

50376
1993

50376
1993

340

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE I

340

Année 1993

THESE

N° d'ordre 1208

de DOCTORAT D'UNIVERSITE

pour l'obtention du grade de

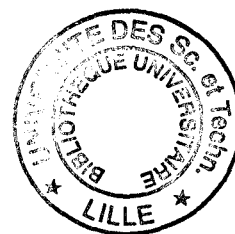
DOCTEUR EN SCIENCES DE LA VIE ET DE LA SANTE

présentée par

Véronique GUYONNET DUPERAT- BIANCHI

ETUDE DES GENES DE MUCINES HUMAINES LOCALISES EN 11P15

MEGACARTOGRAPHIE ET
APPROCHE DE L'ORGANISATION
GENOMIQUE DE MUC5B ET MUC5AC



Présentée le 26 novembre 1993 devant la commission d'examen

JURY

Président : Professeur André VERBERT
Rapporteur: Docteur Dallas SWALLOW
Docteur Alain ZWEIBAUM
Membres : Professeur René CACAN
Professeur Pierre DEGAND
Docteur Nicole PORCHET

Que tous ceux qui m'ont accordé leur amitié, leur soutien, leur aide et leurs conseils trouvent ici l'expression de ma profonde reconnaissance.

Accueillir un étudiant dans un laboratoire n'est déjà pas une chose facile, vous avez réussi le tour de force d'adopter à l'unité INSERM n°16 une bordelaise que vous ne connaissiez pas.

Le travail, la motivation et les aides précieuses que vous m'avez apportés ont fait de moi une lilloise de coeur et une scientifique passionnée. A tous un grand merci.

Monsieur le Professeur Pierre DEGAND

Vous m'avez permis de préparer au sein de votre laboratoire mon DEA puis mon Doctorat, je vous en suis très reconnaissante. Croyez à ma grande joie de voir l'unité recrée, l'équipe entière pouvant ainsi continuer sa recherche sur les mucines humaines.

Monsieur le Docteur Jean-Pierre AUBERT

Tu m'as confié un sujet de recherche dès mon arrivée pour mon stage de DEA. Tu as su, avec humour et sympathie me suivre, m'encourager et m'apporter une aide scientifique très efficace tout au long de ce travail. Je tiens à t'exprimer ma profonde gratitude.

Madame le Docteur Nicole PORCHET

Tu m'as encadrée au cours de mon DEA et de ma thèse avec un grand dévouement, tu as su me transmettre ta passion pour la recherche médicale, ton goût pour la recherche bibliographique si utile tout au long de mes travaux. Avec une réelle admiration, je te remercie pour tout ce que tu as fait.

Madame le Docteur Anne LAINE

Sans ta collaboration et ta disponibilité de chaque instant, il ne m'aurait pas été possible de mener à terme, en raison de mon éloignement géographique, cette étude des gènes de mucines humaines si délicate et laborieuse. Je te remercie vivement et te prie de croire à toute ma sympathie.

Je remercie **Danièle, Pascal, Viviane, Marie-Paule et Michel**, qui chacun dans leurs domaines m'ont apporté une aide technique précieuse et énergique.

Je remercie **Virginie, Marie-Pierre et Jean-Pierre** pour les bons moments de détente passés ensemble et l'ambiance toujours amicale qu'ils ont su maintenir dans le laboratoire.

Je remercie vivement monsieur le **Docteur Joseph HAN** de sa grande générosité qui m'a ainsi permis de réaliser quelques retours vers ma région natale.

Je remercie monsieur le **Docteur Jean-Pierre Kerkaert** et son équipe, plus particulièrement madame le **Docteur Sylvie Zouitina**, madame le **Docteur Maud Collyn** et monsieur **Claude Denis**, de leur efficace et aimable collaboration. Sans eux, l'étude par électrophorèse en champ pulsé n'aurait pu avoir lieu dans d'aussi bonnes conditions, je tiens donc à leur exprimer toute ma reconnaissance.

Je remercie également ceux qui ont participé à l'élaboration de ce mémoire :

Madame Marie-Thérèse MAILLARD et Madame Edith DELALEAU pour leur collaboration dactylographique.

Monsieur Claude VANDEPERRE pour la qualité de la documentation photographique.

Je tiens à remercier très sincèrement pour l'honneur qu'ils me font :

Monsieur le Professeur André VERBERT,

en acceptant de présider ce jury;

Madame le Docteur Dallas SWALLOW et

Monsieur le Docteur Alain ZWEIBAUM,

en acceptant d'être les rapporteurs de cette thèse;

Monsieur le Professeur René CACAN

Monsieur le Professeur Pierre DEGAND

Madame le Docteur Nicole PORCHET, mon tuteur de thèse

qui ont bien voulu juger ce travail.

Le travail présenté dans ce mémoire a été réalisé à l'Unité de Recherche n°16 de l'Institut National de la santé et de la Recherche Médicale, dirigé par Monsieur le Professeur Pierre DEGAND, sous la responsabilité scientifique de Madame le Docteur Nicole PORCHET.

Il a bénéficié de la collaboration de :

Monsieur le Docteur J. P. KERKAERT et son équipe (Unité INSERM 124, Lille),

Monsieur le Docteur Y. S. KIM et J. GUM (Gastrointestinal Research Laboratory de San Francisco),

Madame le Docteur G. MATTEI (Unité INSERM 242, Marseille),

Monsieur le Docteur P. COUILLIN (Unité INSERM 178, Villejuif).

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

PUBLICATIONS

PORCHET N., NGUYEN V.C., DUFOSSE J., AUDIE J.P., GUYONNET DUPERAT V., GROSS M.S., DENIS C., DEGAND P., BERNHEIM A., AUBERT J.P.

Molecular cloning and chromosomal localization of a novel human tracheobronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs.

Biochem. Biophys. Res. Commun., (1991), 175, : 414-422.

PORCHET N., DUFOSSE J., AUDIE J.P., GUYONNET DUPERAT V., PERINI J.M., NGUYEN V.C., DEGAND P., AUBERT J.P.

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

Amer. Rev. Respir. Dis., (1991), 144, : S15-S18.

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

Multiple chromosomal locations of human tracheobronchial mucin genes.

Eur. Resp. J., (1991), 4 : suppl. 382.

GROSS M.S., GUYONNET DUPERAT V., PORCHET N., BERNHEIM A., AUBERT J.P., NGUYEN V.C.

Mucin 4 (MUC-4) gene : regional assignment (3q29) and RFLP analysis.

Ann. Genet., (1992), 35 : 21-26.

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

Degenerate 87 base pair tandem repeats create hydrophilic/hydrophobic alternating domains in human peptide mucins mapped to 11p15.

Biochem. J., (1993), 293 : 329-337.

COMMUNICATIONS

GUYONNET DUPERAT V., LAINE A., AUBERT J.P., PORCHET N.

Les gènes de mucines localisés en 11p15 : approche de l'organisation génomique de MUC 5B et de MUC 5C.

Forum Jeunes Chercheurs, Caen, 7-10 juillet 1992.

Regard sur la Biochimie, 1992.

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

Gènes de mucines humaines, régulation de leur expression.

Association Nationale des Professeurs et Maîtres de Conférences Biochimistes des UER Médicales, Strasbourg, 5 et 6 octobre 1990.

GUYONNET DUPERAT V., AUDIE J.P., LAINE A., AUBERT J.P., PORCHET N.
Mucins encoded by MUC 4 and MUC5 : structural features and tissue expression.
2nd International Workshop on Carcinoma-Associated Mucins, 2-6 août 1992.

AUDIE J.P., DEBAILLEUL V., GUYONNET DUPERAT V., LAINE A., AUBERT J.P., PORCHET N.
Expression and structural features of human gastric mucins ascertained by cDNA cloning.
2nd International Workshop on Carcinoma-Associated Mucins, 2-6 août 1992.

GUYONNET DUPERAT V., AUDIE J.P., DEBAILLEUL V., LAINE A., BUISINE M.P., ZOUITINA S., JANIN A., DEGAND P., AUBERT J.P., PORCHET N.
Human mucin genes mapped to 11p15 : mucin cDNAs, MUC 5A and MUC5C belong to the same gene (MUC5AC) expressed in both human tracheobronchial and gastric mucosae.
Soumis.

Le travail présenté dans cette thèse fera l'objet de 2 publications en cours de rédaction.

ABBREVIATIONS

aa	acide aminé
ADN	Acide déoxyribonucléique
ADNc	ADN complémentaire
ARN	acide ribonucléique
ARNm	ARN messenger
ARNr	ARN ribosomal
ATCC	American Type Culture Collection
ATP	Adénosine triphosphate
BEt	Bromure d'Ethidium
CFTR	Cystic Fibrosis transmembrane conductance regulator
cM	centimorgan
CMH	Complexe Majeur d'Histocompatibilité
cpm	coups par minute
CTP	Cytosine triphosphate
Da	Dalton
ddNTP	Didéoxynucléotique triphosphate
DEPC	Diéthylpyrocarbonate
DMSO	Diméthylsulfoxyde
DNase	Déoxyribonucléase
DO	Densité Optique
DT	DL- α , ϵ - Diaminopimelic acid-thymidine
E. Coli	Escherichia Coli
EDTA	Acide Ethylène Diamine Tétracétique
GTP	Guanosine triphosphate
HGM	Human Gene Mapping
HLA	Human Leucocyte Antigen
Ig	Immunoglobuline
IGF	Insuline-like Growth Factor
kb	kilobase
LB	Luria Bertani
MOPS	Acide-3-(N-morpholino)-propane sulfonique
PAS	Periodic Acide Schiff
pb	paire de bases
PBS	phosphate buffered saline
PCR	Polymerase Chain Reaction
PEG	Poly Ethylène Glycol
PFGE	Pulsed Field Gel Electrophoresis
PMSF	Phényl-Méthane Sulfonyl Fluoride
PRP	Proline Rich Protein

QSP	quantité suffisante pour
RER	Réticulum Endoplasmique Rugueux
RFLP	Restriction Fragment Length Polymorphism
RNase	ribonucléase
SAB	Sérum Albumine bovine
SDS	Sodium Dodécyl Sulfate
ssDNA	Salmon Sperm DNA
Tris	Tris-(hydroxy méthyl) aminoéthane
TSS	Transformation and Storage Solution
TTP	Thymidine triphosphate
U.V.	Ultra-Violet
UDP	Uridine diphosphate
VNTR	Variable Number of Tandem Repeat

SOMMAIRE

AVANT-PROPOS	1
INTRODUCTION	3
I. LE MUCUS	3
I-1. <u>Structure- Fonction</u>	3
I-2. <u>Les pathologies</u>	4
A. <u>ANOMALIES DU MUCUS ET MALADIES RESPIRATOIRES</u>	5
1) <i>La mucoviscidose</i>	5
2) <i>Les pathologies broncho-pulmonaires de l'enfant</i>	6
3) <i>les pathologies bronchiques de l'adulte</i>	6
4) <i>les pathologies des sécrétions nasales</i>	7
5) <i>les pathologies de l'oreille moyenne</i>	7
B. <u>ANOMALIES DU MUCUS ET MALADIES GASTRO-INTESTINALES</u>	8
C. <u>PATHOLOGIES DU MUCUS CERVICAL</u>	9
D. <u>CANCERS EPITHELIAUX</u>	10
II. LES MUCINES	11
A. <u>LE ROLE DES MUCINES</u>	13
1) <i>le rôle des mucines sécrétées</i>	13
2) <i>le rôle des "mucin-like"</i>	13
B. <u>BIOCHIMIE DES MUCINES</u>	14
1) <i>Masse moléculaire</i>	14
2) <i>Composition chimique</i>	14
2a. <i>Composante glycanique</i>	14
2b. <i>Composante peptidique</i>	17
3) <i>Organisation moléculaire : modèles structuraux</i>	20
4) <i>Biosynthèse</i>	23
5) <i>Régulation de la biosynthèse</i>	24
III. LES APOMUCINES	25
A. <u>MUCINES ANIMALES</u>	26
B. <u>MUCINES HUMAINES</u>	27
1) <i>épisialine</i>	28
2) <i>mucines sécrétées</i>	28
C. <u>EXPRESSION DES GENES DE MUCINES</u>	38
1) <i>expression tissulaire</i>	38
2) <i>expression cellulaire</i>	38
D. <u>GENES DE MUCINES</u>	40
IV. LE CHROMOSOME 11 ET SA REGION 11P15	42
A. <u>LE CHROMOSOME 11</u>	42
B. <u>LA REGION 11P15 : CARTOGRAPHIE</u>	42
C. <u>LA REGION 11P15 : GENETIQUE</u>	48

1) les gènes <i>WT1</i> et <i>WT2</i>	49
2) autres oncogènes ou gènes suppresseurs de tumeurs	50
3) facteurs de croissance ou de différenciation	51
4) Méthylation - empreinte génomique parentale	51
STRATEGIE	53
I. MEGACARTOGRAPHIE DE RESTRICTION	53
A. <u>PRINCIPE DE L'ELECTROPHORESE EN CHAMP PULSE</u>	54
B. <u>CHOIX DE LA TECHNIQUE</u>	59
C. <u>CHOIX DES LIGNEES</u>	59
D. <u>PREPARATION DU MATERIEL</u>	60
E. <u>CHOIX DES ENDONUCLEASES DE RESTRICTION</u>	61
F. <u>LES SONDES</u>	65
II. MARCHE SUR LE CHROMOSOME	65
A. <u>PRINCIPE DE LA MARCHE SUR LE GENOME</u>	66
B. <u>CONSTRUCTION D'UNE BANQUE GENOMIQUE CHEVAUCHANTE</u>	66
1) Fractionnement de l'ADN génomique	69
2) Préparation du vecteur	69
3) Ligation	69
C. <u>BANQUE GENOMIQUE EN COSMIDES</u>	70
D. <u>LES SONDES</u>	73
1) <i>JER57</i>	73
2) <i>JER58</i>	78
RESULTATS	84
I. MEGACARTOGRAPHIE DE RESTRICTION	84
A. <u>LES LIGNEES</u>	84
B. <u>LES ENZYMES</u>	86
1) Groupe I : <i>Asc I</i> et <i>Not I</i>	86
2) Groupe II et III : <i>Bss HII</i> , <i>Sac II</i> , <i>Ksp I</i> et <i>Nar I</i>	86
3) Groupe IV : <i>Mlu I</i> et <i>Nru I</i>	87
4) <i>Swa I</i> , <i>Spe I</i> et <i>Ssp I</i>	87
C. <u>RESULTATS</u>	89
D. <u>INTERPRETATION DES RESULTATS</u>	94
1) Le positionnement de la carte dans la région 11p15	94
2) La proximité des séquences exoniques...	95
3) La construction de la cartographie	97
4) Les îlots CpG et la localisation des gènes <i>MUC</i>	109
II. MARCHE SUR LE GENOME	113
A. <u>LA BANQUE GENOMIQUE RESTRINTE DE LA LIGNEE 1W1</u>	113
1) Les clones obtenus	113
2) Carte de restriction de <i>ROM11X-6</i>	113
3) Séquence du fragment (<i>Eco RI/Xba I</i>) hybridant avec <i>JER57</i>	115
B. <u>LA BANQUE GENOMIQUE HUMAINE EN VECTEUR EMBL4 CRIBLEE PAR LA SONDE <i>JER58</i></u>	117

1) <i>Clones obtenus</i>	117
2) <i>Séquence de CEL2A</i>	117
C. <u>LA BANQUE GENOMIQUE HUMAINE EN VECTEUR EMBL4</u>	
<u>CRIBLEE PAR LA SONDE JER57</u>	120
1) <i>Clones obtenus</i>	120
2) <i>Carte de restriction de CEL5</i>	120
3) <i>Séquence des fragments exoniques de CEL5</i>	123
D. <u>LA BANQUE GENOMIQUE HUMAINE EN VECTEUR COSMIDIQUE</u>	
<u>CRIBLEE PAR LA SONDE JER57</u>	134
1) <i>Carte de restriction de BEN1</i>	134
2) <i>Séquence de BEN1A</i>	139
3) <i>Carte de restriction de BEN2</i>	151
DISCUSSION	155
I. RESULTATS ACQUIS	155
A. <u>LA CARTOGRAPHIE</u>	155
B. <u>LES GENES DE MUCINES TRACHEOBRONCHIQUES</u>	155
1) <i>MUC5AC</i>	155
2) <i>MUC5B</i>	156
II. LES INTERROGATIONS	157
A. <u>UN CLUSTER "MUC" ?</u>	157
B. <u>LE GENE MUC5B</u>	157
C. <u>LE GENE DE SOURIS HOMOLOGUE A MUC5B</u>	158
D. <u>L'EXPRESSION DES GENES DE MUCINES</u>	158
III. DISCUSSION	160
A. <u>L'HETEROGENEITE GENIQUE</u>	160
B. <u>EXPRESSION ET REGULATION</u>	164
C. <u>ROLE DES MUCINES DANS LA TUMORIGENESE ?</u>	166
APPENDICE TECHNIQUE	168
I. ANALYSE DE L'ADN GENOMIQUE PAR LA TECHNIQUE DE SOUTHERN BLOT	168
A. <u>PREPARATION DE L'ADN GENOMIQUE</u>	168
1) <i>Lyse cellulaire</i>	168
2) <i>Protéolyse</i>	169
3) <i>Précipitation de l'ADN</i>	169
4) <i>Quantification et appréciation de la pureté</i>	169
B. <u>HYDROLYSE ENZYMATIQUE</u>	169
C. <u>ELECTROPHORESE</u>	169
D. <u>TRANSFERT ET FIXATION</u>	170
E. <u>HYBRIDATION A UNE SONDE NUCLEIQUE</u>	170
1) <i>Préparation de la sonde</i>	170
2) <i>Hybridation</i>	170
3) <i>Lavages</i>	171
4) <i>Autoradiographie</i>	171
F. <u>HYBRIDATION A UNE SONDE OLIGONUCLEOTIDIQUE</u>	171

1) <i>Préparation de la sonde</i>	171
2) <i>Hybridation</i>	172
3) <i>Lavages</i>	172
II. ANALYSE DE L'EXPRESSION PAR LA TECHNIQUE DE NORTHERN BLOT	172
A. <u>PREPARATION DES ARN TOTAUX</u>	172
B. <u>ELECTROPHORESE DES ARN EN MILIEU DENATURANT</u>	174
C. <u>TRANSFERT ET HYBRIDATION</u>	175
III. MEGACARTOGRAPHIE DE RESTRICTION	175
A. <u>PREPARATION DE L'ADN A PARTIR DE CELLULES EN CULTURE</u>	175
B. <u>DIGESTION DES BLOCS PAR DES ENDONUCLEASES DE RESTRICTION</u>	176
C. <u>ELECTROPHORESE EN CHAMP PULSE : TECHNIQUE CHEF</u>	177
1) <i>Appareillage</i>	177
2) <i>Paramètres d'électrophorèse</i>	178
D. <u>LES MARQUEURS DE TAILLE</u>	179
E. <u>TRANSFERT, HYBRIDATION ET AUTORADIOGRAPHIE</u>	179
IV. GENOTHEQUES EN VECTEUR λEMBL4	180
A. <u>CONSTRUCTION DES GENOTHEQUES</u>	180
1) <i>Préparation du phage λEMBL4</i>	181
2) <i>Hydrolyse partielle de l'ADN par Sau 3A</i>	181
3) <i>Purification des fragments</i>	182
4) <i>Quantification</i>	182
5) <i>Ligation</i>	183
6) <i>Packaging</i>	183
7) <i>Infection des bactéries DP50 P₂</i>	183
B. <u>CRIBLAGE DES BANQUES EN VECTEUR λEMBL4</u>	184
C. <u>PREPARATION DES PHAGES RECOMBINANTS</u>	185
1) <i>Préministock</i>	185
2) <i>Ministock</i>	185
3) <i>Préparation rapide de l'ADN phagique</i>	186
4) <i>Amplification et extraction de l'ADN recombinant</i>	186
D. <u>ETABLISSEMENT DES CARTES DE RESTRICTION</u>	187
V. GENOTHEQUE EN VECTEUR COSMIDIQUE pWE 15	188
A. <u>CRIBLAGE DE LA BANQUE EN VECTEUR COSMIDIQUE</u>	188
B. <u>PURIFICATION DE L'ADN DES COSMIDES RECOMBINANTS</u>	189
1) <i>Purification de l'ADN à partir de 3 ml de culture</i>	189
2) <i>Purification de l'ADN à partir de 500 ml de culture</i>	190
C. <u>ETABLISSEMENT DES CARTES DE RESTRICTION</u>	191
VI. SEQUENCE NUCLEOTIDIQUE	191
A. <u>EXTRACTION ET PURIFICATION DU FRAGMENT A SEQUENCER</u>	191
B. <u>PREPARATION DU VECTEUR PLASMIDIQUE</u>	193
1) <i>Hydrolyse</i>	193
2) <i>Déphosphorylation</i>	193
C. <u>LIGATION</u>	194
D. <u>TANSFORMATION DES BACTERIES E. COLI HB101</u>	194

<u>E. HYBRIDATION A UNE SOND E NUCLEIQUE</u>	195
<u>F. AMPLIFICATION DES CLONES PAR PCR</u>	195
<u>G. BOILING : PREPARATION DE L'ADN DES CLONES RECOMBINANTS</u>	197
<u>H. METHODE DE DELETION-MUTATION A L'EXONUCLEASE III</u>	197
<u>I. SEQUENCAGE</u>	200
LEXIQUE DES FORMULES	201
BIBLIOGRAPHIE	204

AVANT-PROPOS

AVANT-PROPOS

Les mucines sont une famille de macromolécules caractérisées par une haute densité en chaînes O-glycanniques et un poids moléculaire élevé. Elles recouvrent la surface des cellules épithéliales au niveau de nombreux tissus : l'oeil, les glandes salivaires, la glande mammaire, le nasopharynx, l'arbre trachéobronchique, le tractus gastro-intestinal, la vésicule biliaire, le pancréas, le rein, l'utérus, l'ovaire, le testicule, la prostate. Les mucines élaborées dans tous ces tissus, et à l'intérieur d'un même tissu sont complexes et hétérogènes au plan structural et antigénique. Actuellement, grâce aux connaissances acquises récemment par la biologie moléculaire, la classification de ces molécules prend en considération deux types de biosynthèse :

- des glycoprotéines incluses dans l'architecture membranaire, appelées aussi glycoprotéines "mucin-like",
- des glycoprotéines sécrétées à la surface épithéliale et responsables de la formation du mucus.

