *MUC5AC*, *MUC5B*, *MUC6* ainsi que *MUC3* et *MUC4*, étaient considérés comme devant être de **grande taille** et coder de volumineuses mucines sécrétées responsables des propriétés physicochimiques typiques du mucus. Par ailleurs, les études d'expression cellulaire réalisées par HIS apportaient des données complémentaires mais à certains égards contradictoires par rapport à ce concept.

Les cellules exprimant les gènes *MUC2* et *MUC5AC* correspondaient à des **cellules caliciformes**, celles exprimant *MUC5B* et *MUC6* étaient le plus souvent des cellules de **glandes muqueuses** c'est à dire des cellules spécialisées dans la biosynthèse, le stockage, et la sécrétion régulée de mucines.

Cependant l'expression de *MUC3* et *MUC4* intéressant de nombreux **types cellulaires épithéliaux non spécialisés** dans ces fonctions laissait entrevoir des mécanismes de biosynthèse différents même s'ils concernaient aussi de très volumineuses apomucines.

L'objectif de notre étude était de comprendre la fonction de l'apomucine MUC4 dans le contexte physiopathologique précédemment décrit, c'est à dire celui d'un marqueur de tumorigenèse potentiel dont le rôle spécifique et/ou concerté par rapport aux autres apomucines devait être élucidé.

Le but de ce travail était de :

- poursuivre le travail de marche sur le chromosome et sur l'ADNc afin de déterminer l'organisation génomique de la partie 5' du gène de mucine humaine *MUC4*
- de cloner la région génomique en amont du gène *MUC4*, région contenant potentiellement des éléments de régulation transcriptionnelle de ce gène

La stratégie d'approche de la problématique nous imposait de **déterminer l'organisation et la séquence de l'ADNc de *MUC4*** de façon à pouvoir :

- y rechercher des domaines fonctionnels,
- obtenir des anticorps dirigés contre des régions uniques, spécifiques et peu glycosylées de l'apomucine *MUC4* de façon à appréhender leurs fonctions,
- aborder le fonctionnement du gène : nombre et structure des transcrits.

Les outils accumulés : sondes, éléments de cartographie et anticorps devaient permettre ensuite la mise en place de l'étude des éléments régulateurs du gène *MUC4*.

VII - STRATEGIE

La reconstitution de la séquence d'ADNc complète du gène MUC4, l'étude de la cartographie et de l'organisation génomique de ce gène, l'étude du nombre et de la structure des différents transcrits nous ont amenés à **cribler un certain nombre de banques** :

des banques génomiques

La marche sur le chromosome est une technique classique qui permet de relier physiquement des séquences clonées pour obtenir des séquences continues sur plusieurs dizaines ou centaines de kilobases recouvrant ainsi la totalité du gène d'intérêt. Cette approche permet de se déplacer en amont ou en aval d'une séquence clonée point de départ, notre point de départ étant la séquence répétitive reconnue par la sonde JER64.

Nous avons utilisé 2 banques génomiques chevauchantes construites à partir d'ADN humain partiellement digéré par les enzymes de restriction *Sau3A/MboI* et cloné :

- soit en vecteur bactériophagique EMBL4
- soit en vecteur cosmique pWE15.

Ainsi nous avons pu sélectionner et étudier

- le phage recombinant ANT55
- les cosmides recombinants LEA2 et LEA47.

des banques d'expression

La banque d'ADNc de muqueuse trachéobronchique humaine (don d'organe) construite au laboratoire en vecteur λ gt11 étant épuisée, nous avons été amenés à construire deux nouvelles banques d'expression, à partir de tissus mucipares d'obtention aisée en milieu hospitalier, et exprimant intensément MUC4. Ont été utilisés un fragment de muqueuse colique et un fragment de muqueuse iléale.

Notre travail a consisté ensuite à sélectionner des clones génomiques ou d'ADNc, à les caractériser par cartographie et séquençage, à reconstituer la séquence codante, à l'étendre par les techniques de RACE-PCR.

Notre objectif était :

- 1) d'étudier le domaine répétitif de *MUC4* :
 - * homogénéité des unités répétées en tandem
 - * recherche de la présence de séquences uniques
 - * polymorphisme VNTR ou de mutation
- 2) de reconstituer les séquences codantes de la région 5' de *MUC4*, alors que notre collègue Nicolas Moniaux, étudiant en thèse, reconstituait la région 3' en aval du domaine répétitif
- 3) identifier et confirmer le site d'initiation de la traduction
- 4) circonscrire et cartographier la région promotrice de *MUC4*.

Cette étude s'est révélée particulièrement longue et délicate dans la mesure où notre travail a été compliqué par l'instabilité notoire et la recombinaison des clones contenant des séquences répétitives, or nous avons été amenés avec N. Moniaux à identifier successivement 4 types de séquences répétitives dans le gène *MUC4*.

RESULTATS

I - CARACTERISATION DES CLONES GENOMIQUES

I.1 - Caractérisation des clones bactériophagiques contenant le gène *MUC4*

Une soixantaine de clones génomiques provenant de la banque construite au laboratoire en vecteur bactériophagique EMBL4 (nomenclature : ANT) ont été tour à tour sélectionnés et amplifiés. Seuls les clones **ANT55 et ANT56** se sont avérés suffisamment stables pour permettre de caractériser une nouvelle séquence répétitive du gène *MUC4*.

I.1.1 - ANT55

L'insert d'ADN génomique contenu entre les deux sites *EcoRI* de ce recombinant a une taille d'environ 15,5 kb. Sa taille n'est pas réductible à l'aide des enzymes de restriction habituellement utilisées pour la cartographie de clones : *BglIII*, *KpnI*, *HindIII*, *XbaI*, *XhoI*. Néanmoins, ***EcoRI* coupe l'insert en 2 fragments, l'un de 13 kb et l'autre de 2,5 kb** (Figure 10). Le grand fragment *EcoRI* de 13 kb est intensément reconnu par la sonde JER64 et contient donc le domaine répétitif.

Le **petit fragment de 2,5 kb** dénommé **HOR1** n'est pas réactif avec la sonde JER64 et ne contient donc pas de séquences répétées typiques de 48 pb. Par contre, il contient un nouveau type de séquence répétitive de 15 pb sur un fragment de 1,7 kb.

I.1.2 - ANT56

Le bactériophage ANT56 a été obtenu par criblage avec la sonde HOR1. La taille de l'insert est d'environ 14 kb. Ce clone recombinant ANT56 s'est avéré construit de manière similaire à celle de ANT55 : *EcoRI* coupe l'insert en deux fragments de 13 kb et 1,2 kb (Figure 11). Le fragment de 13 kb est identique à celui de ANT55. Le fragment de 1,2 kb correspond à HOR1 tronqué de 1,3 kb en 3'.

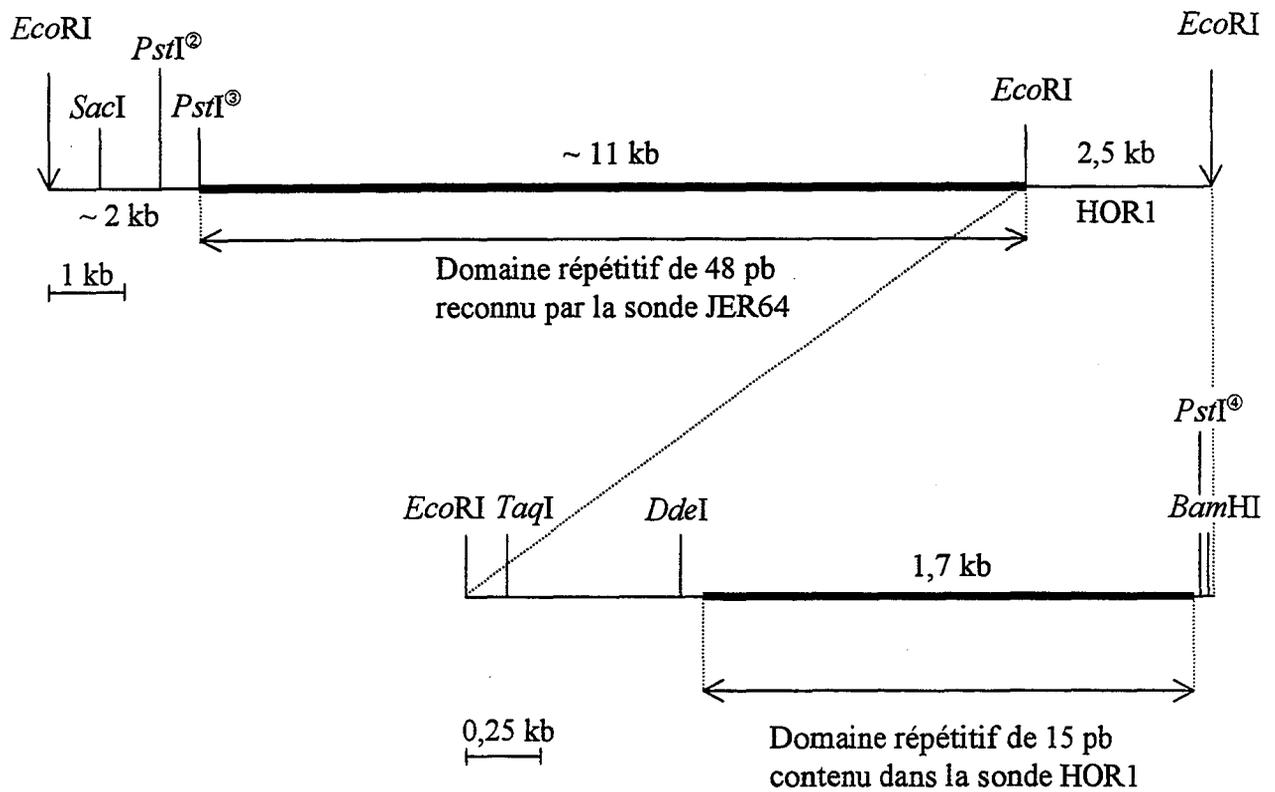


Figure 10 : Cartographie partielle de l'insert du clone bactériophagique ANT55.

Les flèches verticales représentent le site de clonage du vecteur. Les très nombreux sites *DdeI* présents dans le domaine répétitif reconnu par la sonde JER64 n'ont pas été reportés. Seul le site *DdeI* bordant en amont le domaine composé de répétitions de 15 pb est noté.

Remarque : La numérotation des sites *PstI* a été réalisée sur la structure déduite de l'analyse de l'ensemble des clones. Le site *PstI*⁰ est décrit plus loin.

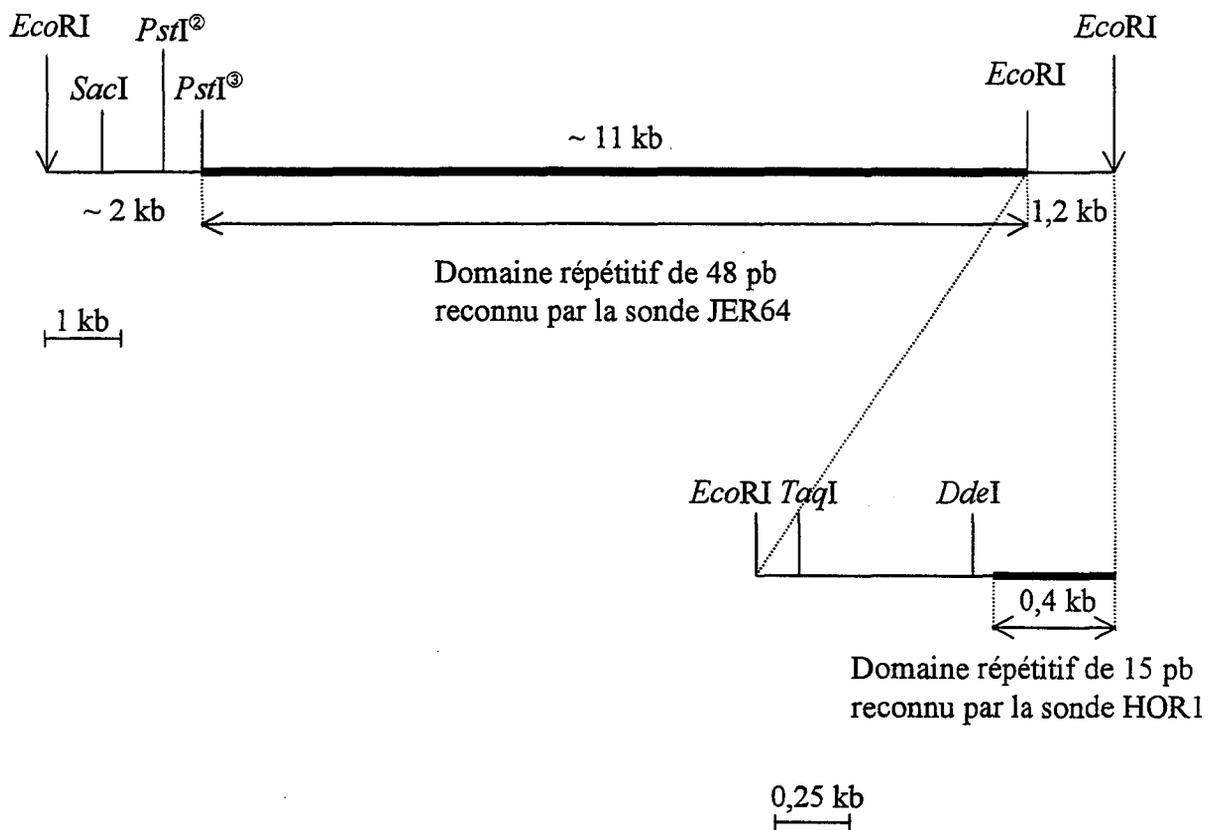


Figure 11 : Cartographie partielle de l'insert du clone bactériophagique ANT56.

Les flèches verticales représentent le site de clonage du vecteur. Les très nombreux sites *DdeI* présents dans le domaine répétitif reconnu par la sonde JER64 n'ont pas été reportés. Seul le site *DdeI* bordant en amont le domaine composé de répétitions de 15 pb est noté.

I.2 - Caractérisation des clones cosmidiques

Un premier criblage de la banque cosmidique à l'aide de la sonde JER64 a permis d'isoler 11 clones positifs dont seulement quatre sont restés stables après le troisième tour d'isolement : Cos 674, 680, 683 et 686. Après analyse en Southern blot de l'ADN cosmidique, il est apparu que Cos674 et Cos686 donnaient un signal d'hybridation très faible par rapport à celui obtenu pour Cos680 et Cos683. Nous avons donc choisi d'étudier ces deux derniers. Nous avons alors débuté l'étude de leur carte de restriction mais, lorsque nous avons voulu relancer les souches en milieu liquide afin d'augmenter notre quantité de matériel, nous n'avons pu que constater la mort de ces souches. Nous avons donc entrepris un nouveau criblage de la banque de cosmides.

49 clones cosmidiques (LEA1 à LEA49) ont été sélectionnés successivement, présentant une réactivité plus ou moins forte vis à vis de la sonde JER64. Ces 49 clones avaient tous la même structure : un grand domaine central reconnu par la sonde JER64 ainsi qu'un deuxième domaine reconnu par la sonde HOR1. Ces clones peuvent être classés en deux familles : l'une dont le fragment central donne un signal positif très important après hybridation avec la sonde JER64 et l'autre dont le fragment central s'hybride avec cette sonde mais beaucoup plus faiblement. L'étude cartographique précise a été réalisée pour quatre d'entre eux : **LEA1**, **LEA2**, **LEA12** et **LEA47**.

I.2.1 – LEA1 et LEA2

La taille estimée de l'ADN inséré dans LEA2 est de 45 kb. La carte de restriction partielle de l'insert est présentée sur la figure 12. La caractéristique principale de cet insert est l'existence d'un très **grand fragment d'environ 19 kb insensible aux enzymes de restriction usuelles** et s'hybridant intensément avec la sonde JER64. La carte de restriction partielle de l'insert cosmidique LEA1 est très ressemblante à celle de LEA2.

I.2.2 - LEA12

L'insert de LEA12 a une taille d'environ 40 kb. Son profil de restriction est identique à celui de l'insert du clone Cos686. Comme pour ce dernier, le signal d'hybridation avec la sonde JER64 est faiblement positif. JER64 reconnaît un grand fragment démuné de sites de

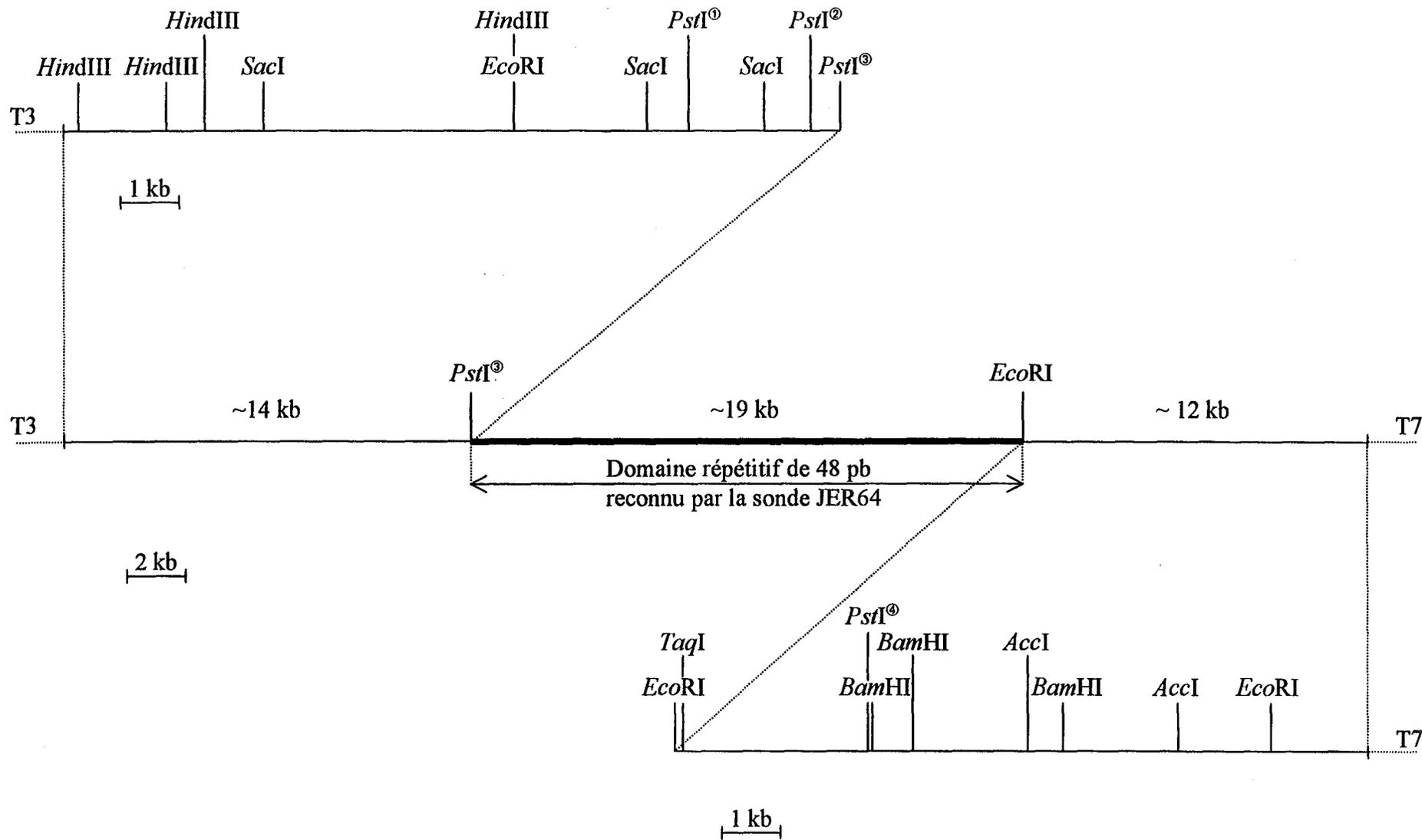


Figure 12 : Cartographie partielle de l'insert du clone cosmidique LEA2.

restriction tout comme dans les clones LEA1 et LEA2 mais avec une intensité beaucoup plus faible (quelques jours d'autoradiographie au lieu de quelques heures). Par contre, les régions encadrant ce domaine sont totalement différentes de celles présentes dans les deux cosmides précédents (Figure 13). Il semblerait que ces clones ne contiennent pas le gène *MUC4* mais une séquence "apparentée" s'hybridant faiblement avec la sonde JER64. Nous n'avons donc pas été plus loin dans l'étude de cette famille de clones.

I.2.3 - LEA47

L'ADN inséré dans le clone recombinant LEA47 représente environ 45 kb. Les clones LEA2 et LEA47 sont **chevauchants** (Figure 14). En effet, LEA47 montre une structure décalée vers le 5' d'environ 5 kb par rapport à LEA2.

A ce stade, il apparaît que la séquence codante répétitive reconnue par la sonde JER64 est **dépourvue des sites usuels de restriction**, et que donc elle a de fortes chances d'être homogène et régulière. De plus, sa taille apparaît, dans les trois clones étudiés, **extrêmement vaste, de l'ordre de 10 à 20 kb**. Nous nous sommes interrogés sur les raisons de l'instabilité de la plupart des clones génomiques que nous avons sélectionnés mais aussi sur les raisons de la "stabilisation" des quelques clones qu'il a été possible d'amplifier successivement et d'analyser. Dans la crainte où les clones stabilisés posséderaient des réarrangements de séquences, tout au long de notre travail, nous avons été amenés à devoir confirmer les résultats obtenus grâce à des études complémentaires réalisées par Southern blot sur l'ADN génomique d'un grand nombre d'individus, de façon à éviter également les éventuelles difficultés d'interprétation liées au polymorphisme.

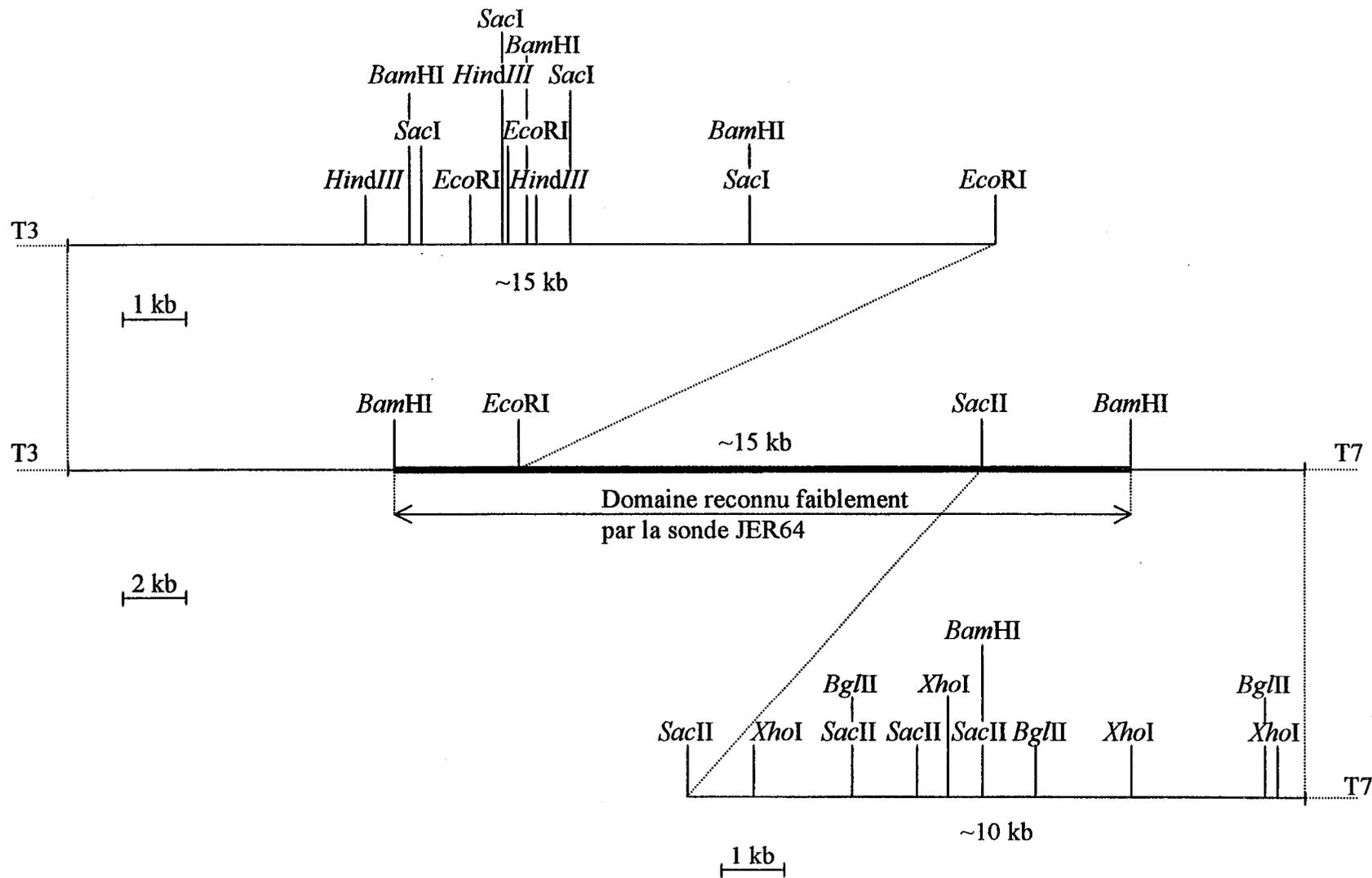


Figure 13 : Cartographie partielle de l'insert du clone cosmidique LEA12.

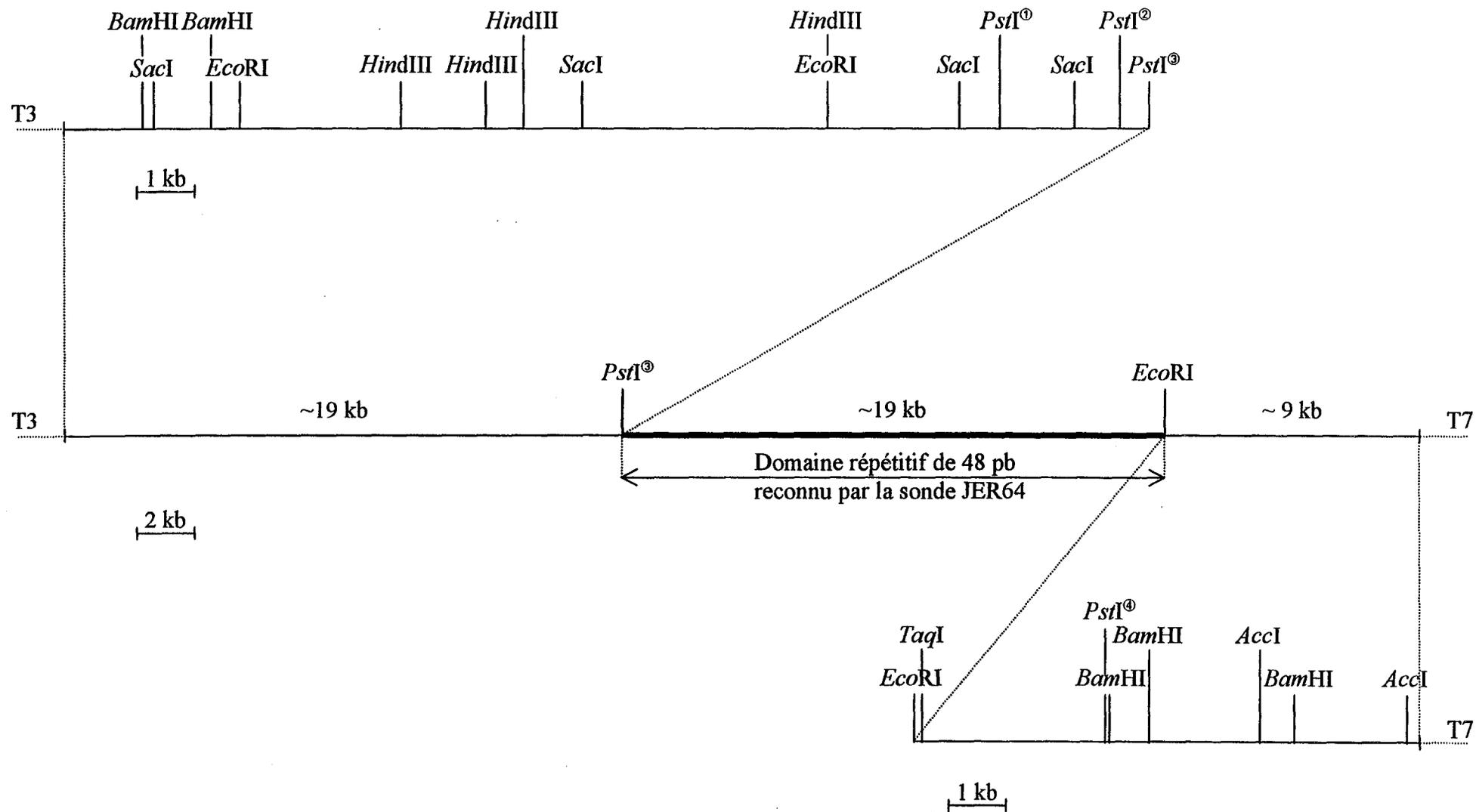


Figure 14 : Cartographie partielle de l'insert du clone cosmique LEA47.

II - LE DOMAINE REPETITIF

II.1 - Le domaine répétitif est homogène et de grande taille

La double digestion enzymatique par *PstI/EcoRI* permet d'isoler un fragment minimum contenant tout le domaine répétitif de 48 pb. En ce qui concerne l'insert de ANT55, ce fragment a une longueur de 11 kb. Il est d'environ 20 kb pour les inserts de LEA2 et de LEA47.

Un petit nombre d'enzymes moins usuelles ont un site de restriction à l'intérieur de la séquence répétitive de 48 pb. L'hydrolyse par l'endonucléase de restriction *RsaI* de LEA1 et LEA2 permet d'obtenir quatre fragments distincts (environ 8, 5, 3,2 et 2 kb) s'hybridant à la sonde JER64 (Figure 15). De plus, la triple digestion enzymatique *PstI/EcoRI/RsaI* donne les mêmes bandes. La somme des tailles des sous-fragments obtenus avec *RsaI* sur le fragment *PstI/EcoRI* de l'insert de LEA1 ou de LEA2 nous permet donc de calculer plus précisément la taille de ce fragment qui est d'environ 18,2 kb.

Cette région ne contenant aucun site pour les endonucléases de restriction conventionnelles, elle n'a pu être clonée en vecteur plasmidique. La détermination directe de la séquence de cette région répétitive s'est donc avérée délicate. Quelques sites de restriction, non présents dans les sites de clonage des vecteurs plasmidiques, existent dans cette région (*DdeI*, *FokI*, *HphI*, *Ksp632I*, *MaeIII*, *MboII*, *MnII*, *SfaNI*). De plus, ces enzymes coupent le domaine répétitif en une multitude de petits fragments.

Cependant, comme un site de restriction pour l'enzyme *DdeI* (CTNAG) existe dans la plupart des répétitions de 48 pb du clone JER64 (dans 36 unités répétitives sur 39), *DdeI* a été utilisée pour digérer la région répétitive. La figure 16, colonne C, montre le fragment *PstI/EcoRI* de 18,2 kb de LEA2 totalement hydrolysé par l'enzyme de restriction *DdeI*. Seuls deux fragments sont observés après électrophorèse en gel d'agarose contenant du BEt. Les tailles de ces fragments, 48 et 96 pb, étaient celles attendues d'après l'analyse de la séquence nucléotidique de JER64. Ces deux bandes s'hybrident à la sonde JER64. La présence d'un fragment de 96 pb est due à l'absence de site *DdeI* entre deux unités répétitives adjacentes. Le fait qu'aucune autre bande supplémentaire ne soit observée suggère fortement que **la région répétitive de l'insert de LEA2 n'est pas interrompue par une séquence unique.**

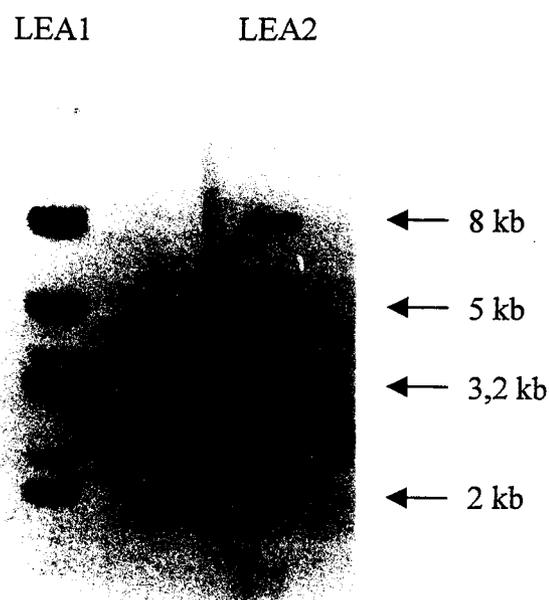


Figure 15 : Détermination de la taille du domaine contenant les répétitions en tandem de 48 pb des cosmides LEA1 et LEA2.

Les inserts cosmidiqes LEA1 et LEA2 ont été hydrolysés par l'enzyme *RsaI*. Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été faite à l'aide de la sonde JER64.

