

50376  
1993  
6

50376  
1993  
6

UNIVERSITE DES SCIENCES ET TECHNIQUES  
DE LILLE FLANDRES-ARTOIS

Année 1993

N° d'ordre : 1036

THESE de DOCTORAT D'UNIVERSITE  
présentée à l'Université de Lille I  
pour l'obtention du grade de

DOCTEUR EN SCIENCES DE LA VIE ET DE LA SANTE



par

José DUFOSSÉ



MUCINES TRACHEOBRONCHIQUES  
HUMAINES  
LOCALISEES EN 11 P 15  
MISE EN EVIDENCE D'UN NOUVEAU TYPE  
D'ORGANISATION PEPTIDIQUE

présentée le 12 janvier 1993 devant la commission d'examen

JURY

Président : Professeur André VERBERT  
Rapporteurs : Docteur Dallas SWALLOW  
Docteur Edith PUCHELLE  
Membres : Professeur Pierre DEGAND  
Professeur Steven BALL  
Docteur Jean-Pierre AUBERT



50376  
1993  
6

50376  
1993  
6



A ma Femme  
A mes Parents  
A toute ma Famille

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

**Articles :**

1) Molecular cloning and chromosomal localization of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs.

N. PORCHET, N. GUYEN VAN CONG, J. DUFOSSE, JP. AUDIE, V. GUYONNET-DUPERAT, MS. GROSS, C. DENIS, P. DEGAND, A. BERNHEIM and JP. AUBERT.

Biochemical and Biophysical Research Communication (1991) vol 175, N°2 (414-422).

2) Structural features of the core proteins of human airway mucins ascertained by cDNA cloning.

N. PORCHET, J. DUFOSSE, JP. AUDIE, V. GUYONNET DUPERAT, N. GUYEN VAN CONG, P. DEGAND and JP. AUBERT

American Review of Respiratory Disease (1991) Vol. 144, (S15-S18)

3) Degenerate 87 base pair tandem repeats create hydrophilic/hydrophobic alternating domains in human peptide mucins mapped on 11 p 15

J. DUFOSSE, N. PORCHET, JP. AUDIE, V. GUYONNET DUPERAT, A. LAINE, I. VAN SEUNINGEN, S. MARRAKCHI, P. DEGAND, JP. AUBERT

Article sous presse à Biochemical Journal

4) Les mucines humaines : pourquoi une telle hétérogénéité peptidique ?

N. PORCHET, J. DUFOSSÉ, P. DEGAND, JP. AUBERT

Medecine et Science  
(1991) Vol. 7, N°10 (1024-1030)

5) Multiple chromosomal locations of human tracheobronchial mucin genes.

JP. AUBERT, JP. AUDIE, V. GUYONNET DUPERAT,  
J. DUFOSSÉ, P. DEGAND, N. PORCHET

Eur. Resp. J. (1991) Vol. 4, (382)

### Communications :

- Gene de mucines tracheo-bronchiques humaines localisé en 11p15: alternance de domaines hydrophiles/hydrophobes.

DUFOSSÉ J., GUYONNET DUPERAT V., AUDIE JP.,  
PORCHET N., DEGAND P & AUBERT JP.

Poster présenté lors de la réunion de l'association nationale des professeurs et maîtres de conférences biochimistes des U.F.R. médicales, les 5 et 6 octobre 1990 à Strasbourg.

- Les mucines trachéobronchiques humaines : Mise en évidence d'alternances de domaines hydrophiles et hydrophobes.

DUFOSSÉ J., PORCHET N., AUBERT JP.

Communication orale présentée lors du XVIIIème Forum des jeunes chercheurs de la société française de biochimie et de biologie moléculaire, le 5 septembre 1991 à TOURS.

*Je tiens à remercier :*

**Monsieur le Professeur Pierre DEGAND,**

vous m'avez accueilli dans votre Unité en 1988, en me permettant de préparer mon DEA puis ma thèse. Je vous en remercie. L'intérêt que vous avez porté à ce travail et les réflexions constructives que vous m'avez faites furent de formidables encouragements.

**Monsieur le Professeur André VERBERT,**

vous me faites l'honneur de présider ce jury de thèse.

**Madame le Docteur Dallas SWALLOW  
et Madame le Docteur Edith PUCHELLE,**

vous avez accepté de juger ce travail et ceci malgré la distance et vos lourdes charges scientifiques.

**Monsieur le Professeur Steven BALL,**

vous me faites l'honneur de siéger à ce jury.

*Je tiens à exprimer ma reconnaissance à :*

**Madame le Docteur Nicole PORCHET**

pour la grande disponibilité, la culture scientifique et le très grand dynamisme dont elle a fait preuve tout au long de ce travail.

**Monsieur le Docteur Jean-Pierre AUBERT**

pour avoir accepté d'être mon tuteur de DEA puis mon directeur de thèse. Vous m'avez accordé votre confiance en me proposant un sujet de recherche ce qui me permet, trois années plus tard, de présenter ce travail devant un Jury. Je vous en remercie.

Je remercie également :

Mr le Docteur Michel CREPIN pour son dynamisme et la très grande disponibilité dont il a fait preuve lors de l'étude PCR.

Mme Danièle PETITPREZ et Mr Pascal MATHON pour l'aide technique très précieuse et efficace qu'ils m'ont apportée.

Mme Edith DELALEAU et Mme Marie-Thérèse MAILLARD pour avoir dactylographié cette thèse.

Mr Claude VANDEPERRE pour la qualité de la documentation photographique de ce mémoire et des publications.

Mr le Docteur DULAT, Chef du Service de Biochimie de l'Hôpital Militaire Scrive, pour les nombreuses facilités qu'il a pu m'accorder lors de mon service militaire.

Tous les membres de l'équipe pour leur collaboration, mais aussi pour la très bonne ambiance qu'ils ont chacun entretenue.



Ce travail a été effectué au sein de l'Unité des Protéines Normales et Pathologiques (Unité INSERM n°16), dirigée par Monsieur le Professeur P. DEGAND, sous la responsabilité scientifique de Monsieur le Docteur J.P. AUBERT.

Il a bénéficié de la collaboration de Monsieur le Docteur N. Guyen Van Cong (Unité CNRS URA 1158).

Cette thèse a bénéficié de l'aide financière du contrat de l'Association pour la Recherche contre le Cancer n° 6780 et n° 6860

## ABREVIATIONS

ADNc	ADN complémentaire des ARN messagers du tissu considéré
ARNm	ARN messenger
Bet	bromure d'éthidium`
BME	$\beta$ mercapto-éthanol
cpm	coups par minute
dCTP	deoxy-cytidine-triphosphate
DEAE	diéthyl-amino-éthyl
DEPC	diéthylpyrocarbonate
DMSO	diméthyl sulfoxyde
dNMP	deoxy-nucléotide-monophosphate
dNTP	deoxy-nucléotide-triphosphate
DO	densité optique
DTT	dithiothreitol
E. Coli	Escherichia Coli
EDTA	acide éthylène diamine tétra acétique
GTE	glucose-Tris-EDTA
hfl	"high frequency lysogeny"
HF	acide fluorhydrique
HPRI	"human placental ribonuclease inhibitor"
IPTG	isopropyl- $\beta$ -D-thiogalactopyrannoside
KAC	acétate de potassium
Kb	kilobase
KDa	Kilodalton
LB	Luria Bertani
MES	Acide 2 [N-morpholino]-éthane sulfonique
Milieu TMN	tris magnésium NaCl
Milieu TFB	"transformation buffer"
MOPS	acide-3-[N-morpholino]-propane sulfonique
NP40	Nonidet P-40
PCR	Polymerase chain reaction
PEG	poly-éthylène glycol
RER	reticulum endoplasmique rugueux
SDS	sodium dodecyl sulfate
SVF	sérum de veau foetal
Tampon GT	thiocyanate de guanidinium



Tampon SSC	citrate de sodium - chlorure de sodium
Tampon STE	sodium-tris-EDTA
Tampon TBE	tris-borate-EDTA
Tampon TBS	"tris buffer saline"
Tampon TE	tris-EDTA
Tampon TMG	tampon magnésium-gélatine
TFMS	acide trifluorométhane sulfonique
Tris	tris-(hydroxy méthyl) aminométhane
Triton X100	isooctylphénoxy polyéthoxyéthanol
X-Gal	5-bromo-4-chloro-3-indolyl- $\alpha$ -D-galactopyrannoside

# SOMMAIRE

	Pages
<b>AVANT-PROPOS</b>	1
<b>INTRODUCTION</b>	3
A - LE MUCUS	3
B - LES MUCINES	4
1) Origine et rôle	4
2) Biosynthèse	6
3) Isolement et purification des mucines	8
4) Masse moléculaire des mucines	9
5) Composition biochimique des mucines humaines	11
a) la chaîne peptidique	
b) la copule glucidique	
6) Modèles structuraux	13
7) Séquences primaires des mucines humaines et apport de la biologie moléculaire	19
* Les mucines intestinales	20
* Les mucines trachéobronchiques	23
· la famille I	
· la famille II	
· la famille III	
* Les "mucin-like" de tumeur mammaire	27
<b>RESULTATS</b>	31
A - STRATEGIE GENERALE	31
B - RESULTATS	32
1) Criblage de la banque d'ADNc en vecteur $\lambda$ gt 11 de muqueuse trachéobronchique humaine	32
2) Criblage de la banque d'ADNc en vecteur $\lambda$ gt 10 de muqueuse trachéobronchique humaine	34
3) Séquence nucléotidique et acides aminés déduites	36
<i>Séquence nucléotidique et peptidique de la famille I</i>	
<i>Séquence nucléotidique et peptidique de la famille II</i>	
<i>Séquence nucléotidique et peptidique de la famille III</i>	
4) Apport de la PCR	76

	Pages
5) Etude de l'expression des mucines trachéobronchiques humaines par Northern blots	76
6) Localisation chromosomique et étude en Southern blots	79
<b>CONCLUSION GENERALE</b>	<b>87</b>
<b>ARTICLE</b>	
<b>APPENDICE TECHNIQUE</b>	<b>89</b>
Préparation des anticorps polyclonaux	89
Criblage de la banque en vecteur $\lambda$ gt11 à l'aide de sondes immunologiques	89
Purification, à partir de l'immunsérum polyclonal antimucine trachéale humaine déglycosylée, des anticorps spécifiques de chaque clone	91
Caractérisation de clones sélectionnés	91
1. Amplification en milieu liquide	
2. Caractérisation de l'insert	
Criblage de la banque en vecteur $\lambda$ gt10	93
Sous clonage des fragments	93
Extraction et purification des fragments	
Amplification et purification du plasmide pBR322	94
Réalisation de sondes nucléotidiques	96
1. "Nick-translation"	
2. "Multiprime"	
Purification des sondes	96
Obtention des séquences nucléotidiques	97
Séquençage	
1. Sous-clonage dirigé à partir de la carte de restriction	
2. Méthode de délétion-mutation à l'exonucléase III	
3. Transformation des bactéries E. Coli JM109 et HB101	
3.1 Technique de Hanahan	
3.2 Technique de Chung	
Localisations chromosomiques	106
Préparation des ARN et Northern blot	106
1. Homogénéisation du matériel biologique	
2. Isolement des ARN cellulaires totaux	
3. Electrophorèse des ARN en milieu dénaturant	
Analyse par la technique du Southern blotting	109
1. Préparation de l'ADN génomique	
2. Analyse en Southern blot des sondes de l'ADN humain	
3. Transfert et Fixation	
4. Hybridation à une sonde nucléique	
<b>BIBLIOGRAPHIE</b>	<b>114</b>

## AVANT-PROPOS

*Le mucus a pu être défini comme "l'une des perfections de la Nature en matière de protection..."*

Toutes les cellules vivantes interagissent avec leur environnement ; ces interactions passent par la membrane plasmique. Néanmoins lorsque les agressions, venant du milieu extérieur, nécessitent une protection supplémentaire, la sécrétion de mucus crée une barrière physique sélective.

Le mucus nous est parvenu du très lointain de l'histoire de l'Evolution et, sa fonction première chez le ver de terre, est de protéger de la déshydratation une peau très perméable, et d'y réguler les échanges gazeux O<sub>2</sub>/CO<sub>2</sub>. Chez l'anguille, la carpe, les cellules à mucus de la peau contrôlent le passage de l'eau et l'homéostasie de la pression osmotique. Chez les mammifères, de nombreux organes, l'oeil, le tractus gastrointestinal, la trachée, les bronches, le pancréas, la vésicule biliaire, l'appareil reproducteur protègent leurs epithelia par la sécrétion de mucus.

*"Néanmoins, la Nature n'a pu prévoir tous les effets néfastes de la Civilisation Occidentale qui concourent tous à limiter le pouvoir protecteur du mucus..."* Ainsi, de nombreuses pathologies humaines sont associées à un dérèglement de la biosynthèse des mucines. De nombreux domaines de la médecine sont intéressés par l'étude des mucines ; c'est le cas de la Pneumologie (mucoviscidose, bronchite chronique, asthme), de la Gastroentérologie (cancer du colon, ulcère gastrique, ulcère duodéal, maladies inflammatoires du tube digestif), de la Gynécologie (stérilité), et enfin de la Cancérologie (cancers épithéliaux).

Les mucines sont les constituants biochimiques majeurs du mucus et à ce titre, depuis de nombreuses années, ont été étudiées comme marqueurs potentiels de la différenciation et de la cancérisation. Les études physicochimiques ont abouti à une définition tout à fait originale de ces glycoprotéines : masse moléculaire élevée et hétérogène, difficultés de purification, haute teneur en chaînes glycaniques elles-mêmes très diversifiées.

L'axe peptidique quant à lui n'a commencé à être connu que depuis 1989, des difficultés technologiques insurmontables n'ayant permis que l'obtention de très courtes séquences peptidiques (22 acides aminés au maximum).

C'est dans ce contexte, que s'est inscrit notre travail de thèse dont l'objectif était l'étude des mucines trachéobronchiques humaines grâce aux technologies de la biologie moléculaire.

# INTRODUCTION

## A - LE MUCUS

Le mucus est une sécrétion très hétérogène organisée en deux couches, l'une fluide au contact de l'épithélium et l'autre superficielle douée de propriétés rhéologiques (viscoélasticité, filance). Il est élaboré, au niveau des muqueuses respiratoire, digestive et génitale, par des cellules épithéliales spécialisées. Il assume de nombreuses fonctions dont la finalité est la protection de l'épithélium de surface à l'égard de nombreuses agressions potentielles :

- de nature chimique comme c'est le cas pour les muqueuses du tractus digestif où se concentrent des pH extrêmes et des activités enzymatiques très puissantes et très variées.

Le mucus et en particulier les mucines exercent ici un pouvoir tampon.

- de nature physique comme la déshydratation des muqueuses.

- de nature mécanique, par exemple lors de la progression du contenu luminal digestif.

- de nature bactérienne, enfin, les muqueuses étant protégées de la colonisation bactérienne grâce au "piège" muqueux.

Le prélèvement de mucus "natif" est très difficile à réaliser chez un sujet sain. C'est pourquoi un grand nombre d'études ont été réalisées sur mucus pathologique.

Le mucus bronchique, par exemple, peut provenir d'aspiration sous bronchofibroscopie (1), d'aspiration par l'orifice d'une trachéotomie (2), de lavages bronchoalvéolaires (3). Une quantité beaucoup plus importante de mucus est souvent obtenue par un recueil des expectorations de sujets hypersécrétants présentant des états pathologiques comme la bronchite chronique, l'asthme ou la mucoviscidose.

La composition biochimique d'un mucus se caractérise par sa richesse en mucines, en eau (90 à 95 %) et en sels minéraux. A côté de ces constituants majoritaires, on trouve d'autres éléments comme :

- des protéines et glycoprotéines : sérumalbumine (4), immunoglobulines (IgA sériques et sécrétoires, IgG)(5), lactotransferrine (4), lysozyme (4), antiprotéases (6) et polypeptides riches en Proline (P.R.P.) décrits dans le mucus trachéobronchique (7).

- des acides nucléiques (8) en quantité peu importante chez un sujet sain, en considérable augmentation quand les sécrétions sont infectées.

- des lipides (phospholipides et glycolipides) (9, 10) dont la quantité varie également avec l'état d'infection des sécrétions.

## **B - LES MUCINES**

### **1) Origine et rôle**

Les mucines sont les constituants biochimiques majeurs du mucus et se définissent comme des populations de macromolécules complexes. Elles sont synthétisées par des cellules spécialisées de l'épithélium, les cellules caliciformes. Ces cellules sont très polarisées et présentent des caractéristiques morphologiques en rapport avec la production massive de mucines (RER concentré au pôle basal, Golgi très développé) (figure n°1). Les mucines, selon les muqueuses, sont également synthétisées par des glandes qui participent très activement à un renouvellement rapide et constant du mucus.

Les mucines ont la propriété, au contact de l'eau, de former un gel et sont responsables des propriétés physicochimiques et biologiques du mucus :

- Au niveau de la muqueuse trachéobronchique, le mucus piège les particules inhalées et les microorganismes et, grâce aux mouvements ciliaires, le film de mucus remonte au niveau du larynx où il est dégluti. Les mucines présenteraient au niveau de leurs chaînes oligosaccharidiques des sites récepteurs pour des adhésines bactériennes.



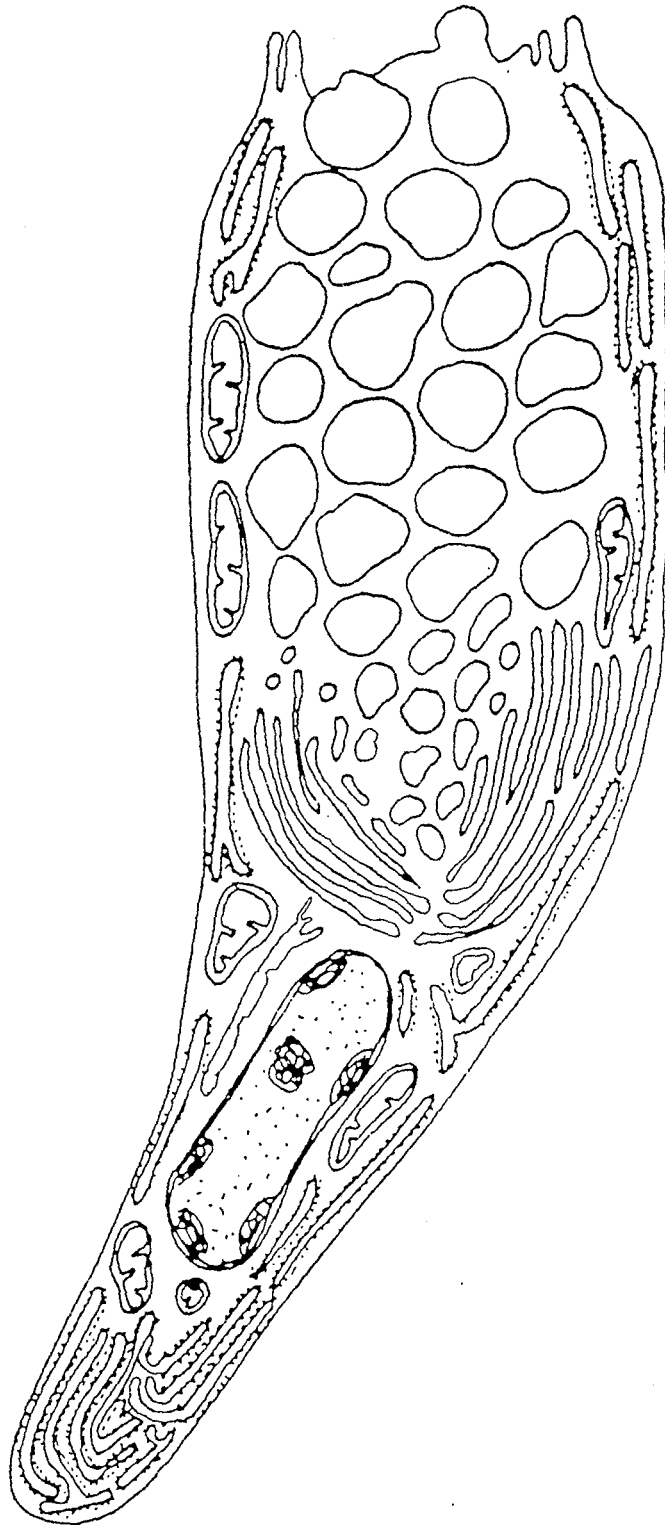


Figure n°1 : morphologie d'une cellule caliciforme.

- Au niveau gastrique, le mucus joue un rôle primordial de lubrification nécessaire au passage d'aliments solides en assurant aussi une protection contre l'abrasion mécanique. Les mucines contribuent à créer une zone tampon entre la lumière gastrique acide (pH 2) et la muqueuse (pH 7) (11).

- Au niveau intestinal, le mucus, grâce au mouvement péristaltique de l'intestin, joue un rôle de filtre sélectif pour l'absorption des nutriments, protège les cellules intestinales des hydrolases, et permet l'évacuation des microorganismes.

- Au niveau de la muqueuse cervicale, les mucines sont synthétisées sous contrôle hormonal. Les changements de structure et de composition du mucus autorisent, en période périovulatoire seulement, la pénétration des spermatozoïdes (12).

Cette présentation du rôle des mucines humaines n'est pas exhaustive. En réalité un certain nombre de fonctions de ces macromolécules ne sont pas encore élucidées. On peut penser que l'expression et la biosynthèse des mucines doivent être modulables en fonction de la nature et de l'intensité du stress.

A côté des mucines, glycoprotéines de sécrétion, il faut distinguer également les glycoprotéines "mucin-like" qui présentent les caractéristiques physico-chimiques des mucines (masse moléculaire élevée, pourcentage élevé de O-glycannes). Ces mucines "membranaires" sont des glycoprotéines ancrées dans la membrane plasmique de nombreuses cellules épithéliales dont l'expression est considérablement accrue dans les cellules transformées. Ceci justifie, pour l'étude diagnostique et pronostique des tumeurs épithéliales, de préparer des anticorps monoclonaux capables de détecter dans le sang des patients les formes solubles de ces antigènes qui sont donc des marqueurs tumoraux.

## **2) Biosynthèse**

La biosynthèse des mucines est un phénomène complexe encore mal connu.

Les étapes de transcription et de traduction sont totalement inexplorées. Les phénomènes post-traductionnels, eux non plus, ne sont encore que très partiellement élucidés et impliqueraient : N-et O-glycosylation, oligomérisation

peptidique, stockage intracellulaire sous forme de grains de mucigène, sécrétion à la fois spontanée non régulée et stimulée.

Les axes peptidiques de mucines sont synthétisés dans les polyribosomes puis acheminés dans la lumière du RER. Une N-glycosylation a pu récemment être démontrée dans le RER nécessitant un lipide-oligosaccharide (Dol P-P-GlcNAc<sub>2</sub>-Man<sub>9</sub>-Glc<sub>3</sub>) (13). La fonction de ces quelques chaînes N-glycanniques est encore inconnue. Le polypeptide est alors dirigé dans le Golgi où les N-glycannes sont convertis en oligosaccharides complexes (addition de GlcNAc, élimination de mannose, action des glycosyltransférases). La O-glycosylation est le phénomène post-traductionnel majeur affectant les mucines puisque 70 à 90 % de la masse moléculaire des molécules achevées sont constituées de O-glycannes. Aucune séquence-code peptidique n'a été mise en évidence. Cependant la présence de résidus proline à proximité des sérines et thréonines, favoriserait l'accès de la Gal NAc transférase (14).

L'addition de résidus de Gal NAc pourrait intervenir à deux niveaux : addition précoce co-traductionnelle, addition majeure post-traductionnelle (15).

La O-glycosylation précoce aurait un rôle sur la conformation du polypeptide et l'addition séquentielle de monosaccharides par des glycosyltransférases spécifiques dans le trans Golgi (16). Les étapes finales du processus voient l'addition à l'extrémité des chaînes oligosaccharidiques d'un nombre variable de résidus de fucose, d'acide sialique, de sulfate dont la densité crée le caractère neutre, acide, voire très acide des molécules de mucines.

Certaines mucines, mucines gastriques, mucines cervicales formeraient un assemblage oligomérique grâce à des liaisons covalentes de type ponts disulfure. Selon DEKKER et Coll, l'oligomérisation pourrait intervenir dans le RER et dépendre de la N-glycosylation précoce (17).

Pour l'instant, il n'existe pas d'éléments expérimentaux permettant d'affirmer que toutes les molécules de mucine d'une muqueuse donnée, ou de différentes muqueuses subissent toutes les mêmes modifications post-traductionnelles.

D'autres modifications, comme l'acylation, ont pu être évoquées.

### 3) Isolement et purification des mucines

En raison de la grande interactivité des mucines avec le milieu chimique ou cellulaire environnant, l'étude de ces macromolécules a été rendue difficile. La préparation, l'isolement et la purification de ces glycoprotéines ont été source de résultats différents publiés par différentes équipes. Nous avons tenté de résumer succinctement dans ce chapitre les différentes méthodologies utilisées.

Le premier problème que l'on rencontre généralement est celui du recueil des sécrétions.

Dans le cas particulier des mucines trachéobronchiques par exemple, on peut obtenir ces sécrétions par aspiration sous bronchofibroscopie ou par l'orifice d'une trachéotomie, par lavages bronchoalvéolaires ou du contenu de kystes bronchogéniques.

Les volumes recueillis ramènent un matériel proche de la sécrétion physiologique mais n'ont pas fourni matière à une étude très poussée des mucines trachéobronchiques. C'est pourquoi, pour réaliser un travail plus complet, il a fallu étudier les expectorations de sujets hypersécrétants. Ces sujets sont atteints de bronchite chronique, d'asthme ou de mucoviscidose.

Le deuxième problème auquel on se heurte est la solubilisation du mucus. Différentes techniques plus ou moins dégradantes sont proposées :

- action d'enzymes protéolytiques
- action d'agents réducteurs
- action mécanique en présence d'agents chaotropiques
- sonication
- dilution par l'eau distillée.

Une fois solubilisées, les mucines doivent être purifiées soit par chromatographie (gel perméation, échange ionique) soit par centrifugation isopycnique en gradient de densité de sels de césium.

Pour vérifier l'état de pureté des préparations de mucines, peu de techniques sont utilisables. La principale consiste à réaliser une électrophorèse SDS-Page en condition non réductrice qui permet de visualiser les éventuelles

protéines contaminantes. En effet, étant donné leur masse élevée, les mucines ne rentrent pas dans un gel de polyacrylamide même très faiblement réticulé. On peut également vérifier l'absence de contaminants nucléiques par électrophorèse en agarose ou spectre UV.

#### 4) Masse moléculaire des mucines

L'estimation des masses moléculaires peut être réalisée par différentes méthodes classiques :

- sédimentation à l'équilibre
- détermination de la  $S^{\circ}_{20W}$  par ultracentrifugation
- diffraction de la lumière
- chromatographie de gel filtration
- microscopie électronique.

Les techniques les moins dégradantes de solubilisation et de purification ont permis d'obtenir les masses moléculaires les plus élevées. L'éventail des techniques de purification et de mesure de masses moléculaires a entraîné la publication de beaucoup de résultats plus ou moins contradictoires (Tableau n°1).

Deux points importants sont à souligner quelle que soit l'origine des mucines :

- la masse moléculaire est élevée de  $2 \cdot 10^5$  à  $3 \cdot 10^7$  daltons
- Il existe une très grande polydispersion des masses dans un même échantillon.

Un dernier point de controverse est à souligner. Il concerne l'utilisation d'agents réducteurs (2-mercaptoéthanol, dithiothreitol, dithioerythritol). L'agent réducteur romprait des ponts disulfure interchaînes libérant des sous-unités, ce qui objectiverait, à l'électrophorèse, une diminution de la masse moléculaire (18, 19). Mais la réduction pourrait également activer des activités protéolytiques contaminantes qui couperaient l'axe peptidique en des sites accessibles, occasionnant aussi une diminution de masse également (20).

AUTEUR	Virmani & Coll 1991 (21)	Thornton & Coll 1990 (22)	Sheehan & Coll 1987 (23)	Marshall & Coll 1978 (24)	Sheehan & Coll 1991 (25)	Creeth & Coll 1977 (26)
Origine de la mucine	mucine trachéale canine	mucine trachéale humaine	mucine cervicale humaine	mucine colique de porc	mucine respiratoire humaine	mucine respiratoire humaine
mode de purification	- chromatographie CL 4B Sepharose - ultracentrifugation en gradient de densité CsCl	- ultracentrifugation en gradient de densité CsCl/GuCl - chromatographie Sepharose CL 2B	- ultracentrifugation en gradient de densité en GuCl	- gel filtration - ultracentrifugation à l'équilibre en CsCl - exclusion en gel filtration Sepharose 2B	-extraction par centrifugation en GuCl 6M en présence d'inhibiteurs de protéases	-solubilisation en CsCl - ultracentrifugation en CsCl
masse moléculaire en daltons	11.10 <sup>6</sup>	14 à 16.10 <sup>6</sup>	6 à 24.10 <sup>6</sup>	10 <sup>6</sup>	3,2.10 <sup>6</sup>	3,3 à 7.10 <sup>6</sup>
remarque		mise en évidence de domaine résistant à la trypsine (T domaines)				diminution de taille après action du βmercaptoéthanol

Tableau n° 1 : Masse moléculaire de mucines

## 5) Composition biochimique des mucines humaines

### a) la chaîne peptidique

La fraction peptidique des mucines humaines est minoritaire puisqu'elle ne rend compte que de 10 à 20 % de leur masse moléculaire. Elle est caractérisée par un très fort pourcentage en résidus hydroxylés. Ce pourcentage peut varier de 19 à 50 % selon l'origine des mucines et surtout selon le protocole de préparation utilisé. Ces variations sont résumées dans le tableau n°2.

Certains acides aminés sont faiblement représentés. C'est le cas de la tyrosine, phénylalanine, méthionine et du tryptophane.

Nous reviendrons plus loin sur les difficultés qui ont été rencontrées pour la détermination de la structure primaire.

### b) la copule glucidique

Les glycanes représentent 70 à 90 % en poids des mucines. Ces dix dernières années, les connaissances structurales acquises dans le domaine des mucines ont concerné essentiellement ces chaînes glycaniques. Plus d'une centaine de structures glycaniques distinctes ont été identifiées et l'on a longtemps attribué l'hétérogénéité des mucines à l'hétérogénéité glycanique.

Pour schématiser cette complexité, on peut signaler que les chaînes oligosaccharidiques peuvent être divisées en 3 parties :

- le noyau ou core

Actuellement 6 cores différents ont été recensés et peuvent être trouvés chez un même individu (34).

- le squelette

Il est constitué d'une association d'unités disaccharidiques de deux types Gal  $\beta$ 1,3 GlcNAc (Type 1) ou Gal  $\beta$ 1,4 GlcNAc (Type 2). Les différents squelettes peuvent être linéaires ou branchés.

- la région périphérique peut correspondre à un antigène de groupe sanguin des systèmes A, B, H et Lewis.



AUTEURS	ROUSSEL et coll (1975) (27)	CREETH et coll (1977) (26)	LAFITTE et coll (1977) (3)	FELDHOFF et coll (1979) (28)	LAMBLIN et coll (1979) (29)	ROSE et coll (1979) (30)	SACHDEV et coll (1980) (1)	BHASKAR et coll (1981) (31)		TARACHNIK et coll (1981) (32)	SNYDER et coll (1982b) (33)	HOUDRET et coll (1981) (20)	
Origine mucus	mucovis- cidence	asthme	normal	asthme	bronchite chronique	asthme	normal	asthme		mucovis- cidence	mucovis- cidence	bronchite chronique	
Fraction étudiée	"S" réduite	18,1* native	"3b" réduite	"1" native	"C <sub>3</sub> b" réduite	MGP 0,3M "HA" réduite	"HO-4" native	"AsEXT" native	"AsEXT" réduite	"IA" natif	"CF-C" native	"Bc" ré- duite en milieu non dissociant	"Bb" ré- duite en milieu dissociant
Acides aminés mol/100mol													
AS(X)	2,2	8,3	2,7	7,0	3,6	4,0	6,9	7,1	5,5	10,6	8,9	2,3	5,3
Thr	28,7	18,1	27,5	17,0	24,4	23,2	16,4	17,9	22,5	9,3	19,8	29,0	24,0
Ser	17,0	12,7	18,6	11,0	16,1	16,9	12,5	14,1	13,7	10,1	15,2	19,5	14,5
GL(X)	4,4	8,1	3,9	7,8	4,2	4,3	9,5	9,7	7,8	8,7	8,1	3,2	6,3
Pro	11,8	10,5	14,7	10,3	11,1	12,8	10,1	12,7	13,7	5,8	12,4	12,5	9,3
Gly	8,8	7,6	6,5	7,2	8,2	7,1	7,1	8,1	7,6	8,8	9,7	5,7	7,0
Ala	10,4	7,2	8,2	8,6	9,2	7,6	7,6	7,1	7,5	8,5	8,1	9,0	7,0
1/2 Cys	-	2,6	0,6	1,5	0,7	1,3	2,3	1,1	1,5	9,2	1,2	0,9	1,5
Val	4,0	4,8	4,0	6,9	3,8	4,2	6,1			6,2	4,0	3,7	4,2
Met	traces	0,4	0,8	1,0	1,2	0,6	0,9			1	0,4	0,6	0
Ile	2,2	2,1	2,1	2,9	2,1	1,7	2,5			2,7	2,4	1,8	2,0
Leu	4,5	4,9	2,3	6,3	4,3	3,8	6,5			6,9	6,5	3,1	4,5
Tyr	traces	1,8	1,1	1,1	1,1	1,6		24,2	20,2	2,2	traces	1,1	1,8
Phe	traces	2,1	1,2	2,1	1,1	2,2	2,6			1,8	traces	1,2	1,9
Lys	1,6	2,1	2,2	2,9	2,7	3,2	3,6			2,7	2,4	1,7	2,2
His	2,1	2,5	1,5	2,1	3,8	2,0	1,9			1,7	0,3	2,8	2,4
Arg	2,1	4,2	2	4,0	2,5	3,5	3,5			3,5	0,5	1,7	3,0
Acides aminés totaux (g/100g)	10,15	29	19,53	28	13	15	23,2	24	18	?	30,3	12,8	18,
OH AA % Ser/Thr	45,7 0,59	30,8 0,70	46,1 0,68	28,0 0,65	40,5 0,66	40,1 0,73	28,9 0,76	30 0,68	36,2 0,61	19,4 1,09	35,0 0,77	48,5 0,67	38,5 0,63

Tableau n°2 : Composition en acides aminés de mucines.

Cette région va également supporter les charges négatives des mucines.

Ce caractère d'acidité relative a permis et permet encore de faire une classification grossière des mucines entières ou des glycopeptides obtenus par hydrolyse par la pronase :

- Fraction neutre : pauvre en résidus acides
- Fraction acide : riche en acides sialiques`
- Fraction très acide : riche en résidus sulfate.

La chromatographie d'échange ionique permet de séparer ces différentes fractions. Il est à noter cependant que ce nombre de fractions variera en fonction du protocole retenu (nombre de tampons du gradient discontinu). Ce résultat montre une nouvelle fois le principe d'une hétérogénéité. Il n'existe pas une molécule de mucines ne contenant que des résidus "fucose" ou que des résidus "acides sialiques".

Les différentes fractions existent dans un équilibre physiologique qui peut être rompu lors d'atteintes pathologiques. Par exemple, PODOLSKI a pu mettre en évidence la disparition d'une fraction acide dans les mucines de sujets atteints de Rectocolite Hémorragique (35).

La présence de mannose dans les compositions glucidiques a longtemps été un sujet de controverse. Le mannose pourrait provenir de glycoprotéines contaminantes, néanmoins récemment il a pu être montré que les mucines pouvaient être N-glycosylées.

## 6) Modèles structuraux

Différents modèles d'organisation structurale des mucines ont été proposés. Cependant quelque soit le modèle, l'hypothèse généralement admise dans la littérature est l'agencement en alternance de domaines peptidiques hydrophobes peu glycosylés et de domaines hydrophiles hyperglycosylés.

Deux modèles principaux sont à considérer :

- le modèle linéaire avec ou sans ponts disulfure interchaînes (figure n°2 et 3)

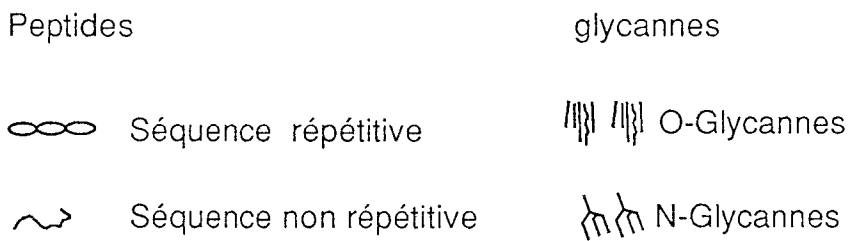
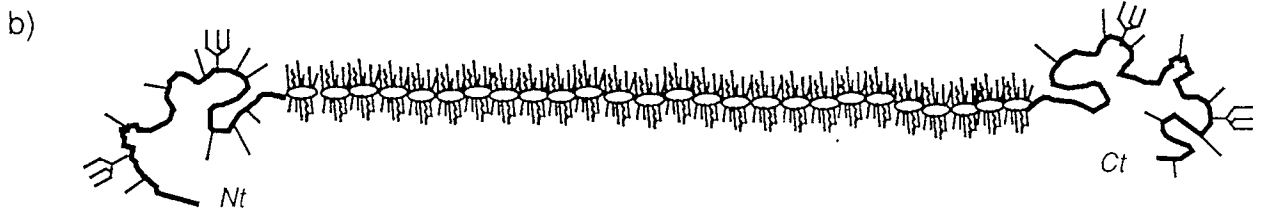
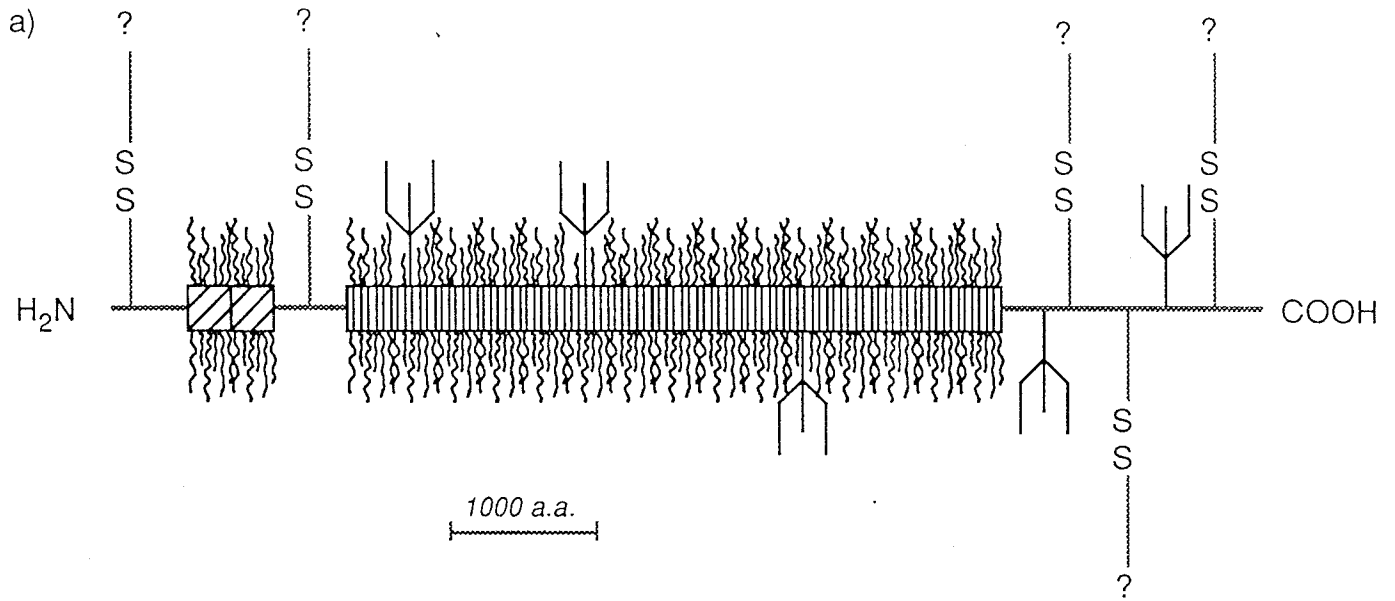
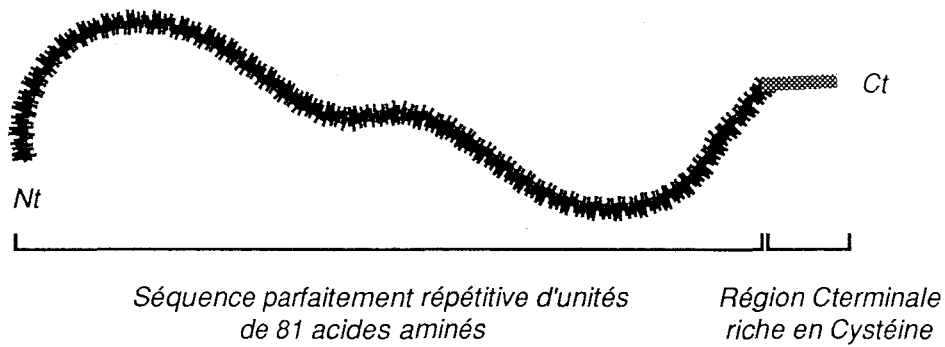
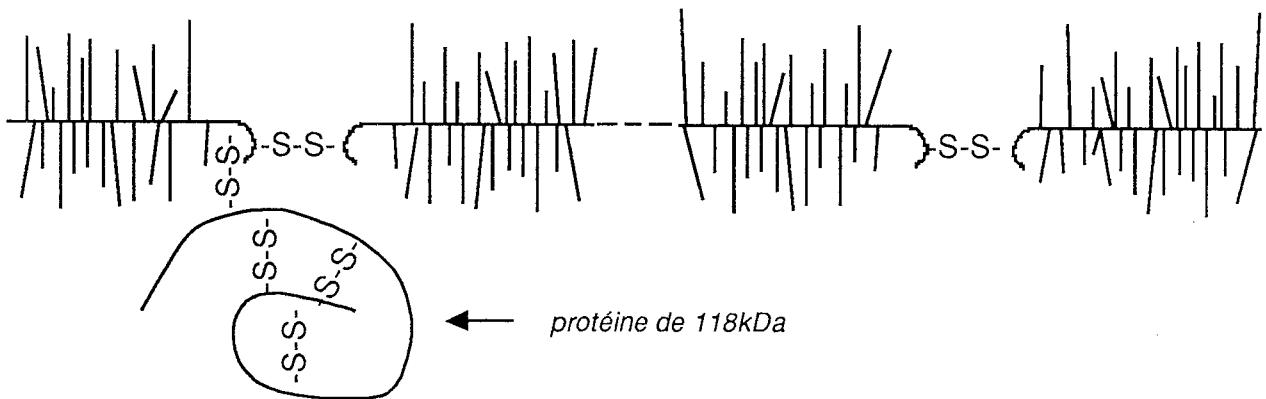


Figure n°2 : Modèles structuraux de mucines intestinales humaines proposés par (a) N.W.Toribara (36) et par (b) Y.S.Kim (37)



a) Modèle structural de mucine sousmaxillaire porcine proposé par A.E.Eckhardt (38)



b) Modèle structural de mucine intestinale de rat proposé par M.Mantle (19)

Figure n°3

-le modèle dit en "moulin à vent" (figure n°4).

Des résultats contradictoires émaillent la littérature. Ils peuvent être en relation avec une origine tissulaire variée (modèles structuraux différents pour les muqueuses gastrique et respiratoire ?) mais aussi et toujours avec la grande diversité des protocoles de préparation des mucines.