Les glycoprotéines "mucin-like" ont été étudiées principalement dans la glande mammaire, l'ovaire, le pancréas, et dans des lignées tumorales établies à partir de ces tissus.

Les mucines sécrétées ont fait l'objet de nombreuses études biochimiques, dans les mucus respiratoire et digestif.

Le mucus est une sécrétion douée de propriétés rhéologiques (viscosité, élasticité, fluidité, filance, adhésivité). Il assume de nombreuses fonctions essentielles à la vie dont la finalité est de créer une interface entre le milieu extérieur et les épithélia soumis à de multiples agressions potentielles.

Les mucines coexistent au sein du mucus sous la forme d'une population très hétérogène de molécules que l'on peut grossièrement classer en mucines neutres (fucomucines), mucines acides (sialomucines), et mucines très acides (sulfomucines). Chaque muqueuse élabore une sécrétion où se définit un équilibre qualitatif et quantitatif de ces trois catégories de mucines. Dans de nombreuses pathologies, observées en pneumologie (mucoviscidose, bronchite chronique, asthme), en gastro-entérologie (cancer du côlon, ulcère gastrique, ulcère duodéal, maladies inflammatoires du tube digestif), en gynécologie (stérilité,

cancer), et en cancérologie (cancers épithéliaux), cet équilibre est rompu et s'accompagne d'altérations de la muqueuse.

Qu'il s'agisse de glycoprotéines "mucin-like" ou de mucines sécrétées, ces macromolécules ont suscité des développements de recherches dans le cadre cancérologique. La caractérisation physico-chimique de certaines de ces molécules, la définition d'épitopes surexprimés ou démasqués au cours de la cancérisation ont suscité la préparation de très nombreux anticorps monoclonaux. On peut espérer de ces outils qu'ils seront utiles à la détection précoce de tumeurs mucisécrétantes, à l'estimation de l'étendue de ces tumeurs, de leur degré de dysplasie, et de leur capacité métastatique. En dehors de ce contexte médical appliqué, l'étude fondamentale de la biosynthèse de macromolécules aussi complexes que les mucines revêt un intérêt scientifique certain.

Jusqu'en 1989, les connaissances structurales concernaient essentiellement les chaînes oligosaccharidiques auxquelles était attribuée l'extrême hétérogénéité des mucines. Depuis, l'apport des techniques de biologie moléculaire a permis, grâce à l'obtention de sondes d'ADNc partiels, une meilleure connaissance de l'axe peptidique et a démontré un deuxième niveau d'hétérogénéité cette fois-ci peptidique, associé à une hétérogénéité génique.

Sept gènes portant le sigle MUC dans la nomenclature du HGM (Human Gene Mapping) ont déjà pu être identifiés, mais, mis à part MUC1, la plupart ne sont pas encore caractérisés ou ne le sont que partiellement. Quatre d'entre eux sont localisés sur le chromosome 11 dans la région p15.5 : MUC2 et très récemment MUC6 ont été découverts par l'équipe de Y. KIM et J. GUM (San Francisco), MUC5B et MUC5AC ont été identifiés par notre groupe. C'est dans ce contexte que s'est situé mon travail de thèse dont l'objectif était d'établir les rapports de proximité de ces gènes par mégacartographie et d'aborder l'étude de l'organisation génomique de MUC5B et MUC5AC grâce aux techniques de marche sur le chromosome.

INTRODUCTION

INTRODUCTION

La thématique de notre groupe intitulée "génétique des apomucines humaines" a été développée à partir du modèle biochimique des "mucines respiratoires" isolées du mucus trachéobronchique et donc elle s'intéresse, a priori, essentiellement aux mucines sécrétées.

I. LE MUCUS

I - 1. Structure - Fonction

Le mucus est une sécrétion très hétérogène qui forme un tapis continu à la surface des épithélia respiratoire, digestif et génital, constituant ainsi une barrière de protection efficace entre l'environnement et la muqueuse. Ce film de mucus est formé de deux phases: une phase aqueuse au contact des cellules, de viscosité faible, proche de celle de l'eau, et une phase superficielle, ou phase gel, se caractérisant par une viscosité et une élasticité élevées. Le mucus est sécrété par des cellules spécialisées, cellules caliciformes, et/ou glandes muqueuses, selon les tissus, sous forme de granules de sécrétion qui sont libérés dans la lumière, formant ensuite au contact de l'eau un gel viscoélastique s'étirant à la surface de l'épithélium.

Le rôle important joué par le mucus dans la protection-défense des épithélia est décrit dans la littérature depuis une vingtaine d'années. Parmi les multiples fonctions du mucus, on peut citer:

- une fonction chimique (antioxydante, antiprotéasique, un pouvoir tampon), comme c'est le cas au niveau des muqueuses du tractus digestif, où se concentrent des pH extrêmes et des activités enzymatiques très puissantes et variées.

- une fonction d'humidification et d'hydratation des muqueuses exposées au contact de l'air (arbre respiratoire).

- une fonction de transport, lors de l'épuration muco-ciliaire, lors de la progression du contenu luminal digestif, ou lors de la progression des spermatozoïdes dans les voies génitales.

- une fonction antibactérienne, par "piégeage" des bactéries (muqueuses colique, respiratoire, génitale).

- une fonction de lubrification de la paroi de conduits (côlon), de canaux (canaux biliaires, pancréatiques).

- une fonction d'amortissements de chocs mécaniques (broyage du bol alimentaire dans la muqueuse gastrique).

Le mucus est composé essentiellement d'eau (95%) emprisonnée dans une matrice gel comme dans une éponge, ou liée aux macromolécules. La fraction non aqueuse est constituée d'électrolytes (1%) et de macromolécules (4%). Dans le cas par exemple de la muqueuse trachéobronchique, ces macromolécules sont soit sécrétées par des cellules spécialisées de la muqueuse et de la sous-muqueuse : mucines, protéoglycannes, lysozyme, transferrine, IgA sécrétoires, inhibiteurs de protéases, soit issues de la transsudation du secteur plasmatique : séralbumine, IgM, IgG, antiprotéases (α_1 -antitrypsine, α_1 -antichymotrypsine, α_2 macroglobuline).

Sont associés :

- des acides nucléiques [1] en quantité peu importante dans le mucus non infecté, en considérable augmentation quand les sécrétions sont le siège d'infections bactériennes.
- des lipides (phospholipides, lipides neutres, glycolipides) [2, 3] dont le taux varie également avec l'état d'infection des sécrétions.

Pour que se maintiennent l'intégrité et la fonctionnalité des épithelia, il faut que soit maintenu ou restauré un mucus adapté au type de fonction qu'il doit remplir. En pathologie humaine, de nombreuses maladies, dites de "civilisation", car touchant davantage des populations à niveau de vie élevé, se sont développées aux dépens de l'ensemble épithelium/mucus : asthme, ulcères gastroduodénaux, cancers épithéliaux par exemple.

I - 2. Les pathologies

Beaucoup de maladies affectent la production du mucus de manière qualitative et/ou quantitative.

Les qualités "in vivo" du mucus sont déterminées par l'interaction de nombreux facteurs. Non seulement la structure et la quantité des mucines sécrétées sont essentielles mais aussi l'environnement ionique et la présence de composants protéiques autres que les mucines (protéines de liaison, par exemple) doivent être déterminants pour conférer au mucus ses propriétés.

A. ANOMALIES DU MUCUS ET MALADIES RESPIRATOIRES

La composition même du mucus respiratoire du sujet sain est mal connue à cause des difficultés à prélever du mucus natif et à savoir distinguer la part respective de l'épithélium de surface et celle des glandes sous-muqueuses. Le mucus provenant d'expectorations de sujets atteints d'hypersécrétions bronchiques est mieux connu [4].

1) *La mucoviscidose*

La mucoviscidose est une exocrinopathie d'origine génétique (maladie autosomique récessive). Elle est dominée par l'existence d'une bronchopathie hypersécrétante, chronique, à potentialité évolutive. On observe une modification des caractéristiques physicochimiques des mucines et un changement des propriétés rhéologiques du mucus. La modification de la composante mucoïde consiste en la variation de l'équilibre existant entre les différentes fractions de mucines (neutres, acides, très acides), avec une augmentation de la composante acide (sialo et sulfo-mucines) [5]. Les résidus terminaux acides de ces chaînes glycaniques semblent constituer des sites préférentiels d'adhésion des bactéries. Successivement, *Staphylococcus aureus* puis surtout *Pseudomonas aeruginosa* colonisent précocément l'épithélium respiratoire des enfants atteints de mucoviscidose. La relative spécificité de l'infection bronchique explique tout l'intérêt porté à l'étude des interactions mucines-bactéries.

La recherche du défaut génétique à l'origine de la mucoviscidose a abouti à la mise en évidence d'une pathologie des canaux "chlore". La protéine membranaire déficiente appelée CFTR est impliquée dans le transport des ions Cl^- [6-8]. La déficience de ces canaux entraînerait une augmentation de la différence de potentiel transépithélial et, sur le plan physique, une diminution du degré d'hydratation des sécrétions bronchiques pouvant aboutir à un déficit de clairance muco-ciliaire. Il a été récemment montré que la protéine CFTR avait aussi une fonction ATPasique [9] et qu'elle n'était pas seulement un canal. Elle intervient dans l'acidification des endosomes et du trans-golgi, et, par là, influe de manière qualitative et quantitative sur la glycosylation et la sulfatation des glycoprotéines membranaires ou sécrétées des cellules épithéliales en culture [10, 11]. L'augmentation du pH sur le site d'activité de glycosyl transférase modifierait, chez les sujets atteints de mucoviscidose, qualitativement et quantitativement l'addition des oses périphériques et des résidus sulfate. Cette hypothèse est en accord avec les taux élevés de fucose, galactose, N-Acétyle glucosamine et sulfate retrouvés dans les sécrétions de ces sujets.

2) Les pathologies broncho-pulmonaires de l'enfant

Les infections bronchiques à répétition du jeune enfant sont fréquentes et relativement difficiles à prévenir. Leur origine est souvent virale. Les virus en cause sont très nombreux et peuvent présenter un tropisme particulier pour les grosses bronches (virus parainfluenzae), pour les petites bronches (virus respiratoire syncytial), ou pour le parenchyme pulmonaire (adénovirus). Ces infections contagieuses sont souvent récidivantes et peuvent entraîner une inflammation bronchique chronique aboutissant à des tableaux sévères d'obstruction, d'emphysème, d'atélectasie, de dilatation des bronches.

Dans la pathologie des infections respiratoires récidivantes, on observe à la fois des anomalies de la mobilité des cellules ciliées et des anomalies biochimiques et rhéologiques du mucus. Outre l'atteinte virale, l'allergie est un autre facteur déclenchant.

3) Les pathologies bronchiques de l'adulte

La pathologie inflammatoire bronchique est associée à une lésion caractéristique, la métaplasie mucipare de l'épithélium. Cette lésion comporte un excès de cellules caliciformes et une raréfaction des cellules ciliées. Des modifications spécifiques du mucus sont observées au cours de la bronchite ; le poids sec des sécrétions augmente en relation avec l'augmentation du volume de la phase gel. Les sulfomucines puis les fucomucines neutres augmentent aux dépens des sialomucines. Après une augmentation initiale, les IgA sécrétoires diminuent progressivement. Les protéoglycannes augmentent progressivement (acide hyaluronique). Aux stades initiaux de la bronchite chronique, la stimulation de la sécrétion de sialomucines, IgA sécrétoires, lysozyme, produit une sécrétion hypervisqueuse, hyperélastique ; aux stades plus avancés, l'apparition des mucines neutres, de la sérum albumine, la réduction des IgA sécrétoires conduisent à une expectoration fluide avec diminution de la viscosité, de l'élasticité en relation avec la perte de la structure fibrillaire du mucus. La surinfection bronchique augmente l'exsudation des protéines plasmatiques, la libération de fibres d'ADN, ce qui réaugmente la viscosité. Mais les protéases bronchiques, bactériennes, leucocytaires achèvent de désorganiser la structure fibrillaire du mucus et l'élasticité s'effondre.

Au cours de l'asthme intrinsèque, l'hypertrophie glandulaire est comparable à celle observée dans la bronchite chronique et les caractères du mucus sont voisins.

Dans l'asthme extrinsèque, au contraire, l'hypertrophie glandulaire est modérée ; les mucines, peu abondantes, possèdent des chaînes oligosaccharidiques plus longues que celles observées dans l'asthme intrinsèque. Ici, l'hypersécrétion bronchique est due à la libération de médiateurs chimiques. Les sécrétions sont denses, déshydratées et peuvent constituer des bouchons muqueux obstruant bronches et bronchioles.

Les connaissances apportées par l'étude de l'expression et de la régulation des gènes de mucines devraient déboucher sur des traitements adaptés aux différentes anomalies du mucus. Pour l'instant l'activité thérapeutique réside en l'hyperhydratation, l'action des agents réducteurs (N-acétylcystéine), et des agents mucorégulateurs (carbocistéine, ambroxol).

4) Les pathologies des sécrétions nasales

Autant les travaux sur la physiopathologie des sécrétions trachéobronchiques sont nombreux, autant les études consacrées aux sécrétions nasales sont rares. Ces sécrétions proviennent de glandes séromuqueuses de la muqueuse nasale et sinusienne, la participation des cellules caliciformes de cette muqueuse étant faible. La symptomatologie des hypersécrétions nasales associe : obstruction, éternuements et rhinorrhée.

Les hypersécrétions nasales peuvent prendre la forme de rhinorrhées séreuses à sécrétions peu visqueuses, peu filantes, de rhinorrhées muqueuses à sécrétions hypervisqueuses et hyperélastiques, de rhinorrhées muco-purulentes constituées de sécrétions hypervisqueuses et hypoélastiques.

Le traitement de ces hypersécrétions nasales reste encore très empirique et très limité. La thérapeutique, dans ce domaine, devrait bénéficier elle aussi de la meilleure connaissance de l'expression tissulaire normale et pathologique des gènes de mucines.

5) Les pathologies de l'oreille moyenne

L'otite séro-muqueuse chez l'enfant est une affection très banale, caractérisée par l'accumulation et la persistance d'un exsudat inflammatoire au niveau de l'oreille moyenne. Le mucus est produit par la métaplasie mucipare de la muqueuse inflammatoire et a des caractéristiques rhéologiques anormales. La trompe d'Eustache est alors le siège d'une obstruction fonctionnelle par inefficacité du transport muco-ciliaire. La stase du mucus entretient la réaction inflammatoire, en particulier par un afflux de cellules macrophagiques qui peut conduire à une altération de la membrane tympanique.

B. ANOMALIES DU MUCUS ET MALADIES GASTRO-INTESTINALES

Tout au long du tube digestif, le mucus présente des propriétés physicochimiques distinctes en rapport avec les différentes fonctions exercées par les différents épithelia dans la digestion.

Le tiers inférieur de l'oesophage, jusqu'au cardia, sécrète un mucus contenant des mucines neutres et des sialomucines. Puis, la muqueuse fundique élabore un épais tapis muqueux riche en mucines neutres; de même, les glandes pyloriques participent à la sécrétion de mucines neutres au niveau de la muqueuse antrale. Ce mucus constitue, à l'état normal, une barrière efficace à la rétro-diffusion des ions H^+ depuis la lumière vers la paroi. Néanmoins ce mucus est normalement dégradé dans son organisation en réseau par la bile et la pepsine. Un juste équilibre entre sécrétion de mucus et altération de ce mucus est nécessaire pour que se maintienne l'intégrité de l'épithélium sous-jacent. Une des altérations pathologiques de la muqueuse gastrique est la métaplasie intestinale (MI). Au cours du processus inflammatoire, l'épithélium mucosécrétant de type gastrique (mucines neutres) est progressivement remplacé par un épithélium mucosécrétant de type intestinal (sulfomucines). Ce processus est fréquemment associé, à des degrés divers (MI type I, II, III) aux différentes pathologies de la muqueuse [12] : ulcère gastrique, duodéal, polypes gastriques, maladie de Ménétrier, gastrite chronique atrophique, reflux gastrooesophagien, cancer. L'analyse biochimique du mucus gastrique chez des malades atteints d'ulcère gastrique a montré que les mucines avaient une faible masse moléculaire [13] associée à une perte des propriétés rhéologiques du mucus.

La muqueuse de l'intestin grêle possède, principalement une fonction d'absorption dévolue aux entérocytes. Les cellules à mucus beaucoup moins nombreuses élaborent des mucines neutres et des sialomucines. Il existe peu d'études du mucus dans les pathologies du grêle, néanmoins, des modifications qualitatives des mucines ont été rapportées dans l'ulcère duodéal, la duodénite idiopathique et la maladie de Crohn.

La muqueuse recto-colique élabore majoritairement des sulfomucines. Il existe un gradient de sécrétion suivant les différents segments anatomiques du côlon [14-16]. Ainsi, dans le côlon gauche et le rectum, les sulfomucines prédominent dans les 2/3 inférieurs des glandes alors que les sialomucines sont majoritaires dans le 1/3 supérieur. Par contre dans le côlon droit, les sulfomucines occupent les 2/3 supérieurs et les sialomucines se trouvent dans la partie profonde de ces glandes. De plus, les proportions de sialomucines suivent la topographie anatomique avec une prédominance dans le côlon gauche [17].

La lésion élémentaire observée dans le côlon a été appelée muqueuse transitionnelle (MT) ; elle associe des modifications morphologiques (épaississement de la muqueuse, glandes

allongées, tortueuses et ramifiées) et des modifications histochimiques (importante réduction des sulfomucines). Cette lésion est observée dans les états précancéreux, dans les carcinomes colo-rectaux, ainsi que dans quelques lésions non néoplasiques : ulcère solitaire du rectum, diverticules.

La maladie de Crohn et la recto-colite hémorragique (RCH) sont deux formes de colite chronique, la première conservant la mucosécrétion par les cellules calciformes, la seconde manifestant une déplétion de sécrétion [18]. Dans les deux cas, la sécrétion est essentiellement celle de sialomucines (au lieu de sulfomucines), néanmoins les choses sont complexes et la qualité des mucines sécrétées varie en fonction des poussées évolutives, de la durée de la maladie et de l'étendue des lésions [19, 20].

Le rôle joué par les mucines dans la genèse des calculs biliaires a été mis en évidence par Lec et coll. [21, 22]. En effet, il existe une corrélation entre le degré d'hypersécrétion de mucines biliaires et l'apparition de "calculs-cholestérol". Au contraire, l'inhibition pharmacologique de cette hypersécrétion de mucines par les salicylés est reliée à une suppression de la formation de ces calculs.

C. PATHOLOGIES DU MUCUS CERVICAL

Dans l'appareil urogénital féminin, la sécrétion de mucus est essentiellement localisée dans le col utérin (endocol), mais l'endomètre, les trompes de Fallope, et les glandes de Bartholin participent également à la production du mucus [23].

Des variations qualitatives et quantitatives du mucus ont été observées au cours du cycle menstruel. En phase folliculaire, les cryptes endocervicales libèrent un mucus fluide dont le volume augmente brutalement en période pré-ovulatoire, puis se maintient jusqu'à l'ovulation. Puis en phase lutéale, le volume de mucus sécrété diminue rapidement ; ses propriétés physiques changent, il devient opaque, épais et visqueux. Les qualités du mucus endocervical varient encore avec l'âge, la parité, la ménopause [24].

Le rôle important joué par le mucus dans la mobilité, la viabilité des spermatozoïdes a été étudié in vitro. Il est probable que des altérations de ce mucus puissent contribuer à certaines formes de stérilité. Les anomalies épithéliales du col (néoplasies intra-épithéliales, adénocarcinomes) sont observées à la jonction de l'endocol (épithélium glandulaire) et de l'exocol (épithélium stratifié). Cette zone de jonction complexe se déplace, sous influence hormonale grâce à un processus de métaplasie. Peu de travaux relatent les variations histochimiques des mucines sur cette zone de jonction. Dans ce domaine, les sondes des

différents gènes "MUC" apporteront vraisemblablement des éléments nouveaux de classification.

D. CANCERS EPITHELIAUX

95 % des tumeurs humaines se développent aux dépens de cellules épithéliales. Ceci justifie que l'on se soit tourné vers l'étude des mucines, synthétisées spécifiquement par les cellules épithéliales, pour rendre compte de la dysplasie et de la cancérisation des épithélia. La caractérisation histochimique des mucines (colorations au PAS, bleu de toluidine ou bleu alcian à différents pH) a permis de mettre en évidence des anomalies qualitatives et/ou quantitatives des mucines sécrétées sur des lésions précancéreuses. L'étude de ces glycoprotéines permet donc d'espérer définir des marqueurs biologiques de diagnostic, de pronostic, de récurrence, d'invasion métastatique. Les molécules les plus étudiées actuellement comme marqueurs tumoraux sont les glycoprotéines membranaires appelées "mucin-like" ou mucines membranaires et, parmi elles, l'épisialine, codée par le gène MUC1. Des concentrations élevées de cette glycoprotéine sont dosées dans le sérum de malades atteints de cancer du sein, de l'ovaire, du pancréas. Le mécanisme responsable de l'amplification de son expression sur la membrane des cellules épithéliales tumorales est encore inconnu. Une glycosylation imparfaite de la protéine (richesse en antigènes Tn ou T) permet d'obtenir des anticorps monoclonaux très utiles pour le diagnostic par immunoscintigraphie [25] ou pour la thérapeutique (drogues véhiculées par l'anticorps monoclonal) [25].

Des études comparables à celles dont "MUC1" a fait l'objet, verront jour à propos des autres "MUC", lorsque l'étude fondamentale des gènes sera plus avancée.

Ce chapitre, qui décrit l'intérêt d'étudier les mucines dans un certain nombre de pathologies, dont la liste n'est pas exhaustive, permet de se rendre compte du rôle essentiel joué par le mucus et par conséquent de souligner l'importance qu'il y a à maintenir son intégrité. Chacune de ces pathologies se caractérise par un changement:

- quantitatif (hypo ou hypersécrétion qui joue sur l'épaisseur de la barrière de mucus) et/ou
- qualitatif (modification des mucines qui perturbe les propriétés rhéologiques du mucus et l'affinité des sucres périphériques vis à vis des ligands naturels).