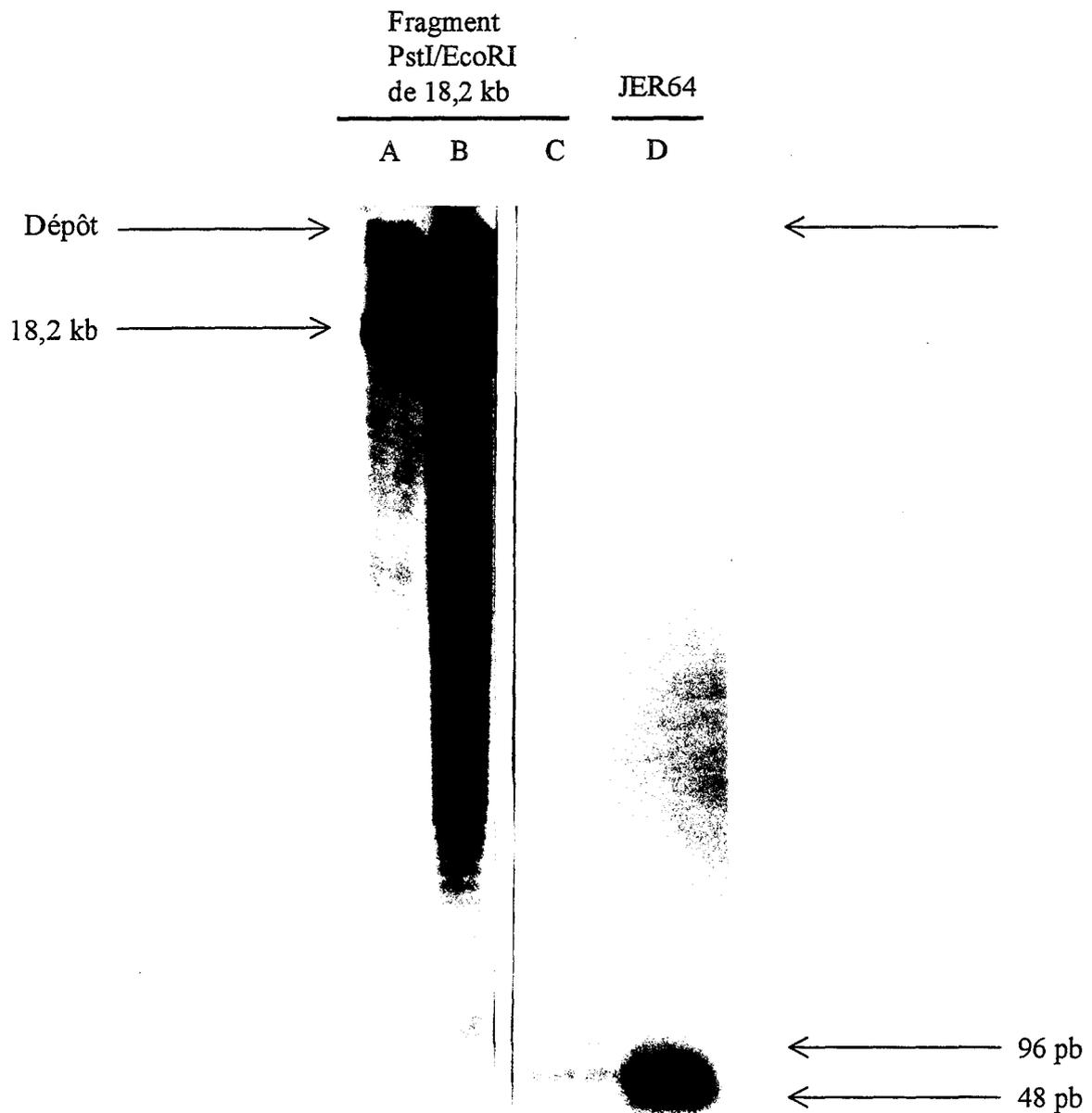


Figure 16 : Caractérisation du domaine contenant les répétitions de 48 pb de l'insert du clone cosmétique LEA2.

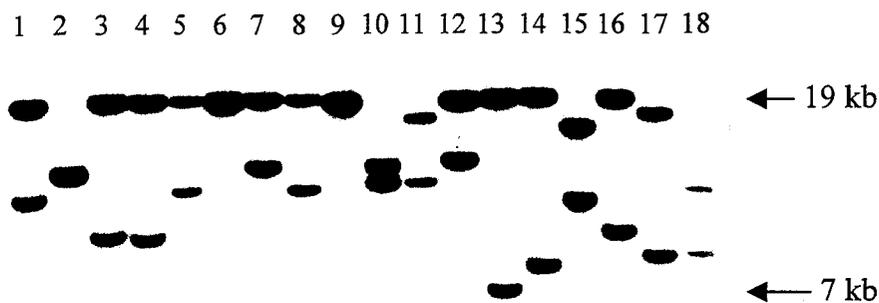
Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été réalisée à l'aide de la sonde JER64. Colonnes A-C, LEA2 a été hydrolysé aux sites *PstI* et *EcoRI* flanquant le domaine répétitif : colonne A, fragment *PstI/EcoRI* ; colonne B, fragment *PstI/EcoRI* partiellement hydrolysé pendant 1 h avec *DdeI* (0,14 U/μg d'ADN) ; colonne C, fragment *PstI/EcoRI* totalement hydrolysé avec *DdeI* (10 U/μg d'ADN). Colonne D, insert JER64 totalement hydrolysé avec *DdeI*.

La digestion enzymatique partielle par *DdeI* du fragment *PstI/EcoRI* de 18,2 kb de LEA2 (Figure 16, colonne B) montre un échelonnement régulier de bandes où chaque bande individuelle est un multiple de 48 pb. Au-dessus d'environ 20 bandes, on observe un continuum homogène rendant difficile l'estimation précise du nombre d'unités répétitives. On observe le même profil de digestion avec ANT55 et LEA47. D'après les calculs, la région répétitive des clones génomiques contient environ 208 (pour le clone phagique) et environ 380 (pour les clones cosmidiques) unités répétées en tandem.

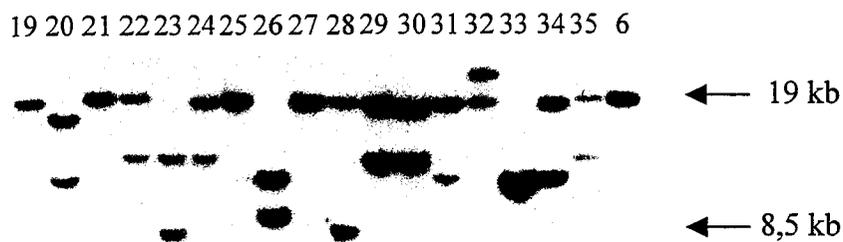
Il apparaît donc que le domaine répétitif de *MUC4* caractérisé par des unités de 48 pb homologues à celles de JER64, est un **grand domaine homogène, non interrompu par une séquence unique**.

II.2 - Le domaine répétitif présente un polymorphisme de type VNTR

Comme nous l'avons vu précédemment, la région contenant les répétitions en tandem de 48 pb est flanquée d'un site *PstI* (numéroté 3 sur les figures 11-13 et 15) à son extrémité 5' et d'un site *EcoRI* à son extrémité 3'. De l'ADN isolé à partir de lymphocytes de 35 volontaires (individus Caucasiens non apparentés du Nord de la France) a donc été hydrolysé par *PstI* et *EcoRI*. **La région répétitive s'est avérée hautement polymorphique** (Figure 17). Après hybridation avec la sonde JER64 des Southern blots obtenus, on observe deux bandes de même intensité chez 26 des 35 individus. Les neuf autres ne présentent qu'une seule bande de forte intensité. Ces deux bandes sont considérées comme étant les formes alléliques du domaine répétitif. Aucune bande supplémentaire indiquant la présence d'un site pour *PstI* ou *EcoRI* dans la région répétitive n'a été trouvée. Dans les conditions utilisées, **treize allèles différents** dont la taille varie **entre 19 kb et 7 kb** ont été observés (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). Le plus grand allèle (allèle A) est le plus commun. Cependant, la plupart des individus possédant le plus grand allèle (25/35) sont hétérozygotes (18/25), l'homozygotie pour cet allèle A ne concernant que sept individus. Chez les personnes hétérozygotes possédant l'allèle A, la taille du second allèle varie de 12 kb (allèle E) à 7 kb (allèle N). Les dix personnes ne possédant pas l'allèle le plus fréquent (allèle A) présentent divers autres allèles (allèles B, C, E, G, H, I, K et L). Deux de ces personnes



A E A A A A A A A E B A A A C A B H
 H E K K H A E H A G G E N M I J L L



A B A A E A A G A A A A A H A A A
 A H A E L E A K A L E E H A H H E A



Figure 17 : Le domaine répétitif de 48 pb présente un polymorphisme de type VNTR.
 L'ADN génomique de 35 individus a été hydrolysé par *Pst*I et *Eco*RI. Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été faite à l'aide de la sonde JER64. Les allèles observés pour chaque individu sont notés de A à N.

sont homozygotes, l'une pour l'allèle E et l'autre pour l'allèle H. Après analyse par Southern blot dans les mêmes conditions, il apparaît que la longueur du fragment *PstI/EcoRI* contenant le domaine répétitif des inserts cosmidiqes LEA2 et LEA47 correspond à celle de l'allèle B. Celle de l'insert bactériophagique ANT55 est la même que celle de l'allèle G.

II.3 - Le domaine répétitif présente un polymorphisme de mutation pour l'enzyme de restriction *RsaI*

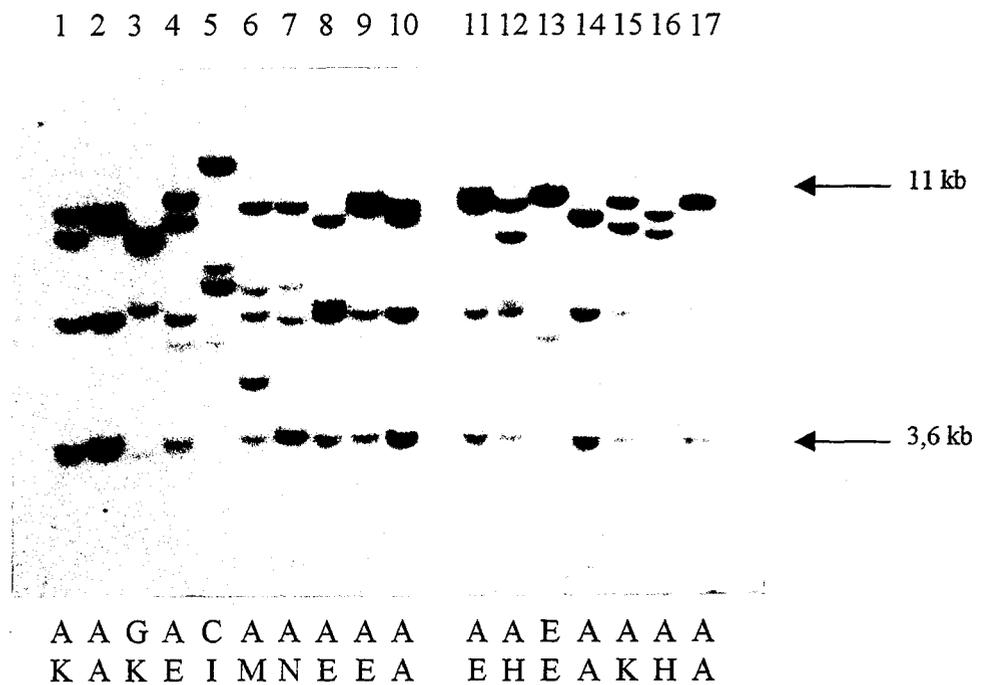
Nous avons observé que l'enzyme *RsaI* coupait de part et d'autre du domaine répétitif comme le faisait la double digestion par *PstI* et *EcoRI* mais qu'en outre des coupures internes existaient aussi. Le domaine répétitif des inserts des clones cosmidiqes LEA1 et LEA2 est en effet coupé en quatre fragments. Une hydrolyse par *RsaI* a pu être effectuée sur l'ADN de 17 des 35 individus étudiés précédemment par l'analyse des fragments coupés par *PstI* et *EcoRI*. Après hybridation avec la sonde JER64 du Southern blot obtenu, on observe pour chacun des individus analysés de deux à cinq fragments plus ou moins intenses (Figure 18).

Le sujet n°14 présente un profil de trois bandes d'intensité à peu près comparable (l'intensité de la bande la plus haute est néanmoins renforcée par rapport aux deux autres) respectivement de 9,5, 5,4 et 3,6 kb. La sommation de ces trois fragments provenant de la digestion par *RsaI* correspond à un fragment théorique de 18,5 kb obtenu par digestion par *PstI-EcoRI* alors que l'analyse préalable avec ces enzymes avait montré chez ce même sujet un profil homozygote pour un allèle A de 19 kb. Ceci montre que :

- l'estimation directe de la taille du fragment *PstI-EcoRI* (19 kb) ou que l'estimation indirecte de la taille de ce fragment par la somme de ses trois sous-fragments coupés par *RsaI* (18,5 kb) est assez concordante ;

- les sites *RsaI* ou *PstI-EcoRI* bornant le domaine répétitif sont très proches ;
- les deux allèles A du sujet n°14 ont la même organisation de leurs sites *RsaI*.

Les sujets n°2 et n°10 offrent le même profil de quatre bandes d'intensité à peu près comparable, de tailles respectives : 10, 9,5, 5,4 et 3,6 kb. Ces deux sujets sont également homozygotes pour l'allèle A. La comparaison de ce profil par rapport au précédent suggère que ces deux sujets soient en fait hétérozygotes pour deux variants A et A⁺. Le variant A a la même constitution que l'allèle A vu précédemment (9,5 + 5,4 + 3,6 = 18,5 kb) alors que l'allèle A⁺ est un peu plus long (10 + 5,4 + 3,6 = 19 kb), les deux allèles A (18,5 kb) et A⁺ (19



Numéros correspondants de la figure 17 4 27 26 24 15 14 13 12 35 6 7 1 2 19 3 34 21

Figure 18 : Le domaine répétitif de 48 pb présente un polymorphisme de mutation.

L'ADN génomique de 17 individus a été hydrolysé par *RsaI*. Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été faite à l'aide de la sonde JER64. Les allèles indiqués sont ceux observés lors de l'étude du polymorphisme de type VNTR avec la sonde JER64.

kb) n'étant pas dissociés dans le système électrophorétique utilisé pour l'analyse des fragments *PstI-EcoRI*. Le rapport d'intensité de marquage des fragments 10 et 9,5 kb (sujets n°2 et n°10) et 9,5 kb (sujet n°14) versus leurs fragments 5,4 et 3,6 kb correspondants permet d'envisager cette interprétation.

Le sujet n°17 présente également une homozygotie pour l'allèle A. Quatre bandes d'inégale intensité sont visibles, de tailles respectives 10, 5,5, 5,4 et 3,6 kb. De la même façon que précédemment, ces deux sujets sont en fait porteurs de deux allèles de type A définis de la façon suivante : $A^+ = 10 + 5,4 + 3,6 = 19$ kb et $A' = 10 + 5,5 + 3,6 = 19,1$ kb. L'intensité de moitié observée dans le marquage des fragments 5,5 et 5,4 kb confirme cette hypothèse, les deux allèles reconnus comme de type A : 19 kb et 19,1 kb étant indissociables en Southern blot.

Il ressort de l'étude structurale des allèles de ces quatre sujets classés AA que :

- l'allèle A possède quatre sites pour *RsaI* capables de le cliver en trois sous-fragments distincts,
- l'allèle A possède une hétérogénéité de situation précise de ces sites puisqu'un polymorphisme de longueur peut être perçu selon les individus sur deux de ses trois sous-fragments.

Les sujets n°1 et n°15 ont été détectés hétérozygotes AK (digestion *PstI-EcoRI*) et présentent un profil de digestion par *RsaI* similaire comportant quatre fragments. Parmi ces fragments, les bandes de longueurs respectives 10, 5,4 et 3,6 kb sont typiques de l'allèle A^+ . La quatrième bande estimée à 8,5 kb correspond donc à l'allèle K qui par ailleurs a été estimé dans le système de digestion par *PstI-EcoRI* à 8,9 kb.

Il ressort donc de cette étude que les allèles A et K ont une distribution différente des sites internes de clivage par *RsaI*.

Les sujets n°12 et n°16 ont été classés hétérozygotes AH (digestion *PstI-EcoRI*). Les profils obtenus après coupure par *RsaI* sont proches mais non similaires. L'ADN n°16 livre cinq fragments dont trois, de tailles respectives 9,5, 5,4 et 3,6 kb, correspondent à l'allèle A. Les deux autres fragments d'intensité de marquage très faible ont été estimés à 8,3 et 5,5 kb soit 13,8 kb en sommation. L'ADN n°12 livre également cinq fragments dont trois, de tailles respectives 10, 5,4 et 3,6 kb, correspondent à l'allèle A^+ . Les deux autres fragments ont une

taille de 8,3 et 5,5 kb. Ces deux fragments correspondent donc à l'allèle H. Or, l'allèle H a été précédemment estimé à 10,5 kb. Il existe donc une surestimation de la taille qui peut s'expliquer par une distribution différente, sur cet allèle H, des positions relatives des sites *Pst*I, *Eco*RI et *Rsa*I bornant le domaine répétitif. La faible intensité de marquage observée à propos du fragment de 5,5 kb étaye l'hypothèse d'un fragment contenant peu de structures identiques au motif de 48 pb de la sonde JER64.

L'ADN du sujet n°13 a livré un profil homozygote de type EE lors de l'analyse par *Pst*I-*Eco*RI. L'hydrolyse par *Rsa*I quant à elle livre deux fragments d'inégale intensité respectivement de 11 kb (très intense) et 5 kb (peu intense). Or, l'allèle E a été estimé à une taille de 12 kb. Ceci montre que cet allèle trouve à 1 kb de l'une de ses extrémités, un site de coupure par *Rsa*I générant le fragment de 11 kb. Un autre site est trouvé à 5 kb du premier. Le fragment de 5 kb ne comporte donc qu'1 kb de séquence reconnue par la sonde ce qui explique le signal d'intensité plus faible. Il faut remarquer également que le site bornant le domaine répétitif de ce côté a disparu.

L'ADN des sujets n°9 et n°11, dont le domaine répétitif est de type hétérozygote AE, offre le même profil de bandes après hydrolyse par *Rsa*I soit : trois bandes de 10, 5,4 et 3,6 kb, typiques de l'allèle A⁺, et deux bandes d'inégale intensité , de 11 et 5 kb, précédemment identifiées comme celles correspondant à l'allèle E.

En ce qui concerne l'individu n°4 (hétérozygote AE), le profil de bandes d'ADN obtenu s'apparente au groupe précédent avec le fragment correspondant à l'allèle A de 9,5 kb au lieu du fragment correspondant à l'allèle A⁺ de 10 kb.

Par contre, l'ADN du sujet n°8, identifié hétérozygote AE, offre un profil différent de ce que l'on attendait. Les bandes de l'allèle A (9,5, 5,4 et 3,6 kb) sont présentes. Par contre, la forte intensité de la bande à 5,4 kb laisse présager la superposition de deux fragments de même taille. Ceci amènerait donc à considérer que l'allèle E chez ce sujet contienne deux fragments de 5,6 et 5,4 kb au lieu des fragments de 11 kb et 5 kb vus précédemment chez le sujet n°13. Cet allèle a une taille proche de l'allèle E (11 kb au lieu de 12 kb) avec lequel il a été confondu. Il s'agit en fait d'un autre allèle.

Chez le sujet n° 6, dont le domaine répétitif est hétérozygote de type AM, les fragments typiques de l'allèle A⁺ sont observés. Nous pouvons en déduire que les fragments supplémentaires (6 et 4,6 kb) caractérisent l'allèle M.

De même chez le sujet n°7 (hétérozygote AN), le fragment de 6,1 kb caractérise l'allèle N.

Dans le cas du sujet n°3 dont le domaine répétitif correspond à une association d'allèles rares G et K, l'interprétation est délicate. Il est possible que la bande intense et large d'environ 8,5-8 kb contienne deux fragments de tailles proches dont l'un corresponde à l'allèle K, déjà observé dans les ADN n°1 et n°15. L'allèle G déduit serait alors composé de trois fragments de 8, 5,5 et 3,5 kb. La faible intensité des fragments de 5,5 et 3,5 kb laisse supposer que ces fragments contiennent peu de séquences répétitives reconnues par la sonde JER64 et que donc les deux sites *RsaI* bornant le domaine répétitif aient disparu, au profit de sites situés dans les deux domaines contigus.

Le profil donné par l'ADN du sujet n°5 n'est pas analysable dans cette étude dans la mesure où les allèles C et J sont rares et ne peuvent être interprétés par comparaison avec d'autres ADN. Ce profil est néanmoins très différent de ce que nous avons déjà vu par ailleurs.

Il apparaît donc que le domaine répétitif de *MUC4* présente un **polymorphisme interindividuel très complexe**. Ce polymorphisme est de **type VNTR**, c'est-à-dire qu'il est dû à une différence dans le nombre de répétitions de 48 pb. Les variations observées sont de forte amplitude puisque la taille du domaine répétitif varie dans un rapport de 1 à (2 à 3). Enfin, il existe également un **polymorphisme de mutation notamment pour l'enzyme de restriction *RsaI***. Chacun des allèles est caractérisé à la fois par la longueur du domaine répétitif et par le nombre et la position de mutations ponctuelles créant des irrégularités au sein des motifs élémentaires.

III - IDENTIFICATION D'UN DEUXIEME DOMAINE REPETITIF NON CODANT DE *MUC4* EN 3' DU DOMAINE REPETITIF (séquence de 15 pb présente dans la sonde HOR1)

III.1 - Caractérisation de la séquence répétitive de 15 pb

Le fragment *EcoRI* de 2,5 kb isolé à partir du clone ANT55 et appelé **HOR1**, a été sous-cloné en plasmide pKS et a pu être complètement séquencé grâce à la méthode de délétions à l'exonucléase III. Il est constitué de deux séquences uniques de 795 et 45 pb encadrant une **séquence répétitive dont le motif de 15 pb est très conservé** (Figure 19). La séquence du domaine répétitif de HOR1 se trouve dans la banque internationale EMBL sous le numéro d'accès AJ000282 (Figure 20). Cette séquence ne présente aucune similarité avec des séquences déjà décrites et en particulier aucune avec celle de JER64 (Figure 19). Dans HOR1, la séquence de 15 pb est répétée 110 fois (soit 1 650 pb).

III.2 - Etablissement de la continuité entre les deux séquences répétitives de 48 pb et de 15 pb

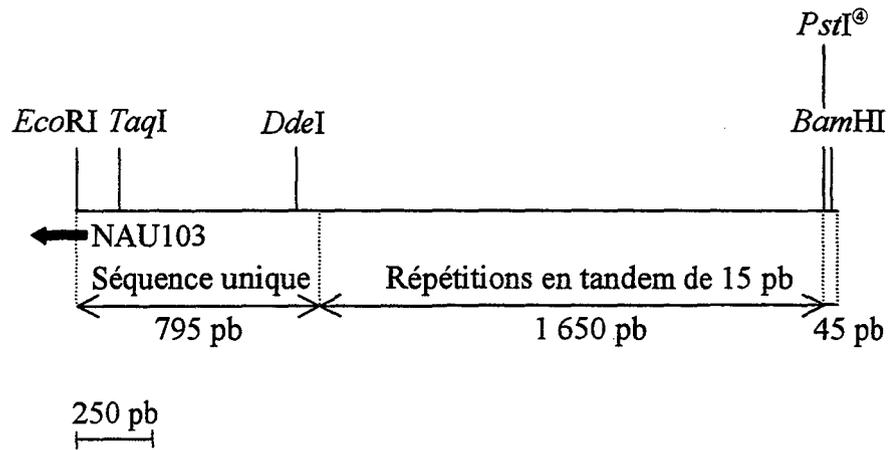
L'oligonucléotide NAU103 situé à l'extrémité 5' de HOR1 (Figure 21) a été choisi et utilisé pour la séquence directe de l'ADN du clone ANT55. En séquençant l'ADN de ANT55 à l'aide de NAU103, nous nous sommes rendus compte que le fragment HOR1 était situé à proximité de répétitions de 48 pb homologues à celles de JER64. Ces répétitions de 48 pb sont en effet situées à 85 pb du site *EcoRI*. Le sens de transcription de JER64 nous a permis d'orienter en aval de la séquence de 48 pb la séquence de HOR1. La séquence consensus du motif de 15 pb est la suivante : GGT GTG GAA GGT ATG.

NAU103 a également été utilisé pour le séquençage direct de l'ADN des cosmides LEA2 et LEA47. Les trois clones génomiques présentent la même séquence unique, entre les deux domaines répétitifs, selon le même agencement.

III.3 - HOR1 : exon ou intron ?

Nous avons longtemps pensé que la séquence répétitive de 15 pb était codante. En effet, une sonde ne contenant que le domaine répétitif de 15 pb avait été hybridée à des

A



B

1 ACT TCC TCA GCA TCC ACA GGT CAC GCC ACC CCT CTT CCT GTC ACC GAC
 2 GGT GTG GAA GGT ATG

Figure 19 : HOR1, fragment *EcoRI* de 2,5 kb isolé du clone bactériophagique ANT55.
 A, schéma d'organisation de HOR1.
 B, séquences nucléotidiques consensus des deux domaines répétitifs de : 1, JER64, 2, HOR1.

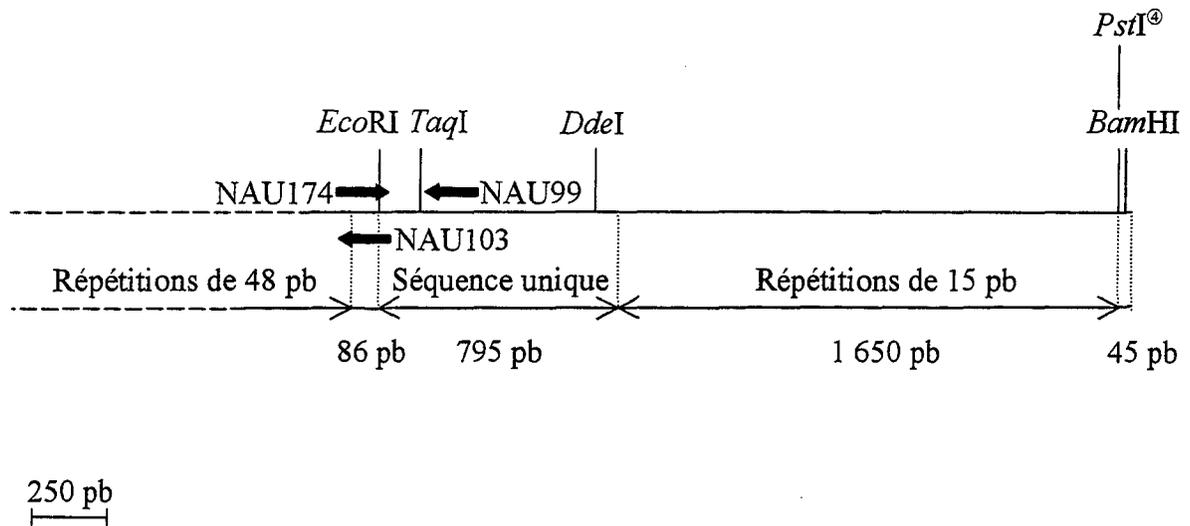
EcoRI

1	GAA	TTC	AGC	CTA	AGG	AGT	ACC	TGA	GAA	CAC	TGG	TGC	ATT	CGC	ATT	
46	AC	C	TGG	TGG	GGC	CGT	GTC	AGG	TCC	CAC	AGG	GGA	GGA	GGT	GAT	GGG
91	TGT	GGT	GGG	TGA	CAG	GCT	CAC	CCT	CCT	TTG	TGC	CGC	AAT	CGA	AAA	
136	GCA	CTG	ATG	TCG	AGA	GTA	GTT	TGG	ATA	TGA	GCA	GGG	GAG	AGA	CAA	
181	GGA	GTT	TCC	AGC	TCC	CTC	TTC	CAG	CTC	CTG	ATT	TCT	TTG	AAT	CTC	
226	TTT	GAC	TCT	CCT	GTT	TTG	TTA	CTG	TAA	GAA	ACA	CCC	CGC	CTT	GTC	
271	TTT	TCA	CGT	GTC	CAG	GAA	TGA	CAA	CAC	CGT	CAC	TGA	AGA	CAG	ACG	
316	GTG	GGA	GAC	GCA	CAG	CCA	CAT	CAC	CAC	CCC	CCA	CAA	CCT	CCC	AGA	
361	CCA	TCA	TTT	CCA	CCA	TTC	CCA	GCA	CTG	CCA	TGC	ACA	CCC	GCT	CCA	
406	CAG	CTG	CCC	CCA	TCC	CCA	TCC	TGC	CTG	AGA	GAG	GTG	AGG	CCA	TAC	
451	AGG	TGA	GGC	CTG	TGC	CTT	TTG	AGG	GGT	GAT	GTA	ACT	GAA	GGC	TCC	
496	CTC	TCA	GCC	TAC	TTC	CCA	CAG	TCT	CTG	CTC	TCT	CGG	GTG	GGG	AGG	
541	GCC	TTA	CCG	AGG	ACA	GGG	ACA	CAG	CAT	CGG	AGT	GCG	TCC	TGA	GGG	
586	CTG	GCT	TTG	TGC	ATG	GCA	CTG	GGC	CAG	GAG	CTG	GAG	ACA	GAG	AAA	
631	TGA	CCC	CAG	TGC	CAT	TCA	GCA	AGG	GAT	AGA	TGG	ACG	GTC	CGG	TAG	
676	CGG	CGG	TTA	GAG	GAC	TCA	TCC	CAG	GGT	CTA	AGT	GCA	CAC	AAT	GGA	
721	AGG	CCC	TAA	GGA	ATG	CAG	AGC	CGG	GGA	TGG	AGG	AGC	ACC	CCA	GGC	
766	AGG	GAG	GAG	GGC	GGG	AAC	AGC	TGG	AAC	AAA	GGT	GTG	GAA	GGT	ATG	
811	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
856	GCC	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
901	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
946	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
991	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
1 036	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 081	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GAT	GTG	GAA	GGT	ATG	
1 126	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
1 171	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 216	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
1 261	GGT	GTG	GAA	GGT	AAG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 306	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	AAG	GGT	GTG	GAA	GGT	ATG	
1 351	GGT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	
1 396	GGT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 441	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTT	GAA	GGT	ATG	
1 486	GGT	GTG	GAA	GTT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 531	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GAT	GTG	GAA	GGT	ATG	
1 576	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 621	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	
1 666	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	AAG	GGT	GTG	GAA	GGT	ATG	
1 711	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 756	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 801	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
1 846	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 891	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
1 936	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
1 981	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 026	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 071	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	
2 116	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 161	GCC	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	
2 206	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 251	ACT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ATG	
2 296	GGT	GTG	GAA	GGT	ATG	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 341	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	
2 386	ACT	GTG	GAA	GGT	ATG	GGT	GTG	GAA	GGT	ATG	GCC	GTG	GAA	GGT	ACG	
2 431	AGT	GTG	GTA	GGT	ATG	GCT	GCA	GAA	AGT	CGT	CCC	GGT	GGC	GGC	TGG	
2 476	GGG	GTG	GAT	CCG	GGG											

Figure 20 : Séquence nucléotidique de l'ADN du clone HOR1.

La position des nucléotides est inscrite sur la gauche. La localisation des oligonucléotides antisens NAU103 et NAU99 est indiquée par des flèches noires. Les flèches blanches délimitent le domaine contenant des répétitions en tandem de 15 pb.

A



B

```

1 TCTTCCCTGTCACCGACTTCCTCAGCATCCACAGGTCAGGCCACCCCTCTTCTGTCA
61 CCAGCCTTTCTCAGTATCCACAGGTCACGCCACCCCTCTTGCTGTACAGTACTACCT
   Répétitions en tandem de 48 pb
121 CAGCTTCCACAGTATCCTCGGACTCCCTCTGAAGATGGAAACATCAGGTAGCTGCCAAC
181 TGCCTCGCCTTTTATGTCTCCAGTGGGCCCTTGGCG|AATTCAGCCTAAGGAGTACCTGA
   EcoRI
241 GAACACTGGTGCATTTCGATTAC ← NAU103
    
```

Figure 21 : Liaison entre les deux séquences répétitives de 48 pb et de 15 pb.

A, représentation schématique de la liaison entre les deux séquences répétitives.

B, séquence obtenue à l'aide de l'oligonucléotide NAU103.

Northern blots, préparés au laboratoire ou commerciaux, et composés d'ARN de divers tissus humains. On observait alors un continuum homogène de transcrits s'étendant de 1 à 20 kb environ et parfaitement superposable au signal observé avec la sonde JER64 (Figure 22). Par ailleurs, un certain nombre d'échantillons de muqueuses coliques ont été étudiés par la technique de RT-PCR et nous a donné des séquences nucléotidiques dont nous savons maintenant qu'elles étaient réarrangées.

Par contre, lorsque cette même sonde a pu être hybridée à un Northern blot dont les échantillons d'ARN ont été préparés dans des conditions dites "douces" (en évitant les cassures mécaniques) et transférés avec un traitement préalable à la soude (pour augmenter le transfert difficile des gros ARN) [Debailleul *et al.* 1998], aucun signal n'a été observé alors que sur les mêmes blots, les transcrits de *MUC4* détectés à l'aide de la sonde JER64 apparaissaient sous forme de bandes bien définies (un transcrit chez les sujets homozygotes, deux transcrits chez les sujets hétérozygotes). De même, l'utilisation d'un oligonucléotide composé de 3 unités de 15 pb utilisé par HIS, donne également un résultat négatif sur tous les échantillons exprimant *MUC4*, qui sont par ailleurs positifs avec un oligonucléotide typique de la séquence de JER64.

Afin de caractériser l'extrémité 3' de l'exon central, c'est à dire l'exon contenant le domaine répétitif de 48 pb, nous avons effectué des RT-PCR en utilisant comme matrice des ARN poly(A)⁺ de trachée humaine et de l'ARN total de côlon humain. Nous avons choisi comme amorce sens un oligonucléotide situé dans la séquence unique juste en aval des répétitions de 48 pb (NAU174) et comme amorce antisens un oligonucléotide situé dans la séquence unique en amont des répétitions de 15 pb (NAU99) (Figure 21). Les différents produits d'amplification ont été clonés en vecteur pCR2.1 et séquencés. La comparaison des séquences d'ADNc et d'ADN génomique a mis en évidence la présence d'un **intron de 333 pb** (Figure 23) situé à 37 pb en aval du domaine répétitif de 48 pb. Les séquences des sites accepteur et donneur d'épissage sont conformes à la règle "GT-AG" [Mount 1982] :

consensus du site donneur :	C/A A G <u>g t</u> a/g a g t
site donneur de l'intron :	C A G <u>g t</u> a g c t
consensus du site accepteur :	c/t n c/t <u>a g</u>
site accepteur de l'intron :	t t t t c a c g t g t c c <u>a g</u>

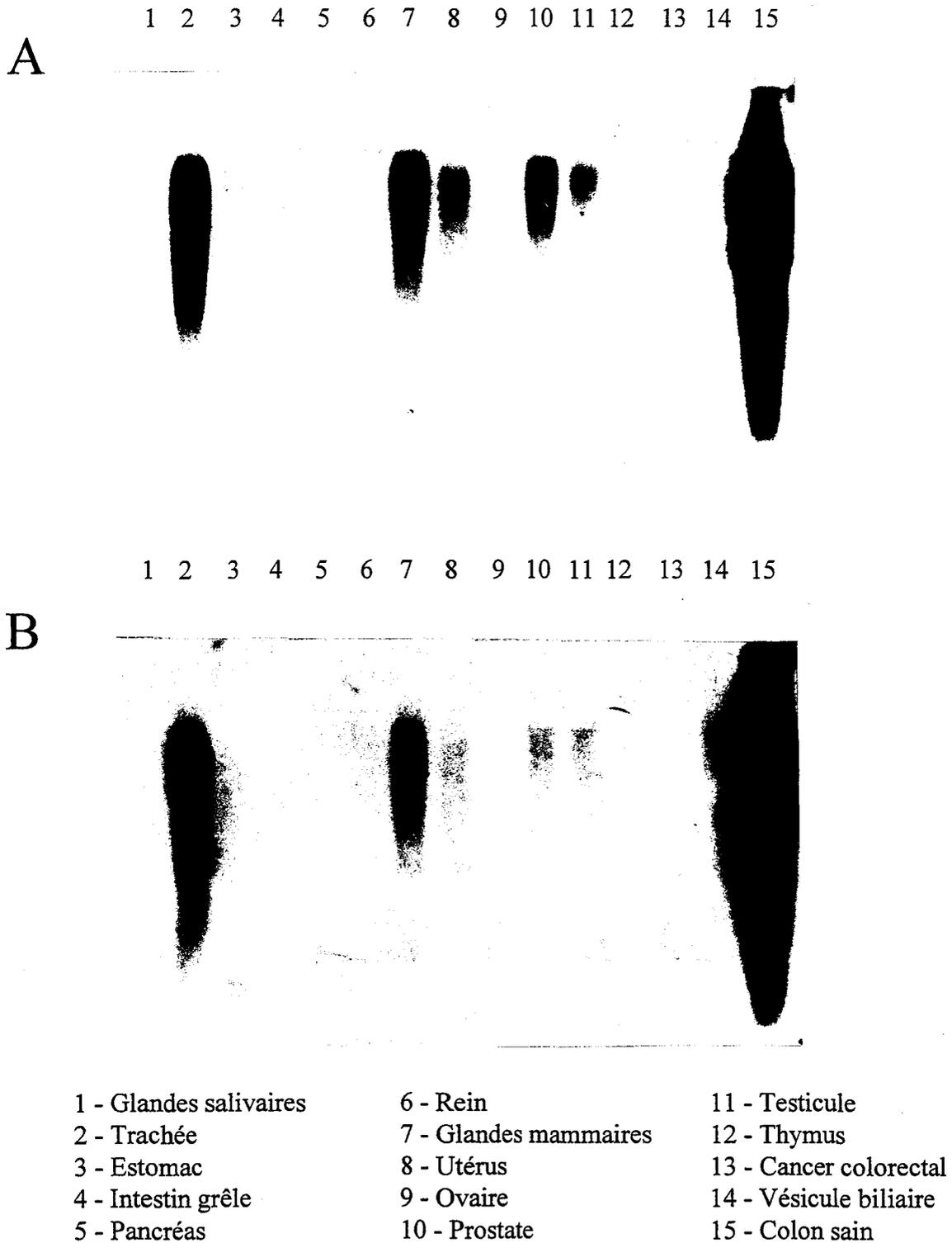


Figure 22 : Analyse par Northern blot de l'expression du gène *MUC4* à partir d'un blot commercial poly(A⁺) (Clontech laboratories).