Le modèle de Carlstedt propose une structure linéaire où alternent des domaines nus et glycosylés enchaînés les uns aux autres par des ponts disulfure (39).

L'action d'agents réducteurs montre une diminution de taille des mucines. Le modèle linéaire contenant des ponts "soufre" interchaînés a été très bien étudié dans le cas des mucines gastriques (17).

La structure des sous-unités est différente chez l'Homme et le rat :

- les monomères sont trois fois plus grands chez l'Homme ;
- les mucines humaines sont moins sulfatées que celles du rat ;

L'alternance de domaines est moins marquée chez le rat puisque le monomère est presque entièrement glycosylé (69).

Le second modèle, dit en "moulin à vent", a été proposé par Allen et al. (figure n°4) (40,41). Les mucines gastriques seraient constituées d'une sous-unité centrale de 70000 D à laquelle seraient associées, par des ponts disulfure, 4 sous-unités de  $0,5 \cdot 10^6$  D. La sous-unité centrale ne serait pas glycosylée. Différentes expériences de protéolyse réalisées sur les sous-unités périphériques tendent à montrer que celles-ci sont constituées d'une région glycosylée flanquée de deux régions non glycosylées. Un modèle similaire a été décrit pour les mucines d'intestin grêle de porc (18). Dans ce modèle, la sous-unité centrale, d'une masse de 90000D, est assemblée à 8 sous-unités, ayant chacune une masse de  $2,4 \cdot 10^5$  D, par l'intermédiaire de ponts disulfure.

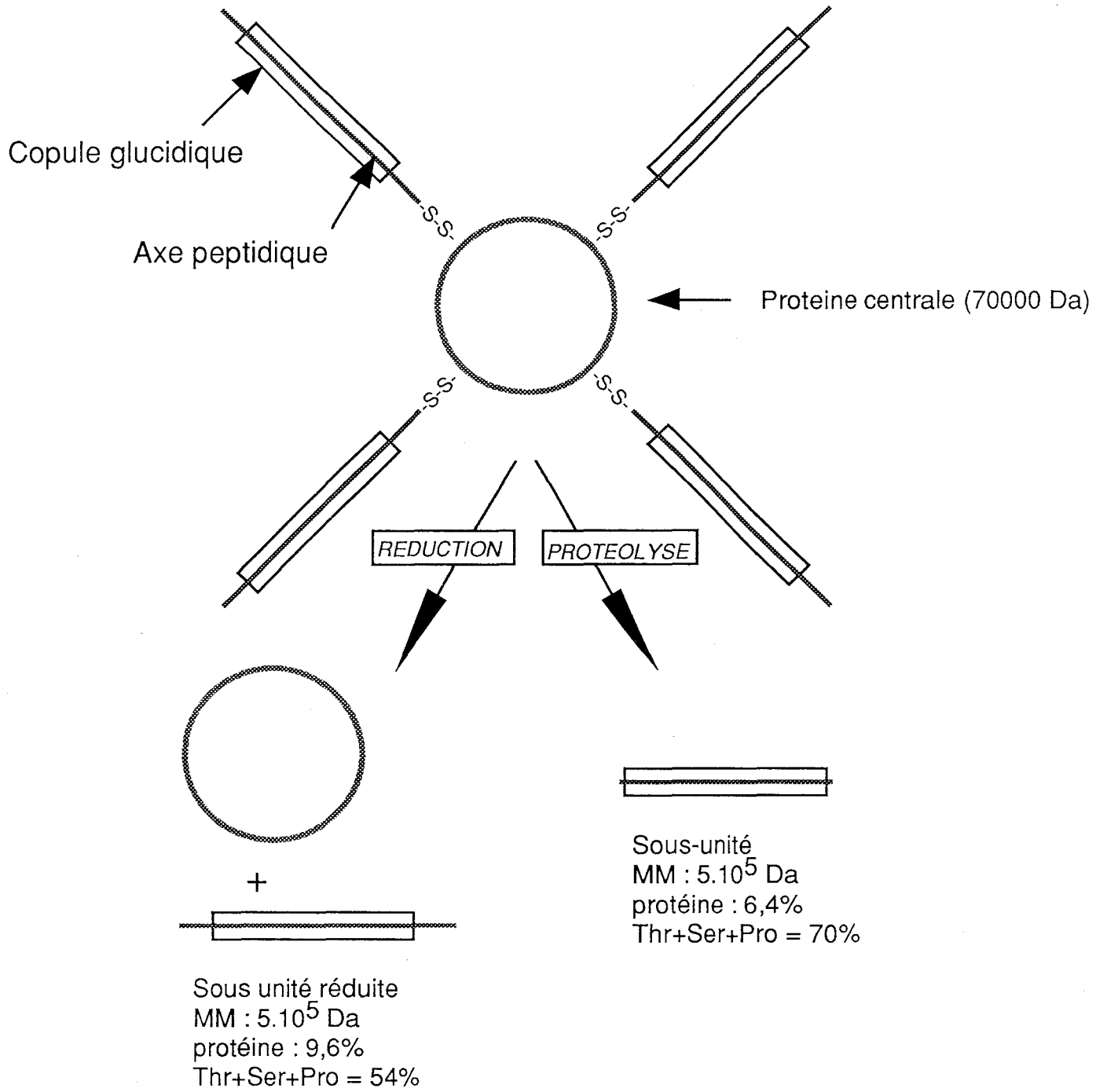


Figure N° 4 : Structure de mucine gastrique de porc proposé par A. Allen (1983)

En 1984, Mantle et al. (42) mettent en évidence la libération d'un peptide de 118000D, par action du  $\beta$  mercaptoéthanol sur les mucines de l'intestin grêle humain (figure n°3b). En 1991, néanmoins une étude de ce peptide révèle qu'il ne constitue pas un peptide de jonction entre deux sous-unités structurales. Le rôle de ce peptide n'est pas connu (19). En 1992 une structure de protéine de liaison a pu être démontrée pour MUC2. Cette structure est codée par un ARNm de mucine intestinale dans une région non répétitive et distante des tandem repeats (43,44).

L'utilisation de la microscopie électronique a permis de faire pencher la balance en faveur du modèle linéaire. Les images obtenues montrent l'existence de longs filaments de mucines cervicales, gastriques ou bronchiques de longueur variant de 200 à 1000 nm (45-48).

En conclusion de cette question des modèles moléculaires, il apparaît que le modèle linéaire où alternent des régions hydrophiles très glycosylées et des régions hydrophobes peu glycosylées est le plus probable. Un faisceau de présomptions étaye cette hypothèse :

- L'action des protéases révèle l'existence de domaines résistants à la protéolyse, dont la masse moléculaire moyenne est comprise entre 200 et 500 kDa. Après protéolyse, les propriétés hydrophobes sont perdues (49).

- Les lipides associés aux mucines de manière non covalente se répartissent de façon irrégulière sur une molécule de mucines (50).

- L'or colloïdal possède une grande affinité pour les parties protéiques dépourvues de chaînes glycaniques. Il constitue un marqueur de localisation de domaines protéiques nus. Son utilisation en microscopie électronique révèle que les sous-unités de mucines cervicales humaines sont constituées d'une alternance de régions glycosylées et de régions nues réparties de manière irrégulière (51).

- Des anticorps dirigés contre des régions nues de mucines cervicales humaines s'hybrident de façon irrégulière (25). Une stratégie identique a permis de visualiser cette alternance sur des mucines d'autres origines (52).

Pour pouvoir prouver définitivement l'existence d'un tel modèle, il faut pouvoir déterminer la séquence primaire d'une ou plusieurs mucines. C'est ce que nous nous sommes efforcé de faire au cours de ce travail de thèse.



## 7) Séquences primaires des mucines humaines et apport de la biologie moléculaire

Les méthodes d'étude des protéines, qui consistent classiquement en l'obtention de peptides grâce à des coupures enzymatiques choisies, puis en la détermination de leur séquence par la dégradation d'Edman, se sont évidemment révélées inopérantes dans l'étude des mucines.

Jusqu'en 1989, n'ont pu être obtenues que de très courtes séquences toujours inférieures à une trentaine d'acides aminés (38, 53-55).

Cela est dû à la présence de très nombreuses chaînes glycaniques mais aussi à la caractéristique de cet axe peptidique que de renfermer 30 à 40 % d'acides aminés hydroxylés engendrant des dérivés labiles lors du séquençage.

Ces difficultés technologiques majeures ont donc entraîné quelques laboratoires à se tourner vers les techniques de l'ADN recombinant.

Schématiquement, les stratégies employées ont comporté les étapes suivantes :

- Préparation d'un anticorps polyclonal ou monoclonal dirigé contre l'axe peptidique, déglycosylé chimiquement par le HF ou le TFMS, d'une mucine ou d'une "mucin-like" préparée à partir d'un tissu sain ou tumoral.

- Confirmation de la qualité de l'anticorps par immunohistochimie.

- Préparation d'une banque d'ADNc en vecteur d'expression à partir de la muqueuse ou lignée cellulaire étudiée.

- Criblage immunologique de ces banques et extraction puis préparation des clones positifs.

- Séquençage. Localisation chromosomique.

- Organisation génomique.

Cette stratégie a déjà permis l'obtention de séquences partielles d'ADNc de mucines animales (glandes sous-maxillaires de porc et de boeuf et muqueuse trachéale canine) (56,21,81), de mucines intestinales humaines et de mucines trachéobronchiques

Dans le cas des "mucin-like", une seule séquence complète celle de l'épisialine appelée aussi PUM ou PEM a été obtenue et confirmée par de nombreuses équipes (57-59).

L'ensemble des résultats préliminaires, apportés par la biologie moléculaire, ont permis de mettre en évidence une extrême hétérogénéité peptidique des mucines humaines.

#### **\* Les mucines intestinales**

En 1989 et 1990, les travaux de Gum et al (60,61) ont montré l'implication de deux gènes dans la biosynthèse des mucines intestinales, l'un situé sur le chromosome 11 en p15 (*MUC2*), l'autre sur le chromosome 7 (*MUC3*).

Le premier gène, en 11 p15, a pu être identifié grâce à trois clones d'ADNc (SMUC 40-41-42) isolés d'une banque de jéjunum humain grâce à un immunsérum dirigé contre l'axe peptidique des mucines d'une lignée tumorale colique LS 174 T. Ces trois clones possèdent le même agencement répétitif d'un domaine élémentaire de 69 nucléotides, soit 23 acides aminés (figure N°5 et 6).

Ces "tandems" répétés 14 fois dans le plus long clone ne sont pas strictement identiques mais possèdent une analogie de 90 %. Néanmoins, certains triplets sont strictement conservés dans tous les domaines répétitifs. La thréonine est fortement représentée (14 acides aminés sur 23) de même que la proline. Pour rendre compte de la teneur en sérine de la mucine intestinale, il faut supposer que la molécule comporte d'autres régions, répétitives ou non, riches en sérine. Deux sites potentiels de N-glycosylation ont été relevés.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
	1	CCA	ACA	ACG	ACA	CCC	ACT	AGC	ACC	ACC	ACC	ATG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACT	GGA	ACA	CAG	ACC
	2	CTA	ACC	CCA	ACA	CCC	ATC	ACC	ACC	ACC	ACT	ACG	GTG	ACC	CCA	ACC	CCT	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	3	CCA	ACA	TCG	ACA	CCC	ATC	AGC	ACC	ACC	ACT	ACG	GTG	ACC	CCA	ACA	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
SMUC	4	CCA	ACC	ACG	ACA	CCC	ATC	ACC	ACC	ACC	ACT	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
40	5	CCA	ACC	COG	ACA	CCC	ATC	ACC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	6	CCA	ACC	COG	ACA	CCC	ATC	ACC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	7	CCA	ACA	TCG	ACA	CCC	ATC	ACC	ACC	AAC	ACT	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCA	ACC	GGC	ACA	CCG	AGT
	1	CCA	ACC	COG	ACA	CCC	ATC	TCC	ACC	ACC	ACT	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	2	CCA	ACC	ACG	ACA	CCC	ATC	ACC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	COG	ACA	CCC	ACC	GGC	ACA	CAG	ACC
SMUC	3	CCA	ACC	ACG	GTA	CTC	ATC	ACC	ACC	ACC	ACT	ACG	ATG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	AGC	ACA	AAG	AGT
41	4	ACA	ACC	GTG	ACA	CCC	ATC	ACC	ACC	ACA	ACT	ACG	GTG	ACC	GCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	5	CCA	ACC	ATG	ATA	CCC	ATC	AGC	ACC	ACC	ACT	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	ACT	GGA	AGC	ACG
SMUC	1	CCA	ACC	ATG	ACA	CCC	ATC	ACC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
42	2	CCA	ACA	ACG	ACA	CCC	ATC	AGC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC

Figure n°5 : Unités nucléotidiques répétitives des clones de mucines intestinales (MUC 2)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23			
SMUC 40	1	Pro	Thr	Thr	Thr	Pro	Ile	<b>Ser</b>	Thr	Thr	Thr	<b>Met</b>	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr		
	2	<b>Leu</b>	Thr	<b>Pro</b>	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr	
	3	Pro	Thr	<b>Ser</b>	Thr	Pro	Ile	<b>Ser</b>	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	4	Pro	Thr	Thr	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	5	Pro	Thr	<b>Pro</b>	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	6	Pro	Thr	<b>Pro</b>	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	7	Pro	Thr	<b>Ser</b>	Thr	Pro	Ile	Thr	Thr	<b>Asn</b>	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	<b>Pro Ser</b>	
SMUC 41	1	Pro	Thr	<b>Pro</b>	Thr	Pro	Ile	<b>Ser</b>	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr		
	2	Pro	Thr	Thr	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	3	Pro	Thr	Thr	<b>Val</b>	<b>Leu</b>	Ile	Thr	Thr	Thr	Thr	Thr	Met	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	<b>Ser</b>	Thr	<b>Lys Ser</b>	
	4	<b>Thr</b>	Thr	<b>Val</b>	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Thr	Val	Thr	<b>Ala</b>	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr
	5	Pro	Thr	<b>Met</b>	<b>Ile</b>	Pro	Ile	<b>Ser</b>	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	<b>Thr Gly Ser</b>	Thr	Thr	
SMUC 42	1	Pro	Thr	<b>Met</b>	Thr	Pro	Ile	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr	
	2	Pro	Thr	Thr	Thr	Pro	Ile	<b>Ser</b>	Thr	Thr	Thr	Thr	Val	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Gly	Thr	Gln	Thr

Figure n°6 : Unités peptidiques répétitives des clones de mucines intestinales (MUC 2)

Le deuxième gène localisé sur le chromosome 7 a été identifié grâce à deux clones (SIB 124-139) isolés de la même banque d'ADNc à l'aide d'un autre immunsérum dirigé contre l'axe peptidique de mucines intestinales humaines. Un agencement en domaines répétitifs est à nouveau observé, le domaine élémentaire étant toutefois différent du précédent et comportant 51 nucléotides soit 17 acides aminés (figure n°7).

L'analyse par *Northern blot* de l'ensemble de ces sondes intestinales, qu'elles proviennent de l'un ou l'autre gène, montre une très large hétérogénéité des ARN messagers codant pour ces mucines.

### \* Les mucines trachéobronchiques

Les études menées par notre groupe de 1989 à 1991 ont permis de localiser trois gènes de mucines trachéobronchiques respectivement sur les chromosomes 11 (en p15), 13 et 3 (en q29) (62-65).

Vingt clones ont été extraits d'une banque d'ADNc de muqueuse trachéobronchique humaine grâce à un immunsérum polyclonal dirigé contre un mélange "équimoléculaire" de glycopeptides de mucines neutres, sialylées, sulfatées, tous déglycosylés par l'acide trifluorométhane sulfonique.

La séquence de quelques uns de ces clones a permis de déduire trois types d'organisation peptidique (figure n°8) :

- la **famille I**, constituée de séquences répétées en tandem :

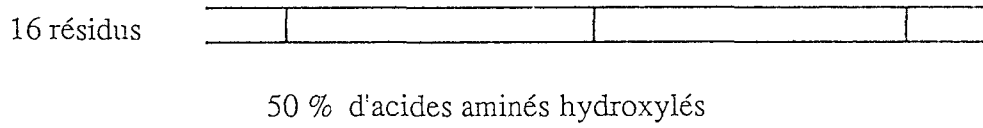
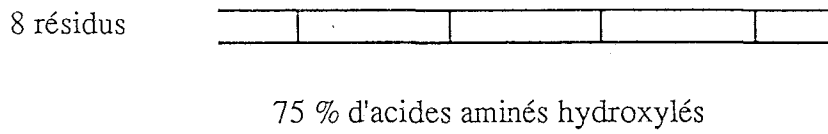
- Le domaine de 8 acides aminés TTSTTSAP est le plus fréquemment rencontré. Dans certains cas, sont intercalées des zones non répétitives, non typiques d'une séquence de mucine. Ce motif est localisé sur le chromosome 11 en p15 (MUC 5C) (Voir les principales répétitions observées dans MUC 5C ci-dessous).

Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
Thr	Thr	Ser	Thr	Thr	Ser	Ala	<b>Ser</b>
Thr	<b>Ala</b>	Ser	Thr	Thr	Ser	<b>Gly</b>	Pro
Thr	Thr	Ser	Thr	Thr	Ser	<b>Gly</b>	Pro

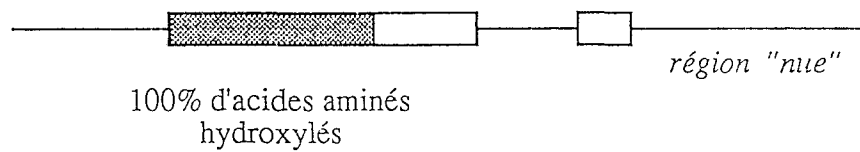
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
SIB 124	1	CAC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TCG	ATC	ACC	ACC	ACC	GAG	GCC	CCC	TCA
	2	AGC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TCG	ATC	ACC	ACC	ACC	GAC	TCG	ATC	GTC
	3	CAC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TTG	ATC	ACC	ATC	ACT	GAG	ACC	ACC	TCA
	4	CAC	AGT	ACT	CCC	AGC	TAC	ACT	ACC	TCA	ATC	ACC	AGC	ACC	AAG	ACC	CCC	TCA
	5	CAC	AGT	ACT	CCC	AGC	TAC	ACT	ACC	TCA	ATC	ACC	ACC	ACT	GAG	ATC	CCA	TCA
	6	CAC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TCA	ATC	ATC	ACC	ACC	GAG	ACC	ACA	TCC
		CAC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TCT	ATG	TTC						
SIB 139	1	AAC	AGT	ACT	GCC	AGC	TTG	ACT	TCT	TCA	ATG	ACC	ACC	ACC	GAG	ACC	ACC	TCA
	2	CAC	AGT	ACT	CCC	AGC	TTG	ACT	TCT	TCA	ATG	ACC	GCC	ACC	GAG	ACC	ACC	TCA
	3	CAC	AGT	ACT	CCC	AGC	TTC	ACT	TCT	TCC	ATC	ACC	ACG	ACC	GAG	ACC	ACC	TCC
	4	AAC	AGT	ACT	GCC	AGC	TTG	ACT	TCT	TCC	ATC	ACC	ACC	ACC	GAG	ACC	ACC	TCC
SIB 124	1	His	Ser	Thr	Pro	Ser	Phe	Thr	Ser	Ser	Ile	Thr	Thr	Thr	Glu	<i>Ala</i>	<i>Pro</i>	Ser
	2	<i>Ser</i>	Ser	Thr	Pro	Ser	Phe	Thr	Ser	Ser	Ile	Thr	Thr	Thr	<i>Asp</i>	<i>Ser</i>	<i>Ile</i>	<i>Val</i>
	3	His	Ser	Thr	Pro	Ser	Phe	Thr	Ser	<i>Leu</i>	Ile	Thr	<i>Ile</i>	Thr	Glu	Thr	Thr	Ser
	4	His	Ser	Thr	Pro	Ser	<i>Tyr</i>	Thr	<i>Thr</i>	Ser	Ile	Thr	<i>Ser</i>	Thr	<i>Lys</i>	Thr	<i>Pro</i>	Ser
	5	His	Ser	Thr	Pro	Ser	<i>Tyr</i>	Thr	<i>Thr</i>	Ser	Ile	Thr	Thr	Thr	Glu	<i>Ile</i>	<i>Pro</i>	Ser
	6	His	Ser	Thr	Pro	Ser	Phe	Thr	Ser	Ser	Ile	<i>Ile</i>	Thr	Thr	Glu	Thr	Thr	Ser
		His	Ser	Thr	Pro	Ser	Phe	Thr	Ser	Ser	<i>Met</i>	<i>Phe</i>						
SIB 139	1	<i>Asn</i>	Ser	Thr	<i>Ala</i>	Ser	<i>Leu</i>	Thr	Ser	Ser	<i>Met</i>	Thr	Thr	Thr	Glu	Thr	Thr	Ser
	2	His	Ser	Thr	Pro	Ser	<i>Leu</i>	Thr	Ser	Ser	<i>Met</i>	Thr	<i>Ala</i>	Thr	Glu	Thr	Thr	Ser
	3	His	Ser	Thr	Pro	Ser	Phe	Thr	Ser	Ser	Ile	Thr	Thr	Thr	Glu	Thr	Thr	Ser
	4	<i>Asn</i>	Ser	Thr	<i>Ala</i>	Ser	<i>Leu</i>	Thr	Ser	Ser	Ile	Thr	Thr	Thr	Glu	Thr	Thr	Ser

Figure n°7 : Unités répétitives des clones de mucines intestinales (MUC 3).

# I Famille I (répétitive)



# II Famille II



# III Famille III (alternée)

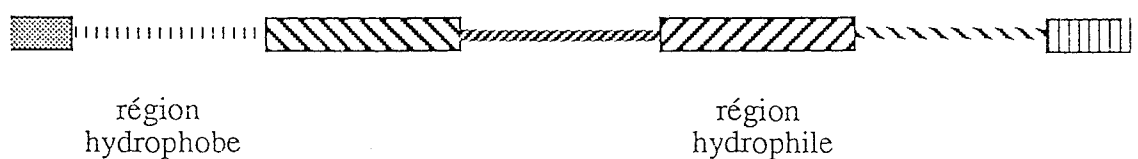


Figure n°8 : Schématisation des familles structurales de mucines trachéobronchiques humaines

- Le domaine de 16 acides aminés (TSSASTGHATPLPVTD) est localisé sur le chromosome 3 (MUC 4). 39 répétitions de ce domaine constitué de 48 nucléotides ont été relevées. Les domaines ne sont pas rigoureusement analogues, néanmoins 7 positions sont parfaitement conservées d'un domaine à l'autre (voir les principales répétitions observées dans MUC 4 ci-dessous).

Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Asp  
 Thr Ser Ser **Val** Ser Thr **Ser** His Ala Thr **Ser** Leu Pro Val Thr Asp  
 Thr Ser Ser Ala Ser Thr Gly **AspThr** Thr Pro Leu **His** Val Thr Asp  
 Thr Ser Ser Ala Ser Thr **Ser** His Ala Thr **Ser** Leu Pro Val Thr Asp

• la **famille II**, localisée à la fois sur les chromosomes 11 et 13, possède un agencement moins structuré, montrant cependant des zones hydrophiles à 100 % de résidus hydroxylés, dont 75 % de thréonine (MUC 5A).

• la **famille III**, localisée sur le chromosome 11 en p15, semble montrer une organisation alternée en domaines hautement glycosylables et en zones peu glycosylables (MUC 5B). Ce raisonnement est vrai pour la séquence peptidique, mais notre travail de thèse a montré que ceci n'était pas vrai pour la séquence nucléotidique.

De l'ensemble de ces résultats, il ressort qu'il existe une grande hétérogénéité structurale des axes peptidiques de mucines trachéobronchiques humaines. Quelques séquences potentielles de N-glycosylation ont été relevées. Cinq acides aminés C-terminaux distincts ont été mis en évidence, mais aucun acide aminé N-terminal n'a pu être caractérisé.

L'analyse de l'expression des mucines, par *Northern blot*, démontre une extrême hétérogénéité en nombre et en taille des transcrits (continuum de 20 kb à 0,4 kb). Elle démontre également une spécificité tissulaire variable : les sondes de la famille III reconnaissent des motifs ubiquitaires communs à l'ensemble des mucines étudiées (muqueuses trachéo-bronchique, colique, gastrique, lignées tumorales mammaires MCF7 et VHB1).



mucines étudiées (muqueuses trachéo-bronchique, colique, gastrique, lignées tumorales mammaires MCF7 et VHB1).

Les sondes de la famille I répétitive se caractérisent, au contraire, par une spécificité tissulaire d'expression plus étroite :

- spécificité limitée à la muqueuse trachéo-bronchique et gastrique pour le motif de 8 acides aminés ;

- spécificité limitée à la muqueuse trachéo-bronchique et à la muqueuse colique pour le motif de 16 acides aminés.

La purification des différents anticorps spécifiques à partir de l'antisérum polyclonal fut réalisée sur les protéines de fusion synthétisées par chacun des 20 clones d'*E. coli*/λgt 11 recombinants. Ces anticorps oligoclonaux permirent alors, sur coupe histologique de muqueuse trachéo-bronchique, de confirmer la positivité des clones :

- seules sont reconnues les cellules sécrétrices de mucines, cellules caliciformes, cellules des glandes muqueuses.

- en aucun cas les cellules séreuses ne sont reconnues, ni le tissu conjonctif.

Sur une seule expérience, les cellules caliciformes ont été reconnues par tous les anticorps et pourraient donc avoir un potentiel de synthèse large alors que les cellules des glandes muqueuses, qui ne sont reconnues que par quelques-uns des anticorps, auraient une capacité sécrétoire plus spécifique (66).

Pour conclure sur la connaissance de ces premières séquences d'ADNc de mucines humaines, la caractéristique de ces glycoprotéines est donc d'être structurée en domaines répétitifs et cela quel que soit le tissu considéré. Néanmoins, d'autres organisations structurales peptidiques pourraient coexister, comme l'alternance de domaines nus / glycosylables.

#### **\* Les mucin-like de tumeur mammaire**

Des difficultés identiques à celles rencontrées avec les "mucines typiques" ont été observées lors de la détermination des séquences des axes peptidiques de

*mucin-like*. Les techniques de génétique moléculaire ont conduit plusieurs équipes à publier la première séquence d'ADNc complète d'une *mucin-like*, l'épisialine ou PEM associée à la tumeur du sein ; son gène *MUC1* est localisé sur le chromosome 1 (67).

Elle est constituée, en grande partie, de la répétition en tandem, 40 à 90 fois selon les sujets (polymorphisme interindividuel), d'un motif de 20 acides aminés contenant 25 % de résidus hydroxylés (figure n°9).

Ces domaines sont rigoureusement identiques au centre de la structure répétitive, le degré d'analogie diminue au fur et à mesure que l'on s'éloigne vers les extrémités N et C terminales. Il existe deux transcrits dont l'un est pourvu d'une région d'ancrage intramembranaire, le deuxième transcrit coderait pour une forme sécrétée de l'épisialine.

Comme nous venons de le voir, la connaissance structurale des mucines n'est qu'à peine ébauchée en 1992. L'existence singulière de très nombreux transcrits complique encore le problème.

De très nombreuses questions restent posées :

- Il semble exister trois familles peptidiques dans les mucines trachéo-bronchiques humaines. Ces familles proviennent-elles d'un même gène ou de trois gènes différents localisés en 11 p15 ?

- Le gène de mucines intestinales (*MUC 2*) localisé en 11 p15 fait-il partie du ou des gènes de mucines trachéo-bronchiques ?

Existe-t-il une régulation tissulaire d'expression de ces gènes ?

- Comment expliquer l'extrême hétérogénéité des ARN messagers ?

- Un ARN messager code-t-il pour un type peptidique ou pour plusieurs ? L'organisation intime de cet ARN messager est-elle identique quelle que soit la muqueuse considérée ?

- Peut-on considérer les mucines "répétitives" comme des *mucin-like* alors que les mucines de type alterné seraient les mucines typiques ?

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MUC 1	1	COG	GGC	TCC	ACC	GOC	CCC	CCA	GOC	CAC	GGT	GTC	ACC	TCG	GOC	COG	GAC	ACC	AGG	COG	CCC
	2	COG	GGC	TCC	ACC	GOG	CCC	GCA	GOC	CAC	GGT	GTC	ACC	TCG	GOC	COG	GAC	ACC	AGG	COG	CCC
	3	TTG	GGC	TCC	ACC	GOC	CCT	CCA	GTC	CAC	AAT	GTC	ACC	TCG	GOC	TCA	GGC	TCT	GCA	TCA	GGC
	4	CTG	GGC	TCC	ACC	ACC	CCG	CCA	GOC	CAC	GAT	GTC	ACC	TCA	GOC	CCG	GAC	AAC	AAG	CCA	GCC
MUC 1	1	Pro	Gly	Ser	Thr	Ala	Pro	Pro	Ala	His	Gly	Val	Thr	Ser	Ala	Pro	Asp	Thr	Arg	Pro	Pro
	2	Pro	Gly	Ser	Thr	Ala	Pro	<b>Ala</b>	Ala	His	Gly	Val	Thr	Ser	Ala	Pro	Asp	Thr	Arg	Pro	Pro
	3	<b>Leu</b>	Gly	Ser	Thr	Ala	Pro	Pro	<b>Val</b>	His	<b>Asn</b>	Val	Thr	Ser	Ala	<b>Ser</b>	<b>Gly</b>	<b>Ser</b>	<b>Ala</b>	<b>Ser</b>	<b>Gly</b>
	4	<b>Leu</b>	Gly	Ser	Thr	<b>Thr</b>	Pro	Pro	Ala	His	<b>Asp</b>	Val	Thr	Ser	Ala	Pro	Asp	<b>Asn</b>	<b>Lys</b>	Pro	<b>Ala</b>

Figure n°9 : Unités répétitives de "mucin-like" humaines (MUC 1).

Dans ce travail, nous avons tenté de répondre à deux de ces questions en réalisant tout d'abord un maximum de séquences d'ADNc dans le but de découvrir si les trois familles peptidiques étaient réellement distinctes.

Le deuxième point de notre étude a été de démontrer l'existence réelle d'une alternance peptidique de domaines hydrophobes et de domaines hydrophiles glycosylables dans les mucines trachéo-bronchiques humaines.

## RESULTATS

### A - STRATEGIE GENERALE

L'unité n°16 INSERM, dans laquelle ce travail a été réalisé, possède une grande expérience dans le domaine des mucines trachéobronchiques humaines. Il existait donc une assez grande quantité de mucines respiratoires sous forme pure. Avant mon arrivée dans le laboratoire, une stratégie avait été mise en place dans le but d'étudier le ou les axes peptidiques de ces mucines grâce aux techniques de biologie moléculaire.

Les mucines trachéobronchiques natives ont été hydrolysées par la pronase. Les glycopeptides obtenus ont été séparés par échange ionique puis déglycosylés chimiquement par le TFMS.

Après vérification de la qualité de déglycosylation de chacune des fractions glycopeptidiques, un pool a été recréé contenant chaque fraction en quantité équimoléculaire.

Ce mélange injecté chez le lapin a permis l'obtention d'un immunosérum polyclonal dirigé contre le ou les axes peptidiques de mucines trachéobronchiques humaines. La qualité de ces anticorps a été vérifiée par immunohistochimie. Ceux-ci ont permis le criblage d'une banque d'ADNc en vecteur d'expression ( $\lambda$ gt11). Les clones positifs radiomarqués ont permis cette fois le criblage d'une banque d'ADNc en vecteur  $\lambda$ gt10.

L'ensemble de ces clones a été séquencé.

Dans ce travail nous nous sommes particulièrement intéressé aux clones appartenant à la famille alternée.

L'ensemble de ces sondes a été testé sur Southern et Northern blot.

## B - RESULTATS

### 1) Criblage de la banque d'ADNc en vecteur $\lambda$ gt11 de muqueuse trachéobronchique humaine

Les détails techniques précis sont donnés dans l'appendice.

Environ 400 000 clones recombinants ont été criblés par l'immunsérum polyclonal. Seuls 5 clones se sont révélés positifs. Les ADNc obtenus ont une taille variant de 0.4 à 1.8 Kb.

CLONE	TAILLE	FAMILLE
JER 58	0,83 Kb	1
JER 61	0,40 Kb	1
JER 47	1,50 Kb	2
JER 28	0,56 Kb	3
JER 57	1,83 Kb	3
JER 62	0,40 Kb	3

Ce rendement peut paraître faible mais le système de clonage utilisé n'offre qu'une chance sur six d'obtenir la bonne traduction peptidique dans le bon cadre de lecture.

A partir de ces 5 protéines de fusion, 5 anticorps oligoclonaux ont été préparés (voir appendice technique).

Chacun de ces anticorps a été testé sur coupes histologiques. Les résultats obtenus sont résumés dans le tableau n°3 et confirment les précédentes expériences.

Dans tous les cas les cellules caliciformes sont marquées de façon plus ou moins intense. Ceci peut être attribué soit à un niveau d'expression différent des différents ARN correspondants, soit à un titre différent des anticorps oligoclonaux.

Anticorps anti-	MUQUEUSE TRACHEALE			MUQUEUSE COLIQUE
	CELLULES CALICIFORMES	CELLULES DES GLANDES MUQUEUSES	CELLULES SEREUSES	
JER58	++++	++++	-	-
JER61	+++	-	-	-
JER47	+++	-	-	-
JER28	+	+	-	-
JER57	++++	++++	-	-
JER62	+++	-	-	-

Tableau n°3 : Anticorps "oligoclonaux" sur différentes coupes histologiques humaines

Par contre, en ce qui concerne les cellules des glandes muqueuses, seuls deux des anticorps oligoclonaux sont capables de les reconnaître (JER 58 et JER 57).

Ce marquage des cellules spécialisées dans la synthèse des mucines et l'absence de ce même marquage dans les cellules séreuses nous permet d'affirmer la positivité de ces cinq premiers clones.

## **2) Criblage de la banque d'ADNc en vecteur $\lambda$ gt10 de muqueuse trachéobronchique humaine**

L'ADNc le plus long obtenu lors du criblage de la 1ère banque avait une taille de 1830 bp. Nous avons donc pensé obtenir des ADNc de plus grande taille qui permettrait de mieux rendre compte de l'organisation des apomucines trachéobronchiques humaines et en particulier de savoir s'il existe des hybrides contenant différents types de séquences.

Un ADNc de grande taille permettrait peut être aussi d'obtenir un hybride MUC2-MUC5 prouvant ainsi l'existence d'un seul gène.

C'est donc avec beaucoup d'espoir que nous avons entamé ce nouveau criblage dans la mesure où nous avons vérifié la représentativité de notre banque (certains inserts atteignent 5 à 6 Kb).

Une autre priorité a été pour nous l'étude plus approfondie de l'organisation de type alterné.

Plusieurs raisons expliquent ce choix :

- cette organisation correspond au modèle couramment admis par la littérature mais non encore prouvé de façon expérimentale ;

- les premières séquences nucléotidiques de clones de cette famille montrent de grandes homologues qui se caractérisent cependant par une traduction peptidique différente ;

- cette organisation de type alterné est présente dans les muqueuses d'origine tissulaire différente.



Nous avons donc entrepris le criblage de cette banque  $\lambda$ gt10 par un mélange de sondes représentatives des trois organisations (JER 28, 29, 30, 57, 61, 58, 47 et 71).

Pour chaque clone positif, une purification est effectuée par un deuxième tour et cette fois chacune des 8 sondes est testée séparément.

16 nouveaux clones ont ainsi été obtenus avec une taille variant de : 50 pb à 1631 pb

11 de ces clones ont été extraits par les sondes JER 58 et JER 61 (famille répétitive I). Ces clones sont de très petites tailles inférieures à 100bp et n'ont donc pas été séquencés

2 clones ont été extraits par la sonde JER 47 (famille II) : JUL 31 et JUL 32

et 3 clones par les sondes JER 28 et JER 57 (famille alternée III) : JUL7, JUL8 et JUL10.

Le criblage a été poursuivi avec un mélange de sondes constitué uniquement de la famille III (JUL 7, 8, 10 et JER 28, 29, 30, 57, 71). 3 clones ont ainsi été obtenus et tous par la sonde JUL 8.

Clone	Taille (kb)	Famille
JUL 7	1,63	3
JUL 10	0,99	3
JUL 31	0,60	2
JUL 32	0,35	2
JUL 8	0,70	3
JUL 33	0,30	3
JUL 35	0,20	3
JUL 36	0,70	3

### 3) Séquence nucléotidique et acides aminés déduite

Les séquences totales ou partielles de tous ces clones nous ont permis de confirmer la classification préétablie en 3 familles.

#### *Séquence nucléotidique et peptidique de la famille I*

Dans cette famille nous ne nous sommes intéressé qu'à deux clones JER 58 et JER 61.

Les compositions en acides aminés de ces 2 clones sont données dans le tableau n°5. Ils possèdent respectivement 55 et 57 % de résidus hydroxylés.

Le rapport Sérine / Thréonine varie de 0.5 à 0.7.

Le nombre d'acides aminés hydrophobes (Isoleucine, Leucine et Valine) est très faible ainsi que les acides aminés dicarboxyliques et diamminés. On note aussi l'absence de résidus Cystéine dans le clone JER 61.

En ce qui concerne le clone JER 58 (figure n°10 a,b), 33 unités répétitives de 8 acides aminés existent avec à quatre reprises l'enchaînement TTSTTSAS et cinq fois la séquence de référence TTSTTSAP.

Le clone JER 61 (figure n°11a,b) révèle 17 unités répétitives. Quatre sont constituées de l'enchaînement TTSTTSAP ; trois de l'enchaînement TASTTSGP. On note une nouvelle fois la présence d'une unité TTSTTSAS.

L'utilisation de ces clones comme sondes radioactives pour l'étude en Southern et en Northern blots sera développée plus loin.

Acides Aminés	JER 58		JER61			
	nbre	%	Nt		Ct	
			nbre	%	nbre	%
Thréonine	112	40,1	25	34,7	26	35,6
Sérine	67	24,0	14	19,4	15	20,6
Proline	43	15,4	11	15,4	12	16,5
Alanine	20	7,2	8	11,1	7	9,6
Glycine	14	5,0	4	5,5	4	5,5
Arginine	4	1,4	2	2,8	2	2,7
Asparagine	1	0,4	0	/	0	/
Ac. Aspartique	0	/	0	/	0	/
Cystéine	0	/	0	/	0	/
Glutamine	0	/	0	/	2	2,7
Ac. Glutamique	2	0,7	1	1,4	0	/
Histidine	0	/	0	/	0	/
Isoleucine	5	1,8	0	/	0	/
Leucine	0	/	4	5,5	3	4,1
Lysine	1	0,4	0	/	0	/
Méthionine	0	/	0	/	0	/
Phénylalanine	0	/	0	/	0	/
Tryptophane	0	/	1	1,4	0	/
Tyrosine	0	/	0	/	0	/
Valine	10	3,6	2	2,8	2	2,7
Total	279	100	72	100	73	100

Tableau n°5

	3		9		15		21		27		33		39		45
1	CCC	TCG	GCT	CCT	ACA	ACC	AGC	ACA	ACT	TCT	GGT	CCT	GGA	ACT	ACC
46	CCA	AGC	CCT	GTT	CCC	ACG	ACC	AGC	ACA	ACC	TCT	GCC	CCT	ACA	ACA
91	AGA	ACA	ACT	TCT	GCT	CCT	AAA	AGC	AGC	ACA	ACC	TCT	GCC	GCT	ACA
136	ACC	AGC	ACA	ACC	TCT	GGT	CCT	GAA	ACT	ACT	CCT	AGA	CCT	GTT	CCT
181	ACC	ACC	AGC	ACA	ACC	TCT	TCT	CCT	ACA	ACC	AGC	ACA	ACC	TCT	GCT
226	CCT	ACA	ACC	AGC	ACA	ACC	TCT	GCT	TCT	ACA	ACC	AGC	ACA	ACC	TCT
271	GGT	GCT	GGA	ACT	ACT	CCC	AGC	CCT	GTT	CCC	ACC	ACC	AGC	ACA	ACC
316	TCT	GCT	CCT	ACA	ACC	AGA	ACA	ACC	TCT	GCC	CCT	ATA	AGC	AGC	ACA
361	ACC	TCT	GCC	ACT	ACA	ACC	AGC	ACA	ACC	TCT	GGT	CCT	GGA	ACT	ACT
406	CCC	AGC	CCT	GTT	CCT	ACC	ACG	AGC	ACA	ACC	TCT	GCT	CCT	ACA	ACC
451	AGC	ACA	ACC	TCT	GGT	CCT	GGA	ACT	ACT	CCC	AGT	GCT	GTT	CCC	ACC
496	ACC	AGT	ATA	ACC	TCT	GTA	CCT	ACA	ACC	AGT	ACA	AAC	TCT	GCC	CCT
541	ATA	AGC	AGC	ACA	ACC	TCT	GCC	ACT	ACA	ACC	AGT	AGA	ATC	TCT	GGT
586	CCT	GAA	ACT	ACT	CCC	AGC	CCT	GTT	CCT	ACC	GCC	AGC	ACA	ACC	TCT
631	GCT	TCT	ACA	ACT	AGC	ACA	ACC	TCT	GGT	CCT	GGA	ACT	ACT	CCC	AGC
676	CCT	GTT	CCT	ACC	ACC	AGC	ACA	ATC	TCT	GTT	CCT	ACC	ACC	AGC	ACA
721	ACT	TCT	GCT	TCT	ACA	ACC	AGC	ACA	ACC	TCT	GCT	TCT	ACA	ACC	AGC
766	ACA	ACC	TCT	GGT	CCT	GGA	ACT	ACT	CCC	AGC	CCT	GTT	CCC	ACC	ACC
811	AGC	ACA	ACC	TCT	GCT	CCC	ACA	ACA	AGC	G					

Figure n°10a : Séquence nucléotidique de JER 58

1					Pro	Ser	Ala	Pro
5	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
13	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
21	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
29	Thr	Thr	Arg	Thr	Thr	Ser	Ala	Pro
37	Lys	Ser	Ser	Thr	Thr	Ser	Ala	Ala
45	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
53	Glu	Thr	Thr	Pro	Arg	Pro	Val	Pro
61	Thr	Thr	Ser	Thr	Thr	Ser	Ser	Pro
69	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
77	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Ser
85	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Ala
93	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
101	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
109	Thr	Thr	Arg	Thr	Thr	Ser	Ala	Pro
117	Ile	Ser	Ser	Thr	Thr	Ser	Ala	Thr
125	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
133	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
141	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
149	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
157	Gly	Thr	Thr	Pro	Ser	Ala	Val	Pro
165	Thr	Ser	Ile	Thr	Ser	Val	Pro	Thr
173	Thr	Ser	Thr	Asn	Ser	Ala	Pro	Ile
181	Ser	Ser	Thr	Thr	Thr	Ser	Ala	Thr
189	Thr	Thr	Ser	Arg	Ile	Ser	Gly	Pro
197	Glu	Thr	Thr	Pro	Ser	Pro	Val	Pro
205	Thr	Ala	Ser	Thr	Thr	Ser	Ala	Ser
213	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
221	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
229	Thr	Thr	Ser	Thr	Ile	Ser	Val	Pro
237	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Ser
245	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Ser
253	Thr	Thr	Ser	Thr	Thr	Ser	Gly	Pro
261	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
269	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
277	Thr	Thr	Ser					

Figure n°10b : Séquence peptidique de JER 58

	3		9		15		21		27		33		39		45
1	CCT	GGG	TCC	TGG	AAC	TAC	TCC	CAG	CCC	TGT	CTC	CCA	CCA	CCA	GCA
46	CAA	CCT	CTG	CTC	CTA	CAA	CCA	GAA	CAA	CCT	CTG	CCT	CTA	CAG	CCA
91	GCA	CAA	CCT	CTG	GTC	CTG	GAA	GTA	CTC	CCA	GCC	CTG	TTC	CCA	CCA
136	CCA	GCA	CAA	CCT	CTG	CTC	CTA	CAA	CCA	GAA	CAA	CCC	CTG	CCT	CTA
181	CAG	CCA	GCA	CAA	CCT	CTG	GTC	CTG	GAA	CTA	CTC	CAG	CCT		

	3		9		15		21		27		33		39		45
1	TGG	AAC	TAC	TCC	CAG	CCC	TGT	TCC	CAC	CAC	CAG	CAC	AAC	CTC	TGC
46	TCC	TAC	AAC	CAG	AAC	AAC	CTC	TGC	CTC	TAC	AGC	CAG	CAC	AAC	CTC
91	TGG	TCC	TGG	AAG	TAC	TCC	CAG	CCC	TGT	TCC	CAC	CAC	CAG	CAC	AAC
136	CTC	TGC	TCC	TAC	AAC	CAG	AAC	AAC	CCC	TGC	CTC	TAC	AGC	CAG	GAC
181	AAC	CTC	TGG	TCC	TGG	AAC	TAC	TCC	CAG	CCC	TGT	TCC	CAC	CAC	AAG
226	CAC	AAC	CTC	TGC	TTC	TAC	G								

Figure n°11a : Séquences nucléotidiques de JER 61 (5' et 3')

Nt						Trp	Val	Leu
	Glu	Leu	Leu	Pro	Ala	Leu	Ser	Pro
	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
	Thr	Thr	Arg	Thr	Thr	Ser	Ala	Ser
	Thr	Ala	Ser	Thr	Thr	Ser	Gly	Pro
	Gly	Ser	Thr	Pro	Ser	Pro	Val	Pro
	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
	Thr	Thr	Arg	Thr	Thr	Pro	Ala	Ser
	Thr	Ala	Ser	Thr	Thr	Ser	Gly	Pro
	Gly	Thr	Thr	Pro	Ala			
Ct	Thr	Pro	Ala	Gln	Pro	Leu	Leu	Leu
	Gln	Thr	Arg	Thr	Thr	Ser	Ala	Ser
	Thr	Ala	Ser	Thr	Thr	Ser	Gly	Pro
	Gly	Ser	Thr	Pro	Ser	Pro	Val	Pro
	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Pro
	Thr	Thr	Arg	Thr	Thr	Pro	Ala	Ser
	Thr	Ala	Ser	Thr	Thr	Ser	Gly	Pro
	Gly	Thr	Thr	Pro	Ser	Pro	Val	Pro
	Thr	Thr	Ser	Thr	Thr	Ser	Ala	Ser
	Thr							

Figure n°11b : Séquences peptidiques de JER 61 (N ET Ct)

## ***Séquence nucléotidique et peptidique de la famille II***

La composition en acides aminés du clone JER 47 dont la séquence (ci-dessous) n'est pas finie est donnée dans le tableau n°6.