Une approche pour le traitement de ces maladies serait d'exercer un contrôle thérapeutique de la biosynthèse et de la sécrétion des mucines, que ce soit pour favoriser la

synthèse d'un type spécifique de mucine afin de rétablir un équilibre perturbé ou pour augmenter ou diminuer le volume des mucines sécrétées. Pour cela, il est indispensable d'avoir une meilleure connaissance des systèmes de régulation de ces macromolécules, au niveau génomique (facteurs de transcription, systèmes activateurs...), au niveau transcriptionnel (dégradation, durée de vie des ARNm, épissage...), au niveau traductionnel et post-traductionnel (mécanismes de glycosylation, sécrétion...).

II. LES MUCINES

Les mucines constituent un groupe distinct de glycoprotéines, différant structurellement des glycoprotéines du sérum et des protéoglycannes; elles se définissent comme des populations de macromolécules complexes.

Elles sont synthétisées et sécrétées par des cellules épithéliales spécialisées, appelées cellules caliciformes au niveau des épithelia bronchique et colique. Elles sont sécrétées sous forme de granules, qui, libérés dans la lumière, forment, au contact de l'eau, un gel viscoélastique à la surface de l'épithélium. Ces cellules sont polarisées, c'est à dire que le noyau et le réticulum endoplasmique sont repoussés vers le pôle basal, alors que les grains de sécrétion sont libérés au pôle apical. Elles présentent des caractéristiques morphologiques en rapport avec la production massive de mucines (RER et Golgi très développés). Dans certains tissus, comme les muqueuses respiratoire, antrale, cervicale, nasale et dans certaines glandes : prostate, glandes salivaires sous-maxillaires, les mucines sont également synthétisées par des glandes qui participent très activement à un renouvellement rapide et constant du mucus (Figure 1).

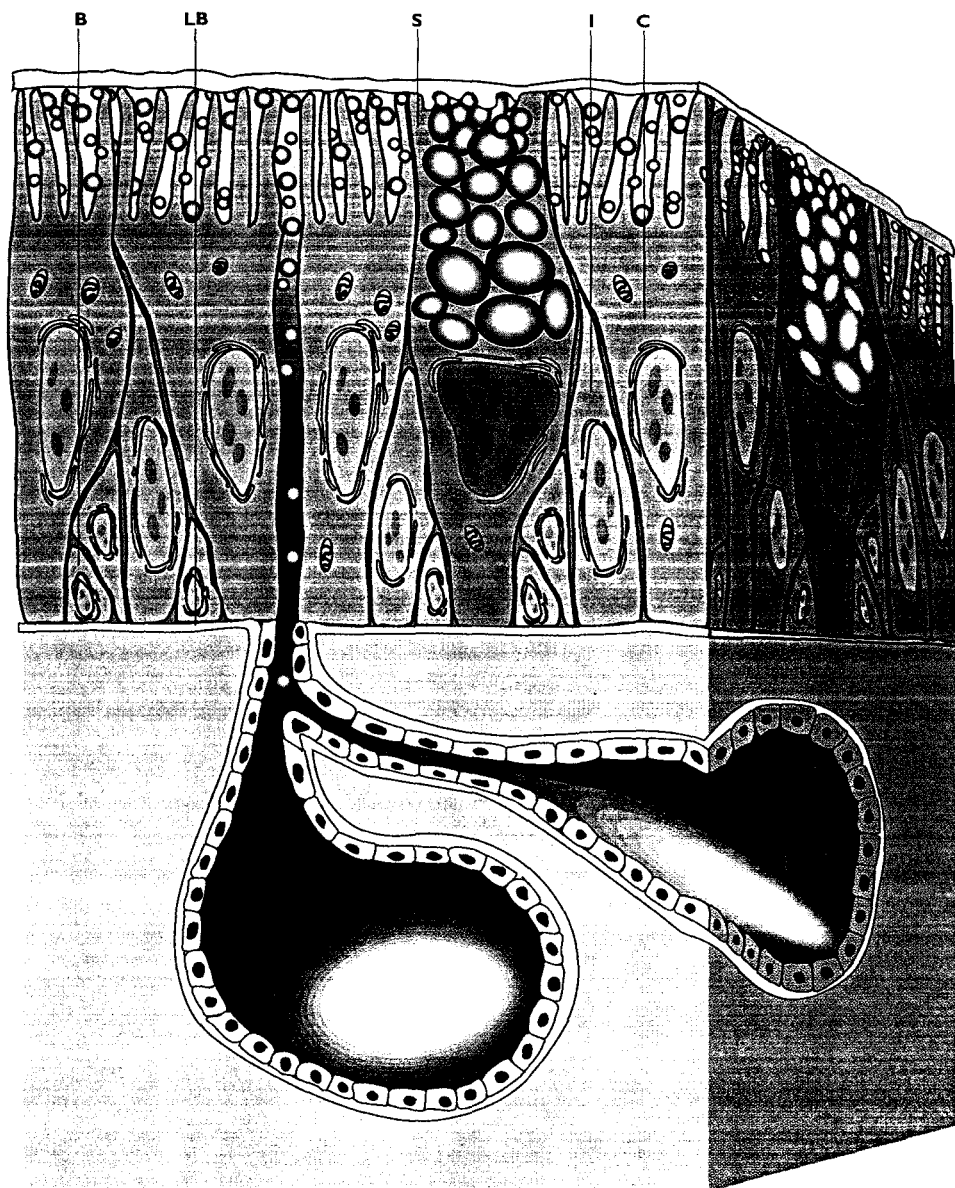


Figure 1 : Architecture de la muqueuse respiratoire.

L'épithélium respiratoire est principalement composé de cellules ciliées (C), de cellules sécrétrices (S), de cellules intermédiaires (I) et de cellules basales (B). L'épithélium de surface est séparé de la sous-muqueuse par une lame basale (LB). Le mucus, sécrété par la sous-muqueuse, recouvre la partie apicale des cils. D'après D. Gaillard, C. Plotkowski et E. Puchelle [26].

A. LE ROLE DES MUCINES

1) *le rôle des mucines sécrétées*

Les mucines sont responsables des propriétés physico-chimiques et biologiques du mucus. Ce sont elles qui lui confèrent son architecture tridimensionnelle en réseau fibrillaire. Ces glycoprotéines de masse moléculaire très élevée forment un enchevêtrement de microfilaments constitués d'un axe polypeptidique hérissé de chaînes glycaniques représentant de véritables pièges pour les molécules, les particules et les bactéries. Très hydrophiles, les mucines contribuent à maintenir une hydratation optimale du mucus.

Outre les propriétés physico-chimiques qu'elles confèrent au mucus, on peut encore citer les rôles suivants :

- Au niveau trachéobronchique, les mucines présentent, par les sucres périphériques de leurs chaînes glycaniques, des sites récepteurs spécifiques pour des adhésines bactériennes [27];

- Au niveau gastrique, les mucines neutres contribuent également à créer une zone tampon entre la lumière gastrique (pH2) et la muqueuse (pH7) [28];

- Au niveau intestinal, les mucines jouent un rôle direct dans la fonction d'absorption des nutriments, des acides gras en particulier [25] et dans la fonction de protection des cellules intestinales vis à vis des hydrolases .

2) *le rôle des "mucin like"*

Les glycoprotéines membranaires dites "mucin like" présentent les caractéristiques physico-chimiques des mucines à savoir : masse moléculaire élevée, richesse en sérine-thréonine, pourcentage élevé de O-glycannes. Ces glycoprotéines membranaires sont ancrées dans la membrane apicale de nombreuses cellules épithéliales. Elles pourraient former une dernière barrière protectrice contre les facteurs extracellulaires (pH acide, hydrolases, inflammation cellulaire) [29]. La densité des mucines membranaires pourrait réguler l'aptitude qu'ont les cellules épithéliales à interagir entre elles ou avec leur environnement: une forte densité limiterait la communication intercellulaire ou empêcherait l'accessibilité des récepteurs aux signaux en général.

B. BIOCHIMIE DES MUCINES

L'isolement, la purification et la caractérisation moléculaire des mucines se sont avérés difficiles en raison de leur propriété d'interagir entre elles ou avec le milieu environnant. L'origine tissulaire multiple des molécules étudiées (muqueuse bronchique, gastrique, colique ; tissus sains ou tissus pathologiques hypersécrétants), la grande diversité des protocoles de préparation des mucines (solubilisation du mucus par des agents chaotropiques variés, purification par chromatographie ou par centrifugation isopycnique en gradient de CsCl) ont fourni des résultats variés quant à la masse moléculaire, et à la composition biochimique des mucines. A partir de ces données, des modèles moléculaires divers ont pu être imaginés.

1) Masse moléculaire

Bien qu'il y ait une grande disparité dans les valeurs des masses moléculaires publiées (Tableau I), il ressort de ces études deux caractéristiques communes à toutes les mucines:

- ces glycoprotéines ont une masse moléculaire très élevée, supérieure ou égale à un million de Daltons;
- il existe une polydispersion des masses au sein d'un même échantillon;
- il existe une polydispersion des masses, d'un organe à l'autre, d'une espèce à l'autre;
- les mucines sont sensibles à l'action des agents réducteurs.

2) Composition biochimique

2.a Composante glycanique

Les mucines sont constituées d'une chaîne peptidique sur laquelle sont branchées de nombreuses chaînes glycaniques par une liaison O-glycosidique.

Tableau I : masse moléculaire de mucines : quelques exemples.

AUTEUR	MUCINE	PREPARATION	M.M (Daltons)	REMARQUES
Creeth & Coll. 1977 [30]	mucine respiratoire humaine	- solubilisation en CsCl - ultracentrifugation en CsCl	3.3 à 7×10^6	diminution de taille après action du β mercaptoéthanol
Marshall & Coll. 1978 [31]	mucine colique de porc	- gel filtration - ultracentrifugation à l'équilibre en CsCl - exclusion en gel filtration Sepharose 2B	10^6	
Sheehan & Coll. 1987 [32]	mucine cervicale humaine	- ultracentrifugation en gradient de densité en GuCl	6 à 24×10^6	
Thornton & Coll. 1990 [33]	mucine trachéale humaine	- ultracentrifugation en gradient de densité CsCl/GuCl - chromatographie Sepharose CL 2B	14 à 16×10^6	mise en évidence de domaines résistants à la trypsine (T-domaines)
Virmani & Coll. 1991 [34]	mucine trachéale canine	- chromatographie Sepharose CL 4B - ultracentrifugation en gradient de densité CsCl	11×10^6	
Sheehan & Coll. 1991 [35]	mucine respiratoire humaine	- extraction par centrifugation en GuCl 6M en présence d'inhibiteurs de protéases	3.2×10^6	

Les chaînes glycaniques représentant 70 à 90% de la masse moléculaire sont caractérisées par une très grande hétérogénéité de structure (plus d'une centaine de structures distinctes sont simultanément présentes dans la sécrétion bronchique d'un seul sujet) et de taille. Elles sont restées longtemps les seules considérées responsables de l'hétérogénéité des mucines. La structure oligosaccharidique est formée de trois parties distinctes:

- le core ou noyau qui comprend le résidu N-acétylgalactosamine lié à l'axe peptidique au niveau d'une sérine ou d'une thréonine et le ou les deux sucres fixés sur cette GalNac [36]. Actuellement six noyaux différents ont été décrits et témoignent de l'origine tissulaire du mucus.

- le squelette formé d'une association linéaire ou branchée d'unités disaccharidiques de deux types, Gal 1,3GlcNac ou Gal 1,4GlcNac.

- la région périphérique qui peut exprimer les antigènes des groupes sanguins A, B, O et Lewis. C'est à l'aide de cette région terminale qui porte les charges négatives des mucines que l'on a pu classer grossièrement les glycopeptides de mucines en fonction de leur caractère plus ou moins acide.

On peut ainsi séparer par chromatographie d'échange d'ions, trois principales fractions qui coexistent au sein du mucus:

- la fraction neutre, pauvre en résidus acides,
- la fraction acide, riche en résidus d'acide sialique,
- la fraction très acide, riche en résidus sulfates.

Le nombre de fractions obtenues ne dépend en réalité que du choix arbitraire du gradient discontinu. Un gradient continu livre un spectre continu de molécules.

Des traces de mannose sont fréquemment observées dans la composition chimique des mucines. Etant donné la difficulté à purifier complètement ces mucines et à les débarrasser de contaminants auxquels elles sont souvent associées, ce mannose a longtemps été considéré comme une contamination de N-glycoprotéines. Pourtant la découverte récente de sites potentiels de N-glycosylation dans la séquence de mucines humaines [37] semble indiquer la présence possible de quelques chaînes glycaniques à liaison N-glycosidique, au milieu d'un grand nombre de chaînes glycaniques à liaison O-glycosidique. Ces chaînes N-glycosidiques pourraient jouer un rôle important dans les étapes initiales de la O-glycosylation.

2.b Composante peptidique

L'axe peptidique des mucines représente, selon l'espèce animale et selon le mucus étudié, de 14 à 37 % [38-41] du poids moléculaire.

La composition en acides aminés est tout à fait originale, les résidus de sérine, thréonine, proline, glycine et alanine représentent de 66 à 75% des acides aminés totaux (Tableau II).

C'est la proportion des acides aminés hydroxylés qui varie le plus d'une étude à l'autre (19 à 48%), le rapport Ser/Thr varie de 0,45 à 1.

L'abondance des acides aminés dicarboxyliques serait corrélée avec l'acidité des mucines, ceci a été démontré dans le cas des mucines coliques [47].

La teneur en acides aminés soufrés (cystéine essentiellement) a fait l'objet de nombreuses controverses : de moins de 0,5% à près de 3% selon les préparations. Or c'est l'interprétation de ces chiffres qui est à la base des modèles moléculaires proposés : avec ou sans pontage (par établissement de liaisons disulfures) inter ou intra-chaînes peptidiques.

Le tryptophane et la méthionine ont été peu étudiés.

Peu de travaux relatent des résultats de séquences peptidiques de mucines. Il s'agit toujours de séquences très partielles de peptides localisés dans des zones peu glycosylées [48-51] (Tableau III).

Tableau II : Composition en acides aminés de quelques préparations de mucines humaines

AUTEURS	ROUSSEL et coll (1975) [40]	CREETH et coll (1977) [30]	LAFITTE et coll (1977) [42]	LAMBLIN et coll (1979) [43]	PRAKOBPHOL et coll (1982) [44]	GOLD et coll (1981) [45]	YUREWICZ et MOGHISSI (1981) [46]
Origine mucus	trachéo- bronchique	trachéo- bronchique	trachéo- bronchique	trachéo- bronchique	salivaire	colique	cervical
pathologie	mucovis- cidose	asthme	-	bronchite chronique	-	-	-
Acides aminés mol/100mol							
AS (X)	2,2	8,3	2,7	3,6	4,9	6,4	3,6
THR	28,7	18,1	27,5	24,4	20,9	20,1	26,6
SER	17,0	12,7	18,6	16,1	16,2	9,7	12,1
GL(X)	4,4	8,1	3,9	4,2	6,9	8,3	5,7
PRO	11,8	10,5	14,7	11,1	23,7	9,3	10,7
GLY	8,8	7,6	6,5	8,2	1,4	6,5	6,8
ALA	10,4	7,2	8,2	9,2	14,4	7,0	9,1
1/2 CYS	-	2,6	0,6	0,7	0,10	-	0,5
VAL	4,0	4,8	4,0	3,8	3,1	6,4	5,0
MET	traces	0,4	0,8	1,2	0,11	-	2,0
ILE	2,2	2,1	2,1	2,1	1,50	4,5	2,3
LEU	4,5	4,9	2,3	4,3	2,80	7,5	5,4
TYR	traces	1,8	1,1	1,1	0,14	1,8	0,3
PHE	traces	2,1	1,2	1,1	0,86	2,6	1,8
LYS	1,6	2,1	2,2	2,7	1,40	4,5	1,4
HIS	2,1	2,5	1,5	3,8	0,87	3,4	1,8
ARG	2,1	4,2	2,0	2,5	0,69	2,1	2,7
Acides aminés totaux (g/100g)	10,15	29	19,53	13	30,4	-	-
OH AA % Ser/Thr	45,7 0,59	30,8 0,70	46,1 0,68	40,5 0,66	37,1 0,78	29,8 0,48	38,7 0,45

Tableau III : Quelques séquences fragmentaires de peptides de mucinesmucines sous-maxillaires ovines, d'après HILL et coll (1977b) [48]

T ₁	1	Ser-	Ser-	Val-	Pro-	Gly-	Glu-	Ser-	Ala-	Thr-	10	Pro-	Gln-	Gln-	Pro-	Gly-	Ala-	Leu-
		Ser-	Glu-	Ser-	Thr-	Thr-	Gln-	Leu-	Pro-	Gly-	Val-	Thr-	Gly-	Thr-	Ser-	Ala-	Val-	
		Thr-	Gly-	Ser-	Glu-	Pro-	Gly-	Leu-	Pro-	Ser-	Thr-	Gly-	Val-	Ser-	Gly-	Leu-	Pro-	
		Gly-	Thr-	(Thr ₂ -	Ser ₄ -	Pro ₅ -	Gly ₄ -	Ala ₃ -	Val-	leu-	Phe ₂ -			Arg-				
T ₄ C	1	Ser-	Ala-	Gly-	Ala-	Gly-	Thr-	Ala-	Gly-	Val-	10	Asp-	Ser-	Glu-	Glu-	Thr-	Ala-	16
																		Arg
T ₄ F	1	Phe-	Ser-	Ser-	Ala-	Gly-	Ile-	Pro-	Ala-	Thr-	10	Pro-	Gly-	Ser-	Thr-	Thr-	Gly-	16
																		Arg

mucines sous-maxillaires de porc, d'après ECKHARDT et coll (1987) [49]

	1	Ile-	Ser-	Val-	Ala-	Gly-	Ser-	Ser-	Gly-	Ala-	10	Pro-	Ala-	Val-	Ser-	Ser-	Gly-	Ala-
		Ser-	Gln-	Ala-	Ala-	Gly-	Thr-	Ser-	Gly-	Ala-	Gly-	Pro-	Gly-	Thr-	Thr-	Ala-	Ser-	
		Ser-	Val-	Gly-	Val-													

mucines trachéobronchiques humaines, d'après ROSE et coll. 1989 [51]

TR3A	Trp-	Phe-	Asp-	Val-	Asp-	Phe-	Pro-	Ser-	Pro-	Gly-	Pro-	His-						
	Gly-	Gly-	Asp-	Lys-	Gln-	Thr-	Tyr-	Asn-	Asn-	Ile								
TR3B	Ala-	Glu-	Ser-	His-	Pro-	Glu-	Val-	Ser-	Ile-	Glu-	His-	Leu						
TR3C			Ser-	His-	Pro-	Glu-	Val-	Ser-	Ile-	Asn-	His-	Leu						
TR3D	Ala-	Ala-	Gly-	Gly-	Ala-	Val-	Cys-	Glu-	Gln-	Pro-	Leu-	Gly-	Leu					

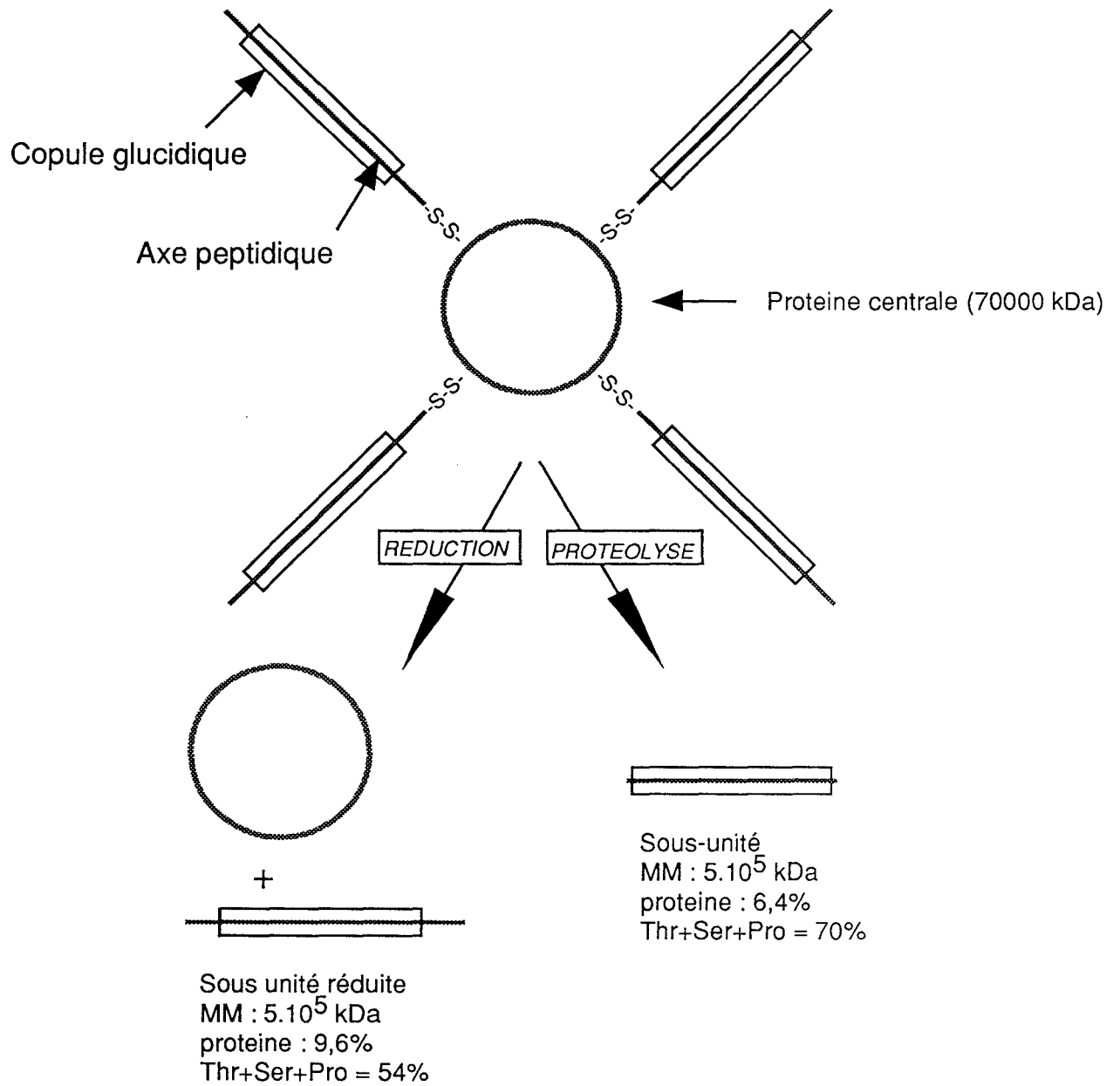
3) Organisation moléculaire: modèles structuraux

Cette caractéristique intrigante de polydispersion des masses moléculaires a été confirmée par la microscopie électronique qui permet de visualiser une polydispersion de longueur des mucines, écartant ainsi les hypothèses de dégradation protéolytique. Les données de microscopie électronique font apparaître les mucines cervicales, gastriques ou bronchiques comme de longs filaments de longueur variable, de 200 à 1 000 nm [52-55].