A, sonde utilisée : JER64 (répétitions de 48 pb).

B, sonde utilisée : répétitions de 15 pb de HOR1.

A

```

1  NAU174
  GCT TCC ACA GTA TCC TCG GAC TCC CCT CTG AAG ATG GAA ACA TCA
exon  ↔ intron
46  Ggtagctgccaaactgcctgccttttatgtctccagtgggcccttggcgaattcagcct
106 aaggagtacctgagaacactgggtgcattcgcattacctgggtggggccgtgtcagggtccca
166 caggggaggagggtgatgggtgtgggtgggtgacaggctcaccctcctttgtgcccgaatcg
226 aaaagcactgatgtcgagagtagtttggatatgagcaggggagagacaaggagtttccag
286 ctccctcttccagctcctgatttctttgaatctctttgactctcctgttttgttactgta
346 agaaacacccccgccttgtcttttcaegtgtccagGA NAU99 ATG ACA ACA CCG TCA CTG
400 AAG ACA GAC GGT GGG AGA CGC ACA GCC A

```

B

```

GCT TCC ACA GTA TCC TCG GAC TCC CCT CTG AAG ATG GAA ACA TCA
A S T V S S D S P L K M E T S 15
▼
GGA ATG ACA ACA CCG TCA CTG AAG ACA GAC GGT GGG AGA CGC ACA
G M T T P S L K T D G G R R T 30
GCC
A 31

```

Figure 23 : Séquences nucléotidique et peptidique déduite des clones obtenus par PCR et RT-PCR entre NAU174 et NAU99.

A, séquence génomique. La position des nucléotides est inscrite sur la gauche. La localisation des oligonucléotides NAU174 et NAU99 est indiquée par des flèches noires. La séquence des exons est en lettres majuscules, celle de l'intron en minuscules.

B, séquence de l'ADNc. La position des acides aminés est inscrite sur la droite. La position de l'intron est indiquée par un triangle noir.

Cet intron est de **classe 1** puisqu'il interrompt la séquence codante entre la première et la seconde base du codon [Patthy 1987].

Nous avons donc réussi à borner l'exon contenant le domaine répétitif du côté 3'. Cet exon se termine par une dernière répétition en tandem dégénérée suivie d'une séquence unique de 31 pb. De façon étonnante, cette **séquence codante répétitive est suivie d'une autre séquence elle aussi répétitive, très régulière et homogène mais intronique**. Cette séquence non codante répétitive a été confirmée dans sa séquence à partir de trois clones génomiques différents, mais elle s'est avérée de longueur variable : elle est de 110 répétitions (1,65 kb) dans le clone bactériophagique et d'environ 160 répétitions (environ 2,4 kb) dans les deux clones cosmiques. Cette différence de taille nous a fait suspecter l'existence d'un polymorphisme interindividuel.

III.4 - Polymorphisme de HOR1

III.4.1 – Polymorphisme de type VNTR

Pour étudier le polymorphisme de HOR1, nous avons réutilisé les Southern blots dont les ADN avaient été hydrolysés par *PstI* et *EcoRI*. En effet, le domaine contenant les répétitions de 15 pb est borné en 5' par le site *EcoRI* (site bornant également en 3' le domaine contenant les répétitions de 48 pb) et en 3' par un site *PstI* (numéroté 4 sur les figures 10, 12 et 14). Après hybridation avec la sonde HOR1 (Figure 24), on observe **12 allèles différents** dont la taille varie entre 6,5 et 1,5 kb (a : 6,5 kb ; b : 6,2 kb ; c : 5,5 kb ; d : 5,2 kb ; e : 5 kb ; f : 4 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). L'allèle le plus fréquent retrouvé chez 27 individus sur les 35 étudiés est l'allèle h. Sur ces 27 individus, 20 sont hétérozygotes et 7 sont homozygotes. Ces derniers sont également homozygotes pour l'allèle A (allèle observé avec la sonde JER64). Le taille du fragment *EcoRI/PstI* contenant le domaine caractérisé par des répétitions de la séquence de 15 pb des deux inserts cosmiques (LEA2 et LEA47) correspond à celle de l'allèle h (3,6 kb). Celle du même fragment de l'insert bactériophagique (ANT55) correspond à celle de l'allèle j (2,8 kb).

III.4.2 - Association allélique entre les deux domaines répétitifs de 48 pb et 15 pb

En comparant les profils obtenus avec les deux sondes répétitives, il est apparu une certaine **concordance allélique entre JER64 et HOR1** (Tableau 5). Il apparaît que les

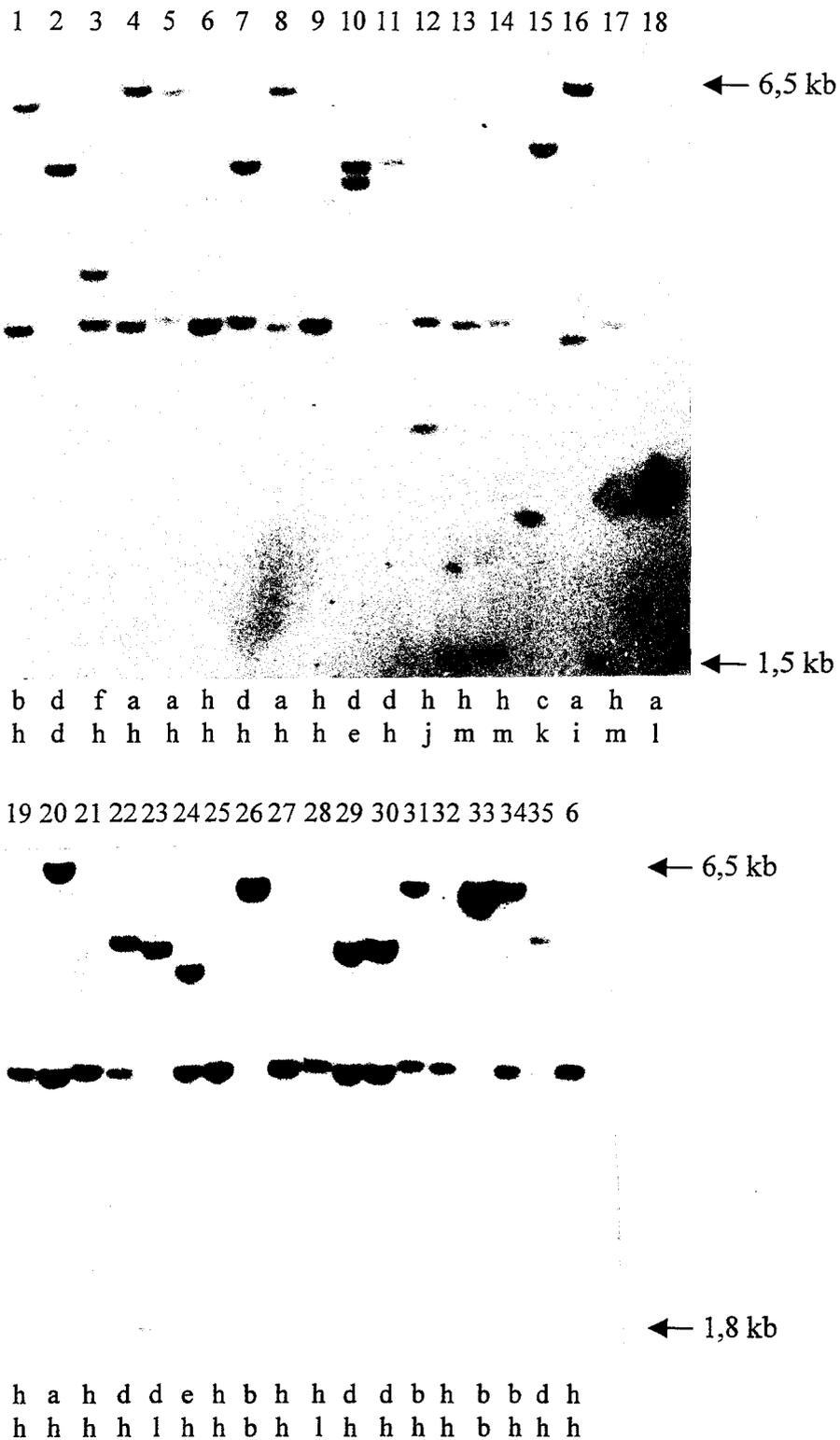


Figure 24 : Le domaine répétitif de 15 pb présente un polymorphisme de type VNTR.
 L'ADN génomique de 35 individus a été hydrolysé par *Pst*I et *Eco*RI. Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été faite à l'aide de la sonde HOR1. Les allèles observés pour chaque individu sont notés de a à m.

Numéro de l'individu	Répétitions de 48 pb	Répétitions de 15 pb	Haplotypes
			<u>Réels</u>
6	A/A	h/h	A-h
9	A/A	h/h	A-h
19	A/A	h/h	A-h
21	A/A	h/h	A-h
25	A/A	h/h	A-h
27	A/A	h/h	A-h
32	A/A	h/h	A-h
2	E/E	d/d	E-d
33	H/H	b/b	H-b
26	G/K	b/b	G-b
			K-b
			<u>Déduits</u>
7	A/E	d/h	A-h
22	A/E	d/h	E-d
29	A/E	d/h	A-h
30	A/E	d/h	E-d
35	A/E	d/h	A-h
			E-d
1	A/H	b/h	A-h
31	A/H	b/h	H-b
34	A/H	b/h	A-h
			H-b
24	A/E	e/h	A-h
			E-e
12	A/E	h/j	A-h
			E-j

Numéro de l'individu	Répétitions de 48 pb	Répétitions de 15 pb	Haplotypes
5	A/H	a/h	A-h H-a
8	A/H	a/h	A-h H-a
3	A/K	f/h	A-h K-f
4	A/K	a/h	A-h K-a
28	A/L	h/l	A-h L-l
14	A/M	h/m	A-h M-m
13	A/N	h/m	A-h N-m
10	E/G	d/e	E-d G-e
23	E/L	d/l	E-d L-l
18	H/L	a/l	<u>Supposés</u> H-a L-l
20	B/H	a/h	H-a B-h
11	B/G	d/h	B-h G-d
17	B/L	h/m	B-h L-m
15	C/I	c/k	<u>Indéterminés</u> C-c ou C-k I-c ou I-k
16	A/J	a/i	A-a ou A-i J-a ou J-i

Tableau 5 : Liaison entre les allèles observés avec la sonde JER64 (en majuscules) et ceux observés avec la sonde HOR1 (en minuscules).

individus homozygotes pour l'allèle détecté à l'aide de la sonde JER64 (9/35) le sont aussi avec la sonde HOR1. De plus, on observe que l'allèle A, qui est l'allèle le plus représenté (25/35), est toujours lié à l'allèle h, détecté à l'aide de la sonde HOR1, sauf pour un individu dont l'haplotype n'a pu être déterminé. Trois individus présentant l'allèle h ne possèdent pas l'allèle A. Par contre, l'allèle h est alors lié à l'allèle B, allèle dont la taille est la plus proche de celle de l'allèle A. Il semblerait donc que la majorité des individus (27/35) présentent un allèle de très grande taille avec la sonde JER64 lié à un allèle de petite taille avec la sonde HOR1. Cependant, les individus possédant un allèle plus petit pour la sonde JER64 présentent un allèle de plus grande taille pour la sonde HOR1 (exemple : l'allèle H lié à l'allèle b chez les individus n°33, 1, 31, 34).

III.4.3 – Etude haplotypique du polymorphisme de type VNTR du gène MUC4

Nous avons tenté d'apprécier les diversités d'organisation des régions répétitives de 48 pb et de 15 pb du gène *MUC4* en analysant les associations haplotypiques de 35 ADN provenant de sujets non apparentés (Tableau 5).

Parmi les 35 sujets, sept homozygotes pour l'allèle A le sont également pour l'allèle h. Ceci montre une certaine fréquence de l'haplotype réel A-h. Un sujet homozygote pour l'allèle E l'est aussi pour l'allèle d ; un autre, homozygote pour l'allèle H, l'est aussi pour l'allèle b. Ceci montre l'existence de deux autres haplotypes réels : E-d et H-b. Enfin, un sujet est hétérozygote pour G/K mais homozygote b/b ce qui montre encore l'existence de deux haplotypes réels G-b et K-b. A partir de ces haplotypes dont l'existence est réelle, nous avons tenté de déduire l'organisation haplotypique des autres ADN.

Cinq sujets sont hétérozygotes A/E et d/h. Nous pouvons postuler que l'organisation haplotypiques est A-h et E-d en fonction des observations précédentes. Trois sujets sont hétérozygotes A/H et b/h. De même, nous proposons les haplotypes A-h et H-b déjà rencontrés. Considérant ensuite, chez neuf sujets hétérozygotes possédant A et h, que ces allèles étaient liés sur le même haplotype, nous en déduisons les haplotypes associés : E-e, E-j, H-a, K-f, K-a, L-l, M-m, N-m. De même, chez deux sujets hétérozygotes possédant E et d, nous proposons comme probable l'association E-d et déduisons les associations G-e et L-l (déjà rencontrées).

Quatre autres ADN présentent des allèles rares à la fois du domaine répétitif de 48 pb et du domaine répétitif de 15 pb. A la lumière des associations haplotypiques déjà envisagées (L-l, H-a et B-h), nous supposons également l'existence des associations B-h, G-d et L-m.

Ces résultats montrent que l'analyse de l'ADN de 35 sujets, soit de 70 chromosomes, révèle l'existence d'au moins 21 associations haplotypiques différentes dans l'organisation du gène *MUC4*, ce qui ne préjuge pas d'une complexité accrue encore par l'analyse des autres domaines constitutifs codants ou non codants du gène.

IV - CARACTERISATION DE LA SEQUENCE EN 5' DU DOMAINE REPETITIF

IV.1 - Construction de deux banques d'ADNc (intestin grêle et côlon)

Deux banques d'expression ont été construites au laboratoire, l'une à partir d'intestin grêle, l'autre à partir de côlon, ces deux tissus exprimant le gène *MUC4*. Seule la banque de côlon a été criblée avec la sonde JER64, la banque d'intestin grêle ayant un titre plus faible : $2,5 \times 10^6$ pfu (intestin grêle) contre $10,3 \times 10^6$ pfu (côlon).

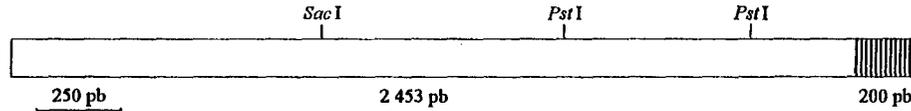
IV.2 - Caractérisation des clones d'ADNc obtenus par criblage de la banque d'expression colique

Dans le cas de la banque d'ADNc, comme vu précédemment lors des expériences de RT-PCR, un certain nombre de clones ont présenté des réarrangements de séquences. Par exemple, le clone JER102 mettait en continuité des répétitions de 48 pb typiques de JER64 avec un polyT suivi du gène de la protéine transmembranaire T245.

Néanmoins, il a été possible d'analyser le clone JER103. JER103 a une longueur de **2 653 pb** (Figure 25). **Les 200 pb situées à l'extrémité 3' consistent en des répétitions de 48 pb** identiques à celles trouvées dans **JER64**. Nous avons donc choisi le cadre de lecture de JER103 en fonction de celui de l'ADNc JER64. La séquence nucléotidique de JER103 révèle la présence d'un **fragment unique codant un peptide riche en résidus Thr et Ser** (21 % Thr et 20 % Ser). Un résidu Cys est trouvé en position 95.

Différents oligonucléotides choisis dans la séquence de JER103 ont été utilisés pour séquencer directement l'ADN des trois clones génomiques (LEA2, LEA47 et ANT55) (NAU170, NAU167, NAU225, NAU247 et NAU224 ; la position de ces oligonucléotides est indiquée dans l'appendice technique, tableau 6). La séquence unique est retrouvée à l'identique sur les cosmides LEA2 et LEA47 ainsi que sur le phage ANT55 (ce dernier ne possédant pas les 439 premières pb) immédiatement en amont du domaine répétitif de 48 pb. Ces deux régions riches en Thr/Ser (la région unique et la région répétitive) étant en continuité, elles sont donc codées par **un seul et même exon**.

A



B

<p>1 A GCA GGA ACT GAA AGT TCT ACC CCA GTG ACC TCA GCA GTC TCA ATA ACA GCT GGA CAG A G T E S S T F V T S A V S I T A G G Q 19</p> <p>59 GAA GGA CAA TCA CGA ACA ACT TCG TGG AGG ACC TCT ATC CAA GAC ACA TCA GCT TCT TGT E G Q S R T T S W R T S I Q D T S A S S 39</p> <p>119 CAG AAC CAC TGG ACT CGG AGC ACG CAG ACC ACC AGG GAA TCT CAA ACC ACC ACC CTA ACA Q N H W T R S T Q T T R E S Q T S T L T 59</p> <p>179 CAC AGA ACC ACT TCA ACT CCT TCT TTC TCT CCA AGT GTA CAC AAT GTG ACA GGG ACT GTT H R T T S T F S F S P S V H N V T G T V 79</p> <p>239 TCT CAG AAG ACA TCT CCT TCA GGT GAA ACA GCT ACC TCA TCC CTC TOT AAT GTC ACA AAC S Q K T S F S G E T A T S S L C S V T T 99</p> <p>299 ACA TCC ATG ATG ACA TCA GAG AAG ATA ACA GTG ACA ACC TCC ACA GGC TCC ACT CIT GGA T S M M T S E K I T V T T S T G S T L G 119</p> <p>359 AAC CCA GGG GAG ACA TCA TCA GTA CCT GTT ACT GGA AGT CIT ATG CCA GTC ACC TCA GCA N P G E T S S V P V T G S L M P V T S A 139</p> <p>419 GCC TTA GTA ACA GTT GAT CCA GAA GGA CAA TCA CCA GCA ACT TTC TCA AGG ACT TCT ACT A L V T V D P E G Q S P A T F S R T S T 159</p> <p>479 CAG GAC ACA ACA GCT TTT TCT AAG AAC CAC CAG ACT CAG AGC GTG GAG ACC ACC AGA GTA Q D T T A F S K N R H Q T Q S F E T T R V 179</p> <p>539 TCT CAA ATC AAC ACC CTC AAC ACC CTC ACA CCG GTT ACA ACA TCA ACT GTT TTA TCC TCA S Q I N T L N T L T P V T T S T G S T L G 199</p> <p>599 CCA AGT GGA TTC AAC CCA AGT GGA ACA GTT TCT CAG GAG ACA TTC CCT TCT GGT GAA ACA P S G F N F S G T V S Q E T F P S G E T 219</p> <p>659 ACC ATC TCA TCC CCT TCC AGT GTC AGC AAT ACA TTC CTG GTA ACA TCA AAG GTG TTC AGA T I S S P S S V S N T F L V T S K V F R 239</p> <p>719 ATG CCA ATC TCC AGA GAC TCT ACT CIT GGA AAC ACA GAG GAG ACA TCA CTA TCT GTA AGT M P I S R D S T L G N T E E T S L S V S 259</p> <p>779 GGA ACC ATT TCT GCA ATC ACT TCC AAA GTT TCA ACC ATA TGG TGG TCA GAC ACT CTG TCA G T I S A T T S K V S T I W W S D T L S 279</p> <p>839 ACA GCA CTC TCC CCC AGT TCT TTA CCT CCA AAA ATA TCC ACA GTC TTC CAC ACC CAG CAG T A L S F S S L P P K I S T A F H T Q Q 299</p> <p>899 AGT GAA GGT GCA GAG ACC ACA GGA CCG CCT CAT GAG AGG AGC TCA TTC TCT CCA GGT GTG S E G A E T T G R P H E R S S P F G V 319</p> <p>959 TCT CAA GAA ATA TTT ACT CTA CAT GAA ACA ACA TGG CCT TCC TCA TCT TCC ACC AAA S Q E I F T L H E T T T W F S F S X 339</p> <p>1019 GGC CAC ACA ACT TGG TCA CAA ACA GAA CTG CCC TCA ACA TCA ACA GGT GCT GCC ACT AGG G H T T W S Q T E L P S T S T G A A T R 359</p> <p>1079 CIT GTC ACA GGA AAT CCA TCT ACA GGG GCA GCT GGC ACT ATT CCA AGG GTC CCC TCT AAG L V T G N P S T G A A G T I P R V P S K 379</p> <p>1139 GTC TCA GCA ATA GGG GAA CCA GGA GAG CCC ACC ACA TAC TCC TCC CAC ACC ACA ACT CTC V S A I G E F G E P T T Y S H S T T L 399</p> <p>1199 CCA AAA ACA ACA GGG GCA GGC GCC CAG ACA CAA TGG ACA CAA GAA ACC GGC ACC ACT GGA P K T T G A G A Q T Q W T G T T G 419</p> <p>1259 GAG GCT CIT CTC AGC AGC CCA AGC TAC AGT GTG ACT CAG ATG ATA AAA AGG ACC ACA TCC E A L L S S F S Y S V T Q M I A C G C S 439</p>	<p>1319 CCA TCT TCT TCA CCT ATG CTG GAT AGA CAC ACA TCA CAA CAA ATT ACA ACG GCA CCA TCA P S S S F M L D R H T S Q Q I T T A F S 459</p> <p>1379 ACA AAT CAT TCA ACA ATA CAT TCC ACA AGC ACC TCT CCT CAG GAA TCA CCA GCT GTT TCC T N H S T I H S T S T S F Q E S P A V S 479</p> <p>1439 CAA AGG GGT CAC ACT CAA GCC CCG CAG ACC ACA CAA GAA TCA CAA ACC ACG AGG TCC GTC Q R G H T Q A P Q T T Q E S Q I T R S V 499</p> <p>1499 TCC CCC ATG ACT GAC ACC AAG ACA GTC ACC ACC CCA GGT TCT TCC TTC ACA GCC AGT GGG S P M T D T K T V T T P G S S F T A S G 519</p> <p>1559 CAC TCG CCG TCA GAA ATT GTT CCT CAG GAC GCA CCC ACC ATA AGT GCA GCA ACA ACC TTT H S F S E I V P Q D A F T I S A A T T F 539</p> <p>1619 GCC CCA GCT CCC ACC GGG GAT GGT CAC ACA ACC CAG GCC CCG ACC ACA GCA CTG CAG GCA A P A F T G D G H T T Q A F T T A L Q A 559</p> <p>1679 ACA CCC AGC AGC CAT GAT GCC ACC CTG GGG CCC TCA GGA GGC ACG TCA CIT TCC AAA ACA T P S S H D A T L G F S G G T S L S K T 579</p> <p>1739 GGT GCC CIT ACT CTG GCC AAC TCT GTA GTG TCA ACA CCA GGG GGC CCA GAA GGA CAA TGG G A L T L A N S V V S T P G F E G Q W 599</p> <p>1799 ACA TCA GGC TCT GCC ACC TCA CCT GAC ACA GCA GCA GCC ATA ACC CAT ACC CAG CAG T S A S A S T S P D T A A A M T H T H Q 619</p> <p>1859 GCT GAG AGC ACA GAG GCC TCT GGA CAA ACA CAG ACC AGC GAA CCG GCC TCC TCA GGG TCA A E S T E A S G Q T Q T S E F A S S G S 639</p> <p>1919 CCA ACC ACC TCA GCG GGC ACA GCT ACC CCT TCC TCA TCC GGG GCG AGT GGC ACA ACA CCT R T Y S A G T A T F S S S G A S G T T P 659</p> <p>1979 TCA GGA AGC GAA GGA ATA TCC ACC TCA GGA GAG ACG ACA AGG TTT TCA TCA AAC CCC TCC S G S E G I S T S G E T T R F S S N P S 679</p> <p>2039 AGG GAC AGT CAC ACA ACC CAG TCA ACA ACC GAA TTG CTG TCC GCC TCA GCC AGT CAT GGT R D S H T T Q S T T E L L S A S A S H G 699</p> <p>2099 GCC ATC CCA GTA AGC ACA GGA ATG GCG TCT TCG ATC GTC CCC GGC ACC TTT CAT CCC ACC A I P S S T G M A S S I V F G T F H P T 719</p> <p>2159 CTC TCT GAG GCC TCC ACT GCA GGG AGA CCG ACA GGA CAG TCA AGC CCA ACT TCT CCC AGT L S E S T A G R F T G Q S P S 739</p> <p>2219 GCC TCT CCT CAG GAG ACA GCC GCC ATT TCC CCG ATG GCC CAG ACT CAG AGG ACA AGA ACC A S P Q E T A A I S R M A Q T R T 759</p> <p>2279 AGC AGA GGG TCT GAC ACT ATC AGC CTG GCG TCC CAG GCA ACC GAC ACC TTC TCA ACA GTC S R G S S D T I S L A S Q A T D T F S T V 779</p> <p>2339 CCA CCC ACA CCT CCA TCG ATC ACA TEC AGT GGG CIT ACA TCT CCA CAA ACC CAG ACC CAC P P T P F S I T S S G L T S P Q T Q T H 799</p> <p>2399 ACT CTG TCA CCT TCA GGA TCT GGT AAA ACC TTC ACC ACG CTC ATC AGC AAC GCC ACC T L S P S G S G K T F T T A L I S N A T 819</p> <p>2459 CCT CIT CCT GTC ACC TAC GCT TCC TCG GCA TCC ACA GGT CAC ACC CCT CIT CAT GTC P L F V T Y A S S A S T G H T T F L H V 839</p> <p>2519 ACC GAT GCT TCC TGA TCA TCC ACA GGT CAC GCC ACC CCT CIT CCT GTC ACC AGC CCT TCC T D A S S V S T G H A W T F L F T S 859</p> <p>2579 TCA GTA TCC ACA GGT CAC ACC CCT CIT CCT GTC ACC GAC GCT TCC GTA TCC ACA S V S T G H T T P L F V T D A S S V T S 879</p> <p>2639 GGC CAC CCC ACC CCA G H P T P 884</p>
--	---

Figure 25 : Le clone d'ADNc JER103.

A, Cartographie partielle de JER103. Le rectangle blanc représente la séquence unique, le rectangle hachuré verticalement représente le domaine composé de répétitions en tandem de 48 pb.

B, Séquences nucléotidique et peptidique déduite de JER103. La position des nucléotides est inscrite sur la gauche, celle des acides aminés correspondants sur la droite. La localisation des oligonucléotides NAU124 et NAU138 (oligonucléotides utilisés pour la première expérience de 5' RACE-PCR), est indiquée par des flèches noires. Trois sites potentiels de N-glycosylation sont encadrés. La position du seul résidu Cys de cette séquence est indiquée par un astérisque.

IV.3 - Obtention de clones d'ADNc par la technique de 5' RACE-PCR

Deux oligonucléotides antisens, **NAU124** et **NAU138**, ont été utilisés pour allonger la séquence du côté 5' par une expérience de RACE-PCR réalisée à partir d'ARN total de cølon humain. Un fragment d'ADNc de **358 pb (RAC1 ; Figure 26)** a été obtenu et sa séquence déterminée. La séquence de son extrémité 3' chevauchant la séquence de l'extrémité 5' de JER103, RAC1 correspond bien à un ADNc de *MUC4*. Cette expérience a été suivie par une seconde expérience de RACE-PCR utilisant deux nouvelles amorces déduites de la séquence de RAC1 (**NAU155** et **NAU156**). Nous avons ainsi cloné un nouvel ADNc de **300 pb (RAC2)**. Une autre expérience de RACE 5' utilisant les oligonucléotides **NAU168**, **NAU169** et **NAU204** (choisis dans la séquence de RAC2) a permis l'obtention d'un clone de **183 pb (RAC3)**. Deux autres clones RAC, **RAC4 (140 pb)** et **RAC5 (353 pb)** ont ensuite été obtenus à l'aide de 4 nouveaux oligonucléotides (**NAU287**, **NAU288**, **NAU289** et **NAU327**). Ces cinq clones RAC nous ont permis d'allonger la séquence de l'ADNc de *MUC4* du côté 5' de **942 pb (Figure 27)**.

Afin de vérifier la réalité de ces séquences, nous avons effectué des expériences de RT-PCR sur de nouveaux échantillons biologiques : des RT-PCR utilisant les amorces **NAU162** et **NAU138** ont été réalisées à partir d'ARN de muqueuses trachéobronchique et colique humaines. Des produits d'amplification de **563 pb** ont été obtenus et clonés. La séquence nucléotidique obtenue est 100 % similaire à la séquence précédemment déduite par compilation des séquences des clones RAC2 et RAC1.

IV.4 - Comparaison des séquences génomiques et des séquences ADNc en amont du domaine répétitif

La séquence codante de RAC1 est identique à celle trouvée dans les clones génomiques LEA2 et LEA47, immédiatement en amont de la séquence correspondant au clone JER103. La séquence de la région 3' de RAC2 est retrouvée sur les clones génomiques LEA2 et LEA47 sur 219 pb seulement. Par contre, les 81 premières pb de RAC2 ne sont pas retrouvées dans ces clones génomiques. La présence d'une région intronique a donc été suspectée.

La sonde RAC3 ne s'hybride pas aux fragments d'ADN génomique des clones cosmidiques LEA2 et LEA47 analysés par Southern blot, indiquant que la séquence 5'

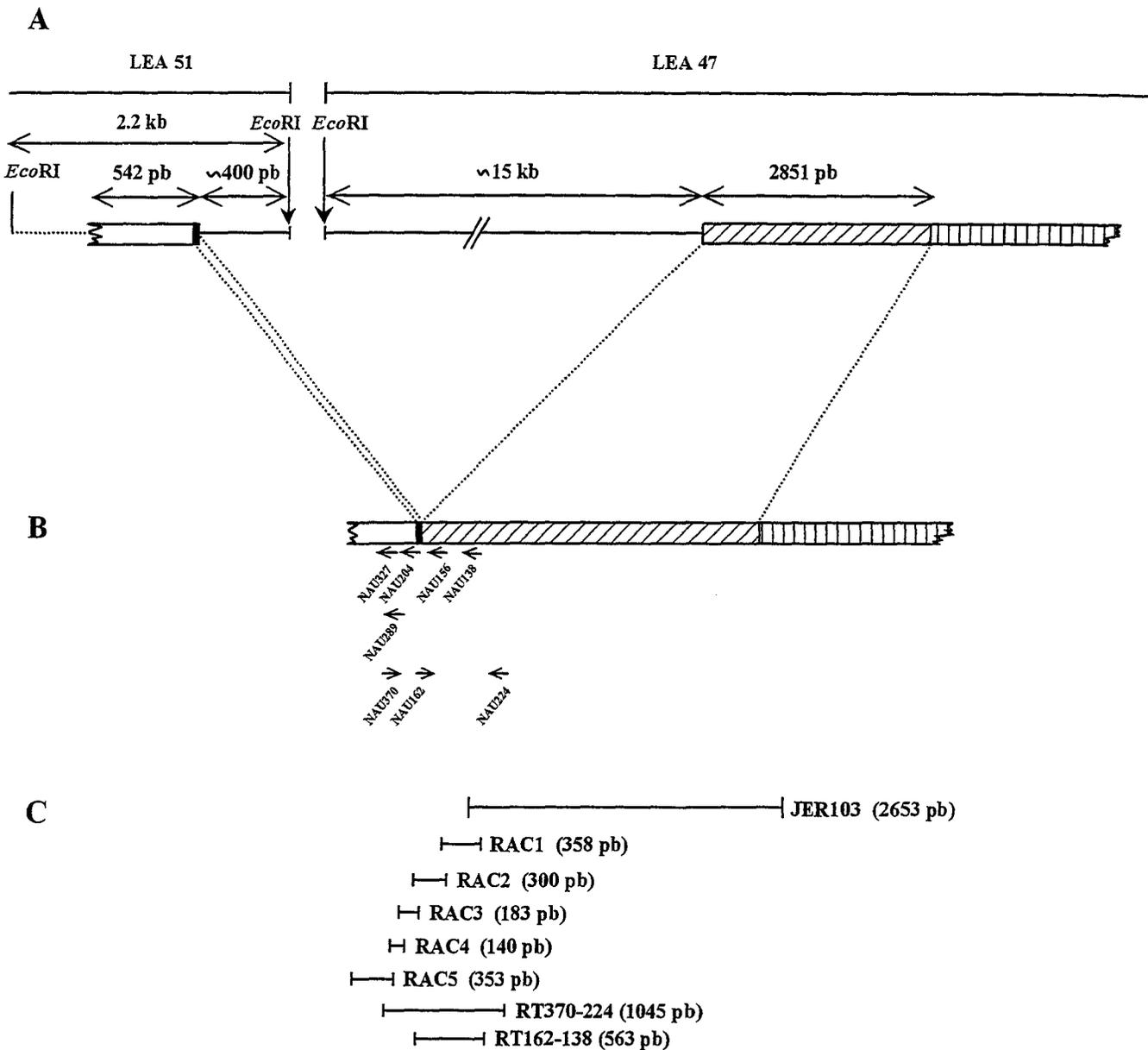


Figure 26 : Organisation de la région 5' terminale de MUC4.