**Nt**

CCA AAG ACC AGC ACA AGC CAT GTT TCC ATA TCC AAG ACA ACC CAC TCC  
Pro Lys Thr Ser Thr Ser His Val Ser Ile Ser Lys Thr Thr His Ser

CAA CCA GTC ACC AGC GAC TGT CAT CCC CGG TGC ACC TGG ACC AAA TGG  
Gln Pro Val Thr Arg Asp Cys His Pro Arg Cys Thr Trp Thr Lys Trp

TTT GAT GTG GAC TTT CCA TCC CCT GGA CCC CAC GGT GGG ACA AGG AAA  
Phe Asp Val Asp Phe Pro Ser Pro Gly Pro His Gly Gly Thr Arg Lys

CCT ACA ACA ACA TCA TCA GGA GTG --- ---  
Pro Thr Thr Thr Ser Ser Gly Val --- ---

CCC GTG ACC TCC ACA TCT GTG ACA GCT CCT AGC ACC CTA GTG GGA GAG  
Pro Val Thr Ser Thr Ser Val Thr Ala Pro Ser Thr Leu Val Gly Glu

CCA CCA GCC CAA CTC AGA GGT CAC CTC CTC TTG GCA GAA ATC CAG GAC  
Pro Pro Ala Gln Leu Arg Gly His Leu Leu Leu Ala Glu Ile Gln Asp

AAC CAC TTT GGT GAC AAG CAG CAT AAC CTC CAC TAC ACA GAC CAG CAC  
Asn His Phe Gly Asp Lys Gln His Asp Leu His Tyr Thr Asp Gln His

AAC CTC TGC CCC TAC

Asn Leu Cys Pro Tyr **Gt**

La sonde JER 47 a permis l'obtention de 2 clones JUL 31 et JUL 32.

La séquence du clone JUL 31 est incomplète : 414 bases sur 600. 90 % de la séquence nucléotidique est constituée des deux seules bases C et A (figure n°12 et le tableau n°7).

La traduction peptidique d'une telle séquence montre la présence de 90 % de résidus hydroxylés et 10 % de résidus alanine.



Acides Aminés	JER 47			
	Nt		Ct	
	nbre	%	nbre	%
Thréonine	11	19,6	5	9,4
Sérine	8	14,3	3	5,7
Proline	6	10,7	5	9,4
Alanine	0	/	3	7,7
Glycine	4	7,1	3	5,7
Arginine	3	5,4	1	1,9
Asparagine	0	/	2	3,8
Ac. Aspartique	3	5,4	4	7,5
Cystéine	2	3,6	1	1,9
Glutamine	1	1,8	4	7,5
Ac. Glutamique	0	/	2	3,8
Histidine	4	7,1	5	9,3
Isoleucine	1	1,8	1	1,9
Leucine	1	1,8	7	13,2
Lysine	4	7,1	1	1,9
Méthionine	0	/	0	/
Phénylalanine	2	3,6	1	1,9
Tryptophane	2	3,6	1	1,9
Tyrosine	0	/	2	3,8
Valine	4	7,1	3	5,7
Total	56	100	53	100

Tableau n°6

Figure N° 12 : Séquence nucléotidique et peptidique de JUL 31

```

      10      20      30      40      50      60
      |      |      |      |      |      |
TCGTGACCCAGATTTAGCAATGTTGCACACCACCACCACCACCACCACCACCACCACCAG
ArgAspProAspLeuAlaMETLeuHisThrThrThrThrThrThrThrThrThrThrThrSer

      70      80      90      100     110     120
      |      |      |      |      |      |
CACCGCCACCGCCACCGCCACCGCCACCGCCACCGCCACCGCCACCACCAGCACCAGCAC
ThrAlaThrAlaThrAlaThrAlaThrAlaThrAlaThrAlaThrThrSerThrSerThr

      130     140     150     160     170     180
      |      |      |      |      |      |
CAGCACCACCACCACCACCACCACCAGCACCACCAGCACCACCACCACCACCACCACCAC
SerThrThrThrThrThrThrThrSerThrThrSerThrThrThrThrThrThrThrThr

      190     200     210     220     230     240
      |      |      |      |      |      |
CACCACCACCACCACCAGCACCGCCACCGCCACCGCCACCGCCACCGCCACCGCCACCAC
ThrThrThrThrThrSerThrAlaThrAlaThrAlaThrAlaThrAlaThrAlaThrThr

      250     260     270     280     290     300
      |      |      |      |      |      |
CAGCACCACCACCAGCACCAGCACCAGCACCACCACCACCACCACCACCAGCACCACCAG
SerThrThrThrSerThrSerThrSerThrThrThrThrThrThrThrThrSerThrThrSer

      310     320     330     340     350     360
      |      |      |      |      |      |
CACCACCACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCACCACCAC
ThrThrThrThrThrThrThrThrThrThrThrThrThrThrSerThrThrThrThrThr

      370     380     390     400
      |      |      |      |
CAGCACCACCAGCACCACCAGCACCACCACCACCACCACCACCACCACCG
SerThrThrSerThrThrSerThrThrThrThrThrThrHisHisHisAr

```

Acides aminés	JUL 31		JUL 32	
	nbre	%	nbre	%
Thréonine	94	68,6	28	22,0
Sérine	17	12,4	5	4,2
Proline	1	0,7	13	11,0
Alanine	14	10,2	5	4,2
Glycine	0	/	5	4,2
Arginine	2	1,5	9	7,6
Asparagine	0	/	6	5,1
Ac. Aspartique	2	1,5	2	1,7
Cystéine	0	/	6	5,1
Glutamine	0	/	6	5,1
Ac. Glutamique	0	/	3	2,5
Histidine	4	2,9	4	3,4
Isoleucine	0	/	5	4,2
Leucine	2	1,5	6	5,1
Lysine	0	/	4	3,4
Méthionine	1	0,7	1	0,9
Phénylalanine	0	/	2	1,7
Tryptophane	0	/	3	2,5
Tyrosine	0	/	2	1,8
Valine	0	/	5	4,2
Total	137	100	118	100

Tableau n°7 : Composition en acides aminés de JUL 31 et JUL 32

Le rapport Sérine / Thréonine est très loin de la moyenne attendue puisqu'il est de 0,15 avec 17 Sérines pour 109 Thréonines. Cette séquence très particulière répond très bien à la définition des mucines caractérisées par une grande richesse en acides aminés hydroxylés. Elle permet d'envisager la présence d'une touffe de glycosylation très importante. De plus ce déséquilibre entre les résidus Sérine et Thréonine suggère la présence d'autres séquences cette fois riche en Sérine pour pouvoir retrouver la composition en acides aminés typique des mucines.

N'ayant malheureusement pu obtenir la séquence complète, nous ne savons pas si ce domaine hautement glycosylable constitue tout le clone. A la lecture du gel de séquence, il nous a semblé que ce domaine serait suivi d'un domaine moins glycosylable.

La séquence nucléotidique du clone JUL 32 a été totalement déterminée (figure n°13). Elle présente une homologie de 35 pb en 3' OH de sa séquence avec celle de JER 47.

La composition en acides aminés déduite se caractérise par un pourcentage d'acides aminés hydroxylés de 27 % avec un rapport Sérine / Thréonine de 0,19. La Proline représente 10 % de la composition totale et les résidus Cystéine 5 % (tableau n°7).

Quatre des six Cystéine sont situées en 5' de la séquence alors que les deux autres sont situées en 3'. Parallèlement 12 des 15 résidus hydrophobes sont situés dans les régions 5' et 3'. 81 % des résidus hydroxylés sont concentrés au centre de ce clone.

On ne peut pas encore parler d'une organisation peptidique alternant des zones glycosylables hydrophiles et des zones hydrophobes peu glycosylables (figure n°14) caractérisant la famille III que nous verrons dans le paragraphe suivant.

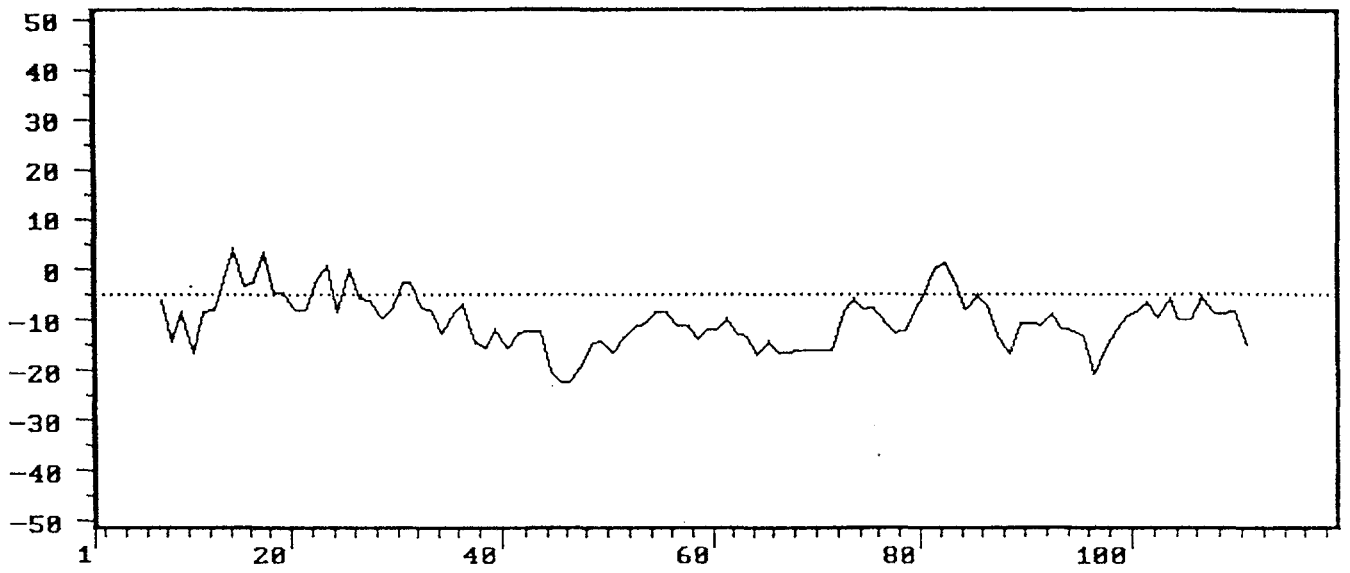
On peut cependant se poser la question de savoir si cette famille II est réelle ou si cette famille est intermédiaire entre la famille répétitive et la famille alternée.

Figure N°13: séquence nucléotidique et peptidique de JUL 32

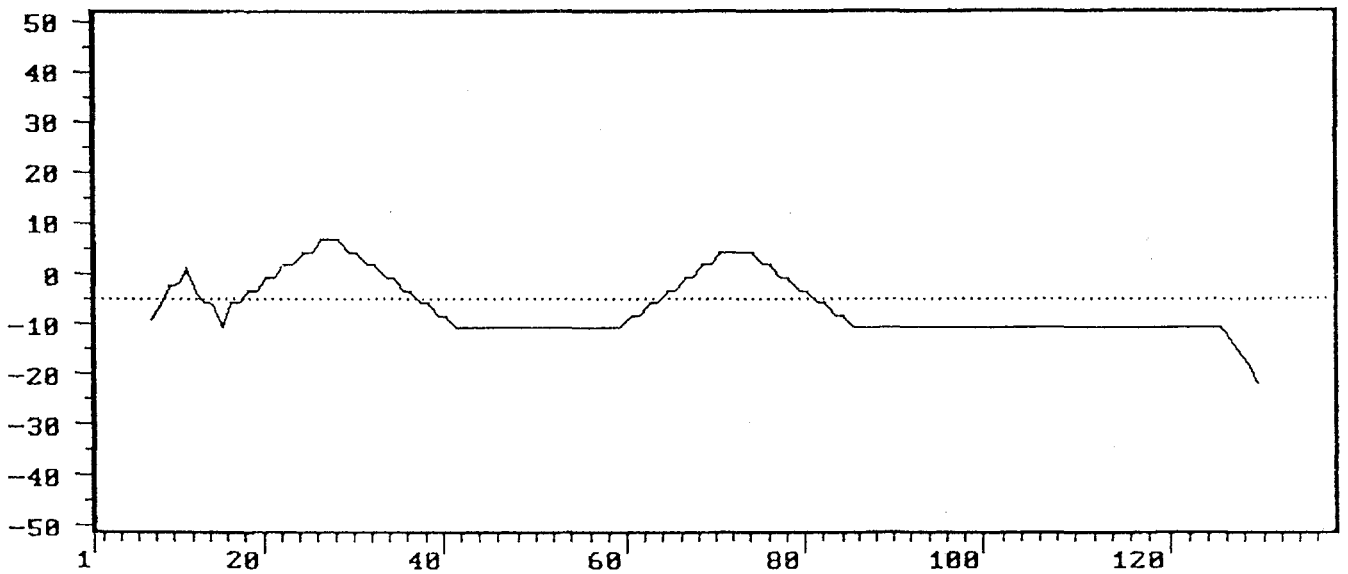
10	20	30	40	50	60
CCCTGAACAAGAACCAGCTCCCACCCATCTGCTACAACCTATGAGATCCGCATCCAGTGTT					
LeuAsnLysAsnGlnLeuProProIleCysTyrAsnTyrGluIleArgIleGlnCysCys					
70	80	90	100	110	120
GCGAGACGGTGAACGTGTGCAGAGACATCACCAGACTGCCAAAGACCGTTCGCAACGACAC					
GluThrValAsnValCysArgAspIleThrArgLeuProLysThrValAlaThrThrArg					
130	140	150	160	170	180
GGCCGACTCCACATCCAACCGGAGCTCAGACCCAGACCACCTTCACCACACACATGCCCT					
ProThrProHisProThrGlyAlaGlnThrGlnThrThrPheThrThrHis <u>MET</u> ProSer					
190	200	210	220	230	240
CGGCCTCCACAGAGCAACCCACGGCAACCTCCAGGGGTGGGCCACAGCAACCAGCGTCA					
AlaSerThrGluGlnProThrAlaThrSerArgGlyGlyProThrAlaThrSerValThr					
250	260	270	280	290	300
CACAGGGCACCCACACCACACTAGTCACCAGAACTGTCATCCCCGGTGCACCTGGACAA					
GlnGlyThrHisThrThrLeuValThrArgAsnCysHisProArgCysThrTrpThrLys					
310	320	330	340		
AGTGGTTCGACGTGACTTCCCTCCCCGGACCCCATGGTTGAGAAAAGG					
TrpPheAspValThrSerLeuProArgThrProTrpLeuArgLys					

Figure N°14 : Courbes d'hydrophilies obtenues à partir des séquences des clones a) JUL 32 et b) JUL 31

a)



b)



### ***Séquence nucléotidique et peptidique de la famille III***

Comme nous l'avons signalé précédemment, c'est surtout à cette famille que nous nous sommes intéressé.

Cas particulier des clones JUL8, JUL 33, JUL 35 et JUL 36 :

Ces clones ont été extraits de la banque en  $\lambda$ gt 10 avec la sonde nucléique JER 57.

Nous avons préféré ne pas mener plus en avant l'étude de ces clones pour les raisons suivantes :

- A part la séquence du clone JER 35, toutes les séquences ont des cadres de lectures fermés.

- Les homologues de séquences nucléotidiques de ces clones et JER 57 sont très faibles.

- Alors que l'ensemble des clones de la famille alternée a été localisé sur le chromosome 11, la séquence du clone JUL 8 a été localisée sur le chromosome 16.

On notera cependant que la traduction peptidique de la séquence JER 35 a toutes les caractéristiques d'une séquence de mucine (figure n°15 et tableau n°8).

5 clones ont tout particulièrement retenu notre attention :

JER 62	180 bp	
JER 57	1830 bp	
JER 28	561 bp	banque $\lambda$ gt 11
JUL 7	1631 bp	
JUL 10	991 bp	banque $\lambda$ gt 10

Les séquences nucléotidiques et peptidiques sont données dans les Figures n°16 à 24. Les compositions en acides aminés sont présentées dans les tableaux n°9 et n°10.

GGG Gly	GCC Ala	TTA Leu	CAC His	TGG Trp	TTG Leu	CAT His	TCA Ser	TTC Phe	CTC Leu	CAA Gln	GAG Glu	AAG Lys	TTC Phe	ATA Ile	TTC Phe
TCA Ser	CAG Gln	GTG Val	GGA Gly	TTA Leu	GGA Gly	CAC His	TTC Phe	CAG Gln	TCA Ser	CCA Pro	GCT Ala	CGC Arg	TGC Cys	TGT Cys	CCT Pro
CCA Pro	CCG Pro	CCA Pro	CCA Pro	CCT Pro	CAA Gln	CTG Leu	GGA Gly	AAT Asn	CTC Leu	TCG Ser	GCC Ala	ACC Thr	ACC Thr	ACC Thr	ACC Thr
ACT Thr	GCC Ala	ACC Thr	ACC Thr	TCC Ser	ATA Ile	GCC Ala	TCC Ser	ACG Thr	GCA Ala	CAT His	GGG Gly	TCC Ser	TCC Ser	ATC Ile	GCC Ala
CTC Leu	GGC Gly	CTC Leu	CAC His	GAC Asp											

Figure n°15 : séquences nucléotidique et peptidique de JUL 35



Acides Aminés	JUL 35	
	nbre	%
Thréonine	8	11,7
Sérine	8	11,7
Proline	7	10,2
Alanine	7	10,1
Glycine	6	8,7
Arginine	1	1,4
Asparagine	1	1,4
Ac. Aspartique	1	1,4
Cystéine	2	2,9
Glutamine	4	5,9
Ac. Glutamique	1	1,4
Histidine	5	7,2
Isoleucine	3	4,3
Leucine	8	11,7
Lysine	1	1,4
Méthionine	0	/
Phénylalanine	4	5,9
Tryptophane	1	1,4
Tyrosine	0	/
Valine	1	1,4
Total	69	100

Tableau n°8

Figure N°16: Séquence nucléotidique de JER 57

	3	9	15	21	27	33	39	45							
1	CTG	GGC	CTC	GAG	TGC	CGT	GCC	AGC	CCA	GCC	TGG	TGT	CCC	TGG	GGG
46	AGT	TGG	GCC	AGG	TCG	TGG	AAT	GCA	GCC	TGG	ACT	TTG	GCC	TGG	TCT
91	GCA	GGA	ACC	GTG	AGC	AGG	TGG	GGA	AGT	TCA	AGA	TGT	GCT	TCA	ACT
136	ATG	AAA	TCC	GTG	TGT	TCT	GCT	GCA	ACT	ACG	CCA	CTG	CCC	CAG	CAC
181	CCC	GGC	CAC	CAG	CTC	TAC	GCC	ATG	CCC	TCC	TCC	ACT	CCG	GGG	ACG
226	ACC	TGG	ATC	CTC	ACA	GAG	CTG	ACC	ACA	ACA	GAC	CAC	TAC	GAC	TGC
271	ATC	CAC	TGG	ATC	CAC	GGC	CAC	CCC	GTC	CTC	CAC	CCC	GGG	AAC	AGC
316	TCC	CCC	TCC	CAA	AGT	GCT	GAC	CAG	CCC	GGC	CAC	CAC	ACC	CAC	AGC
361	CAC	CAG	TTC	CAA	AGC	CAC	TTC	CTC	CTC	CAG	TCC	AAG	GAC	TGC	AAC
406	CAC	CCT	TCC	AGT	GCT	GAC	AAC	GAC	AGC	CAC	CAA	ATC	CAC	AGC	TAC
451	CAG	GTT	ACA	CCC	ATC	CCC	TCC	TCC	ACC	CTT	GGG	ACC	ACC	GGG	ACC
496	CTC	CCA	GAA	CAG	ACC	ACA	CAC	CCC	GTG	GCC	ACC	ATG	TCC	ACA	ATC
541	CAC	CCC	TCC	TCC	ACT	CCG	GAA	ACC	ACC	CAC	ACC	TCC	ACA	GTG	CTG
586	ACC	ACG	AAG	GCC	ACC	ACG	ACA	AGG	GCC	ACC	AGT	TCC	ACG	TCC	ACC
631	CCC	TCC	TCC	ACT	CCG	GGG	ACG	ACC	TGG	ATC	CTC	ACA	GAG	CTG	ACC
676	ACA	GCA	GCC	ACT	ACA	ACT	GCA	GGA	CCG	TCC	TCC	ACC	CCA	GGG	ACC
721	ACC	TGG	ATC	CTC	ACA	GAG	CTG	ACC	ACA	ACA	GCC	ACT	ACG	ACT	GCG
766	TCC	ACT	GGA	TCC	ACG	GCC	ACC	CTG	TCC	TCC	ACC	CCA	GGG	ACC	ACC
811	TGG	ATC	CTC	ACA	GAG	CCG	AGC	ACT	ACA	GCC	ACC	GTG	ACG	GTC	CCA
856	CCG	GAT	CCA	CAG	CCA	CCG	CCT	CCA	CCC	AGG	CAA	CTG	CTG	GCA	CCC
901	ACA	TGT	GAG	CAC	CAC	GGC	CAC	GAC	ACC	CAC	AGT	CAC	CAG	CTC	CAA
946	AGC	CAC	GTG	CCC	TCC	TCC	AGT	CCA	GGG	ACT	GCA	ACT	GCC	CTT	CCA
991	GCA	CTG	AGA	AGC	ACA	GCC	ACC	ACA	CCC	ACA	GCT	ACC	AGC	TTT	ACA
1036	GCC	ATG	GCC	CTC	CTC	CTC	CCT	GGG	CAC	CAC	CTG	GAC	CCG	CCT	ATC
1081	ACA	GAC	CAC	CAC	ACC	CAC	GGC	CTA	CCA	TGT	CCA	CAG	CAA	CCC	TCC
1126	TCA	CTC	CAG	AGA	CTG	TCC	ACA	CCT	CCA	CAG	TGC	TTA	CGC	CAC	GGC
1171	CAC	CAC	AAC	CGG	CGA	CCG	GCT	CTG	TGG	CCA	CCC	CCT	CCT	CCA	CCC
1216	CAG	CAA	CAG	CTC	ACA	CTA	CCA	AAG	TGC	CGA	CTA	CCA	CAA	CCA	CGG
1261	CTT	CAC	AGC	CAC	CCC	CTC	CTC	CAG	CCC	AGG	GAC	GGG	ACT	CAC	GCC
1306	TCC	AGT	GTG	GAT	CAG	CAC	AAC	CAC	CAC	ACC	CAC	AAC	CAC	CAC	ACC
1351	CAC	AAC	CCG	TGG	CTC	CAC	GGT	GAC	CCC	CTC	CTC	CAT	CCC	GGG	GAC
1396	CAC	CCA	CAC	CGC	CAG	AGT	GCT	GAC	CAC	CAC	CAC	CAC	AAC	TGT	GCC
1441	ACT	GGT	TCT	ATG	GCA	ACA	CCC	TCC	TCT	AGC	ACA	CAG	ACC	AGT	GGT
1486	ACT	CCC	CCA	TCA	CTG	ACC	ACC	ACG	GCC	ACT	ACG	ATC	ACG	GCC	ACC
1531	GCG	TCC	ACC	ACC	AAC	CCC	TCC	TCA	ACT	CCA	GGG	ACA	ACA	CCC	ATC
1576	CCC	CCA	GTG	CTG	ACC	AGC	AGC	CAC	CAC	ACC	CGC	AGC	CAC	CAG	CTC
1621	CAA	AGC	ACT	TCC	TCC	TCC	AGT	CCA	AGG	ACT	GCA	ACC	ACC	CTT	CCA
1666	GTG	CTG	ACA	AGC	ACA	GCC	ACA	AAA	TCC	ACA	GCT	ACC	AGC	TTT	ACA
1711	CCC	ATC	CCC	TCC	CTC	ACC	CTG	TGG	ACC	ACG	TGG	ACC	GTC	CCA	GCA
1756	CAG	ACC	ACC	ACA	CCC	ATG	TCC	ACC	ATG	TCC	ACA	ATC	CAC	ACC	TCC
1801	TCT	ACT	CCA	GAG	ACC	ACC	CAC	ACC	TCC	ACG					





Figure N°19: Séquence nucléotidique de JER 28

	3		9		15		21		27		33		39		45
1	CCC	TCC	ACC	CCA	GGG	ACC	ACC	TGG	ATC	CTC	ACA	GAG	CCG	AGC	ACT
46	ACA	GCC	ACC	GTG	ACG	GGT	CCC	ACC	GGA	TCC	ACG	GCC	ACC	GCC	TCC
91	TCC	ACC	CAG	GCA	ACT	GCT	GGC	ACC	CCA	CAT	GTG	AGC	ACC	ACG	GCC
136	ACG	ACA	CCC	ACA	GTC	ACC	AGC	TCC	AAA	GCC	ACT	CCC	TTC	TCC	AGT
181	CCA	GGG	ACT	GCA	ACC	GCC	CTT	CCA	GCA	CTG	AGA	AGC	ACA	GCC	ACC
226	ACA	CCC	ACA	GCT	ACC	AGC	TTT	ACA	GCC	ATC	CCT	CTC	TCC	TGG	GCA
271	CCA	ACT	GGA	CCG	CTA	TCA	CAG	ACC	ACC	ACA	CCA	CGG	CCA	CCA	TGT
316	CCA	CAG	CCA	CAC	CCT	CCT	CCA	CTC	CAG	AGA	CTG	TCC	ACA	CCT	CCA
361	CCA	GTG	CTT	ACC	ACC	ACG	CCA	CCA	CCA	ACC	GGG	CAC	CGG	CTC	TGT
406	GGC	CAC	CCC	CTC	CTC	CAC	CCC	AGC	AAC	AGC	TCA	CAC	TAC	CAA	AGT
451	GCT	GAC	TAC	CAC	AAC	CAC	GGC	TTC	ACA	GCC	ACC	CCC	TCC	TCC	AGC
496	CCA	GGG	ACG	GCA	CGC	ACG	CTT	CCA	GTG	TGG	ATC	AGC	ACA	ACC	ACC
541	ACA	CCC	ACA	ACC	AGA	GGT	TCC								

	Pro	Ser	Thr	Pro	Gly	Thr	Thr	Trp	Trp	Ile	Leu	Thr	Glu	Pro	Ser	Thr
	Gly	Pro	Thr	Gly	Ser	Thr	Ala	Thr	Ala	Ser	Ser	Thr	Ala	Thr	Val	Thr
	Thr	Gln	Ala	Thr	Ala	Gly	Thr	Pro	His	Val	Ser	Thr	Thr	Ala	Thr	Thr
1	Pro	Thr	Val	Thr	Ser	Ser	Lys	Ala	Thr	Pro	Phe	Ser	Ser	Pro	Gly	Thr
	Ala	Thr	Ala	Leu	Pro	Ala	Leu	Arg	Ser	Thr	Ala	Thr	Thr	Pro	Thr	Ala
	Thr	Ser	Phe	Thr	Ala	Ile	Pro	Leu	Ser	Trp	Ala	Pro	Thr	Gly	Pro	Leu
	Ser	Gln	Thr	Thr	Thr	Pro	Arg	Pro	Pro	Cys	Pro	Gln	Pro	His	Pro	Pro
2	Pro	Leu	Gln	Arg	Leu	Ser	Thr	Pro	Pro	Pro	Val	Leu	Thr	Thr	Thr	Pro
	Pro	Pro	Thr	Gly	His	Arg	Leu	Cys	Gly	His	Pro	Leu	Leu	His	Pro	Ser
	Asp	Ser	Ser	His	Tyr	Gln	Ser	Ala	Asp	Tyr	His	Asn	His	Gly	Phe	Thr
1	Ala	Thr	Pro	Ser	Ser	Ser	Pro	Gly	Thr	Ala	Arg	Thr	Leu	Pro	Val	Trp
	Ile	Ser	Thr	Thr	Thr	Thr	Pro	Thr	Thr	Arg	Gly	Ser				

1 : Domaine glycosylable

2 : Domaine nu

Figure n°20 : Séquence peptidique de JER 28

1	Leu Gly Leu Glu Cys Arg Ala Ser Pro Ala Trp Cys Pro Trp Gly
	Ser Trp Ala Arg Ser Thr Asn Ala Ala Trp Thr Leu Ala Trp Ser
	Ala Gly Thr Val Ser Arg Trp Gly Ser Ser Arg Cys Ala Ser Thr
	Met Lys Ser Val Cys Ser Ala Ala Thr Thr Pro Leu Pro Gln His
2	Pro Gly His Gln Leu Tyr Ala Met Pro Ser Ser Thr Pro Gly Thr
	Thr Trp Ile Leu Thr Glu Leu Thr Thr Thr Asp His Tyr Asp Cys
	Ile His Trp Ile His Gly His Pro Val Leu His Pro Gly Asn Ser
	Ser Pro Ser Gln Ser Ala Asp Gln Pro Gly His His Thr His Ser
1	His Gln Phe Gln Ser His Phe Leu Leu Gln Ser Lys Asp Cys Asn
	His Pro Ser Ser Ala Asp Asn Asp Ser His Gln Ile His Ser Tyr
	Gln Val Thr Pro Ile Pro Ser Ser Thr Leu Gly Thr Thr Gly Thr
	Leu Pro Glu Gln Thr Thr His Pro Val Ala Thr Met Ser Thr Ile
1	His Pro Ser Ser Thr Pro Glu Thr Thr His Thr Ser Thr Val Leu
	Thr Thr Lys Ala Thr Thr Thr Arg Ala Thr Ser Ser Thr Ser Thr
	Pro Ser Ser Thr Pro Gly Thr Thr Trp Ile Leu Thr Glu Leu Thr
	Thr Ala Ala Thr Thr Thr Ala Gly Pro Ser Ser Thr Pro Gly Thr
2	Thr Trp Ile Leu Thr Glu Leu Thr Thr Thr Ala Thr Thr Thr Ala
	Ser Thr Gly Ser Thr Ala Thr Leu Ser Ser Thr Pro Gly Thr Thr
	Trp Ile Leu Thr Glu Pro Ser Thr Thr Ala Thr Val Thr Val Pro
	Pro Asp Pro Gln Pro Pro Pro Pro Pro Arg Gln Leu Leu Ala Pro
2	Thr Cys Glu His His Gly His Asp Thr His Ser His Gln Leu Gln
	Ser His Val Pro Ser Ser Ser Pro Gly Thr Ala Thr Ala Leu Pro
	Ala Leu Arg Ser Thr Ala Thr Thr Pro Thr Ala Thr Ser Phe Thr
	Ala Met Ala Leu Leu Leu Pro Gly His His Leu Asp Pro Pro Ile
1	Thr Asp His His Thr His Gly Leu Pro Cys Pro Gln Gln Pro Ser
	Ser Leu Gln Arg Leu Ser Thr Pro Pro Gln Cys Leu Arg His Gly
	His His Asn Arg Arg Pro Ala Leu Trp Pro Pro Pro Pro Pro Pro
	Gln Gln Gln Leu Thr Leu Pro Lys Cys Arg Leu Pro Gln Pro Arg
2	Leu His Ser His Pro Leu Leu Gln Pro Arg Asp Gly Thr His Ala
	Ser Ser Val Asp Gln His Asn His His Thr His Asn His His Thr
	His Asn Pro Trp Leu His Gly Asp Pro Leu Leu His Pro Gly Asp
	His Pro His Arg Gln Ser Ala Asp His His His His Asn Cys Ala
1	Thr Gly Ser Met Ala Thr Pro Ser Ser Ser Thr Gln Thr Ser Gly
	Thr Pro Pro Ser Leu Thr Thr Thr Ala Thr Thr Ile Thr Ala Thr
2	Ala Ser Thr Thr Asn Pro Ser Ser Thr Pro Gly Thr Thr Pro Ile
	Pro Pro Val Leu Thr Ser Ser His His Thr Arg Ser His Gln Leu
1	Gln Ser Thr Ser Ser Ser Ser Pro Arg Thr Ala Thr Thr Leu Pro
	Val Leu Thr Ser Thr Ala Thr Lys Ser Thr Ala Thr Ser Phe Thr
1	Pro Ile Pro Ser Leu Thr Leu Trp Thr Thr Trp Thr Val Pro Ala
	Gln Thr Thr Thr Pro Met Ser Thr Met Ser Thr Ile His Thr Ser
	Ser Thr Pro Glu Thr Thr His Thr Ser Thr

1 : Domaine glycosylable

2 : Domaine nu

Figure n°21 : Séquence peptidique de JER 57

2	{	Val Phe Cys Cys Asn Tyr Gly His Cys Pro Ala Pro Arg Pro Pro
		Ala Leu Arg Pro Cys Pro Pro Pro Leu Arg Gly Arg Pro Gly Ser
1	{	Ser Thr Glu Leu Ala Pro Thr Ile Ser His Tyr Asp Cys Ser Pro
		Leu Asp Ser Thr Ala Thr Pro Ser Thr Pro Gly Thr Ala Pro Pro
2	{	Pro Gln Val Leu Thr Ser Pro Ala Thr Thr Pro Thr Ala Thr Ser
		Ser Lys Ala Thr Ser Ser Ser Ser Pro Arg Thr Ala Thr Thr Leu
1	{	Pro Val Leu Thr Ser Thr Ala Thr Lys Ser Thr Ala Thr Ser Val
		Thr Pro Ile Pro Ser Ser Thr Leu Gly Thr Thr Gly Thr Leu Pro
2	{	Glu Gln Thr Thr Thr Pro Val Ala Thr Met Ser Thr Ile His Pro
		Ser Ser Thr Pro Glu Thr Thr His Thr Ser Thr Val Leu Thr Thr
1	{	Lys Ala Thr Thr Thr Arg Ala Thr Ser Ser Thr Ser Thr Pro Ser
		Ser Thr Pro Gly Thr Thr Trp Ile Leu Thr Glu Leu Thr Thr Ala
2	{	Ala Thr Thr Thr Ala Gly Thr Gly Pro Thr Ala Thr Pro Ser Ser
		Thr Pro Gly Thr Thr Trp Ile Leu Thr Glu Leu Thr Thr Thr Aal
1	{	Thr Thr Thr Ala Ser Thr Gly Ser Thr Ala Thr Leu Ser Ser Thr
		Pro Gly Thr Thr Trp Ile Leu Thr Glu Pro Ser Thr Thr Ala Thr
2	{	Val Thr Val Pro Thr Gly Ser Thr Ala Thr Ala Ser Ser Thr Gln
		Ala Thr Ala Gly Thr Pro His Val Ser Thr Thr Ala Thr Thr Pro
1	{	Thr Val Thr Ser Ser Lys Pro Leu Leu Leu Ser Pro Gly Leu Gln
		Leu Pro Ser Ala Leu Arg Ser Thr Ala Thr Asn Thr His Ser Tyr
2	{	Gln Leu Tyr Ser His Pro Leu Leu Leu Pro Gly Thr Thr Trp Thr
		Arg Leu Ser Gln Thr Thr Thr Pro Thr Ala Thr Met Ser Thr Thr
1	{	Ala Thr Pro Ser Ser Thr Pro Glu Thr Val His Thr Ser Thr Val
		Leu Thr Ala Thr Ala Thr Thr Thr Gly Ala Pro Ala Leu Trp Pro
2	{	Pro Pro Pro Pro Pro Gln Glu Gln Leu Thr Leu Pro Lys Cys Arg
		Leu Pro Gln Pro Pro Leu His Ser His Pro Leu Leu Gln Pro Arg
1	{	Asp Gly Thr His Ala Ser Ser Val Asp Gln His Asn His His Thr
		His Asn His His Thr His Asn Gln Trp Leu His Gly Asp Pro Leu
2	{	Leu His Pro Gly Asp His Pro His Arg Gln Ser Ala Asp His His
		His His Asn Cys Ala Thr Gly Ser Met Ala Thr Pro Ser Ser Ser
1	{	Thr Gln Thr Ser Gly Thr Pro Pro Ser Leu Thr Thr Thr Ala Thr
		Thr Ile Thr Ala Thr Gly Ser Thr Thr Asn Pro Ser Ser Thr Pro
2	{	Gly Thr Thr Pro Ile Pro Pro Val Leu Thr Ser Thr Ala Thr Thr
		Pro Ala Ala Thr Ser Ser Lys Ala Thr Ser Ser Ser Ser Pro Arg
1	{	Thr Ala Thr Thr Leu Pro Val Leu Thr Ser Thr Ala Thr Lys Ser
		Thr Ala Thr Ser Phe Thr Pro Ile Pro Ser Ser Pro Cys Gly Pro
2	{	Arg Gly Pro Ser

1 : Domaine glycosylable

2 : Domaine nu

Figure n°22 : Séquence peptidique de JUL 7



1	{	Pro Thr Lys Ala Thr Thr Thr Arg Ala Thr Ser Ser Met Ser Thr
		Pro Ser Ser Thr Pro Gly Met Thr Trp Ile Leu Thr Glu Leu Thr
		Thr Ala Ala Thr Thr Thr Ala Ala Thr Ala Pro His Cys Gln Pro
2	{	Val Leu His Pro Arg Asp His Leu Asp Pro His Arg Ala Gln His
		Tyr Ser Thr Val Thr Val Pro Thr Gly Ser Gln Pro Thr Ala Ser
1	{	Ser Thr Arg Gly Thr Ala Gly Thr Leu Lys Val Leu Thr Ser Asp
		His His Thr His Ser His Gln Leu Gln Ser His Ser Leu Leu Gln
2	{	Ser Arg Asp Cys Asn Arg Pro Ser Ser Thr Glu Lys His Ser His
		His Thr His Ser Tyr Gln Arg Tyr Ser His Pro Leu Phe Leu Pro
		Gly Thr Ala Trp Thr Arg Leu Ser Gln Thr Thr Thr Pro Thr Ala
1	{	Thr Met Ser Thr Ala Thr Pro Ser Ser Thr Pro Glu Thr Val His
		Thr Ser Thr Val Leu Thr Thr Thr Ala Thr Thr Thr Arg Ala Thr
		Ala Leu Trp Pro Pro Pro Pro Pro Pro Gln Glu Gln Leu Thr Leu
2	{	Pro Lys Cys Arg Leu Pro Gln Pro Arg Leu His Ser Tyr Pro Leu
		Leu Gln Pro Arg Asp Gly Thr His Ala Ser Ser Val Asp Gln His
		Asn His His Thr His Asn Gln Arg Leu His Gly Asp Pro Leu Leu
		His Pro Gly Thr Thr His Thr Ala Thr Val Leu Thr Thr Thr Thr
1	{	Thr Thr Val Pro Leu Val Leu Trp Gln His Pro Pro Leu Ala His
		Arg Pro Val Val Leu Pro Pro Ser Leu Thr Thr Thr Ala Thr Thr
		Ile Thr Ala Thr Gly Ser Thr Thr Asn Pro Ser Ser Thr Pro Gly
		Thr Thr Pro Ile Pro Pro Val Leu Thr Thr Thr Ala Asn His Thr
2	{	Cys Ser His Gln Gln His Ser Asp Ser Leu Leu Cys Pro Arg Glu

1 : Domaine glycosylable

2 : Domaine nu

Figure n°23 : Séquence peptidique de JUL 10

Figure N°24: Séquence nucléotidique et peptidique de JER 62

```
      10      20      30      40      50      60
      |      |      |      |      |      |
CAGCAGGGACCCCTTCAAGATGTGCCTCAACTACGAGGTGCGTGTGCTCTGCTGCGAGACC
GlnGlnGlyProPheLysMETCysLeuAsnTyrGluValArgValLeuCysCysGluThr

      70      80      90      100     110     120
      |      |      |      |      |      |
CCCAAAGGCTGCCCGTGACCTCCACACCTGTGACAGCTCCTAGCACCCCTAGTGGGAGA
ProLysGlyCysProValThrSerThrProValThrAlaProSerThrProSerGlyArg

      130     140     150     160     170     180
      |      |      |      |      |      |
GCCACCAGCCCAACTCAGAGCACTTCCTCCTTGGCAGAATCAGGACAACCACTTTGTGACA
AlaThrSerProThrGlnSerThrSerSerTrpGlnAsnGlnAspAsnHisPheValThr
```

Acides aminés	JER 28		JER 57		JER 62	
	nbre	%	nbre	%	nbre	%
Thréonine	48	25,7	127	20,8	9	15,2
Sérine	24	12,8	81	13,3	7	11,9
Proline	33	17,6	76	12,5	7	11,9
Alanine	17	9,1	43	7,0	2	3,4
Glycine	12	6,4	27	4,4	3	5,1
Arginine	6	3,2	17	2,8	2	3,4
Asparagine	1	0,5	10	1,6	3	5,1
Ac. Aspartique	2	1,1	15	2,5	1	1,7
Cystéine	2	1,1	11	1,8	4	6,7
Glutamine	5	2,7	29	4,7	5	8,5
Ac.Glutamique	0	/	9	1,5	2	3,4
Histidine	8	4,3	55	9,0	1	1,7
Isoleucine	3	1,6	14	2,3	0	/
Leucine	12	6,4	49	8,0	2	3,4
Lysine	1	0,5	5	0,8	2	3,4
Méthionine	0	/	7	1,1	1	1,7
Phénylalanine	3	1,6	4	0,7	2	3,4
Tryptophane	3	1,6	15	2,5	1	1,7
Tyrosine	2	1,1	3	0,5	1	1,7
Valine	5	2,7	13	2,2	4	6,7
Total	187	100	610	100	59	100

Tableau n°9

Acides aminés	JUL 7		JUL 10	
	nbre	%	nbre	%
Thréonine	147	27,0	75	22,7
Sérine	76	14,0	34	10,3
Proline	76	14,0	40	12,1
Alanine	48	8,8	22	6,7
Glycine	26	4,8	10	3,0
Arginine	13	2,4	15	4,5
Asparagine	7	1,3	5	1,6
Ac. Aspartique	7	1,3	9	2,7
Cystéine	8	1,5	5	1,6
Glutamine	14	2,6	16	4,8
Ac. Glutamique	8	1,5	5	1,6
Histidine	26	4,8	30	9,1
Isoleucine	9	1,6	3	0,9
Leucine	42	7,7	32	9,7
Lysine	7	1,3	4	1,2
Méthionine	3	0,5	3	0,9
Phénylalanine	2	0,4	1	0,3
Tryptophane	6	1,1	4	1,2
Tyrosine	4	0,7	4	1,2
Valine	15	2,7	13	3,9
Total	544	100	330	100

Tableau n°10

Mise à part le clone JER62 (de taille trop réduite) les quatre autres clones présentent tous la même particularité d'une alternance de domaines glycosylables hydrophiles et de domaines hydrophobes moins glycosylables.