L'action de protéases (trypsine, pronase) sur les mucines révèle la présence de grands domaines (200 à 500 kDa) résistants à la protéolyse et correspondant à des domaines hyperglycosylés. Les points de sensibilité à l'action enzymatique sont eux situés dans des régions peu glycosylées dites "nues". C'est ainsi qu'est né le concept de l'organisation des mucines en alternance de domaines glycosylés et de domaines nus. Des expériences utilisant la microscopie électronique ont permis de visualiser sur les mucines des zones affines pour les lipides [56] ou pour l'or colloïdal [57]. Il a donc été conclu que les domaines nus étaient très vraisemblablement hydrophobes et irréguliers en taille. Différents modèles d'organisation structurale des mucines ont été proposés pour rendre compte à la fois de cette hypothèse et également de la sensibilité des mucines aux agents réducteurs. Deux modèles principaux, associant à cette structure alternée hydrophile-hydrophobe des pontages soit intermoléculaires, soit intramoléculaires par des ponts disulfures ont été proposés respectivement par Allen et par Carlstedt [58-60].

Le modèle de Allen [61] dit en "moulin à vent" propose, pour les mucines gastriques de porc, quatre sous-unités réunies à une protéine de liaison (70×10^3 D) grâce à des ponts disulfure (Figure 2).

Le modèle de Carlstedt, en accord avec les données de la microscopie électronique propose une structure linéaire où alternent des domaines "nus" et des domaines glycosylés enchaînés les uns aux autres par des ponts disulfure (Figure 3) [58, 33]. Ce modèle observé aussi bien pour les mucines cervicales et les mucines trachéobronchiques humaines [33] que pour les mucines intestinales de rat [62, 63], apparaît être le modèle le plus vraisemblable. Cependant la preuve absolue d'une telle organisation ne sera obtenue qu'après avoir élucidé totalement la séquence primaire d'une ou plusieurs mucines.



**Figure 2 : Modèle dit en "moulin à vent" proposé par A. Allen [59-61].
Structure de mucine gastrique de porc.**

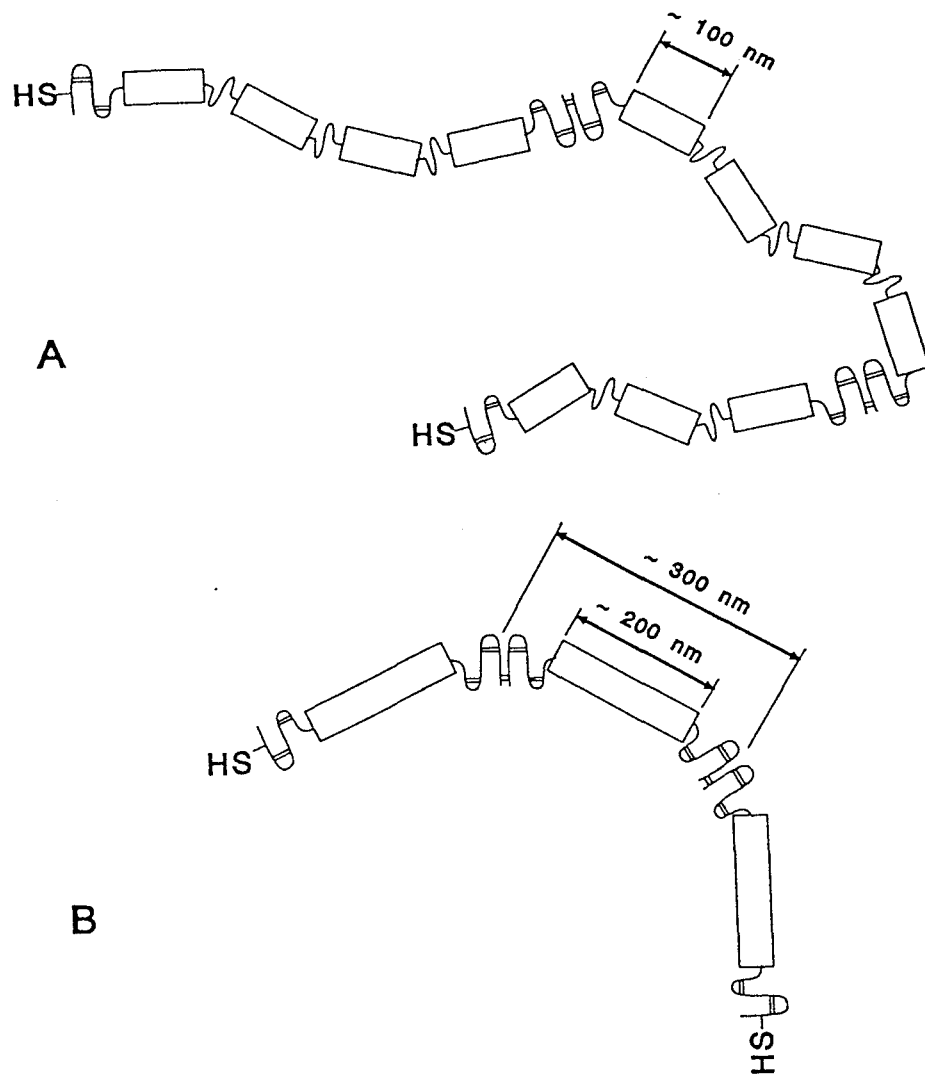


Figure 3 : Modèle proposé par Carlstedt.

A : Structure de mucine cervicale humaine [58].

B : Structure de mucine gastrique de rat [62].

4) Biosynthèse

Les mucines sont synthétisées puis sécrétées par des cellules exocrines épithéliales spécialisées.

La biosynthèse des mucines est encore mal connue. Ceci est surtout dû à l'absence d'outils, par exemple des anticorps, qui soient capables de témoigner de l'évolution des intermédiaires de biosynthèse.

La biosynthèse des mucines est un phénomène complexe impliquant la transcription des gènes de mucines, les mécanismes post-transcriptionnels, la traduction des ARNm, l'oligomérisation peptidique, les modifications post-traductionnelles comme la N-glycosylation, la O-glycosylation et peut-être d'autres, le transport intracellulaire, le stockage et enfin la sécrétion. Les étapes primaires de transcription et de traduction ainsi que leurs mécanismes de régulation sont encore inconnus. Seules les étapes post-traductionnelles ont été un peu plus étudiées.

Les chaînes peptidiques de mucines sont synthétisées dans les polyribosomes puis acheminées dans la lumière du RER. C'est à ce niveau que s'effectuerait une étape précoce de N-glycosylation, formant quelques chaînes N-glycanniques dont la fonction est encore imprécise. Seule une hypothèse a pu être émise, selon laquelle, la N-glycosylation contribuerait à l'orientation du peptide vers le Golgi. Les N-glycannes y seraient convertis en oligosaccharides complexes. L'initiation de la O-glycosylation est due à l'action d'enzymes de haute spécificité, les UDP-GalNAc-polypeptide- α -N-acétylgalactosaminyltransférases qui permettraient de lier les résidus de GalNAc à l'apomucine [64]. Aucune séquence-code permettant de signaler les sites potentiels de O-glycosylation n'a été mise en évidence sur le peptide. Cependant la présence de résidus proline à proximité des sérines et thréonines favoriserait par la conformation du peptide l'accès de la GalNAc-transférase. L'étape suivante d'élongation forme le squelette par addition successive d'un certain nombre d'unités disaccharidiques grâce à l'action de glycosyltransférases. L'étape finale consiste en la fixation au squelette de résidus tels que le fucose, le galactose, la N-acétylgalactosamine ou l'acide N-acétylneuraminique. Ce sont ces sucres périphériques qui confèrent aux mucines leur caractère polyanionique.

Certaines mucines, comme les mucines gastriques et les mucines cervicales formeraient un assemblage oligomérique grâce à des liaisons covalentes de type pont disulfure. Selon Dekker et Coll., l'oligomérisation pourrait intervenir dans le RER et dépendre de la N-glycosylation précoce [65].

Les mucines ainsi synthétisées sont stockées, compactées, dans des granules de sécrétion puis excrétées par exocytose à la surface apicale de la cellule. La sécrétion peut être stimulée rapidement sous certaines conditions, quand l'épithélium a besoin d'une sur-protection. La sécrétion du mucus est probablement sous contrôle neuronal et hormonal [66].

5) Régulation de la biosynthèse

La synthèse et la sécrétion des mucines doivent être soumises à des systèmes complexes de régulation quantitative et qualitative. Ces mécanismes permettent de maintenir une certaine épaisseur de mucus à la surface épithéliale et de maintenir constantes les propriétés du mucus assurant la protection de la muqueuse. Par ailleurs, il faut également que les propriétés des mucines puissent s'adapter aux variations physiologiques, comme c'est le cas, par exemple, durant la digestion, dans le tractus gastrointestinal. On peut alors se demander si au cours de ces changements, ce sont les mêmes polypeptides de mucines qui sont exprimés, et si des mécanismes post-traductionnels différents sont à l'origine de nouvelles propriétés physico-chimiques; ou si les modifications viscoélastiques produites sont dues à l'expression d'apomucines différentes. A quels niveaux s'effectue la régulation ?

Les sucres terminaux des chaînes glycaniques interviennent dans la reconnaissance des particules étrangères (virus, bactéries, allergènes...) suivant la vieille théorie "clef-serrure" ou plus précisément "sucre-lectines" [67]. Une certaine spécificité de reconnaissance portée par les sucres périphériques apparaît particulière à chaque individu. En effet, chaque individu ne possède pas le même degré de sensibilité à divers agents infectieux ou toxiques. Cette spécificité des sucres périphériques est-elle simplement due à une spécificité des glycosyltransférases ou à une spécificité d'expression des apomucines qui orienteraient différemment la O-glycosylation? Quel est ou quels sont les facteurs génétiques responsables de cette spécificité interindividuelle ? Les gènes de mucines, les gènes des glycosyltransférases ou les deux familles de gènes? L'exemple des antigènes de groupes sanguins serait plutôt en faveur d'une spécificité des glycosyltransférases; mais comme peu d'études analysent la relation apomucine/glycosylation, toutes les hypothèses sont envisageables.

III. LES APOMUCINES

Les techniques 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 de Edman, se sont révélées inopérantes dans l'étude des mucines pour de nombreuses raisons :

- l'accessibilité limitée des enzymes à leur cible peptidique en raison de la densité des chaînes glycaniques;
- la déglycosylation enzymatique trop partielle;
- la déglycosylation chimique efficace mais agressive pour le peptide (coupure des résidus de méthionine, hydrolyse acide);
- le fractionnement des peptides impossible en raison des très fortes homologues existant entre eux;
- et surtout, la grande labilité des dérivés phénylthiohydantoïnes engendrés par la dégradation de Edman qui rend impossible l'interprétation des résultats au delà du vingtième cycle.

Toutes ces raisons ont constitué des obstacles insurmontables lors de la détermination de la séquence peptidique des mucines.

Depuis 1989, certains laboratoires se sont tournés vers les techniques de l'ADN recombinant. Schématiquement, toutes les stratégies employées comportaient les étapes suivantes:

- préparation d'un anticorps polyclonal ou monoclonal dirigé contre l'axe peptidique, déglycosylé chimiquement par l'acide fluorhydrique ou l'acide trifluorométhane sulfonique, d'une mucine ou d'une "mucin like" préparée à partir d'un tissu sain ou tumoral;
- confirmation de la qualité des anticorps par immuno-transfert ou immunohistochimie;
- préparation d'une banque d'ADNc construite dans un vecteur d'expression à partir de la muqueuse ou de la lignée cellulaire étudiée;
- criblage immunologique de ces banques et extraction-purification des clones positifs;

- séquençage et localisation chromosomique.

Cette stratégie a permis l'obtention des premières séquences partielles d'ADNc de mucines animales isolées de glandes sous-maxillaires de porc [68], de boeuf [69], de muqueuse intestinale de rat [70] et de mucines humaines (intestinales, trachéobronchiques) ainsi que la séquence complète d'une mucine mammaire membranaire humaine [71-73].

L'analyse des séquences peptidiques déduites a permis de démontrer des traits structuraux communs à toutes les mucines :

- la richesse en sérine, thréonine, alanine, glycine, proline. Ces cinq acides aminés représentent 75% de l'ensemble des acides aminés. La sérine et la thréonine, sites potentiels de la O-glycosylation représentent de 25 à 75% de l'ensemble des résidus d'acides aminés.

- l'organisation en séquences peptidiques répétitives.

En 1973, Pigman et son équipe [69] émettaient l'hypothèse que l'axe peptidique de la mucine sous-maxillaire de boeuf était composée de plusieurs séquences répétées. Ce concept a été confirmé dans toutes les séquences partielles d'ADNc animales ou humaines publiées. En général, ces séquences répétées en tandem sont localisées dans la partie centrale de la chaîne polypeptidique et contiennent tous les sites potentiels de O-glycosylation; elles peuvent s'étendre sur plus de 2000 acides aminés.

A. MUCINES ANIMALES

La mucine la mieux connue actuellement est la mucine sous-maxillaire porcine [68, 74]. L'axe peptidique comporte quatre régions.

L'extrémité N-terminale comporte 120 résidus riches en sérine et thréonine. Cette région est suivie d'une région répétitive comportant 130 à 200 domaines élémentaires de 81 acides aminés. La région répétitive est suivie ensuite d'une séquence unique de 450 acides aminés essentiellement sérine, thréonine, glycolle, alanine. L'extrémité C terminale de structure globulaire contient 30 résidus de cystéine engagés dans des ponts disulfure. Seuls, les résidus de thréonine de cette molécule seraient glycosylés. Ces éléments structuraux permettent de prédire une masse moléculaire d'au moins 2×10^6 Daltons.

D'autres mucines animales sont beaucoup moins bien caractérisées (Tableau IV):

Référence	Origine	Dénomination	Taille du domaine peptidique (AA)	Séquence consensus
[68]	mucine sous-maxillaire porcine	PSM	81	GAGP GTTASSVGV TETARPSVAGSGSGTTG SGSTGSSSGSPGATGASIGQPETSRI RVAG AVSSGASQAAGTS
[69]	mucine sous-maxillaire bovine	BSM	11	GTTVAPGSSNT
[75]	mucine tégumentaire de <i>xenopus laevis</i>	FIM A.1 FIM B.1	9 11	VPTTPETTT GESTPAPSETT
[76]	mucine de truite "arc en ciel"	PSGP	13	DDATSEAATGPSG
[77]	homologue souris épisialine	-	21	DSASSPVAHGGTSSPATSPLR
[78]	leucosialine "mucin-like" rat	W3/13 antigen	9	T A S A DPPVT V I T A G M A (dégénérée)
[70]	mucine intestinale de rat	-	6	TTTPDV

B. MUCINES HUMAINES

Actuellement, la seule séquence d'ADNc complète connue est celle de l'épisialine, une mucine membranaire humaine, ainsi que celle de son homologue de souris [72, 77, 85-87].

1) *Episialine*

Les cellules épithéliales tumorales du pancréas, du sein, de l'ovaire expriment une même mucine membranaire, l'épisialine, appelée aussi MAM-6, PEM ou PUM [71-73]. L'axe peptidique de cette mucine est caractérisé par la répétition en tandem, 40 à 90 fois selon les sujets (polymorphisme interindividuel) [88], d'un motif de 20 acides aminés contenant 25% de résidus hydroxylés. Les molécules achevées diffèrent d'un tissu à l'autre seulement par quelques différences portant sur le nombre de chaînes oligosaccharidiques [85-87]. Les domaines peptidiques 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 transmembranaire, le deuxième transcrit coderait pour une forme sécrétée de l'épisialine. Ces mucines membranaires sont petites en comparaison des mucines sécrétées, et ne contiennent que trois résidus de cystéine par monomère. Aussi, sont elles probablement peu oligomérisées.

Le gène humain MUC1 (Figure 4) codant pour cette "mucin-like" est localisé sur le chromosome 1 (1q21-24) [89].

Tableau V : Séquences consensus des domaines répétitifs de mucines humaines.

NOM	A.A.	pb	SEQUENCE CONSENSUS	Ref.
MUC1	20	60	PDTRPAPGSTAPPAHGVTSA	[71]
MUC2	23	69	PTTTPITTTTTVTPTPTGTQT	[79]
MUC3	17	51	HSTPSFTSSITTTETTS	[80]
MUC4	16	48	TSSASTGHATPLPVD	[81]
MUC5A			dégénérée	[37, 82]
MUC5B	29	87	dégénérée	[37, 83]
MUC5C	8	24	TTSTTSAP	[37, 82]
MUC6	169	507	SPFSSSTGPMTATSFOTTTTTYPTPSHPOTTLPHTVPPFSTSL VTPSTGTYITPHTAOMATSASIHSTPTGTIPPPTTLKATGS THTAPPMTPTTSGTSQAHSSFSTAKTSTSLHSHTSSSTHPE VTPTSTTTITPNPTSTGTSTPVAHTTSATSSRLPTPFTHS PPTGS	[84]

2) Mucines sécrétées

♦ MUC2

Les travaux de J. Gum et Y. Kim [79, 80] ont permis d'identifier trois clones d'ADNc partiels (SMUC 40, 41, 42) isolés d'une banque d'ADNc de jéjunum humain grâce à un immunosérum dirigé contre l'axe peptidique des mucines d'une lignée tumorale colique (LS 174T). Ces trois clones possèdent le même arrangement répétitif d'un domaine élémentaire de 69 nucléotides, soit 23 acides aminés (Figure 5). Ces domaines ne sont pas strictement identiques dans les trois clones mais possèdent entre eux une homologie de 90%. Néanmoins, certains triplets sont strictement conservés dans tous les domaines. Les acides aminés thréonine et proline sont fortement représentés. On peut compter 14 résidus de thréonine par unité répétitive. Deux sites potentiels de N-glycosylation ont été relevés.

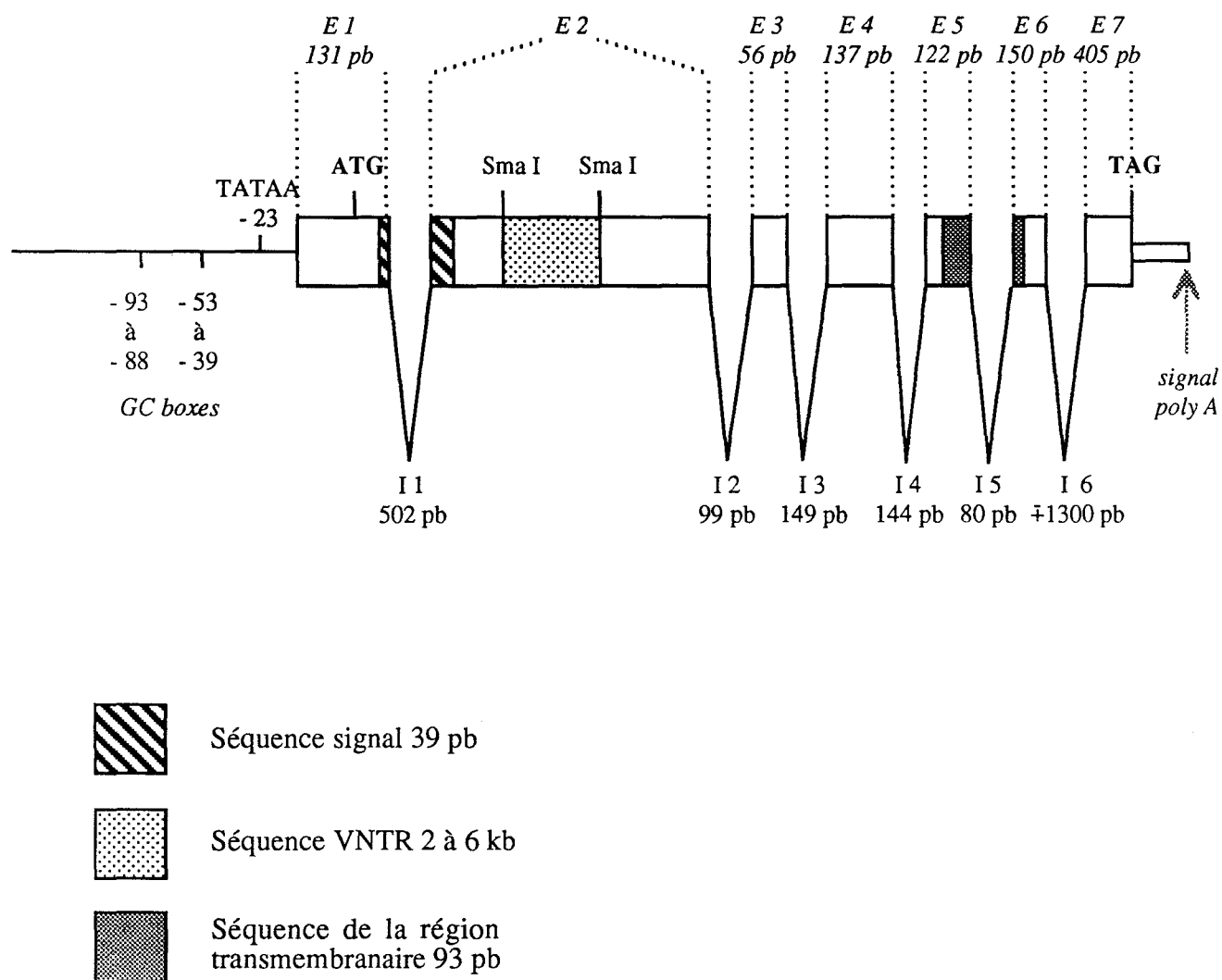


Figure 4 : Schéma de l'organisation du gène MUC1.

Le gène contient 7 exons (E); il a une taille de 4 à 7 kb qui varie en fonction de la taille de la région VNTR.

D'après S. J. Gendler [90].

A

	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	CCG	ACA	CCC	ATC	ACC	ACC	ACC	ACC	ACG	GTG	ACC	CCA	ACC	CCA	ACA	CCC	ACC	GGC	ACA	CAG	ACC
	6	CCA	ACC	CCG	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	CCG	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	CCG	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

B

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

Figure 5 : A : Unités nucléotidiques répétitives des clones de mucines intestinales [79].