A, Représentation schématique de la structure exon-intron de l'extrémité 3' du clone LEA51 et de l'extrémité 5' du clone LEA2. Le rectangle ouvert à gauche du schéma représente la région 5' non traduite, le rectangle noir la région codant le peptide signal, le rectangle hachuré diagonalement la séquence unique codant une région riche en Thr/Ser et le rectangle hachuré verticalement la région 5' terminale du domaine contenant les répétitions de 48 pb. La ligne interrompue représente la région intronique.

B, Illustration de la compilation des séquences des différents clones d'ADNc obtenus. Les oligonucléotides sont indiqués par des flèches horizontales (leur position précise est donnée dans l'appendice technique).

C, Position et longueur du clone JER103 et des différents clones d'ADNc obtenus par les expériences de 5' RACE et de RT-PCR.

```

1  GTACAGCCCCAAGGTGCTCCCTCTGGGGCCCTTTCTTCCCCATTCTTCCCAGCAGCCCA
61  AAGCTCTGGTGGGACAGGGGCAGCCCTGGGGAGGGAGGAGAGGACCCAGGAACCCGGCT
121  AGGAGGGTGGCCACCCATTTCCAGTGTGACCTGTTCCCATTTCCCCATGTCTCCTCCCA
181  TCCCTCCCGCCACTCAGCTCAGGCTGATGAGAAGCAGAGCAACGGGTGTATCGGTGTTTT
241  CTTTCTGGTGGGGTAGTGGGGTGGGGCTGAGGAGAGAAAAGGGTGATTAGCGTGGGGCC
301  CCGCCCTCTTTTGTCTCTTCCCAGGTTCCCTGGCCCTTTCGGAGAAACGCACTTGGTTT
361  GGGCCAGCCGCTGAGGGGACGGGCTCACGTCTGCTCCTCACACTGCAGCTGCTGGGCCG
421  TGGAGCTTCCCCAGGGAGCCAGGGGGACTTTTGCCGCAGCCATGAAGGGGGCACGCTGGA
481  GGAGGGTCCCCTGGGTGTCCCTGAGCTGCCTGTGTCTCTGCCTCCTTCCGCATGTGGTCC
541  CAGGAACCACAGAGGACACATTAATAACTGGAAGTAAAACCTCCTGCCCCAGTCACCTCAA
601  CAGGCTCAACAACAGCGACACTAGAGGGACAATCAACTGCAGCTTCTTCAAGGACCTCTA
661  ATCAGGACATATCAGCTTTCATCTCAGAACCACAGACTAAGAGCACGGAGACCACCAGCA
721  AAGCTCAAACCGACACCCTCACGCAGATGATGACATCAACTCTTTTTTCTTCCCCAAGTG
781  TACACAATGTGATGGAGACTGTTACGCAGGAGACAGCTCCTCCAGATGAAATGACCACAT
841  CATTTCCCTCCAGTGTACCAACACACTCATGATGACATCAAAGACTATAACAATGACAA
901  CCTCCACAGACTCCACTCTTGGAAACACAGAAGAGACATCAACAGCAGGAACTGAAAGTT
961  CTACCCCAAGTGACCTCAGCAGTCTCAATAACAGCTGGACAGGAAGGACAATCACGAACAA
1021  CTTCC

```

NAU357 ←
 NAU376 ←
 RAC5 ↗
 RAC4 ↗
 NAU162 →
 RAC3 ↗
 RAC2 ↗
 RAC1 ↗ NAU138 ←

Figure 27 : Séquence compilée des différents clones d'ADNc obtenus par des expériences de 5' RACE-PCR.

Les flèches blanches indiquent la position 3' terminale des clones RAC. La localisation des oligonucléotides NAU357, NAU376, NAU162 et NAU138 est indiquée par des flèches noires.

obtenue en compilant les séquences des clones RAC3, RAC4 et RAC5 n'est pas contenue dans ces cosmides. Il semblerait donc que cette séquence codante soit située sur l'ADN génomique à **au moins 15 kb** en amont de la séquence RAC1-2 et que le fragment de 15 kb situé à l'extrémité 5' de LEA47 corresponde à un **large intron** (Figure 26). La taille de 15 kb est déduite de la taille globale de l'insert cosmidique ainsi que de la carte de restriction partielle de cet insert.

Par ailleurs, nous avons vérifié par analyse en Southern blot, l'existence et la taille de la séquence unique codant de nombreux résidus Thr et Ser située en amont du domaine répétitif. Trois sites de restriction pour l'enzyme *PstI* permettent de couper cette séquence en deux fragments s'hybridant à la sonde JER103 (les sites *PstI* sont numérotés 1-3 sur les figures 10-12 et 14). Nous avons donc hybridé le Southern blot dont les ADN ont été hydrolysés par *EcoRI* et *PstI* avec la sonde JER103. Les deux fragments *PstI* se sont avérés d'une taille constante de 2 kb et 0,5 kb chez tous les individus testés (Figure 28). La séquence juste en amont du domaine répétitif est donc parfaitement confirmée et ne présente pas de polymorphisme de taille d'un individu à l'autre.

IV.5 - Obtention de nouveaux clones génomiques contenant l'extrémité 5' terminale du gène de mucine humaine *MUC4*

Afin d'obtenir l'extrémité 5' du gène *MUC4*, la **banque génomique en vecteur cosmidique** pWE15 a été criblée au laboratoire avec la **sonde RAC3**. Un nouveau clone appelé **LEA51** a été isolé et sa carte partielle de restriction a été réalisée (Figure 29). L'insert de LEA51 a une taille d'environ 40 kb. Il ne s'hybride pas à la sonde JER64 et ne contient donc pas de répétitions de 48 pb. Un **fragment clivé par *EcoRI* de 2,2 kb** situé à l'**extrémité 3'** de ce clone s'hybride à la sonde RAC3. Ce fragment a été sous-cloné en vecteur pKS et partiellement séquencé. Sa partie 3' consiste en une **région intronique d'environ 400 pb** (Figure 26). Les **542 pb** en amont de cette région intronique présentent **100 % de similarité** de séquence avec les 542 pb de séquence codante obtenues par compilation des **ADNc RAC3-5**.

Nous avons déduit la **jonction exon-intron** par comparaison des séquences génomiques et des séquences d'ADNc. Les séquences des sites donneur et accepteur d'épissage sont eux aussi conformes à la **règle "GT-AG"** proposée par Mount [Mount 1982] :

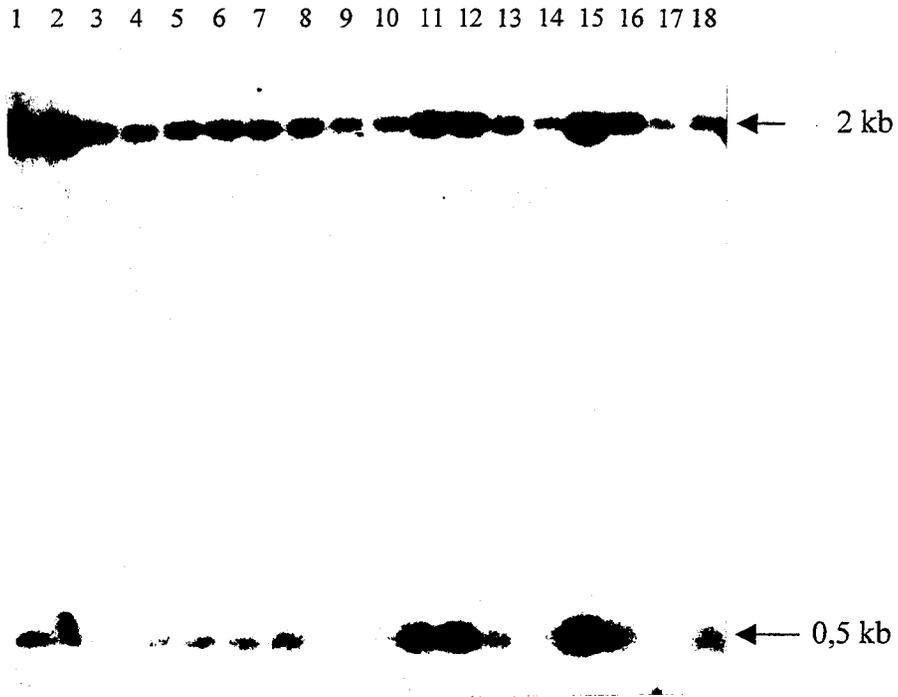


Figure 28 : La séquence unique riche en Thr/Ser en amont du domaine répétitif ne présente pas de polymorphisme de type VNTR.

L'ADN génomique de 18 individus a été hydrolysé par *Pst*I et *Eco*RI. Après séparation électrophorétique en gel d'agarose, l'analyse du Southern blot a été faite à l'aide de la sonde JER103.

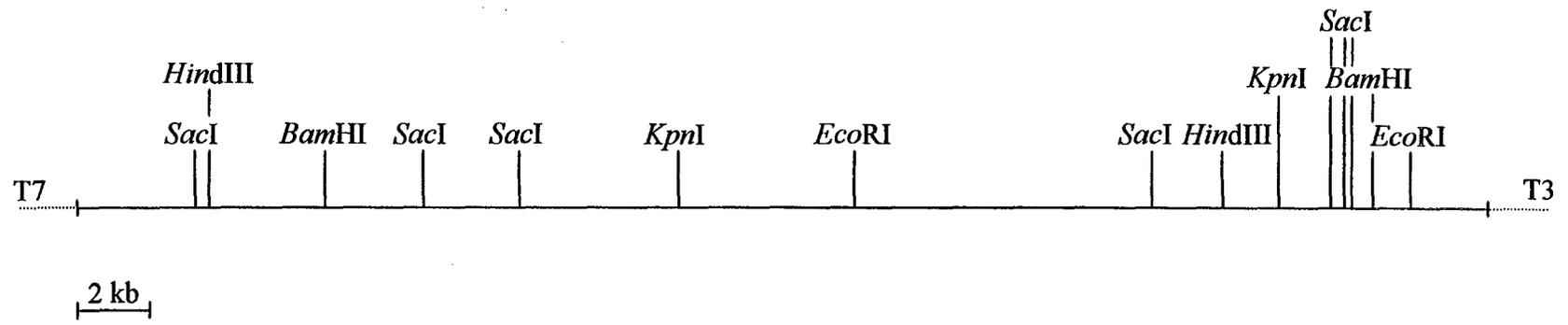


Figure 29 : Cartographie partielle de l'insert du clone cosmique LEA51.

consensus du site donneur : C/A A G **g t** a/g a g t
 site donneur de l'intron : C A G **g t** a a g t
 consensus du site accepteur : c/t n c/t **a g**
 site accepteur de l'intron : t t t g g t t c a c t c c **a g**

Cet intron est également de **classe 1**.

Le clone LEA51 contenant dans sa partie 3' terminale la partie 5' terminale du gène *MUC4*, il semble donc tout à fait probable qu'il contienne également les régions de régulation transcriptionnelle de ce gène. En parallèle du sous-clonage des différents fragments de LEA51, nous avons utilisé le kit "GenomeWalker" (cf. appendice technique) afin d'obtenir de nouveaux clones génomiques également susceptibles de contenir le promoteur du gène *MUC4*. Le kit contient cinq banques constituées d'ADN génomique hydrolysé par cinq enzymes différentes : *EcoRV* ; *ScaI* ; *DraI* ; *PvuII* ; *SspI*. Un adaptateur a été ligué aux 2 extrémités des différents fragments. Une première réaction d'amplification est effectuée à partir d'une amorce s'hybridant à l'adaptateur (AP1) et de l'oligonucléotide NAU376 (Figure 27 ; sa position est donnée dans l'appendice technique). Une deuxième réaction d'amplification est effectuée à partir de deux amorces plus internes : une amorce s'hybridant à l'adaptateur (AP2) et l'oligonucléotide NAU357 (Figure 27). Cette expérience nous a permis d'amplifier et de cloner un nouveau fragment de 2,8 kb dans la banque génomique dont les fragments ont été obtenus après hydrolyse par *PvuII* (Figure 30). Ce fragment de 2,8 kb contient un site de restriction pour l'enzyme *EcoRI* qui le clive en 2 sous-fragments de 1,5 kb et 1,3 kb. Ce site est le même que celui situé à l'extrémité 3' de LEA51. Ces sous-fragments ont été sous-clonés et séquencés. Un deuxième clone de 650 pb (provenant de la banque génomique hydrolysée par *EcoRV*) a été obtenu à l'aide des oligonucléotides NAU598 et NAU599 choisis dans la région 5' terminale du clone précédent. La séquence des deux clones ainsi que les 460 pb comprises entre NAU357 et l'ATG potentiel sont compilées figure 31.

Nous avons recherché dans ces séquences les régions potentiellement régulatrices à l'aide du programme PC/Gene et du programme MatInspector V2.2 (disponible sur le serveur Transfac à l'adresse Internet : http://transfac.gbf.de/cgi-bin/matSearch/matsearch_pub.pl). On observe **4 boîtes TATA** dont deux, notées sur la figure 31, sont corrélées avec un signal de coiffe placé en aval. Il existe également **une boîte CCAAT**, **quatre boîtes GC**, **une boîte SRE** ("Serum Responsiv Element"), **un site de fixation pour HNF-1** ("Hepatic Nuclear Factor-1"), **un site de fixation pour USF** ("Upstream Stimulating Factor") ainsi qu'**un site de fixation pour Sp1**. Ces éléments permettent de penser que la région contenant le promoteur de *MUC4* a été effectivement identifiée.

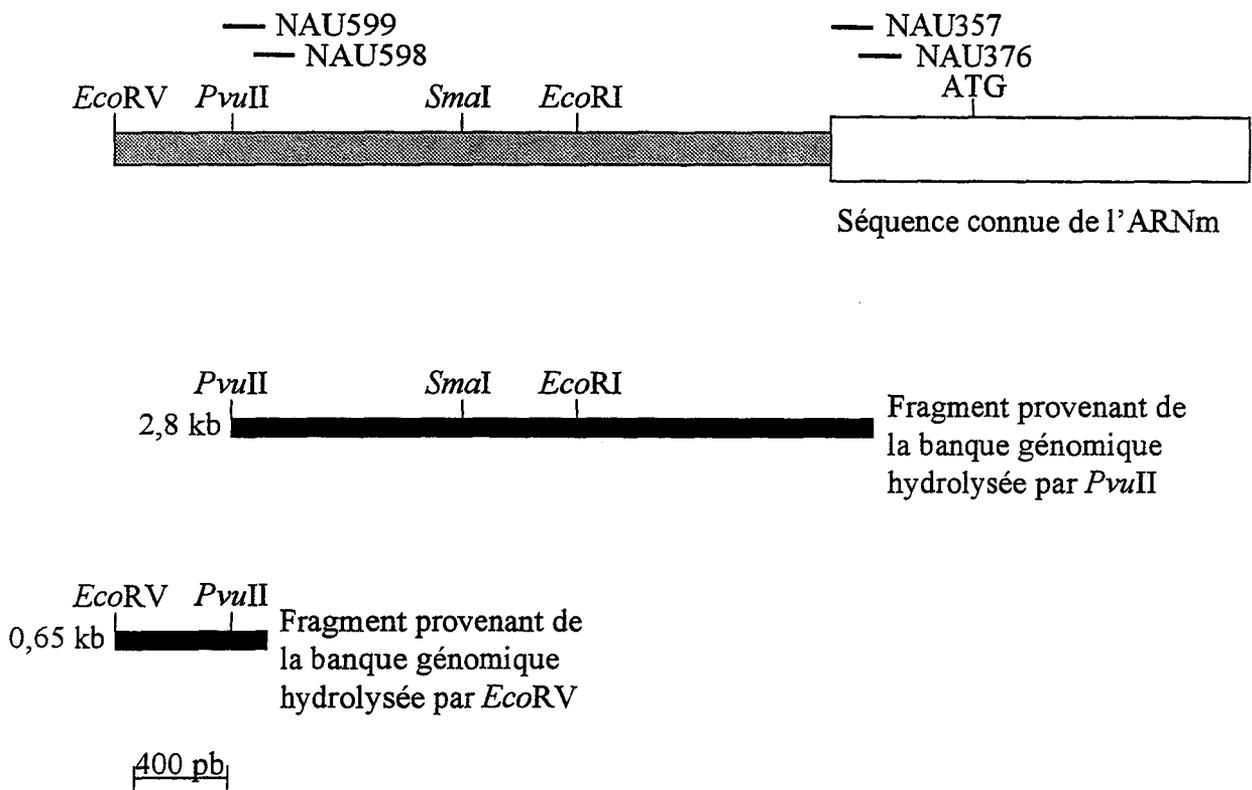


Figure 30 : Obtention de la région génomique en amont de la séquence connue de l'ARNm.

1 ACTATAGGGCACGCGTGGTTCGACGGCCGGGCTGTGGTCTGTGGGCGGTCGGCCAGGTGATTGCCCTTCTTTGCTTTGGAG
81 CCAGGCAGAACAGAGCTCAAATTCAGCTTTGGGACAGTGTGTGGCCAGAAAGGCTGTAAGCTGGGAOTTAACACTCTCC
161 AGTTTCCCTAGATGGTCCGACCTGGCCGGTACTCGGGGATGATGAAGCTCCTGGAGCCTCAGCGTCTCACTTGAGAGG
241 TAGGGCAGCGCTTTGTACTTTCACAGGGCCATGGGACAGGCTCAGGGCCTGGCACTCTAGTGACCGGCTTCATGCTGGGT
321 CACTGACTAGGGAATCCACACACCTTTTTAGTCCAGACTGAGTAGGTAATCAGACTTAATAAACATCCAGCTTCTGCT
401 CGAGCTCAGCCTAGTTGTCTAAGTCTGACAAGATTGAACTTTATTCACAGCACGAGGATAAATGATATAAAGGTAAGC
481 AATTTAGGGGACCCAAAAACTAGGGACTCTACTTGGGTAGCTTAGAAGGGGAATCTCTTGGAGATCCTGGCTTTG
561 CTGCTCAAACCCAGCTGTCTTCAAGCCAAGTCTGGGAGAATCTCATGGATCTTATCCTTGTCCAAATTTACTTACAC
641 CCCACATCCACACTAAGCTCTTCCTGCCCTCCCGTGGAAATGAACTTACACCTGACTTCAGACTCCTGCCCTATGCAGA
721 CCCAGCAGTATGCGGGCCTAGGAGGTTTTTCGGCCACAAGGAAGTAAAGAGGCACTCTTTTGTCCCTCTCATTACCC
801 CCATCTGTCTCCCATCGGATGCCTTGGGAGGAGAGAAGGAAGGACAAATGGGTGTCCCGTGGCCCAACCCAGATGAAGAC
881 AGAGCCATTTATCAGAGACCCAGACGCATCCTAACTCTGGAAAAATGGCATATTGAGGGAGCTGGAAGACAGAAAGG
961 AAACGGGCTGGGCAGGGAAGAGGTGAACAGGAGTCCGACCGCCAGTCTCCAGGATCTCAGCTCTTAATCAGGATCC
1041 TATAAGACTCAAAGGGGACGCGAAAGACCTGGGCTTGGAAAATTCACCTCCGGACTCCTCCTGTATCCTGGCTGGCCCT
1121 CGTTTGTGTCTCCCTCCAGGGATGATGCCCGATGGAAAGGAGAGCCCGTTCTCCACAGGTGTCTCCACAAACAGGTG
1201 TCTCAACAACAGGTGTCTGAACAACAGGTGTCTCACAACAGGTGTCTTAATCTCCCGACGCCCTCAGCATCTACAGG
1281 CTCCTCCACCCCTACAACCTTTAGATAGGTGTTATCCCATTTTACAGGTGAACGAGGAAAACAGCTCGAAGAGGCTG
1361 TGACTTTGTGGTGACCGGTGGCAGAGCTGGGATCAAGCCCGGTTTGTAAAGCCCTCCTGGAGTCTCGGCGGTACAG
1441 CAGCTTCGGGAGAGCGCTGTGTCTCTGGCCTTCTGGCATGGGAAAGAGGTGAGCTACCTGTGTTCGGTCAACAAGAT
1521 GTCCAGCGCTTAGTTTGGTCTCACCTGGATGCTCCTTTCTCCTTAGAGATGAAAACTGTTGCTGAAAGCGTCAACA
1601 AACAAAACCTGTGAAAGGTAATGGGGGACCCACGAGCAAGAGCTGACAGAAAAATGACCAGGGCCGGGTGCAGTGCT
1681 CACGCTGTAATCCAGCACTTTGGGAGGCCAAGGCGGGGATCATCTGAGTTCAGGGGTTGAGACCAACCTGGCCAA
1761 CATGGTAAAACCTGTCCCTACTAAAAATACAAAAAAATTAAGCCGGGCTGGTGGCAGGTGCCTGTAGTCCAGCTACT
1841 AGGGAGGCTGAGCCAGGGGATGGCCGTAACCCGGGAGGCGGAGCCGCAAGTGAGCTGAGATGGTCCACTGCCTCCAG
1921 CCTGGGCGACAGGCAAGACTCCATCTCAAAAAAAAAGTCAAAAATGATCACGAAAGGCCAAGCTGGGCTTGGTGT
2001 TGCGAATTCCTGCAGTAGTCCCTCCTGCGCCCGTCCAGGCGCCACAGGATGCTCCACATGACTCTTTACCTTTATG
2081 ATTTCTTACCTCCAGGAAAGAACAGGGTCAAGCTGTCCACCCAGAGCTGGAGGGACTGTTGGGGGACCACGCCACTTCC
2161 CCACGGTATCCGCATTATCATGCGGTGGCCAGGCTTTGGCCTGGGCCATGAACAGACTTTCAGCTTCTAGAAGCAAGTG
2241 GTCACAGGGCCTCCGTGACCCCTCGGAGACAGCGCCACAGCGTGGTTGCTACTCCTTGGTGGACAAGGCTGGCTCCGCT
2321 TGGGAATGCAGAAACCTGAGTCTTGTCTGGACCCACAGTGTCTGAGCATCGAAGGTCCTGCTGGGACGAGTCA
2401 CCCCACAGACCCCCACATTCGGAATCAAGTCTTGGGAATCAGAAGCAGAGATTACACTGTGGCAAAGAGGGGAGCG
2481 GAGGTATCCCCGTTTCTCTTGCAGAGAGCTACCCAACCTGCCCTCTTCTGCTTTACAACCCACTTGGGGTCCCTTTCC
2561 CTCTCGCTTGTGCTGTGAGTTCAGAGGTTGGAACTGGGGTGGAGCAGAGCCCCAGGTCTCCAGGAGCAGCTCCACC
2641 TCTCAGAGTTACAGTCTCTCTCTACCGAGTGTTCCTTCAACCCACTGCCCGTCCCGGCTCCCTCAGGCCAGTCT
2721 TCCCAGACTGTCCCTGGCTCCCTGCACCTGGCTCCAGCTCCCTGCCCTGGAAGAAGAAGTCCACCATCTGGTGAGGGT
2801 CTGCCACGCATCTGGGGTCTGCATGCGTCTGCACAGAGGTTCTGCCACGAGTCTGCACTGAGGTTCTGCCATGCGTCTG
2881 GAGTCTGCCACGGCTCTGGGGTCTGCCACGCGTCTGGAGTCTGCCATGGGTCTGGGGTCTGCCATGTCTGGGATCTGC
2961 CGTGCCTCTGGGGTCTGCCACGCATCTGCACAGGTTCTGCCACGCGTCTGGGGTCTGCCATGCGTCTGCACAGGTT
3041 CTGCCACGCGTCTGGGGTCCGCATGCGTCTGGGGTCTGCCACGCGTCTGGGGTCTGCCACGCACCTGCACAGGAT
3121 ATTCCAAGCACACCACCTTTCAAAGACGGCATAAAGTGCATTTCTGTTCTGCCCTCACCCACATGTTTCCCCG
3201 TTCCAATTCCTTGGGGTCTGCTTAAGTCTCACTCTCTTTCCCCATTATACAGCCCAAGGTCGCTCCCTCTGGGGCC
3281 CTTTCTTCCCCATTCTTCCAGCAGCCCAAAGCTCTGGTGGGACAGGGGACGCCCTGGGGAGGGAGGAGGACCCAGG
3361 AACCCGGCTAGGAGGGTGGCCCAACCAATTCAGTGTGACCTGTTCCCAATCCCCATGTCTCTCCATCCCTCCCCG
3441 ACTCAGCTCAGGCTGATGAGAAAGCAGAGCAACGGGTGATCGGTGTTTTCTTCTGGTGGGGTAGTGGGGTGGGGCTGA
3521 GGAGAGAAAAGGGTATTAGCGTGGGGCCCGCCCTCTTTGTCTCTTCCAGGTTCCCTGGCCCTTCGGAGAAACGC
3601 ACTTGGTTCGGCCAGCCGCTGAGGGGACGGGCTCAGTCTGCTCCTCACACTGCAGCTGCTGGGCCGTGGAGCTTCCC
3681 CAGGGAGCCAGGGGACTTTTGGCCAGCCATG

Figure 31 : Séquence des clones génomiques en amont de l'ATG potentiel.

La position des oligonucléotides utilisés pour l'obtention de ces clones est indiquée par des flèches noires. Les différents éléments de régulation potentiels sont encadrés. L'ATG potentiel est souligné deux fois.

V - ANALYSE DES SEQUENCES NUCLEOTIDIQUE ET PEPTIDIQUE DEDUITE DE LA PARTIE 5' DE *MUC4*

La compilation des séquences des différents ADNc partiels obtenus successivement nous a permis de déterminer la **totalité de la séquence codante de la partie 5' de l'ADNc de *MUC4* et sa liaison avec la région répétitive centrale de 48 pb**. Cette séquence se trouve dans la banque internationale EMBL sous le numéro d'accès AJ000281. Une région 5' non traduite de 460 pb est suivie d'une région codante dont le cadre de lecture ouvert sur 978 résidus d'aa est suivi de celui de la région centrale caractérisée par les répétitions en tandem de 16 aa (Figure 32).

Un résidu Met situé au nt 1 sur la figure 32 est compris dans un **contexte optimal pour l'initiation de la traduction** tel qu'il a été décrit par **Kozak** [Kozak 1987] :

séquence consensus de Kozak :

G C C G C C A / G C C A T G G

séquence bordant le codon d'initiation putatif de *MUC4* : G C C G C A G C C A T G A.

La courbe d'hydrophobie [Kyte et Doolittle 1982] des 200 premiers résidus de la séquence peptidique déduite montre que les 27 résidus N-terminaux codés par l'exon 1 sont très hydrophobes, suggérant ainsi qu'ils constituent le **peptide signal potentiel** (Figure 33). Trois sites potentiels de clivage existent entre les aa 23 et 24, 27 et 28 et 28 et 29. La séquence a été comparée à celles contenues dans la banque de données GenBank. Il existe un haut degré de similarités entre les peptides signaux de *MUC4* et *ASGP-1* de rat [Wu *et al.* 1994], de même qu'entre ceux de *MUC1* [Gendler *et al.* 1990] et *Muc1* de souris [Spicer *et al.* 1991], et *MUC2* [Gum *et al.* 1994] et *Muc2* de rat [Ohmori *et al.* 1994] (Figure 34). Une **similarité de 62 % entre *ASGP-1* de rat et *MUC4*** est observée au niveau nucléotidique. Au niveau protéique, la similarité est de 59 %. Cette similarité entre les deux peptides signaux est particulièrement frappante si on considère la région C-terminale de ces peptides signaux : 12 résidus sur les 15 sont parfaitement conservés, soit une similarité de 80 %.

La région comprise entre le nt 83 et le nt 2 934 ainsi que le domaine répétitif de type JER64, forment **un seul grand exon**. Cette région (nt 83/nt 2 934) code une **séquence unique** de 951 résidus typique des apomucines, comprenant 21,9 % de Thr et 19,8 % Ser. La

1681	CCA AAA ACA ACA GGG GCA GGC GCC CAG ACA CAA TGG ACA CAA GAA ACG GGG ACC ACT GGA	580
	P K T T G A G A Q T Q W T Q E T G T T G	
1741	GAG GCT CTT CTC AGC AGC CCA AGC TAC AGT GTG ACT CAG ATG ATA AAA ACG GCC ACA TCC	600
	E A L L S S P S Y S V T Q M I K T A T S	
1801	CCA TCT TCT TCA CCT ATG CTG GAT AGA CAC ACA TCA CAA CAA ATT ACA ACG GCA CCA TCA	620
	P S S S P M L D R H T S Q Q I T T A P S	
1861	ACA AAT CAT TCA ACA ATA CAT TCC ACA AGC ACC TCT CCT CAG GAA TCA CCA GCT GIT TCC	640
	T N H S T I H S T S T S P Q E S P A V S	
1921	CAA AGG GGT CAC ACT CAA GCC CCG CAG ACC ACA CAA GAA TCA CAA ACC ACG AGG TCC GTC	660
	Q R G H T Q A P Q T T Q E S Q T T R S V	
1981	TCC CCC ATG ACT GAC ACC AAG ACA GTC ACC ACC CCA GGT TCT TCC TTC ACA GCC AGT GGG	680
	S P M T D T K T V T T P G S S F T A S G	
2041	CAC TCG CCC TCA GAA ATT GTT CCT CAG GAC GCA CCC ACC ATA AGT GCA GCA ACA ACC TTT	700
	H S P S E I V P Q D A P T I S A A T T F	
2101	GCC CCA GCT CCC ACC GGG GAT GGT GAC ACA ACC CAG GCC CCG ACC ACA GCA CTG CAG GCA	720
	A P A P T G D C H T T Q Q A P T A L Q A	
2161	ACA CCC AGC AGC CAT GAT GCC ACC CTG GGG CCC TCA GGA GGC ACG TCA CTT TCC AAA ACA	740
	T P S S H D A T L G P S G G T S L S K T	
2221	GGT GCC CTT ACT CTG GCC AAC TCT GTA GTG TCA ACA CCA GGG GGC CCA GAA GGA CAA TGG	760
	G A L T L A N S V V S T P G G P E G Q W	
2281	ACA TCA GCC TCT GCC AGC ACC TCA CCT GAC ACA GCA GCA GCC ATG ACC CAT ACC CAC CAG	780
	T S A S A S T S P D T A A A M T H T H Q	
2341	GCT GAG AGC ACA GAG GCC TCT GGA CAA ACA CAG ACC AGC GAA CCG GCC TCC TCA GGG TCA	800
	A E S T E A S G Q T Q T S E P A S S G S	
2401	CGA ACC ACC TCA GCG GGC ACA GCT ACC CCT TCC TCA TCC GGG GCG AGT GGC ACA ACA CCT	820
	R T T S A G T A T P S S S G A S G T T P	
2461	TCA GGA AGC GAA GGA ATA TCC ACC TCA GGA GAG ACG ACA AGG TTT TCA TCA AAC CCC TCC	840
	S G S E G I S T S G E T T R F S S N P S	
2521	AGG GAC AGT CAC ACA ACC CAG TCA ACA ACC GAA TTG CTG TCC GCC TCA GCC AGT CAT GGT	860
	R D S H T T Q S T T E L L S A S A S H G	
2581	GCC ATC CCA GTA AGC ACA GGA ATG GCG TCT TCG ATC GTC CCC GGC ACC TTT CAT CCC ACC	880
	A I P V S T G M A S S I V P G T F H P T	
2641	CTC TCT GAG GCC TCC ACT GCA GGG AGA CCG ACA GGA CAG TCA AGC CCA ACT TCT CCC AGT	900
	L S E A S T A G R P T G Q S S P T S P S	
2701	GCC TCT CCT CAG GAG ACA GCC GCC ATT TCC CGG ATG GCC CAG ACT CAG AGG ACA AGA ACC	920
	A S P Q E T A A I S R M A Q T Q R T R T	
2761	AGC AGA GGG TCT GAC ACT ATC AGC CTG GCG TCC CAG GCA ACC GAC ACC TTC TCA ACA GTC	940
	S R G S D T I S L A S Q A T D T F S T V	
2821	CCA CCC ACA CCT CCA TCG ATC ACA TCC AGT GGG CTT ACA TCT CCA CAA ACC CAG ACC CAC	960
	P P T P P S I T S S G L T S P Q T Q T H	
2881	ACT CTG TCA CCT TCA GGA TCT GGT AAA ACC TTC ACC ACG GCC CTC ATC AGC AAC GCC ACC	980
	T L S P S G S G K T F T T A L I S N A T	
2941	CCT CTT CCT GTC ACC GAC ACT TCC TCA GCA TCC ACA GGT CAC H GCC ACC CCT CTT GCT GTC	1000
	P L P V T D T S S A S T G H A T P L A V	
3001	AGC AGT GCT ACC TCA GCT TCC ACA GTA TCC TCG GAC TCC CCT CTG AAG ATG GAA ACA TCA	1020
	S S A T S A S T V S S D S P L K M E T S	
3061	GGA ATG ACA ACA CCG TCA CTG AAG ACA GAC GGT GGG AGA CGC ACA GCC A	1036
	G M T T P S L K T D G G R R T A	

Figure 32 : Séquences nucléotidiques compilées et séquences peptidiques déduites des clones d'ADNc RAC1-5, JER103 et RT174-99.

La position des nucléotides est indiquée par les nombres de gauche et celle des acides aminés par ceux de droite. La position des oligonucléotides NAU370 et NAU224 est indiquée par les flèches noires. La séquence du peptide signal potentiel est soulignée. Les astérisques indiquent la position des résidus de cystéine (lignes 20 et 260). Les triangles noirs représentent la position des introns. Les flèches blanches délimitent les trois unités répétitives trouvées au début de la région riche en Thr/Ser. Trois sites potentiels de *N*-glycosylation sont encadrés. La séquence est présentée avec une seule unité répétitive de 48 pb. Le nombre de répétitions (n) varie d'environ 145 (allèle N) à environ 395 (allèle A) dans cette étude.