Si nous comparons tout d'abord les séquences peptidiques déduites, on peut noter quatre points communs entre ces différents clones :

1) Les acides aminés hydroxylés représentent entre 34 et 41 % des acides aminés totaux avec un rapport Ser/Thr de 1/2.

2) Les résidus proline atteignent 11 à 18 % et l'histidine 10 %.

3) Les résidus Cystéine dont on discute souvent la fonctionnalité sont présents de façon constante de 1 à 1,5 %.

4) Le tryptophane qui est l'un des acides aminés le plus hydrophobe représente toujours plus de 1 % de la composition et atteint 2,3 % dans JER 57.

A côté de ces points communs concernant la composition en acides aminés, les différents peptides montrent la même organisation interne dans laquelle la notion d'alternance de domaines est toujours présente.

Si on considère tout d'abord la notion d'hydrophilie et d'hydrophobie de chaque séquence peptidique, on peut couper chacune de celles-ci en zones hydrophiles et en zones hydrophobes.

Les longueurs de ces zones varient en fonction de la séquence étudiée et à l'intérieur même d'une seule séquence (Figure n°25).

Si on considère maintenant le pourcentage de sérine et de thréonine de chacun des domaines des différents clones, il apparaît clairement que :

- Les zones hydrophiles sont potentiellement des zones hautement glycosylées avec un pourcentage moyen en résidus hydroxylés de 47,3 % (valeur maximale 54,2 % ; valeur minimale 25,4 %).

- Les zones hydrophobes sont moins glycosylables avec un pourcentage moyen en Sérine, Thréonine de 15,7 % (valeur maximale 33,3 %, valeur minimale 0%).

Ces deux notions d'alternance sont de plus parfaitement superposables (figure n°25). La notion d'hydrophilie se trouvera d'ailleurs amplifiée lors de la glycosylation réelle.

Les zones hydrophiles varient en taille de 28 résidus (JUL 10) à 325 acides aminés (JUL 7). Il faut cependant considérer que les domaines N et C terminaux peuvent être plus longs car les clones ne représentent pas un ARN messenger complet.

Les zones hydrophobes ou "nues" varient en taille de 11 résidus (JER 57) à 194 résidus (JER 57).

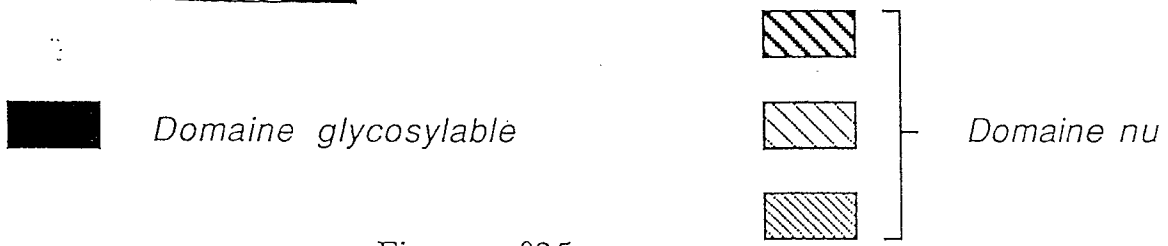
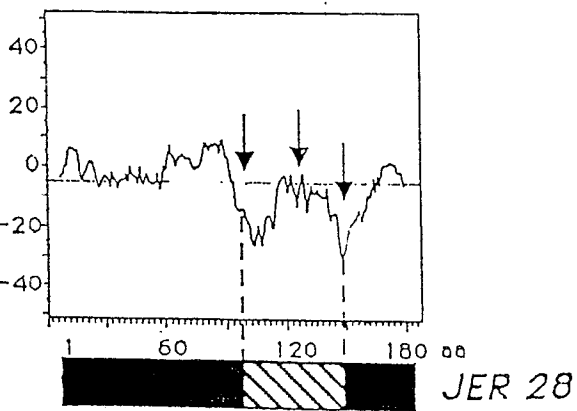
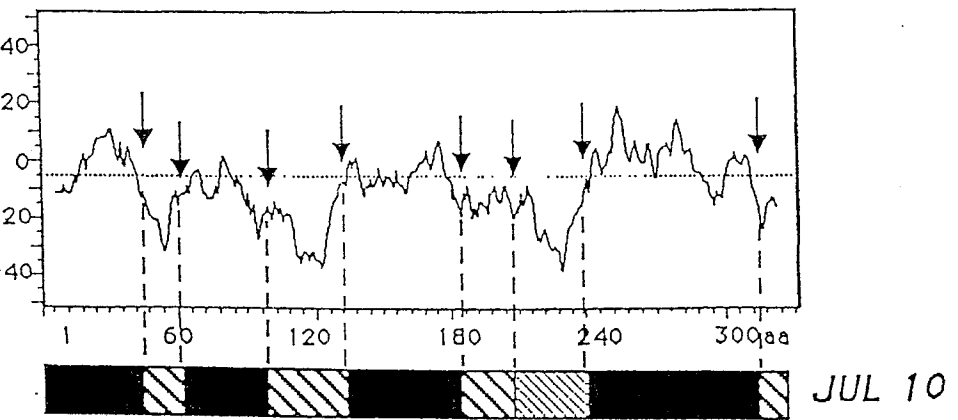
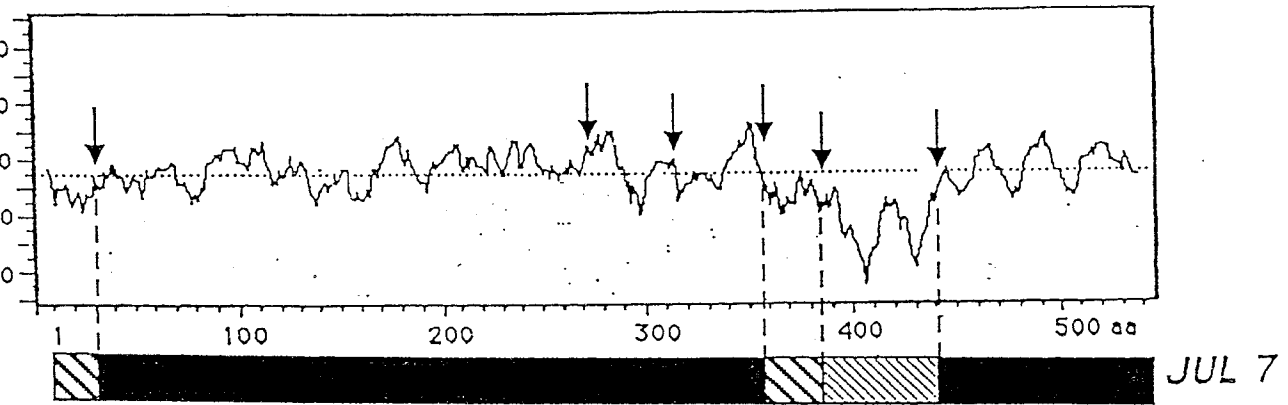
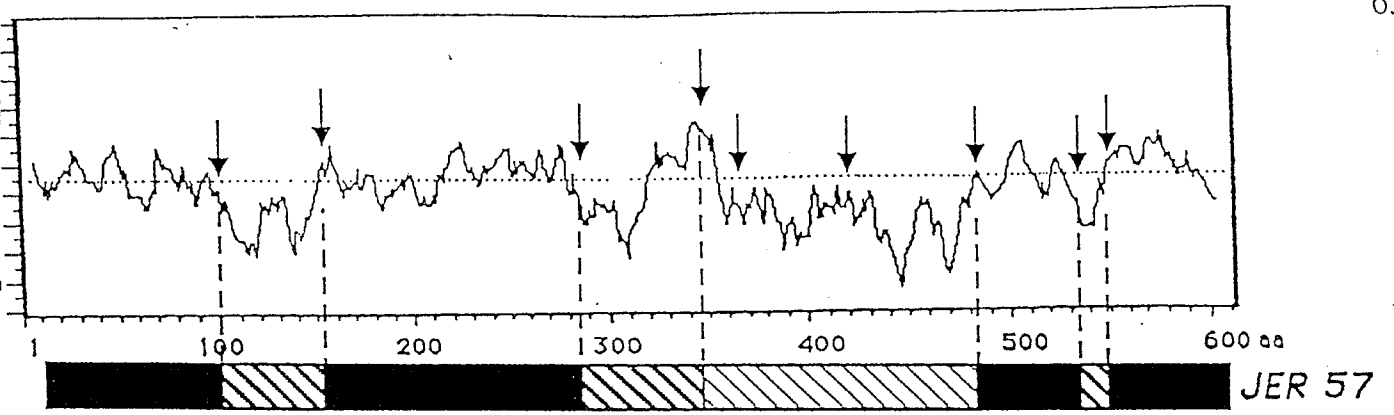


Figure n°25

Un dernier point corroborant la notion d'alternance de domaines est la répartition particulière d'un certain nombre de résidus comme la proline ou l'histidine.

Dans les quatre clones étudiés, des structures polyproline de 3 à 6 résidus sont uniquement retrouvées dans les zones hydrophobes. La même observation peut être faite pour les résidus histidine que l'on retrouve sous forme de polyhistidine de 2 à 4 résidus.

Par rapport à tout ce que nous avons vu précédemment nous ne sommes plus en présence de séquences répétitives pures comme pour MUC1, MUC2, MUC3, MUC4, MUC5C.

Ce résultat apporte donc la preuve définitive d'existence d'une organisation alternée : touffe potentielle de glycosylation - zone "nue".

Si nous regardons maintenant de près les séquences nucléotidiques qui donnent naissance à ce type d'organisation peptidique, on s'aperçoit que c'est une dégénérescence d'un motif répétitif qui est à l'origine de cette alternance.

Si on compare tout d'abord les clones entre eux, on s'aperçoit que l'homologie la moins bonne est de 71 % (JUL10/JER57) alors que la meilleure est de 96 % (JER28/JER57).

Une analyse très détaillée de ces quatre structures nucléotidiques a permis de montrer que le motif de base de la répétition était de 87. Cette séquence est donnée sur la figure n°26.

Sur 55 séquences répétitives possibles (pour les 4 clones), seules 22 possèdent les 87 paires de bases. 27 sont plus courtes à cause de délétions et 6 sont plus longues à cause d'insertions. Le plus long motif est fait de 102 paires de bases et le plus court n'en contient que 25 (figure n°26 à 29).

Si maintenant nous détaillons un des clones, on s'aperçoit que JER 28 est constitué de sept tandem repeats.

Quatre délétions sont visibles (figure n°29) :



34 TGTCCCTGGGGGAGTTGGGCCAGGTCGTGGAATGCAGCCTGGACTTTGGCCTGGTCTGCAGGAAACCGTGAGCAAGGTGGGGAAAGTTCA  
 121 AGATGTGCTTCAACTATGAAATCCGTGTGTTCTGCTGCAACTACGCCACTGCCCCAGCACCCCGGCCACCAGCTCTACGCCATGCC  
 204 TCTCCACTCCGGGGACGACCTGGATCCTCACAGAGCTGACCAACAACAGACCACTACGACTGCATCCACTGGATCCACGGCCACCCC ← G  
 294 TCTCCACCCCGGGAACAGCTCCCCTCCAAAGTGCTGACCAGCCCGGCCACCACACCCACAGCCACCAGTTCCAAAGGCCACTTCC  
 382 TCTCCAGTCCAAGGACTGCAACCACCCCTTCCAGTGCTGACAACGACAGCCACAAAATCCACAGCTACCAGGTTCACACCCATCCCC  
 469 TCTCCACCCCTTGGGACCAACCGGGACCCCTCCCAAGAACAGACCAACAACCCCGTGGCCACCATGTCCACAATCCACCCC  
 547 TCTCCACTCCGGAAACCAACCACACCTCCACAGTGCTGACCACGAAGGCCACCACGACAAGGGCCACCAGTTCCACGTCCACCCC  
 634 TCTCCACTCCGGGGACGACCTGGATCCTCACAGAGCTGACCACAGCAGCCACTACA ACTGCAGGA ██████████ ← G  
 703 TCTCCACCCCAAGGGACCACTGGATCCTCACAGAGCTGACCAACAACAGCCACTACGACTGCCTCCACTGGATCCACGGCCACCCCTG  
 790 TCTCCACCCCAAGGGACCACTGGATCCTCACAGAGCCGAGCACTACAGCCACCCTGACGGTCC ██████████ ← G  
 874 CCTCCACCCAGGGCAACTGCTGGCACC ██████████ ACATGTGAGCACAACGGCCACGACCCACAGTCAACAGCTCCAAAGCCACGTGCC  
 954 TCTCCAGTCCAGGGACTGCAACTGCCCTTCCAGCACTGAGAAGCACAGCCACCACACCCCAAGCTACCAGCTTTACAGCCATGCC ← G  
 1041 TCTCCTCCTCCCTGGGGACCACTGGACCCGCCCTATCACAGACCACCAACCCACACAGCAACCCTCCTCACTCCAGAGACTGTCCAC  
 1144 CCTCCACAAG ██████████ T GCTTACGCCACGGCCACCCACAACCGGCAGCCGGCTCTGTGGCCACCCCC  
 1204 TCTCCACCCCAAGCAACAGCTCACACTACCAAAAGTGCCGACTACCACAACCCAC ██████████ GGCTTCAAGCCACCCCC  
 1277 TCTCCAGCCCAAGGGACGGGACTCACGCCCTCCAGTGTGGATCAGCACAAACCACACACCCACAACCCAGGCTCCACGGGTGACCCCC  
 1373 TCTCCATCCCGGGGACCAACCACACCGCCAGAGTGCTGACCACCAACCACCAACTGT ██████████ GCCACTGGTTCTATGGCAACACC  
 1443 TCTCTAGCACACAGACAGACAGTGGTACTCCCCCATCACTGACCACCAACGGCCACTACGATCACGGGCCACC ██████████ GCGTCCACCCACCC  
 1444 TCTCAACTCCAGGGACAACACCCCATCCCCCAAGTGCTGACCAGCA ██████████ GCCACACACCCCGCAGCCACCAGCTCCAAGCA ██████████ CT  
 1450 TCTCCTCCCAAGGACTGCAACCACCCCTTCCAGTGCTGACAAGCACAGCCACC ██████████ AAAATCCACAGCTACCAGCTTTACACCCATCCC  
 1750 TCCGTACCCCTGTGGACCACTGGGACCCGTCCTCCAGCACAGACCAACCA ██████████ CCCATGTCCACCATGTCCACAATCCACAC  
 1754 TCTCTACTCCAGAGACCAACCACACCTCCACG

AGT
GGCCTACCATGTC
ACACCCACAACCCGT

Figure n°26 : Séquence nucléotidique de JER 57 : mise en évidence des unités répétitives de 87 Pb.

1 GTGTTCTGCTGCAACTACGGGCCACTGCCCCAGCACC CCGGGCCACCAGCTCTACGGCCATGCC  
 13 CCTCCACTCCGGGGACGACCTGGA  
 14 CCTCCACAGAGCTGGCACCTACAATCAGCCACTACGACTGCAGTCCACTGGATTCCACGGCCACCCCGTCTACCCCGGGAACAGCC  
 175 CCCCCTCC  
 211 CCTCCAGTCCAAGGACTGCAACCAACCCCTTCCAGTGCTGACAA GCACAGCCACCAAA TCCACAGCTACCAGCGTTACAACCATCCCC  
 228 CCTCCACCCCTTGGGACCAACCGGACCCCTCCAGAAACA GACCAAC  
 404 CCTCCACTCCGGAGAACCAACCCACACCTCCACAGTGCTGACCAACGAGGCCAACACGACAAGGGCCACCAGTTCCACGTTCCACCCCC  
 433 CCTCCACTCCGGGGACGACCTGGATCCCTCACAGAGCTGACCAACAGCAGCCACTACAACCTGCAGGCCACTGGCCCCACGGCCACCCCG  
 560 CCTCCACCCCAAGGACCAACCTGGATCCCTCACAGAGCTGACCAACAACAGCCACTACGACTGCCTCCACTGGATCCACGGCCACCCCTG  
 647 CCTCCACCCCAAGGACCAACCTGGATCCCTCACAGAGCTGACCAACAACAGCCACTACAGCCACCCTGACGGTGCACCACCGGATCCACAAGCCACCCG  
 754 CCTCCACCCCAAGGCAACTGCTGGCCACCCACATGTGAAGCAACCAAGGCCACGACCAACCCACAGTCAACAGCTCCAAAGCCA  
 923 CCTCCTCAGTCCAGGACTGCAACTGCCCTTCAGCACTGAGAAAGCACAGCCACCAACAACCCACAGCTACCAGCTTTACAGCCATCCCC  
 920 CCTCCTCCTGG  
 997 CCTCCACTCCAGAGACTGTCCACACCTCCACAGTGCTTACC GGCCACGGCCACCAACAACCGGGGC  
 1095 CCTCCACCCCAAGGAAACAGCTCACACTACCAAAAGTGCCGACTACCAACAACCAACGGCTTCAAG  
 1154 CCTCCAGCCCAAGGACGGGCACTCACGGCCTCCAGTGTTGGATCA GCAACAACCAACAACCCACAACCAACCAAGCTCCACGGTGAACCC  
 1256 CCTCCATCCCGGGACCAACCCACACCCGCCAGAGTGCTGACCAACCAACCAACAAC  
 1329 CCTCTAGCACACAGAACCAAGTGGTACTCCCCCATCACTGACCAACACGGCCACTACGATCACGGCCACCGGCTCCACCAACCAACCC  
 1428 CCTCAAATCCAGGGACAACAACCCATCCCCCAAGTGCTGACCAAGCACGGCCACCAACAACCCGCAGCCACCAAGCTCCAAAGCCACTTCC  
 1513 CCTCCAGTCCACAGGACTGCAACCAACCCCTTCCAGTGCTGACAA GCAACAGCCACA AAA TCCACAGCTTACAACCATCC  
 1600 CCTCACCCTGTGGACCAAGTGGACCGTCCG

CACCCACAACCAAGT

Figure n°27 : Séquence nucléotidique de JUL 7 : mise en évidence des unités répétitives de 87 Pb.

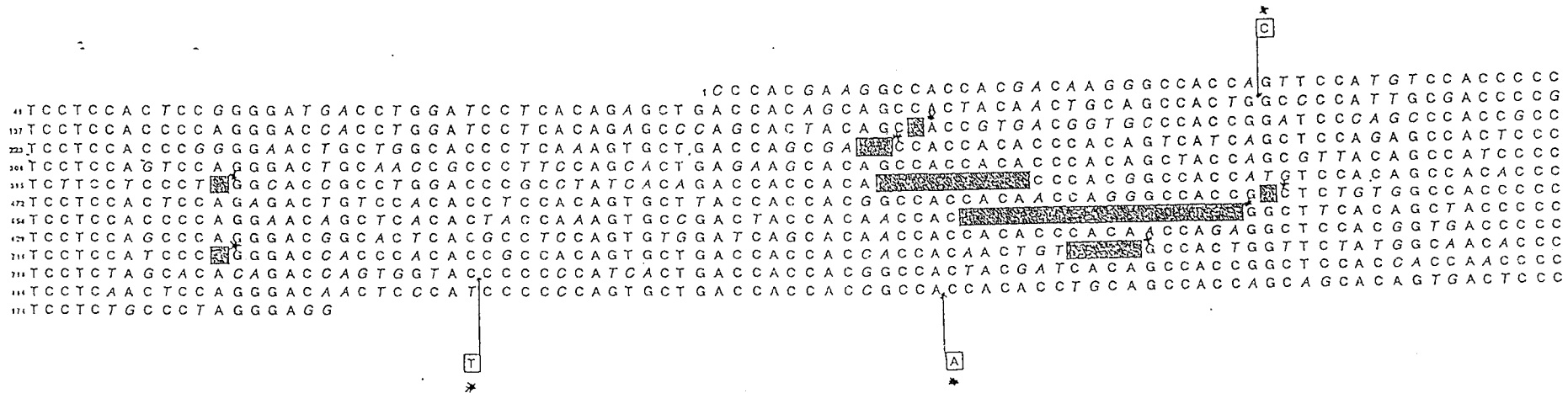


Figure n°28 : Séquence nucléotidique de JUL 10 : mise en évidence des unités répétitives de 87 Pb

1 T C C T C C A C C C C A G G G A C C A C C T G G A T C C T C A C A G A G C C G A G C A C T A C A G C C A C C G T G A C G G G T C C C A C C G G A T C C A C G G C C A C C G C C  
 14 T C C T C C A C C C A G G C A A C T G C T G G C A C C C C C A C A T G T G A G C A C C A C G G C C A C G A C A C C C A C A G T C A C C A G C T C C A A A G C C A C T C C C  
 172 T T C T C C A G T C C A G G G A C T G C A A C C G C C C T T C C A G C A C T G A G A A G C A C A G C C A C C A C A C C C A C A G C T A C C A G C T T T A C A G C C A T C C C T  
 211 C T C T C C T G G G C A C C A A C T G G A C C G C T A T C A C A G A C C A C C A C A C C A C G G C C A C C A T G T C C A C A G C C A C A C C C  
 250 T C C T C C A C T C C A G A G A C T G T C C A C A C C T C C A C C A G T G C T T A C C A C C A C G C C A C C A C C A A C C G G G C T A C C G G C T C T G T G G C C A C C C C C  
 411 T C C T C C A C C C C A G C A A C A G C T C A C A C T A C C A A A G T G C T G A C T A C C A C A A C C A C C A C C C A C A C C C A C A C C C A C A C C A G A G G T T C C  
 447 T C C T C C A G C C C A G G G A C G G C A C G C A C G C T T C C A G T G T G G A T C A G C A C A A C C A C C A C A C C C A C A C C C A G A G G T T C C

Figure n°29 : Séquence nucléotidique de JER 28 : mise en évidence des unités répétitives de 87 Pb

- La première (en position 116) introduit la délétion d'un résidu thréonine.

- Les trois autres délétions entraînent un changement du cadre de lecture. Dans ce cas les résidus hydroxylés sont remplacés par des résidus proline et histidine.

Pour les trois autres clones JER57, JUL7 et JUL10, les mêmes remarques peuvent être faites. La plupart des insertions ou délétions entraînent un changement du cadre de lecture.

Ces situations sont repérables par \* sur les figures.

Il est à noter que sur l'ensemble des clones, les délétions sont le mécanisme le plus fréquent. Un nombre variable de nucléotides sont concernés par ces délétions (de 1 à 62 paires de bases). La longueur des insertions varie de 1 à 15 paires de bases.

La plupart des modifications dans la répétition parfaite entraîne un changement du cadre de lecture. Ceci se produit trente fois sur quarante deux insertions ou délétions. Cependant, quatre fois ces changements sont annulés car trop proches l'un de l'autre : une insertion compensant une délétion ou inversement.

Cette grande fréquence de changement du cadre de lecture entraîne la complète disparition d'une répétition au niveau peptidique. Ce phénomène est amplifié par le fait qu'il existe également de nombreuses mutations ponctuelles même dans les tandem repeats complets.

Nous avons également relié ces changements de cadre de lecture à la création d'une alternance de zones hydrophiles et hydrophobes. Chaque insertion ou délétion entraînant un changement du cadre de lecture correspond parfaitement à un changement dans la courbe d'hydrophilie (tableau n°11).

L'observation des séquences nucléotidiques montre que le codon le plus fréquent est ACC qui code la Thréonine, la lecture dans les trois cadres d'un double codon ACC ACC donnerait :

dans le 1er cadre : deux résidus de Thréonine

dans le 2e cadre : un résidu de Proline

Clone	Nombre de nucléotides i : insertion d : délétion	Position du nucléotide	Changement du cadre et de lecture	Position de l'acide aminé	Alternance hydrophile (H) - hydrophobe (h)
JER57	1i	295	+	99	H->h
	1d	456	+	152	h->H
	9d	516	0	172	0
	19d	699	+ ) 0	233	0
	1i	701	+ )	234	0
	1d	852	+	284	H->h
	3d	872	0	291	0
	1i	1043	+	348	h->+h
	13i	1098	+	366	h->+h
	27d	1155	0	385	0
	16d	1258	+	420	0
	15i	1345	0	449	0
	4d	1437	+	479	h->H
	2d	1594	+	532	H->h
	4d	1627	+	543	h->H
3i	1638	0	546	0	
9d	1767	0	589	0	
JUL7	29d	87	+	29	h->H
	21d	182	0	61	0
	9d	371	0	124	0
	8d	832	+	278	0
	10d	932	+	311	0
	1d	1061	+	354	H->h
	16d	1145	+	382	h->+h
	15i	1223	0	408	0
4d	1311	+	437	h->H	
JUL10	1i	118	+	40	H->h
	1d	186	+	62	h->H
	2d	269	+	90	H->h
	1d	405	+	135	h->H
	9d	441	0	147	0
	1d	541	+	181	H->h
	16d	610	+	204	h->+h
	1d	726	+	242	h->H
	4d	773	+ )	258	0
	1i	823	+ ) 0	275	0
1i	937	+	313	H->h	
JER28	3d	115	0	39	0
	1d	285	+	95	H->h
	15d	314	0	105	0
	1d	394	+	132	0
	16d	468	+	156	h->H

Tableau n° 11 : Récapitulatif des insertions et des délétions des séquences nucléotidiques de la famille alternée.

dans le 3e cadre : un résidu d' Histidine.

A partir de ce modèle, la figure n°30 donne une bonne illustration des conséquences de l'introduction d'une délétion ponctuelle au début d'un domaine glycosylable du clone JUL 7.

- lue dans le 1er cadre de lecture, la séquence se caractérise par sa richesse en acides aminés hydroxylés, un taux de Proline relativement faible et un taux standard pour l'Histidine. Nous sommes dans un domaine glycosylable.

- lue dans le deuxième cadre de lecture ; la séquence montre un pourcentage en acides aminés hydroxylés de 15,8 %. Nous sommes dans un domaine peu glycosylable qui se caractérise par son pourcentage en acides aminés Proline : 36,5 %. Nous sommes dans une zone riche en Proline.

- lue dans le troisième cadre de lecture, la séquence révèle que le pourcentage en acides aminés hydroxylés est de 12,7 %. Nous sommes dans un domaine peu glycosylable caractérisé par sa richesse en acides aminés Histidine. Nous sommes dans une zone riche en Histidine.

La modification du cadre de lecture s'accompagne d'une variation du degré d'hydrophilie du peptide (figure n°31) : dans le 1er cadre, le peptide est plutôt hydrophile, lu dans le 2e cadre, le peptide est plutôt hydrophobe et enfin, lu dans le 3e cadre de lecture, le peptide est nettement hydrophobe.

N'ayant pas réussi à relier les trois types de séquences peptidiques, nous ne savons toujours pas si elle dérivent toutes de l'expression d'un même gène.

L'ARNm codé par ce gène hypothétique pourrait être de structure parfaitement répétitive en son centre et de structure répétitive dégénérée de chaque côté. Cette hypothèse est quand même relativement peu probable car sur le chromosome 11, les répétitions quasi parfaites concernent 24 bases et les autres concernent 87 bases. Ce genre de dégénérescence a déjà été observé dans le cas de MUC2 mais sans changement dans le nombre de bases concernées.

Nous pencherions donc plutôt pour l'existence d'au moins deux gènes sur le chromosome 11 en p15 donnant naissance à des peptides d'organisation totalement différente et peut être également de fonctionnalité différente.

880 900 920 940 960  
 AAGCACAGCCACCAACACCCACAGCTACCAGCTTTACAGCCATCCCCTCCTCCCTTGCACCACCTGGACCCGCCTATCACAGACC  
Ser Thr Ala Thr Asn Thr His Ser Tyr Gln Leu Tyr Ser His Pro Leu Leu Leu Pro Gly Thr Thr Trp Thr Arg Leu Ser Gln Thr  
 980 1000 1020 1040  
 ACCACACCCACGGCCACCATGTCCACAGCCACACCCCTCCTCCACTCCAGAGACTGTCCACACCTCCACAGTGCTTACCGCCACGGCCA  
Thr Thr Pro Thr Ala Thr Met Ser Thr Ala Thr Pro Ser Ser Thr Pro Glu Thr Val His Thr Ser Thr Val Leu Thr Ala Thr Ala  
 1060  
 CCACAACCGGGGCAC  
Thr Thr Thr Gly Ala

Thr: 34,9%  
 Ser: 12,7%  
 Ala: 9,5%  
 Leu: 9,5%  
 Pro: 6,3%  
 His: 4,7%

880 900 920 940 960  
 AAGCACAGCCACCAACACCCACAGCTACCAGCTTTACAGCCATCCCCTCCTCCCTTGCACCACCTGGACCCGCCTATCACAGACC  
Ala Gln Pro Pro Thr Pro Thr Ala Thr Ser Phe Thr Ala Ile Pro Ser Ser Ser Leu Ala Pro Pro Gly Pro Ala Tyr His Arg Pro  
 980 1000 1020 1040  
 ACCACACCCACGGCCACCATGTCCACAGCCACACCCCTCCTCCACTCCAGAGACTGTCCACACCTCCACAGTGCTTACCGCCACGGCCA  
Pro His Pro Arg Pro Pro Cys Pro Gln Pro His Pro Pro Pro Leu Gln Arg Leu Ser Thr Pro Pro Gln Cys Leu Pro Pro Arg Pro Pro  
 1060  
 CCACAACCGGGGCAC  
Gln Pro Gly His

Pro: 36,5%  
 Thr: 7,9%  
 Ser: 7,9%  
 Ala: 7,9%  
 Gln: 7,9%  
 His: 3,1%

880 900 920 940 960  
 AAGCACAGCCACCAACACCCACAGCTACCAGCTTTACAGCCATCCCCTCCTCCCTTGGCACCACCTGGACCCGCCTATCACAGACC  
Lys His Ser His Gln His Pro Gln Leu Pro Ala Leu Gln Pro Ser Pro Pro Pro Trp His His Leu Asp Pro Pro Ile Thr Asp His  
 980 1000 1020 1040  
 ACCACACCCACGGCCACCATGTCCACAGCCACACCCCTCCTCCACTCCAGAGACTGTCCACACCTCCACAGTGCTTACCGCCACGGCCA  
His Thr His Gly His His Val His Ser His Thr Leu Leu His Ser Arg Asp Cys Pro His Leu His Ser Ala Tyr Arg His Gly His  
 1060  
 CCACAACCGGGGCAC  
His Asn Arg Gly

His: 28,6%  
 Pro: 15,8%  
 Leu: 9,5%  
 Ser: 7,9%  
 Thr: 4,8%  
 Arg: 4,7%

Figure n°30 : Traduction dans les trois cadres de lecture de la séquence JUL 7 (de la base 880 à 1069)



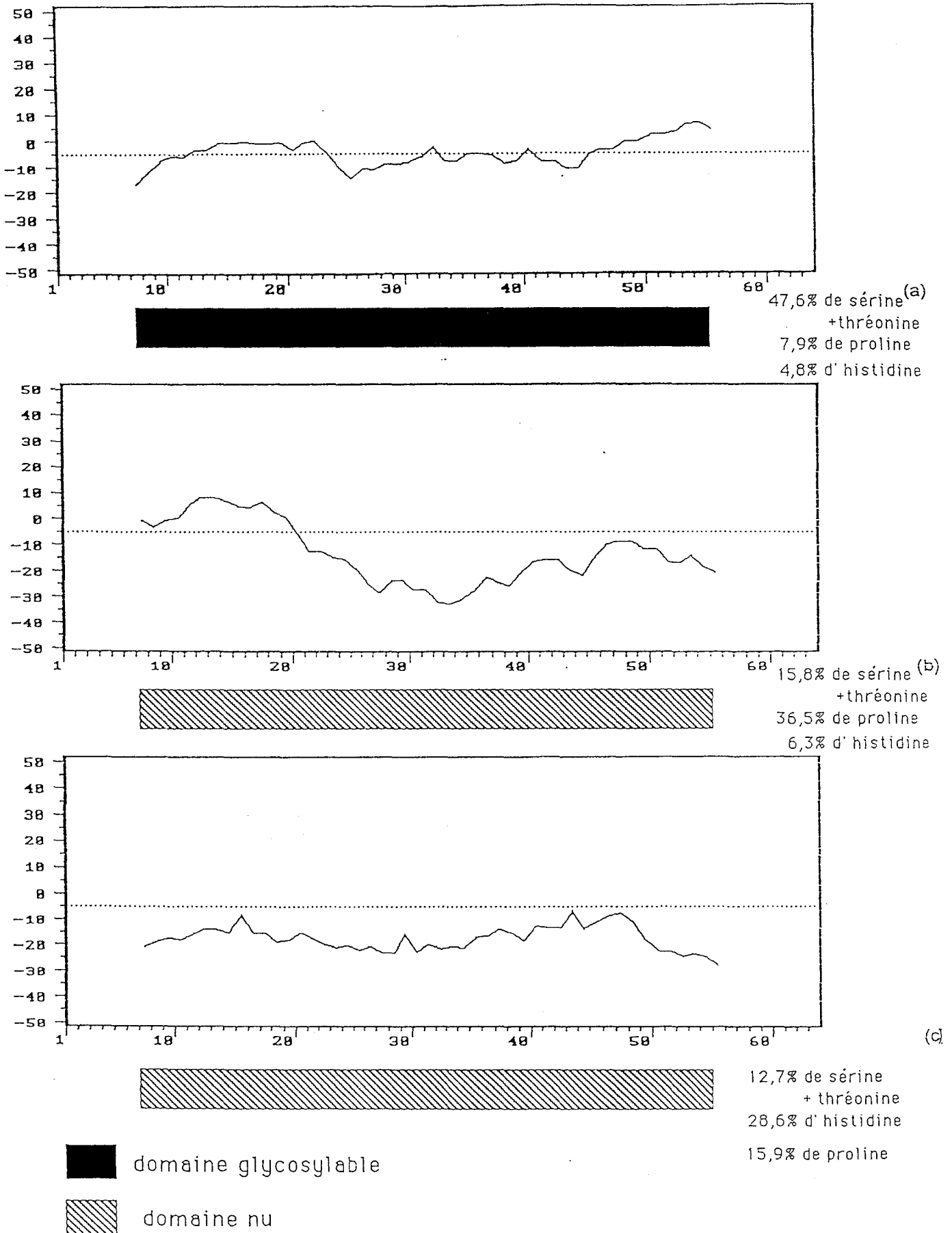


Figure n°31 : Courbes d'hydrophilie des trois séquences peptidiques de JUL 7 (lecture dans les trois cadres)

#### **4) Apport de la PCR**

L'outil PCR aurait dû être l'arme absolue pour étudier l'éventuelle continuité des différentes séquences sur un seul ARNm ou l'existence d'ARNm différents pour chaque structure.

Nous avons commencé ce travail de PCR dès l'obtention des premières séquences en choisissant des couples d'amorce de séquence rare. Pratiquement systématiquement nous avons obtenu une multitude d'amplifiats de petites tailles (<1 Kb). Le résultat n'est donc pas surprenant car au fur et à mesure de la découverte de nouvelles séquences, les amorces de séquences rares se sont révélées ne plus être rares du tout. L'ensemble des résultats obtenus par ce type d'expérience est donc totalement inexploitable.

#### **5) Etude de l'expression des mucines trachéobronchiques humaines par Northern blots**

L'étude a été réalisée sur :

- l'ARN de deux muqueuses trachéales humaines
- l'ARN d'une muqueuse gastrique humaine
- l'ARN d'une muqueuse colique humaine.

Les résultats sont reportés sur la figure n°32 et résumés dans le tableau n°12

Toutes les sondes utilisées révèlent un large continuum sur Northern blots de trachée humaine démontrant une grande diversité de taille des ARN messagers de 0.4 Kb à plus de 18 Kb à mettre en relation avec l'extraordinaire hétérogénéité des précurseurs peptidiques des mucines trachéales humaines (68).

Les sondes nucléiques correspondant aux clones JER 58 et JER 61 (famille I) ne reconnaissent pas les ARN messagers de muqueuses coliques humaines.

L'organisation de type répétitif à 8 acides aminés est donc spécifique de la muqueuse trachéale humaine et de la muqueuse gastrique.



JER 57

MUC 5B



JER 57



JER 47

MUC 5A



JER 58

MUC5C

1 : Trachée  
2 : Colon

Figure n°32

	Trachée humaine		Muqueuse gastrique	Muqueuse colique
	1	2		
JER 28	+++	+++	+	++
JER 47	+++	+++	+++	+
JER 57	+++	+++	+	++
JER 58	++	+++	+++	+/-
JER 61			+++	+/-
JER 62			+/-	+/-

Tableau n°12 : Etude en Northern blot des sondes extraites de la banque en  $\lambda$ gt 11

- +++ : Continuum intense
- ++ : Continuum moyen
- +
- +/- : Absence de signal significatif

Les sondes nucléiques JER 47 (famille II), JER 28 et 57 (famille III) révèlent un continuum sur les Northern blots de muqueuses gastriques et coliques humaines. Cependant JER 47 donne un marquage plus intense sur la muqueuse gastrique alors que les sondes de la famille III reconnaissent beaucoup mieux la muqueuse colique.

## **6) Localisation chromosomique et étude en Southern blot**

Les sondes JER 47, JER 57 et JER 58 ont été localisées sur le chromosome 11 en p15 (48). Le ou les gènes portent maintenant le nom officiel de MUC 5A, MUC 5B et MUC 5C. De plus, la sonde JER 47 a permis de visualiser un gène en relation avec les mucines sur le chromosome 13 (MUC 5L).

Tous les autres clones cités dans cette étude ont été localisés sur le chromosome 11. L'hybridation in situ n'a pas été réalisée mais il est plus que probable qu'ils se situent tous en p15.

En ce qui concerne l'utilisation de ces sondes sur Southern blots, nous avons pu relevé deux types de profils :

- les sondes JER 47, 58, 61, JUL 31, 32 (famille I et II) montrent l'existence d'une bande à 21 Kb pour la coupure Bam HI, d'une bande à 8,2 Kb pour la coupure Pst I et un gros fragment (>35 Kb) pour la coupure EcoRI (figure n°33).

- les sondes de la famille III montrent un profil caractérisé par la présence de 3 fragments Bam HI : 1,8 ; 1,3 et 0,3 Kb (figure n°34).

Les enzymes utilisés ici sur différents ADN n'ont pas permis de montrer l'existence d'un polymorphisme de restriction comme cela a été le cas pour MUC 4. Une étude plus approfondie est en cours de réalisation en collaboration avec le Dr D. Swallow (MRC Londres).