B : Unités peptidiques répétitives des clones de mucines intestinales.

En 1990, cette même équipe a localisé à partir de ces clones, le gène MUC2, sur le chromosome 11 en p15 [91]. Des études de liaisons génétiques sur des familles du CEPH (Centre d'Etude du Polymorphisme Humain) ont révélé une liaison entre MUC2 et les gènes HRAS, INS, TH et HBBC de la région 11p15.

Ce gène est caractérisé par une expression typique en continuum d'ARNm. Il s'exprime à la fois dans l'intestin grêle, le côlon et les voies aériennes [92, 93].

Ce gène contient au moins deux domaines répétitifs majeurs [94]. Le plus grand de ces domaines contient environ une centaine de "tandem repeats" identiques à ceux décrits précédemment, de 23 acides aminés. L'analyse biochimique de cette séquence indique que 78% des thréonines sont glycosylées dans cette mucine [95]. On peut donc s'attendre à ce que cette région de 2300 acides aminés contienne approximativement 1000 chaînes oligosaccharidiques.

La seconde région répétitive de 347 résidus se caractérise par une grande richesse en résidus de thréonine. Sa séquence diffère des autres séquences répétitives de mucines; elle est composée de plusieurs "tandems" de longueur variable de 7 à 40 résidus.

Le domaine C-terminal de MUC2 de 984 résidus contient 9% de résidus de cystéine [96]. La séquence du domaine N-terminal n'est pas encore connue.

La mucine issue du gène MUC2 contient donc plus de 4500 acides aminés. Elle est formée de deux domaines riches en cystéine de part et d'autre de la région centrale répétitive (Figure 6).

♦ MUC3

La banque d'ADNc de jéjunum humain qui a permis d'isoler MUC2 a également été criblée à l'aide d'un autre immunsérum, dirigé contre l'axe peptidique de mucines intestinales humaines "saines". Gum et coll. ont ainsi isolé deux clones d'ADNc partiels (SIB 124,139) [80] caractérisés, eux aussi, par une organisation en domaines répétitifs. Le domaine élémentaire est différent du précédent, il comporte 51 nucléotides, soit 17 acides aminés.

Le gène correspondant à ces ADNc partiels est localisé sur le chromosome 7 en q22 [97], il répond à la dénomination de MUC3. Malheureusement, la connaissance de MUC3 se limite actuellement à la description de ses "tandem repeats".

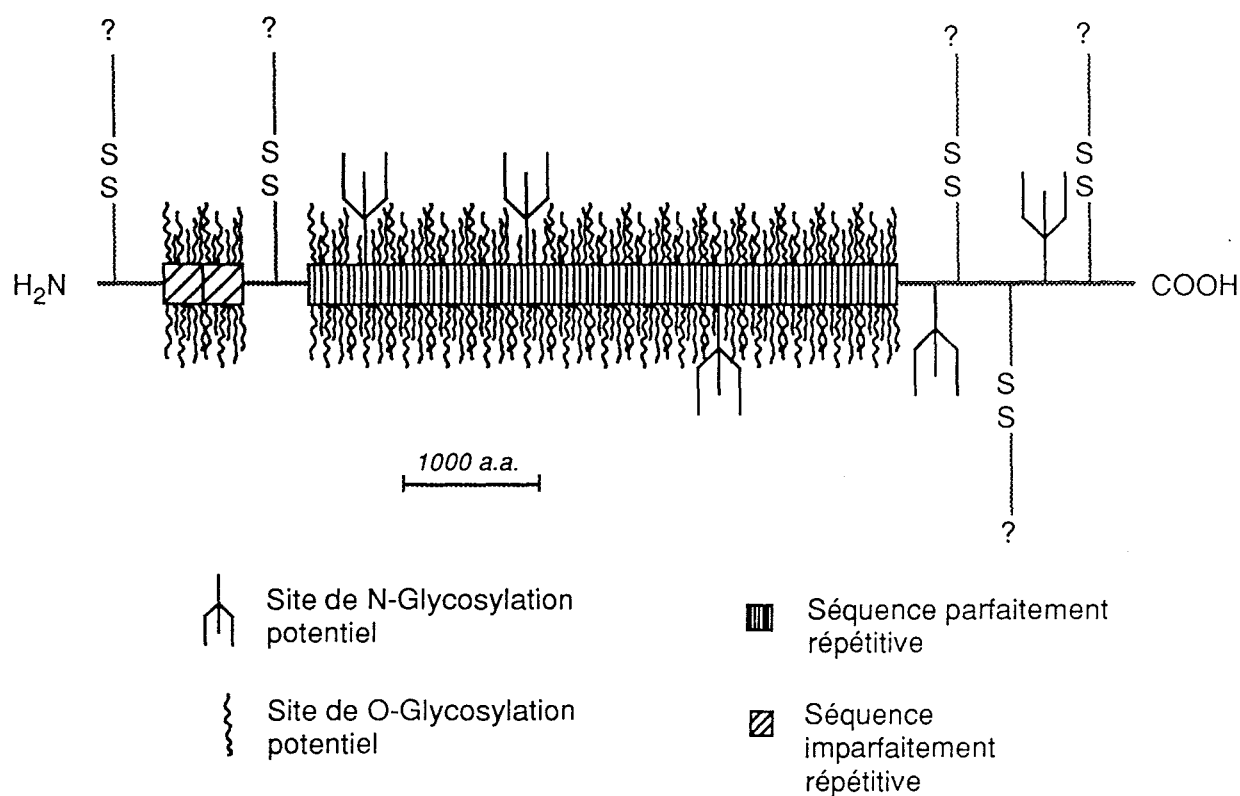


Figure 6 : Schéma de l'organisation de la mucine issue du gène MUC2 d'après N. W. Toribara [94].

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 ARNm codant pour ces mucines visualisée sous forme de continuum ou "smear".

Avant mon arrivée au laboratoire, l'équipe à laquelle je me suis intégrée avait déjà commencé, depuis 1989, l'étude des apomucines trachéobronchiques humaines par les techniques de biologie moléculaire (Figure 7) [37, 98-100]. Le criblage d'une banque d'ADNc de muqueuse trachéobronchique humaine grâce à un immunosérum polyclonal dirigé contre un mélange "équimolaire" de glycopeptides de mucines neutres, sialylées et sulfatées, tous déglycosylés chimiquement, avait permis d'extraire une vingtaine de clones [101]. Certains de ces clones avaient été séquencés, et ainsi révélé différents types d'organisations peptidiques, toujours basées sur un motif répétitif. La nomenclature proposée par le HGM comportait la distinction entre MUC4 et MUC5 car deux chromosomes distincts étaient concernés. Mais à l'intérieur du groupe "MUC5", trois sous-groupes étaient distingués MUC5A, MUC5B, MUC5C dans la mesure où aucune information ne permettait de prédire s'il s'agissait d'un, de deux ou de trois gènes distincts.

◆ MUC4

Le clone ADNc JER64 (1,83kb) possède un arrangement répétitif d'un domaine peptidique de 16 acides aminés, soit 48 nucléotides [81]. Les domaines ne sont pas rigoureusement analogues, néanmoins 8 positions sont parfaitement conservées d'un domaine à l'autre (Figure 8). La sonde JER64 a permis par hybridation in situ de localiser le gène MUC4 correspondant sur le chromosome 3, en q29 [81]. La sonde JER64 a détecté un important polymorphisme de restriction avec toutes les enzymes analysées (Bam HI, Hind III, Pst I, Eco RI et Taq I) [102]. Le polymorphisme observé de type "variable number of tandem repeats" (VNTR) est en rapport avec la présence de séquences nucléotidiques répétées en tandem dans le gène MUC4.

◆ MUC5A

Le clone d'ADNc JER47 a posé des problèmes de sous-clonage et de séquençage, pendant longtemps, seules les séquences des deux extrémités étaient connues (Figure 8). Elles sont constituées de séquences peptidiques déduites, très riches en thréonine mais non organisées en "tandem repeats". La séquence peptidique TR-3A décrite par ROSE [51] sur des préparations peptidiques issues de mucines trachéobronchiques humaines y est présente.

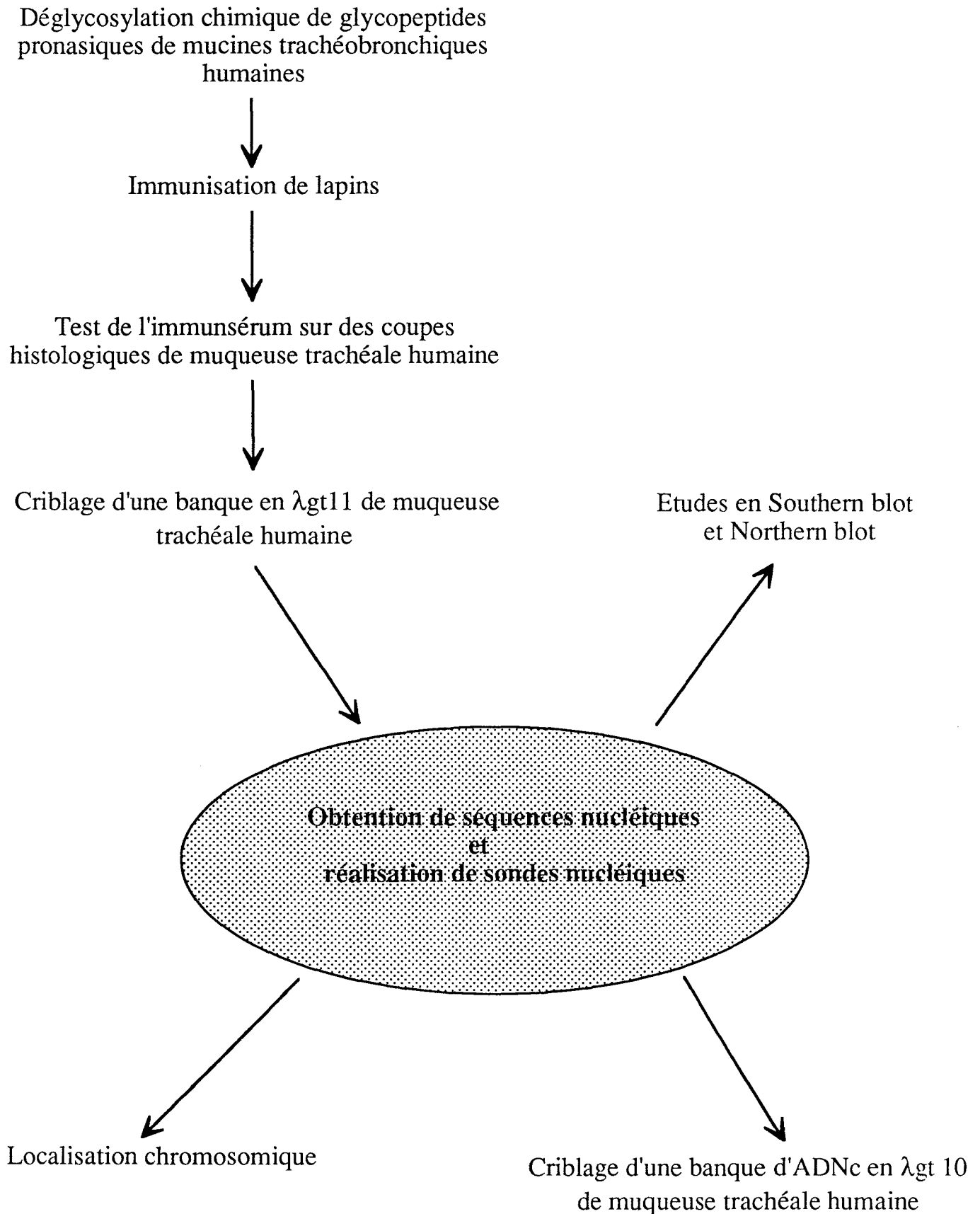


Figure 7 : Stratégie utilisée par notre équipe pour l'étude des apomucines trachéobronchiques humaines.

Le gène MUC5A correspondant à l'ADNc JER47 est localisé, par hybridation in situ, sur le chromosome 11 en p15 [103]. La sonde JER47 détecte par analyse en Southern blot de lignées somatiques hybrides (homme/rongeur), une bande majeure Bam HI/Bam HI de 21kb, localisée sur le chromosome 11 et également une bande mineure de 4kb localisée sur le chromosome 13. Le fragment mineur correspond probablement à une séquence apparentée aux mucines trachéales [103] présente sur ce chromosome 13.

◆ MUC5B

Deux clones JER57 (1.83kb) et JER28 (0.56kb) de séquences peptidiques déduites différentes, isolés de l'immunocriblage de la banque ADNc de muqueuse trachéobronchique humaine, présentent le même motif répétitif dégénéré de 87 nucléotides, soit 29 acides aminés [83]. Le criblage d'une banque en λ gt10, à l'aide de la sonde JER57, a permis d'isoler deux autres clones, JUL7 et JUL10, codant pour des peptides différents mais toujours organisés à partir du même domaine nucléotidique primaire.

Ces quatre clones sont caractérisés au niveau nucléique par une structure répétitive dégénérée et au niveau peptidique par une alternance de zones potentiellement glycosylables et de zones nues. Cette image des mucines était suspectée depuis longtemps mais les premiers résultats concernant la séquence primaire de ces macromolécules ne décrivaient jusqu'à présent que des séquences répétitives hautement glycosylables. Cette alternance peptidique dérive en réalité de la dégénérescence des "tandem repeats" modifiant le cadre de lecture de l'ARNm. Elle concerne des mutations ponctuelles d'un tandem à l'autre et également des insertions et délétions qui modifient la longueur du domaine [83]. Jusqu'à maintenant les séquences de mucines décrites MUC1, MUC2, MUC3, MUC4 et MUC5C présentaient une homologie presque parfaite entre les répétitions et une constance dans la longueur de ces répétitions. Ces quatre clones de la famille MUC5B ont donc suscité au sein du groupe un vif intérêt du fait de leurs particularités nucléiques et peptidiques. C'est pourquoi, lors de mon arrivée dans cette équipe, on me confia l'étude de l'organisation génomique de ces clones.

La connaissance génomique de JER57 se limitait en 1990 à sa localisation chromosomique. Comme MUC2 et MUC5A, JER57 est localisé en 11p15, il définit ainsi le locus MUC5B [103].

A

```

3       9       15      21      27      33      39      45
1 ACT TCC TCA GTA TCC ACA GGC CAC GGC ACC TCT CTT CCT GTC ACC GAC
49 ACT TCC TCA GCA TCC ACA AGT CAC GGC ACC TCT CTT CCT GTC ACT GAC
97 ACT TCC TCA GCA TCC ACA AGT CAC GGC ACC TCT CTT CCT GTC ACC GAC
145 GCT TCC TCA GTA TCC ACA GGT GAC ACC ACC CCT CTT CCT GTC ACC GAC
193 ACT TCC TCA GCA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
241 GCT TCC TCA GTA TCC ACA GGT GAC ACC ACC TCT CTT CCT GTC ACC GAC
289 CTT TCC TCA GTA TCC ACA GGT GAC GGC ACC CCT CTT CCA GTC ACC GAC
337 CCT TCC TCA GCA TCC TCA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
385 GCT TCC TCA CTC TCC ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
433 GCT TCC TCA GTA TCC ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
491 GCT TCC TCA GCA TCC ACA GGT CAC ACC ACC TCT CTT CCT GTC ACC GAC
529 GCT TCC TCA GCA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
577 ACT TCC TCA GTA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
625 GCT TCC TCA GTA TCC ACA GGT GAC GGC ACC CCT CTT CCA GTC ACC GAC
673 CTT TCC TCA GTA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
721 CCT TCC TCA GCA TCC TCA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
769 GCT TCC TCA GTA TCC ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
817 CCT TCC TCA GCA TCC TCT GGT GAC GGC ACC TCT CTT CCA GTC ACC GAC
865 CTT TCC TCA CTC TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
913 CTT TCC TCA GCA TCC ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
961 ACT TCC TCA GTA TCT ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
1009 GCT TCC TCA GTA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1057 GCT TCC TCA GTA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
1105 CTT TCC TCA GCA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
1153 ACT TCC TCA GTA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
1201 ACT TCC TCA GTA TCC ACA GGT GAC ACC ACC TCT CTT CCA GTC ACC GAC
1249 ACT TCC TCA GTA TCC ACA AGC CAC GGC ACC TCT CTT CCA GTC ACC GAC
1297 CCT TCC TCA GCA TCC ACA GGT GAC ACC ACC CCT CTT CCA GTC ACC GAC
1345 ACT TCC TCA GTA TCC ACA AGC CAC GGC ACC TCT CTT CCA GTC ACC GAC
1393 ACT TCC TCA GCA TCC ACA AGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
1441 ACT TCC TCA GCA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1489 ACT TCC TCA GCA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1537 ACT TCC TCA GCA TCC ACA GGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
1585 CCT TCC TCA GCA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1633 CCT TCC TCA GCA TCC ACA AGT CAC GGC ACC TCT CTT CCA GTC ACC GAC
1681 ACT TCC TCA GCA TCC ACA AGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1729 ACT TCC TCA GCA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1777 ACT TCC TCA GCA TCC ACA GGT CAC GGC ACC CCT CTT CCA GTC ACC GAC
1825 ACT TCC
    
```

```

Thr Ser Ser Val Ser Thr Gly His Ala Thr Ser Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Asp
Ala Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly His Ala Thr Ser Leu Pro Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Pro Leu His Val Thr Ser
Leu Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Ser
Pro Ser Ser Ala Ser Ser Gly His Ala Thr Ser Leu Pro Val Thr Asp
Ala Ser Ser Leu Ser Thr Gly His Ala Thr Ser Leu His Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Leu Leu His Val Thr Asp
Ala Ser Ser Ala Ser Thr Gly His Thr Thr Ser Leu Pro Val Thr Asp
Ala Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly Asp Thr Thr Pro Leu His Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Pro Leu His Val Thr Ser
Leu Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Ser
Pro Ser Ser Ala Ser Ser Gly His Ala Thr Ser Leu Pro Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Ser Leu Pro Val Thr Ile
Pro Ser Ser Ala Ser Ser Gly Asp Ala Thr Ser Leu Pro Val Thr Ser
Leu Ser Ser Leu Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Ser
Leu Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Asp
Thr Ser Ser Val Ser Thr Gly His Ala Thr Ser Leu Leu Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Pro Leu His Val Thr Asp
Ala Ser Ser Val Ser Thr Gly His Ala Thr Ser Leu His Val Thr Asp
Ala Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Ser
Pro Ser Ser Ala Ser Thr Gly Asp Thr Thr Pro Leu Pro Val Thr Asp
Thr Ser Ser Val Ser Thr Gly Asp Thr Thr Pro Leu Leu Val Thr Asp
Thr Ser Ser Val Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Asp
Thr Ser Ser Val Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Asp
Pro Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Leu Val Thr Asp
Thr Ser Ser Ala Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Ser
Pro Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Ser
Pro Ser Ser Ala Ser Thr Ser His Ala Thr Ser Leu Pro Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Leu Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Leu Val Thr Asp
Thr Ser Ser Ala Ser Thr Gly His Ala Thr Pro Leu Pro Val Thr Asp
Thr Ser
    
```

B

```

10       20       30       40       50       60
CCAAAGACCAGCACAAAGCCATGTTCCATATCCAAGACAACCCACTCCCAACCAGTCACCAGAGACTGT
ProLysThrSerThrSerHisValSerIleSerLysThrThrHisSerGlnProValThrArgAspCys

70       80       90       100      110      120      130
CATCTCCGGTGCACCTGGACCAAGTGGTTGACGTAGACTTCCCATCCCCTGGACCCACGGCGGGGAC
HisLeuArgCysThrTrpThrLysTrpPheAspValAspPheProSerProGlyProHisGlyGlyAsp

140      150      160
AAGGGAAACCTACAACAACATCATCAGG
LysGlyAsnLeuGlnGlnHisHisGln

10       20       30       40       50       60
CCCAAGGCTGCCCGTGACCTCCACATCTGTGACAGCTCCTAGCACCCCTAGTGGGAGAGCCACCAGC
ProLysGlyCysProValThrSerThrSerValThrAlaProSerThrProSerGlyArgAlaThrSer

70       80       90       100      110      120      130
CCAACTCAGAGCACCTCCTCTGGCAGAAATCCAGGACAACCACTTTGGTGACAAGCAGCATAACCTCC
ProThrGlnSerThrSerSerTrpGlnLysSerArgThrThrThrLeuValThrSerSerIleThrSer

140      150      160      170
ACTACACAGACCAGCACAACTCTGCCCTACAAG
ThrThrGlnThrSerThrThrSerAlaProThr
    
```

Figure 8 : Séquences nucléotidiques et peptidiques déduites de clones de mucines trachéobronchiques.

A : JER64 (MUC4)

B : JER47 (MUC5A).

◆ MUC5C

Parmi les 20 clones isolés de la banque d'ADNc, sept clones (JER36, 51, 54, 55, 58, 61 et 63) séquencés ou partiellement séquencés présentent une séquence répétée en tandem de 8 acides aminés de type **TTSTTSAP** plus ou moins conservée. Le locus correspondant MUC5C est lui aussi localisé en 11p15 [103].

Les sondes de mucines trachéobronchiques comme celles de mucines intestinales manifestent une expression tissulaire originale qui toujours revêt l'aspect "inédit" d'un large continuum d'ARNm.

◆ MUC6

La dernière sonde d'ADNc de mucine humaine a été isolée récemment en 1993, par le groupe de San Francisco [84]. A partir d'une banque d'ADNc commerciale préparée en vecteur d'expression à partir d'estomac humain, un clone a été isolé par immunocriblage. Sa structure nucléotidique, elle aussi répétitive, de 507 paires de bases, code pour un domaine peptidique de 169 acides aminés. Il s'agit du plus long domaine peptidique décrit.

Cette séquence s'hybride, elle aussi, au chromosome 11 en p15 [84].

De ces premières études de clones d'ADNc, il ressort qu'il existe une grande hétérogénéité structurale des axes peptidiques des mucines humaines qui présentent cependant, quelques caractéristiques communes. Le peptide primaire semble divisé en trois régions:

- les régions N- et C-terminales ne sont pas fortement O-glycosylées et contiennent tous les sites de N-glycosylation et tous les résidus de cystéine; elles sont constituées d'une séquence unique.

- la région centrale est formée de séquences répétées en "tandem" et possède tous les sites potentiellement O-glycosylables.