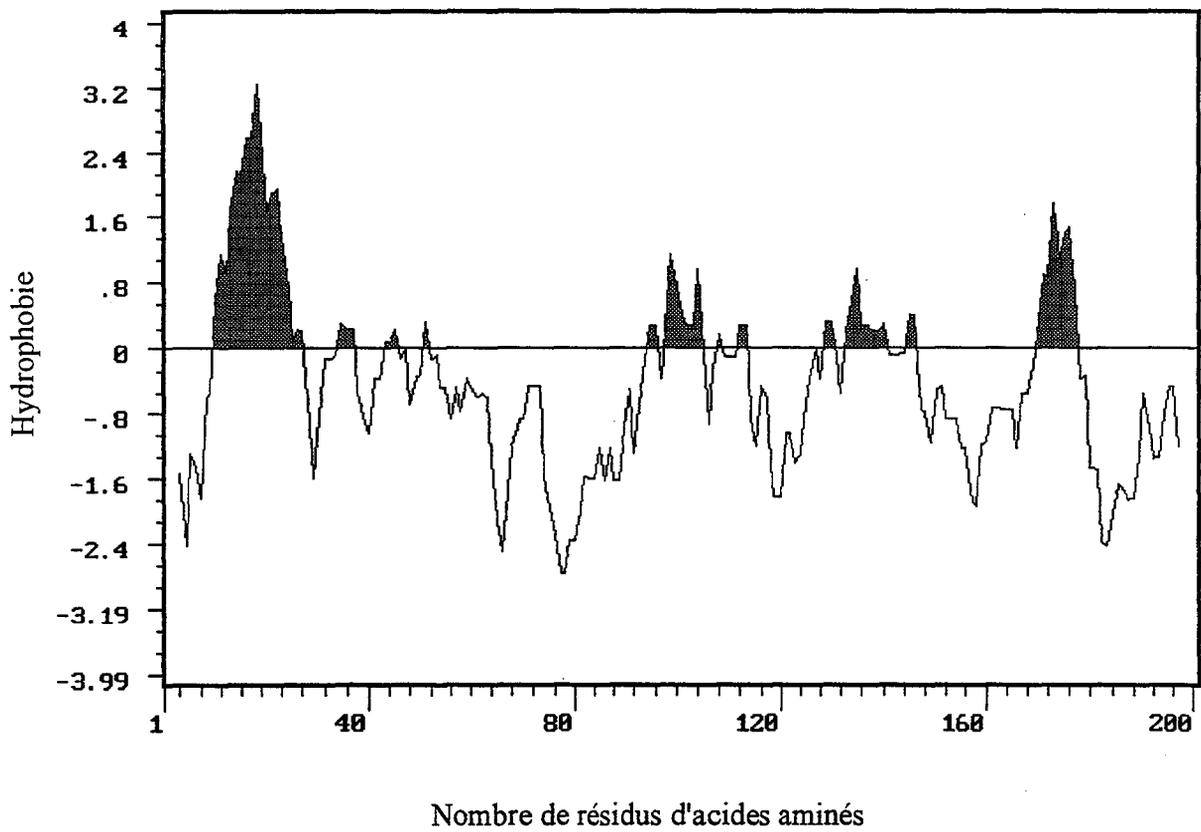


Figure 33 : Courbe d'hydrophobie des 200 premiers résidus de MUC4 [Kyte et Doolittle 1982].

A

MUC1	M T P G T Q S P F F L L L L L T V L T V V	21
Muc1 de souris	M T P G I R A P F F L L L L L A S L K G F	21
MUC2	M G L P L A R L A A V C L A L S L A G G	20
Muc2 de rat	M G L P L A R L V A V C L V L A L A K G	20
MUC4	M K - - - G A R W R R V P W V S L S C L C L C L L P H V V P	27
ASGP-1 de rat	M R G P H G V S W - R V P W L C L S C L C S C L L L L P V N	29

B

MUC4	A T G A A G - - - - - G G G G C A C G C T G G A G G	21
ASGP-1 de rat	A T G C G A G G G C C T C A T G G A G T C T C A T G G - - -	27
MUC4	A G G G T C C C C T G G G T G T C C C T G A G C T G C C T G	51
ASGP-1 de rat	A G A G T T C C C T G G C T G T G T C T G A G C T G C C T G	57
MUC4	T G T C T C T G C C T C C T T C C G C A T G T G G T C C C A	81
ASGP-1 de rat	T G T T C T T G C C T C C T C C T G C T A C C T G T A A A T	87

Figure 34 : Comparaison entre différents peptides signaux de mucines.

A, Alignement des séquences peptidiques déduites du peptide signal de MUC1 avec celui de son homologue de souris, du peptide signal de MUC2 avec celui de son homologue de rat, et du peptide signal de MUC4 avec celui de l'ASGP-1 de rat. Les tirets représentent les intervalles introduits dans les séquences afin de permettre un alignement optimum. Les acides aminés identiques sont encadrés. Les nombres sur la droite indiquent la position des acides aminés.

B, Similarités entre les séquences nucléotidiques du peptide signal de MUC4 et celui de l'ASGP-1 de rat. Les nucléotides identiques sont encadrés. Les nombres sur la droite indiquent la position des nucléotides.

séquence TXXP, qui est considérée comme le site majeur de *O*-glycosylation [Elhammer *et al.* 1993], est retrouvée 21 fois. Cette région est également riche en Pro (7,5 %). Trois sites potentiels de *N*-glycosylation sont également trouvés dans cette séquence aux positions 235, 260 et 622. Cette séquence ne comporte qu'un seul résidu Cys, en position 256. La séquence de cette région unique riche en Thr/Ser ne présente pas de similarités significatives avec les répétitions de 48 pb de JER64. Un examen plus approfondi a révélé l'existence dans cette séquence de trois sous-régions similaires entre elles. Ces trois unités répétées commencent au résidu 43. Les deux premières répétitions ont une longueur de 126 résidus, la troisième a une longueur de 130 résidus. Ces sous-régions sont typiques des mucines : elles contiennent 24 % de Thr et 20 % de Ser.

Il existe, dans la séquence unique de *MUC4*, des régions présentant des similarités au niveau nucléique avec la séquence de l'*ASGP-I*. La région comprise entre les nucléotides 111 et 394, comprenant une grande partie de la répétition n°1, présente 58 % de similarités avec la région s'étendant du nucléotide 777 au nucléotide 1 062 de l'*ASGP-I* [Wu *et al.* 1994]. Cette région correspond à une partie de la deuxième répétition de l'*ASGP-I*. La région comprise entre les nucléotides 290 et 319 de *MUC4* (répétition n°1) est 83 % similaire à celle comprise entre les nucléotides 4 193 et 4 222 de l'*ASGP-I* (douzième répétition). Enfin, la région comprise entre les nucléotides 588 et 699 de *MUC4* (répétition n°2) est 68 % similaire à celle comprise entre les nucléotides 4 776 et 4 887 de l'*ASGP-I* (treizième répétition).

La séquence de la figure 32 ne correspond pas exactement à celle que nous avons précédemment publiée [Nollet, Moniaux *et al.* 1998]. En effet, dans le but de choisir des amorces en dehors du domaine répétitif afin de pouvoir quantifier l'expression de *MUC4* au niveau de l'endocol, Gipson *et al.* ont cloné et séquencé 2 679 pb en amont du domaine répétitif [Gipson *et al.* 1999]. Leur séquence publiée ne comporte ni la région 5' non traduite ni la région codant les 86 premiers aa. On observe deux différences entre leur séquence et la nôtre : une guanidine présente au nt 2 436 de notre séquence n'est pas retrouvée dans leur séquence qui contient par contre une cytosine en plus au niveau du nt 2 515 de notre séquence. La séquence peptidique déduite publiée très récemment par Gipson diffère donc de la nôtre pour 26 aa. Après vérification de cette partie de la séquence, il s'est effectivement avéré que la bonne séquence était celle de Gipson. La séquence que nous avons communiquée à la banque internationale EMBL a immédiatement été corrigée.

VI - VERIFICATION DU ROLE POTENTIEL DU PEPTIDE SIGNAL

Afin de confirmer le rôle potentiel de peptide signal de la séquence peptidique identifiée, le clone RT370-224 (1 045 pb) (Figure 26) a été utilisé comme matrice pour réaliser des expériences de transcription et traduction *in vitro*, en présence ou en absence de microsomes (cf. appendice technique). Le but poursuivi est de comparer la taille des produits de traduction obtenus en l'absence de microsomes (peptide non mature) ou en présence de microsomes (peptide mature : clivage du peptide signal, glycosylation). Les peptides obtenus ont été étudiés par électrophorèse (Figure 35).

En absence de microsomes (colonne 3), un produit unique de traduction non mature, correspondant à 324 résidus d'aa d'après les calculs, apparaît sous une masse moléculaire apparente de 40 kDa. En présence de microsomes (colonne 1), trois peptides sont présents, de masses moléculaires apparentes 43 kDa, 41 kDa et 39 kDa. Le plus petit fragment de 39 kDa obtenu en présence de microsomes est de taille significativement plus basse (colonne 1) que celui du produit de traduction non mature (40 kDa) obtenu en l'absence de microsomes (colonne 3). Ceci indique qu'en présence de membranes, un clivage du produit de traduction a été réalisé par une signal peptidase. Ceci confirme donc le rôle putatif de peptide signal de la séquence identifiée dans le clone RT370-224.

Les peptides de masse moléculaire plus élevée doivent correspondre à des peptides matures *N*-glycosylés. Pour vérifier cette hypothèse, nous avons soumis ces produits de traduction à l'action de l'endoH (endo- β -N-acétylglucosaminidase H EC 3.2.1.96). Une réduction de taille, confirmant la présence de *N*-glycannes, a été observée (colonne 2) compatible avec la possibilité que :

- le peptide de 43 kDa correspond à un peptide portant deux chaînes *N*-glycanniques mais amputé du peptide signal ;
- le peptide de 41 kDa correspond lui aussi à un peptide amputé du peptide signal mais porteur d'une seule chaîne *N*-glycannique ;
- le peptide de 39 kDa correspond au peptide traduit non *N*-glycosylé et amputé du peptide signal.

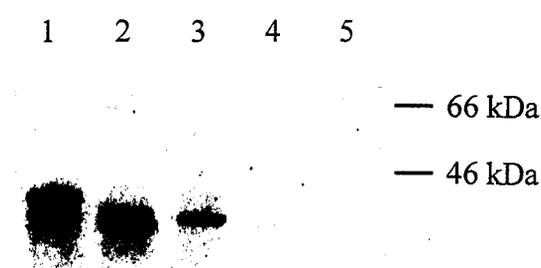


Figure 35 : Marquage à la Met [^{35}S] du produit d'expression du clone RT370-224 synthétisé à l'aide d'un système acellulaire en absence ou en présence de microsomes et d'endo H.

Les réactions de transcription et de traduction ont été effectuées à l'aide du kit "TNT[®] Coupled Reticulocyte Lysate Systems" (Promega). Les produits de la réaction ont été traités comme indiqué dans l'appendice technique.

Colonne 1, incubation avec 1 μg de RT370-224 et microsomes.

Colonne 2, incubation avec 1 μg de RT370-224, microsomes et endo H.

Colonne 3, incubation avec 1 μg de RT370-224.

Colonne 4, incubation avec 1 μg de vecteur pCR2.1.

Colonne 5, incubation avec 1 μg de vecteur pCR2.1 et microsomes.

La taille des marqueurs protéiques colorés Rainbow[™] (Amersham) est indiquée sur la droite.

DISCUSSION

Lorsque le travail d'étude de l'organisation génomique de *MUC4* m'a été confié, nous avions à notre disposition au laboratoire :

- une sonde d'ADNc (JER64) contenant le domaine répétitif spécifique de 48 pb de *MUC4* [Porchet *et al.* 1991] ; aucune autre séquence n'était connue ;
- une connaissance de l'expression tissulaire de *MUC4* nous permettant de choisir les tissus à partir desquels nous allions préparer les banques d'expression et réaliser les expériences de RT-PCR ;
- une connaissance limitée des enzymes de restriction délimitant le domaine répétitif de *MUC4*, laissant cependant pressentir un possible polymorphisme interindividuel [Gross *et al.* 1992].

La première partie de notre étude a comporté un long travail de criblage de banques génomiques destiné, à partir du domaine répétitif de *MUC4*, à reconstituer le gène entier.

I - LE DOMAINE REPETITIF DE *MUC4*

I.1 - Clonage du domaine répétitif de *MUC4*

La première partie de ce travail a été consacrée au criblage tour à tour de deux banques génomiques, l'une construite en vecteur bactériophagique, l'autre en vecteur cosmique, afin d'obtenir les premiers recombinants contenant le gène *MUC4*. Nous nous sommes heurtés à de gros problèmes liés à l'instabilité des recombinants, plus particulièrement avec les clones bactériophagiques. Cette instabilité était détectée par un affaiblissement progressif du signal d'hybridation des clones positifs avec la sonde JER64. De même, des modifications sensibles de la carte de restriction des ADN recombinants, et en particulier un raccourcissement net du domaine répétitif reconnu par la sonde, étaient observées au cours des étapes successives d'amplification du titre bactériophagique et au cours des étapes de préparation. Il semble donc que les gènes de mucines, par leur caractère répétitif (VNTR), constituent un "point chaud" de recombinaison qui contribue à l'instabilité des clones. Ces difficultés ont été signalées par

ailleurs à propos de *MUC2* [Toribara *et al.* 1991]. Néanmoins, des **clones génomiques (phagiques et cosmidiques) contenant *MUC4*** ont pu être sélectionnés et étudiés.

I.2 - Le domaine répétitif de *MUC4* est homogène et de grande taille

Nous avons pu montrer que le domaine répétitif de *MUC4* est totalement contenu dans une bande *PstI/EcoRI* de 11 kb dans les clones phagiques et d'environ 18,2 kb dans les clones cosmidiques. La digestion enzymatique de cette bande par *DdeI*, pour laquelle un site de restriction existe dans la plupart des unités élémentaires du domaine répétitif, nous a permis de mettre en évidence que le fragment *PstI/EcoRI* n'est constitué que de répétitions de 48 pb puisque aucun fragment supérieur à 96 pb (correspondant à deux unités répétitives) n'a été observé. La région répétitive des clones cosmidiques contient environ 380 unités répétitives de 48 pb et est **ininterrompue et localisée sur un même exon**. Ces résultats, obtenus par l'étude d'ADN cloné, ont été confirmés par l'étude d'ADN génomique non cloné provenant de divers sujets.

Les gènes de mucines humaines sont caractérisés par un grand domaine répétitif codant un peptide typiquement riche en résidus d'acides aminés hydroxylés. Comme pour *MUC1*, *MUC2* et *MUC7*, la région répétitive de *MUC4* n'est interrompue par aucune séquence unique contrairement aux domaines répétitifs de *MUC5B* et de *MUC5AC* qui sont interrompus plusieurs fois par des sous-domaines riches en cystéine. Dans le cas de *MUC4*, comme dans ceux de *MUC1*, *MUC5B* et *MUC7* pour lesquels l'organisation génomique est connue, il s'avère que la séquence répétitive est toujours contenue dans un seul exon de taille variable : *MUC1* : 3,2 à 5,9 kb, *MUC5B* : 10,7 kb, *MUC7* : 2,2 kb. Il apparaît donc que, en ce qui concerne les **gènes de mucines humaines, le domaine répétitif de *MUC4*** (pour les plus grands allèles) **est le plus grand décrit jusqu'à présent**. Cependant, il faut encore signaler que le domaine répétitif de la **mucine sous-maxillaire de porc** est codé par un exon inhabituellement grand (32,8-34 kb) [Eckhardt *et al.* 1997].

I.3 - Le domaine répétitif de *MUC4* est polymorphique

Gross *et al.* [Gross *et al.* 1992] avaient suggéré que le polymorphisme observé entre les individus par analyse en Southern blot à l'aide de la sonde JER64 était de type VNTR, s'appuyant pour cela sur la structure répétitive de cet ADNc. Cependant, les sites de restriction des enzymes utilisées dans cette étude ne flanquaient pas précisément le domaine répétitif. Nous avons choisi, pour notre étude de polymorphisme, d'utiliser *PstI* et *EcoRI*, ces enzymes

flanquant directement la région répétitive de 48 pb (*Pst*I est situé 272 pb en amont et *Eco*RI 85 pb en aval du domaine répétitif). Il était ainsi possible d'isoler le domaine répétitif en entier et d'étudier ses variations chez 35 individus non apparentés.

Treize allèles différents, d'une taille variant entre 19 et 7 kb, ont pu être détectés sur les 70 chromosomes analysés. Chez ces mêmes sujets, au contraire, les séquences uniques situées en amont du domaine répétitif ont la même longueur. Ces séquences sont également retrouvées à l'identique dans tous les clones génomiques étudiés ou les clones d'ADNc étudiés. En aval du domaine répétitif, il existe une autre région répétitive mais qui s'est révélée être intronique. Cette dernière présente également un polymorphisme de longueur avec les enzymes *Pst*I et *Eco*RI. Un site de restriction pour *Eco*RI est situé entre les deux domaines répétitifs. Or, les allèles observés avec les deux sondes répétitives n'ont jamais la même longueur. Le polymorphisme observé n'est donc pas un polymorphisme de mutation du site de restriction *Eco*RI mais bien un **polymorphisme de type VNTR**.

La taille des allèles observés pour le domaine répétitif codant varie entre 7 et 19 kb et correspond à une variation du nombre de répétitions de 48 pb d'environ 145 à 395 unités. Une variation d'une telle importance a également été mise en évidence pour *MUC1*. Selon les individus, les allèles de *MUC1* contiennent entre 20 et 125 unités répétées en tandem [Gendler *et al.* 1990]. Ceci amène à penser que la taille de l'axe peptidique n'est pas cruciale à la fonction de la molécule [Gendler et Spicer 1995]. Néanmoins, il semble que les individus possédant de petits allèles de *MUC1* aient une plus grande susceptibilité à développer un carcinome gastrique dans la population étudiée par Carvalho *et al.* [Carvalho *et al.* 1997]. Les variations du nombre de répétitions pourraient donc, dans certains cas, avoir une signification fonctionnelle. Cependant, les allèles de *MUC4* sont codominants et presque tous les individus de notre étude possèdent au moins un grand allèle de *MUC4*. Il est donc possible que des individus ne possédant pas au moins un allèle codant une protéine *MUC4* dont le domaine répétitif soit suffisamment grand, aient un défaut de protection des cellules épithéliales et soient ainsi plus exposés à développer certaines pathologies. Des études de polymorphisme concernant des patients atteints de maladie de Crohn sont actuellement en cours au laboratoire.

Le polymorphisme des domaines répétitifs codants des apomucines est une caractéristique intrigante. Le fait que de longues séquences répétitives soient rassemblées en

un seul exon constitue un facteur favorisant les événements de "crossing-over" inégaux et donc le polymorphisme. Il semble donc que, ni la longueur précise de cette région, ni la fidélité de répétition de la séquence de l'unité élémentaire, ne soient des exigences à sa fonctionnalité. Néanmoins, la proportion, plus que le nombre en valeur absolue de thréonine et de sérine codées par l'ensemble de ces domaines, reste constante. Ceci semble indiquer qu'il existe néanmoins des systèmes de maintien de certaines position d'acides aminés (Figure 9).

II - HOR1 : DEUXIEME DOMAINE REPETITIF DE *MUC4*

La région intronique juste en aval du domaine répétitif de *MUC4* présente également un polymorphisme de type VNTR (Figure 36). Deux autres régions introniques répétitives présentant un polymorphisme VNTR ont été mises en évidence dans la partie 3' du gène *MUC4* [Moniaux, 1999a ; Moniaux *et al.* 1999b]. Un polymorphisme de type VNTR a également été observé pour l'intron 36 de *MUC5B* [Desseyn *et al.* 1999].

Des facteurs de transcription peuvent se lier à des séquences introniques présentant un polymorphisme VNTR [Nakamura *et al.* 1998]. Par exemple, des membres de la famille rel/NF- κ B (facteurs de régulation de la transcription) se lient à des séquences VNTR présentes dans la région 3' flanquant le gène *HRAS* [Trepicchio *et al.* 1992]. Ces séquences varient en taille de 1 000 à 2 500 pb (35-90 répétitions) ; plusieurs de ces allèles sont retrouvés fréquemment mais quelques allèles rares existent aussi. La présence d'un ou de deux allèles rares chez un même individu semble être associée avec un plus grand risque de développer des cancers de différents tissus (côlon, sein, vessie) [Krontiris *et al.* 1993]. Le gène de l'insuline contient une séquence VNTR (répétitions de 14 pb) dans sa région 5'. Lucassen *et al.* ont constaté une association entre la longueur du VNTR et une susceptibilité à développer un diabète insulino-dépendant [Lucassen *et al.* 1993]. Un troisième VNTR pouvant influencer la transcription est celui trouvé dans l'intervalle DH-JH de la chaîne lourde des immunoglobulines humaines. La séquence consensus de 50 pb de ce VNTR contient une séquence similaire au motif de fixation du facteur de transcription myc/HLH [Trepicchio *et al.* 1993]. Bien qu'aucun effet sur la transcription n'ait été démontré, Ogilvie *et al.* ont rapporté une association entre un polymorphisme VNTR du second intron du gène du transporteur de la sérotonine et une susceptibilité à une dépression grave [Ogilvie *et al.* 1996]. Meloni *et al.* fournissent une preuve supplémentaire que les polymorphismes de type VNTR peuvent jouer un rôle significatif dans la régulation de la transcription : ils rapportent que des répétitions polymorphiques de 4 nucléotides, localisées dans le premier intron du gène de la

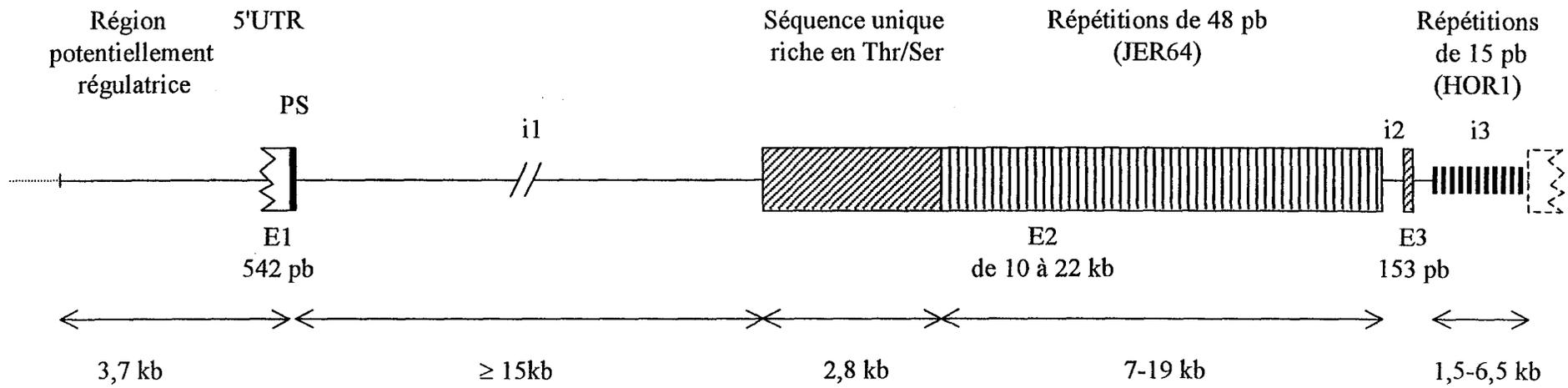


Figure 36 : Représentation schématique de l'organisation génomique de la région 5' ainsi que des deux domaines répétitifs du gène *MUC4*.

tyrosine hydroxylase, jouent le rôle d'élément "enhancer" [Meloni *et al.* 1998]. Ils suggèrent que les protéines de la famille Fos-Jun forment un complexe avec cette séquence répétée.

Il semble donc plausible, au vu de la littérature, que les séquences introniques VNTR des gènes de mucines puissent avoir une action régulatrice. Nous avons utilisé le programme MatInspector V2.2 afin de rechercher dans la séquence de HOR1 des sites de fixation potentiels pour des facteurs de régulation [Quandt *et al.* 1995]. 110 boîtes E2 (une dans chaque unité répétée) ont été trouvées. Ces boîtes E2 (CACCTG) sont des sites de liaison pour différents facteurs d'activation mais sont également la cible d'un facteur de répression : la protéine δ EF1 [Sekido *et al.* 1994]. Cette protéine, surtout exprimée dans le mésoderme et le système nerveux, serait impliquée dans l'embryogenèse [Funahashi *et al.* 1993]. 75 sites de fixation potentiels pour la protéine MZF1 ("myeloid zinc finger") sont également retrouvés dans la séquence répétitive de HOR1. Cette région présentant un polymorphisme VNTR, sa taille et donc le nombre de sites de fixation potentiels pour des facteurs de régulation, est très variable et pourrait ainsi influencer le niveau de transcription de *MUC4*.

Si la région intronique présentant un polymorphisme de type VNTR possède des fonctions régulatrices, nous pouvons également nous demander si la distance entre cette région et le promoteur n'influe pas sur sa fonction de régulation. Or, l'analyse de l'ADN de 35 sujets, soit de 70 chromosomes, révèle l'existence d'au moins 21 associations haplotypiques entre les deux domaines répétitifs (caractérisés par les sondes JER64 et HOR1). Ceci se traduit par une distance variable entre le promoteur et la région supposée régulatrice, bien qu'il soit étonnant de constater qu'un grand allèle de l'un des VNTR soit associé à un petit allèle de l'autre et inversement, ce qui aurait pour finalité de maintenir un écart à peu près constant.

III - LA REGION 5' TERMINALE DE *MUC4*

III.1 - Obtention de la première séquence unique de *MUC4*

La sonde JER64 nous a permis d'isoler un clone d'ADNc de *MUC4*. Ce clone contient une séquence unique située en 5' du domaine répétitif de 48 pb, qui a pu progressivement être complétée par une technique de RACE-PCR. Nous avons ainsi reconstitué la **séquence codante amino-terminale complète de *MUC4***. La comparaison de cette séquence codante

avec la séquence génomique nous a permis de déterminer également l'organisation génomique de la partie 5' de *MUC4*.

III.2 – Organisation génomique de la partie 5' de *MUC4*

De 5' en 3', l'organisation de la partie 5' de *MUC4* est la suivante :

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 sont très hydrophobes et correspondent au **peptide signal** de *MUC4*, comme nous l'avons démontré par des expériences de transcription et de traduction *in vitro* en présence de microsomes pancréatiques de chien. Un **intron d'au moins 15 kb** est situé à la jonction entre le peptide signal potentiel et le peptide *MUC4* mature.

Le **second exon** de *MUC4* code un peptide pouvant être divisé en quatre sous-domaines distincts :

- une première région composée de **trois répétitions imparfaites de 126 résidus** et riche en résidus de thréonine et de sérine,
- une deuxième région de 554 résidus correspondant à une **séquence unique également riche en thréonine et en sérine**,
- un troisième domaine, le plus important en taille, composé de **répétitions en tandem de 16 résidus** homologues aux répétitions de JER64,
- et enfin un quatrième domaine constitué d'une **séquence unique de 26 résidus**.

Le **troisième exon** de *MUC4* code un **peptide de 51 acides aminés** également riche en thréonine et en sérine [Moniaux et al. 1999b], suivi du **troisième intron** contenant les **répétitions de 15 pb de HOR1**.

De ce point de vue, *MUC4* ressemble aux autres gènes de mucines. La région N-terminale de *MUC1* est aussi riche en résidus de sérine, thréonine et proline [Gendler *et al.* 1990]. La région unique en amont du domaine répétitif central de *MUC2*, *MUC5AC* et *MUC5B* est elle aussi riche en acides aminés hydroxylés [Toribara et al. 1991 ; Klomp *et al.* 1995 ; Desseyn *et al.* 1998b]. Par contre, elle contient aussi des domaines riches en cystéine caractéristiques des mucines dont les gènes sont localisés en 11p15.5 et qui sont absents dans *MUC4*. Comme pour *MUC4*, ces régions ainsi que le domaine répétitif central sont codés par le même exon.

III.3 – Premier anticorps dirigé contre une séquence unique de MUC4

Un produit de PCR obtenu à partir de JER103 et d'une taille de 1 184 pb (nt 1 657 – nt 2 841 ; Figure 32) a été cloné afin d'obtenir une protéine de fusion GST-MUC4_{Ag} (70,9 KDa) qui a ensuite été utilisée au laboratoire par Isabelle Van Seuningen pour fabriquer un anticorps polyclonal de lapin. La région contre laquelle est dirigé cet anticorps est externe au domaine répétitif mais demeure encore riche en résidus hydroxylés (34 %). Il est donc possible que les épitopes reconnus soient plus ou moins masqués par la glycosylation. Actuellement, cet outil est utilisé de manière efficace en Western blot sur des extraits cellulaires provenant de la lignée MM39 connue pour exprimer MUC1 et MUC4 [Lo-Guidice *et al.* 1997]. Par contre, des améliorations techniques sont encore nécessaires pour son utilisation en immunohistochimie (déglycosylation préalable nécessaire sur coupes).

III.4 - Peptide signal de MUC4 et similarités avec celui de l'ASGP-1

Les peptides signaux sont de courtes séquences N-terminales nécessaires à la translocation des chaînes polypeptidiques naissantes à travers la membrane du RE [Kreil 1981]. En dépit du fait que ces courtes séquences N-terminales clivées contiennent toujours un regroupement de résidus hydrophobes flanqués de résidus hydrophiles, les séquences primaires en elles-mêmes diffèrent à la fois dans leur longueur et dans leur structure [Haeuptle *et al.* 1989 ; Bird *et al.* 1990 ; Sakaguchi *et al.* 1992]. La comparaison des structures primaires de séquences signal d'une grande variété de protéines de cellules eucaryotes (protéines sécrétées et protéines de membranes) et de protéines d'origine procaryote ne révèle qu'une homologie structurale limitée [Blobel *et al.* 1979].

Néanmoins, les structures primaires des **séquences signal de certaines protéines homologues** de différentes espèces se sont révélées remarquablement **conservées au cours de l'Evolution**. C'est le cas par exemple des protéines du lait (**caséines, β -lactoglobulines et α -lactalbumines**) [Gaye et Mercier 1981]. Les séquences de la région 5' non codante et du peptide signal de différentes protéines de rat et de souris appartenant à la famille des **protéines riches en proline** (protéines contenant, comme les mucines, des répétitions en tandem) présentent un haut degré de similarité (> 94 %) contrairement aux régions N- et C-terminales [Clements *et al.* 1985]. De même, en ce qui concerne la **stathérine humaine et le peptide riche en histidine** (deux protéines des glandes sous-maxillaires jouant un rôle dans la protection non-immunitaire de la cavité buccale et dérivant d'un gène ancestral commun), la

majorité de la séquence codante ne montre pas de similarités significatives, sauf au niveau du peptide signal et de la séquence nucléotidique codant les 14 premiers résidus N-terminaux [Dickinson *et al.* 1987]. La **conservation inhabituelle de la séquence entourant le site d'initiation de la traduction** a laissé supposer que cette séquence contribuerait à la **régulation spécifique et/ou à la stabilité et à la conformation de l'ARNm** [Clements *et al.* 1985].

Ainsi, comme le **peptide signal potentiel de MUC4 présente un haut degré de similarité avec celui de l'ASGP-1**, nous pouvons postuler que les apomucines MUC4 et ASGP-1 de rat dérivent d'un **ancêtre commun**. Dans notre laboratoire, parallèlement à ce travail s'intéressant à la région 5' de *MUC4*, le travail réalisé par Nicolas Moniaux permettait de reconstituer l'intégralité de la séquence 3' [Moniaux *et al.* 1999b]. Ainsi, la partie C-terminale de l'apomucine MUC4 a une longueur de 1 156 résidus d'aa et présente un haut degré de similarité avec le complexe de sialomucines de rat (SMC). SMC est un complexe hétérodimérique composé d'une sous-unité de mucine (ASGP-1) et d'une sous-unité transmembranaire (ASGP-2). La même organisation est retrouvée dans MUC4 où la présence d'un site de clivage protéolytique (Gly Asp Pro His) pourrait permettre de cliver le précurseur MUC4 en 2 sous-unités, MUC4 α et MUC4 β . Comme ASGP-2, qui se lie au récepteur tyrosine kinase p185^{neu}, MUC4 β possède deux domaines de type EGF, une séquence transmembranaire et un site potentiel de phosphorylation.

MUC4, homologue humain de SMC de rat, serait donc une glycoprotéine de surface hétérodimérique et bifonctionnelle de 1 à 2,12 μ m selon la taille du domaine répétitif central. Ces résultats confèrent un nouveau rôle biologique à MUC4 en tant que **ligand potentiel de ErbB2**.

IV - MUC4, RESSEMBLANCES ET DIFFERENCES AVEC LES AUTRES GENES D'APOMUCINES HUMAINES

IV.1 - MUC4 et les gènes d'apomucines localisés en 11p15.5

Les quatre gènes de mucines humaines *MUC2*, *MUC5AC*, *MUC5B* et *MUC6* sont regroupés sur le chromosome 11 en p15.5 [Pigny *et al.* 1996a], alors que les 6 autres gènes sont dispersés à travers le génome humain. Un domaine typique, riche en résidus de cystéine,

semble être caractéristique d'au moins trois (*MUC2*, *MUC5AC* et *MUC5B*) des quatre gènes de mucines localisés en 11p15.5 [Desseyn *et al.* 1997c]. Ces trois mucines feraient partie de la même famille de protéines et dériveraient d'un ancêtre commun selon un schéma évolutif proposé par Jean-Luc Desseyn [Desseyn *et al.* 1998a] (Figure 5).

Aucune des régions riches en résidus de cystéine retrouvées dans les mucines formant le gel de mucus (gènes localisés en 11p15.5) n'a été observée, ni dans la partie N-terminale ni dans la partie C-terminale [Moniaux *et al.* 1999b] de *MUC4*. De plus, **l'expression cellulaire de *MUC4* le distingue des gènes localisés en 11p15.5**. En effet, *MUC4* est exprimé dans des cellules épithéliales autres que les cellules mucosécrétantes. Ainsi, *MUC4* est exprimé dans toutes les cellules de l'appareil respiratoire et dans toutes les cellules de l'intestin chez l'adulte [Audié *et al.* 1993] ou chez le fœtus [Buisine *et al.* 1998b ; 1999].

IV.2 - *MUC4* et les gènes d'apomucines membranaires

Les résultats décrits par Moniaux *et al.* indiquent que *MUC4* peut être classée dans le groupe des mucines transmembranaires au même titre que *MUC1*. Il est donc probable que *MUC4* possède quelques-unes des fonctions attribuées à *MUC1*. *MUC4* en situation membranaire, c'est-à-dire s'élevant de 1 à 2 μm au-delà de la surface apicale au-dessus du glycocalyx, jouerait les mêmes rôles de protection et d'anti-adhérence que ceux décrits à propos de *MUC1* [Hilkens *et al.* 1992]. *MUC4* possède également un site de clivage protéolytique potentiel. *MUC4* pourrait exister, comme SMC, sous trois isoformes : une forme transmembranaire, une forme sécrétée et une forme soluble. *MUC4* aurait donc à jouer un rôle réel dans le gel de mucus recouvrant les cellules épithéliales. Ceci a été démontré de façon indirecte pour *MUC1* à propos de l'étude comparative des souris "knock out" CF^{-/-}, *Muc1*^{-/-} et enfin les doubles "knock out" CF^{-/-} et *Muc1*^{-/-}. D'après ces travaux, il semble clairement, tout au moins chez la souris, que *Muc1* joue un rôle décisif sur le volume, l'épaisseur et la densité du mucus colique de façon directe ou indirecte [Parmley et Gendler 1998]. Néanmoins, il apparaît clairement que *MUC4* n'appartient pas à la famille des mucines capables, par la formation de multimères, de réaliser l'assemblage moléculaire qui initie la formation de la phase gel du mucus. Cependant, le mucus contient également, en plus de la phase gel, une phase aqueuse au contact des cellules. *MUC4*, comme *MUC1*, pourrait donc être un composant important de cette phase aqueuse. *MUC4* contrôlerait la quantité et la consistance de cette phase et ainsi ses fonctions de protection de l'épithélium sous-jacent. Il

pourrait aussi jouer un rôle dans l'adhérence de la phase gel supérieure et moduler ainsi son efficacité ou sa clairance.