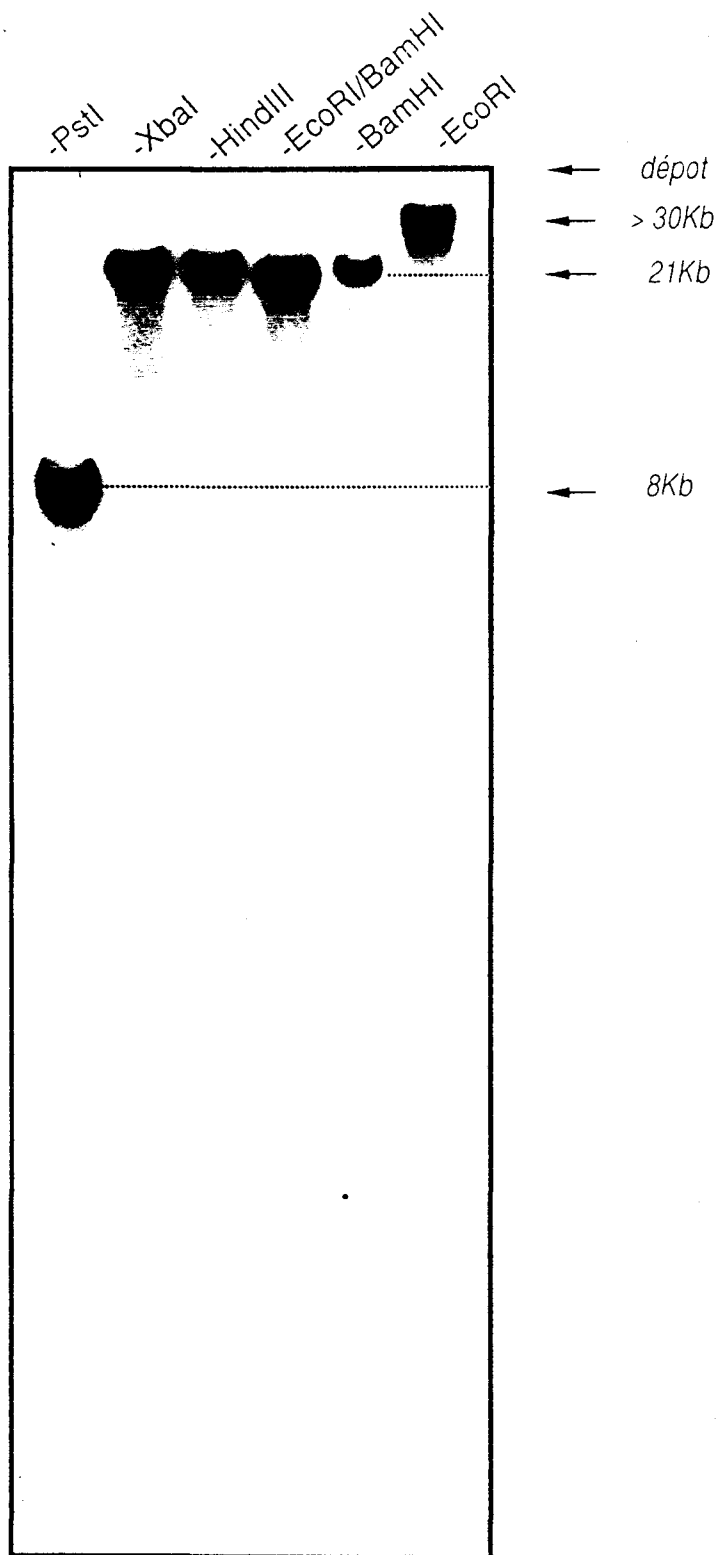


Figure n 33 : Sonde JER 47 sur Southern blot

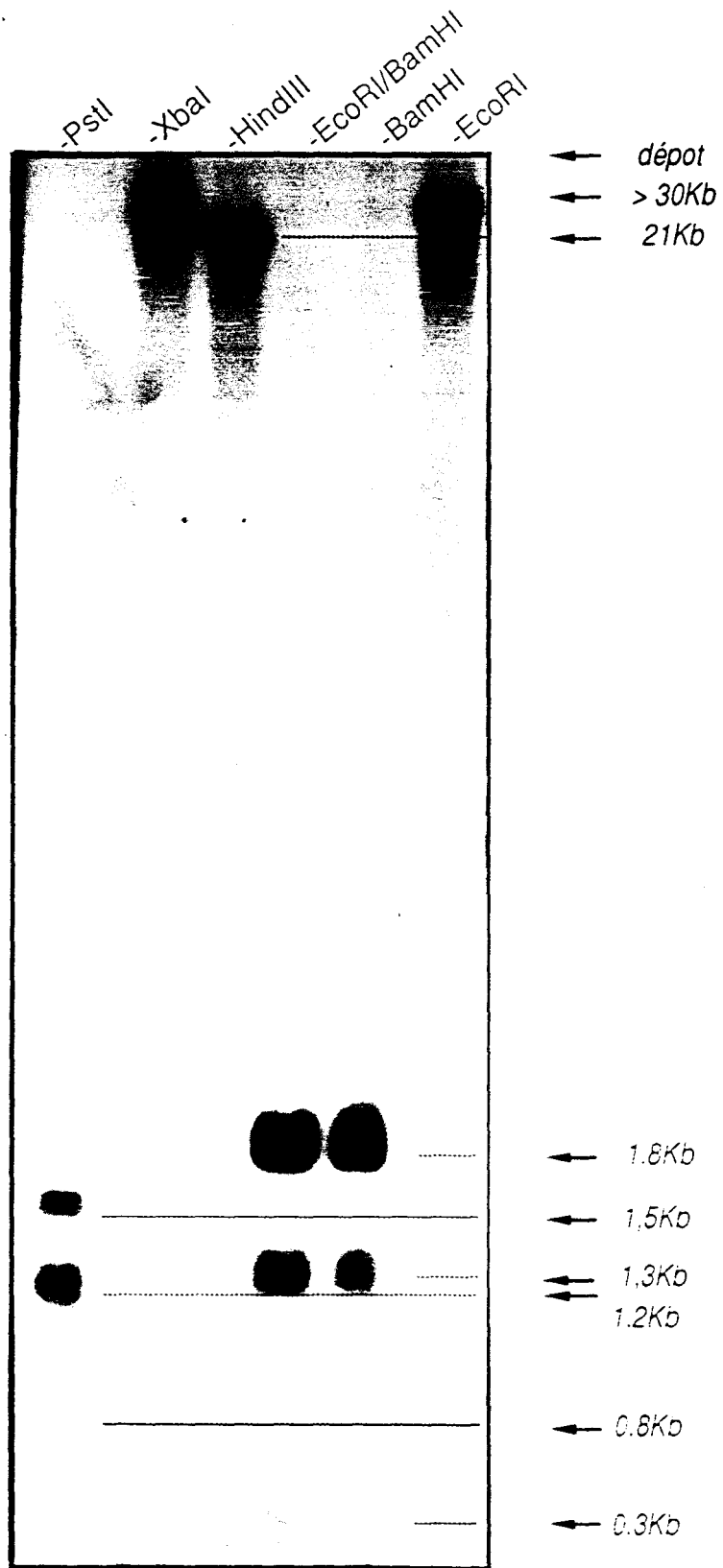
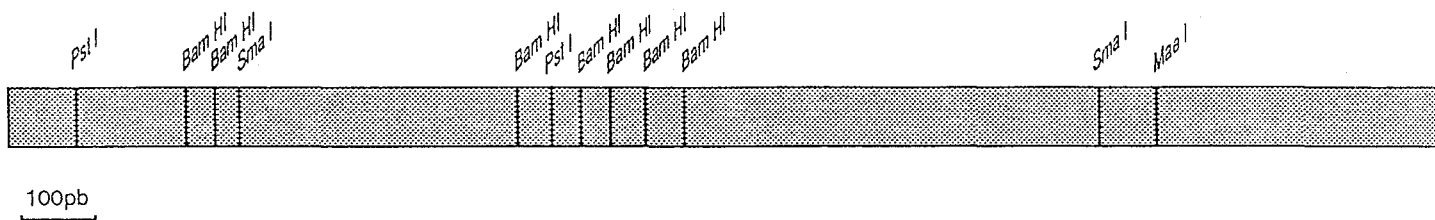


Figure n°34 : Sonde JER 57 sur Southern blot

En observant la carte de restriction du clone JER 57 ci-dessus, nous pouvons voir la présence de très nombreux sites de coupure par l'enzyme Bam HI et l'enzyme Pst I. Nous avons cependant tenté de localiser les différents fragments de JER 57 sur les bandes précédemment décrites :



- un fragment de 440 pb correspondant à l'extrémité 3' du clone JER 57. Ce fragment est libéré par une hydrolyse Sma I.

- un fragment de 89 pb correspondant à l'extrémité 5' de ce même clone libéré par une hydrolyse Pst I.

- un fragment de 90 pb central de la base 730 à la base 820 obtenu par amplification PCR sur le fragment JER 57.

Le fragment 440 pb révèle un profil parfaitement identique à celui obtenu avec la sonde entière (figure n°35).

Cette séquence est donc commune à l'ensemble des fragments 1,8 ; 1,3 et 0,3Kb révélés par l'hydrolyse Bam HI.

Ceci implique que celle-ci ou tout au moins une partie de celle-ci soit répétée sur le génome ou que dans ce fragment, il existe plusieurs exons répartis sur les 3 bandes.

Le fragment de 89 pb révèle les bandes de haute masse moléculaire obtenues par les hydrolyses Xba I, Hind III et Eco RI. Par contre dans le cas de la coupure Bam HI, seule la bande à 1,8 Kb est révélée et dans le cas de l'hydrolyse par Pst I, le fragment 1,3 Kb n'est pas révélé (figure n°36).



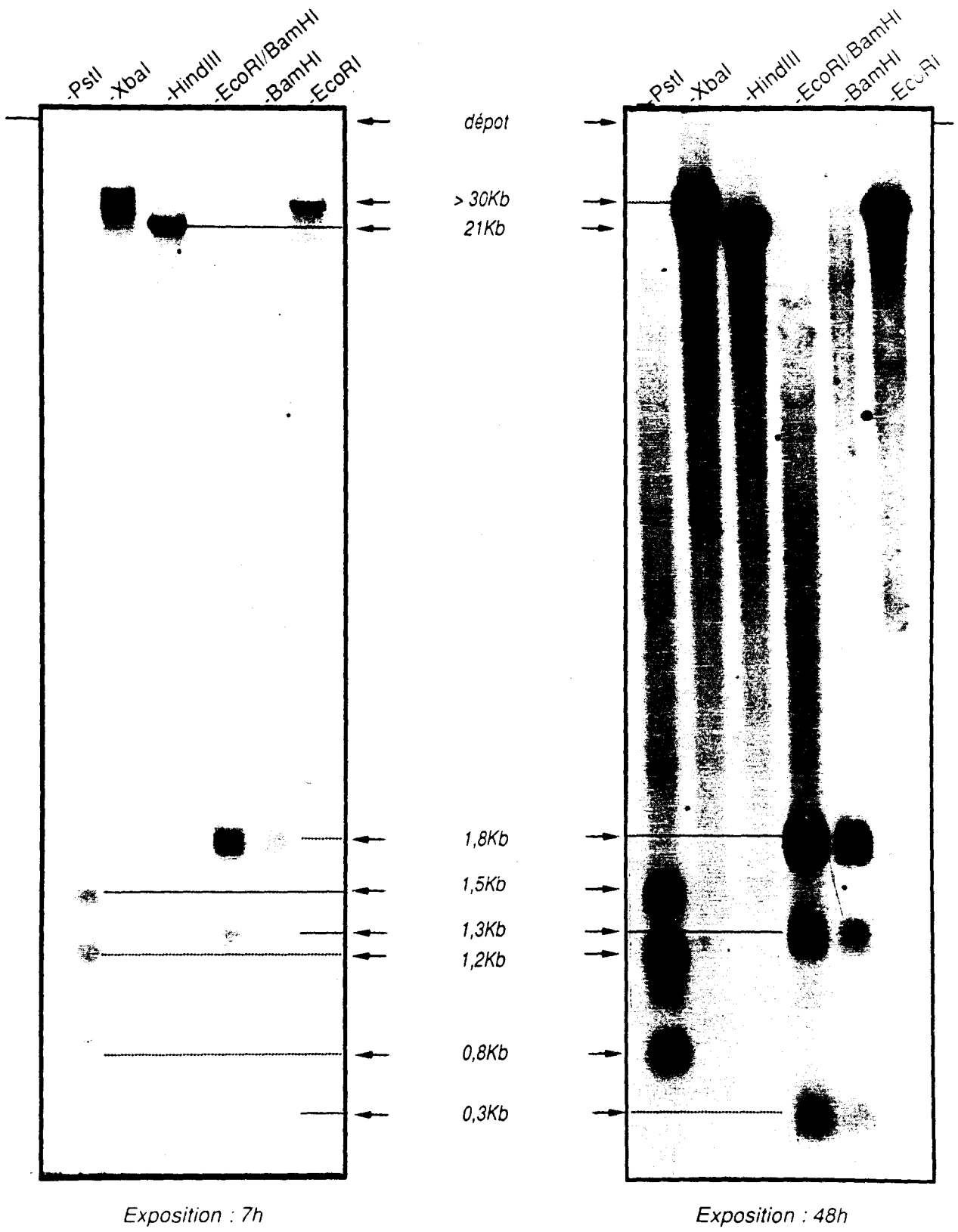


Figure n°35 : Sonde 440pb sur Southern blot

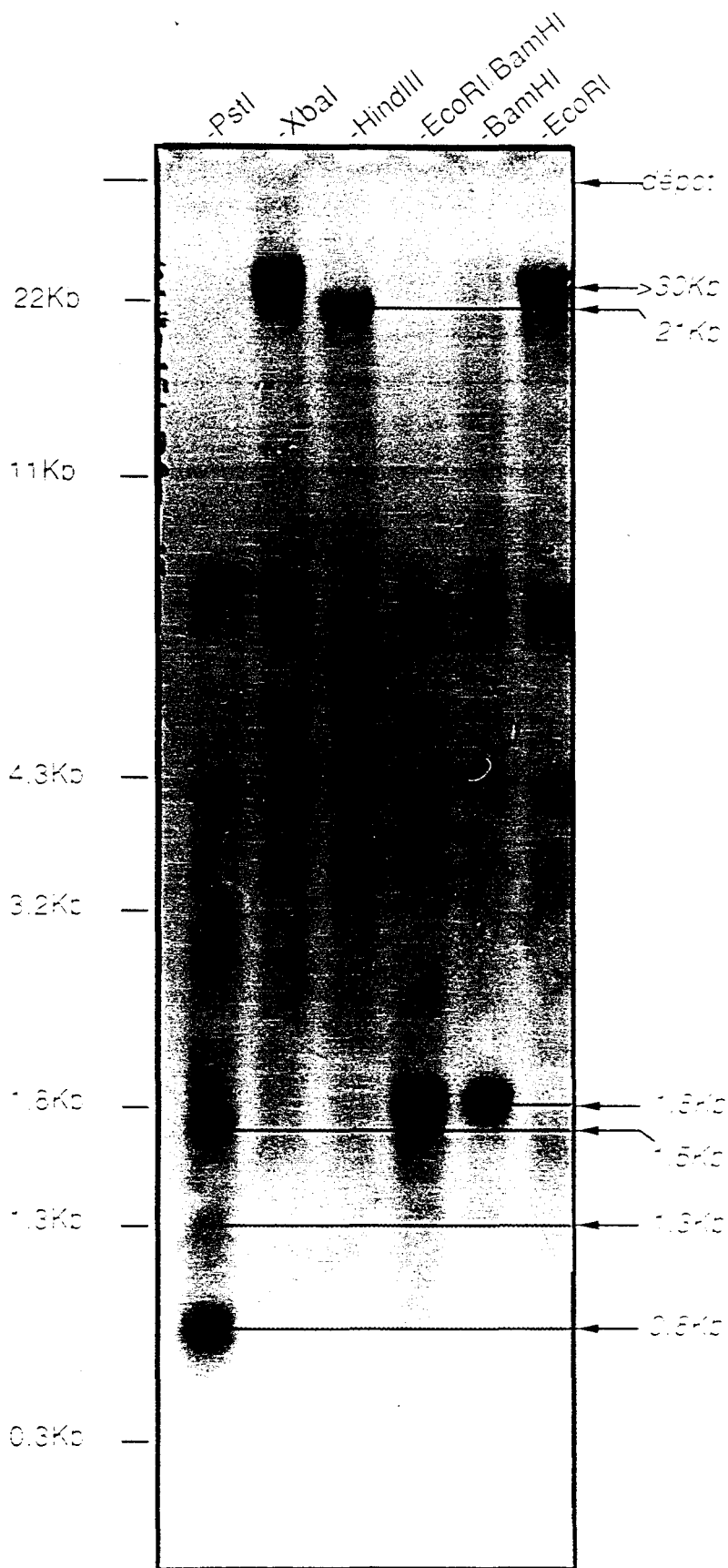


Figure n°36 : Sonde 89pb sur Southern blot

La partie 5' du clone JER 57 serait donc localisé uniquement sur le plus gros fragment Bam HI.

Le fragment de 90 pb ne révèle que les fragments de haute masse moléculaire engendrés par les hydrolyses Xba I, Hind III et Eco RI. Cette sonde ne révèle aucun fragment Bam HI (figure n°37). Ceci peut s'expliquer par la présence d'un site Bam HI au milieu de cette séquence. La multitude de ces sites de restriction implique une fragmentation des motifs exoniques en fragments de taille inférieure à 0.3 Kb très difficiles à détecter sur les Southern blots.

Cette étude préliminaire ne nous a pas permis de répondre aux questions :

- Les exons de la famille III sont-ils proches des exons des autres familles ?
- Existe-t-il un seul gène MUC 5 ou plusieurs gènes en 11 p 15 dont MUC 2 ?

Des études menées dans le groupe par électrophorèse en champs pulsés ont montré que les sondes représentatives des 3 familles se situent toutes sur un fragment Sma I de 220 Kb.

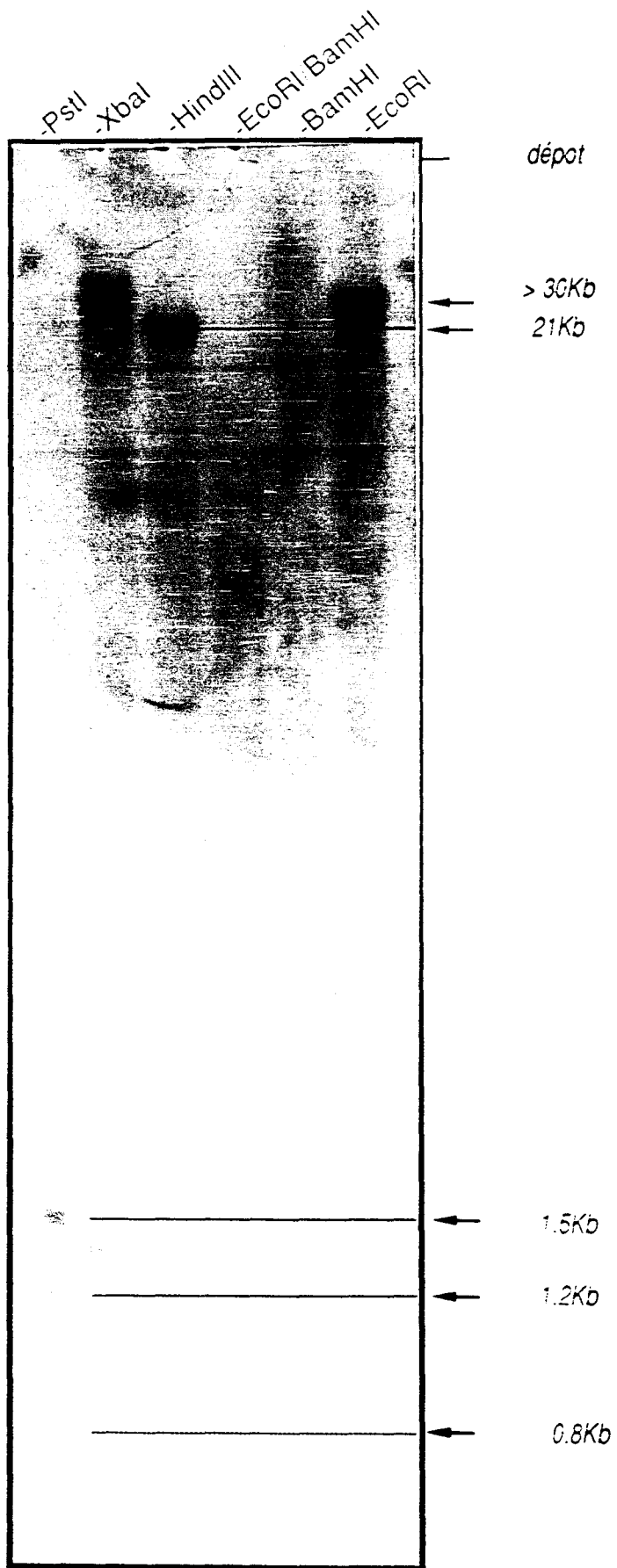


Figure n°37 : Sonde 90pb sur Southern blot

## CONCLUSION GENERALE

Ce travail nous a permis de démontrer l'existence de structures peptidiques alternant les zones potentiellement glycosylables et les zones nues dans les mucines trachéobronchiques humaines.

Cette image de mucines était suspectée depuis longtemps mais les premiers résultats concernant la séquence primaire de ces macromolécules ne décrivaient que des séquences répétitives hautement glycosylables.

Cette opposition entre ces deux conformations n'est réelle qu'au niveau peptidique. Les séquences de type alterné dérivent en réalité d'une structure répétitive dégénérée au niveau de l'ARN messenger et certainement du gène. Cette dégénérescence entraîne l'alternance des zones.

Tous les gènes de mucines actuellement décrits présentent donc, tout au moins dans leur partie centrale, des répétitions en tandem. Il semble donc logique de penser qu'à l'origine nous étions en présence de quelques tandem repeats qui petit à petit se sont dupliqués.

Cette duplication a peut être été nécessaire pour répondre aux agressions extérieures de plus en plus fréquentes permettant ainsi aux mucines d'acquérir de nouvelles propriétés.

Les deux gènes actuellement les mieux décrits sont MUC1 et MUC2. Dans ces deux cas l'homologie entre les répétitions est presque parfaite au centre de la structure pour dégénérer au fur et à mesure que l'on s'éloigne de ce centre. Il semble cependant que la longueur de la répétition reste constante. Dans le cas de la famille alternée décrite dans ce travail, la dégénérescence concerne aussi la longueur du tandem grâce à des insertions et des délétions. MUC5B est-il un gène à part dans le monde des mucines ? ou MUC5B fait-il partie d'un seul gène MUC5, situé loin du centre de la partie répétitive ?

Cette partie répétitive de MUC5 est constituée de 24bp alors que dans ce type alterné la répétition est de 87bp. Ce résultat est peut-être à mettre en relation

avec celui obtenu par MUC2 qui lui aussi présente deux domaines répétitifs distincts (36).

Au cours de ce travail nous n'avons malheureusement pas pu mettre en évidence une éventuelle continuité entre les différents MUC5. La technique PCR qui devait être la solution parfaite pour réaliser ce travail n'est pas utilisable dans notre domaine de recherche.

Beaucoup de questions restent posées.

Pourquoi tant de gènes ? Ont-ils une origine commune ? Pourquoi tant de transcrits ? Le phénomène "d'édition" prend-t-il part dans la régulation de l'expression des mucines ? Cette complexité n'est pas sans nous faire penser au système de l'immunité nécessitant une réelle hétérogénéité pour pouvoir répondre très rapidement à toutes sollicitations extérieures.

Les études de régulation de l'expression des gènes de mucines vont nécessiter encore beaucoup d'énergie et un grand nombre d'années. Cependant l'augmentation du nombre disponible de sondes et d'anticorps de plus en plus spécifiques devrait permettre d'atteindre plus rapidement la spécificité d'expression tissulaire et cellulaire ainsi que tout dérèglement. Les mucines pourraient devenir des marqueurs précieux de cancérisation et de différenciation.

**Degenerate 87 base pair tandem repeats create  
hydrophilic/hydrophobic alternating domains in human peptide  
mucins mapped to 11 p 15**

José DUFOSSE <sup>1</sup>, Nicole PORCHET <sup>1, 2</sup>, Jean-Pierre AUDIE <sup>1</sup>, Véronique  
GUYONNET DUPERAT <sup>1</sup>, Anne LAINE <sup>1</sup>, Isabelle VAN-SEUNINGEN <sup>1</sup>,  
Slaheddine MARRAKCHI <sup>2</sup>, Pierre DEGAND <sup>1, 2</sup>, and Jean-Pierre  
AUBERT <sup>1, 2\*</sup>

1) Unité 16 INSERM, Place de Verdun, 59045 Lille Cedex FRANCE

2) Laboratoire de Biochimie de l'Hôpital C. HURIEZ, CHRU Lille, FRANCE

\*To whom correspondence should be addressed

Phone Number 33 20 52 94 84

Fax Number 33 20 52 70 83

## SUMMARY

A human tracheobronchial  $\lambda$ gt 11 cDNA library was screened using antiserum prepared against the deglycosylated protein backbone of human tracheobronchial mucins. Two cDNAs designated JER 28 and 57 obtained from this immunoscreening were used to isolate two other cDNA clones JUL 7 and JUL 10 from a human tracheobronchial  $\lambda$ gt 10 cDNA library. These four clones (561 bp, 1830 bp, 1631 bp, 991 bp) which mapped to chromosome 11 p 15 were all found to contain degenerate 87 base pair tandem repeats which encode non repetitive peptides. Numerous deletions or insertions in an otherwise virtually perfect 87 base pair tandem repeat create many shifts in reading frame which completely destroy the repetitive peptide structure. The peptide is composed of alternate hydrophobic and hydrophilic domains which probably differ in the extent to which they are glycosylated. The mRNAs are expressed both in the respiratory and in the digestive tracts. These human mucin probes may be important in assessing the abnormal mucins associated with inflammatory diseases or carcinoma from human mucosae.



## INTRODUCTION

Mucins are secretory glycoproteins which constitute the viscous gel that covers mucosal epithelia of respiratory, gastrointestinal and reproductive tracts. The protective and lubricative properties of mucus are due to the viscous and viscoelastic properties of mucins, the major glycoprotein components of mucus. Mucins are characteristically heavily glycosylated (70 to 80 percent of their molecular weight) via the hydroxyl groups on serine and threonine which are major constituents of their protein backbone [For a recent review see STROUS and DEKKER, 1992,(1)]. The composition of these molecules and their high level of heterogeneity made the elucidation of their primary peptide structures extremely difficult. Until 1989, no more than twenty or thirty consecutive amino acids were identified by Edman degradation (2).

Several authors have however proposed structural models : Allen's "windmill" model in which four highly glycosylated subunits were linked together by a link peptide (3) ; a flexible thread model without link peptide presented by Carlstedt (4) and Slayter (5) in which subunits were attached end to end by disulphide bridges. These structural models have been constructed to explain the diverse physico-chemical properties of these molecules : abrupt loss of viscosity of mucus gel observed after contact with reducing agents which cleave disulfide bonds between cysteine residues (6), possible existence of a central link protein (7), sensitivity to different proteolytic enzymes (8). Although the arrangement and the size of the mucin subunits are controversial, most authors agree that the peptide cores contain heavily glycosylated domains (resistant to enzymatic attack and rich in hydroxy amino acid residues) and also, poorly glycosylated regions (so called "naked regions") mainly consisting of hydrophobic amino acids such as leucine, isoleucine and valine and which are sensitive to proteolytic enzymes.

The isolation of mucin cDNA clones has introduced a new view of the structure of mucins : a repetitive model involving numerous tandem repeats of a short peptide sequence comprising a large part of the coding region. At present, all the published peptide sequences deduced from mucin cDNA clones : porcine submaxillary gland mucin (9), breast or pancreatic tumor "mucin-like" (10-12), human intestinal mucins (13, 14), human tracheobronchial mucins (15, 16), contain this repetitive peptide structure.

Here we report for the first time, the complete peptide sequences deduced from four different human tracheobronchial cDNA clones which clearly demonstrate another type of peptide organization. Preliminary information concerning two other short clones has been published elsewhere (16).

In the clones, the peptide organization arises from numerous base deletions or insertions in the nucleotide sequence of a perfect 87 base pair tandem repeat organization. These nucleotide sequence modifications create many shifts in the reading frame which completely destroy the repetitive peptide structure. The peptide is composed of two distinct alternating domains : one is rich in serine and threonine and thus potentially heavily glycosylated and hydrophilic and the other is likely to be poorly glycosylated and represent a "naked" hydrophobic domain.

## EXPERIMENTAL PROCEDURES

**RNA purification and Northern blot analysis.** Total RNA from different human mucosae was purified using the guanidine isothiocyanate/cesium chloride method (17). Poly (A<sup>+</sup>) mRNAs were selected on 2 cycles of oligo (dT) cellulose affinity chromatography using standard procedures.

RNA was fractionated through 0,9% agarose gel containing glyoxal/dimethyl sulfoxide. Electrophoresis buffer (10mM sodium phosphate pH7.0) was recirculated. Transfer to reinforced cellulose nitrate membrane BA-S83 (Schleicher and Schuell) was achieved by capillary blotting. 28S and 18S rRNAs and a mixture of RNAs from Bethesda Research Laboratories were both used as molecular weight markers after membrane staining with methylene blue.

Probe hybridization was performed at 42°C for 16 h with 50% formamide, 3X SSC, 10µg/ml sheared salmon sperm DNA, 1X Denhardt's solution, HEPES buffer 50mM pH7.0, 12 percent dextran sulfate and a minimum of 10<sup>6</sup>cpm/ml of radiolabeled probe. The membranes were washed in 2X SSC, 0,1% SDS at 65°C and autoradiographed at -80°C for 1 to 24h.

Three short oligonucleotides (16bp or 17bp) were hybridized to Northern blots under the following conditions :

Prehybridization was carried out at ten degrees below the T<sub>m</sub> for at least 6 hours (T<sub>m</sub> of each oligonucleotide was for 56°C to 58°C) in 6X SSPE, 10X Denhardt's solution and 0.05 percent SDS buffer. Hybridization was at the same temperature for 15 hours in 6X SSPE, 10X Denhardt's solution, 0.5 percent SDS, 0.1mg/ml Salmon sperm DNA, 10 percent Dextran sulphate using 1pmole/ml of 5'end radiolabeled probe ( $\gamma^{32}\text{P}$  ATP). Post hybridization washings were first carried out in 4X SSPE, 0.1 percent SDS, 5°C below T<sub>m</sub> of each oligonucleotide. To guarantee the signal specificity, more stringent temperatures conditions were then used : a 5mn washing at one degree below the T<sub>m</sub> temperature.

**Polyclonal immune serum raised against human tracheobronchial apomucins.** Human bronchial sputum was collected from a single patient with blood group O suffering from chronic bronchitis. Pronase resistant glycopeptides from mucins were separated by anion exchange chromatography according to the acidic content of their carbohydrate chains. All details concerning preparation, purification and biochemical

analysis of mucin glycopeptides were described in (16), (18), (19). Five fractions corresponding to neutral, acidic, or very acidic glycopeptides were selected and then were deglycosylated by trifluoromethane sulfonic acid in the presence of phenylalanine (as scavenger) and Triton X 100. The precise protocols adapted for each fraction were described previously (16, 20). Two New Zealand white rabbits were immunized with an antigenic mixture containing 100 $\mu$ g of each deglycosylated fraction. (Protocol of immunization in 16).

***Construction and screening of the  $\lambda$ gt 10 and  $\lambda$ gt 11 cDNA libraries.***

cDNAs were prepared from normal human tracheobronchial mucosa poly (A<sup>+</sup>) RNAs using the cDNA Synthesis System Plus manufactured by Amersham Corp. Synthesis of the first cDNA strand was done using oligo(dt)<sub>12-18</sub> as primer for reverse transcription and the second cDNA strand was synthesized by using RNase H, DNA polymerase I and T4 DNA polymerase.

The cDNA was methylated with Eco RI methylase ; Eco RI linkers (pGGAATTCC) were ligated, cut by Eco RI endonuclease and separated by gel filtration. cDNA were ligated into unique Eco RI restriction site of  $\lambda$ gt 10 or  $\lambda$ gt 11. The recombinant molecules were packaged into phage by using *in vitro* packaging kit from Stratagene. The  $\lambda$ gt 10 and  $\lambda$ gt 11 cDNA libraries contained  $1.5 \times 10^6$  and  $2.5 \times 10^6$  recombinants respectively. The  $\lambda$ gt 11 library was screened with the polyclonal rabbit immunoserum raised against tracheobronchial apomucins. The secondary antibodies were peroxidase labeled goat antirabbit immunoglobulins and the detection was performed with 4-chloro -1 naphthol as substrate. JER 57 and JER 28 were selected because of a strongly positive signal of their fusion protein with immunoprobes. The  $\lambda$ gt 10 cDNA library was screened with the radiolabeled JER 57 probe; two clones were isolated which are called JUL 7 and JUL 10.

***Sequencing of cDNAs.*** The cDNA clones were sequenced by Sanger's method (21) using Sequenase T4 DNA polymerase under conditions recommended by the supplier (United States Biochemical Corp). The cDNA inserts were sequenced either by directional subcloning or by using exonuclease III - deleted clones. The length of deletions were estimated by gel electrophoresis. In each case, sequences were done on both strands.

Three oligonucleotide probes were chosen in particular domains of clones corresponding to degenerate domains

NAU 23 : 5'GGT GGC CGT GGG GCC A 3'

this sequence was chosen to be able to hybridize to 568-585 nucléotides from JUL 7.

NAU 24 : 5'T CGC AAT GGG GCG CAG T 3'

this sequence was complementary to nucleotide sequence of JUL 10 (positions 115-131).

NAU 28 : 5' AC GGG TTG TGG GTG TGG 3'

this sequence was complementary to nucleotide sequence of JER 57 (positions 1344-1360).

***Southern blot analysis and chromosome mapping.*** Human genomic DNA was extracted from lymphocytes by SDS-proteinase K digestion and phenol-chloroform extraction. DNA was digested by the indicated enzymes and was fractionated by 1% agarose gel electrophoresis. Southern blotting was performed on Nylon N<sup>+</sup> membranes (Amersham Corp). Probes were labeled with <sup>32</sup>P dCTP using a commercially available random primed labeling kit (Boehringer Mannheim).

DNA samples of 25 human-rodent hybrids were digested with Bam HI, electrophoresed and transferred as previously. The complete procedure including the source and description of hybrid cells, chromosome studies and *in situ* hybridization was described previously (22, 23).

***Hydrophilic/Hydrophobic index.*** Hydropathy curves were computed using the method described by Kyte and Doolittle (24).

## RESULTS

**Isolation of tracheobronchial mucin cDNAs ; immunological characterization.** The same polyclonal rabbit antiserum which was previously used for the isolation of the tandem repeat probes from MUC-4 (the gene located on chromosome 3q29) (15, 25) and MUC-5 (the gene located on chromosome 11p15) (16, 20) was used in the present study. This immunoserum was directed against a mixture of TFMS deglycosylated mucin peptides. (16). It reacts strongly with the  $\beta$ galactosidase fusion-proteins produced by JER 57 and JER 28. As was previously done for MUC-4 and MUC-5C cDNA probes (20), purified antibodies were prepared from the immunoserum by specific adsorption on these two fusion-proteins. Purified antibodies from JER 57 and those from JER 28 were able to label mucous cells from human submaxillary glands and both goblet cells and mucous cells from human bronchial (submitted) and tracheal (20) mucosae. Peroxidase labeling was situated around the goblet cell nucleus. The mucous cells show a cup-shaped peripheral labeling. No serous cell were labeled. The secretion granules containing glycosylated mucins were not labeled.

On the twenty positive clones identified in our original screening, two appeared to have distinctive structural features : they cross-hybridized with each other, but not with other clones, this constituting the third family of our sequence classification (16, 20) Furthermore the fusion proteins were also shown to bind an anti-respiratory mucin monoclonal antibody which has been isolated independently using a quite different apomucin preparation (26). The larger of these two clones, JER 57 was used as a radiolabeled probe to screen the  $\lambda$ gt 10 cDNA library and two other clones JUL 7 and JUL 10 were selected among other positive clones due to their large size. Therefore, we analyzed the details of their structures.

The complete sequence of each clone was determined and shown on figure 1. The insert sizes were : 561 bp for JER 28, 1830 bp for JER 57, 1631 bp for JUL 7, 991 bp for JUL 10.

The deduced amino acid sequences and the percentages of each amino acid are shown in figure 2.

```

34 TGTCCCTGGGGGAGTTGGGCCAGGTCTGTGGAATGCAGCCTGGACTTTGGCCTGGTCTGCAGGAACCGTGAGCAAGTGGGGAAAGTTCA
121 AGATGTGCTTCAACTATGAAATCCGTGTGTTCTGCTGCAACTACGCCACTGCCCCAGCACCCCGGCCACCAAGCTCTACGCCATGCC
208 T CCTCCACTCCGGGGACGACCTGGATCCTCACAGAGCTGACCACAACAGACCACTACGACTGCATCCACTGGATCCACGGCCACCCC
296 T CCTCCACCCCGGGAAACAGCTCCCCCTCCCAAAGTGCTGACCAGCCCGGCCACCACCCACAGCCACCAAGTTCCAAAGCCACTTCC
383 T CCTCCAGTCCAAGGACTGCAACCAACCCTTCCAGTGCTGACAACGACAGCCACCAAAATCCACAGCTACCAAGTTACAACCCATCC
469 T CCTCCACCCCTTGGGGACCACCGGGACCCTCCCAAGAACAGACCAACAACA [REDACTED] CCGTGGCCACCATGTCCACAATCCACCCC
647 T CCTCCACTCCGGAAACACCCACACCTCCACAGTGCTGACCACGAAGGCCACCACGACAAGGGCCACCAGTCCACGTCCACCC
634 T CCTCCACTCCGGGGACGACCTGGATCCTCACAGAGCTGACCACAGCAGCCACTACAACCTGCAGGA [REDACTED] CC
703 T CCTCCACCCCAAGGGACCACCTGGATCCTCACAGAGCTGACCACAACAGCCACTACGACTGCGTCCACTGGATCCACGGCCACCCCTG
790 T CCTCCACCCCAAGGGACCACCTGGATCCTCACAGAGCCGAGCACTACAGCCACCGTGACGGTCC [REDACTED] CCACCCGGATCCACAGCCACC [REDACTED]
874 T CCTCCACCCCAAGCAACTGCTGGCACCC [REDACTED] ACATGTGAGCACCCACGGCCACGACCCACAGTCCACAGCTCCAAAGCCACGTGCC
958 T CCTCCAGTCCAGGGACTGCAACTGCCCTTCCAGCACTGAGAAAGCACAGCCACCACACCACAGCTACCAAGCTTTACAGCCATGCC
1046 T CCTCCCTCCCTGGGGACCACCTGGACCCGCCCTATCACAGACCAACCAACCCAC [REDACTED] CACAGCAACCCCTCCTCACTCCAGAGACTGTCCAC
1146 A CCTCCACAG [REDACTED] TGGTTACGCCACGGCC [REDACTED] ACCACAACCGGCCGACCCGGCTCTGTGGCCACCC
1206 T CCTCCACCCCAAGCAACAGCTCACACTACCAAAAGTGC [REDACTED] GACTACCAACAACCAC [REDACTED] GGCTTCAAGCCACCC
1277 T CCTCCAGCCCAAGGGACGGGACTCACGCCCTCCAGTGTGGATCAGCAACAACCAC [REDACTED] CACACCACAACCACCGGCTCCACGGTGAACCC
1379 T CCTCCACTCCCGGGGACCAACCACACCCGCCAGAGTGCTGACCACCAACCACCAACTGTT [REDACTED] GCCACTGGTTCATGGCAACACCC
1462 T CCTCTAGCACACAGACCAAGTGGTACTCCCCCATCACTGACCACCAACGGCCACTACGATCACGGCCACC [REDACTED] GCGTCCACCAACCAACC
1549 T CCTCAACTCCAGGGACAAACACCACTCCCCCAGTGCTGACCAGCA [REDACTED] GCCACCAACCCCGCAGCCACCAGCTCCAAAGCA [REDACTED] CT
1630 T CCTCCTCCCAAGGACTGCAACCAACCCTTCCAGTGCTGACAAGCACAGCCACAAATCCACAGCTACCAGCTTTACACCCATCC
1720 T CCTCACCCCTGTGGACCACGTGGACCGTCCAGCACAGACCAACA [REDACTED] CCAATGTCCACCATGTCCACAATCCACACC
1798 T CCTCTACTCCAGAGACCACCCACACCTCCACG

```

AGT

GGCCTACCATGTC

ACACCCACAACCCGT

Figure n°1a

1GTGTTCTGCTGCAACTACGGCCA CTGCCAAGCACC CCGGCCACCAGCTCTACGGCCATGCC

83 T CCTCCACTTCGGGGACGACCTGGA  
 88 T CCTCCACAGAGCTGGCACCTACAATCAGGCCACTACGACTGCA GTCCA CTGGATTC CACGGCCACCCCGTCTACC CCGGGAACAGCC  
 176 C CCCCTCC  
 241 T CCTCCAGTCCAAGGACTGCAACCACCCCTTCCAGTGTGCTGACAA G CACA GCCACC AAATCCACAGCTACCAGCGTTACA C C C ATCC C C  
 328 T CCTCCACCCCTTGGGACCACCCGGGACCCCTCC CAGAACAGACCA C  
 406 T CCTCCACTCCGGAGACCACCCACACCTCCACAGTGTGCTGACCA C GAAGGCCACCACGACAAGGGCCACCAGTTCCACGTCACCC C C  
 493 T CCTCCACTTCGGGGACGACCTGGATCCTCACAAGAGCTGACCA C AGCAGGCCACTACAACTG CAGGCCACTGGCC C C C CAGGCCACCC C G  
 580 T CCTCCACCC CAGGGACCACCTGGATCCTCACAAGAGCTGACCA C AACAGGCCACTACGACTGCCTCCACTGGATCCAC C GGCACCCCTG  
 667 T CCTCCACCC CAGGGACCACCTGGATCCTCACAAGAGCCGAGCACTACA GCCACC GTGACGGTGCC CACC GGATCCACA GGCACCCGCC  
 754 T CCTCCACCC CAGGCAACTGCTGGCACC C CACATGTGA GCA C C C GGC CACGACA C C CACA GTCA C CAGCTCC AAGCCA C  
 823 T CCTCCTCAGTCCAGGACTGCAACTGCCTTCAGCACTGAGAAGCACA CAGCCACCAACA C C C CACAGCTACCAGCTTTACA GCCATC C C C  
 920 T CCTCCTCCTGG  
 997 T CCTCCACTTCAGAGACTGTCCA CACCTCCACAAGTGTCTAACCGCCACGGCCACCA C AACC GGGGC  
 1093 T CCTCCACCC CAGGAA CAGCTCACACTACCAAAGTGC C GACTACCA C AACCACGGCTTCA CAG  
 1154 T CCTCCAGCC CAGGGACGGCACTCACGGCTCCAGTGTGGATCAGCA C AACCACCA C C C C AACCACCAAGCTCCACGGGTGACCC C C  
 1256 T CCTCCATCC CGGGACCACCCACACCCGC CAGAGTGTGCTGACCA C CACCACCA C A C  
 1339 T CCTCTAGCACACAGACCAGTGGTACTCC C C C CACTCACTGACCA C CACGGCCACTACGATCACGGCCACC GGCTCCACCACCA C C C C  
 1428 T CCTCAACTCCAGGGACAACACC C C C C C CAGTGTGACCA CAGCAGGCCACCA C C C C C CAGCCACCAGCTCCA AAGCCACTTC C  
 1513 T CCTCCA GTCCAAGGACTGCAACCACCCCTTCCAGTGTGCTGACAA G CACA GCCACAA AATCCACAGCTACCAAGCTTTACA C C C ATCC C C  
 1600 T CCTCACCCCTGTGGACCA C GTGGGACC GTCCG

CACCCACAACCAGTG

Figure n°1b



```

49 TCCTCCACTCCGGGGATGACCTGGATCCTCACAGAGCTGACCAAGCAAGCCACTACAACCTGCAAGCCACTGGGCCCATTTGGGACCCCG
137 TCCTCCACCCAGGGGACCACTGGATCCTCACAGAGCCGAGCACTACAGCCACCGGTGCCCACCGGATCCCAGGCCACCGGCC
223 TCCTCCACCCGGGAACTGCTGGCAACCCTCAAGTGTGCTGACCAGCGAAGCCACCAACCCACAGTCATCAGCTCCAGAGCCACTCCC
308 TCCTCCAGTCCAGGGACTGCAACC GCCCTTCCAGCACTGAGAAAGCAAGCCACCAACAACCCACAGCTACCAAGCGTTACAGCCATCCC
395 TCTTCCCTCCCTGGGCACCGCCTGGACCCGCCTATCACAGACCACCAACAAGCCACCGGCCACCAATGTCACAAGCCACACCC
472 TCCTCCACTCCAGAGACTGTCCACAACCTCCACAGTGTCTACCAACCAAGGCCACCAACAAGGCCACCAACAAGGCCACCAACAAGGCC
558 TCCTCCACCCAGGAACAGCTCACACTACCAAGTGCCTGACTACCAACAACCACAGCCACCAACAAGGCCACCAACAAGGCCACCAACAAGGCC
629 TCCTCCAGCCAGGGACGGCACTCACGGCTCCAGTGTGGATCAGCAACAACCACCAACAACCACCAACAACCACCAACAACCACCAACAACC
714 TCCTCCATCCCAGGGACCAACCAACCGCCACAGTGTGCTGACCACCAACCACCAACAACCACCAACAACCACCAACAACCACCAACAACC
796 TCCTCTAGCACACAGACCAAGTGGTACCCGCCATCACTGACCACCAACCACCAACAACCACCAACAACCACCAACAACCACCAACAACC
886 TCCTCAAATCCAGGGACAACTCCCATCCGCCAGTGTGACCAACCAACCACCAACAACCACCAACAACCACCAACAACCACCAACAACC
974 TCCTCTGCCCTAGGGAGG

```

C

T

A

Figure n°1c

1 T C C T C C A C C C C A G G G A C C A C C T G G A T C C T C A C A G A G C C G A G C A C T A C A G C C A C C G T G A C G G G T C C C A C C G G A T C C A C G G C C A C C G C C  
88 T C C T C C A C C C A G G C A A C T G C T G G C A C C C C A C A T G T G A G C A C C A C G G C C A C G A C A C C C A C A G T C A C C A G C T C C A A A G C C A C T C C C  
172 T T C T C C A G T C C A G G G A C T G C A A C C G C C C T T C C A G C A C T G A G A A G C A C A G C C A C C A C A C C C A C A G C T A C C A G C T T T A C A G C C A T C C C T  
259 C T C T C C T G G G C A C C A A C T G G A C C G C T A T C A C A G A C C A C C A C A C C A C G G C C A C C A T G T C C A C A G C C A C A C C C  
330 T C C T C C A C T C C A G A G A C T G T C C A C A C C T C C A C C A G T G C T T A C C A C C A C G C C A C C A C C A A C C G G G C A C C G G C T C T G T G G C C A C C C C C  
416 T C C T C C A C C C C A G C A A C A G C T C A C A C T A C C A A A G T G C T G A C T A C C A C A A C C A C A C C C A C A C C C A C A A C C A G A G G T T C C  
487 T C C T C C A G C C C A G G G A C G G C A C G C A C G C T T C C A G T G T G G A T C A G C A C A A C C A C C A C A C C C A C A A C C A G A G G T T C C