Cette subdivision des polypeptides de mucines est en accord avec les résultats des études de digestions protéolytiques. Toutes les mucines étudiées présentaient un glycopeptide résistant à la protéolyse, contenant les O-glycannes et une partie sensible à la protéolyse où étaient retrouvés les résidus de cystéine. Seules les mucines cervicales et trachéobronchiques font exception à ce modèle, par le fait qu'elles possèdent plus d'une touffe de glycosylation (Figure 3). Cette caractéristique est maintenant confirmée pour les mucines trachéobronchiques humaines par la séquence peptidique déduite de MUC5B qui démontre une alternance de zones glycosylables et de zones nues.

C. EXPRESSION DES GENES DE MUCINES

1) expression tissulaire

L'analyse de l'expression tissulaire par Northern blot montre pour l'ensemble des sondes, à l'exception de MUC1, une extrême hétérogénéité en nombre et en taille de transcrits. Le profil d'hybridation montre un continuum de plus de 20 kb à moins de 1 kb, d'une parfaite homogénéité, sans que l'on puisse deviner la moindre bande discrète. Cette homogénéité de signal est conservée même dans des conditions de stringence maximale. Elle est la même que l'expression soit forte (observable en quelques heures d'exposition autoradiographique) ou faible (décelable après plusieurs jours). Ces observations suscitent deux types d'interrogations : artefacts ?, mécanisme génétique d'expression nouveau ?

L'expression semble, avec la plupart des sondes, obéir à une loi du tout ou rien : expression très intense ou alors très faible selon le tissu étudié.

2) expression cellulaire

L'étude par hybridation in situ utilisant des sondes oligonucléotidiques marquées au soufre 35 et correspondant à chacun des gènes de mucines a été réalisée au laboratoire [104]. Il a ainsi pu être démontré une spécificité cellulaire de l'expression de ces gènes : par exemple, MUC4 s'exprime indifféremment dans les cellules caliciformes et les glandes mucipares de la plupart des muqueuses, alors que, au contraire, MUC5B a une expression tissulaire limitée aux glandes sous-maxillaires et aux bronches, et une expression cellulaire limitée aux glandes.

L'ensemble des résultats sont compilés dans le tableau VI suivant :

Tableau VI : Profil d'hybridation in situ des gènes MUC sur les muqueuses humaines d'après J. P. Audié [104].

MUQUEUSES HUMAINES	SONDES				
	MUC2	MUC3	MUC4	MUC5B	MUC5C
GL. SOUS-MAXILLAIRES	—	—	—	++	—
BRONCHE	E : ++ Gl : —	E : — Gl : —	E : ++ Gl : ++	E : — Gl : ++	E : +++ Gl : +++
FUNDUS	E : — Gl : —	E : + Gl : —	E : + Gl : —	E : — Gl : —	E : ++++ Gl : —
ANTRE	E : — Gl : +	E : ++ Gl : +	E : + Gl : —	E : — Gl : —	E : ++++ Gl : —
JEJUNUM	E : ++++ Gl : ++++	E : +++ Gl : —	E : + Gl : +	E : — Gl : —	E : — Gl : —
ILEON	E : ++++ Gl : ++++	E : +++ Gl : —	E : + Gl : +	E : — Gl : —	E : — Gl : —
COLON	E : ++++ Gl : ++++	E : + Gl : —	E : + Gl : +	E : — Gl : —	E : — Gl : —
ENDOCOL	E : + Gl : +	E : — Gl : —	E : ++ Gl : ++	E : + Gl : +	E : ++ Gl : ++

E : épithélium

Gl : glandes

Remarque : l'étude de MUC6 est en cours.

D. GENES DE MUCINES

Actuellement, en fonction des localisations chromosomiques au moins quatre gènes de mucines humaines sont suspectés, sept, si l'on considère les profils d'expression tissulaire; et on peut supposer que la découverte de sondes de mucines gastriques (dont la recherche est en cours dans notre laboratoire) ou de mucines cervicales puisse encore élever ce chiffre.

Tableau VII : Localisation chromosomique des différents gènes de mucines humaines.

Gène	mucine	Localisation	Année	Auteurs	Ref
MUC1	membranaire	1q21-q24	1987	Swallow D.M. et al	[89]
MUC2	intestinale	11p15	1989	Griffiths B. et al	[91]
MUC3	intestinale	7q22	1990	Gum J.R. et al	[97]
MUC4	trachéobronchique	3q29	1991	Porchet N. et al	[102]
MUC5A MUC5B MUC5C MUC5L	trachéobronchique	11p15 13	1990	V.C.Nguyen et al	[103]
MUC6	gastrique	11p15	1993	Toribara N. et al	[84]

Sur ces 7 gènes potentiels, plusieurs seraient localisés en 11p15. Les clones d'ADNc correspondant à MUC5A, MUC5B et MUC5C, MUC2, MUC6 sont caractérisés par une séquence peptidique différente et possèdent chacun, une expression tissulaire qui leur est propre. Pourtant, ils ont tous la même localisation génomique en 11p15. Sont-ils tous issus d'un même gène volumineux dont l'expression serait différente d'une muqueuse à l'autre; ce qui impliquerait de nombreux promoteurs, un épissage alternatif complexe ? ou sont-ils chacun issus d'un gène différent soumis à une régulation tissulaire propre? La vérité se situe-t-elle entre ces deux extrêmes ? Plus simplement, ceci pose la question : combien y a-t-il de gènes de mucines en 11p15 ?

Ces multiples questions, le fait que potentiellement plusieurs gènes de mucines soient concentrés en 11p15, nous ont incités à étudier plus particulièrement cette région du génome humain. L'étude cartographique de cette région a donc fait l'objet d'une grande partie de mon travail de thèse.

Les gènes MUC1, MUC2 et MUC4, les seuls pour lesquels une étude génétique de polymorphisme ait été effectuée, ont révélé un important polymorphisme de restriction. Ce polymorphisme semble dû à la variation, d'un sujet à l'autre, du nombre des "tandem repeats" (VNTR). Ces variations génétiques dans le nombre de nucléotides répétés ont déjà été décrites pour des séquences d'ADN "minisatellite" non-codantes, fréquemment trouvées dans les génomes eucaryotes [105]. Les séquences répétées des mucines sont uniques par le fait qu'elles font partie de la séquence codante du gène. Dans MUC1, toutes les répétitions sont situées dans un exon [86]. Le seul autre exemple de gène humain qui soit caractérisé par un tel polymorphisme, dû à la présence de séquences répétées en tandems, sont les gènes de la famille des P.R.P (protéines riches en proline) [106-109]. Ces gènes localisés sur le chromosome 12 sont formés de 5 à 15 tandems de 63 nucléotides, tous situés dans un exon. Ils montrent un polymorphisme inter-individuel, probablement dû à des événements de crossing-over inégaux, semblables à ceux postulés pour l'ADN "minisatellite".

Les variations alléliques des gènes de mucines seraient également la conséquence d'évènements de crossing-over dans les séquences répétées. Ces événements affecteraient uniquement le nombre de répétitions. En fonction des allèles, le nombre de tandems peut varier énormément. Pour MUC1, par exemple, il a été observé en fonction des allèles, 20 à 125 tandems. Swallow et Coll. ont démontré que les variations alléliques du gène MUC1 corrélaient directement avec le polymorphisme protéique [88]. Les variations du nombre de tandems vont modifier la taille du polypeptide et la longueur de la région glycosylée, deux facteurs responsables de la propriété des mucines à former un gel. Le polymorphisme des gènes de mucines affecteraient donc les propriétés rhéologiques du mucus. Ce polymorphisme pourrait également être utilisé comme marqueur RFLP associé aux pathologies des gènes de voisinage en 11p15 : susceptibilité au cancer, maladies métaboliques, maladies mentales.

IV. LE CHROMOSOME 11 ET SA REGION 11P15

A. LE CHROMOSOME 11

Le chromosome 11 est l'un des plus étudiés après les chromosomes 1 et X: 153 gènes clonés y ont été recensés par le HGM11 (Londres 1991). En effet, le chromosome 11 est connu pour contenir de nombreux gènes responsables de maladies génétiques ou des gènes de susceptibilité aux maladies mentales, métaboliques, au cancer.

Trois régions 11p15, 11p13, 11q23 sont porteuses de gènes impliqués dans la croissance et la différenciation cellulaire. Des mutations germinales en font des gènes de susceptibilité au cancer, des mutations somatiques : délétions, pertes d'allèles, duplications, translocations, déclenchent le processus de tumorigénèse. A la fois des oncogènes, des gènes suppresseurs de tumeurs, des facteurs de croissance, sont concernés par ces anomalies chromosomiques; un certain nombre de ces gènes sont encore à cloner.

Ces gènes qui suscitent un grand intérêt dans le monde médical et scientifique sont à l'origine de la recherche de nouveaux marqueurs tout le long du chromosome afin de construire une carte à la fois physique et cytogénétique qui contribuerait ainsi à la caractérisation moléculaire de ces gènes. C'est dans cette optique que Hori et coll. ont isolé 168 marqueurs cosmidiques [110] et Gerhard et coll. 1001 nouveaux marqueurs du chromosome 11 (1992) [111]. Les équipes du CEPH et du Généthon dans le cadre du programme "Génome Humain" ont publié [112-114], en 1992, pour le chromosome 11, 72 nouveaux loci couvrant ainsi 161cM sur les 245cM, longueur génétique estimée du chromosome 11. Sa taille physique est de 144Mb, il y a donc 1 marqueur toutes les 2Mb (Figures 9 et 10).

B. LA REGION 11P15 : CARTOGRAPHIE

La région 11p15, en situation télomérique, est particulièrement riche en gènes comme d'ailleurs toutes les régions télomériques [116]; elle est la région du chromosome la plus impliquée dans les phénomènes tumoraux et les maladies génétiques.

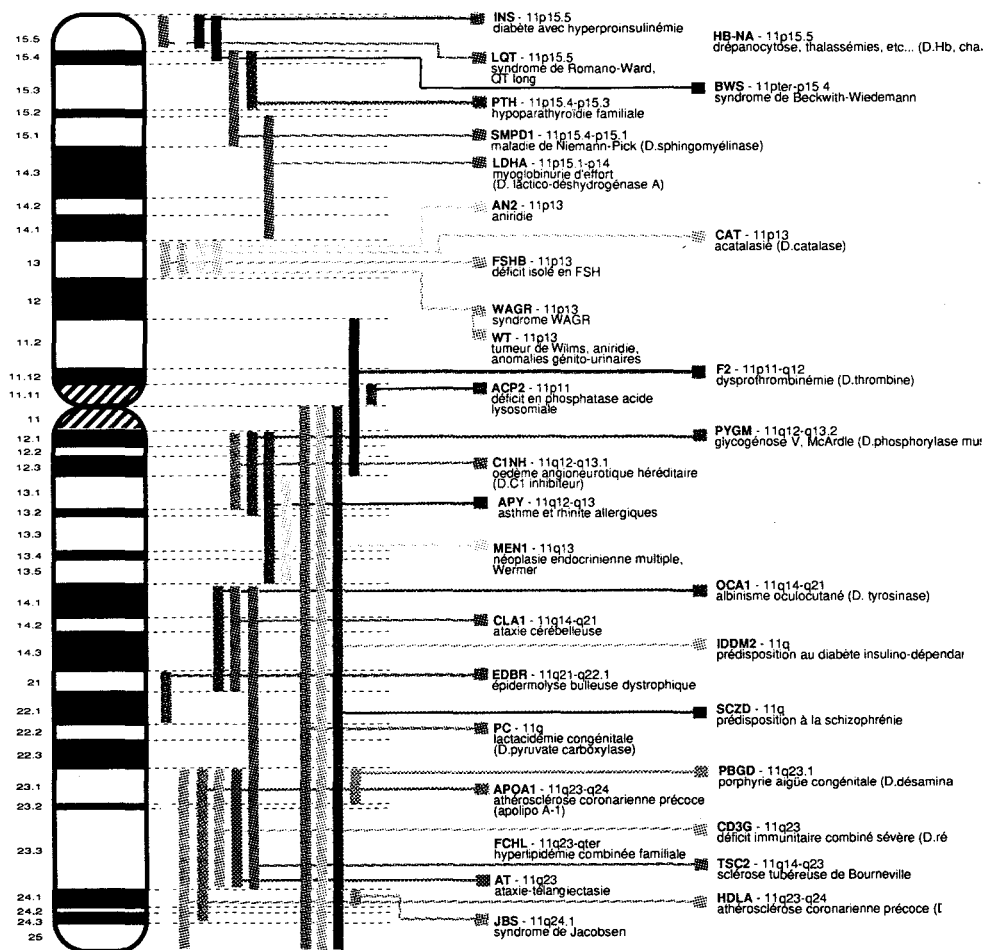


Figure 9 : Carte des locus morbides du chromosome 11 humain
d'après J. C. Kaplan [115].

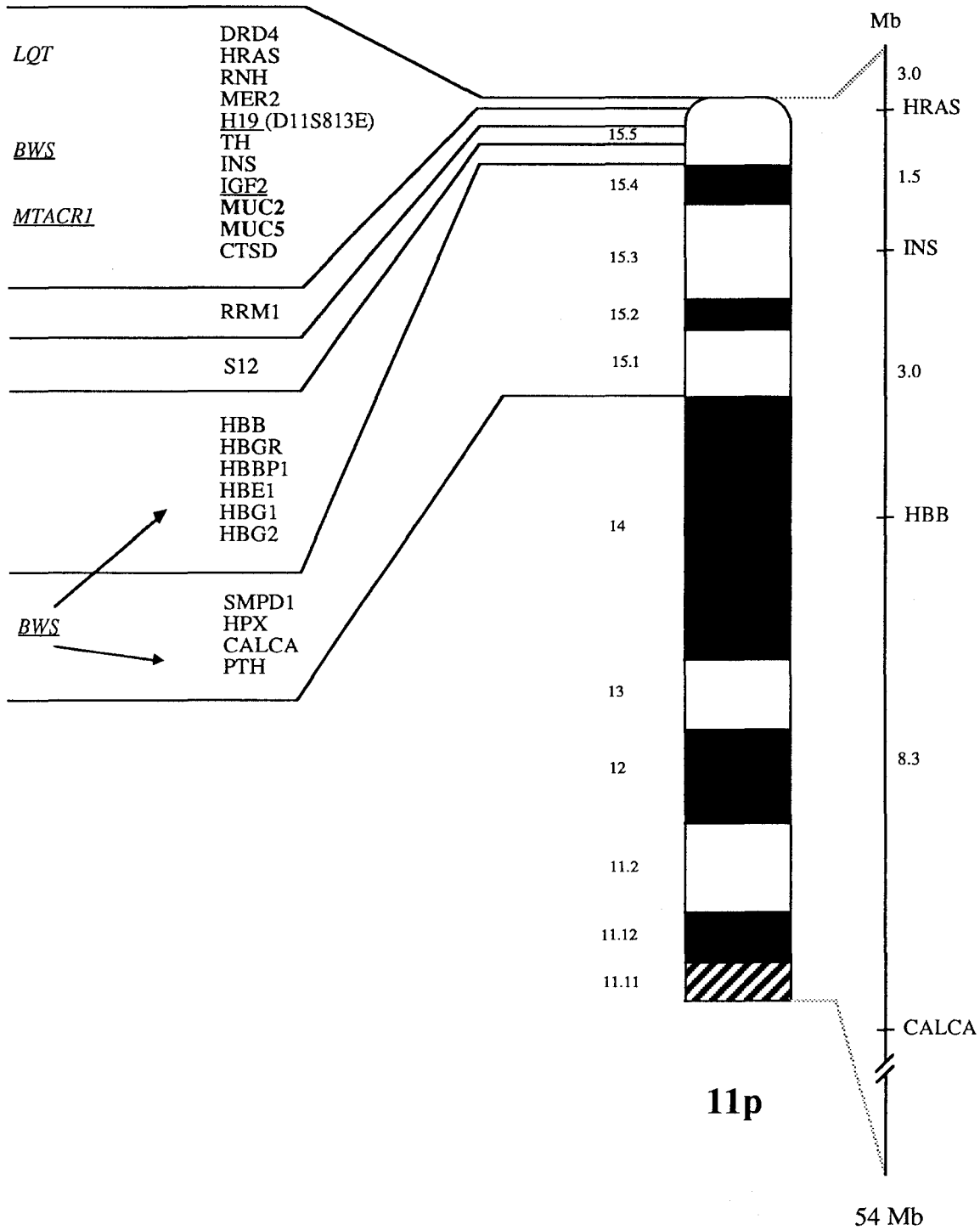


Figure 10 : Bras court du chromosome 11 humain
 d'après C. Junien et V. Van Heyningen[117] (HGM 11. Londres 1991).
 Les loci soulignés sont soumis à une empreinte parentale.

les marqueurs de référence de la région 11p15 sont, du télomère vers le centromère [117] :

HRAS	p15.5	oncogène H. RAS
INS	p15.5	insuline
D11S454	p15.5) sondes anonymes
D11S12	p15.5	
HBB	p15.5	β globine
PTH	p15.2 - 15.1	parathormone

Les autres gènes identifiés et localisés dans cette région sont :

- le locus de la tyrosine hydroxylase (TH) et le locus de l'IGF2 situés à proximité du locus INS [118],
- le locus de l'inhibiteur de ribonucléase (RNH) situé à 90 kb de HRAS [119, 120, 121],
- le locus du récepteur D4 de la dopamine (DRD4) distal de 2cM de HRAS [122],
- un gène apparenté au gène de la kératine riche en soufre (KRN1L) localisé en 11p15.5 [123],
- un gène de protéine lymphocytaire (PNIA4) localisé en 11p15 [124],
- le gène de la sphingomyélinase (SMPD1) localisé en p15.4 - 15.1 [125]
- le gène de la cathepsine D localisé en 11p15.5 [126].

Les ADNc MUC5A, 5B, 5C ont été localisés en 11p15 (p15.3 - p15.5) par hybridation in situ [103].

L' ADNc MUC2 a également été localisé par hybridation in situ en 11p15. Par analyse de liaison génétique, il a pu être localisé précisément à proximité des loci INS et TH (0% de recombinaison méiotique) alors que les loci HRAS et HBB sont plus éloignés [91] (Figure 11).

Les cartes comparées du chromosome 11 humain et des régions homologues de la souris montrent qu'il existe 5 groupes synténiques correspondant au bras court du chromosome 11 (Figure 12) [117]. Cela signifie qu'il existe une conservation en terme de distances physiques mais aussi d'organisation, de fonctionnalité, peut être de régulation, au cours de l'évolution pour les gènes d'un même groupe. Le gène de mucine MUC2, pourrait donc se situer dans le groupe synténique désigné par TH/IGF2/H19 (Figure 12).

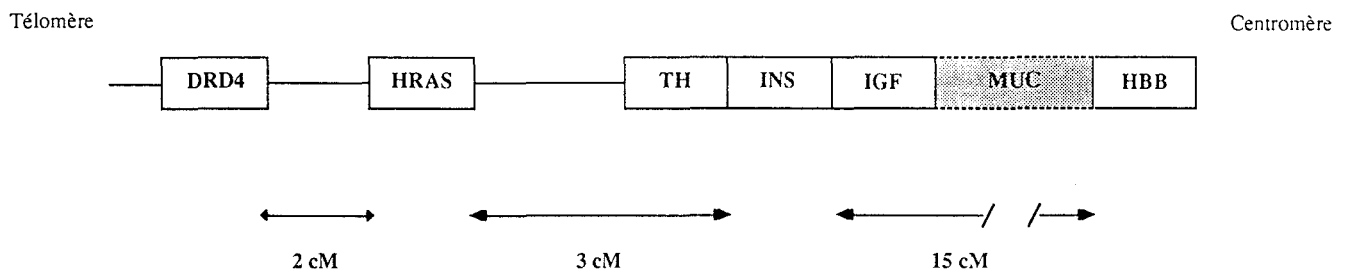


Figure 11 : Schéma de la région 11p15.5 d'après C. Junien et V. Van Heyningen [117].

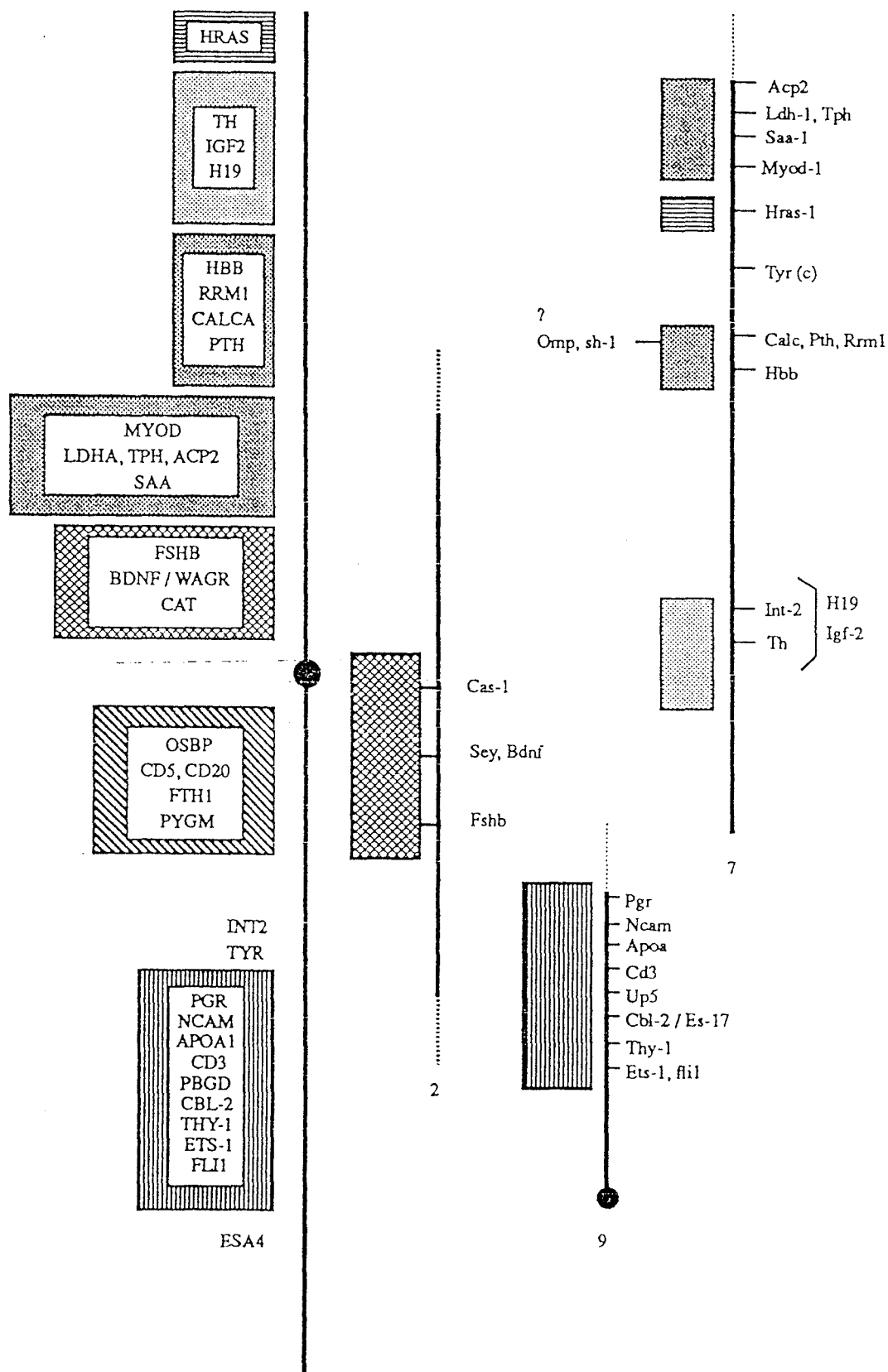


Figure 12 : Carte comparative du chromosome 11 humain et des régions homologues chez la souris d'après C. Junien et V. Van Heyningen [117].
Chaque rectangle différent représente un groupe synténique conservé.