MUC4 possède, comme SMC et comme les homologues de MUC3 chez la souris et le rat, deux domaines de type EGF. Par homologie avec l'ASGP-2, MUC4 β servirait de ligand à ErbB-2 par un mécanisme intramembranaire. Son rôle pourrait donc être différent des ligands solubles classiques. MUC4, comme SMC, pourrait jouer un rôle dans les processus de prolifération, de réparation et/ou de différenciation des épithéliums, comme le suggèrent les travaux réalisés au laboratoire sur l'expression des mucines fœtales et embryonnaires.

V - LA REGION PROMOTRICE DE *MUC4*

Au cours de ce travail de thèse, nous avons cloné et séquencé 3 710 pb en amont du site potentiel d'initiation de la traduction. De nombreux sites potentiels de fixation de facteurs de transcription y sont localisés (Figure 31). Cette région potentiellement promotrice a été sous-clonée en différents fragments. Ces derniers ont été utilisés au sein du laboratoire par Isabelle Van Seuningen pour des expériences de transfection afin d'étudier leur activité transcriptionnelle. Les premiers résultats montrent une **activité positive pour un fragment de 566 pb contenant la boîte TATA en position 1 040-1 054** sur la figure 31. Un fragment contenant les 460 pb juste en amont de l'ATG potentiel possède également une activité positive. Des expériences complémentaires sont actuellement en cours au laboratoire.

Les connaissances apportées par l'étude de la régulation de *MUC4* devraient nous permettre de comprendre les mécanismes entraînant des variations du niveau d'expression de *MUC4* dans divers cancers ainsi que les mécanismes permettant la synthèse de MUC4 dans des tissus où la glycoprotéine n'est normalement pas exprimée (pancréas, vésicule biliaire).

VI – PERSPECTIVES

Le clonage et le séquençage d'un gène sont des étapes nécessaires et préalables à l'étude de la fonction de la protéine codée par ce gène. De nombreuses perspectives s'ouvrent maintenant pour comprendre les fonctions de MUC4.

L'obtention de nouveaux outils (anticorps polyclonaux et monoclonaux spécifiques de domaines fonctionnels) ainsi que l'identification du promoteur, des régions cis-régulatrices et des facteurs protéiques transrégulateurs, permettront de reconstituer les voies de signalisation transduites par MUC4 ou modulées par MUC4. De même, ils permettront de comprendre son rôle dans la différenciation épidermoïde (en synergie avec MUC1) et dans la différenciation glandulaire (en synergie avec les apomucines dont les gènes sont situés en 11p15.5) chez l'adulte, dans le renouvellement ou la réparation des épithéliums chez l'embryon ou le fœtus au cours du développement ou dans la genèse des cancers de type épidermoïde ou de type mucineux.

Ces résultats contribueront peut-être à permettre le développement de nouveaux marqueurs tumoraux. A terme ils permettront, peut-être, de proposer des solutions thérapeutiques aux sujets atteints de carcinomes de la sphère pancréato-biliaire ou de carcinomes du sein ou du col utérin.

APPENDICE TECHNIQUE

I – EXTRACTION ET PURIFICATION DE L'ADN GENOMIQUE A PARTIR DU SANG

L'ADN génomique est préparé à partir d'un échantillon de sang (10 ml sur EDTA) selon la méthode de Jeanpierre [Jeanpierre 1987]. Les échantillons sont stockés à +4°C si la préparation intervient dans les jours suivants le prélèvement, ou à -20°C dans les autres cas.

I.1 – Lyse cellulaire

La lyse des hématies est réalisée par addition de 40 ml d'une solution de lyse (Tris-HCl 10 mM pH 7,4, EDTA 10 mM) aux 10 ml de sang transférés dans un tube Falcon de 50 ml. Après une attente de 10 min, les tubes sont centrifugés 10 min à 3 000 g à +4°C afin d'éliminer dans le surnageant, l'hémoglobine et les autres constituants solubles. L'opération de lavage est répétée plusieurs fois (5 à 6) jusqu'à obtention d'un culot de débris cellulaires et de noyaux pratiquement blanc (1^{ère} à 3^{ème} fois : 3 000 g ; 4^{ème} à 6^{ème} fois : 1 700 g).

I.2 – Protéolyse

14 ml de chlorure de guanidium 6 M (dénaturant protéique) sont ajoutés au culot qui est ensuite homogénéisé par agitation dans un automate type Turbula pendant 30 min. Puis 2 ml de sarcosyl 10 % (détergent anionique) et 150 µl de protéinase K à 10 mg/ml, préparée extemporanément, sont ajoutés à chaque tube. Celui-ci est alors incubé une nuit à 37°C.

I.3 – Précipitation de l'ADN

Au tube renfermant l'ADN est ajouté 1 ml d'acétate d'ammonium 7,5 M puis le contenu du tube est déversé dans un bécher renfermant 33 ml d'éthanol absolu froid. L'ADN précipite sous la forme d'un filament qui est enroulé autour d'une pipette Pasteur. Celle-ci est rincée 5 fois dans 4 ml d'éthanol froid à 70 %. L'ADN ainsi purifié est resuspendu dans 4 ml de TE 1X, le tube étant mis sous agitation douce pendant 48 heures pour homogénéiser la solution. Les préparations d'ADN sont ensuite conservées en TE 1X à 4°C.

I.4 – Quantification et appréciation de la pureté

La concentration en ADN est évaluée par mesure au spectrophotomètre de la DO à 260 nm en appliquant la règle : 1 unité de DO à 260 nm correspond à une concentration de 50 µg/ml d'ADN double brin. La pureté du matériel est appréciée par mesure des DO à 280 nm (contamination protéique) et à 230 nm (contamination saline) et calcul des rapports $R_1 = DO_{260}/DO_{280}$ et $R_2 = DO_{260}/DO_{230}$. R_1 doit se situer idéalement entre 1,7 et 2. Si R_1 et/ou R_2 est inférieur ou égal à 1, cela indique une contamination de l'échantillon.

II - MARQUAGE DES SONDÉS

II.1 - Sondes nucléiques : technique du multi-amorçage au hasard

L'ADN double brin à marquer (25 ng dans 10 µl) est tout d'abord dénaturé 5 min dans de l'eau bouillante puis gardé dans la glace. La réaction d'élongation, amorcée par un cocktail d'hexanucléotides, est réalisée pendant 30 min à 37°C par le fragment Klenow de l'ADN polymérase (kit "Random Primed DNA Labeling", Boehringer Mannheim) en présence de dNTPs dont du dCTP [$\alpha^{32}\text{P}$]. La sonde double brin est ensuite purifiée par gel-filtration sur colonne de Sephadex G50 en tampon d'élution STE 1X-SDS 0,1 %, puis dénaturée par 1/10 (v/v) de NaOH 3 M pendant 10 min à température ambiante et neutralisée par addition d'HCl 3 M (5/6^{ème} du volume de soude).

II.2 - Sondes oligonucléotidiques : marquage en 5'OH

L'oligonucléotide (5 pmoles dans un volume maximal de 3 µl) est dénaturé 5 min à 95°C, mis 5 min dans la glace puis mélangé à 1 µl de tampon de phosphorylation 10X, 5 µl de dATP [$\gamma^{32}\text{P}$] et 1 µl de T4 polynucléotide kinase (kit "5' end labeling", Amersham). Le mélange est incubé 30 min à 37°C. La réaction est stoppée par un passage de 5 min à 95°C puis 5 min dans la glace et addition de 1/10^{ème} du volume de STE 10X. L'oligonucléotide est ensuite purifié par gel-filtration sur "Quick Spin™ Column Sephadex® G25" (Boehringer Mannheim) puis dénaturé 5 min à 95°C et mis 3 min dans la glace. Un volume égal de formamide est alors ajouté.

III – ANALYSE PAR LA TECHNIQUE DE SOUTHERN BLOT DE L'ADN

L'ADN est hydrolysé par les enzymes de restriction appropriées à raison d'environ 10 U/µg d'ADN dans les conditions de tampon et de température requises pour chaque enzyme.

Les fragments d'ADN sont ensuite séparés par électrophorèse en gel d'agarose (0,8 à 2 %) en tampon borate puis transférés sur membrane et fixés.

III.1 - Transfert

L'ADN est tout d'abord dénaturé par 2 bains de 15 min dans un tampon de dénaturation (NaCl 1,5 M ; NaOH 0,5 M) puis neutralisé par 2 bains de 15 min dans un tampon de neutralisation (Tris-HCl 0,5 M pH 7,5 ; NaCl 3 M). L'ADN est alors transféré sur une membrane de Nylon Hybond™-N+ (Amersham) par capillarité à l'aide d'un tampon SSC 20X pendant une nuit. L'ADN est ensuite fixé sur la membrane par une exposition aux UV.

III.2 - Hybridation

III.2.1 - Sonde nucléique

La membrane humidifiée par du tampon SSC 3X est tout d'abord préhybridée pendant au moins 2 heures à 65°C dans un tampon de préhybridation, afin de saturer les sites libres de la membrane. Elle est ensuite hybridée à l'aide d'un tampon d'hybridation, pendant une nuit à la même température, en présence de la sonde marquée au ³²P (entre 100 000 et 150 000 cpm par couloir).

III.2.2 - Sonde oligonucléotidique

Les étapes techniques sont identiques à celles employées pour utiliser une sonde double brin mais la température de préhybridation et d'hybridation ainsi que les compositions des tampons sont différentes :

- température de préhybridation et d'hybridation : 42°C
- tampon d'humidification : SSPE 6X
- tampon de préhybridation et d'hybridation (cf. composition des tampons).

III.3 - Lavages et autoradiographie

L'excès de sonde est éliminé de la membrane par rinçage dans un tampon SSC 3X (cas d'une sonde d'ADN double brin) ou SSPE 6X (cas d'une sonde oligonucléotidique) ; la membrane est ensuite lavée 2 fois 15 min à 65°C par un tampon SSC 0,1X-SDS 0,1 % (cas d'une sonde d'ADN double brin) ou à 45°C, par un tampon SSPE 4X-SDS 0,1 % (cas d'une sonde oligonucléotidique). Après un nouveau rinçage en tampon SSC 3X ou SSPE 6X, la membrane est séchée sur papier et mise en contact avec un film autoradiographique pendant des temps pouvant aller de quelques min à plusieurs jours. La membrane peut ensuite être

déshybridée par des bains bouillants de SDS (0,1 à 0,5 %) et réutilisée après contrôle de la déshybridation par autoradiographie.

IV – EXTRACTION ET PURIFICATION DES ARN TOTAUX ET DES ARNm A PARTIR DE TISSUS

IV.1 – Préparation des ARN totaux par la technique dite "classique"

La technique utilisée a été adaptée à partir de la technique décrite par Chirgwin [Chirgwin *et al.* 1979]. Des précautions particulières doivent être prises lors de la préparation des ARN, ceux-ci étant très sensible aux RNases, en particulier à la RNase A. Celle-ci est ubiquitaire, extrêmement active et très résistante (un traitement à 90°C pendant une heure n'altère pas son activité). Toute préparation et manipulation d'ARN doit être effectuée le plus stérilement possible : les tampons et l'eau sont traités au DEPC (diéthylpyrocarbonate) puis autoclavés ; le matériel est aussi autoclavé quand cela est possible, sinon il est traité une nuit dans la soude 1 N puis rincé soigneusement à l'eau DEPC juste avant son utilisation. De plus, le port de gants permet d'éviter un apport exogène de RNases. L'ensemble du protocole utilisé a été détaillé dans le mémoire de thèse de Virginie Debailleul [Debailleul 1997]. Les transcrits des apomucines (d'une taille inhabituellement grande) préparés par cette technique classique, étaient caractérisés par un profil d'expression par Northern blot présentant un haut degré d'hétérogénéité (continuum homogène de transcrits).

IV.2 – Protocole original adapté à la préparation de grands ARN

Cette technique dite "optimisée" reprend les grandes étapes de celle décrite précédemment (les tampons et les quantités utilisés sont les mêmes). Cependant, quelques modifications ont été apportées par Virginie Debailleul : les échantillons d'ARN sont préparés dans des conditions dites "douces" (en évitant les cassures mécaniques) et transférés avec un traitement préalable à la soude (pour augmenter le transfert difficile des gros ARN) [Debailleul *et al.* 1998]. Dans ces conditions, les transcrits des mucines apparaissent sous forme de bandes et non plus sous forme de continuum.

IV.3 – Sélection des ARN poly(A⁺)

Deux méthodes sont utilisées :

- affinité sur colonne d'oligo dT cellulose (Pharmacia Biotech) [Aviv et Leder 1972],

- affinité sur billes magnétiques (utilisation d'un kit commercialisé par Promega : "PolyA Tract mRNA Isolation System III" utilisé dans les conditions recommandées par le fabricant).

V – ANALYSE DES ARN TOTAUX OU MESSAGERS PAR LA TECHNIQUE DE NORTHERN BLOT

V.1- Electrophorèse des ARN en milieu dénaturant

Les échantillons d'ARN (environ 10 µg) sont dénaturés par 10 min de chauffage à 68°C dans un tampon contenant 50 % de formamide désionisée, 18 % de formaldéhyde désionisée et du tampon MOPS 0,02 M. Les ARN sont ensuite séparés par électrophorèse en gel d'agarose 0,9 % contenant 18 % de formaldéhyde et du tampon MOPS 0,02 M. L'électrophorèse est effectuée en tampon MOPS 0,02 M à 30 volts pendant 16 heures.

V.2 – Transfert

Dans la technique "optimisée", le gel est préalablement trempé 20 min dans de la soude 0,05 N. Il est ensuite rincé et équilibré en SSC 20X. Le matériel ribonucléique est transféré sur une membrane de type "HybondTM-N⁺" (Amersham) par transfert sous vide pendant 1 heure. La fixation covalente de l'ARN sur le support est assurée par chauffage de la membrane à 80°C pendant 30 min sous vide partiel puis par une exposition aux UV.

V.3 - Hybridation

La membrane est humidifiée en tampon SSPE 6X et préhybridée au moins 2 heures à 42°C avec un tampon de préhybridation/hybridation préparé selon le protocole préconisé par Clontech. Elle est ensuite hybridée après remplacement du tampon "Clontech" par contact avec la sonde (1.000.000 à 1.500.000 cpm par couloir) durant une nuit.

V.4 - Lavage et autoradiographie

La membrane est rincée par du tampon SSPE 6X et lavée 2 fois pendant 15 min à 65°C par du tampon SSPE 0,1X-SDS 0,1 %. Elle est alors exposée au film autoradiographique durant plusieurs jours.

VI - CRIBLAGE DES BANQUES GENOMIQUES

VI.1 – Banque génomique construite en vecteur bactériophagique λ EMBL4

Une banque génomique a été construite en vecteur lambdaïde EMBL4 dans notre équipe. Ce vecteur permet de cloner des fragments d'une vingtaine de kilobases. Les fragments d'ADN génomique insérés ont été obtenus par une hydrolyse par l'enzyme de restriction *Sau3A* dans des conditions d'hydrolyse partielle.

Le vecteur est mis en contact avec une culture bactérienne. Après incubation une nuit à 37°C, on observe des plages de lyse contenant chacune environ 10⁶ phages recombinants. Les clones d'intérêt sont sélectionnés parmi l'ensemble des clones de la banque à l'aide d'une ou plusieurs sondes nucléiques.

VI.1.1 - Criblage de la banque

L'ADN phagique est transféré sur un disque de nitrocellulose déposé à la surface de chaque boîte et laissé en contact 30 s. Les disques sont alors retirés délicatement et traités comme suit :

- contact 2 min avec un tampon de dénaturation (NaOH 0,2 M ; NaCl 1,5 M) ;
- contact 4 min avec un tampon de neutralisation (Tris HCl 0,5 M pH 7,5 ; NaCl 3 M) ;
- cuisson 2 heures à 80°C sous vide afin de fixer l'ADN sur la membrane ;
- incubation au moins 2 heures dans du tampon de préhybridation 3X ;
- incubation au moins 4 heures avec la sonde nucléique dans le même tampon ;
- lavages : 3 fois 5 min par du SSC 0,1X, SDS 0,1 % à 55°C ;
- révélation par autoradiographie.

VI.1.2 - Préparation des phages recombinants

Les recombinants révélés positifs par autoradiographie sont prélevés. Leur titre phagique est augmenté en milieu liquide par diverses étapes.

VI.1.2.1 - Préministock

Un recombinant positif prélevé de la boîte de Pétri est mélangé à 200 μ l de culture bactérienne en présence de MgSO₄ 10 mM. Après 15 min de contact, le mélange est versé dans un flacon contenant 5 ml de LB MgSO₄ et placé sous agitation à 37°C jusqu'à obtention de la lyse.

Lorsque la lyse est terminée, on ajoute $1/100^{\text{ème}}$ du volume de chloroforme et on laisse en contact 15 min. Après centrifugation, le surnageant est récupéré et constitue notre préministock auquel sont ajoutées quelques gouttes de chloroforme.

VI.1.2.2 - Ministock

Dans un flacon de 250 ml, on met 2 ml de LB-MgSO₄, 50 µl de préministock et 2 ml de relance bactérienne. Après 10 min à 37°C, on ajoute 16 ml de LB-MgSO₄. Le mélange est alors placé à 37°C sous agitation jusqu'à obtention de la lyse. On ajoute ensuite du chloroforme ($1/100^{\text{ème}}$) et on laisse en contact 15 min. Le surnageant obtenu après centrifugation est conservé avec quelques gouttes de chloroforme et constitue notre ministock de phages.

Avant de réaliser l'amplification proprement dite et la purification de l'ADN phagique, on effectue une préparation rapide de cet ADN phagique afin de vérifier la qualité des phages recombinants.

VI.1.2.3 - Préparation rapide de l'ADN phagique

A 3 ml de ministock sont ajoutés 150 µl de DNase I (1 mg/ml dans NaCl 0,15 M ; glycérol 50 %) et 10 µl de RNase à 10 mg/ml (en tampon Tris-HCl pH 7,5 ; NaCl 0,015 M). Puis le tout est incubé à 37°C pendant 30 min.

Au mélange précédent, on ajoute 65 µl de Tris-HCl 1 M pH 8,0, 130 µl d'EDTA 0,5 M pH 8,0 et 17 µl de protéinase K (20 mg/ml). Le tout est incubé 5 min à 37°C. Puis, 330 µl de Sarcosyl à 5 % sont ajoutés au mélange qui est alors incubé 30 min à 37°C.

Les protéines sont éliminées par extraction par un mélange phénol/chloroforme. Après précipitation par $1/10^{\text{ème}}$ du volume d'acétate de sodium 2 M pH 5,5 et 2 volumes d'alcool absolu à -20°C, le culot d'acides nucléiques est redissous par du TE 1X. La solution est dialysée 2 heures sur une membrane Millipore, contre du TE 1X.

L'ADN phagique ainsi purifié est hydrolysé par les enzymes *EcoRI*, *BamHI* et *EcoRI/BamHI*. Les fragments sont ensuite séparés par électrophorèse en gel d'agarose 0,8 %. A partir de ces gels sont réalisés des Southern blots qui permettent de vérifier la présence de l'insert dans le phage recombinant.

VI.1.2.4 - Amplification et extraction de l'ADN recombinant

L'amplification est réalisée dans les conditions décrites précédemment sous un volume de 500 ml à partir du ministock.

La purification des phages demande les étapes supplémentaires suivantes :

La précipitation des phages se fait à 4°C en présence de NaCl et de PEG. Après centrifugation, le culot est repris par 8 ml de TMN. Puis 8 ml de chloroforme sont ajoutés au mélange qui est centrifugé 10 min à 3 800 g. On ajoute alors 9,4 g de surnageant à 7 g de CsCl et on sépare l'ADN par ultracentrifugation pendant 2 heures à 4°C à 300 000 g. L'ADN phagique est prélevé à la seringue. Le CsCl est éliminé par dialyses successives :

- 2 x 30 min contre du Tris 20 mM, EDTA 10 mM, NaCl 100 mM
- 2 x 30 min contre du Tris 20 mM, EDTA 1 mM, NaCl 100 mM.

On ajoute alors 25 µl de protéinase K à 20 mg/ml ainsi que 63 µl de SDS. Après une incubation de 2 heures à 37°C, l'ADN est purifié par extraction par le phénol/chloroforme puis précipité par 1/10^{ème} du volume d'acétate de sodium 2 M pH 5,5 et 2 volumes d'alcool absolu à -20°C. Il est alors enroulé sur une baguette de verre et rincé à l'alcool à 70 %. L'ADN est ensuite dissous dans du TE 1X, pendant au moins 24 heures et dosé au spectrophotomètre à 260 nm. L'ADN recombinant ainsi obtenu, purifié en grande quantité, peut alors être analysé.

VI.2 – Banque génomique construite en vecteur cosmique pWE15

La banque génomique utilisée est une banque commerciale d'ADN génomique humain construite en vecteur cosmique pWE15 (Stratagène). L'hydrolyse partielle par l'enzyme de restriction *MboI* permet l'obtention de fragments d'ADN génomique de 35 à 45 kb qui sont ensuite clonés dans un site *BamHI*. Les inserts peuvent être excisés par l'enzyme *NotI*.

VI.2.1 - Criblage de la banque

Un aliquot de la banque diluée (1/10⁶) est étalé sur des boîtes contenant un milieu LB-Agar Ampicilline 50 µg/ml. Les boîtes sont placées durant une nuit à l'étuve à 37°C. Après apparition des clones, une empreinte de la position des clones sur chaque boîte est obtenue par simple contact avec une membrane de nitrocellulose. L'ADN est ensuite dénaturé par la soude 0,5 M puis neutralisé dans un tampon Tris-HCl 1 M pH 7,6 et fixé sur la membrane par un traitement par le tampon NaCl 1,5 M, Tris-HCl 1 M pH 7,6. Cette fixation deviendra irréversible par cuisson à 80°C sous vide pendant 2 heures. Les membranes sont préhybridées dans du tampon d'hybridation 3X pendant au moins 2 heures à 42°C et hybridées pendant une nuit avec une sonde marquée au ³²P (500 000 cpm/disque). Après 3 lavages de 5 min en SSC 0,1 X, SDS 0,1 %, les membranes sont rincées en SSC 3X et soumises pendant 24-48 heures à exposition autoradiographique.

Un 2^{ème} tour de criblage, réalisé dans les mêmes conditions, est effectué avec les clones positifs afin de les isoler.

VI.2.2 - Purification de l'ADN des cosmides recombinants

Une petite quantité d'ADN cosmique est d'abord préparée afin d'identifier les clones d'intérêt (kit "circle spin midi", BIO101).

Les clones positifs sont mis en culture à 37°C pendant une nuit dans 30 ml de LB contenant 50 µg/ml d'ampicilline. Après centrifugation, le culot est repris par différents tampons permettant la lyse des cellules, la dégradation de l'ARN, ainsi que la précipitation des protéines et de l'ADN bactérien. La purification de l'ADN cosmique est réalisée par passage sur une résine. L'ADN est ensuite élué par du TE 1X.

VI.2.3 – Etablissement des cartes de restriction

Pour construire la carte de restriction des inserts clonés en vecteur cosmique, on utilise l'enzyme *NotI* qui permet de libérer l'insert. Puis on effectue des hydrolyses totales et partielles de l'ADN cosmique par des enzymes de restriction différentes. Ces digestions partielles utilisent des concentrations variables d'enzyme. Les fragments de différentes tailles sont ensuite séparés par électrophorèse en gel d'agarose puis sont transférés sur une membrane de Nylon N⁺. Le cosmide pWE15 possède, entre les sites *NotI* de chaque côté de l'insert et le site de clonage, des promoteurs T3 et T7. L'hybridation avec les sondes oligonucléotidiques correspondantes permet d'orienter les fragments de digestion partielle par rapport aux extrémités du vecteur de clonage.

VII – CRIBLAGE DES BANQUES D'EXPRESSION

L'ARN total préparé au laboratoire à partir de muqueuse de côlon ou d'intestin grêle humain a été utilisé comme matrice pour la synthèse d'ADNc. La synthèse de l'ADNc double brin et le clonage en vecteur phagique λ gt11 ont été réalisés dans les conditions définies par le fournisseur (kits "cDNA synthesis system" et "cDNA cloning system – λ gt11", Amersham). Seule la banque de côlon a été criblée avec la sonde JER64 (cf. VI.1), la banque d'intestin grêle ayant un titre plus faible ($2,5 \times 10^6$ pfu contre $10,3 \times 10^6$ pfu).

VIII – SOUS CLONAGE EN VECTEUR PLASMIDIQUE PKS

Le vecteur utilisé pour le sous-clonage des fragments cosmidiqes ou plasmidiqes est le phagémide *pBluescript*[®] II KS (+) (Stratagène). Il a une taille de 2 961 pb.

VIII.1 - Purification des fragments

Deux méthodes de purification ont été utilisées :

VIII.1.1 - Méthode "GENECLEAN"

Les fragments d'ADN sont séparés par électrophorèse en gel d'agarose 0,8% contenant du BEt (0,5 µg/ml), et en tampon TEA pour éviter l'action des DNases. Les marqueurs de masse moléculaire utilisés pour la calibration sont les fragments du phage 13i hydrolysé par *EcoRI* (7 fragments de 31, 22, 11, 4,3, 3,2, 1,3 et 0,5 kb) et le marqueur de taille commercial "100-bp ladder" contenant des fragments d'ADN multiples de 100 pb (Pharmacia Biotech). Les fragments d'intérêt sont extraits de l'agarose dont les bandes sont découpées au scalpel. L'ADN est extrait et purifié à l'aide du kit "GENECLEAN II"[®] (Bio 101 Inc.) : l'agarose est liquéfié en ajoutant trois volumes de NaI 6 M et en incubant 5 min à 50°C. L'ADN est ensuite adsorbé sur la matrice de silice ("Glassmilk") pendant 5 min sur de la glace : 5 µl de "Glassmilk" sont ajoutés pour 5 µg d'ADN (ou moins) et 1 µl de "Glassmilk" supplémentaire par 0,5 µg d'ADN au delà des 5 µg. Après centrifugation, on récupère le culot d'ADN/matrice et on le lave trois fois par des solutions de 300 µl de New wash dilué (selon les conditions du fournisseur). Après centrifugation, le culot est resuspendu dans 5 à 10 µl de TE 1X et placé environ 5 min à 50°C. Après 30 s de centrifugation à 17 500 g, le surnageant contenant l'ADN est transféré dans un nouveau tube. Une deuxième élution de l'ADN peut être effectuée et permet d'obtenir 10-20 % supplémentaires du matériel élué la première fois.

VIII.1.2 - Méthode "SeaPlaque"

Les fragments d'ADN à purifier sont séparés par électrophorèse en gel d'agarose SeaPlaque[®]GTG[®] (FMC BioProducts) à 0,8 % en tampon borate et contenant du BEt (0,5 µg/ml). Cet agarose a la particularité de se liquéfier à basse température. Après repérage sous UV, les bandes d'agarose d'intérêt sont excisées au scalpel et placées dans un tube de 1,5 ml. On ajoute 50 µl de TE 1X et 100 µl de tampon de purification de PCR (Promega). Les tubes sont mis au bain-marie à 65°C pendant 10 min. On ajoute ensuite au gel liquéfié 1 ml de résine (Wizard[™] PCR Preps DNA Purification System, Promega). On retourne plusieurs fois

les tubes ; l'ADN s'adsorbe sur la matrice. On filtre ensuite les mélanges sur des minicolonnes (Wizard™ Minicolumns, Promega). Les filtres sont lavés 2 fois par 1 ml d'isopropanol 80 % puis essorés par centrifugation pendant 2 min à 17 500 g. On dépose sur les colonnes 50 µl de TE 1X. Après au moins 1 min de contact, les produits d'amplification purifiés sont élués par centrifugation pendant 30 s à 17 500 g. Le produit d'amplification, s'il est unique, peut également être purifié directement sans passer par l'étape d'électrophorèse. La purification se fait par addition de 100 µl de tampon de purification et 1 ml de résine. L'ADN s'adsorbe sur la résine puis le mélange est filtré sur les minicolonnes. Après les lavages à l'isopropanol 80 %, l'ADN est élué par du TE 1X.

VIII.2 - Hydrolyse du vecteur

Le vecteur est tout d'abord clivé par une ou deux enzymes de restriction appartenant au site de clonage multiple. Le vecteur est ensuite purifié par extraction au phénol/chloroforme ou purifié comme indiqué ci-dessus. Il est ensuite précipité, repris par du TE 1X et dosé en gel d'agarose par comparaison au témoin 13i.

VIII.3 – Déphosphorylation

Lorsque le vecteur est hydrolysé par une enzyme unique, on le déphosphoryle afin d'éviter qu'il ne se referme sur lui-même. On utilise pour cela une phosphatase (Boehringer Mannheim). Le vecteur ouvert est incubé 1 heure à 37°C avec un tampon 10X (Tris-HCl 0,5 M ; EDTA 1 mM pH 8,5) et 1 U de phosphatase alcaline (Boehringer Mannheim). Le vecteur déphosphorylé est ensuite purifié et repris en TE 1X. Il est ensuite dosé après électrophorèse en gel d'agarose.

VIII.4 – "Ligation" rapide

L'ADN purifié est mélangé avec 50 ng de vecteur, 2 µl de tampon 5X (kit "Rapid DNA Ligation", Boehringer Mannheim) et de l'eau stérile (qsp 10 µl). La ligation est réalisée dans un rapport molaire insert/vecteur de 3/1. On ajoute 10 µl de tampon de ligation 2X et 1 µl (5 U/µl) de T4 ADN ligase. Le mélange est incubé 5 à 10 min à température ambiante.

VIII.5 - Transformation

1 µl de produit de "ligation" est ajouté à 19 µl de bactéries JM109 ou HB101 (Promega) vendues compétentes. Après 30 min de contact dans la glace, le mélange est placé 45 s à 42°C puis 2 min dans la glace. 150 µl de SOC sont ajoutés, puis les tubes sont incubés

1 h à 37°C. Les bactéries sont ensuite étalées sur boîtes comme précédemment sur un milieu LB-agar contenant de l'ampicilline (25 µg/ml pour les HB101 et 100 µg/ml pour les JM109). Les boîtes sont placées à 37°C pendant une nuit.

VIII.6 - Recherche des clones d'intérêt

30 min avant d'étaler sur boîte de Pétri une culture de bactéries JM109, on imprègne des boîtes LB-agar contenant 100 µg/ml d'ampicilline avec 20 µl d'une solution de X-Gal à 50 mg/ml et 100 µl d'IPTG à 24 mg/ml. La souche JM109 possède un gène codant une β-galactosidase défective, dépourvue des aa 11 à 41 (gène LacZΔM15), utile pour le test de l'α-complémentation. Le vecteur *pBluescript*[®] II KS (+) possède quant à lui, en plus d'un gène de résistance à l'ampicilline, un site de clonage multiple (encore appelé "polylinker") introduit au début du gène LacZ. Pour mettre en évidence la présence (vecteur sauvage sans insert) ou non (clonage d'un insert dans le gène LacZ) d'une activité β-galactosidase, deux réactifs sont indispensables : X-Gal et IPTG. X-Gal est un β-galactoside qui peut être hydrolysé par la β-galactosidase libérant ainsi, en plus du galactose, une substance colorée en bleu. Finalement, les colonies blanches sont témoins de la présence de recombinants et les bleues d'absence de recombinants. L'IPTG est un inducteur par dérégulation de la synthèse de β-galactosidase.

Dans le cas des bactéries HB101, il est nécessaire pour identifier les clones recombinants d'intérêt de posséder la sonde correspondant à un fragment de l'insert à sous-cloner. On procède alors comme suit.

VIII.6.1 - Empreintes plasmidiques

Un disque de nitrocellulose est appliqué sur chaque boîte de Pétri pendant 1 min. L'ADN est dénaturé dans de la soude 0,5 N pendant 5 min. Le disque est ensuite neutralisé pendant 5 min dans un bain de Tris-HCl pH 6,0-7,0 1M ; NaCl 1,5 M. Le disque de nitrocellulose est séché sur papier et le matériel est fixé par la chaleur (1 heure à 80°C sous vide).

VIII.6.2 - Hybridation à une sonde nucléique

Le principe de la méthode est identique à celui précédemment décrit si ce n'est une préhybridation et une hybridation à 42°C dans un tampon d'hybridation 3X pendant au moins 2 heures. L'hybridation est obtenue en présence de la sonde ajoutée à raison de 250 000 cpm

par disque. Enfin, les filtres ou disques sont lavés par 4 bains de 5 min à 55°C dans un tampon de SSC 0,1X-SDS 0,1 % et sont soumis à autoradiographie pendant 6 heures ou une nuit.

IX - PREPARATION DE L'ADN PLASMIDIQUE RECOMBINANT

(kit "Wizard™ Minipreps DNA Purification System", Promega)

IX.1 - Lyse bactérienne par un détergent

Les clones positifs sont mis en culture dans 5 ml de LB-agar contenant de l'ampicilline. 1,5 ml de culture sont centrifugés 5 min à 14 000 g. Le surnageant est éliminé à la trompe à vide et le culot de bactéries est remis en suspension dans du tampon (200 µl de Tris-HCl pH 7,5 50 mM ; EDTA 10 mM ; RNase A 100 mg/ml). Les bactéries sont ensuite lysées (200 µl de NaOH 0,2 M ; SDS 1 %) et les protéines sont précipitées (200 µl d'acétate de potassium pH 4,8 1,32 M) puis centrifugées. Après centrifugation (10 min à 17 500 g), le surnageant contenant les plasmides est récupéré.

IX.2 - Purification de l'ADN plasmidique

1 ml de résine "Wizard™ Minipreps DNA Purification" est ajouté au surnageant. Le tout est filtré sur minicolonne. Après lavage par 2 ml d'une solution de "Wash" (NaCl, 200 mM ; Tris-HCl pH 7,5 20 mM ; EDTA 5 mM) diluée dans de l'éthanol 95 % (selon les conditions du fournisseur), la colonne est essorée par centrifugation pendant 2 min.

IX.3 - Elution de l'ADN de la matrice

50 µl de TE 1X chaud (65°C) sont appliqués sur la colonne pendant 1 min. L'ADN plasmidique est élué de la colonne par 20 s de centrifugation.