Figure n°1d

a)

				5					10					15
1	Val	Phe	Cys	Cys	Asn	Tyr	Gly	His	Cys	Pro	Ala	Pro	Arg	Pro
16	Ala	Leu	Arg	Pro	Cys	Pro	Pro	Pro	Leu	Arg	Gly	Arg	Pro	Gly
31	Ser	Thr	Glu	Leu	Ala	Pro	Thr	Ile	Ser	His	Tyr	Asp	Cys	Ser
46	Leu	Asp	Ser	Thr	Ala	Thr	Pro	Ser	Thr	Pro	Gly	Thr	Ala	Pro
61	Pro	Gln	Val	Leu	Thr	Ser	Pro	Ala	Thr	Thr	Pro	Thr	Ala	Thr
76	Ser	Lys	Ala	Thr	Ser	Ser	Ser	Ser	Pro	Arg	Thr	Ala	Thr	Thr
91	Pro	Val	Leu	Thr	Ser	Ser	Thr	Ala	Thr	Lys	Ser	Thr	Ala	Thr
106	Thr	Pro	Ile	Pro	Ser	Ser	Thr	Leu	Gly	Thr	Thr	Gly	Thr	Leu
121	Glu	Gln	Thr	Thr	Thr	Pro	Val	Ala	Thr	Met	Ser	Thr	Ile	His
136	Ser	Ser	Thr	Pro	Glu	Thr	Thr	His	Thr	Ser	Thr	Val	Leu	Thr
151	Lys	Ala	Thr	Thr	Thr	Arg	Ala	Thr	Ser	Ser	Thr	Ser	Thr	Pro
166	Ser	Thr	Pro	Gly	Thr	Thr	Trp	Ile	Leu	Thr	Glu	Leu	Thr	Thr
181	Ala	Thr	Thr	Thr	Ala	Gly	Thr	Gly	Pro	Thr	Ala	Thr	Pro	Ser
196	Thr	Pro	Gly	Thr	Thr	Trp	Ile	Leu	Thr	Glu	Leu	Thr	Thr	Thr
211	Thr	Thr	Thr	Ala	Ser	Thr	Gly	Ser	Thr	Ala	Thr	Leu	Ser	Ser
226	Pro	Gly	Thr	Thr	Trp	Ile	Leu	Thr	Glu	Pro	Ser	Thr	Thr	Ala
241	Val	Thr	Val	Pro	Thr	Gly	Ser	Thr	Ala	Thr	Ala	Ser	Ser	Thr
256	Ala	Thr	Ala	Gly	Thr	Pro	His	Val	Ser	Thr	Thr	Ala	Thr	Thr
271	Thr	Val	Thr	Ser	Ser	Lys	Pro	Leu	Leu	Leu	Ser	Pro	Gly	Leu
286	Leu	Pro	Ser	Ala	Leu	Arg	Ser	Thr	Ala	Thr	Asn	Thr	His	Ser
301	Gln	Leu	Tyr	Ser	His	Pro	Leu	Leu	Leu	Pro	Gly	Thr	Thr	Trp
316	Arg	Leu	Ser	Gln	Thr	Thr	Pro	Thr	Ala	Thr	Met	Ser	Thr	Thr
331	Ala	Thr	Pro	Ser	Ser	Thr	Pro	Glu	Thr	Val	His	Thr	Ser	Thr
346	Leu	Thr	Ala	Thr	Ala	Thr	Thr	Thr	Gly	Ala	Pro	Ala	Leu	Trp
361	Pro	Pro	Pro	Pro	Pro	Gln	Glu	Gln	Leu	Thr	Leu	Pro	Lys	Cys
376	Leu	Pro	Gln	Pro	Pro	Leu	His	Ser	His	Pro	Leu	Leu	Gln	Pro
391	Asp	Gly	Thr	His	Ala	Ser	Ser	Val	Asp	Gln	His	Asn	His	His
406	His	Asn	His	His	Thr	His	Asn	Gln	Trp	Leu	His	Gly	Asp	Pro
421	Leu	His	Pro	Gly	Asp	His	Pro	His	Arg	Gln	Ser	Ala	Asp	His
436	His	His	Asn	Cys	Ala	Thr	Gly	Ser	Met	Ala	Thr	Pro	Ser	Ser
451	Thr	Gln	Thr	Ser	Gly	Thr	Pro	Pro	Ser	Leu	Thr	Thr	Thr	Ala
466	Thr	Ile	Thr	Ala	Thr	Gly	Ser	Thr	Thr	Asn	Pro	Ser	Ser	Thr
481	Gly	Thr	Thr	Pro	Ile	Pro	Pro	Val	Leu	Thr	Ser	Thr	Ala	Thr
496	Pro	Ala	Ala	Thr	Ser	Ser	Lys	Ala	Thr	Ser	Ser	Ser	Ser	Pro
511	Thr	Ala	Thr	Thr	Leu	Pro	Val	Leu	Thr	Ser	Thr	Ala	Thr	Lys
526	Thr	Ala	Thr	Ser	Phe	Thr	Pro	Ile	Pro	Ser	Ser	Pro	Cys	Gly
541	Arg	Gly	Pro	Ser										

Amino acid composition.  
=====

48 Ala	8 Cys	26 His	3 Met	147 Thr
13 Arg	14 Gln	9 Ile	2 Phe	6 Trp
7 Asn	8 Glu	42 Leu	76 Pro	4 Tyr
7 Asp	26 Gly	7 Lys	76 Ser	15 Val

b)

				5					10					15
1	Pro	Thr	Lys	Ala	Thr	Thr	Thr	Arg	Ala	Thr	Ser	Ser	Met	Ser
16	Pro	Ser	Ser	Thr	Pro	Gly	Met	Thr	Trp	Ile	Leu	Thr	Glu	Leu
31	Thr	Ala	Ala	Thr	Thr	Thr	Ala	Ala	Thr	Ala	Pro	His	Cys	Asp
46	Val	Leu	His	Pro	Arg	Asp	His	Leu	Asp	Pro	His	Arg	Ala	Gln
61	Tyr	Ser	Thr	Val	Thr	Val	Pro	Thr	Gly	Ser	Gln	Pro	Thr	Ala
76	Ser	Thr	Arg	Gly	Thr	Ala	Gly	Thr	Leu	Lys	Val	Leu	Thr	Ser
91	His	His	Thr	His	Ser	His	Gln	Leu	Gln	Ser	His	Ser	Leu	Leu
106	Ser	Arg	Asp	Cys	Asn	Arg	Pro	Ser	Ser	Thr	Glu	Lys	His	Ser
121	His	Thr	His	Ser	Tyr	Gln	Arg	Tyr	Ser	His	Pro	Leu	Phe	Leu
136	Gly	Thr	Ala	Trp	Thr	Arg	Leu	Ser	Gln	Thr	Thr	Thr	Pro	Thr
151	Thr	Met	Ser	Thr	Ala	Thr	Pro	Ser	Ser	Thr	Pro	Glu	Thr	Val
166	Thr	Ser	Thr	Val	Leu	Thr	Thr	Thr	Ala	Thr	Thr	Thr	Arg	Ala
181	Ala	Leu	Trp	Pro	Pro	Pro	Pro	Pro	Pro	Gln	Glu	Gln	Leu	Thr
196	Pro	Lys	Cys	Arg	Leu	Pro	Gln	Pro	Arg	Leu	His	Ser	Tyr	Pro
211	Leu	Gln	Pro	Arg	Asp	Gly	Thr	His	Ala	Ser	Ser	Val	Asp	Gln
226	Asn	His	His	Thr	His	Asn	Gln	Arg	Leu	His	Gly	Asp	Pro	Leu
241	His	Pro	Gly	Thr	Thr	His	Thr	Ala	Thr	Val	Leu	Thr	Thr	Thr
256	Thr	Thr	Val	Pro	Leu	Val	Leu	Trp	Gln	His	Pro	Pro	Leu	Ala
271	Arg	Pro	Val	Val	Leu	Pro	Pro	Ser	Leu	Thr	Thr	Thr	Ala	Thr
286	Ile	Thr	Ala	Thr	Gly	Ser	Thr	Thr	Asn	Pro	Ser	Ser	Thr	Pro
301	Thr	Thr	Pro	Ile	Pro	Pro	Val	Leu	Thr	Thr	Thr	Ala	Asn	His
316	Cys	Ser	His	Gln	Gln	His	Ser	Asp	Ser	Leu	Leu	Cys	Pro	Arg

Amino acid composition.  
=====

22 Ala	5 Cys	30 His	3 Met	75 Thr
15 Arg	16 Gln	3 Ile	1 Phe	4 Trp
5 Asn	5 Glu	32 Leu	40 Pro	4 Tyr
9 Asp	10 Gly	4 Lys	34 Ser	13 Val

Figure n°2

c)

					5					10					15
1	Leu	Gly	Leu	Glu	Cys	Arg	Ala	Ser	Pro	Ala	Trp	Cys	Pro	Trp	Gly
16	Ser	Trp	Ala	Arg	Ser	Thr	Asn	Ala	Ala	Trp	Thr	Leu	Ala	Trp	Ser
31	Ala	Gly	Thr	Val	Ser	Arg	Trp	Gly	Ser	Ser	Arg	Cys	Ala	Ser	Thr
46	Met	Lys	Ser	Val	Cys	Ser	Ala	Ala	Thr	Thr	Pro	Leu	Pro	Gln	His
61	Pro	Gly	His	Gln	Leu	Tyr	Ala	Met	Pro	Ser	Ser	Thr	Pro	Gly	Thr
76	Thr	Trp	Ile	Leu	Thr	Glu	Leu	Thr	Thr	Thr	Asp	His	Tyr	Asp	Cys
91	Ile	His	Trp	Ile	His	Gly	His	Pro	Val	Leu	His	Pro	Gly	Asn	Ser
106	Ser	Pro	Ser	Gln	Ser	Ala	Asp	Gln	Pro	Gly	His	His	Thr	His	Ser
121	His	Gln	Phe	Gln	Ser	His	Phe	Leu	Leu	Gln	Ser	Lys	Asp	Cys	Asn
136	His	Pro	Ser	Ser	Ala	Asp	Asn	Asp	Ser	His	Gln	Ile	His	Ser	Tyr
151	Gln	Val	Thr	Pro	Ile	Pro	Ser	Ser	Thr	Leu	Gly	Thr	Thr	Gly	Thr
166	Leu	Pro	Glu	Gln	Thr	Thr	His	Pro	Val	Ala	Thr	Met	Ser	Thr	Ile
181	His	Pro	Ser	Ser	Thr	Pro	Glu	Thr	Thr	His	Thr	Ser	Thr	Val	Leu
196	Thr	Thr	Lys	Ala	Thr	Thr	Thr	Arg	Ala	Thr	Ser	Ser	Thr	Ser	Thr
211	Pro	Ser	Ser	Thr	Pro	Gly	Thr	Thr	Trp	Ile	Leu	Thr	Glu	Leu	Thr
226	Thr	Ala	Ala	Thr	Thr	Thr	Ala	Gly	Pro	Ser	Ser	Thr	Pro	Gly	Thr
241	Thr	Trp	Ile	Leu	Thr	Glu	Leu	Thr	Thr	Thr	Ala	Thr	Thr	Thr	Ala
256	Ser	Thr	Gly	Ser	Thr	Ala	Thr	Leu	Ser	Ser	Thr	Pro	Gly	Thr	Thr
271	Trp	Ile	Leu	Thr	Glu	Pro	Ser	Thr	Thr	Ala	Thr	Val	Thr	Val	Pro
286	Pro	Asp	Pro	Gln	Pro	Pro	Pro	Pro	Pro	Arg	Gln	Leu	Leu	Ala	Pro
301	Thr	Cys	Glu	His	His	Gly	His	Asp	Thr	His	Ser	His	Gln	Leu	Gln
316	Ser	His	Val	Pro	Ser	Ser	Ser	Pro	Gly	Thr	Ala	Thr	Ala	Leu	Pro
331	Ala	Leu	Arg	Ser	Thr	Ala	Thr	Thr	Pro	Thr	Ala	Thr	Ser	Phe	Thr
346	Ala	Met	Ala	Leu	Leu	Leu	Pro	Gly	His	His	Leu	Asp	Pro	Pro	Ile
361	Thr	Asp	His	His	Thr	His	Gly	Leu	Pro	Cys	Pro	Gln	Gln	Pro	Ser
376	Ser	Leu	Gln	Arg	Leu	Ser	Thr	Pro	Pro	Gln	Cys	Leu	Arg	His	Gly
391	His	His	Asn	Arg	Arg	Pro	Ala	Leu	Trp	Pro	Pro	Pro	Pro	Pro	Pro
406	Gln	Gln	Gln	Leu	Thr	Leu	Pro	Lys	Cys	Arg	Leu	Pro	Gln	Pro	Arg
421	Leu	His	Ser	His	Pro	Leu	Leu	Gln	Pro	Arg	Asp	Gly	Thr	His	Ala
436	Ser	Ser	Val	Asp	Gln	His	Asn	His	His	Thr	His	Asn	His	His	Thr
451	His	Asn	Pro	Trp	Leu	His	Gly	Asp	Pro	Leu	Leu	His	Pro	Gly	Asp
466	His	Pro	His	Arg	Gln	Ser	Ala	Asp	His	His	His	His	Asn	Cys	Ala
481	Thr	Gly	Ser	Met	Ala	Thr	Pro	Ser	Ser	Ser	Thr	Gln	Thr	Ser	Gly
496	Thr	Pro	Pro	Ser	Leu	Thr	Thr	Thr	Ala	Thr	Thr	Ile	Thr	Ala	Thr
511	Ala	Ser	Thr	Thr	Asn	Pro	Ser	Ser	Thr	Pro	Gly	Thr	Thr	Pro	Ile
526	Pro	Pro	Val	Leu	Thr	Ser	Ser	His	His	Thr	Arg	Ser	His	Gln	Leu
541	Gln	Ser	Thr	Ser	Ser	Ser	Ser	Pro	Arg	Thr	Ala	Thr	Thr	Leu	Pro
556	Val	Leu	Thr	Ser	Thr	Ala	Thr	Lys	Ser	Thr	Ala	Thr	Ser	Phe	Thr
571	Pro	Ile	Pro	Ser	Leu	Thr	Leu	Trp	Thr	Thr	Trp	Thr	Val	Pro	Ala
586	Gln	Thr	Thr	Thr	Pro	Met	Ser	Thr	Met	Ser	Thr	Ile	His	Thr	Ser
601	Ser	Thr	Pro	Glu	Thr	Thr	His	Thr	Ser	Thr					

Amino acid composition.

=====

43 Ala	11 Cys	55 His	7 Met	127 Thr
17 Arg	29 Gln	14 Ile	4 Phe	15 Trp
10 Asn	9 Glu	49 Leu	76 Pro	3 Tyr
15 Asp	27 Gly	5 Lys	81 Ser	13 Val

d)

					5					10					15
1	Pro	Ser	Thr	Pro	Gly	Thr	Thr	Trp	Ile	Leu	Thr	Gly	Pro	Ser	Thr
16	Thr	Ala	Thr	Val	Thr	Gly	Pro	Thr	Gly	Ser	Thr	Ala	Thr	Ala	Ser
31	Ser	Thr	Gln	Ala	Thr	Ala	Gly	Thr	Pro	His	Val	Ser	Thr	Thr	Ala
46	Thr	Thr	Pro	Thr	Val	Thr	Ser	Ser	Lys	Ala	Thr	Pro	Phe	Ser	Ser
61	Pro	Gly	Thr	Ala	Thr	Ala	Leu	Pro	Ala	Leu	Arg	Ser	Thr	Ala	Thr
76	Thr	Pro	Thr	Ala	Thr	Ser	Phe	Thr	Ala	Ile	Pro	Leu	Ser	Trp	Ala
91	Pro	Thr	Gly	Pro	Leu	Ser	Gln	Thr	Thr	Thr	Pro	Arg	Pro	Pro	Cys
106	Pro	Gln	Pro	His	Pro	Pro	Pro	Leu	Gln	Arg	Leu	Ser	Thr	Pro	Pro
121	Pro	Val	Leu	Thr	Thr	Thr	Pro	Pro	Pro	Thr	Gly	His	Arg	Leu	Cys
136	Gly	His	Pro	Leu	Leu	His	Pro	Ser	Asp	Ser	Ser	His	Tyr	Gln	Ser
151	Ala	Asp	Tyr	His	Asn	His	Gly	Phe	Thr	Ala	Thr	Pro	Ser	Ser	Ser
166	Pro	Gly	Thr	Ala	Arg	Thr	Leu	Pro	Val	Trp	Ile	Ser	Thr	Thr	Thr
181	Thr	Pro	Thr	Thr	Arg	Gly	Ser								

Amino acid composition.

=====

17 Ala	2 Cys	8 His	0 Met	48 Thr
6 Arg	5 Gln	3 Ile	3 Phe	3 Trp
1 Asn	0 Glu	12 Leu	33 Pro	2 Tyr
2 Asp	12 Gly	1 Lys	24 Ser	5 Val

Figure n°2

*Comparison of the four deduced peptide sequences*. The following common features are noticed :

- (1)-hydroxy amino acids represent 34 to 41 percent of the total amino acids, with a Ser/Thr ratio : 1/2.
- (2)-proline constitutes 11 to 18 percent and histidine 10 percent of the total amino acids.
- (3)-cysteine is present at a constant percentage of 1 to 1.5 percent.
- (4)-tryptophan, one of the most hydrophobic amino acids, is present at about 1 percent in three of the four cDNA clones, and at 2.3 percent in JER 57.

Besides common features in their amino acid composition, and the fact that no tandemly repeated sequences can be noticed, the four peptides show a similar internal organization in which an alternation of domains can be seen.

Figure 3 shows the hydrophilic/hydrophobic index of peptide stretches and demonstrates that each sequence can be cut into "hydrophilic" zones and "hydrophobic" zones which alternate in a somewhat manner. For example, JUL 7 contains two large hydrophilic domains and two short hydrophobic domains. The JUL 10 clone, in contrast shows shorter hydrophilic/hydrophobic zones. In JER 57 clone, the three hydrophobic domains exhibit very variable size (from 10 to 193 amino acids). Concerning JER 28 the shortest clone, only a few domains are detectable, two hydrophilic and one hydrophobic (figure 3).

Consideration of the percentage of serine and threonine of each domain from each clone (figure 4), indicates that the hydrophilic domains are potentially heavily glycosylated with 47.3 percent of serine and threonine on average. (range 25.4-54.2). The hydrophobic domains are probably less heavily glycosylated with 15.7 percent of serine and threonine on average. (range 0-33.3).

"Hydrophilic domains" vary in size from 20 to 357 amino acid residues, although domains located close to N- or C-terminal part of the clone might be longer on the corresponding full length mRNA. Hydrophobic or "naked" regions vary from 33 to 203 amino acids.

A third point worth noting is the striking partition of some residues like proline and histidine. In the four clones, polyproline structures

comprising 3 to 6 consecutive proline residues are only found in hydrophobic domains. The same observation can be made for histidine residues which are also mainly found in the hydrophobic domains in some cases as polyhistidine structures of two to four consecutive residues.

### ***Nucleotide sequence analysis of mucin cDNAs.***

1) *Southern blot analysis, chromosomal localization and in situ hybridization* : Human genomic DNA from the lymphocytes of 9 healthy volunteers was digested with Eco RI, Bam HI, Hind III, Xba I or Pst I. In each instance, the four cDNA probes hybridized to the same restriction fragments (figure 5) : a single band > 30 kb was seen with Eco RI, digestion with Bam HI revealed three fragments at 1.8, 1.3 and 0.3 kb, a single 25 kb fragment was seen with Hind III, a single broad band > 30 kb with Xba I, and then four fragments were cut by Pst I, the highest at 1.4 kb, a broad band at 1-1.2 kb, and two small bands at 0.9 and 0.7 kb. In order to map these four cDNAs, hybridization to a panel of 25 independant human-rodent somatic cell hybrids was performed. No cross-hybridization with hamster or mouse DNA was noticed. There was a complete correlation between the presence of BamHI hybridization bands and the presence of human chromosome 11 in all hybrids. So, the four sequences were localized on chromosome 11. Complete analysis including *in situ* hybridization was carried out with the JER 57 probe which was assigned to 11 p 15. This work has been described in detail elsewhere (23).

2) *Comparison of the nucleotide sequences* : The four nucleotide sequences showed a high level of similarity (figure 1) with a percentage from 71% (JUL 10 compared to JER 57) to 96% (JER 28 compared to JER 57).

In spite of the fact that no repetitive peptide sequence could be deduced from the nucleotide sequences, we observed that some nucleotides were constantly repeated in the sequence, for example three consecutive triplets CCC-CTC-CTC. Using these observations, we were able to show that the sequences of the four clones were based upon a succession of imperfect tandem repeats (figure 1). Compiling the 55 complete tandem repeat sequences, the most frequent is a 87 base pair structure, its sequence is shown on figure 6. However only 22 units have the perfect 87 base pair sequence, 27 are shorter because of nucleotide

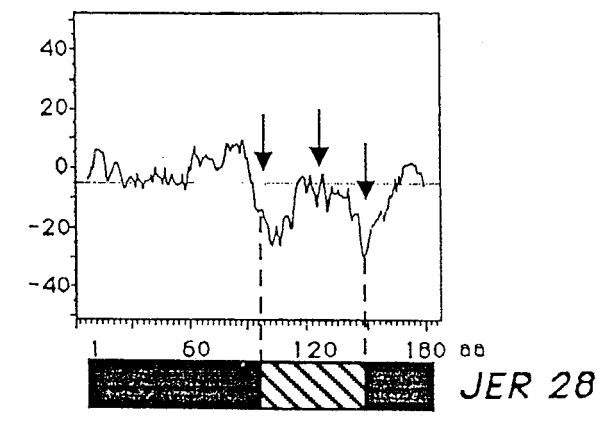
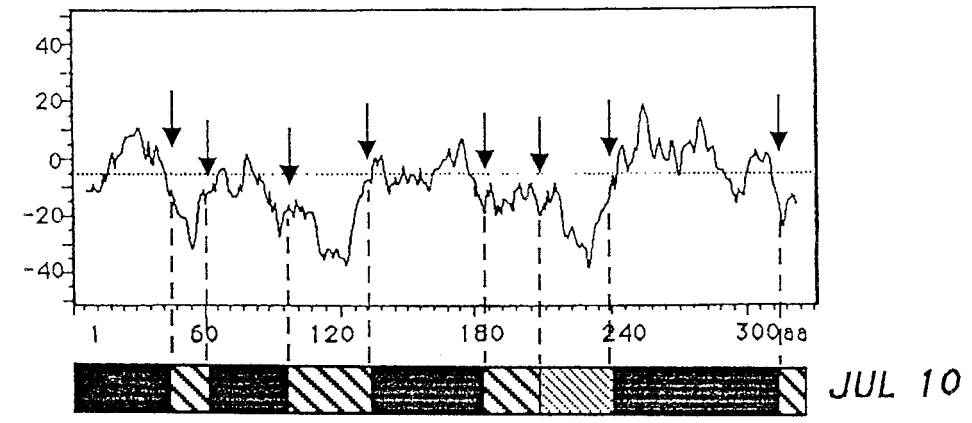
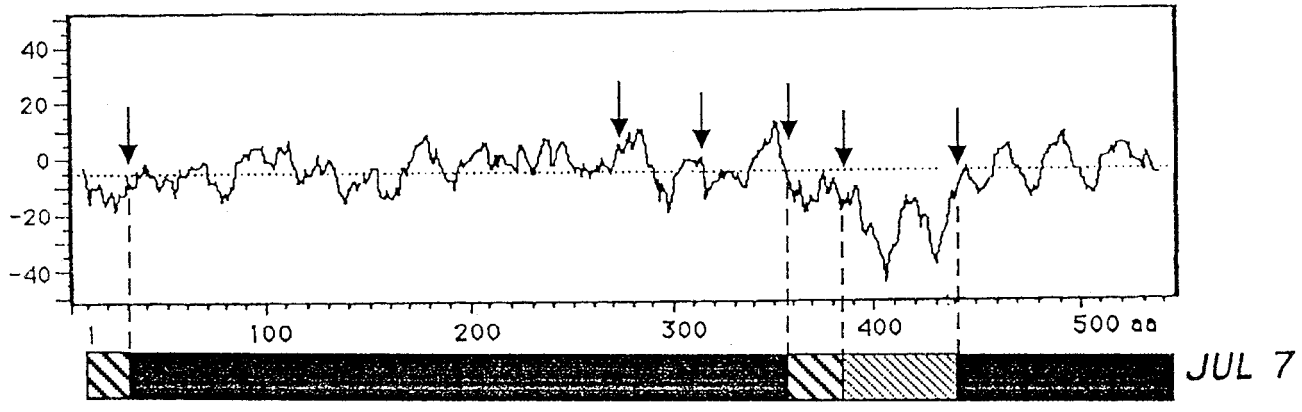
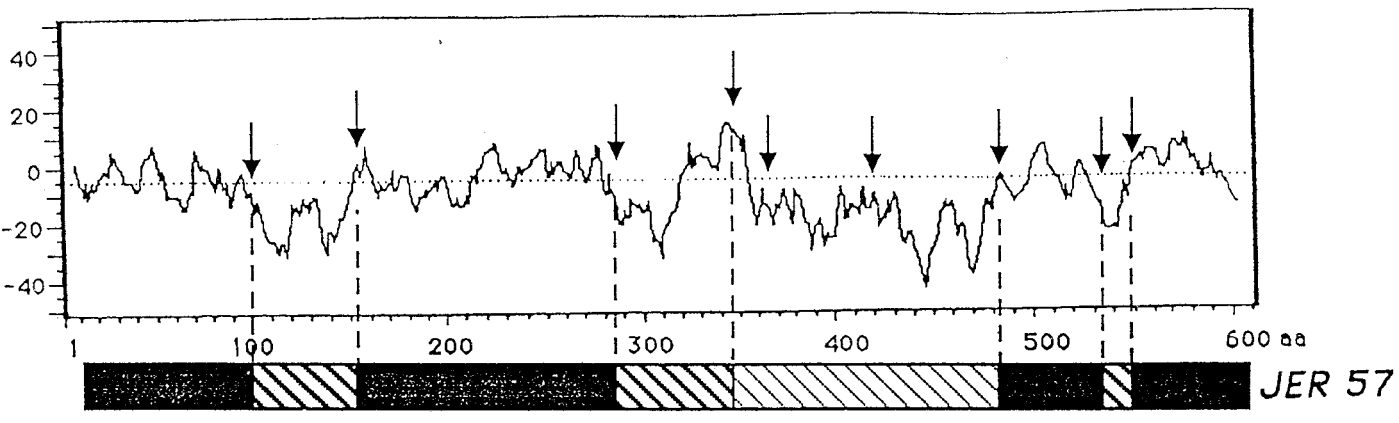


Figure n°3

clone	domains (amino acid positions)	hydrophilic = H hydrophobic = h	Ser + Thr percentage
<b>JER 57</b>	1 to 99	H	25.2
	100 to 152	h	22.6
	153 to 284	H	52.9
	285 to 348	h	26.5
	349 to 478	very h	10.6
	479 to 532	H	53.7
	533 to 543	h	36.3
	544 to 610	H	53.7
<b>JUL 7</b>	1 to 29	h	0.0
	30 to 354	H	49.4
	355 to 382	h	7.1
	383 to 437	very h	10.9
	438 to 544	H	50.4
<b>JUL 10</b>	1 to 40	H	50.0
	41 to 62	h	4.5
	63 to 90	H	42.7
	91 to 135	h	26.6
	136 to 181	H	54.2
	182 to 204	h	4.3
	205 to 242	very h	12.8
	243 to 313	H	40.8
	314 to 330	h	23.4
<b>JER 28</b>	1 to 95	H	45.1
	96 to 156	h	23.3
	157 to 187	H	48.3

Figure n°4



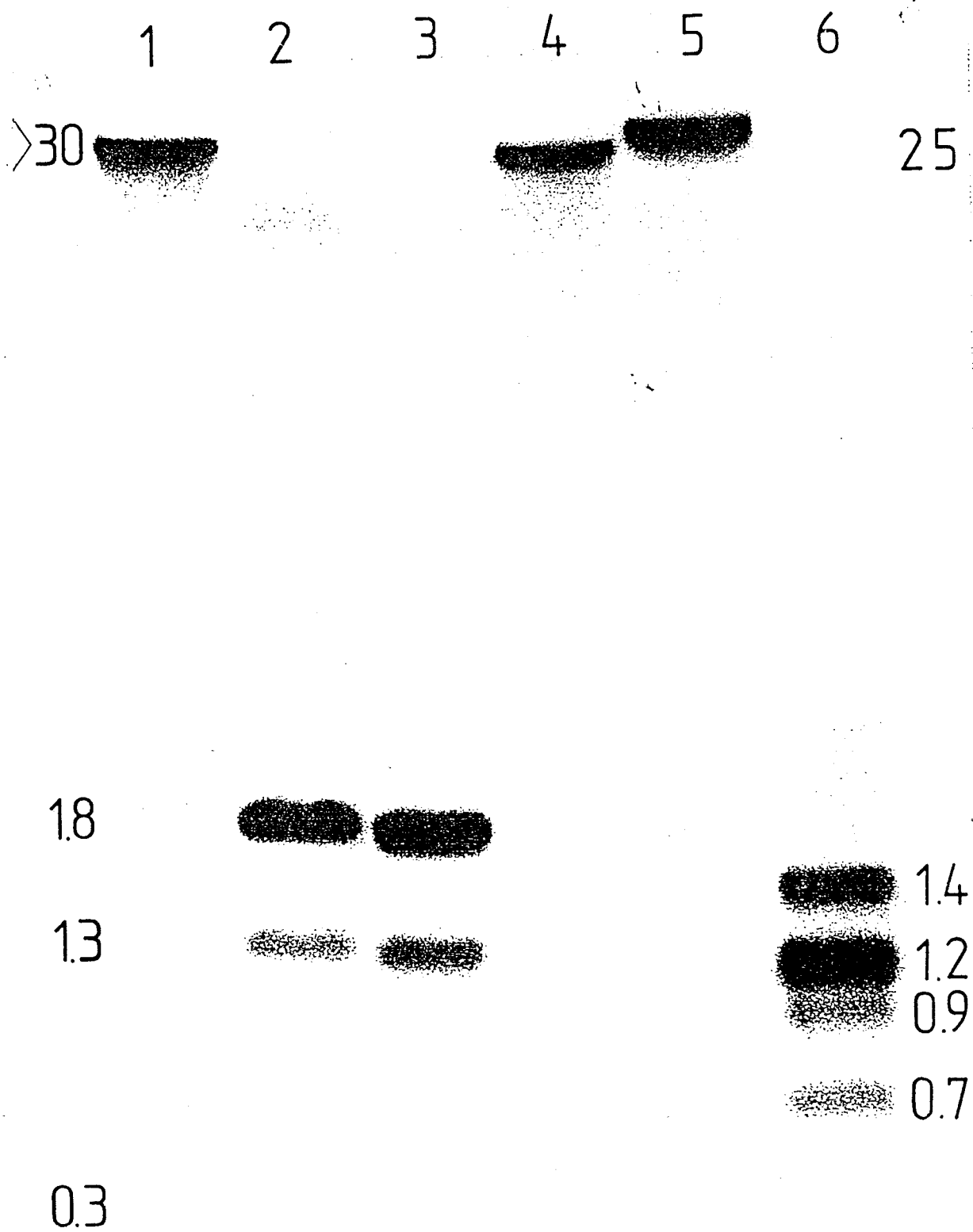


Figure n°5

10                      20                      30                      40  
|                      |                      |                      |  
1 TCCTCCACCCCAGGGACTGCCACACCCTCACAGTGCTGACCACC  
46 ACAGCCACCACACCACAGCCACCGGCTCCACAGCCACCCCC

Figure n°6

clone	nucleotide position	number of nucleotides i = inserted d = deleted	reading frameshift	amino acid position	peptide hydrophilic (H) hydrophobic (h) alternation
JER 57	295	1 i	+	99	H→h
	456	1 d	+	152	h→H
	516	9 d	0	172	0
	699	19 d	+ } 0	233	0
	701	1 i	+ }	234	0
	852	1 d	+	284	H→h
	872	3 d	0	291	0
	1043	1 i	+	348	h→more h
	1098	13 i	+	366	h→more h
	1155	27 d	0	385	0
	1258	16 d	+	420	0
	1345	15 i	0	449	0
	1437	4 d	+	479	h→H
	1594	2 d	+	532	H→h
	1627	4 d	+	543	h→H
	1638	3 i	0	546	0
	1767	9 d	0	589	0
JUL 7	87	29 d	+	29	h→H
	182	21 d	0	61	0
	371	9 d	0	124	0
	832	8 d	+	278	0
	932	10 d	+	311	0
	1061	1 d	+	354	H→h
	1145	16 d	+	382	h→more h
	1223	15 i	0	408	0
1311	4 d	+	437	h→H	
JUL 10	118	1 i	+	40	H→h
	186	1 d	+	62	h→H
	269	2 d	+	90	H→h
	405	1 d	+	135	h→H
	441	9 d	0	147	0
	541	1 d	+	181	H→h
	610	16 d	+	204	h→more h
	726	1 d	+	242	h→H
	773	4 d	+ } 0	258	0
	823	1 i	+ }	275	0
937	1 i	+	313	H→h	
JER 28	115	3 d	0	39	0
	285	1 d	+	95	H→h
	314	15 d	0	105	0
	394	1 d	+	132	0
	468	16 d	+	156	h→H

Figure n°7

A

B

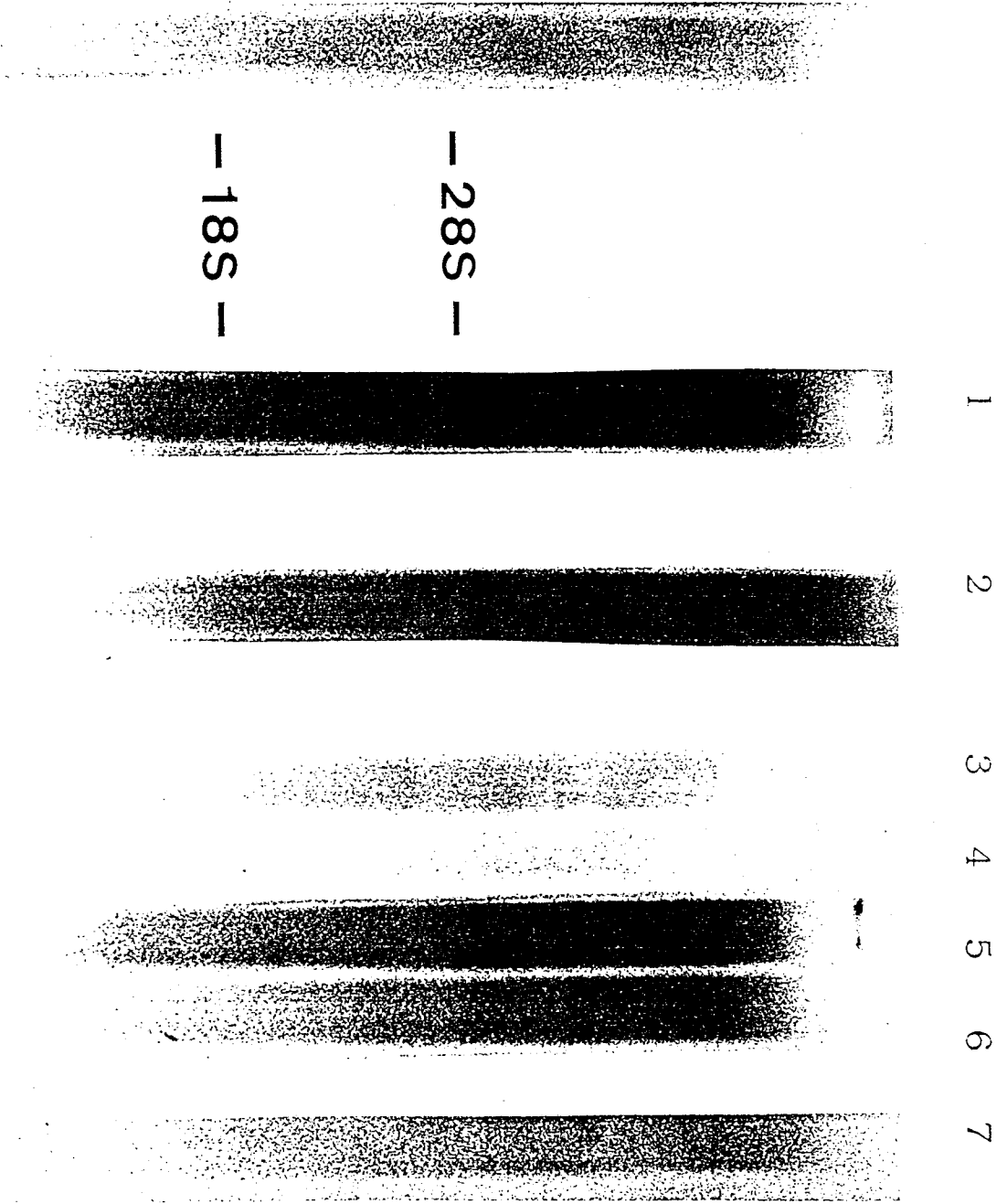


Figure n°8

deletions, and 6 are longer because of nucleotide insertions. The shortest is constituted of 25 nucleotides, the longest of 102 nucleotides.

By analysing the clones in this way, one can make the following observations about each clone : JER 28 is constituted of seven tandem repeats. Four nucleotide deletions are detected (figure 1d). The first one (following nucleotide position n°115) results in the deletion of a threonine residue. The three other deletions introduce a shift in reading frame which results in the peptide sequence by the substitution of many serine and threonine residues by proline and histidine residues.

In the three other clones, most of nucleotide insertions or deletions introduce a shift in reading frame (figure 1\*). Deletions are more frequent modifications than insertions. Variable numbers of nucleotides are deleted (from 1 to 62) or inserted (from 1 to 15). Most of modifications in nucleotide tandem-repeat create shifts in reading frame : out of 42 modifications, 30 shifts in reading frame occur. Four shifts in the reading frame nullify each other in two cases because of a deletion immediately followed by an insertion which restores the reading frame. Comparison of the bounds of each domain on hydrophobic curves of the proteins and the nucleotide sequence reveals that out of 26 shifts in the reading frames (arrows on hydrophobic curves), 21 coincide with inversion in hydrophobic curves (figure 7).

The 204 first nucleotides of JER 57 do not exhibit any homology with the 87 base pair tandem repeat and may be considered as a very long inserted structure, but because this structure coincides with the N-terminal part of the clone, it may be incomplete.

### ***RNA blot analysis.***

2µg poly A<sup>+</sup> RNA from different human mucosae : tracheal mucosa, bronchial mucosa, gastric mucosa (antrum and fundus), colonic mucosa (ascending colon, transverse colon, descending colon) were subjected to northern blot analysis. The same hybridization signal was obtained on each blot with the four cDNA probes (figure 8). In all cases, the message was uniformly polydisperse in size (up 20 kb to 1 kb). A very strong signal after only 2 hours of autoradiography was observed for tracheal or bronchial mucosae. A longer exposure time was required (from 24h to 48h) for the digestive mucosae tested. The signal was resistant to very stringent washing conditions (1 X SSC, 0.1% SDS, 65°C). Several other human tissues were also tested in the same way : heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas using a

northern blot obtained from Clontech Laboratories, Inc. No hybridization signal could be observed, even after 1 week exposure (data not shown).

An Eco RI-PstI fragment from JER 57 which comprises the first 93 nucleotides was also used to probe the bronchial mucosa northern blot and produced same hybridization signal as the complete probe.

Three short oligonucleotides chosen to be able to recognize a perfect 87 bp tandem repeat (NAU 23) or inserted domains of degenerate tandem repeats (NAU 24, NAU 28) produce the same polydisperse hybridization signal as the corresponding entire probes

**Homology search.** Computer searches for similarities of the deduced amino acid sequences with proteins registered in the University of Geneva Protein (Swiss-Prot) revealed no statistically significant sequence homology with any protein including MUC 1 (10, 12), MUC 2 and 3 (13, 14), MUC 4 and 5 (15, 16), porcine (9), bovine (32), submaxillary mucins.

## DISCUSSION

Here, we report the complete analysis of four mucin cDNAs which all map to chromosome 11 p15 and which constitute a new mucin peptide family. The special feature of these peptides is the complete lack of tandem repeat organization, in spite of the fact that nucleotide sequence contains numerous internal repeats. The peptide structure in each case appears as an irregular alternation of potentially heavily glycosylated domains and potentially naked domains. The twenty clones originally isolated from a  $\lambda$ gt 11 cDNA library can be classified into four families : JER 64 (called MUC4 in Human Gene Mapping nomenclature) was mapped to 3q29 (15, 25) ; the three other clone families all mapped to 11p15 (27). Complete analysis of one of them is presented in this paper. This family comprises four clones JER 28 and JER 57 (from the  $\lambda$ gt 11 cDNA library) and JUL 7 and JUL 10 (from the  $\lambda$ gt 10 cDNA library). There are six lines of evidence for this that these four clones encode human mucins.

1) The immunoserum used for their isolation was raised against human tracheobronchial apomucins. It is unlikely that the immunoserum contains significant levels of antibodies directed against other tracheobronchial proteins because of the extensive pronase digestion and drastic TFMS treatment which are used in the preparation of the apomucins.

2) Purified antibodies prepared from fusion-proteins recognized only mucous cells and goblet cells from tracheal and bronchial epithelia (20).

3) Northern blot analysis showed that the mRNA detected by the four cDNA clones was large and very polydisperse and characteristic of the patterns seen with other mucin probes. It was expressed in various mucosae but not in the other tissues tested.

4) The amino acid composition and in particular the high percentage of hydroxy amino acids the four deduced peptide sequences is characteristic of mucins.

5) The probes map to chromosome 11p15 which is the same region as two other mucin cDNA families : intestinal cDNAs encoding for 23 amino acid tandem repeat peptides and products of the gene MUC 2 (13), and tracheobronchial cDNAs encoding for 8 amino acid tandem repeat peptides (23).

6) Nucleotide sequences of the four cDNAs presented here show the typical internal tandem repeat organization described for all animal or human mucin structures.

However the peptide sequences deduced from cDNAs of this mucin family contrast to other previously described, by a completely lack of tandem repeats. This is due to numerous deletions or insertions in a repetitive nucleotide structure which introduce shifts in reading frame.

A 89 bp DNA probe and 16 or 17 bp oligonucleotides which all only recognized inserted regions from degenerate tandem repeats gave on RNA blots the same polydisperse signal as the entire four cDNA probes.

This clearly indicates that these nucleotide structures are also present into bronchial mucosa RNA from another individual and not only into our cDNA library.

This phenomenon raises several questions : why is there such degeneracy ? Is it related to biological functions within the mucosa ? Is it related to the position of these cDNA stretches within the corresponding full length mRNA, as was previously observed for MUC1 (28). In this case, the full length protein may be composed of a repetitive domain with non repetitive domains located in the N- and C-terminal parts. If this is the case the full length mRNA must be very long, since the sequences presented here have a total of more than 5 kb.

The existence of alternating hydrophilic and hydrophobic domains is perfectly consistent with structural studies on mucins : Sheehan and Carlstedt have shown, by electron microscopy, that antibodies raised against naked domains bind to cervical mucin molecules in an irregular manner (29). The same observation has been made on gastric, bronchial and salivary mucins (30). Furthermore, Shankar et al have provided evidence for presence of hydrophobic zones in human airway mucins by the use of a fluorescent hydrophobic probe (31).