C. LA REGION 11P15 : GENETIQUE

Trois grandes disciplines médicales sont intéressées par la génétique de la région 11p15 : endocrinologie, psychiâtrie, cancérologie.

Le polymorphisme du gène IGF2, la région VNTR située en amont du promoteur du gène de l'insuline apportent des marqueurs de prédisposition au diabète, à l'obésité [127]. Les anomalies génétiques affectant les gènes de la parathormone et de la calcitonine relèvent de la cancérologie.

Le polymorphisme du gène de la tyrosine hydroxylase, du gène du récepteur D4 de la dopamine de même que d'autres loci non identifiés entre HRAS et INS semblent impliqués dans les maladies mentales : schizophrénie, autisme, syndromes dépressifs, retards mentaux [128, 129].

Le cadre cancérologique est de loin le plus intéressé par les anomalies génétiques survenant dans la région 11p15 :

	Gène de prédisposition	Pertes d'allèles
Rétinoblastome	13q14	13q14
Ostéosarcome	13q14	13q14
Néphroblastome		11p13, 11p15
	(WAGR)	11p15
	(BWS)	11p15
	(familiale)	16q
Rhabdomyosarcome	(BWS)	11p15
Corticosurrénaome	(BWS)	11p15
Hépatoblastome	(BWS)	11p15
Cancer de la vessie	?	11p15
Von Hippel Lindau	3p	
Cancer rénal		3p
Neurinome de l'acoustique (NF2)	22q	
Méningiome		22q
Neurofibromatose	17q11	
MEN2A	10q	
Cancer médullaire thyroïde		1p
Phéochromocytome		1p, 22
MEN1	11q	
Insulinome		11q
Polypose colique familial (FAP)	5q21	
Carcinome colorectal		5q, 17p, 18p, 22
Cancer du sein	?	13q, 11p
Cancer du poumon	?	
à petites cellules		3p, 13q, 17p
autres types		3p (13q, 17p)
Mélanome	1	multiples
Sclérose tubéreuse de Bourneville (TSC)	9q, 11q	
Ataxie téléangiectasie (AT)	11q	

Tableau VIII : Localisation des gènes de susceptibilité au cancer, associés ou non à un syndrome malformatif, et pertes d'allèles dans les tumeurs [130].

L'analyse de liaison génétique a permis de définir l'existence de nombreux oncogènes (H.RAS) ou gènes suppresseurs de tumeurs (antioncogène 3, ADCR, BCCR₃). D'autres n'ont pas encore été clonés ou n'ont pas reçu de localisation physique précise :

1) Les gènes *WT1* et *WT2*

Le locus *WT1* (11p13) et *WT2* (11p15.5) [131] sont impliqués simultanément dans des mécanismes d'initiation et de progression tumorale responsables de :

- la tumeur de Wilms (*WT*) ou néphroblastome de l'enfant,
- le syndrome de Beckwith-Wiedeman (*BWS*), caractérisé par un syndrome malformatif multi-organe associé à une prédisposition au cancer (néphroblastome, corticosurréalome, hépatome, rhabdomyosarcome) [132],
- le néphrome mésoblastique congénital, autre tumeur infantile,
- le syndrome *WAGR*, c'est à dire associant aniridie, malformations génito-urinaires et retard mental [133].

Ces gènes *WT1* et *WT2* pourraient être des gènes suppresseurs de tumeurs. Selon la théorie de Knudson, un premier événement germinale constituerait une susceptibilité au cancer, transmise génétiquement, le deuxième événement somatique déclencherait le processus de tumorigénèse [134]. Un troisième gène *WT3* localisé sur le chromosome 16 [135] participerait aussi à cette cascade d'évènements [136].

	Gènes concernés		
	<i>mutations germinales</i>	<i>mutations somatiques</i> ⇒ <i>perte d'allèles</i>	<i>surexpression</i>
WT Familiaux	WT3	WT2	
WT Bilatéraux		WT2 + WT1?	
WAGR	WT1	WT2	
BWS	WT2	WT2	
WT Sporadiques		WT2 + WT1	
Néphroblastome congénital	?	?	IGF2

Ces gènes ou d'autres gènes suppresseurs de tumeurs localisés en 11p15 pourraient être également responsables de rhabdomyosarcome embryonnaire [137], carcinome de la

corticosurrénale [138] observés de façon sporadique ou familiale, pouvant être associés ou non au syndrome de Beckwith-Wiedeman.

2) Autres oncogènes ou gènes suppresseurs de tumeurs

Tumeurs de la vessie : des mutations de l'oncogène H.RAS ainsi que des pertes d'hétérozygotie en 11p15 leur sont fréquemment associées [139].

Tumeurs mammaires : des pertes d'allèles, translocations, délétions dans la région de H.RAS sont observées chez l'homme et chez le rat (tumeurs expérimentales) [140].

Cancers du poumon : (carcinome à cellules squameuses, cancer à petites cellules, cancers non à petites cellules) : des pertes d'allèles en 11p15, près du locus IGF2, ou près du locus de la tumeur de Wilms sont observées [141, 142].

Tumeurs neuroendocriniennes : l'activation de l'oncogène Ttg1 (11p15) serait impliquée dans la genèse de ces tumeurs [143].

Tumeurs parathyroïdiennes, thyroïdiennes : l'activation d'un oncogène situé dans le locus D11S 287E serait en cause [144]. Par ailleurs, des réarrangements du gène de la parathormone aboutissant à la perte de régulation du gène PTH (11p15.2) ont été observés dans des adénomes parathyroïdiens sporadiques ou associés au MEN1 [145].

Tumeurs osseuses : plusieurs translocations impliquant toujours la région 11p15 ont été observées dans les tumeurs osseuses à cellules géantes.

Leucémies : une famille d'oncogènes des cellules T, les rhombotines, responsables de leucémies, ont été localisées en 11p15 (Rhom1) et en 11p13 (Rhom2). Une autre classe de protéines riches en soufre (gène Ttg2, 11p15) constituent aussi des facteurs de croissance lymphocytaire [146, 147].

Tumeurs cervicales : un gène suppresseur de tumeur HTS1 voit l'expression de l'un de ses trois messagers (l'ARNm de 2.8 kb) disparaître dans certaines lignées de carcinome cervical [148].

Tumeurs testiculaires : le locus MTACR1 montre une perte d'hétérozygotie dans de nombreux cas de tumeurs testiculaires [149].

3) facteurs de croissance ou de différenciation

Outre le gène IGF₂, des facteurs de croissance et de différenciation, impliqués dans le développement embryonnaire des nerfs crâniens et la morphogénèse des muscles squelettiques (gène ASM1) ont été localisés en 11p15 [150]. Des translocations impliquant la région 11p15 ont pour conséquence sur ces tissus le développement de tumeurs embryonnaires.

4) Méthylation - empreinte génomique parentale

La régulation de l'expression des gènes de la région 11p15 est soumise, à un degré élevé, au phénomène de méthylation. Un exemple concret de l'importance de la méthylation est celui donné par le gène de la calcitonine (CALCA) situé en 11p15 à proximité du locus PTH. Le niveau de méthylation de ce gène est un bon marqueur du niveau général de méthylation des gènes de la région. Dans certaines leucémies (leucémies aiguës myéloblastiques, leucémies myéloïdes chroniques), il existe une bonne corrélation entre l'hyperméthylation du gène CALCA et l'accélération de la transformation tumorale [151]. Ce phénomène est également observé dans la tumeur de Wilms et le rétinoblastome où la méthylation des îlots CpG est associée à la prédisposition du gène à la tumorigénicité [152, 153].

La méthylation intervient également dans un cadre plus complexe de régulation de l'expression des gènes de la région 11p15, celui de l'empreinte génomique parentale [130, 154]. Ce marquage différentiel du génome paternel et maternel affecte la région 11p15 humaine et chez la souris, le groupe synténique homologue au groupe humain TH-IGF2-H19 [117].

Les preuves de ce phénomène sont données par l'observation spécifique de pertes d'allèles maternels dans la tumeur de Wilms, le rhabdomyosarcome, le syndrome de Beckwith-Weideman, et de pertes d'allèles paternels dans toutes les tumeurs testiculaires associées à une perte d'hétérozygotie des marqueurs du locus MTACR1 [149].

En conclusion, la région 11p15 apparaît comme un point chaud de mutations, délétions, réarrangements, observés dans de nombreux types de cancer. En outre, cette région comporte un certain nombre d'oncogènes et de gènes suppresseurs de tumeur interagissant entre eux, par les produits des gènes ? par une synergie ou une concertation de régulation ?

interagissant également avec des gènes de la région 11p13. La méthylation et l'empreinte génomique parentale ajoutent encore à la complexité de régulation de cette région.

Dans ces conditions, comment ne pas imaginer que les gènes "MUC", dont on connaît déjà la relation au cancer, et qui sont situés dans cette région, ne soient pas concernés eux aussi par les mêmes événements de mutations ou de dérégulation qui touchent les gènes de voisinage ?

STRATEGIE

STRATEGIE

Nous avons choisi d'étudier la région 11p15 qui rassemble les gènes correspondant aux ADNc MUC2, MUC5A, MUC5B, MUC5C et MUC6 et dont le mode d'organisation génomique est inconnu. Afin de relier physiquement ces séquences et d'établir ainsi une carte moléculaire de cette région, notre recherche a suivi deux axes:

I. la réalisation d'une **mégacartographie de restriction** de la région 11p15, grâce à des enzymes coupant rarement et à la séparation électrophorétique en champ pulsé;

II. la **marque sur le chromosome** à l'aide de sondes correspondant à MUC5A, MUC5B, MUC5C, pour caractériser l'un des gènes de mucines trachéobronchiques humaines.

L'élaboration de la stratégie a constitué une partie importante de mon travail de thèse.

I. MEGACARTOGRAPHIE DE RESTRICTION

La mégacartographie était une technique toute nouvelle au laboratoire et un domaine encore inexploré dans le cas des gènes de mucines. Il a donc fallu dans un premier temps, mettre au point cette technique, qui a consisté :

- à choisir parmi les lignées cellulaires dont nous disposons celles, dont l'analyse de l'ADN par électrophorèse en champ pulsé, donnaient des profils interprétables (absence d'anomalie du caryotype sur le chromosome 11 en p15) ;

- à choisir des endonucléases de restriction permettant de localiser un gène (localisation des îlots CpG);

- à choisir parmi les nombreuses techniques de séparation par électrophorèse en champ pulsé la plus adéquate en fonction de la taille des fragments que nous voulions étudier.

Cette stratégie avait trois objectifs qui étaient de déterminer : le **nombre** de gènes par le nombre d'îlots CpG, la **taille** de ces gènes, et la **distance** entre ces gènes.

Depuis quelques années, des techniques de clonage de grands fragments d'ADN ont été développées grâce à l'électrophorèse de l'ADN en champ pulsé. Les nouvelles techniques électrophorétiques utilisant un champ pulsé permettent ainsi de faire le lien entre l'étude de l'ADN, par analyse génétique (analyse de liaison, cytogénétique), et l'étude de la structure fine de nombreux gènes pris individuellement, par les méthodes de clonage (Figure 13). Particulièrement dans notre cas, cette technologie établit le lien entre la localisation commune des ADNc de mucines en 11p15 et leurs séquences différentes.

Ce type d'électrophorèse permet de séparer des molécules d'ADN de grandes tailles, pouvant aller actuellement jusqu'à quelques millions de paires de bases avec une très bonne résolution. Elles ont d'ores et déjà abouti à l'établissement de cartes précises de régions importantes de chromosomes à l'échelle de la mégabase, comme celle du complexe HLA chez l'homme [155, 156], de la myopathie de Duchenne [157, 158] et du chromosome d'*Escherichia coli* [159] de même qu'à la séparation de petits chromosomes comme ceux de la levure [160].

A. PRINCIPE DE L'ELECTROPHORESE EN CHAMP PULSE

Les techniques classiques d'électrophorèse sont fondées sur l'effet tamisant des supports solides utilisés (gel d'agarose) qui joue le rôle de différentiel de migration, les molécules d'ADN étant d'autant plus freinées qu'elles sont plus grandes. Il y a alors proportionnalité entre la taille et la mobilité des fragments d'ADN [161]. Ce type de technique permet de séparer des fragments d'ADN de taille pouvant aller jusqu'à quelques milliers, voire quelques dizaines de milliers de paires de bases. Au delà de 50 000 pb, la taille des fragments est supérieure au diamètre des plus larges pores du support (un fragment de 50 kb a une longueur de 18 microns, les pores d'un gel d'agarose à 1 % ont un diamètre de 1 micron). Tout effet de tamisage différentiel est alors annulé, et la mobilité électrophorétique de l'ADN ne dépend plus de sa longueur.

C'est Schwartz et Cantor [162] qui en 1982-83 proposent pour améliorer la séparation des grands fragments, de modifier un tout autre paramètre que le support: la nature du champ électrique. Ils sont à l'origine de l'électrophorèse en champ pulsé.

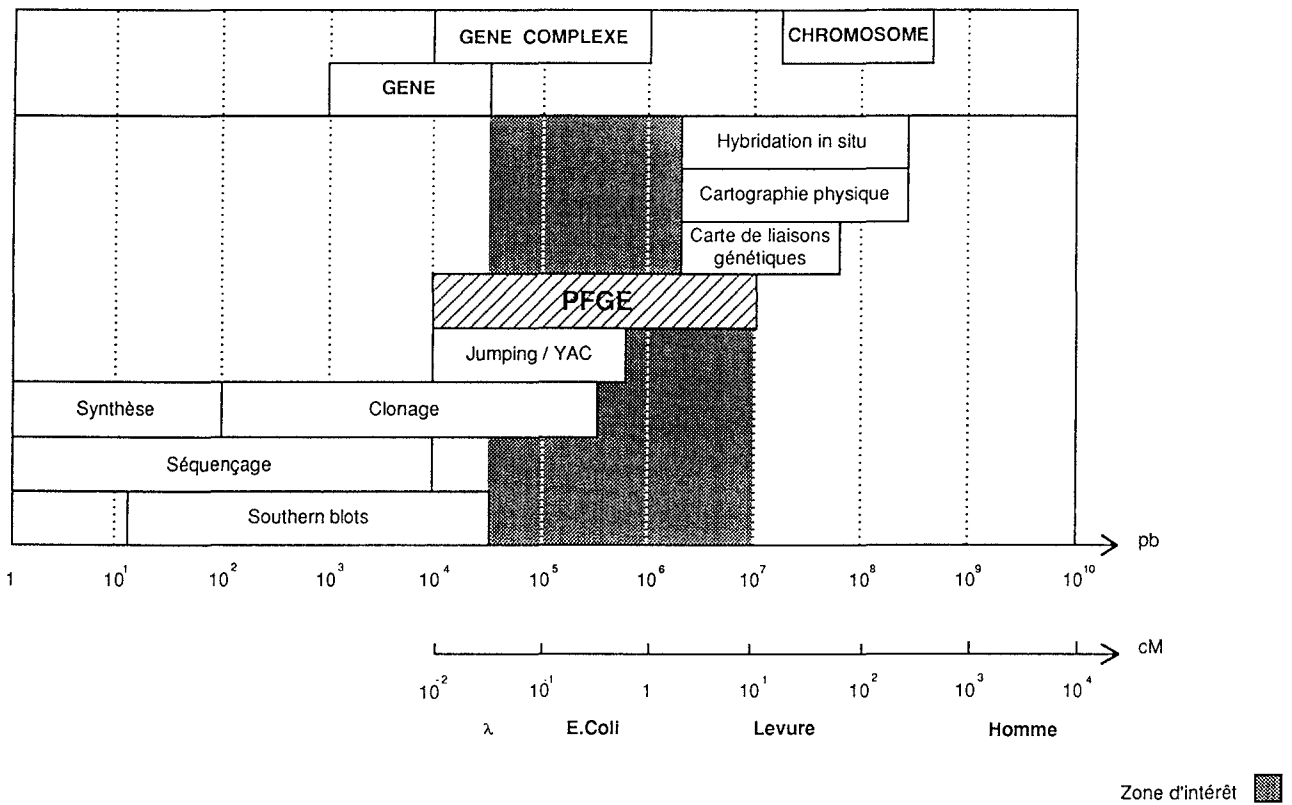


Figure 13 : Résolution du PFGE.

Comparaison avec d'autres techniques de génétique moléculaire.

Au lieu d'un champ électrique unidirectionnel, uniforme et continu, l'électrophorèse pulsée utilise deux champs d'orientation différente appliqués de façon alternative. Des électrodes étant placées sur les divers côtés de l'appareil utilisé, un pulseur permet d'alterner l'établissement des champs électriques et de choisir le temps d'application de chacun d'eux (temps de pulse). Entre deux applications successives, les molécules subissent un phénomène de réorientation lié à la différence de direction entre les champs. Or, il se trouve que la vitesse de réorientation des molécules d'ADN est étroitement liée à leur taille. Plus elle est grande, plus le temps de réorientation sera long.

La durée d'application du champ électrique dans une direction donnée (la durée du pulse) a donc deux composantes : l'une de réorientation, au cours de laquelle on peut considérer que la molécule ne migre pas, et l'autre, de migration proprement dite (Figure 14).

- a: Les fragments d'ADN s'orientent selon la direction D_1 donnée par un champ E_1 .
- b: Les fragments migrent indépendamment de leur taille suivant la direction D_1 .
- c: Sous l'effet d'un autre champ électrique E_2 , les fragments se réorientent selon la direction D_2 .
- d: Migration dans cette nouvelle direction D_2 .

Si les deux champs E_1 et E_2 sont sollicités alternativement, les fragments seront forcés à migrer selon un cycle: orientation - migration (D_1)
réorientation - migration (D_2)

La vitesse de migration étant la même pour tous les fragments de grande taille, leur séparation dépendra de leur vitesse de réorientation et donc de leur taille.

L'intervalle de taille pour lequel la séparation des molécules d'ADN est optimale dépend du temps de pulse choisi. On doit donc régler ce paramètre en fonction de la taille des molécules que l'on tente de séparer: temps de pulse d'autant plus élevé que la taille des molécules à séparer est grande.

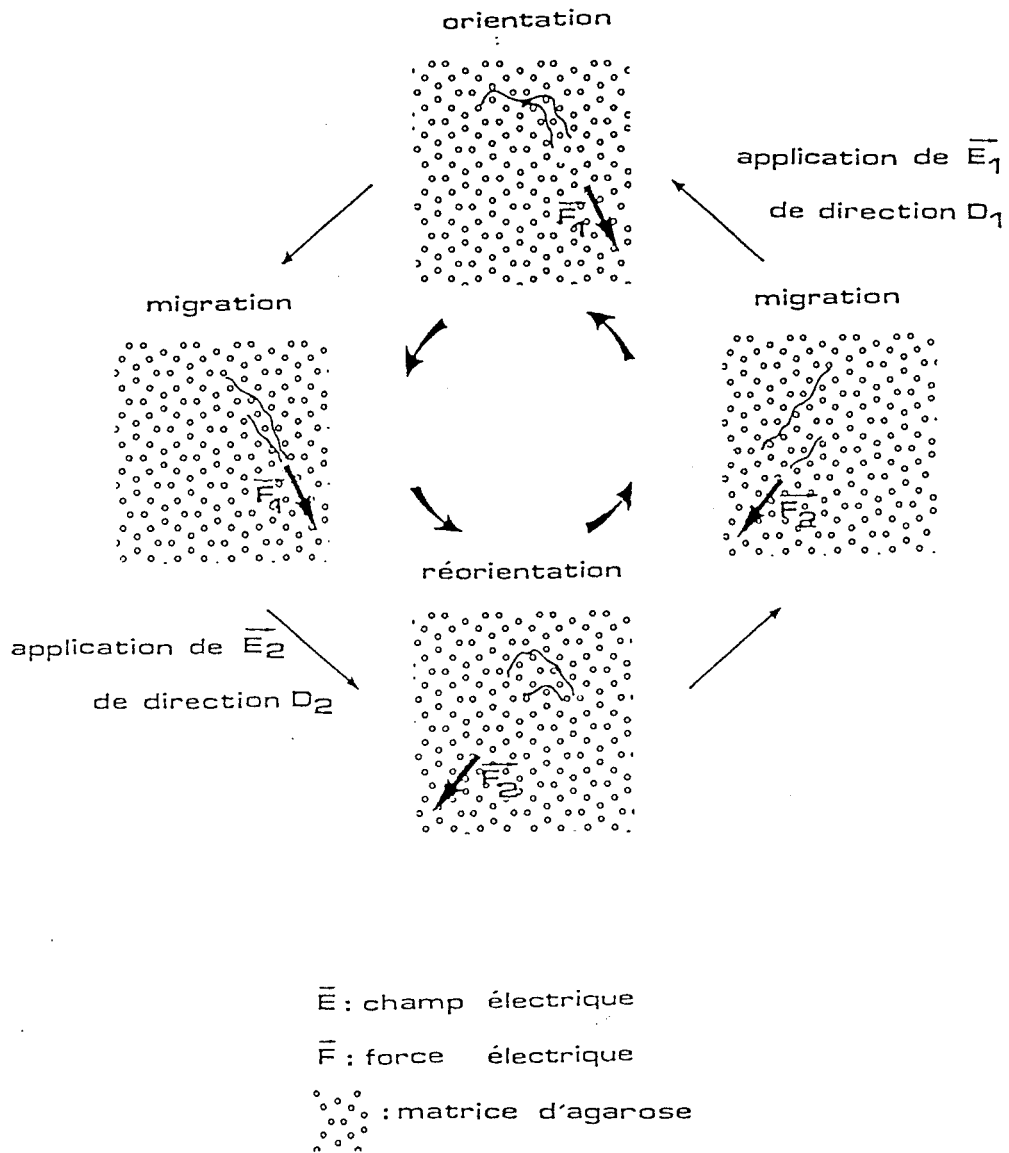


Figure 14 : Mécanisme hypothétique de la séparation des fragments d'ADN par PFGE.