IX.4 - Détermination de la taille des inserts

2 µl de chaque solution de plasmides sont hydrolysés pendant 1 heure à 37°C dans un tube contenant 6 µl de TE 1X, 1 µl de tampon 10X adapté à l'enzyme (Boehringer Mannheim) et 1 µl d'enzyme (ou 2 × 0,5 µl de 2 enzymes différentes) libérant l'insert du site de clonage. L'ADN de l'insert et du vecteur sont ensuite séparés par électrophorèse sur gel d'agarose.

Une hydrolyse sur une plus grande quantité (35 µl) permet, après purification comme précédemment (cf. VIII

.1), d'avoir assez de matériel pour utiliser ces inserts comme sonde sur des Southern blots et/ou des Northern blots.

X – CONSTRUCTION DE MUTANTS PAR DELETION

Cette technique permet de déléter les régions distales d'un gros fragment d'ADN et ainsi de pouvoir séquencer ses régions internes sans sous-clonage préalable. On utilise l'exonucléase III qui élimine les nucléotides un à un dans le sens 3'→5' à partir d'une extrémité 3' rentrante. Par contre, une extrémité 3' sortante est protégée de la dégradation. L'ADN recombinant est hydrolysé successivement par 2 enzymes de restriction. La première hydrolyse libère une extrémité 3' sortante du côté du vecteur. La deuxième hydrolyse libère une extrémité 5' sortante, sensible à l'exonucléase III, du côté de l'insert. Après précipitation de l'ADN, le culot est repris par un tampon ExoIII (Tris 66 mM pH 8,0 ; MgCl₂ 6,6 mM).

Après 5 min à 37°C, on ajoute 150 U d'Exonucléase III. Toutes les 30 s, on prélève 2,5 µl que l'on ajoute à 7,5 µl du mélange nucléase S₁ (2,5 U par tube)-tampon S₁ (NaCl 500 mM; AcoNa pH 4,6 2 mM ; glycérol 10 % ; ZnSO₄ 2 mM). Les tubes sont incubés 30 min à 30°C. On ajoute alors 1 µl de tampon stop S₁ (Tris pH 8,0 0,8 M ; EDTA 20 mM ; MgCl₂ 80 mM). On met le tout 10 min à 70°C puis dans la glace. On prélève 2 µl par tube pour étude électrophorétique. On ajoute ensuite 1 µl du mélange suivant : 20 µl d'eau ; 6 µl de MgCl₂ 1 M ; 3 µl de Tris pH 7,6 0,1 M et 5 U de Klenow (Boehringer Mannheim). On incube 5 min à 37°C. On ajoute 1 µl du mélange des 4 nucléotides (500 pmoles de chaque) et on laisse le tout 15 min à température ambiante. Enfin, 40 µl du mélange de ligation suivant sont ajoutés : 650 µl d'eau ; 100 µl de tampon 10X (Tris-HCl 660 mM ; MgCl₂ 50 mM ; dithioérythritol 10 mM; ATP 10 mM ; pH 7,5) ; 250 µl de PEG 30 % et 5 U de T4 DNA ligase (Boehringer Mannheim). Le tout est incubé 2 heures à température ambiante. On obtient alors, à partir du même fragment d'ADN, une série de plasmides recircularisés contenant un insert d'ADN de taille variable que l'on transforme dans des bactéries.

XI - SEQUENÇAGE NUCLEOTIDIQUE

L'ADN des clones obtenus est séquencé sur chaque brin. La technique utilisée est celle de Sanger [Sanger *et al.* 1977]. L'ADN plasmidique ou cosmique est tout d'abord dénaturé par de la soude en présence d'EDTA puis précipité par de l'acétate de sodium 3 M et de l'éthanol. Le vecteur monobrin est ensuite hybridé à un oligonucléotide synthétique (0,5 pmole/µl) correspondant aux amorces T3 et T7 du plasmide pKS, aux amorces T7 et M13 reverse du vecteur pCR2.1, ou aux amorces spécifiques de *MUC4*. Quatre réactions de polymérisation en parallèle sont réalisées par la T7 ADN polymérase ("Sequenase version 2.0", Amersham) ou par une polymérase thermostable (kit "SequiTherm™ Cycle

Sequencing", Epicentre Technologies). La synthèse du second brin est effectuée en présence de dCTP, dGTP, dTTP et de dATP marqué en α par du ^{35}S , plus stable que le ^{32}P et donnant des images plus nettes. Chacune des 4 réactions est ensuite poursuivie en présence de l'une des 4 bases sous forme di-désoxyribonucléoside triphosphate à très faible concentration qui arrête, au hasard, la polymérisation. Les fragments de tailles différentes sont ensuite séparés par électrophorèse en gel de polyacrylamide à 6 % en tampon borate contenant de l'urée (SequaGel-6™, National Diagnostics). La lecture directe de la séquence est ensuite réalisée après autoradiographie pendant une nuit à plusieurs jours. Cette méthode permet de lire des séquences de 200 à 400 pb. En ce qui concerne la séquence directe à partir d'ADN cosmique, elle s'effectue à partir de 2 μg d'ADN. 5 pmoles d'oligonucléotide (au lieu de 0,5 pmoles pour les plasmides) sont hybridés à l'ADN à 37°C pendant 30 min (65°C pendant 2 min pour les plasmides).

Remarque : Au cours de ce travail, en fin d'étude, quelques séquences ont pu être réalisées par séquençage automatique sur LI-COR 4000L (ScienceTec).

XII - ANALYSE DES SEQUENCES

Les données ont été traitées à l'aide des programmes informatiques de PC/GENE, et récemment grâce au serveur NCBI disponibles sur Internet : <http://www.ncbi.nih.gov/>.

XIII - SYNTHÈSE D'ADNC

XIII.1 - Méthode d'amorçage aléatoire

De l'ARN poly(A)⁺ de trachée humaine (0,5 μg) (Clontech) ou de l'ARN total préparé au laboratoire à partir de muqueuse de côlon humain (1 μg) sous un volume de 12,5 μl (eau DEPC) est mis en présence de 1 μl d'un cocktail d'hexanucléotides puis dénaturé par chauffage pendant 2 min à 70°C. Le mélange est ensuite mis dans la glace 5 min puis centrifugé quelques secondes. On y ajoute des dNTPs (1 μl à 10 mM chaque), 20 U d'inhibiteur de RNases (0,5 μl), 4 μl de tampon de transcriptase inverse (5X) et 200 U (1 μl) de transcriptase inverse MMLV du kit (kit "1st STRAND™ cDNA Synthesis" ; Clontech). Le tout est incubé pendant 1 heure à 42°C. L'ARN, la polymérase et les DNases sont ensuite détruits par chauffage pendant 5 min à 94°C. Après centrifugation, les 20 μl de solution d'ADNc complétés par 80 μl d'eau DEPC sont conservés à -80°C. Une analyse par PCR de contrôle est effectuée sur 5 μl d'ADNc en utilisant des amorces spécifiques de la G3PDH.

XIII.2 - Méthode RACE-PCR

La RACE-PCR a initialement été décrite par Frohman *et al.* [Frohman *et al.* 1988]. Nous avons utilisé le kit "5'/3' RACE" (Boehringer Mannheim).

XIII.2.1 - Synthèse de l'ADNc

Dans un microtube sont mélangés puis centrifugés :

tampon de transcription inverse (5X)	4 µl
dNTPs (10 mM chaque)	2 µl
oligonucléotide spécifique (5 pmol/µl)	2,5 µl
ARN poly(A) ⁺ (0,5 µg) ou total (2µg)	
eau DEPC	qsp 19 µl
transcriptase inverse AMV (20 U/µl)	1 µl

Le mélange est incubé pendant 1 heure à 55°C puis 10 min à 65°C. Il est centrifugé et purifié. Les amorces spécifiques de *MUC4* que nous avons utilisées sont : NAU124, NAU155, NAU168 et NAU287 (Tableau 6).

XIII.2.2 - Purification de l'ADNc

Aux 20 µl d'ADNc sont ajoutés 100 µl d'une solution tampon (thiocyanate de guanidine 3 M ; Tris-HCl 10 mM ; éthanol 5 % (v/v) ; pH 6,6), pour permettre l'adsorption des acides nucléiques sur microcolonne. Le mélange est passé sur un filtre (kit "High Pure PCR Product Purification", Boehringer Mannheim) et l'ADNc retenu est lavé 2 fois par 0,5 ml d'une solution de lavage (NaCl 20 mM ; Tris-HCl 2 mM ; pH 7,5 ; 4 volumes d'éthanol). Après élimination des contaminants par centrifugation pendant 30 secondes à 14 000 g, l'ADNc purifié est élué par 50 µl d'une solution de Tris-HCl, 10 mM pH 8.

XIII.2.3 - Synthèse de la queue poly(A)

Dans un microtube de 0,5 ml placé sur un lit de glace sont mélangés 19 µl d'ADNc purifié, 2,5 µl d'une solution tampon 10X (Tris-HCl, 100 mM ; MgCl₂, 15 mM ; KCl, 500 mM, pH 8,3) et 2,5 µl de dATP (2 mM). L'ADNc est dénaturé à 94°C pendant 3 min puis remis sur la glace et centrifugé. On ajoute 1 µl de l'enzyme terminale-transférase (10 U) qui va catalyser l'addition de dATP à l'extrémité 3'OH de l'ADNc. L'incubation se fait à 37°C pendant 30 min. La terminale-transférase est ensuite inactivée par chauffage à 70°C pendant 10 min. Le mélange, après centrifugation, est conservé à -80°C. 5 µl de ce mélange seront utilisés pour chacune des réactions de PCR.

Désignation	Séquence (de 5' en 3')	Position	Orientation
NAU99	TGGCTGTGCGTCTCCCA	3 093/3 109	AS
NAU103	GTAATGCGAATGCACCAGTG	244/263	AS
NAU124	TGTGTCTTGGATAGAGGTCCTCCA	565/588	AS
NAU138	GGAAGTTGTTTCGTGATTGTCCTTCCTG	538/564	AS
NAU155	TGTGTACACTTGGGGAAG	308/325	AS
NAU156	GAGTTGATGTCATCATCTGCGTGAGGG	275/301	AS
NAU162	GAATTCGTGTGAAGGGGGCACGCTGGA	2/19	S
NAU167	TGGAAAGTGACGTGCCTCCTGAGG	2 192/2 215	AS
NAU168	CAGGAGTTTACTTCCAGTT	105/124	AS
NAU169	ATGTGTCCCTCTGTGGTTCCT	81/100	AS
NAU170	CCGTGGTGAAGGTTTTACCA	2 901/2 920	AS
NAU174	AGCTTCCACAGTATCCTCGGA	3 015/3 035	S
NAU204	ACCACATGCGGAAGGAGGCAGAGACAC	51/77	AS
NAU224	GCTGTTGTGTCCTGAGTAGA	955/974	AS
NAU225	TTCTTGTGTGGTCTGCGGGGC	1 939/1 959	AS
NAU247	TTGACAGAGTGTCTGACCACC	1 301/1 321	AS
NAU287	CTTCATGGCTGCGGCAAAAGT	-15/6	AS
NAU288	TGGCTCCCTGGGGAAGCTCCA	-41/-21	AS
NAU289	CAGTGTGAGGAGCAGACGTGA	-76/-56	AS
NAU327	GTGCGTTTCTCCGAAGGGCCA	-130/-109	AS
NAU357	GTCCCACCAGAGCTTTGGGCT	-407/-387	AS
NAU370	TCTGCTCCTCACACTGCAGCTGCT	-71/-48	S
NAU376	AAATGGTGGGCCACCCTCCTA	-342/-321	AS
NAU598	ATATTCCACGGGAGGCAGGAAGAGCTT	655/681	AS
NAU599	AGTGTGGATGTGGGGTGTAAAGTAAATT	628/654	AS

Tableau 6 : Oligonucléotides utilisés.

Les chiffres indiquant la position des oligonucléotides sont ceux retrouvés sur la figure 30, sauf pour les oligonucléotides présents dans une séquence intronique : NAU103 (Figure 19) et NAU598-NAU599 (Figure 29). AS : antisens, S : sens.

XIV – REACTIONS D'AMPLIFICATION

XIV.1 - Principe

La réaction de polymérisation en chaîne (PCR pour "Polymerase Chain Reaction") a été mise au point au milieu des années 80 par K. Mullis (prix Nobel de chimie en 1993). Elle utilise un couple d'amorces qui encadre la région que l'on veut amplifier sélectivement, des dNTPs, de l'ADN monobrin (essentiellement de l'ADNc) ou double brin (ADN génomique ou un vecteur et son insert) et une polymérase thermostable de type Taq Polymérase et son tampon.

XIV.2 – Conditions techniques générales

Nous avons utilisé cette méthode pour amplifier de l'ADN à partir de plasmide (moins d'1 µl de solution de préparation de plasmide), de cosmide (20 ng par réaction) et de l'ADNc. Les oligonucléotides ont une taille comprise entre 18 et 27 nucléotides. Ils sont synthétisés par Eurogentec (Liège, Belgique) et sont utilisés à des concentrations de 5 pmol/µl. Les amorces sont choisies de façon à ce qu'elles ne puissent pas s'hybrider entre elles, s'apparier sur elles-mêmes (en épingle à cheveux), qu'elles ne contiennent pas de séquences répétées et qu'elles ne puissent pas s'hybrider à plusieurs endroits du matériel à amplifier. Il est préférable que la composition en bases soit équilibrée (éviter les longues répétitions de GC) et que les températures de fusion (T_m) des 2 amorces ne soient pas trop différentes l'une de l'autre. Nous avons utilisé, pour choisir les amorces et calculer les T_m , le programme informatique PCR Plan de PC/Gene. Les réactions de PCR ont été faites sur un volume final de 50 µl avec une solution de dNTPs (Boehringer Mannheim) à 2,5 mM chaque (5 µl par réaction) dans un thermocycleur Perkin Elmer 480. Un corollaire indésirable et inattendu de la puissance d'amplification de la PCR est qu'une contamination même mineure du matériel de départ peut être amplifiée. Nous avons donc préparé les réactions sous une hotte à flux laminaire en utilisant des embouts cotonnés pour les pipettes automatiques. Nous avons aussi fait des témoins négatifs en remplaçant le matériel nucléique par de l'eau. L'enzyme thermostable employée pour les PCR "classiques" est la Taq polymérase de chez Boehringer Mannheim à 1 U/µl. Pour chaque réaction, nous avons effectué un "hot start", c'est à dire que le tube est placé dans l'appareil PCR alors que sa température est supérieure à 90°C.

XIV.3 - Réactions

XIV.3.1 - Amplification d'ADN génomique

Dans un microtube on mélange 200 ng d'ADN génomique ou 20 ng de cosmide, 10 pmoles de chaque amorce, 5 µl de dNTPs (10 mM total), 5 µl de tampon 10X fourni (Tris-HCl, 100 mM ; MgCl₂, 15 mM ; KCl, 500 mM, pH 8,3), de l'eau stérile (qsp 48 µl) et 2 unités de Taq polymérase (Boehringer Mannheim). On recouvre le mélange avec 60 µl d'huile minérale (Sigma) afin d'éviter l'évaporation. Après une dénaturation de 2 min à 94°C, on effectue 30 cycles :

94°C : 1 min (dénaturation des brins)

T_m + 2°C : 1 min (hybridation)

72°C : 30 s à 2 min (élongation) selon la taille à amplifier.

On termine par une élongation finale plus longue de 15 min à 72°C.

XIV.3.2 - Amplification d'ADNc

XIV.3.2.1 - Minibanque d'ADNc

Un mélange est réalisé contenant dans un volume final de 50 µl (qsp en eau stérile) 8 µl d'ADNc, 10 pmoles de chaque amorce, 5 µl de dNTPs (10 mM total), 5 µl de tampon 10X et 2 U de Taq polymérase. L'ADNc a été amplifié à l'aide de diverses amorces : NAU162, NAU138, NAU370, NAU224, NAU174 et NAU99 (Tableau 6). On effectue les mêmes cycles que ceux décrit plus haut.

XIV.3.2.2 - RACE-PCR

Une première réaction d'amplification est réalisée sur 5 µl d'ADNc ayant une queue poly(A) avec 1 µl (12,5 pmoles) d'une amorce "dT-anchor", 12,5 pmoles de la seconde amorce antisens et spécifique du transcrit à amplifier (choisie en amont de l'amorce utilisée pour la synthèse de l'ADNc), 5 µl de dNTPs (10 mM chaque), 2 U de Taq polymérase, 5 µl de tampon d'amplification 10X et de l'eau stérile (qsp 50 µl). 4 amorces spécifiques de *MUC4* ont été utilisées : NAU138, NAU156, NAU169 et NAU288 (Tableau 6). Une seconde réaction d'amplification est réalisée ensuite sur 1 µl de la première réaction de PCR. Les amorces utilisées sont, pour l'amorce antisens : soit la même que précédemment, soit une amorce antisens spécifique du transcrit choisie en amont de l'amorce précédente (NAU138, NAU156, NAU204, NAU289 et NAU327 ; Tableau 6) et pour l'amorce sens : l'amorce

"anchor" du kit de RACE-PCR. Le mélange réactionnel est soumis aux cycles "classiques" d'amplification, la température d'hybridation de la seconde PCR étant augmentée de 2°C par rapport à la première.

XV - CLONAGE DES FRAGMENTS AMPLIFIES

Un échantillon de la PCR est quantifié après électrophorèse en gel d'agarose contenant du BEt (0,5 µg/ml) en comparaison avec le marqueur de taille moléculaire 13i.

XV.1. – Clonage en vecteur plasmidique T/A

La Taq DNA Polymérase, comme la plupart des polymérases thermostables, ajoute un nucléotide en 3' des fragments amplifiés, le plus souvent un A. Cette activité d'addition d'un nucléotide semble indépendante de l'information contenue dans le brin copié. Une telle activité catalytique de type terminale transférase a été trouvée pour d'autres polymérases dont l'AMV [Clark 1988]. Les T-vecteurs sont pourvus d'un dT non apparié, ce qui permet une ligation facile des fragments de PCR et du vecteur. En pratique, ils peuvent être préparés de la façon suivante : le vecteur de clonage est hydrolysé par une enzyme de restriction à coupure franche (on emploie généralement les enzymes *SmaI* ou *EcoRV*) et on l'incube à 72°C avec une Taq polymérase et du dTTP. Nous avons choisi comme vecteur le plasmide pCR[®]2.1 (Invitrogen) qui a une taille de 3,9 kb.

XV.2 - Ligation

Nous avons utilisé le kit "TA cloning" (Invitrogen). Un mélange final de 10 µl contenant 0,5-2 µl du produit de PCR purifié, 1 µl de tampon de ligation 10X, 2 µl (50 ng) de vecteur pCR[®]2.1 et 1 µl d'enzyme T4 ADN ligase (4 U) est incubé à 14°C pendant une nuit.

XV.3 – Transformation bactérienne

2 µl de produit de "ligation" sont additionnés à 50 µl de bactéries InvαF' (Invitrogen) en présence de β-mercaptoéthanol (2µl). L'ensemble est laissé en contact 30 min dans la glace. On effectue ensuite un choc thermique à 42°C pendant 30 s, puis le tube est remis dans la glace pendant 2 min. 250 µl de SOC (milieu de culture riche) sont ajoutés, puis les bactéries sont placées dans l'incubateur 1 h à 37°C sous agitation à 225 rpm avant d'être étalées sur des boîtes de LB-agar ampicilline (50 µg/ml). Les bactéries sont incubées à 37°C pendant une nuit. Seules les bactéries possédant le plasmide seront résistantes à l'antibiotique et pourront pousser sur la boîte.

XV.4 - Sélection des clones recombinants

En ce qui concerne les bactéries InvoF', le principe de détection des clones recombinants est le même que pour le clonage en vecteur pKS mais ici les bactéries sont étalées sur des boîtes LB-agar contenant 50 µg/ml d'ampicilline et 30 µl d'une solution de X-gal (50 mg/ml).

XVI – CLONAGE DES REGIONS REGULATRICES

Le kit "Genome Walker™" (Clontech) permet d'obtenir des clones d'ADN génomiques en amont ou en aval d'une séquence connue telle que celle d'un ADNc [Siebert *et al.* 1995]. Le kit contient 5 banques constituées d'ADN génomique hydrolysé par 5 enzymes différentes : *EcoRV* ; *ScaI* ; *DraI* ; *PvuII* ; *SspI*. Un adaptateur a été fixé aux 2 extrémités des différents fragments. Une première réaction d'amplification est effectuée à partir d'une amorce s'hybridant à l'adaptateur (AP1) et d'une amorce spécifique du gène d'intérêt (GSP1). 1 µl d'une des 5 banques est ajouté à un mélange contenant : 5 µl de tampon 10 X, 1 µl de dNTP (10 mM chaque), 2,2 µl de Mg(OAc)₂ (25 mM), 10 pmoles d'AP1, 10 pmoles de GSP1 et 1 µl du tampon "Advantage-GC Genomic Polymerase" 50X (Clontech). Ce tampon contient un mélange de 2 ADN polymérases thermostables permettant des réactions d'amplification sur de longues distances (LD PCR). La majeure partie de l'amplification est effectuée par la *Tth* ADN polymérase alors que la seconde ADN polymérase possède une activité exonucléasique 3'→5' corrigeant ainsi les nucléotides mal incorporés. Le mélange réactionnel est soumis à différents cycles :

7 cycles :

94°C 25 s

70°C 4 min

32 cycles :

94°C 25 s

65°C 4 min

Et enfin 65°C pendant 4 min supplémentaires. Le produit de PCR est alors dilué et utilisé comme matrice pour une deuxième réaction d'amplification à partir de 2 amorces plus internes : une amorce s'hybridant à l'adaptateur (AP2) et une amorce spécifique du gène d'intérêt (GSP2). 1µl d'une dilution au 50^{ème} de la première PCR est ajouté à 49 µl d'un mélange identique au précédent (avec AP2 et GSP2 comme amorces). Le nombre de cycles est diminué : 5 au lieu de 7 et 22 au lieu de 32. Les fragments d'ADN ainsi obtenu débutent

dans une séquence connue en 5' de GSP2 et se poursuivent dans l'ADN génomique adjacent et non connu. Ces fragments sont alors clonés et analysés.

XVII – TRANSCRIPTION ET TRADUCTION *IN VITRO*

RT370-224 est un produit de RT-PCR (NAU370/NAU224 ; Tableau 6) contenant un ADNc codant un peptide dont la longueur prédite est de 34 kDa. Ce peptide comprend la séquence signal amino-terminale de 27 résidus et deux sites potentiels de *N*-glycosylation. L'ADNc correspondant est cloné en vecteur pCR2.1 sous le contrôle du promoteur T7. La transcription et la traduction *in vitro* de ce fragment ont été effectuées à l'aide du kit "TNT[®] Coupled Reticulocyte Lysate Systems" (Promega). A 0,2-2 µg d'ADN plasmidique sont ajoutés : 25 µl de lysat de réticulocytes de lapin TNT[®], 2 µl de tampon, 1 µl d'ARN polymérase (T3 ou T7), 1 µl d'un mélange d'aa ne contenant pas de méthionine, 4 µl de méthionine marquée au ³⁵S (environ 1000 Ci/mmol ; Amersham), 40 U d'inhibiteur de RNases (1 µl) et de l'eau DEPC (qsp 50 µl). Le tout est incubé pendant 1 à 2 heures à 30°C. Les protéines synthétisées sont analysées par SDS/PAGE (électrophorèse en gel de polyacrylamide/SDS) en gradient 10-30 % (p/v) et révélées par autoradiographie. Les marqueurs protéiques colorés Rainbow[™] (Amersham) dont la taille varie entre 14,3 et 220 kDa sont utilisés afin de déterminer la masse moléculaire. Les modifications co- et post-traductionnelles des protéines ainsi synthétisées (ex : *N*-glycosylation ou clivage du peptide signal) peuvent également être étudiées *in vitro* en ajoutant au mélange réactionnel ci-dessus des membranes de microsomes pancréatiques de chien (2,5 µl pour 25 µl) (Promega). Afin de déterminer l'importance de la *N*-glycosylation des produits traduits, les peptides marqués au ³⁵S ont été incubés toute la nuit à 37°C avec 10 mU d'endoH (Boehringer Mannheim) dans un tampon contenant du SDS (0,02 %) et du citrate de sodium (0,1 M, pH 5,5). L'endoH clive entre les deux résidus de GlcNAc internes des complexes oligosaccharidiques riches en mannose. Après migration en gel SDS, le produit glycosylé ainsi clivé acquiert un poids moléculaire apparent comparable à celui du produit non-glycosylé analysé sur le même gel.

COMPOSITION DES TAMPONS**Denhardt's 50X**

Ficoll 400	1 g
Polyvinylpyrrolidone	1 g
Sérum albumine bovine	1 g
H ₂ O stérile	qsp 100 ml
Filtrer sur filtre 0,22 µm	

Eau DEPC

DEPC	1 ml
H ₂ O désionisée	qsp 1 l
Ajouter le DEPC sous hotte, agiter vivement, laisser au minimum une nuit avant autoclavage	

Formamide désionisée

Résine AG 501X8	5 g
Formamide	1 ml
Sous agitation, à l'abri de la lumière. Récupérer la formamide et recommencer avec 5 g de résine neuve	

LB

Bactotryptone	10 g
Extrait de levures	5 g
NaCl	5 g
H ₂ O	qsp 1 l
pH 7,2	

LB-Agar

LB contenant 15 g d'Agar pour 1 l

MOPS 10X

MOPS	10 g
Acétate de sodium 3M	4,2 ml
EDTA 0,5 M pH 8,0	5 ml
H ₂ O DEPC	qsp 250 ml
Ajuster à pH 7,0 (avec NaOH 3M) et autoclaver	

SOC

Bactotryptone (poids/volume)	2 %
Extrait de levures (poids/volume)	0,5 %
NaCl	10 mM
KCl	25 mM
MgCl ₂	10 mM
MgSO ₄	10 mM
Glucose	20 mM

Solution d'IPTG

IPTG	0,12 g
H ₂ O stérile	qsp 5 ml

Solution de X-Gal

X-Gal	50 mg
N,N' diméthylformamide	qsp 1 ml

SSC 20X

NaCl	175,3 g
Citrate trisodique	88,2 g
H ₂ O	qsp 1 l
Ajuster à pH 7,0	

SSPE 20X

NaCl	174 g
NaH ₂ PO ₄ , H ₂ O	27,6 g
EDTA	7,4 g
H ₂ O stérile	qsp 1 l
Ajuster à pH 7,4 (avec NaOH 10N)	

STE 10X

NaCl 3 M	166,6 ml
Tris-HCl 1 M pH 8,0	50 ml
EDTA 0,5 M	10 ml
H ₂ O	qsp 500 ml

Tampon Borate 10X

Acide borique	55 g
Tris	108 g
EDTA	9,3 g
H ₂ O	qsp 1 l
pH 8,3	

Tampon Borate 10X (séquence)

Acide borique	55 g
Tris	108 g
EDTA	6,72 g
H ₂ O	qsp 1 l
pH 8,3	

Tampon de purification des produits de PCR

KCl	372,8 mg
Tris-HCl, pH 8,8	121,14 mg
MgCl ₂ , 6 H ₂ O	30,50 mg
Triton X-100	100 mg
H ₂ O stérile	qsp 100 ml

Tampon de préhybridation d'une sonde nucléique (Southern blot)

SSC 20X	30 ml
Denhardt's 50X	10 ml
SDS 10 %	5 ml
H ₂ O stérile	qsp 100 ml

Tampon d'hybridation d'une sonde nucléique (Southern blot)

SSC 20X	30 ml
Denhardt's 50X	10 ml
SDS 10 %	5 ml
ssDNA	2,5 ml (25 mg)
Sulfate de dextran 50 %	10 g
H ₂ O stérile	qsp 100 ml

Tampon de préhybridation et d'hybridation ARN Clontech

SSPE 20X	25 ml
Denhardt's 50X	20 ml
ssDNA	1 ml (10 mg)
Formamide désionisée	50 ml
SDS	2 g
H ₂ O stérile	qsp 100 ml

Tampon de préhybridation et d'hybridation d'une sonde oligonucléotidique

SSPE 20X	30 ml
Denhardt's 50X	20 ml
ssDNA	1 ml (10 mg)
Sulfate de dextran	10 g
SDS 10 %	0,5 ml
H ₂ O	qsp 100 ml

Tampon d'hybridation 3X

Formamide désionisée	50 ml
EDTA 0,5 M	0,3 ml
Denhardt's 50X	2 ml
Hepes 500 mM	10 ml
SSC 20X	15 ml
ssDNA	2,5 ml (25 mg)
H ₂ O	qsp 100 ml

TE 10X

Tris	12,1 g
EDTA	2,9 g
H ₂ O stérile	qsp 100 ml
Ajuster à pH 7,0 (avec HCl 1N)	

TEA 10X

Tris	148,4 g
Acide acétique	11,42 g
H ₂ O	qsp 1 l
pH 8,0	

TMN 0,5 M

Tris-HCl	20 mM
NaCl	0,5 M
MgSO ₄	10 mM

BIBLIOGRAPHIE

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

Allen A., Pain R. H., Robson T. R. (1976) Models for structure of the gastric mucous gel. *Nature* 264 : 88-89

Allen A. (1983) Mucus - a protective secretion of complexity. *TIBS* 8 : 169-173

Allen A., Hutton D. A., Leonard A. J., Pearson J. P., Sellers L. A. (1986) The role of mucus in the protection of the gastroduodenal mucosa. *Scand. J. Gastroenterol.* 21 : 71-77

Arias E. B., Verhage H. G., Jaffe R. C. (1994) Complementary deoxyribonucleic acid cloning and molecular characterization of an estrogen-dependent human oviductal glycoprotein. *Biol. Reprod.* 51 : 685-694

Aruffo A., Dietsch M. T., Wan H., Hellstrom K. E., Hellstrom I. (1992) Granule membrane protein 140 (GMP140) binds to carcinomas and carcinoma-derived cell lines. *Proc. Natl. Acad. Sci. U S A* 89 : 2292-2296

Asker N., Baeckstrom D., Axelsson M. A. B., Carlstedt I., Hansson G. C. (1995) 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

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

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

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

Audié J. P., Tétaert D., Pigny P., Buisine M. P., Janin A., Aubert J. P., Porchet N., Boersma A. (1995) Mucin gene expression in the human endocervix. *Hum. Reprod.* 10 : 98-102

Aviv H., Leder P. (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. U S A* 69 : 1408-1412

Baekstrom D., Hansson G. C., Nilsson O., Johansson C., Gendler S. J., Lindholm L. (1991) Purification and characterization of a membrane-bound and a secreted mucin-type glycoprotein carrying the carcinoma-associated sialyl-Le^a epitope on distinct core proteins. *J. Biol. Chem.* 266 : 21537-21547

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

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

Bansil R., Stanley E., Lamont J. T. (1995) Mucin biophysics. *Ann. Rev. Physiol.* 57 : 635-657

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

Baron M., Norman D. G., Campbell I. D. (1991) Protein modules. *Trends Biochem. Sci.* 16 : 13-17

Bell G. I., Selby M. J., Rutter W. J. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* 295 : 31-35

Bell S. L., Khatri I. A., Xu G., Forstner J. F. (1998) Evidence that a peptide corresponding to the rat Muc2 C-terminus undergoes disulphide-mediated dimerization. *Eur. J. Biochem.* 253 : 123-131

Berg E. L., Robinson M. K., Mansson O., Butcher E. C., Magnani J. L. (1991) A carbohydrate domain common to both sialyl Le(a) and sialyl Le(X) is recognized by the endothelial cell leukocyte adhesion molecule ELAM-1. *J. Biol. Chem.* 266 : 14869-14872

Bird P., Gething M. J., Sambrook J. (1990) The functional efficiency of a mammalian signal peptide is directly related to its hydrophobicity. *J. Biol. Chem.* 265 : 8420-8425

Blobel G., Walter P., Chang C. N., Goldman B. M., Erickson A. H., Lingappa V. R. (1979) Translocation of proteins across membranes: the signal hypothesis and beyond. *Symp. Soc. Exp. Biol.* 33 : 9-36

Bobek L. A., Tsai H., Biesbrock A. R., Levine M. J. (1993) Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (*MUC7*). *J. Biol. Chem.* 268 : 20563-20569

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

Boshell M., Lalani E. N., Pemberton L., Burchell J., Gendler S., Taylor-Papadimitriou J. (1992) 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

Braga V. M., Pemberton L. F., Duhig T., Gendler S. J. (1992) Spatial and temporal expression of an epithelial mucin, *Muc-1*, during mouse development. *Development* 115 : 427-437

Bresalier R. S., Kim Y. S. (1989) Malignant neoplasms of the large and small intestine. Editions M. H. Sleisenger, J. S Fordtran, *Gastrointestinal Disease* pp. 1519-1560

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

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

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

Burchell J., Wang D., Taylor-Papadimitriou J. (1984) 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

Burchell J., Gendler S., Taylor-Papadimitriou J., Girling A., Lewis A., Millis R., Lampion D. (1987) Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin. *Cancer Res.* 47 : 5476-5482

Campion J. P., Porchet N., Aubert J. P., L'Helgoualc'h A., Clément B. (1995) UW-preservation of cultured human gallbladder epithelial cells : phenotypic alterations and differential mucin gene expression in the presence of bile. *Hepatology* 21 : 223-231

Carlson D. M. (1968) Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. *J. Biol. Chem.* 243 : 616-626

Carlstedt I., Lindgren H., Sheehan J. K. (1983) The macromolecular structure of human cervical-mucus glycoproteins. *Biochem. J.* 213 : 427-435

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

Carraway K. L. 3rd, Rossi E. A., Komatsu M., Price-Schiavi S. A., Huang D., Guy P. M., Carjaval M. E., Fregien N., Carraway C. A., Carraway K. L. (1999) An intramembrane modulator of the ErbB2 receptor tyrosine kinase that potentiates neuregulin signaling. *J. Biol. Chem.* 274 : 5263-5266

Carvalho F., Seruca R., David L., Amorim A., Seixas M., Bennett E., Clausen H., Sobrinho-Simões M. (1997) MUC1 gene polymorphism and gastric cancer – an epidemiological study. *Glycoconj. J.* 14 : 107-111

Chang S. K., Dohrman A. F., Basbaum C. B., Ho S. B., Tsuda T., Toribara N. W., Gum J. R., Kim Y. S. (1994) Localization of mucin (MUC2 and MUC3) messenger RNA and peptidic expression in human normal intestine and colon cancer. *Gastroenterology* 107 : 28-36

Charest P. M., Roth J. (1985) Localization of sialic acid in kidney glomeruli: regionalization in the podocyte plasma membrane and loss in experimental nephrosis. *Proc. Natl. Acad. Sci. U S A* 82 : 8508-8512

Chirgwin J. M., Przybyla A. E., MacDonald R. J., Rutter W. J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18 : 5294-5299

Clark J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* 16 : 9677-9686

Clements S., Mehansho H., Carlson D. M. (1985) Novel multigene families encoding highly repetitive peptide sequences. Sequence analyses of rat and mouse proline-rich protein cDNAs. *J. Biol. Chem.* 260 : 13471-13477