Thus for the first time, the nucleotide sequences of this mucin family are consistent with a classical structural model of mucin proteins.

It is noteworthy that these new mucin sequences map to the same region of chromosome 11 as the (two) other classes of mucin clones we have characterised and also the MUC 2 gene. The lack of sequence identity with MUC 2 and difference in tissue expression indicates that none of the tracheobronchial mucin clones are part of MUC 2 and suggest that at least one more gene (MUC 5) is located in this region. However, it is not yet clear whether all the classes of tracheobronchial mucin clones which map to 11p15 are products of a single or several genes, or how large the region is. We hope that restriction mapping of large size DNA fragments using PFGE will help to answer these questions.

*Acknowledgment.* This work was supported by the Association pour la Recherche sur le Cancer. We thank Danièle Petitprez and Pascal Mathon for their excellent technical assistance, Dr. Dallas Swallow for helpful discussions and reading the manuscript and Dr Nguyen Van Cong for the chromosomal localization.

## REFERENCES

1. STROUS G. J., DEKKER J. (1992) *Crit. Rev. Biochem. Mol.*, **27**, 57-92.
2. ROSE M.C., KAUFMAN B., MARTIN B. (1989) *J. Biol. Chem.*, **264**, 8193-8199.
3. ALLEN, A. (1981) in *Structure and Function of Gastrointestinal Mucus* (Johnson, L.R. ed.), Raven Press, New-York.
4. CARLSTEDT I., LINDGREN H., and SHEEHAN J. K. ULMSTEN U., WINGERUP L. (1983) *Biochem. J.*, **213**, 427-435.
5. SLAYTER H. S., LAMBLIN G., LE TREUT A., GALABERT C., HOUDRET N., DEGAND P., ROUSSEL P. (1984) *Eur. J. Biochem.*, **142**, 209-218.
6. ALLEN A., SNARY D. (1972) *Gutt.*, **13**, 666-672.
7. SLOMIANY A., OKAZAKI K., TAMURA S., SLOMIANY B. L. (1991) *Arch. Biochem. Biophys.*, **286**, 383-391.
8. SCAWEN M., ALLEN A. (1977) *Biochem J.*, **163**, 363-368.
9. TIMPTE C. S., ECKHARDT A. E., ABERNETHY J. L., HILL R. L. (1988) *J. Biol. Chem.*, **263**, 1081-1088.
10. GENDLER S., TAYLOR-PAPADIMITRIOU J., DUHIG T., ROTHBARD J., BURCHELL J. (1988) *J. Biol. Chem.*, **263**, 12820-12823.
11. SIDDIQUI J., ABE M., HAYES D., SHANI E., YUNIS E., KUFE D. (1988) *Proc. Natl. Acad. Sci. USA.*, **85**, 2320-2323.
12. LAN M. S., BATRA S. K., QI W. N., MEZGAR R. S., and HOLLINGSWORTH M. A. (1990) *J. Biol. Chem.*, **265**, 15294-299.
13. GUM J. R., BYRD J. C., HICKS J. W., TORIBARA N. W., LAMPORT D. T. A., and KIM Y. S. (1989) *J. Biol. Chem.*, **264**, 6480-6487.
14. GUM J. R., HICKS J. W., SWALLOW D. M., LAGACE R. L., BYRD J. C., LAMPORT D. T. A., and SIDDIKI B., and KIM Y. S. (1990) *Biochem. Biophys. Res. Comm.*, **171**, 407-415.
15. PORCHET N., NGUYEN V.C., DUFOSSE J., AUDIE J.P., GUYONNET DUPERAT V., GROSS M.S., DENIS C., DEGAND P., BERNHEIM A. and AUBERT J.P. (1991) *Biochem. Biophys. Res. Comm.*, **175**, 414-422.
16. AUBERT J.P., PORCHET N., CREPIN M., DUTERQUE-COQUILLAUD M., VERGNES G., MAZZUCA M., DEBUIRE B., PETITPREZ D., and DEGAND P. (1991) *Am. J. Respir. Cell. Mol. Biol.*, **5**, 178-185.
17. GLISIN V., ORKVENJAKOV R., and BYUS C. (1974) *Biochemistry.*, **13**, 2633-2637.
18. DEGAND P., ROUSSEL P., LAMBLIN G. (1973) *Biochem. Biophys. Acta.*, **320**, 318-330.
19. LAMBLIN G., DEGAND P., ROUSSEL P., HAVEZ R., HARTEMANN E., FILLIAT M. (1972) *Clin. Chim. Acta.*, **36**, 329-340.
20. CREPIN M., PORCHET N., AUBERT J. P., and DEGAND P. (1990) *Biorheology.*, **27**, 471-484.

21. SANGER F., NICKLEN, S. and COLSON A. R. (1977), *Proc. Natl. Acad. Sci. USA.*, **74**, 5463-5467.
22. NGUYEN V. C., WEIL D., FINAZ C., COHEN-HAGUENAUER O., GROSS M. S., JEGOU-FOUBERT C., TAND M. F., COCHET C., GROUCHY J., FREZAL J. (1986) *Ann. Génét. (Paris)*, **29**, 20-26.
23. NGUYEN V. C., AUBERT J. P., GROSS M. S., PORCHET N., DEGAND P., and FREZAL J. (1990) *Hum. Genet.*, **86**, 167-172.
24. KYTE J., DOOLITTLE R.F. (1982) *J. Mol. Biol.*, **157**, 105-132.
25. GROSS M. S., GUYONNET-DUPERAT V., PORCHET N., BERNHEIM A., AUBERT J. P., NGUYEN V. C. (1992) *Ann. Genet.*, **35**, 21-26.
26. PERINI J.M., VANDAMME-CUBADDA N., AUBERT J.P., PORCHET N., MAZUCCA M., LAMBLIN G., HERCOVICS A., and ROUSSEL P. (1991) *Eur. J. Biochem.*, **196**, 321-328.
27. PORCHET N., DUFOSSE J., AUDIE J. P., GUYONNET DUPERAT V., PERINI J. M., NGUYEN VAN CONG, DEGAND P., AUBERT J. P. (1991) *Am. Rev. Respir. Dis.*, **144**, 15-18.
28. LIGTENBERG M. J. L., VOS H. L., GENNISSSEN M. C., HILKENS J. (1990) *J. Biol. chem.*, **265**, 5573-5578.
29. SHEEHAN J. K., CARLSTEDT I. (1990) *Biochem. J.*, **265**, 169-178.
30. SHEEHAN J. K., BOOT-HANDFORD R. P., CHANTLER E., CARLSTEDT I., THORNTON D. J. (1991) *Biochem. J.*, **274**, 293-296.
31. SHANKAR V., NAZIRUDDIN B., REYES DE LA ROCHA S., SACHDEV G. P. (1990) *Biochemistry.*, **29**, 5856-5864.
32. BHARGAVA A. K., WOITACH H. T., DAVIDSON E. A., BHAVANANDAN V. P. (1990) *Proc. Natl. Acad. Sci. USA.*, **87**, 6798-6802.

## LEGENDS FOR FIGURES

Figure 1 - *Nucleic acid sequences of :*

- a) JUL 7
- b) JUL 10
- c) JER 57
- d) JER 28

\* indicates a shift in the reading frame

Figure 2 - *Deduced amino acid sequences and amino acid compositions of*

- a) JUL 7
- b) JUL 10
- c) JER 57
- d) JER 28

Figure 3 - *Hydropathy curves of JER 57, JUL 7, JUL 10, JER 28.*

Black rectangles represent the potentially heavily glycosylated domains which are hydrophilic domains. Diagonally shaded rectangles indicate the potentially naked domains which are hydrophobic (darker) or very hydrophobic (lighter) domains.

Figure 4 - *Hydroxy amino acid percentage of each peptide domain from JER 57, JUL 7, JUL 10 and JER 28.*

According to hydropathy curves (figure 3), each peptide sequence of each clone was cut into peptide domains. Arrows on figure 3 indicate the bounds of each domain. Serine and threonine percentage is calculated.

Figure 5 - *Southern blot analysis .*

Human genomic DNA blots were analyzed using JER 57 as a probe. DNA prepared from the lymphocytes of one healthy donor (JPA) was digested by Eco RI (lane 1), Bam HI (lane 2), Eco RI-Bam HI (lane 3). Hind III (lane 4), Xba I (lane 5), Pst I (lane 6). 20 µg of digested DNA were loaded per lane.

Figure 6 - *Nucleotide sequence analysis.*

Nucleotide sequence of the four cDNA clones were computed and nucleotide sequence representing the most frequently 87 base pair tandem repeat was deduced. Here is its sequence presented.

Figure 7 - *Degenerescence into tandem repeat nucleotide sequences.*

Each tandem repeat of each clone was compared to the virtually 87 base pair most frequent tandem repeat. All nucleotide insertions or deletions are listed.

Figure 8 - *RNA blot analysis.*

Poly (A<sup>+</sup>) RNA samples (2 µg) from human mucosae were analyzed as described under "Materials and methods".

In part A, bronchial mucosa was probed with 90 bp Eco RI - Pst I fragment from JER 57. Autoradiography exposure time was 15<sup>h</sup>.

In part B, tracheal mucosa (lane 1), bronchial mucosa (lane 2), gastric mucosae (fundus in lane 3, antrum in lane 4), colonic mucosae (ascending colon in lane 5, descending colon in lane 6, transverse colon in lane 7) were probed with JER 57.

Hybridization signals were obtained after autoradiography exposure time 3<sup>h</sup> for respiratory mucosae samples and 24<sup>h</sup> for digestive mucosae samples.

## APPENDICE TECHNIQUE

### PREPARATION DES ANTICORPS POLYCLONAUX

Des mucines trachéobronchiques humaines ont été hydrolysées par la pronase puis déglycosylées par le T.F.M.S. selon le protocole défini par Crépin et al.(66). De manière à obtenir un immunosérum ayant la plus large représentativité, les produits de déglycosylation des sous fractions de mucines sialilées, sulfatées et fucosylées ont été mélangés en égale proportion. Le mélange a été injecté à deux lapins selon un protocole proche de Vaitukaitis (1981) (70). Le sérum recueilli après la troisième saignée a pu être utilisé comme sonde immunologique.

La première injection était de 500  $\mu$ g de matériel, les injections de rappel à 15 jours d'intervalle étaient de 200  $\mu$ g en présence d'un volume d'adjuvant complet de Freund.

Plusieurs saignées de 10 à 15 ml ont été réalisées. L'immunosérum est obtenu après centrifugation du sang total pendant 5 minutes à 2500 trs/mn, puis il est réparti en aliquots et conservé à -20°C.

La spécificité de l'immunosérum a été vérifiée sur différentes coupes histologiques de muqueuses trachéales humaines (66).

### CRIBLAGE DE LA BANQUE EN $\lambda$ gt11 A L'AIDE DE SONDAS IMMUNOLOGIQUES

Chacune des deux souches E. Coli (Y 1088 et Y 1090) sont relancées en LB Amp. 50  $\mu$ g/ml, maltose 0,4%, une nuit à 37°C.

0,4 ml des relances sont mis en culture 3 heures à 37°C dans 20 ml de LB Amp. 50  $\mu$ g/ml, maltose 0,4 % suivis d'une centrifugation 10 minutes à 3000 trs/min. Les culots bactériens sont repris dans 6 ml de MgSO<sub>4</sub> 0,01M. Le contact entre les bactériophages et les deux souches bactériennes est réalisé comme suit :

- un aliquot de la banque est mis en contact avec 30  $\mu$ l d'E. Coli Y 1088 pendant 15 minutes,

- 150  $\mu$ l de la culture E. Coli Y 1090 et 3 ml de Top Agar sont ajoutés. Le tout est étalé sur milieu solide (LB Amp 50  $\mu$ g/ml)

- Les boîtes où bactéries (Y1088 et Y1090) et bactériophages ont été mis en contact sont incubées 3 heures à 42°C puis une nuit à 37°C.

- Après trois heures à 42°C, les nitrocelluloses saturées par l'IPTG vont servir de support de transfert des protéines synthétisées par la souche bactérienne (71,72).

- Le lendemain, ces nitrocelluloses repérées sont incubées en tampon TBS (Tris 50mM, pH 8, NaCl 0,15M) SVF 20 % pendant 30 minutes.

- Elles sont mises en contact avec les anticorps primaires (4 heures) puis secondaires (1 heure) dilués en tampon TBS, SVF 20 %. Ces anticorps secondaires sont des anticorps de chèvre anti-anticorps de lapin, ils sont marqués à la peroxydase. Entre chaque contact, ces nitrocelluloses subissent 3 lavages de 7 minutes en tampon TBS, TBS NP40 0,1%, TBS.

- La réaction colorée (chloronaphtol et eau oxygénée) obtenue est bleu-violacée, elle est stoppée par immersion des nitrocelluloses dans l'eau.

- Les populations (clones) positives apparaissent sous forme d'anneaux où la coloration est plus intense à la périphérie. Elles objectivent une réaction entre l'anticorps primaire et la protéine antigène synthétisée.

- Grâce aux repères portés sur les nitrocelluloses, les clones réagissant positivement sont repiqués (cure-dents) au niveau des boîtes et conservés (tampon TMG + chloroforme). Une amplification de la réponse est réalisée par un nouveau criblage.

## **PURIFICATION, A PARTIR DE L'IMMUNSERUM POLYCLONAL ANTIMUCINE TRACHEALE HUMAINE DEGLYCOSYLEE, DES ANTICORPS SPECIFIQUES DE CHAQUE CLONE**

La technique se confond, en un premier temps, avec le criblage immunologique en vecteur  $\lambda$ gt11 mais elle en diffère après l'incubation avec l'immunsérum dilué. Les rondelles de nitrocellulose sont traitées comme suit :

- Lavage 7 minutes en tampon TBS
- Lavage 7 minutes en tampon TBS + NP40 à 0,1 %
- Lavage 7 minutes en tampon TBS.

Les rondelles sont alors mises en présence d'un tampon d'élution (Tris-HCl 10mM pH7,6 MgCl<sub>2</sub> 4M, sérum albumine bovine 1 %) comme décrit par TIMPTE et coll. (1988) (56). Après rinçage des rondelles pour ce même tampon, le produit de décrochage est dialysé contre du Tris-HCl 10mM pH8, NaCl 0,15M, EDTA 1mM, Triton X100 0,1%. Les anticorps récupérés sont concentrés et testés sur coupes histologiques de muqueuses trachéale et colique humaine.

## **CARACTERISATION DE CLONES SELECTIONNES**

### **1. Amplification en milieu liquide**

De manière à augmenter la concentration en particules phagiques, un préministock puis un ministock sont réalisés comme suit :

1ère étape : le préministock

- La souche bactérienne Y1088 est préparée la veille du contact par une étape de croissance en milieu LB + Ampicilline 50  $\mu$ g/ml + Maltose 0,4 %.

Le clone positif est prélevé à la micropipette sous forme d'une carotte et incubé en milieu Mg SO<sub>4</sub> 10 mM, en présence d'E. Coli Y 1088.

Après 15 minutes de contact à 37°C sans agitation, le mélange est ajouté à 5ml de LB-ampicilline 50  $\mu$ g/ml, CaCl<sub>2</sub> 5 mM et incubé à 37°C jusque lyse complète des bactéries Y 1088. 100  $\mu$ l de chloroforme sont alors ajoutés, après 15 minutes de contact, le milieu est centrifugé 10 minutes à 7000 trs/mn (J21 Beckman). Le



surnageant contenant les phages est récupéré et conservé en présence de traces de chloroforme.

Ceci constitue notre préministock de phages.

2ème étape : le ministock :

- 1 ml de la souche bactérienne (cf preministock) est mise en culture dans la 20 ml de LB + Ampicilline 50 µg/ml + maltose 0,4 %, 3h à 37°C.

- La concentration en Mg<sup>2+</sup> est amenée à 10mM.

- 2 ml de la souche bactérienne est mise en contact dans 2 ml de LB + ampicilline 50 µg/ml + Mg<sup>2+</sup> (10mM) + 100 µl de Préministock.

- Après 15 minutes de contact, on ajoute 16 ml de LB + ampicilline 50 µg/ml + Ca Cl<sub>2</sub> (10mM).

- La lyse de bactéries est théoriquement obtenue après 3 heures d'incubation à 37°C.

- 200 µl de chloroforme sont ajoutés.

- Le milieu est centrifugé 10 minutes à 7000 trs/mn. Le surnageant constitue le ministock.

## 2. Caractérisation de l'insert

- 3 ml de ministock sont mis en contact avec 1,5 µg de désoxyribonucléase I et 10 µg de ribonucléase A, 30 minutes à 37°C.

- 65 µl de tampon Tris 1M pH 8,0, 130 µl d'EDTA 500 mM, pH 8,0 et 0,15 mg de Protéinase K sont ajoutés.

- 330 µl de sarcosyl à 5% sont ajoutés (destruction des membranes).

- L'ensemble est mis 30 minutes à 37°C.

- Protéines et enzymes sont éliminés par une extraction phénolique. Deux purifications par le chloroforme / alcool isoamylique sont alors réalisées.

- La phase supérieure est précipitée par de l'acétate de sodium et de l'alcool éthylique à -80°C. Elle est alors centrifugée à 10000tr/min pendant 30 minutes.

- Les différents culots de précipitation sont séchés, repris et dialysés en tampon TE (Tris-HCl 10 mM pH 8, EDTA 1mM) puis hydrolysés par Eco RI 30 minutes à 37°C. La taille de l'insert est visualisée en gel d'agarose en présence de BEt.

## CRIBLAGE DE LA BANQUE EN $\lambda$ GT10

Le principe de ce criblage est le même que celui de la banque en  $\lambda$ gt11.  
Le protocole diffère légèrement.:

- les bactéries souches (MN 514 ou C 600 hfl+) sont préparées la veille par deux étapes de croissance en milieu LB maltose 0,4 %.

- Après centrifugation des bactéries à 3000 tr/min 10 minutes, le culot bactérien est repris par une solution de  $Mg SO_4$  0,01 M.

- 100  $\mu$ l de bactéries et 100  $\mu$ l de phages sont mis en contact à 37°C pendant 15 minutes. Les bactéries infectées sont étalées sur milieu solide et incubées une nuit à 37°C.

- Les rondelles de nitrocellulose servent de support au transfert du matériel nucléique. Celui-ci est dénaturé en solution NaOH 0,5M, 5 minutes, puis neutralisé en solution Tris 0,05M pH 7,0, NaCl 1,5 M, 5 minutes.

- L'ADN monobrin est fixé définitivement sur les rondelles par une cuisson à 80°C sous vide 2 heures.

- Les rondelles seront rincées en solution SSC 3X, préhybridées à 42°C 1h30 puis hybridées avec la sonde nucléique une nuit à 42°C sous agitation permanente.

- Les rondelles sont lavées comme suit :

- 2 fois 5 minutes en solution SSC 3X

- 4 fois 5 minutes en solution SSC 0,1 X SDS 0,1 % à 55°C

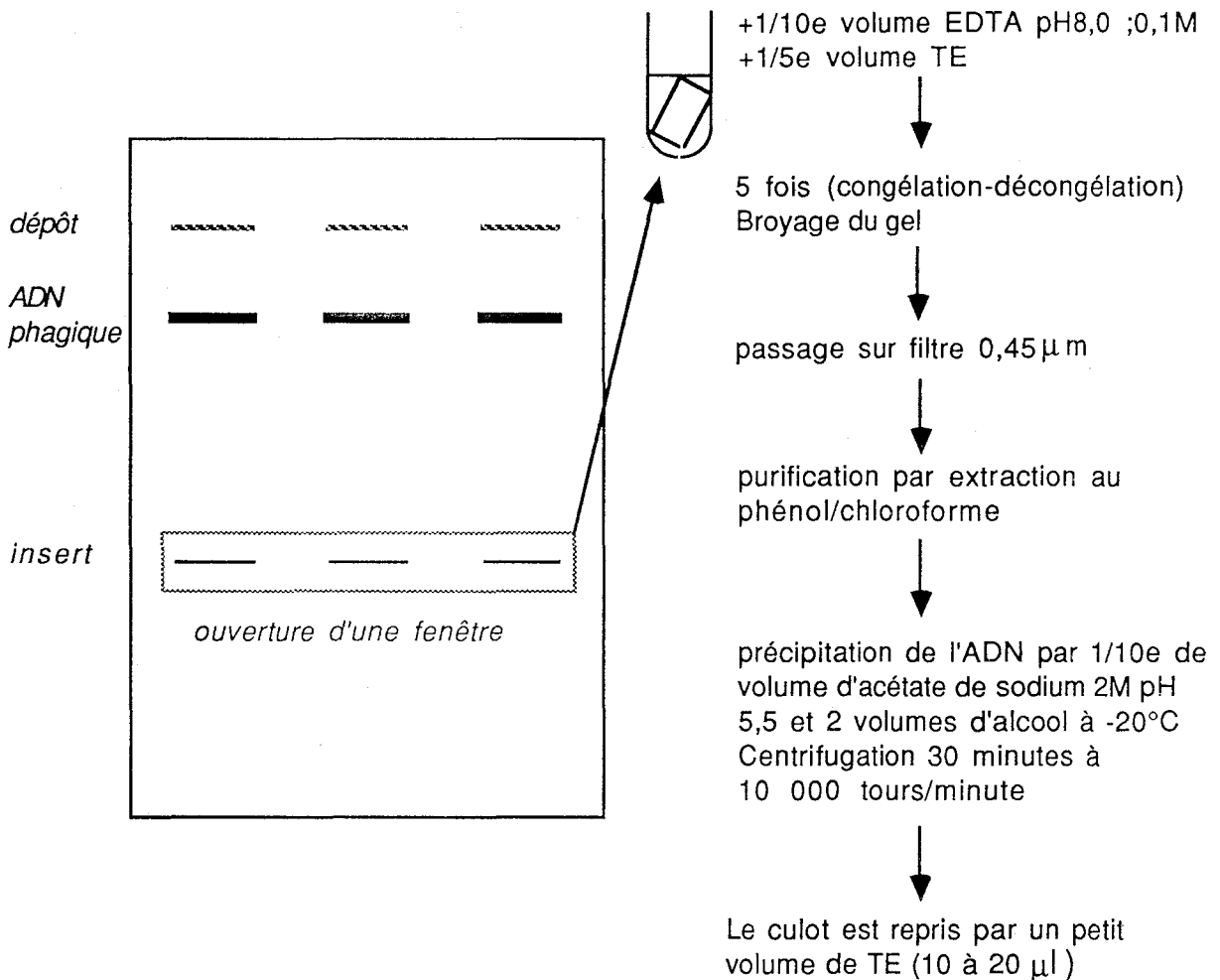
- 1 fois 5 minutes en solution SSC 3X.

- Après un séchage rapide, les rondelles sont mises en autoradiographie. Les clones potentiellement positifs seront contrôlés par un deuxième tour de criblage.

## SOUS CLONAGE DES FRAGMENTS

### Extraction et purification des fragments

Le fragment est libéré par une hydrolyse Eco RI puis isolé grâce à une électrophorèse en gel d'agarose. Coloré par le BEt, en tampon TEA (Tris 0,4 M, EDTA 0,01 M, acide acétique 1 %). La migration est réalisée à faible voltage (20 volts) une nuit et la purification du fragment réalisée comme suit :



## AMPLIFICATION ET PURIFICATION DU PLASMIDE pBR322

Après une phase d'amplification 85 mg de chloramphénicol sont ajoutés et ceci pour un volume de 500 ml de culture bactérienne. Le milieu est incubé 12 heures à 37°C sous agitation forte (blocage de la synthèse des protéines). Ce milieu bactérien est alors centrifugé 10 minutes à 6000 trs/mn et à 4°C. Le culot est repris par 10 ml de tampon "GTE" (glucose 50mM, Tris 25 mM pH 8,0 EDTA 10 mM pH 8,0). Après dissolution du culot, 2 ml de lysosyme en solution à 10 mg/ml dans le GTE sont ajoutés et laissés 10 minutes dans la glace. 40 ml de NaOH 0,2N, SDS 1 % sont ajoutés et le milieu est laissé dans la glace puis 20 ml d'acétate de potassium pH 5,8 sont ajoutés. Après 15 minutes dans la glace le milieu est centrifugé 10 minutes à 10000 trs/mn (J21 Beckman). Le surnageant est filtré sur laine de verre. On lui ajoute 50 ml d'isopropanol. Après une centrifugation de 10 minutes à 5000 trs/mn, le culot est repris par 7 ml de tampon TE. Le milieu est

neutralisé par 300  $\mu$ l de Tris 2M auxquels sont ajoutés 100  $\mu$ l de BEt et 10 g de chlorure de césium. Ce mélange subit une ultracentrifugation à 68000 trs/mn pendant 4 heures 30 minutes à 16°C.

La bande d'ADN plasmidique visible est récupérée à la seringue. Le BEt est éliminé par un volume d'isopropanol alors que la phase inférieure est dialysée 24 heures contre du tampon TE.

L'adialysable est incubé 30 minutes à 37°C en présence de ribonucléase A (0,1 mg/ml) qui sera ensuite éliminée par :

- 1 volume de phénol (centrifugation 5 minutes à 2000 trs/mn)
- 1 volume de phénol / chloroforme (v/v) puis une centrifugation identique
- 1 volume de chloroforme et une nouvelle centrifugation.

On trouve alors, dans la phase supérieure l'ADN plasmidique qui sera précipité par 1/10<sup>e</sup> de volume d'acétate de sodium 2M pH5,5 et 2 volumes d'éthanol à -20°C. La centrifugation qui suit est réalisée à 11000 trs/mn pendant 30 minutes.

Le culot est repris en tampon TE. La qualité et la quantité d'ADN sont contrôlées par spectrométrie U.V..

L'extraction du fragment à partir du plasmide est en tous points semblable à une extraction d'un fragment d'origine phagique. Le marquage du fragment est décrit dans le paragraphe suivant.

## REALISATION DE SONDES NUCLEOTIDIQUES

Deux techniques ont été utilisées :

1. La "**Nick translation**" en présence d'un nucléotide marqué :  $\alpha^{32}\text{P}$  dCTP (Kit Amersham)
2. Le "**multiprime**" qui est un marquage utilisant des amorces hexanucléotidiques et une extension par le fragment de Klenow en présence de  $\alpha^{32}\text{P}$  dCTP.

### 1. La "**Nick translation**"

Cette méthode est utilisée en général lorsque le fragment d'ADN à marquer est intégré dans un vecteur, par exemple un plasmide. Le principe de cette méthode est de mettre en contact le fragment d'ADN que l'on veut marquer avec de

faibles quantités de DNase I, qui opère des coupures simples brins dans l'ADN, et de l'ADN polymérase I qui est capable de catalyser les réactions couplées : élimination du nucléotide en 5' et remplacement de ce nucléotide par polymérisation à partir de l'extrémité 3'OH.

La présence, dans le milieu d'incubation de déoxynucléotides marqués, permet l'incorporation de radio-activité dans la matrice d'ADN.

**Le protocole** : 200 ng de plasmide sont ajoutés à 2  $\mu$ l de tampon 10X contenant un mélange de dATP, dGTP et dTTP ; en présence de 2  $\mu$ l d'enzyme et de 3  $\mu$ l de  $\alpha^{32}\text{P}$  dCTP. Le volume est complété par de l'eau distillée à 20  $\mu$ l. La réaction est effectuée pendant 2 heures à 15°C puis stoppée par un chauffage à 65°C 5 minutes

## 2. Le "Multiprime"

25 ng de fragment d'ADN sont dénaturés par la chaleur (10 minutes à 95°C) en présence d'un mélange d'amorces (hexanucléotiques). Ceux-ci vont s'hybrider tout le long de chacun des deux brins.

Les brins complémentaires sont synthétisés à partir de l'extrémité 3'OH de l'amorce, par le fragment de Klenow de la Polymérase I avec incorporation de  $\alpha^{32}\text{P}$ dCTP ajoutée au milieu réactionnel. (Random Primed DNA Labelling kit Boehringer)

## PURIFICATION DES SONDAS

- Elle est réalisée par chromatographie sur Séphadex G50 en tampon d'élution NaCl 0,15M, EDTA 0,01M, Tris 0,05M pH 7,5, SDS 0,1%.

- La sonde est colorée au rouge de phénol et déposée. La séparation de la sonde et des dNTPs libres est suivie au compteur Geiger. La sonde est progressivement éluée du gel alors que les dNTPs restent dans la partie supérieure de la colonne. Les sondes sont récupérées, dénaturées en présence de soude 300 mM, 10 minutes à température du laboratoire. Un volume d'acide chlorhydrique 3 M est ajouté de façon à obtenir une molarité finale en soude de 50 mM. Un aliquot de la sonde est compté en radioactivité.

## OBTENTION DES SEQUENCES NUCLEOTIDIQUES

L'obtention des séquences nucléotidiques est l'étape définitive de notre étude, et ceci pour deux raisons :

- Grâce au code génétique, une traduction peptidique de cette séquence sera réalisée.

- La séquence nucléotidique permet de caractériser la sonde nucléique qui sera ultérieurement utilisée. Elle est réalisée par l'intermédiaire de différents vecteurs.

Le premier vecteur est le bactériophage M13 qui se multiplie au niveau de colonies d'E.Coli infestées en produisant, dans un premier temps, une forme répliquative double brin circulaire. La forme répliquative est introduite dans une bactérie compétente TG1 par une étape de transformation.

Dans un deuxième temps, lorsque la concentration de la forme double brin atteint environ 200 copies, une forme monobrin est synthétisée et secrétée. Celle-ci est récupérée et utilisée pour le séquençage selon la technique de Sanger et Coll. (1977) (73).

Le second vecteur est le plasmide Bluescript. Nous avons choisi ce vecteur pour son très bon rendement de préparation. Ce vecteur n'étant pas simple brin, une étape supplémentaire de dénaturation du plasmide est nécessaire.

### **Séquençage :**

La technique utilisée est celle de Sanger (1977) (73). L'enzyme utilisée est la "Klenow" (kit Séquénase, version 2.O., United States, Biochemical Corporation).

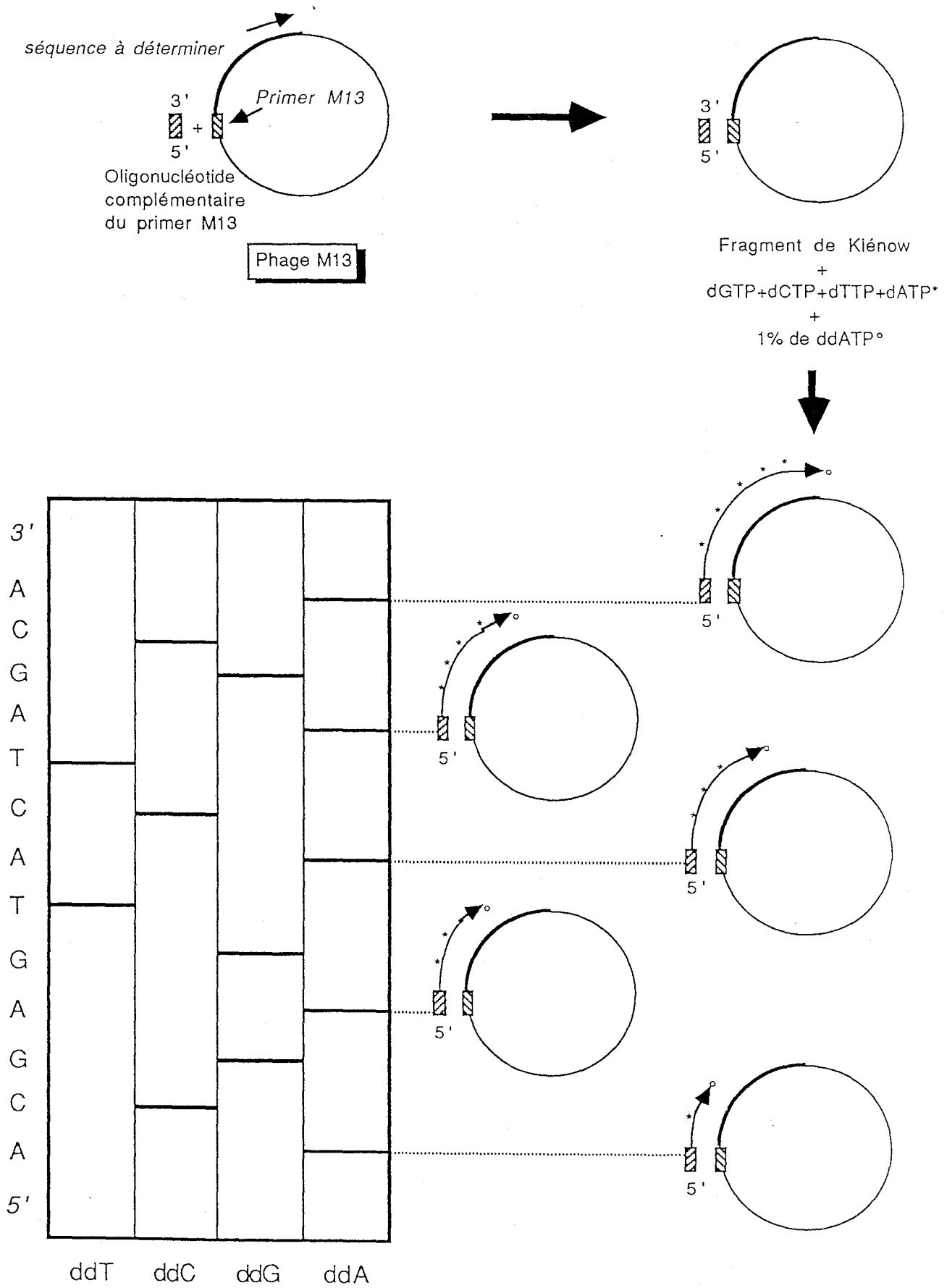


Figure n°38 : Séquençage par la technique de Sanger

Le principe de cette technique est le suivant (figure n°38) : un oligonucléotide complémentaire à la séquence monobrin s'hybride dans une région connue qui précède l'insert. Cet oligonucléotide constitue l'amorce à la polymérisation par l'ADN polymérase 1 ou fragment de Klenow. Cette enzyme est dépourvue de toute activité exonucléasique. Après avoir initié la polymérisation, la réaction est arrêtée par des analogues structuraux : les di-déoxynucléotides triphosphate. La compétition entre les analogues structuraux et les déoxynucléotides permet un arrêt de la réaction au hasard. L'ADN est alors analysé sur gel de polyacrylamide.

La technique de séparation électrophorétique est principalement responsable de la limite de séquençage d'un fragment. Actuellement nous ne pouvons séquencer que 300 pb dans un sens. Le fragment ne doit donc pas dépasser 600 pb en taille. Dans ce cas le fragment pourra être directement ligué au vecteur M13 ou pSK. La ligation est réalisée comme suit :

On utilise de 50 à 100 ng de vecteur digéré par Eco RI. Dans le cas d'un sous clonage dirigé, le vecteur est digéré par les deux enzymes de restriction générant des extrémités compatibles avec celles du fragment.

Les masses moléculaires des vecteurs sont respectivement pour :

M13	6,2 kb
pSK	3,1 kb

La ligation se fait en présence du vecteur, de l'insert purifié à raison de 3 moles pour 1 mole de vecteur, de PEG à 7,5 % et de 5 unités de ligase 2h à température ambiante.

La taille moyenne de nos fragments étant nettement supérieure à 600 pb, il est alors nécessaire de digérer le fragment avant de le sous cloner. Nous avons adopté deux techniques différentes permettant d'obtenir la séquence dans les deux sens.

### **1. Le sous clonage dirigé à partir de la carte de restriction.**

Cette première technique a pour but de découper le fragment trop long en plusieurs fragments de taille inférieure à 600 pb (figure n°39). Ces fragments sont clonés, puis séquencés séparément. La séquence totale est alors reconstituée.



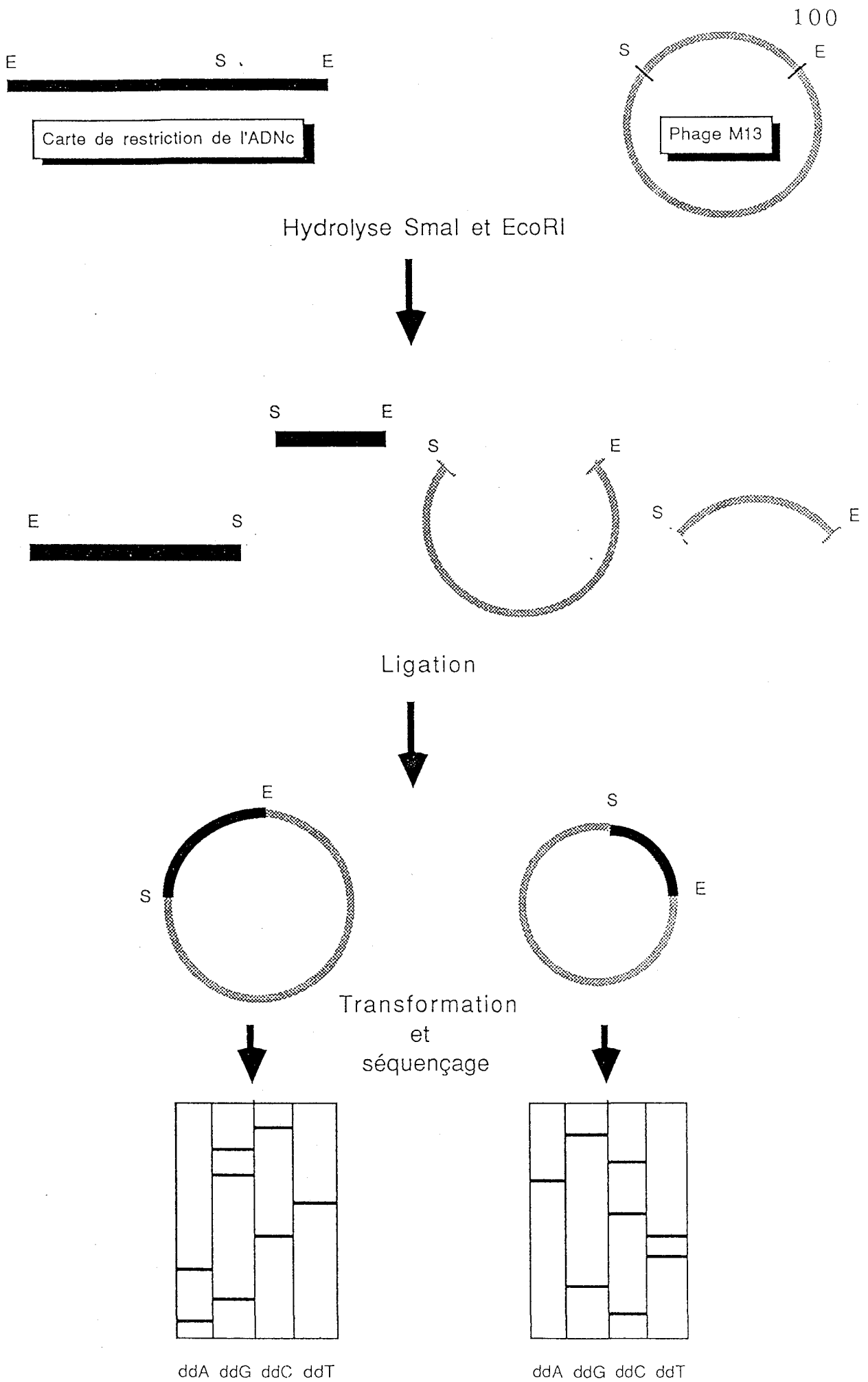


Figure n°39 : Technique de sous clonage dirigé

Si un seul enzyme est utilisé pour libérer le fragment, le clonage ne peut être dirigé: il faudra séquencer plusieurs clones pour espérer obtenir les deux sens.

Si le fragment a été libéré par deux enzymes différents, le sous clonage peut être dirigé et le séquençage dans les deux sens sera alors facilité.

Cependant, la carte de restriction ne permet pas toujours d'extraire des fragments de taille inférieure à 600 pb. Il faut alors utiliser la méthode de délétion-mutation à l'exonucléase III.

## 2. Méthode de délétion-mutation à l'exonucléase III.

Cette deuxième technique permet de générer des mutations par délétions unidirectionnelles (figure n°40).

Une cinétique d'hydrolyse par l'exonucléase III du fragment permet l'obtention de sous fragments dont la taille est vérifiée sur gel d'agarose. Le sous clonage de ces fragments, suivi du séquençage, permet de reconstituer la séquence de fragments de taille supérieure à 5Kb.

Le protocole a été décrit par Henikoff 1984 (74)

- L'ADN à séquencer doit être cloné dans un vecteur M13 ou autre dans un site qui laisse un maximum de sites entre l'insert et le site d'hybridation du primer.

- 5 µg d'ADN recombinant sont digérés par 2 enzymes de restriction entre l'insert et le site d'hybridation du primer (les deux enzymes utilisées ne doivent pas avoir de sites dans l'ADN à séquencer). Une enzyme produit une extrémité 5' sortante côté insert et l'autre une extrémité 3' sortante côté primer.

- L'ADN est purifié par un mélange phénol-chloroforme puis précipité.

- L'ADN est repris en solution TE. La concentration finale est de 0.2 µg/ml.

Après une préincubation de 5 minutes à 37°C en présence de tampon Exo III ( Tris HCl pH 8 ; 6,6 mM ; MgCl<sub>2</sub> 0,66 mM), une cinétique est réalisée avec 150 unités d'exonucléase III par pmole d'extrémité 3' enfouie.

- L'enzyme est inactivé par une incubation à 70°C 10 minutes.

- L'ADN monobrin de chaque prélèvement est hydrolysé par 20 unités de nucléase S1, 20 minutes à température ambiante. Après avoir vérifié la longueur de la délétion sur gel d'agarose, l'ADN est mis en présence de 5 unités de Klenow et 250 pmoles des 4 nucléotides dATP, dCTP, dGTP et dTTP 10 minutes 37°C.

- L'ADN est recircularisé par 2 U de ligase, en présence de PEG 7,5 %, 4 à 6 heures à 14°C puis cloné.