Crépin M., Porchet N., Aubert J. P., Degand P. (1990) Diversity of the peptide moiety of human mucins. *Biorheology* 27 : 471-484

Debailleul V. Expression des gènes de mucines humaines. Etude de la stabilité, de l'hétérogénéité et du polymorphisme des ARNm. Thèse d'Université, Lille, 1997

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

Desseyn J. L. Organisation du gène de mucine humaine *MUC5B*. Bases moléculaires d'une nouvelle classification des mucines. Thèse d'Université, Lille, 1997a

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

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

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

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

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

Devine P. L., Warren J. A., Ward B. G., McKenzie I. F. C., Layton G. T. (1990) Glycosylation and the exposure of tumor associated epitopes on mucins. *J. Tumor Marker Oncol.* 5 : 11-26

Devine P. L., McKenzie I. F. C. (1992) Mucins : structure, function, and associations with malignancy. *Bioessays* 14 : 619-625

Dickinson D. P., Ridall A. L., Levine M. J. (1987) Human submandibular gland statherin and basic histidine-rich peptide are encoded by highly abundant mRNA's derived from a common ancestral sequence. *Biochem. Biophys. Res. Commun.* 149 : 784-790

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

- Eckhardt A. E., Timpfe C. S., DeLuca A. W., Hill R. L.** (1997) The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin. *J. Biol. Chem.* 272 : 33204-33210
- Elhammer A. P., Poorman R. A., Brown E., Maggiora L. L., Hoogerheide J. G., Kezdy F. J.** (1993) The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides. *J. Biol. Chem.* 268 : 10029-10038
- Elstein M.** (1978) Functions and physical properties of mucus in the female genital tract. *Br. Med. Bull.* 34 : 83-88
- Escande F.** Caractérisation de la région 5' du gène MUC5AC. DEA Sciences de la Vie et de la Santé, Lille, 1998
- Fahim R. E. F., Specian R. D., Forstner G. G., Forstner J. F.** (1987) Characterization and localization of the putative "link" component in rat small intestinal mucin. *Biochem. J.* 243 : 631-640
- Fox M. F., Lahbib F., Pratt W., Attwood J., Gum J., Kim Y., Swallow D. M.** (1992) Regional localization of the intestinal mucin gene MUC3 to chromosome 7q22. *Ann. Hum. Genet.* 56 : 281-287
- Frohman M. A., Dush M. K., Martin G. R.** (1988) Rapid production of full-length cDNAs from rare transcripts : amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85 : 8998-9002
- Funahashi J., Sekido R., Murai K., Kamachi Y., Kondoh H.** (1993) Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. *Development* 119 : 433-446
- Fung P. Y., Longenecker B. M.** (1991) Specific immunosuppressive activity of epiglycanin, a mucin-like glycoprotein secreted by a murine mammary adenocarcinoma (TA3-HA). *Cancer Res.* 51 : 1170-1176
- Gaillard D., Plotkowski C., Puchelle E.** (1992) Mucus et protection de la muqueuse respiratoire. Edition et communications médicales pp. 23-33
- Gaye P., Mercier J. C.** (1981) The functional significance of signal sequences: their role in the transfer of lactoproteins across the endoplasmic reticulum membrane. *Reprod. Nutr. Dev.* 21 : 199-208

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

Gendler S. J., Taylor-Papadimitriou J., Duhig T., Rothbard J., Burchell J. (1988) 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

Gendler S. J., Lancaster C. A., Taylor-Papadimitriou J., Duhig J., Peat T., Burchell J. M., Pemberton L., Lalani E. N., Wilson D. (1990) Molecular cloning and expression of a human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* 265 : 15286-15293

Gendler S. J., Spicer A. P., Lalani E. L., Duhig T., Peat N., Burchell J., Pemberton L., Boshell M., Taylor-Papadimitriou J. (1991) Structure and biology of a carcinoma-associated mucin, MUC1. *Am. Rev. Respir. Dis.* 144 : S42-S47

Gendler S. J., Spicer A. P. (1995) Epithelial mucin genes. *Annu. Rev. Physiol.* 57 : 607-634

Gerard C., Eddy R. L., Shows T. B. (1990) The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat structure. Evidence for a common mucin in airway and gastrointestinal tissue. *J. Clin. Invest.* 86 : 1921-1927

Gipson I. K., Ho S. B., Spurr-Michaud S. J., Tisdale A. S., Zhan Q., Torlakovic E., Pudney J., Anderson D. J., Toribara N. W., Hill J. A. 3rd (1997a) Mucin genes expressed by human female reproductive tract epithelia. *Biol. Reprod.* 56 : 999-1011

Gipson I. K., Inatomi T. (1997b) Mucin genes expressed by the ocular surface epithelium. Editions Elsevier Science Ltd, *Progress in Retinal and Eye Research* 16 : 81-98

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

Girling A., Bartkova J., Burchell J., Gendler S., Gillett C., Taylor-Papadimitriou J. (1989) A core protein epitope of the polymorphic epithelial mucin detected by the monoclonal antibody SM-3 is selectively exposed in a range of primary carcinomas. *Int. J. Cancer* 43 : 1072-1076

Gollub E. G., Goswami S., Kouba D., Marom Z. (1993) Regulation of mucin gene expression in secretory epithelial cells. *Biochem. Biophys. Res. Commun.* 197 : 667-73

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

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

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

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

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

Gum J. R. (1992a) Mucin genes and the proteins they encode : structure, diversity, and regulation. *Am. J. Respir. Cell. Mol. Biol.* 7 : 557-564

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

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

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

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

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

Haeuptle M. T., Flint N., Gough N. M., Dobberstein B. (1989) A tripartite structure of the signals that determine protein insertion into the endoplasmic reticulum membrane. *J. Cell Biol.* 108 : 1227-1236

Handa K., Nudelman E. D., Stroud M. R., Shiozawa T., Hakomori S. (1991) Selectin GMP-140 (CD62; PADGEM) binds to sialosyl-Le(a) and sialosyl-Le(x), and sulfated glycans modulate this binding. *Biochem. Biophys. Res. Commun.* 181 : 1223-1230

Hanski C., Drechsler K., Hanisch F. G., Sheehan J., Manske M., Ogorek D., Klusmann E., Hanski M. L., Blank M., Xing P. X., McKenzie I. F. C., Devine P. L., Riecken E. O. (1993) 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

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

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

Hewetson A., Chilton B. S. (1997) Molecular cloning and hormone-dependent expression of rabbit Muc1 in the cervix and uterus. *Biol. Reprod.* 57 : 468-477

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

Hilkens J., Kroezen V., Bonfrer J. M., De Jong-Bakker M., Bruning P. F. (1986) MAM-6 antigen, a new serum marker for breast cancer monitoring. *Cancer Res.* 46 : 2582-2587

Hilkens J., Buijs F. (1988) Biosynthesis of MAM-6, an epithelial sialomucin. Evidence for involvement of a rare proteolytic cleavage step in the endoplasmic reticulum. *J. Biol. Chem.* 263 : 4215-4222

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

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

Hollingsworth M. A., Closken C., Harris A., McDonald C. D., Pahwa G. S., Maher L. J. 3rd (1994a) 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

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

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

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

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

Inatomi T., Tisdale A. S., Zhan Q., Spurr-Michaud S., Gipson I. K. (1997) Cloning of rat Muc5AC mucin gene : comparison of its structure and tissue distribution to that of human and mouse homologues. *Biochem. Biophys. Res. Commun.* 236 : 789-797

Ioannides C. G., Fisk B., Jerome K. R., Irimura T., Wharton J. T., Finn O. J. (1993) Cytotoxic T cells from ovarian malignant tumors can recognize polymorphic epithelial mucin core peptides. *J. Immunol.* 151 : 3693-3703

Irimura T., McIsaac A. M., Carlson D. A., Yagita M., Grimm E. A., Menter D. G., Ota D. M., Clary K. R. (1990) Soluble factor in normal tissues that stimulates high-molecular-weight sialoglycoprotein production by human colon carcinoma cells. *Cancer Res.* 50 : 3331-3338

Jany B. H., Gallup M. W., Yan P. S., Gum J. R., Kim Y. S., Basbaum C. B. (1991) Human bronchus and intestine express the same mucin gene. *J. Clin. Invest.* 87 : 77-82

Jeanpierre M. (1987) A rapid method for the purification of DNA from blood. *Nucleic Acids Res.* 15 : 9611

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

Jerome K. R., Domenech N., Finn O. J. (1993) Tumor-specific cytotoxic T cell clones from patients with breast and pancreatic adenocarcinoma recognize EBV-immortalized B cells transfected with polymorphic epithelial mucin complementary DNA. *J. Immunol.* 151 : 1654-1662

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

Khatri I. A., Forstner G. G., Forstner J. F. (1993) Suggestive evidence for two different mucin genes in rat intestine. *Biochem. J.* 294 : 391-399

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

Khatri I. A., Forstner G. G., Forstner J. F. (1998) Susceptibility of the cysteine-rich N-terminal and C-terminal ends of rat intestinal mucin muc 2 to proteolytic cleavage. *Biochem. J.* 331 : 323-330

Kindon H., Pothoulakis C., Thim L., Lynch-Devaney K., Podolsky D. K. (1995) Trefoil peptide protection of intestinal epithelial barrier function : cooperative interaction with mucin glycoprotein. *Gastroenterology* 109 : 516-523

Klings-Levan K., Gum J. R., Gendler S. J., Kim Y. S., Hansson G. C. (1996) Chromosomal mapping of three mucin genes in the rat. *Mammalian Genome* 7 : 248-250

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

Komatsu M., Carraway C. A., Fregien N. L., Carraway K. L. (1997) Reversible disruption of cell-matrix and cell-cell interactions by overexpression of sialomucin complex. *J. Biol. Chem.* 272 : 33245-33254

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

Kovarik A., Lu P. J., Peat N., Morris J., Taylor-Papadimitriou J. (1996) Two GC boxes (Sp1 sites) are involved in regulation of the activity of the epithelium-specific MUC1 promoter. *J. Biol. Chem.* 271 : 18140-18147

Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15 : 8125-8148

Kreil G. (1981) Transfer of proteins across membranes. *Annu. Rev. Biochem.* 50 : 317-348

Krontiris T. G., DiMartino N. A., Colb M., Parkinson D. R. (1985) Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumour DNAs of cancer patients. *Nature* 313 : 369-374

Krontiris T. G., Devlin B., Karp D. D., Robert N. J., Risch N. (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N. Engl. J. Med.* 329 : 517-523

Kyo K., Parkes M., Takei Y., Nishimori H., Vyas P., Satsangi J., Simmons J., Nagawa H., Baba S., Jewell D., Muto T., Lathrop G. M., Nakamura Y. (1999) Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3. *Hum. Mol. Genet.* 8 :307-311

Kyte J., Doolittle R. F. (1982) A simple method for displaying the hydrophathic character of a protein. *J. Mol. Biol.* 157 : 105-132

Lalani E. N., Berdichevsky F., Boshell M., Shearer M., Wilson D., Stauss H., Gendler S. J., Taylor-Papadimitriou J. (1991) Expression of the gene coding for a human mucin in mouse mammary tumor cells can affect their tumorigenicity. *J. Biol. Chem.* 266 : 15420-15426

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

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

Lancaster C. A., Peat N., Duhig T., Wilson D., Taylor-Papadimitriou J., Gendler S. (1990) Structure and expression of the human polymorphic epithelial mucin gene : an expressed VNTR unit. *Biochem. Biophys. Res. Commun.* 173 : 1019-1029

Lapensée L., Paquette Y., Bleau G. (1997) Allelic polymorphism and chromosomal localization of the human oviductin gene (MUC9). *Fertil. Steril.* 68 : 702-708

Lasky L. A., Singer M. S., Dowbenko D., Imai Y., Henzel W. J., Grimley C., Fennie C., Gillett N., Watson S. R., Rosen S. D. (1992) An endothelial ligand for L-selectin is a novel mucin-like molecule. *Cell* 69 : 927-938

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

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

Li J. D., Dohrman A. F., Gallup M., Miyata S., Gum J. R., Kim Y. S., Nadel J. A., Prince A., Basbaum C. B. (1997) Transcriptional activation of mucin by *Pseudomonas aeruginosa* lipopolysaccharide in the pathogenesis of cystic fibrosis lung disease. *Proc. Natl. Acad. Sci. U S A* 94 : 967-972

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

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

Ligtenberg M. J. L., Gennissen A. M. C., Vos H. L., Hilkens J. (1991) A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episialin mRNA. *Nucleic Acids Res.* 19 : 297-301

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

Ligtenberg M. J. L., Kruijshaar L., Buijs F., Van Meijer M., Litvinov S. V., Hilkens J. (1992b) Cell-associated episialin is a complex containing two proteins derived from a common precursor. *J. Biol. Chem.* 267 : 6171-6177

Linsley P. S., Kallestad J. C., Horn D. (1988) Biosynthesis of high molecular weight breast carcinoma associated mucin glycoproteins. *J. Biol. Chem.* 263 : 8390-8397

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

Lo-Guidice J. M., Merten M. D., Lamblin G., Porchet N., Houvenaghel M. C., Figarella C., Roussel P., Perini J. M. (1997) Mucins secreted by a transformed cell line derived from human tracheal gland cells. *Biochem. J.* 326 : 431-437

Lucassen A. M., Julier C., Beressi J. P., Boitard C., Froguel P., Lathrop M., Bell J. I. (1993) Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat. Genet.* 4 : 305-310

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

Mantle M., Forstner G. G., Forstner J. F. (1984) Antigenic and structural features of goblet cell mucin of human small intestine. *Biochem. J.* 217 : 159-167

- Mantle M., Allen A.** (1989) Gastrointestinal mucus. Editions J. S. Davison, Wright, *Gastrointestinal secretion* pp. 202-229
- Mayadas T. N., Wagner D. D.** (1992) Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc. Natl. Acad. Sci. USA* 89 : 3531-3535
- McCool D. J., Forstner J. F., Forstner G. G.** (1995) Regulated and unregulated pathways for MUC2 mucin secretion in human colonic LS180 adenocarcinoma cells are distinct. *Biochem. J.* 312 : 125-133
- McGuckin M. A., Walsh M. D., Hohn B. G., Ward B. G., Wright R. G.** (1995) Prognostic significance of MUC1 epithelial mucin expression in breast cancer. *Hum. Pathol.* 26 : 432-439
- McNeer R. R., Price-Schiavi S., Komatsu M., Fregien N. R., Carraway K. L.** (1997) Sialomucin complex in tumors and tissues. *Front. Biosci.* 2 : d449-d459
- McNeer R. R., Huang D., Fregien N. L., Carraway K. L.** (1998) Sialomucin complex in the rat respiratory tract : a model for its role in epithelial protection. *Biochem. J.* 330 : 737-744
- Meerzaman D., Charles P., Daskal E., Polymeropoulos M. H., Martin B. M., Rose M. C.** (1994) Cloning and analysis of cDNA encoding a major airway glycoprotein human tracheobronchial mucin (MUC5). *J. Biol. Chem.* 269 : 12932-12939
- Meloni R., Albanese V., Ravassard P., Treilhou F., Mallet J.** (1998) A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. *Hum. Mol. Genet.* 7 : 423-428
- 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., Seigler H. F.** (1984) Detection of a pancreatic cancer-associated antigen (DU-PAN-2 antigen) in serum and ascites of patients with adenocarcinoma. *Proc. Natl. Acad. Sci. USA* 81 : 5242-5246
- Middleton-Price H., Gendler S., Malcolm S.** (1988) Close linkage of PUM and SPTA within chromosome band 1q21. *Ann. Hum. Genet.* 52 : 273-278
- Moniaux N.** Caractérisation de l'extrémité C-terminale de MUC4 (apomucine humaine) : formes solubles et membranaires, approche des relations structure/fonctions. Thèse d'Université, Lille, 1999a
- Moniaux N., Nollet S., Porchet N., Degand P., Laine A., Aubert J. P.** (1999b) Complete sequence of the human mucin MUC4: a putative cell membrane-associated mucin. *Biochem. J.* 338 : 325-333

- Moriarty J., Skelly C. M., Bharathan S., Moody C. E., Sherblom A. P.** (1990) Sialomucin and lytic susceptibility of rat mammary tumor ascites cells. *Cancer Res.* 50 : 6800-6805
- Mount S. M.** (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* 10 : 459-772
- Nakamori S., Ota D. M., Cleary K. R., Shirotani K., Irimura T.** (1994) MUC1 mucin expression as a marker of progression and metastasis of human colorectal carcinoma. *Gastroenterology* 106 : 353-361
- Nakamura Y., Koyama K., Matsushima M.** (1998) VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. *J. Hum. Genet.* 43 : 149-152
- Neutra M. R., Forstner J. F.** (1987) Gastrointestinal mucus : synthesis, secretion, and function. Editions L. R. Johnson. New York : Raven, *Proceedings of the Gastrointestinal Tract* pp 975-1009
- Nguyen P. L., Niehans G. A., Cherwitz D. L., Kim Y. S., Ho S. B.** (1996) Membrane-bound (MUC1) and secretory (MUC2, MUC3, and MUC4) mucin gene expression in human lung cancer. *Tumour Biol.* 17 : 176-192
- Nguyen V. C., Aubert J. P., Gross M. S., Porchet N., Degand P., Freizal J.** (1990) Assignment of human tracheobronchial mucin gene(s) to 11p15 and a tracheobronchial mucin-related sequence to chromosome 13. *Hum. Genet.* 86 : 167-172
- Nielsen P. A., Bennett E. P., Wandall H. H., Therkildsen M. H., Hannibal J., Clausen H.** (1997) Identification of a major human high molecular weight salivary mucin (MG1) as tracheobronchial mucin MUC5B. *Glycobiology* 7 : 413-419
- Nollet S., Moniaux N., Maury J., Petitprez D., Degand P., Laine A., Porchet N., Aubert J. P.** (1998) Human mucin gene MUC4: organization of its 5'-region and polymorphism of its central tandem repeat array. *Biochem. J.* 332 : 739-748
- Offner G. D., Nunes D. P., Keates A. C., Afdhal N. H., Troxler R. F.** (1998) The amino-terminal sequence of MUC5B contains conserved multifunctional D domains : implications for tissue-specific mucin functions. *Biochem. Biophys. Res. Commun.* 251 : 350-355
- Ogata S., Uehara H., Chen A., Itzkowitz S. H.** (1992) Mucin gene expression in colonic tissues and cell lines. *Cancer Res.* 52 : 5971-5978
- Ogilvie A. D., Battersby S., Bubb V. J., Fink G., Harmar A. J., Goodwin G. M., Smith C. A.** (1996) Polymorphism in serotonin transporter gene associated with susceptibility to major depression. *Lancet* 347 : 731-733

Ohmori H., Dohrman A. F., Gallup M., Tsuda T., Kai H., Gum J. R., Kim Y. S., Basbaum C. B. (1994) Molecular cloning of the amino-terminal region of a rat MUC2 mucin gene homologue. Evidence for expression in both intestine and airway. *J. Biol. Chem.* 269 : 17833-17840

Oosterkamp H. M., Scheiner L., Stefanova M. C., Lloyd K. O., Finstad C. L. (1997) Comparison of MUC-1 mucin expression in epithelial and non-epithelial cancer cell lines and demonstration of a new short variant form (MUC-1/Z). *Int. J. Cancer* 72 : 87-94

Parmley R. R., Gendler S. J. (1998) Cystic fibrosis mice lacking Muc1 have reduced amounts of intestinal mucus. *J. Clin. Invest.* 102 : 1798-1806

Parry G., Beck J. C., Moss L., Bartley J., Ojakian G. K. (1990) Determination of apical membrane polarity in mammary epithelial cell cultures : the role of cell-cell, cell-substratum, and membrane-cytoskeleton interactions. *Exp. Cell. Res.* 188 : 302-311

Patthy L. (1987) Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* 214 : 1-7

Peat N., Gendler S. J., Lalani N., Duhig T., Taylor-Papadimitriou J. (1992) Tissue-specific expression of a human polymorphic epithelial mucin (MUC1) in transgenic mice. *Cancer Res.* 52 : 1954-1960

Pemberton L. F., Taylor-Papadimitriou J., Gendler S. J. (1992) Antibodies to the cytoplasmic domain of the MUC1 mucin show conservation throughout mammals. *Biochem. Biophys. Res. Commun.* 185 : 167-175

Pemberton L. F., Rughetti A., Taylor-Papadimitriou J., Gendler S. J. (1996) The epithelial mucin MUC1 contains at least two discrete signals specifying membrane localization in cells. *J. Biol. Chem.* 271 : 2332-2340

Perez-Vilar J., Eckhardt A. E., Hill R. L. (1996) Porcine submaxillary mucin forms disulfide-bonded dimers between its carboxyl-terminal domains. *J. Biol. Chem.* 271 : 9845-9850

Perez-Vilar J., Hill R. L. (1998) The carboxyl-terminal 90 residues of porcine submaxillary mucin are sufficient for forming disulfide-bonded dimers. *J. Biol. Chem.* 273 : 6982-6988

Pigny P., Pratt W. S., Laine A., Leclercq A., Swallow D. M., Nguyen V. C., Aubert J. P., Porchet N. (1995) The MUC5AC gene : RFLP analysis with the JER58 probe. *Hum. Genet.* 96 : 367-368

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

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

Porchet N., Nguyen V. C., Dufossé J., Audié J. P., Guyonnet Dupérat V., Gross M. S., Denis C., Degand P., Bernheim, A., Aubert J. P. (1991) 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

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

Quandt K., Frech K., Karas H., Wingender E., Werner T. (1995) MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research.* 23 : 4878-4884

Reddy M. S., Bobek L. A., Haraszthy G. G., Biesbrock A. R., Levine M. J. (1992) Structural features of the low-molecular-mass human salivary mucin. *Biochem. J.* 287 : 639-643

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

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

Rose M. C., Kauffman B., Martin B. M. (1989) Proteolytic fragmentation and peptidic mapping of human carboxyamidomethylated tracheobronchial mucin. *J. Biol. Chem.* 264 : 8193-8199

Rose M. C. (1992) Mucins : structure, function, and role in pulmonary diseases. *Am. J. Physiol.* 263 : L413-L429

- Rossi E. A., McNeer R. R., Price-Schiavi S. A., Van den Brande J. M. H., Komatsu M., Thompson J. F., Carothers Carraway C. A., Fregien N. L., Carraway K. L.** (1996) Sialomucin complex, a heterodimeric glycoprotein complex. Expression as a soluble, secretable form in lactating mammary gland and colon. *J. Biol. Chem.* 271 : 33476-33485
- Roussel P., Lamblin G., Lhermitte M., Houdret N., Laffite J. J., Perini J. M., Klein A., Scharfman A.** (1988) The complexity of mucins. *Biochimie* 70 : 1471-1482
- Roussel P., Lamblin G.** (1996) Human mucosal mucins in diseases. Editions J. Montreuil, J. F. G. Vlieghehart, H. Schachter, *Glycoproteins and disease* pp. 351-393
- Sakaguchi M., Tomiyoshi R., Kuroiwa T., Mihara K., Omura T.** (1992) Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. *Proc. Natl. Acad. Sci. U S A* 89 : 16-19
- Sanger F., Nicklen S., Coulson A. R.** (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U S A* 74 : 5463-5467
- Sekido R., Murai K., Funahashi J., Kamachi Y., Fujisawa-Sehara A., Nabeshima Y., Kondoh H.** (1994) The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol. Cell. Biol.* 14 : 5692-5700
- Shankar V., Gilmore M. S., Elkins R. C., Sachdev G. P.** (1994) A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization. *Biochem. J.* 300 : 295-298
- Shankar V., Gilmore M. S., Sachdev G. P.** (1995) Further evidence that the human MUC2 gene transcripts in the intestine and trachea are identical. *Biochem. J.* 306 : 311-312
- Shankar V., Pichan P., Eddy R. L., Tonk V., Nowak N., Sait S. N. J., Shows T. B., Schultz R. E., Gotway G., Elkins R. C., Gilmore M. S., Sachdev G. P.** (1997) Chromosomal localization of a human mucin gene (MUC8) and cloning of the cDNA corresponding to the carboxy terminus. *Am. J. Respir. Cell. Mol. Biol.* 16 : 232-241
- Sheehan J. K., Oates K., Carlstedt I.** (1986) Electron microscopy of cervical, gastric and bronchial mucus glycoproteins. *Biochem. J.* 239 : 147-153
- Sheehan J. K., Thornton D. J., Howard M., Carlstedt I., Corfield A. P., Paraskeva C.** (1996) Biosynthesis of the MUC2 mucin : evidence for a slow assembly of fully glycosylated units. *Biochem. J.* 315 : 1055-1060
- Shekels L. L., Lyftogt C., Kieliszewski M., Filie J. D., Kozak C. A., Ho S. B.** (1995) Mouse gastric mucin : cloning and chromosomal localization. *Biochem. J.* 311 : 775-785

- Shekels L. L., Hunninghake D. A., Tisdale A. S., Gipson I. K., Kieliszewski M., Kozak C. A., Ho S. B.** (1998) Cloning and characterization of mouse intestinal MUC3 mucin : 3' sequence contains epidermal-growth-factor-like domains. *Biochem. J.* 330 : 1301-1308
- Sheng Z. Q., Hull S. R., Carraway K. L.** (1990) Biosynthesis of the cell surface sialomucin complex of ascites 13762 rat mammary adenocarcinoma cells from a high molecular weight precursor. *J. Biol. Chem.* 265 : 8505-8510
- Sheng Z. Q., Wu K., Carraway K. L., Fregien N. L.** (1992) Molecular cloning of the transmembrane component of the 13762 mammary adenocarcinoma sialomucin complex. A new member of the epidermal growth factor superfamily. *J. Biol. Chem.* 267 : 16341-16346
- Sherblom A. P., Carraway K. L.** (1980) A complex of two cell surface glycoproteins from ascites mammary adenocarcinoma cells. *J. Biol. Chem.* 255 : 12051-12059
- Shimizu M., Yamauchi K.** (1982) Isolation and characterization of mucin-like glycoprotein in human milk fat globule membrane. *J. Biochem. (Tokyo)* 91 : 515-524
- Shimizu Y., Shaw S.** (1993) Mucins in the mainstream. *Nature* 366 : 630
- Shirotani K., Taylor-Papadimitriou J., Gendler S. J., Irimura T.** (1994) 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
- Siebert P. D., Chenchik A., Kellogg D. E., Lukyanov K. A., Lukyanov S. A.** (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23 : 1087-1088
- Slyter H. S., Lamblin G., Letreut A., Galabert C., Houdret N., Degand P., Roussel P.** (1984) Complex structure of human bronchial mucus glycoprotein. *Eur. J. Biochem.* 142 : 209-218
- Spicer A. P., Parry G., Patton S., Gendler S. J.** (1991) 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
- Steck P. A., Nicolson G. L.** (1983) Cell surface glycoproteins of 13762NF mammary adenocarcinoma clones of differing metastatic potentials. *Exp. Cell. Res.* 147 : 255-267
- Strous G.J., Dekker J.** (1992) Mucin-type glycoproteins. *Crit. Rev. Biochem. & Mol. Biol.* 27 : 57-92
- Swallow D. M., Gendler S., Griffiths B., Corney G., Taylor-Papadimitriou J., Bramwell M. E.** (1987a) The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. *Nature* 328 : 82-84

Swallow D. M., Gendler S., Griffiths B., Kearney A., Poveys S., Sheer D., Taylor-Papadimitriou J. (1987b) 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

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

Taylor-Papadimitriou J., Peterson J. A., Arklie J., Burchell J., Ceriani R. L., Bodmer W. F. (1981) Monoclonal antibodies to epithelium-specific components of the human milk fat globule membrane : production and reaction with cells in culture. *Int. J. Cancer* 28 : 17-21

Thornton D. J., Sheehan J. K., Lindgren H., Carlstedt I. (1991) Mucus glycoproteins from cystic fibrotic sputum. Macromolecular properties and structural "architecture". *Biochem. J.* 276 : 667-675

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

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

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

Trepicchio W. L., Krontiris T. G. (1992) Members of the rel/NF-kappa B family of transcriptional regulatory proteins bind the HRAS1 minisatellite DNA sequence. *Nucleic Acids Res.* 20 : 2427-2434

Trepicchio W. L., Krontiris T. G. (1993) IGH minisatellite suppression of USF-binding-site- and E mu-mediated transcriptional activation of the adenovirus major late promoter. *Nucleic Acids Res.* 21 : 977-985

Troxler R. F., Offner G. D., Zhang F., Iontcheva I., Oppenheim G. O. (1995) Molecular cloning of a novel high molecular mucin (MG1) from human sublingual gland. *Biochem. Biophys. Res. Commun.* 217 : 1112-1119

Van de Bovenkamp J. H., Hau C. M., Strous G. J., Buller H. A., Dekker J., Einerhand A. W. (1998) 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

Van de Wiel-Van Kemenade E., Ligtenberg M. J., de Boer A. J., Buijs F., Vos H. L., Melief C. J., Hilkens J., Figdor C. G. (1993) Episialin (MUC1) inhibits cytotoxic lymphocyte-target cell interaction. *J. Immunol.* 151 : 767-776

Van Klinken B. J. W.; Dekker J., Büller H. A., Einerhand A. W. C. (1995) Mucin gene structure and expression : protection vs. adhesion. *Am. J. Physiol.* 269 : G613-G627

Van Klinken B. J. W., Oussoren E., Weenink J. J., Strous G. J., Buller H. A., Dekker J., Einerhand A. W. C. (1996) The human intestinal cell lines Caco-2 and LS174T as models to study cell-type specific mucin expression. *Glycoconj. J.* 13 : 757-768

Van Klinken B. J. W., Van Dijken T. C., Oussoren E., Buller H. A., Dekker J., Einerhand A. W. C. (1997) Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem repeat region. *Biochem. Biophys. Res. Commun.* 238 : 143-148

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

Van Klinken B. J. W., Einerhand A. W. C., Duits L. A., Makkink M. K., Tytgat K. M. A. J., Renes I. B., Verburg M., Büller H. A., Dekker J. (1999) Gastrointestinal expression and partial cDNA cloning of murine muc2. *Am. J. Physiol.* 276 : G115-G124

Van Seuningen I., Perrais M., Pigny P., Desseyn J. L., Porchet N., Aubert J. P., Laine A. (1999) Transcription of the human MUC5B mucin gene is initiated at different promoters and is regulated by Sp1. *Soumis à publication*

Vandehaute B., Buisine M. P., Debailleul V., Clément B., Moniaux N., Dieu M. C., Degand P., Porchet N., Aubert J. P. (1997) Mucin gene expression in biliary epithelial cells. *J. Hepatol.* 27 : 1057-1066

Velcich A., Palumbo L., Selleri L., Evans G., Augenlicht L. (1997) Organization and regulatory aspects of the human intestinal mucin gene (MUC2) locus. *J. Biol. Chem.* 272 : 7968-7976

Verweij C. L., Hart M., Pannekoek H. (1988) Proteolytic cleavage of the precursor of von Willebrand factor is not essential for multimer formation. *J. Biol. Chem.* 263 : 7921-7924

Vinall L. E., Hill A. S., Pigny P., Pratt W. S., Toribara N., Gum J. R., Kim Y. S., Porchet N., Aubert J. P., Swallow D. M. (1998) Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. *Hum. Genet.* 102 : 357-366

Voorberg J., Fontijn R., Van Mourik J. A., Pannekoek H. (1990) Domains involved in multimer assembly of von willebrand factor (vWF): multimerization is independent of dimerization. *EMBO J.* 9 : 797-803

Voorberg J., Fontijn R., Calafat J., Janssen H., Van Mourik J. A., Pannekoek H. (1991) 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

Vos H. L., de Vries Y., Hilkens J. (1991) The mouse episialin (Muc1) gene and its promoter : rapid evolution of the repetitive domain in the protein. *Biochem. Biophys. Res. Commun.* 181 : 121-130

Vos H. L., Devarayalu S., de Vries Y., Bornstein P. (1992) Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene family. *J. Biol. Chem.* 267 : 12192-12196

Wagner D. D., Saffaripour S., Bonfanti R., Sadler J. E., Cramer E. M., Chapman B., Mayadas T. N. (1991) Induction of specific storage organelles by von Willebrand factor propolypeptide. *Cell* 64 : 403-413

Watson C. J., Gordon K. E., Robertson M., Clark A. J. (1991) Interaction of DNA-binding proteins with a milk protein gene promoter in vitro : identification of a mammary gland-specific factor. *Nucleic Acids Res.* 19 : 6603-6610

Wesley A., Mantle M., Man D., Qureshi R., Forstner G., Forstner J. (1985) Neutral and acidic species of human intestinal mucin. *J. Biol. Chem.* 260 : 7955-7959

Wesseling J., Van der Valk S. W., Vos H. L., Sonnenberg A., Hilkens J. (1995) Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components. *J. Cell. Biol.* 129 : 255-265

Williams C. J., Wreschner D. H., Tanaka A., Tsarfaty I., Keydar I., Dion A. S. (1990) Multiple protein forms of the human breast tumor-associated epithelial membrane antigen (EMA) are generated by alternative splicing and induced by hormonal stimulation. *Biochem. Biophys. Res. Commun.* 170 : 1331-1338

Williams S. E., Turnberg L. A. (1980) Retardation of acid diffusion by pig gastric mucosa : a potential role in mucosal protection. *Gastroenterology* 79 : 299-304

Wise R. J., Pittman D. D., Handin R. I., Kaufman R. J., Orkin S. H. (1988) The propeptide of von Willebrand factor independently mediates the assembly of von Willebrand multimers. *Cell* 52 : 229-236

Wreschner D. H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J., Zaretsky J., Kothes P., Weiss M., Lathe R. (1990) *Eur. J. Biochem.* 189 : 463-473

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

Xing P. X., Prenzoska J., Apostolopoulos V., Karkaloutsos J., McKenzie I. F. C. (1997) Monoclonal antibodies to a MUC4 peptide react with lung cancer. *Int. J. Oncol.* 11 : 289-295

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

Yang P., Yin X., Rutishauser U. (1992) Intercellular space is affected by the polysialic acid content of NCAM. *J. Cell. Biol.* 116 : 1487-1496

Yoshida A., Suzuki M., Ikenaga H., Takeuchi M. (1997) Discovery of the shortest sequence motif for high level mucin-type O-glycosylation. *J. Biol. Chem.* 272 : 16884-16888

Zhang K., Sikut R., Hansson G. C. (1997) A MUC1 mucin secreted from a colon carcinoma cell line inhibits target cell lysis by natural killer cells. *Cell. Immunol.* 176 : 158-65

Zotter S., Hageman P. C., Lossnitzer A., Mooi W. J., Hilgers J. (1988) Tissue and tumor distribution of human polymorphic epithelial mucin. *Cancer Rev.* 11-12 : 55-101

Zrihan-Licht S., Vos H. L., Baruch A., Elroy-Stein O., Sagiv D., Keydar I., Hilkens J., Wreschner D. H. (1994) 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