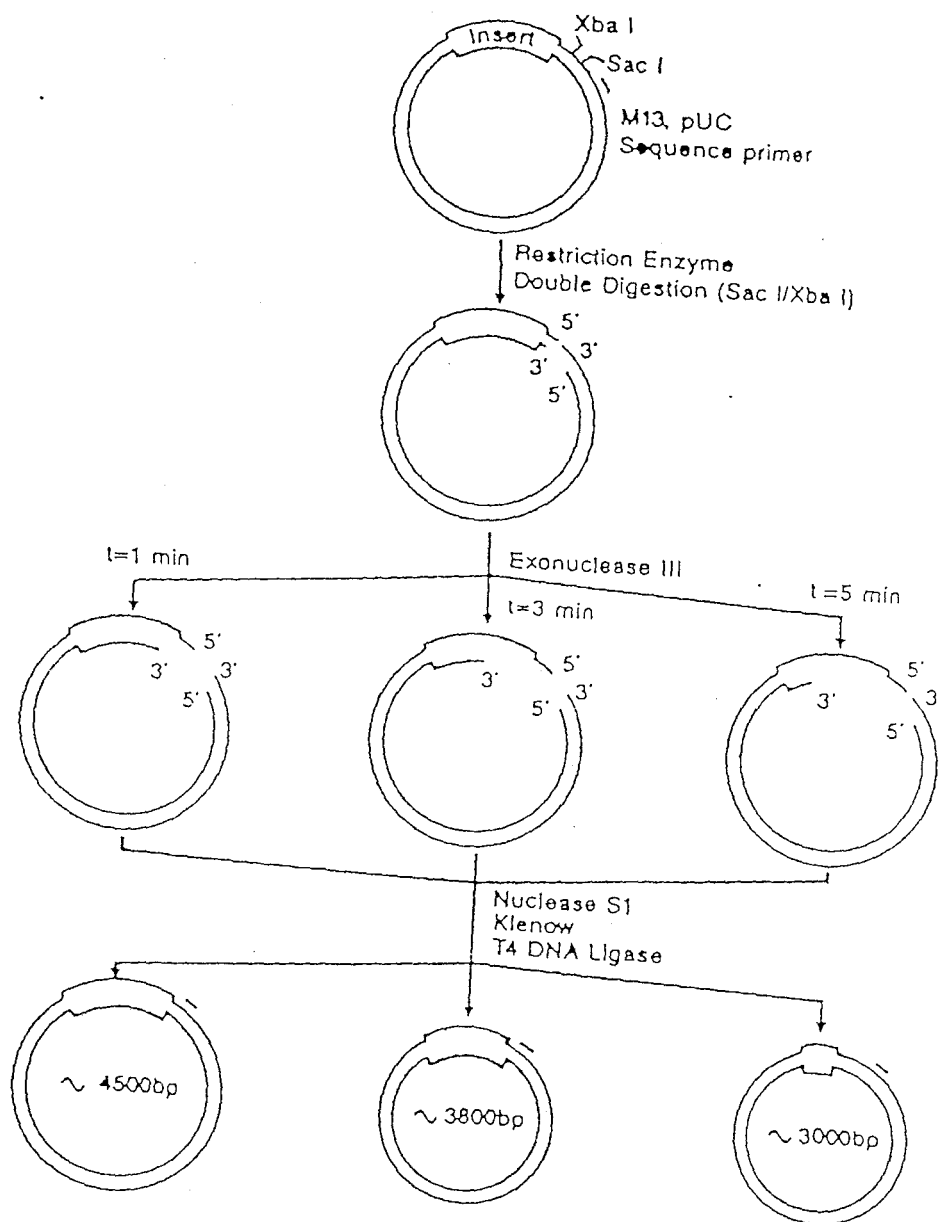


Figure n° 40 : Technique de mutation-délétion à l'exonucléase III

- Sur quelques clones, l'ADN est préparé puis hydrolysé de manière à libérer l'insert. La taille de celui-ci est visualisée sur gel d'agarose.

### **3. Transformation des bactéries E Coli JM109 et HB101**

Les trois souches bactériennes utilisées sont : HB101 pour le plasmide SK et JM109 et TG1 pour le phage M13. Deux techniques de transformation ont été utilisées :

#### **3.1. Technique de Hanahan (75)**

##### **3.1.1 La préparation des bactéries**

La veille de la transformation, les bactéries souches sont relancées deux fois en LB. Le lendemain, elles sont préparées comme suit :

- 0,5 ml de la relance est mis en culture dans 10 ml de milieu SOB, 2 heures à 37°C en agitation, puis placées dans la glace 15 minutes.

- Un aliquot de la culture (6 ml) est centrifugé à 3000 trs/mn pendant 10 minutes. Le culot est repris par 6 ml de TMN (Tris 20 mM pH 7,5, NaCl 0,5 M, Mg SO<sub>4</sub> 10 mM).

- La solution est centrifugée à 2500 trs/mn pendant 10 minutes. Le culot est repris par 6 ml de TFB (RbCl 100 mM, MnCl<sub>2</sub> 45 mM, CaCl<sub>2</sub> 50 mM, THC 3 mM, K-MES 10 mM) puis laissé dans la glace 15 minutes.

- La solution est centrifugée à 2500 trs/mn pendant 10 minutes. Le culot est repris par 300 µl de TFB. A la solution, sont ajoutés :

- 10 µl de DMSO, la solution est laissée 10 minutes dans la glace.

- 10 µl de DTT : 10 minutes dans la glace

- 10 µl de DMSO : 10 minutes dans la glace.

Les bactéries sont compétentes pour la transformation.

##### **3.1.2 La transformation**

Le produit de ligation (2µl) est mis en contact avec les bactéries compétentes (50µl) pendant 30 minutes dans la glace puis il subit un choc thermique à 42° C pendant 1 minute 30 secondes très précises ; 250µl de SOC sont ajoutés au milieu.

### 3.2 Technique de Chung (76)

Récemment utilisée, cette technique est plus rapide que la précédente, mais donne de moins bons rendements :

#### 3.2.1 La préparation des bactéries

La préparation des bactéries est réalisée par contact de la souche (relance 4 heures en 2XTY) avec un tampon TSS (LB contenant 10 % de PEG, 5 % de DMSO, un sel de magnésium 50mM).

#### 3.2.2 La transformation

La transformation se fait par contact avec l'ADN dans la glace, le temps n'est pas crucial et le choc thermique à 42° C n'est pas nécessaire.

Les techniques divergent alors en fonction des bactéries transformées :

- **dans le cas de la transformation de la souche HB 101** par le plasmide pBR 322, le milieu est incubé 45 minutes à 37°C puis étalé sur des boîtes d'Agar-ampicilline 25 µg/ml.

- Le lendemain, les clones positifs contrôlés par hybridation avec une sonde nucléique, sont mis en culture une nuit à 37°C en milieu LB Ampicilline 25 µg/ml.

- 1,4 ml de la culture est centrifugé 5 minutes à 10000 trs/mn. Le culot est repris par 300 µl de tampon Boiling (Sucrose 0,8 %, Triton 0,5 %, EDTA 50 mM pH 8,0 Tris pH 8,0) et 50 µl de lysozyme (10 mg/ml).

- Le tout est mis à ébullition 40 secondes puis centrifugé 15 minutes à 10000 trs/mn.

- Au surnageant est ajouté 5 µl de RNase A (Solution à 10 mg/ml), le milieu est incubé 20 minutes à 37°C.

- 5 µl de Protéinase K (solution 25 mg/ml) est ajouté au milieu. Celui-ci est incubé 30 minutes à 37°C.

- Les protéines sont éliminées par un volume de phénol/chloroforme (V/V). Centrifugation 10 minutes à 10000trs/mn.

- La phase supérieure est récupérée et l'ADN précipité par 1/10ème de volume en acétate de sodium 2M pH 5,5 et 2,5 volumes d'éthanol. Après 20 minutes de centrifugation à 10000 trs/mn, le culot est rincé par 500 µl d'éthanol à 70 % puis centrifugé 15 minutes à 10000trs/mn.

- Le culot est séché puis repris par 25 µl de TE.

- Pour être séquencé, l'ADN devra être dénaturé comme suit :

- A 12 µl de la solution d'ADN est ajouté 4 µl d'une solution de NaOH N et 4 µl d'une solution d'EDTA M. Le milieu est incubé 30 minutes à 37°C.

- L'ADN est précipité par 2 µl d'acétate de sodium 3M pH 5,5 et 100 µl d'éthanol à -20°C. Après 15 minutes à -80°C, le milieu est centrifugé 30 minutes à 10000 trs/mn.

- Le culot est rincé par 100 µl d'éthanol à 70 %, puis centrifugé 15 minutes à 10000 trs/mn. Il est séché et repris par un petit volume de TE. L'ADN est prêt à être séquencé.

- **dans le cas de la transformation de la souche TG1 ou JM109** par l'ADN du bactériophage M13, des aliquots de 10 à 100 µl de phages recombinants sont dilués par 100 µl de bactéries-souche ; 3 ml de Top Agar 2XTY ainsi que 40 µl de X-Gal (à 20 mg/ml dans du DMF) et 40 µl d'IPTG (à 48 mg/ml d'eau) sont ajoutés avant étalement sur boîte de milieu 2XTY. Après incubation d'une nuit à 37°C, les clones recombinants apparaissent (plages blanches pour M13, colonies pour pBR 322).

- Le lendemain, un certain nombre de clones recombinants correspondant à chaque préparation sont sélectionnés (**sonde spécifique**) et amplifiés.

- La souche bactérienne TG<sub>1</sub> est mise en culture une nuit à 37°C et une série de dilutions de cette bactérie est mise en contact avec un certain nombre de bactériophages repiqués au cure-dent sur le milieu solide.

- Le contact bactéries-bactériophages a lieu 5 à 7 heures à 37°C et sous agitation douce.

- Le mélange est centrifugé 10 minutes à 4200 trs/mn.

- Chacun des surnageants est mélangé à 300 µl de NaCl 2,5 M, PEG à 20 %. Il est mélangé par retournements et laissé 30 minutes à température du laboratoire.

- Il subit une centrifugation de quelques secondes à 12000 trs/mn à l'issue de laquelle le surnageant est éliminé (trompe à vide).

- Une alternance de centrifugations de quelques secondes et d'aspirations permet d'éliminer le surnageant.

- L'ultime étape pour enlever toute trace liquide est l'utilisation d'un tube capillaire.

- Chaque culot est dissous en tampon TE (100 µl). 50 µl de phénol est ajouté, le milieu est agité au vortex et laissé au repos 5 minutes. Le déphasage est obtenu par 5 minutes de centrifugation.

- La phase supérieure est récupérée, l'ADN est précipité par 1/10ème de volume d'acétate de sodium 2M pH 5,5 et 2 volumes d'éthanol à -20°C. Après une centrifugation 30 minutes à 12000 trs/mn, le culot est repris par 500 µl d'éthanol à 70 % et centrifugé 10 minutes à 10000 trs/mn.

- Le culot est séché puis repris par une faible quantité de TE X. L'ADN est prêt à être séquencé.

## LOCALISATIONS CHROMOSOMIQUES

Les sondes nucléotidiques réalisées à partir de chacun de nos clones ont permis d'établir les localisations chromosomiques de chaque insert.

La localisation chromosomique est réalisée en deux temps : d'abord la détermination du chromosome concerné puis une localisation plus précise de la région concernée sur le chromosome par hybridation *in situ*.

On extrait l'ADN de différentes lignées hybrides de hamsters ou de souris possédant plusieurs chromosomes humains.

Les ADN sont hydrolysés par une enzyme de restriction, ici Bam HI; après migration électrophorétique, le matériel est transféré sur membrane puis fixé. Après hybridation avec une sonde nucléique, lavage et révélation de l'autoradiogramme, une localisation précise est possible par recoupement de l'ensemble des signaux révélés. La première partie de ce travail est réalisée dans notre laboratoire. L'hybridation *in situ* est réalisée par le docteur N'Guyen Van Cong (IGR Villejuif).

## PREPARATION DES ARN ET NORTHERN BLOT

### 1. Homogénéisation du matériel biologique

Au moment de l'utilisation, la muqueuse (trachéale) humaine est déposée dans un mortier contenant de l'azote liquide et réduite en poudre. Cinq volumes de tampon GT sont alors ajoutés (isothiocyanate de guanidine 6M, citrate de sodium

5mM pH 7,0 ,  $\beta$ -mercaptoéthanol 0,1M, sarcosyl 0,5%). Afin de diminuer la viscosité, L'ADN est cassé par des passages successifs au travers d'aiguilles de diamètres décroissants.

## 2. Isolement des ARN cellulaires totaux

16 g de chlorure de césium solide sont ajoutés à l'homogénat et son volume est ajusté à 40 ml en tampon GT. Il est réparti dans 6 tubes, déposé délicatement au dessus de 3,2 ml d'une solution de chlorure de césium 5,7M, EDTA 0,1M (pH 7,5) DEPC 0,1%. Les tubes sont centrifugés dans une ultra-centrifugeuse Beckman L8 (Rotor : SW41) à 29500 tr/mn pendant 17 heures à 18°C. Chacun des culots est dissous dans 1 ml de TE ; SDS 1% après élimination délicate des surnageants. Les fractions sont regroupées par 3.

Les ARN sont alors extraits par 2 volumes de chloroforme n-butanol (4 : 1). La phase organique est réextraite par 1 volume de TE, SDS 1%. Les 2 phases aqueuses sont ensuite mélangées et réparties en 2 volumes égaux. Les ARN sont précipités par 0,1 volume d'acétate de sodium 3M (pH 5,2) et 2,2 volumes d'éthanol absolu, puis conservés à -80°C. Les ARN sont récupérés après centrifugation à 10000 tr/min pendant 30 min (Rotor JS13.1).

Les précipités obtenus sont ensuite lavés successivement par 5 ml d'éthanol, NaCl 0,2M (70 : 30) puis trois fois par 5 ml d'éthanol 95% et enfin par 5ml d'éthanol absolu. Chacun des culots est séché sous courant d'azote puis résolubilisé dans 500  $\mu$ l d'eau stérile. Un aliquot de 10  $\mu$ l de chaque échantillon est prélevé et dilué dans 1 ml d'eau pour la quantification.

La pureté de chaque échantillon est appréciée par deux critères :

- le rapport D. O. à 260 nm sur D. O. à 230 nm (la contamination par les solvants). Il doit être supérieur à 2.

- le rapport D. O. à 260 nm sur D. O. à 280 nm ( la contamination par les protéines). Il doit être supérieur à 1,75.



### 3. Electrophorèse d'ARN en milieu dénaturant

L'étude en Northern blot a pour but de visualiser un ARNm. Cette visualisation permet de déterminer la taille de cet ARNm, de mettre, éventuellement en évidence, l'existence de plusieurs messagers codés par un même gène.

L'ARN est dénaturé selon un protocole dérivé de LEHRACH et coll (1977) (77) par chauffage à 68° C en milieu dénaturant dont la composition est la suivante:

- formamide désionisée sur AG 501 X 8	500 µl
- formaldéhyde désionisée sur AG 501 X 8	178 µl
- MOPS 10 X	100 µl
- eau stérile DEPC QSP	1 000 µl

Des aliquots équivalents à 25 µg et 50 µg de chaque préparation d'ARN sont déposés sur un gel d'agarose à 0,9 % ainsi préparé :

- agarose	1,2 g
- MOPS 10X	15 ml
- eau stérile DEPC	109 ml

La migration est réalisée une nuit à température ambiante, sous agitation magnétique, à 30 V en tampon MOPS X préparé par dilution d'une solution mère de MOPS 10X :

- MOPS (acide-3-N-morpholino propane sulfonique)	10 g
- acétate de sodium 3M	4,2 ml
- EDTA 0,5 M	5 ml
- eau stérile traitée DEPC QSP	250 ml
après ajustement à pH 7,0 par la soude 3 M	

Après lavage par 6 bains successifs de tampon SSC 10X pendant 10 mn, sous agitation douce, destinés à éliminer le formaldéhyde, l'intégrité des ARN est contrôlée par visualisation sous lampe UV des bandes d'ARN ribosomales colorées par le bromure d'éthidium (BET) (solution aqueuse à 10 mg/ml) :

ARNr 18 S	2,06kb
ARNr 28 S	5,30 kb

## **"Northern blot" et hybridation**

La réalisation du "Northern blot" comporte, outre l'étape de fractionnement électrophorétique de l'ARN total en milieu dénaturant précédemment décrit, une étape de transfert et de fixation sur membrane de nitrocellulose.

La masse moléculaire de chacun des ARN ultérieurement révélés par la sonde sera estimée grâce à une courbe de calibration (linéaire sur papier semi-logarithmique) établie par mesure des distances de migration de chacune des bandes colorées au bromure d'éthidium d'un standard de masse moléculaire (BRL 9 bandes s'échelonnant de 0,24 à 9,5 kb)

Le transfert des bandes isolées sur le gel d'agarose est réalisé classiquement (MANIATIS et coll, 1982) (78) par mouvement de capillarité d'un tampon SSC 20X (citrate trisodique 2H<sub>2</sub>O : 0,3M, NaCl : 3M, ajusté à pH 7,0) au travers d'une membrane de type Hybond C Extra (Amersham) constituée d'une matrice de nylon revêtue de nitrocellulose. La fixation covalente de l'ARN sur le support est assurée de cuisson de la membrane à 80°C pendant 2 heures sous vide partiel.

## **ANALYSE PAR LA TECHNIQUE DU SOUTHERN BLOTTING**

L'étude en Southern blot a pour but de visualiser un fragment de génome. La visualisation des fragments génomiques, correspondant à chacun de nos inserts, a pour but de nous renseigner sur l'organisation génomique de nos inserts ainsi que de mettre en évidence d'éventuels polymorphismes de restrictions exoniques (R.F.L.P.).

### **1 - PREPARATION DE L'ADN GENOMIQUE**

#### **1.1 Extraction de l'ADN génomique**

L'ADN est extrait selon la technique décrite par M. JEANPIERRE (79) l'avantage de cette technique, par rapport aux techniques classiques, réside dans sa simplicité et sa rapidité évitant les phases phénoliques, au profit d'une extraction au chlorure de guanidium.

### 1.1.1 Lyse cellulaire

A partir de sang congelé prélevé sur anticoagulant, le sang est décongelé rapidement à 37°C, puis réparti en aliquot de 10 ml dans des tubes de 50 ml.

Une lyse cellulaire est effectuée à l'aide de 40 ml d'une solution Tris 10 mM pH 7,4 ; EDTA 10 mM pendant 10 minutes.

Une centrifugation de 10 minutes à 3000 g permet d'éliminer avec le surnageant, l'hémoglobine et les autres constituants solubles. L'opération de lavage est répétée plusieurs fois (5 à 6 fois) jusqu'à l'obtention d'un culot de débris cellulaires et de noyaux pratiquement blanc.

### 1.1.2 Protéolyse

Au culot sont ajoutés 14 ml d'hydrochlorure de guanidium 6 M filtré qui est un dénaturant protéique.

Le culot est homogénéisé dans cette solution en agitant de façon ménagée, par retournement dans un turbula (30 mn), le vortex est déconseillé car trop énergique.

Puis 2 ml de sarcosyl à 10 % qui est un détergent anionique et, 150 µl de protéinase K à 10 mg/ml (préparé extemporanément) sont ajoutés à chaque préparation et incubés une heure à 60°C ou une nuit à 37°C selon la commodité.

### 1.1.3 Précipitation de l'ADN

L'ADN est enfin précipité par de l'éthanol absolu froid en complétant à 50 ml chaque tube, auquel a préalablement été ajouté 1 ml d'acétate d'ammonium 7,5 M.

L'ADN est recueilli sur une baguette de verre puis, rincé par 4 fois à l'éthanol à 70% (V/V) et resuspendu dans du TE (4 ml pour un tube de 10 ml de sang). Dissolution 48 h (turbula).

### 1.1.4 Quantification et qualité

La concentration d'ADN extrait est calculée après lecture spectrophotométrique de la densité optique (DO) dans une cuve en quartz de 1 cm d'épaisseur à 260 nm et à 280 nm. Sa pureté est appréciée par le rapport 260/280 qui doit être voisin de 2.

Une DO de 1 correspond à une concentration de 50 µg d'ADN par ml. Les quantités d'ADN obtenues sont généralement de 20 mg/ml.

La qualité de l'ADN peut être contrôlée par migration électrophorétique sur un gel d'agarose à 0,8 % dans un tampon TBE (Tris 0,09 M, borate de sodium 0,09 M pH 8,2, EDTA 0,002 M) qui permet une migration rapide à fort voltage, 2 V/cm pendant 1 h 30 environ.

Dans le gel est ajouté 0,5 mg/ml de bromure d'éthidium (BEt) qui est un agent intercalant permettant de visualiser l'ADN aux ultraviolets.

L'ADN de bonne qualité doit migrer sous la forme d'une bande d'environ 50 kb ou plus, la présence d'une traînée signifie que l'ADN est dégradé.

Les solutions d'ADN sont stockées à 4°C.

## 1.2. Digestion de l'ADN génomique

La digestion enzymatique se fait dans les conditions optimales indiquées par le fabricant. On digère de 10 à 40 ng d'ADN génomique par expérience dans un volume final de 50 µl, contenant 1/10 de tampon adéquat (10 X) et 2,5 à 10 U/ng d'enzyme au total par ng d'ADN.

Le temps de coupure est de 24 heures avec addition d'enzyme au bout de 6 heures.

## **2 - ANALYSE EN SOUTHERN BLOT DE L'ADN HUMAIN (figure n°41)**

### **2.1. Hydrolyse**

L'ADN est hydrolysé par les enzymes Pst I, Xba I, Hind III, Bam HI et Eco RI. D'autres enzymes possédant un site dans la séquence JER 57 ont été utilisés puis délaissés à cause d'hydrolyses partielles de l'ADN. 40 µg d'ADN ont été hydrolysés pour chaque enzyme (5U/mg) dans un tampon adéquat à 37°C quelques heures.

### **2.2. Electrophorèse**

Les fragments d'ADN hydrolysés sont séparés par une électrophorèse en gel d'agarose (1 % en tampon : Phosphate ( $\text{Na}_2 \text{HPO}_4$  6,8 mM,  $\text{Na H}_2 \text{PO}_4$  3,2 mM pH 6,8)) contenant du BEt (0,5mg/ml) une nuit à 20 volts.

## **3. TRANSFERT ET FIXATION**

La réalisation du Southern blot comporte une étape préalable de dénaturation de l'ADN. Pour cela, le gel est trempé 2x30 minutes dans du tampon de dénaturation (1,5 M NaCl; 0,5 M NaOH). puis il est rincé par de l'eau desionisée. Il est enfin neutralisé par deux bains de 30 minutes dans un tampon de neutralisation (0,5 M Tris HCl pH 7,5; 3 M NaCl). Le transfert est effectué par capillarité pendant une nuit, en utilisant du tampon SSC 6X (80). L'ADN est fixée de façon covalente par une irradiation UV à 330 nm 3 minutes.

## **4. HYBRIDATION A UNE SONDE NUCLEIQUE**

La membrane est préhybridée dans un tampon de préhybridation (SSC 6X, Denhardt's 5X, SDS 0,5%) 2 heures à 65°C. Le tampon de préhybridation est alors remplacé par un tampon d'hybridation (SSC 6X, Denhardt's 5X, SDS 0,5%, sperme de saumon, sulfate de dextran 10%) et de la sonde nucléotidique.

Après une nuit à 65°C, la membrane est lavée comme suit:

- Un rinçage en SSC 3X
- Un lavage en SSC 0,1X, SDS 0,1%, 10 minutes à 65°C
- Un rinçage en SSC 3X

La membrane est époncée puis mise en autoradiographie.

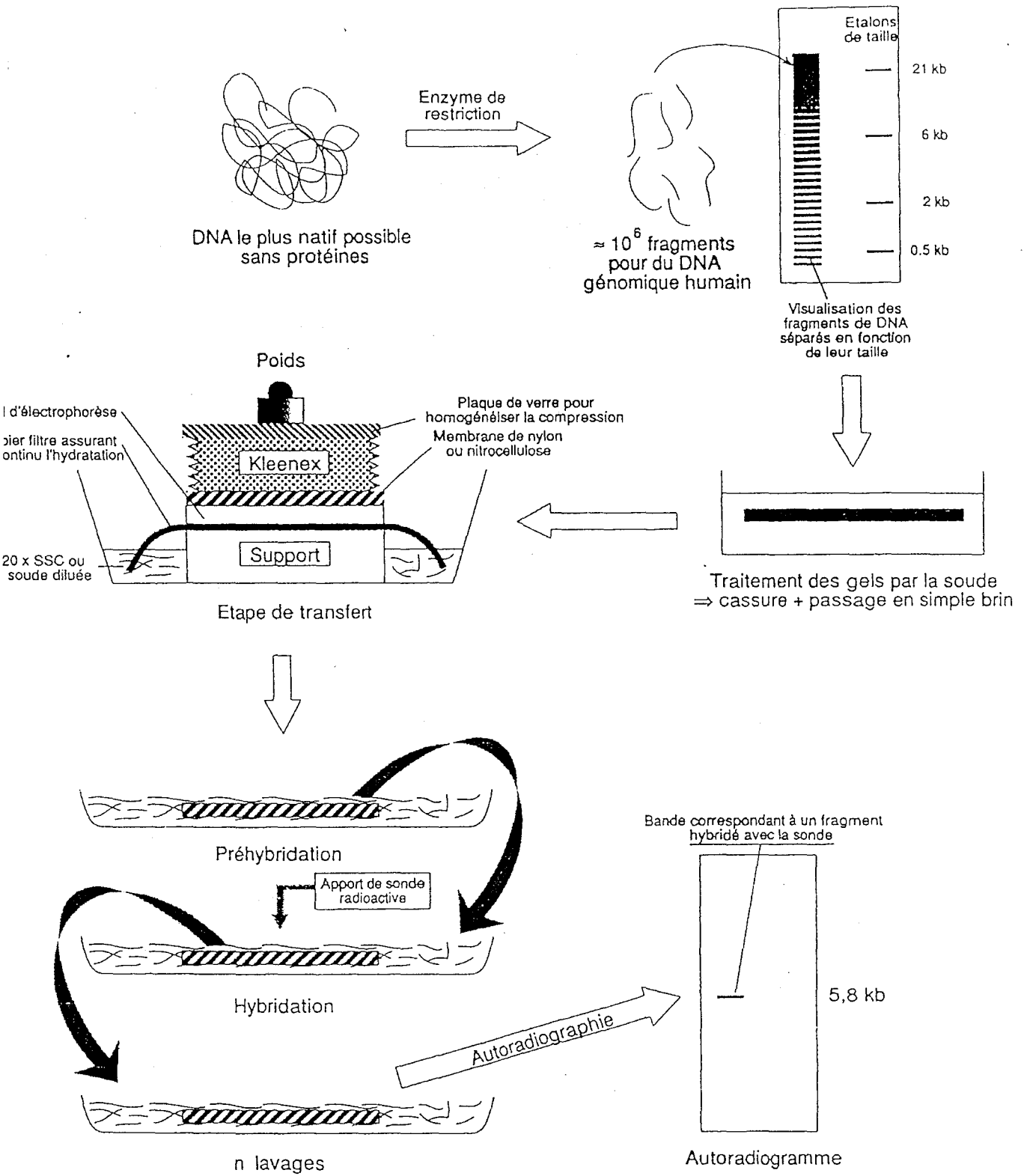


Figure n° 41 : Technique de Southern blot

## BIBLIOGRAPHIE

1. SACHDEV G.P., MYERS F.J., HORTON F.O., FOX O.F., WENG G., ROGERS R.M. (1980)  
Isolation, chemical composition and properties of the major mucin component of normal human tracheobronchial secretions.  
*Biochem. Med.* 24, 82-94.
2. SNYDER C.E., NADZIEJKO C.E., HERP A. (1982)  
Binding of basic proteins to glycoproteins in human bronchial secretions.  
*Int. J. Biochem.* 14, 895-898.
3. LAFITTE J.J., LAMBLIN G., LHERMITTE M., HUMBERT P., DEGAND P., ROUSSEL P., (1977)  
Etude des glycoprotéines bronchiques humaines de type mucinique obtenues par lavage de bronches macroscopiquement saines.  
*Carbohydr. Res.* 56, 383-389.
4. KOTLAR H.K., HARBITZ O., JENSSEN A.O., SMIDSROD O. (1980)  
Quantitation of proteins from patients with chronic obstructive lung disease.  
II-Determination of albumin, transferrin, 1 acid glycoprotein, IgG, IgM, Lysozyme and C<sub>3</sub> Complement factor.  
*Eur. J. Resp. Dis.* 61, 233-239.
5. HARBITZ O., JENSSEN A.O., SMIDSROD O. (1980)  
Quantitation of proteins in sputum from Patients with chronic obstructive lung disease.  
I-Determination of Immunoglobulin A.  
*Eur. J. Resp. Dis.* 61, 84-94.
6. STOCKLEY R.A., BURNETT D. (1980)  
Alpha1-antichymotrypsin in infected and non-infected sputum.  
*Am. Rev. Resp. Dis.* 122, 81-88.
7. BAILLEUL V., RICHET C., HAYEM A., DEGAND P. (1977)  
Propriétés rhéologiques des sécrétions bronchiques : mise en évidence et rôle de polypeptides riches en proline (PRP).  
*Clin. Chim. Acta.* 74, 115-123.
8. PICOT R., DAS I., REID L. (1978)  
Pus, deoxyribonucleic acid and sputum viscosity  
*Thorax* 33, 235-242.
9. HOUDRET N., PERINI J.M., GALABERT C., SCHARFMAN A., HUMBERT P., LAMBLIN G., ROUSSEL P. (1986)  
The high lipid content of respiratory mucin in cystic fibrosis is related to infection.  
*Biochim. Biophys. Acta* 880, 54-61.

10. SLOMIANY A., WITAS H., AONO M., SLOMIANY B.L. (1983)  
Covalently linked fatty acids in gastric mucus glycoprotein of cystic fibrosis patients.  
*J. Biol. Chem.* 258, 8535-8538.
11. WILLIAMS S.E., TURNBERG L.A. (1980)  
Retardation of acid diffusion by pig gastric mucus : a potential role in mucosal protection.  
*Gastroenterology*, 79, 299-304.
12. WOLF D.P., SOKOLOSKI J.E., LITT M. (1980)  
Composition and function of human cervical mucus.  
*Biochim. Biophys. Acta.* 630, 543-558.
13. KORNFELD R., KORNFELD S. (1985)  
Assembly of asparagine-linked oligosaccharides.  
*Annu. Rev. Biochem.* 54, 631-664.
14. STROUS G.J. A.M. (1979)  
Initial glycosylation of proteins with acetyl galactosaminyl serine linkages  
*Proc. Natl. Acad. Sci.* 76, 2694-2698.
15. ABEIJON C., HISCHBERG C.B. (1987)  
Subcellular site of synthesis of the N-Acetyl galactosamine ( $\alpha$ 1-0) serine (or threonine) linkage in rat liver.  
*J. Biol. Chem.* 262, 4153-4159.
16. SCHACHTER H., BROCKHAUSEN I. (1989)  
The biosynthesis of branched O-Glycans in mucus and related topics.  
Chantler and Ratcliffe N.A. Eds ; Symposia of the Society for Experimental Biology p1-26.
17. DEKKER J., ENDE A.V.D., AELMANS P.H., STROUS G.J.A.M. (1991)  
Rat gastric mucin is synthesized and secreted exclusively as filamentous oligomers  
*Biochem. J.*, 279, 251-256.
18. MANTLE M., MANTLE D., ALLEN A. (1981)  
Polymeric structure of pig small intestinal mucus glycoprotein.  
*Biochem. J.* 195, 277-285.
19. MANTLE M. (1991)  
Effects of hydrogen peroxide, mild trypsin digestion and partial reduction on rat intestinal mucin and it's disulphide-bound 118kDa glycoprotein  
*Biochem. J.* 274, 679-685.
20. HOUDRET N., LETREUT A., LHERMITTE M., LAMBLIN G., DEGAND P., ROUSSEL P. (1981)  
Comparative action of reducing agents on fibrillar human bronchial mucus under dissociating and non dissociating conditions.  
*Biochim. Biophys. Acta* 668, 413-419.



- 21 VIRMANI A.K., SHANKAR V., GILMORE M.S., GRAVES D.C., SACHDEV G.P. (1991)  
Translation of messenger RNA from canine tracheal epithelial cells : identification of mucin core protein  
*Am. J. Respir. Cell. Mol. Biol.* 5, 149-154.
- 22 THORNTON D.J., DAVIES J.R., KRAAYENBRINK M., RICHARDSON P.S., SHEEHAN J.K., CARLSTEDT I. (1990)  
Mucus glycoproteins from "normal" human tracheobronchial secretion.  
*Biochem. J.* 265, 179-186.
- 23 SHEEHAN J.K., CARLSTEDT I. (1987)  
Size heterogeneity of human cervical-mucus glycoproteins. Studies performed with rate-zonal centrifugation and laserlight scattering.  
*Biochem. J.* 245, 757-762.
- 24 MARSCHALL T., ALLEN A. (1978)  
Isolation and characterisation of the high molecular weight glycoproteins from pig colonic mucus.  
*Biochem. J.* 173, 569.
25. SHEEHAN J.K., BOOT-HANDFORD R.P., CHANTLER E., CARLSTEDT I., THORNTON D.J. (1991)  
Evidence for shared epitopes within the "naked" protein domains of human mucus glycoproteins  
*Biochem. J.* 274, 293-296.
- 26 CREETH J.M., BHASKAR K.R., HORTON J.R., DAS T., LOPEZ-VIDRIER M.T., REID L. (1977)  
The separation and characterization of bronchial glycoproteins by density-gradient methods.  
*Biochem. J.* 167, 557-569.
- 27 ROUSSEL P., LAMBLIN G., DEGAND P., WALKER-NASIR E., JEANLOZ R.W. (1975)  
Heterogeneity of the carbohydrate chains of sulfated bronchial glycoproteins isolated from a patient suffering from cystic fibrosis.  
*J. Biol. Chem.* 250, 2114-2122.
- 28 FELDHOFF P.A., BHAVANANDAN V.P., DAVIDSON E.A. (1979)  
Purification, properties and analysis of human asthmatic bronchial mucin.  
*Biochemistry* 18, 2430-2436.
- 29 LAMBLIN G., LHERMITTE M., DEGAND P., ROUSSEL P., SLAYTER H.J. (1979)  
Chemical and physical properties of human bronchial mucus glycoproteins.  
*Biochimie* 61, 23-43.

- 30 ROSE M.C., LYNN W.S., KAUFMAN B. (1979)  
Resolution of the major components of human lung mucosal gel and their capabilities for reaggregation and gel formation.  
*Biochemistry* 18, 4030-4037.
- 31 BHASKAR K.R., REID L. (1981)  
Application of density gradient methods for the study of mucus glycoprotein and other macromolecular components of the sol and gel phases of asthmatic sputa.  
*J. Biol. Chem.* 256, 7583-7589.
- 32 TARACHNIK N.F., BLACKBURN P., CERAMI A. (1981)  
Biochemical and rheological characterization of sputum mucins from a patient with cystic fibrosis.  
*J. Biol., Chem.* 256, 7161-7165.
- 33 SNYDER C.E., NADZIEJKA C.E., HERP A. (1982)  
Isolation of bronchial mucins from cystic fibrosis sputum by use of citraconic anhydride.  
*Carbonyd. Res.* 105, 87-93.
- 34 ROUSSEL P., LAMBLIN G., LHERMITTE M., HOUDRET N., LAFITTE J. J., PERINI J.M., KLEIN A., SCHARFMAN A. (1988)  
The complexity of mucins  
*Biochimie* 70, 1471-1482.
- 35 PODOLSKY D.K., ISSELBACHER K.J. (1984)  
Glycoprotein composition of colonic mucosa  
*Gastroenterology* 87, 991-998.
- 36 TORIBARA N.W., GUM J.R., CULHANE P.J., LAGACE R.E., HICKS J.W., PETERSEN G.M., KIM Y.S. (1991)  
MUC2 Human small intestinal mucin gene structure.  
*J. Clin. Invest.* 88, 1005-1013.
- 37 KIM Y.S., GUM J.R., BYRD J.C., TORIBARA N.W. (1991)  
The structure of human intestinal apomucins.  
*Am. Rev. Respir. Dis.* 144, S10-S14.
- 38 ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOUMADJE A., JOHNSON N.C. Jr., HILL R.L. (1987)  
Structural properties of porcine submaxillary gland apomucin.  
*J. Biol. Chem.* 262, 11339-11344.
- 39 CARLSTEDT I., LINDGREN H., SHEEHAN J.K. (1983)  
The macromolecular structure of human cervical mucus glycoproteins.  
*Biochem. J.* 213, 427-435.
- 40 ALLEN A., PAIN R.H., ROBSON T.R. (1976)  
Model for the structure of the gastric mucous gel.  
*Nature* 264, 88-89.

41. ALLEN A. (1978)  
Structure of gastrointestinal mucus glycoproteins and the viscous and gel forming properties of mucus  
Brit. Med. Bull. 34, 28-33.
42. 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.
43. XU G., HUAN L.J., KHATRI I., WANG D., BENNIC 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.
44. XU G., HUAN L.J., KHATRI I., SAJJAN U.S., Mc COOL D., WANG D., JONES C., FORSTNER G., FORSTNER J. (1992)  
Human intestinal Mucin-Like Protein (M.L.P.) is homologous with rat MLP in the Cterminal region, and is encoded by a gene on chromosome 11 p 15.5.  
Biochem. Biophys. Res. Com. 183, 821-828.
45. SLAYTER 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.
46. ROSE M.C., VOTER W.A., BROWN C.F., HAUFMAN B. (1984)  
Structural features of human tracheo-bronchial mucin glycoproteins.  
Biochem. J. 222, 371-377.
47. MIKKELSEN A., STOKKE B.T., CHRISTENSSEN B.E., EKGSAETER A. (1985)  
Flexibility and length of human bronchial mucin studied using low-shear viscometry, birefringence, relaxation analysis and electron microscopy.  
Biopolymers, 24, 1683-1704.
48. SHEEHAN J.K, OATES K., CARLSTEDT I. (1986)  
Electron microscopy of cervical, gastric and bronchial mucus glycoproteins.  
Biochem. J., 239, 147-153.
49. SMITH B.F., LAMONT J.F. (1984)  
Hydrophobic binding properties of bovine gallbladder mucin.  
J. Biol. Chem. 259, 12170-12177.
50. WOODWARD H., HORSEY B., BHAVANANDAN V.P., DAVIDSON E.A. (1982)  
Isolation, purification and properties of respiratory mucus glycoproteins.  
Biochemistry 21, 694-701.

51. SHEEHAN J.K., CARLSTEDT I. (1990)  
Electron microscopy of cervical mucus glycoproteins and fragments therefrom.  
*Biochem. J.* 265, 169-178.
52. BARA J., GAUTIER R., LE PENDU J., ORIOL R. (1988)  
Immunochemical characterization of mucins. Polypeptide (M<sub>1</sub>) and polysaccharide (A and Le<sup>b</sup>) antigens.  
*Biochem. J.* 254, 185-193.
53. DEGAND P., HAVEZ R., BISERTE G. (1972)  
Séquence en acides aminés au voisinage du point d'attache d'un glycopeptide de mucine bronchique humaine.  
*C.R. Acad. Sci.* 275, 491-493.
54. HILL H.D., SCHWYZER M., STEINMAN H.M., HILL R.L. (1977)  
Ovine submaxillary mucin primary structure and peptide substrates of UDP-N-Acetyl galactosamine mucine transferase.  
*J. Biol. Chem.* 252, 3799-3804.
55. ROSE M.C., KAUFMAN B., MARTIN B. (1989)  
Proteolytic fragmentation and peptide mapping of human carboxyamino methylated tracheobronchial mucin.  
*J. Biol. Chem.* 264, 8193-8199.
56. TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L. (1988)  
Porcine submaxillary gland apomucin contains tandemly repeated identical sequences of 81 residues.  
*J. Biol. Chem.* 263, 1081-1088.
57. GENDLER S.J., BURCHELL J.M., DUHIG T., LAMPORT D., WHITE R., PARKER M., TAYLOR PAPADIMITRIOU J. (1987)  
Cloning of partial cDNA encoding differentialion and tumor associated mucin glycoproteins expressed by human mammary epithelium.  
*Proc. Natl. Acad. Sci. USA*, 84, 6060-6066.
58. WRESCHNER D.H., HAREUVENI M., TSARFATY I. SMORODINSKY N., HOREV J., ZARETSKY J., KOTKES P., WEISS M., LATHE R., DION A., KEYDAR J. (1990)  
Human epithelial tumor antigen cDNA sequences.  
*Eur. J. Biochem.* 189, 463-473.
59. SIDDIQUI J., ABE M., HAYES D., SHANI E., YUNIS E., KUFÉ D. (1988)  
Isolation and sequencing of a cDNA coding for the human DF<sub>3</sub> breast carcinoma associated antigen.  
*Proc. Natl. Acad. Sci. USA*, 85, 2320-2323.
60. GUM J.R., BYRD J.C., HICKS J.W., TORIBARA N.W., LAMPORT D.T.A., KIM Y.S. (1989)  
Molecular cloning of human intestinal mucin cDNAs  
*J. Biol. Chem.*, 264, 6480-6487.

61. 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. Commu.* 171, 407-415.
62. NGUYEN V.C., AUBERT J.P., GROSS M.S., PORCHET N., DEGAND P., FREZAL J. (1990)  
Assignment of human tracheo bronchial mucin gene(s) to 11p15 and a tracheobronchial mucin-related sequence to chromosome 13.  
*Hum. Genet.*, 86, 167-172.
63. AUBERT J.P., PORCHET N., CREPIN 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.
64. PORCHET N., NGUYEN V.C., DUFOSSÉ J., AUDIE J.P., GUYONNET-DUPERAT V., GROSS M.S., DENIS C., DEGAND P., BERNHEIM A., AUBERT J.P. (1991)  
Molecular cloning and chromosomal localisation of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs.  
*Biochem. Biophys. Res. Commun.* 175, 414-422.
65. VAN CONG N., AUBERT J.P., GROSS M.S., PORCHET N., DEGAND P., FREZAL 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.
66. CREPIN M., PORCHET N., AUBERT J.P., DEGAND P. (1990)  
Diversity of the peptide moiety of human mucins.  
*Biorheology* 27, 471-484.
67. SWALLOW D.M., GENDLER S., GRIFFITH B., KEARNEY A., POVEY S., SHEER D., TAYLOR PAPADIMITRIOU J. (1987)  
The hypervariable gene locus PUM, which codes for the tumor associated epithelial mucins, is located on chromosome 1, within the region 1q 21-24  
*Ann. Hum. Genet.* 51, 289.
68. PERRINI J.M., VANDAMME-CUBADDA N., AUBERT J.P., PORCHET N., MAZZUCA M., LAMBLIN G., HERSCOVICS A., ROUSSEL P. (1991)  
Multiple apomucin translation products from human respiratory mucosa mRNA  
*Eur. J. Biochem* 196, 321-328.
69. DEKKER J., VAN BEURDEN -LAMBERO W.M.O., STROUS G.J. (1989)  
Biosynthesis of gastric mucus glycoprotein of the rat.  
*J. Biol. Chem.* 264, 10431-10437.

- 70 VAITUKAITIS J.L. (1981)  
Production of antisera with small doses of immunogen : multiple intradermal injections.  
Methods Enzymol. 73, 46-52.
- 71 YOUNG R.A., DAVIS R.W. (1983)  
Efficient isolation of genes by using antibody probes.  
Proc. Natl. Acad. Sci. USA 80, 1194-1198.
- 72 YOUNG R.A., DAVIS R.W. (1983)  
Yeast RNA polymerase II genes : isolation with antibody probes.  
Science 222, 778-782.
- 73 SANGER F., NICKLEN S., COULSON A.R. (1977)  
Sequencing with chain-terminating inhibitors terminating.  
Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 74 HENIKOFF S. (1984)  
Unidirectional digestion with exonuclease III. Creates targeted breakpoints for DNA sequencing.  
Gene 28, 351-359.
- 75 HANAHAN D. (1983)  
Studies on transformation of Escherichia Coli with plasmids  
J. Mol. Biol. 166, 557-580.
- 76 CHUNG C.T., NIEMELA S.L., MILLER R.H. (1989)  
One-step preparation of competent Escherichia Coli  
Proc. Natl. Acad. Sci. USA 86, 2172-2175.
- 77 LEHRACH H., DIAMOND D., WOZNEY J.M., BOEDTKER H. (1977)  
RNA molecular weight determination by gel electrophoresis under denaturing conditions, cortical reexamination.  
Biochemistry 16, 4743.
- 78 MANIATIS T., FRITSCH E.F., SAMBROOK J. (1982)  
Molecular cloning : a laboratory manual, Cold Spring Harbor Laboratory, New-York.
- 79 JEANPIERRE M. (1987)  
A rapid method for the purification of DNA from blood.  
Nucleic Acids Res. 15, 9611.
- 80 SOUTHERN E. (1975)  
Detection of specific sequences among DNA fragments separated by gel electrophoresis.  
J. Mol. Biol. 98, 503.
- 81 BHARGAVA A.K., WOITACH J.T., DAVIDSON E.A., BHAVANANDAN V.P. (1990)  
Cloning and cDNA sequence of a bovine submaxillary gland mucin-like protein containing two distinct domains.  
Proc. Natl. Acad. Sci. USA 87, 6798-6802.