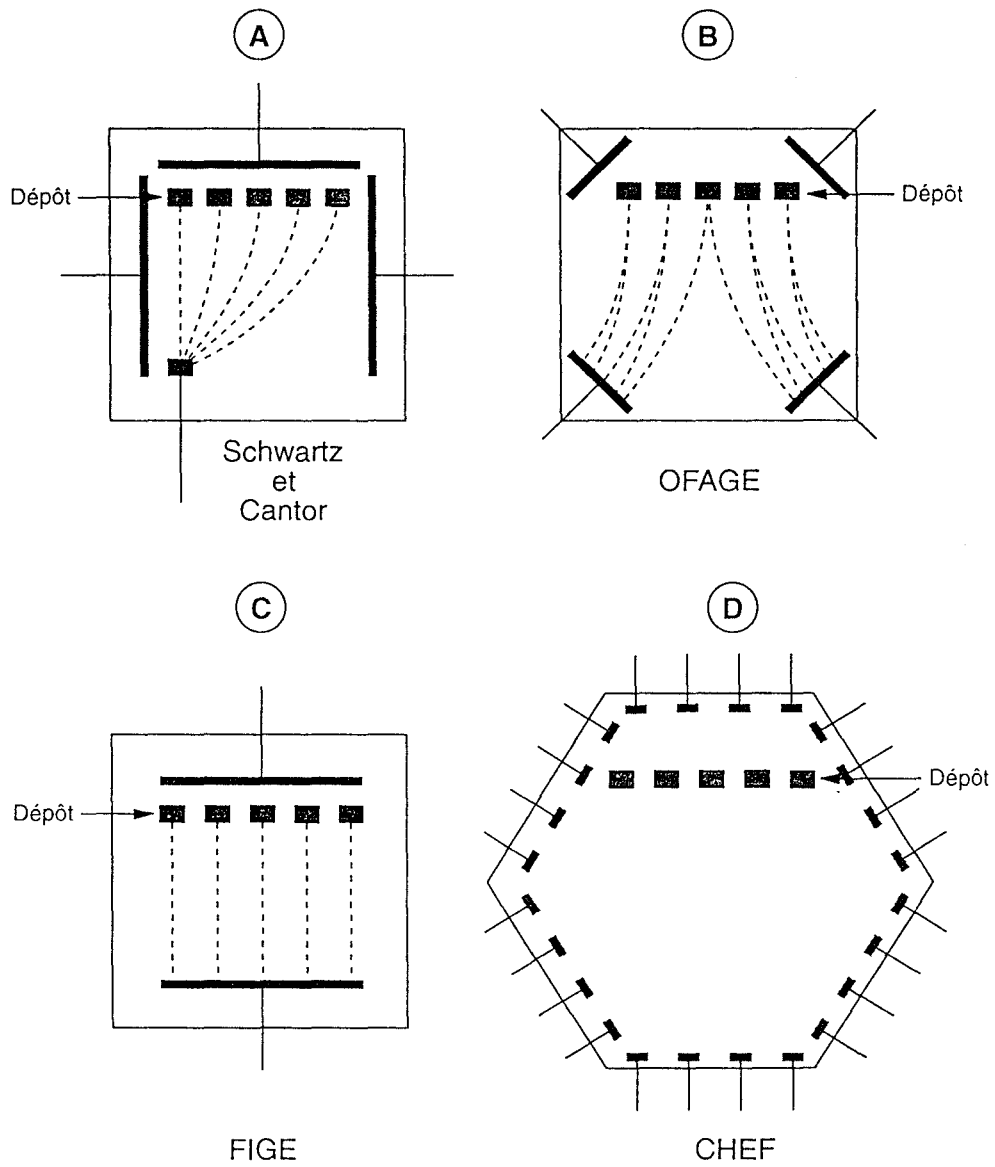


Figure 15 : Les principaux types d'électrophorèse en champ pulsé.

B. CHOIX DE LA TECHNIQUE

A partir de ces bases théoriques (Figure 15a) se sont différenciées trois techniques principales :

- la technique **OFAGE** (Orthogonal Field Alternation Gel Electrophoresis) utilise deux champs orthogonaux non homogènes alternatifs ; l'angle de réorientation est de 90° (Figure 15b) [163];
- la technique **FIGE** (Field Inversion Gel Electrophoresis) mise au point par Carle et Olson en 1986 [164] emploie un champ électrique homogène dont l'angle de réorientation est de 180° (Figure 15c);
- la technique **CHEF** (Contour-clamped Homogeneous Electric Field) conçue par Chu et coll. utilise au contraire de la technique OFAGE des champs alternatifs strictement homogènes, orientés à 120° [165] (Figure 15d). L'homogénéité des champs, l'angle de réorientation de 120° et la disposition relative des électrodes et du gel entraînent des trajectoires de migration absolument droites indépendantes de la position de l'échantillon et permettent des comparaisons très précises de plusieurs échantillons analysés simultanément. Cette technique permet encore de séparer des fragments dont la taille peut atteindre 12 Mb. Il s'agit de la technique la plus performante actuellement, aussi a-t-elle été choisie pour mener à bien notre étude.

D'autres techniques ont été décrites, mais sont peu utilisées (RFGE, CFGE, PHOGE, PACE, ZIFE, SPFGE...) Pour une bonne résolution couvrant toutes les gammes de masse moléculaire, il est possible sur certains appareillages de panacher les techniques au cours d'une même migration, par exemple séparer d'abord les gros fragments en CHEF et séparer ensuite les petits fragments en FIGE.

C. CHOIX DES LIGNEES

Le choix du matériel biologique de départ a été guidé, vers des lignées cellulaires en culture, par le fait que nous ne disposions pas, en quantité suffisante, de tissus ou de sang frais d'un seul sujet. Grâce à notre collaboration avec le groupe de J. P. Kerkaert de l'Unité INSERM 124, un certain nombre de lignées cellulaires dont les principales caractéristiques sont résumées dans le tableau ci-dessous ont pu être choisies. Il s'agit de lignées lymphoblastiques,

érythroblastiques ou épithéliales dont le caryotype ne comporte aucune anomalie cytogénétique liée au chromosome 11 en p15 :

Lignée	Caractéristiques	Caryotype	Références
HBL100	épithéliale mammaire humaine	normal	HTB124 (ATCC)
Karpas	lymphoblastique B	t(4; 11) (q21 ; q23) et t(14; 18)	[166]
K562	érythroblastique	trisomie du 11 et t(9; 22)	CCL243 (ATCC)
170	lymphoblastique B	normal	INSERM U 124
194	lymphoblastique B	normal	INSERM U 124

Une caractéristique commune à toutes ces lignées est le fait que l'on observe un taux important de méthylation des îlots CpG correspondant à des gènes dont la fonction est non essentielle pour la croissance des cellules en culture. Dans de telles lignées établies depuis longtemps, plus de la moitié des îlots CpG sont méthylés [167]. Ce phénomène est moins important dans des lignées lymphoblastiques nouvellement établies.

La méthylation des îlots CpG les rend résistants à l'hydrolyse par les enzymes de restriction à site de coupure rare [168] et produit ainsi pour ces lignées cellulaires des profils de digestion partielle. Ceci ne facilite pas la localisation des îlots CpG et donc des gènes; mais favorise en revanche l'obtention de fragments partiels de grande taille qui permettent d'étendre la carte à grande échelle au delà des îlots CpG normalement coupés.

D. PREPARATION DU MATERIEL

Il faut noter au préalable que quelque soit la technique employée, l'électrophorèse en champ pulsé, nécessite un soin particulier quant à l'homogénéité ionique du tampon d'électrophorèse et au maintien d'une température constante tout au long de l'expérience.

La qualité de l'analyse par cette technique va dépendre, entre autres, de la qualité des fragments. Aussi faut-il éviter les forces de cisaillement intervenant au cours des divers stades de préparation en incluant l'échantillon biologique dans de l'agarose qui joue un rôle protecteur [162, 169]. Toutes les opérations de lyse des membranes cellulaires ou nucléaires, de déprotéinisation

permettant d'extraire l'ADN à l'état pur et de coupures par des enzymes de restriction sont alors réalisées dans cette matrice solide.

E. CHOIX DES ENDONUCLEASES DE RESTRICTION

La cartographie du génome humain par PFGE nécessite le fractionnement de l'ADN en fragments de taille appropriée, c'est à dire compris entre 30 et 10 000 kb. Ceci est réalisable grâce à l'utilisation d'enzymes de restriction particulières. En effet, la longueur des fragments générés par la majorité des enzymes de restriction est en moyenne deux fois plus petite que ce qu'elle devrait être pour pouvoir relier des gènes éloignés les uns des autres sur une région d'un chromosome. Tel est le cas pour MUC5B, MUC5A, MUC5C et MUC2, lorsqu'on les étudie par Southern blot avec des enzymes classiques; il est impossible de relier ces séquences entre elles. Aussi est-il nécessaire d'utiliser des enzymes à sites de coupures relativement rares si l'on souhaite établir une cartographie d'une grande région chromosomique.

Les endonucléases de restriction qui permettent de générer de grands fragments reconnaissent de longues séquences (octa ou hexa-nucléotides) riches en dinucléotides CG (5' cytosine phosphate guanine 3'), qui sont rarement distribués dans le génome. En effet, l'enchaînement des bases CG est sous-représenté dans le génome humain (25 % seulement de la proportion statistique attendue si l'enchaînement des dinucléotides était livré au seul hasard) [170]. De plus 80 % des séquences CpG sont méthylables sur la cytosine. Ces deux facteurs font que les enzymes sensibles à la méthylation de l'ADN et reconnaissant des sites riches en CpG, coupent très rarement, mais pas au hasard [168]. Elles coupent préférentiellement dans des régions riches en CpG et non méthylées appelées : "îlots CpG" ou "îlots HTF" (îlot HTF : HpaII Tiny Fragment) [171, 172].

La localisation des îlots CpG, d'une taille approximative de 1.5 kb s'avère intéressante à plus d'un titre [173]. Assez fréquents (il y en aurait environ 30 000 par génome haploïde), ils sont généralement situés dans la région 5' de gènes. Ils indiquent la présence de gènes car ils leur sont presque toujours associés [174]. Ils ont donc un grand intérêt pour aider à repérer des gènes dans l'ADN, d'autant plus que leur position est facilement décelable, puisqu'ils concentrent les sites de coupure d'enzymes à site rare, très utilisées pour la cartographie à grande échelle de l'ADN [175]. Ces structures jouent donc un rôle de balise bien utile lorsque l'on pense que les gènes n'occupent que quelques pour cent de l'ADN [176]. Généralement non méthylés, sauf dans le cas du chromosome X inactif, de cellules tumorales, de sites fragiles et de lignées cellulaires en culture

depuis quelques années, les îlots CpG jouent sans doute un rôle dans l'activité des gènes qui leurs sont associés [177].

La recherche des îlots CpG constituent une étape indispensable dans "la chasse aux gènes" [173, 178, 179]. Cette approche, maintenant classique, a joué un rôle important dans l'identification du gène impliqué dans la mucoviscidose [180] et a permis de découvrir douze nouveaux gènes dans la vaste région du CMH, chez l'homme [181]. Nous l'avons donc appliquée à la cartographie de la région 11p15, afin de déterminer le nombre de gènes MUC, ainsi que la distance entre ces gènes en se basant sur le nombre d'îlot CpG et la distance qui les sépare.

De par leurs différentes caractéristiques de coupures, on peut classer les enzymes à coupure rare en fonction de la teneur en G et C et du nombre de dinucléotides CpG présents dans leur site de reconnaissance. Ce classement est fort intéressant et aide ainsi à faire un choix judicieux des enzymes à utiliser pour identifier, localiser des gènes et cartographier une région du génome (Tableau IX) [173, 175, 182, 183].

Tableau IX : Classification des enzymes utilisées pour la cartographie du génome humain [173]. Ces résultats sont issus de l'étude de 37 îlots CpG humains.

Caractéristiques du site					
Enzyme^α	Nombre de CpG	(G + C)	site	Proportion des îlots CpG reconnus	Remarques
Groupe I	2	8/8		93	Pratiquement tous les sites du groupe I sont localisés dans les îlots CpG, mais seulement une minorité des îlots possède un de ces sites.
NotI			GCGGCCGC		
AscI			GGCGCGCC		
Groupe II	2	6/6		76	La plupart de ces sites sont dans les îlots CpG; Ceux qui n'y sont pas sont souvent méthylés. Chaque îlot possède au moins un site pour une de ces enzymes. Un "cluster" de ces sites est une bonne preuve de la présence d'un îlot CpG.
BssHII			GCGCGC		
EagI			CGGCCG		
SacII / KspI			CCGCGG		
Groupe III	1	6/6		41	Comme pour le groupe II, à peu près chaque îlot possède au moins un site pour une de ces enzymes, mais il existe beaucoup plus de sites en dehors des îlots CpG, sujets à une méthylation variable.
NaeI			GCCGGC		
NarI			GGCGCC		
SmaI			CCCGGG		
Groupe IV	2	4/6		4	La plupart de ces sites sont en dehors des îlots CpG et sont sujets à une méthylation variable. Ces enzymes sont surtout utilisées pour relier physiquement des marqueurs génétiques entre eux.
MluI			ACGCGT		
NruI			TCGCGA		
PvuI			CGATCG		
SplI			CGTACG		
Groupe V	1	4/6		5	Ces enzymes sont utilisées comme celles du groupe IV.
SalI			GTCGAC		
XhoI			CTCGAG		

^α Toutes les enzymes sont bloquées par la méthylation.

Ces divers groupes d'enzymes ont des propriétés distinctes et par conséquent des utilisations différentes dans l'analyse du génome.

Les sites des enzymes du **groupe I** sont localisés exclusivement dans les îlots CpG, mais ne sont pas présents dans chaque îlot. L'existence d'un site NotI est donc une preuve solide de la présence d'un îlot CpG, mais cette enzyme ne permet pas de localiser tous les îlots. Par la rareté de ses sites de reconnaissance dans le génome, cette enzyme est une des plus utilisées pour réaliser des cartographies à grande échelle.

Les enzymes du **groupe II** sont les plus favorables pour identifier la position des îlots CpG, car 75% de leurs sites de reconnaissance sont dans les îlots et que les autres sites sont probablement méthylés et par conséquent insensibles à l'hydrolyse. Chaque îlot possède au moins un site pour une de ces enzymes.

Seule une faible proportion des sites des enzymes du **groupe III** est associée aux îlots CpG. La présence de ces sites peut confirmer l'existence d'un îlot CpG, mais n'est pas suffisante à elle seule pour affirmer leur existence.

Les enzymes SacII, NaeI et NarI coupent à des sites préférentiels qui dépendent de l'environnement du site de reconnaissance [New England Biolabs catalog p135 (1991)].

La plupart des sites des enzymes du **groupe IV** sont dans les régions inter-îlots et sont sujets à une méthylation variable. Ces enzymes sont cependant utiles pour établir des liaisons physiques entre différents marqueurs (linkage), car elles génèrent de grands fragments de restriction au delà des îlots CpG.

En dehors des cinq groupes de cette classification, quelques enzymes récemment commercialisées peuvent être aussi très utiles. C'est le cas de l'enzyme Sma I ou de Pac I, dont les sites de reconnaissance ne comportent pas de C,G. Ces enzymes, insensibles à la méthylation, doivent pouvoir générer des fragments identiques quelque soit le niveau de méthylation des lignées et permettre d'obtenir des balises précieuses entre les îlots CpG.

F. LES SONDÉS

Pour réaliser cette cartographie, nous avons utilisé nos sondes de mucines trachéobronchiques, JER47 (MUC5A), JER57 (MUC5B), JER58 (MUC5C) décrites en détail dans le paragraphe suivant et les sondes de mucines digestives MUC2 et MUC6. L'étude correspondant à ces deux derniers ADNc a été accomplie d'abord avec des oligonucléotides 48-mer NAU37 et NAU47 dont les séquences sont les suivantes :

NAU37 : 5' GGT CTG TGT GCC GGT GGG TGT TGG GGT TGG GGT CAC CGT GGT GGT GGT 3'
(séquence tirée de SMUC41, MUC2)

NAU47 : 5' CAT CTG TGC GTG GGT AGG GGT GAT GAC TGT GTG AGT ACT TGG AGT CAC 3'
(séquence tirée du clone 642 bp de MUC6)

puis avec les plasmides recombinants de MUC2 et MUC6, en collaboration avec l'équipe de Y. S. Kim du Gastrointestinal Research Laboratory de San Francisco. La sonde MUC6 a été isolée et publiée très récemment (1993). Elle n'a malheureusement été en notre possession qu'en juin dernier. Les résultats de cartographie de cette région MUC6 sont donc encore assez préliminaires.

Cette étude cartographique a permis de livrer des balises utiles pour le travail de marche sur le chromosome.

II. MARCHÉ SUR LE CHROMOSOME

Ce deuxième axe d'étude avait pour objectif de caractériser le ou les gènes de la famille des ADNc MUC5, dont nous possédions trois familles de sondes ADNc (JER47, JER57 et JER58).

L'existence d'homologie entre les chromosomes 11 et 13 décelable par la sonde JER47, nous a incités à étudier isolément le chromosome 11. Pour cela nous avons tenté la construction de 4 banques différentes :

- Nous avons envisagé la construction d'une génothèque restreinte à partir d'une lignée hybride homme-hamster ne comportant que le chromosome 11 humain. Malheureusement, en raison des problèmes de stabilité de cette lignée, le chromosome 11 humain s'est révélé trop sous-représenté par rapport au génome du hamster. Ce matériel a dû être abandonné.

- Nous avons donc construit une banque génomique restreinte à partir d'une lignée hybride homme-souris (lignée 1W1) contenant les chromosomes 11 et X humains.

- Nous avons également réalisé une banque génomique restreinte à partir de fragments d'ADN génomique humain coupé par BamHI et sélectionnés à la taille moléculaire révélée par la sonde JER57 (0.3kb, 1.2 - 1.3 kb et 1.7 - 1.8 kb).

- La construction d'une banque génomique totale, chevauchante a enfin été réalisée.

Seule cette dernière banque a permis l'obtention des clones attendus. Nous ne développerons donc que cette méthodologie.

A. PRINCIPE DE LA MARCHE SUR LE GENOME

La marche sur le génome permet de relier physiquement sur le chromosome des séquences clonées pour obtenir des séquences continues sur plusieurs centaines de kb recouvrant ainsi la totalité d'un gène. A l'aide de cette méthode on peut se déplacer de part et d'autre d'une séquence clonée, qui sert de point de départ. Les extrémités de cette séquence sont sous-clonées et utilisées à leur tour pour cribler une banque chevauchante (Figure 16), les clones génomiques positifs vont à leur tour être cartographiés et leurs extrémités sous-clonées.

B. CONSTRUCTION D'UNE BANQUE GENOMIQUE CHEVAUCHANTE

La marche sur le chromosome a comporté plusieurs étapes dont la première a été la construction des banques génomiques. Quelque soit le matériel de départ utilisé, le principe de construction est le même (Figure 17).

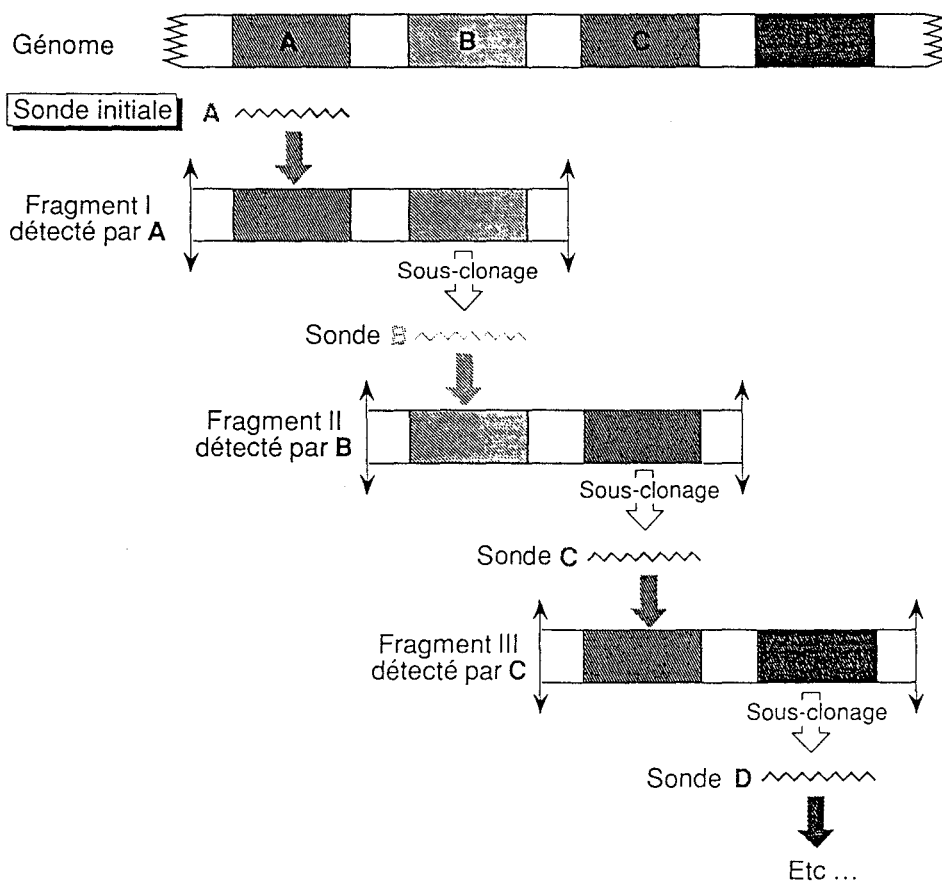


Figure 16 : Principe de la marche sur le génome.

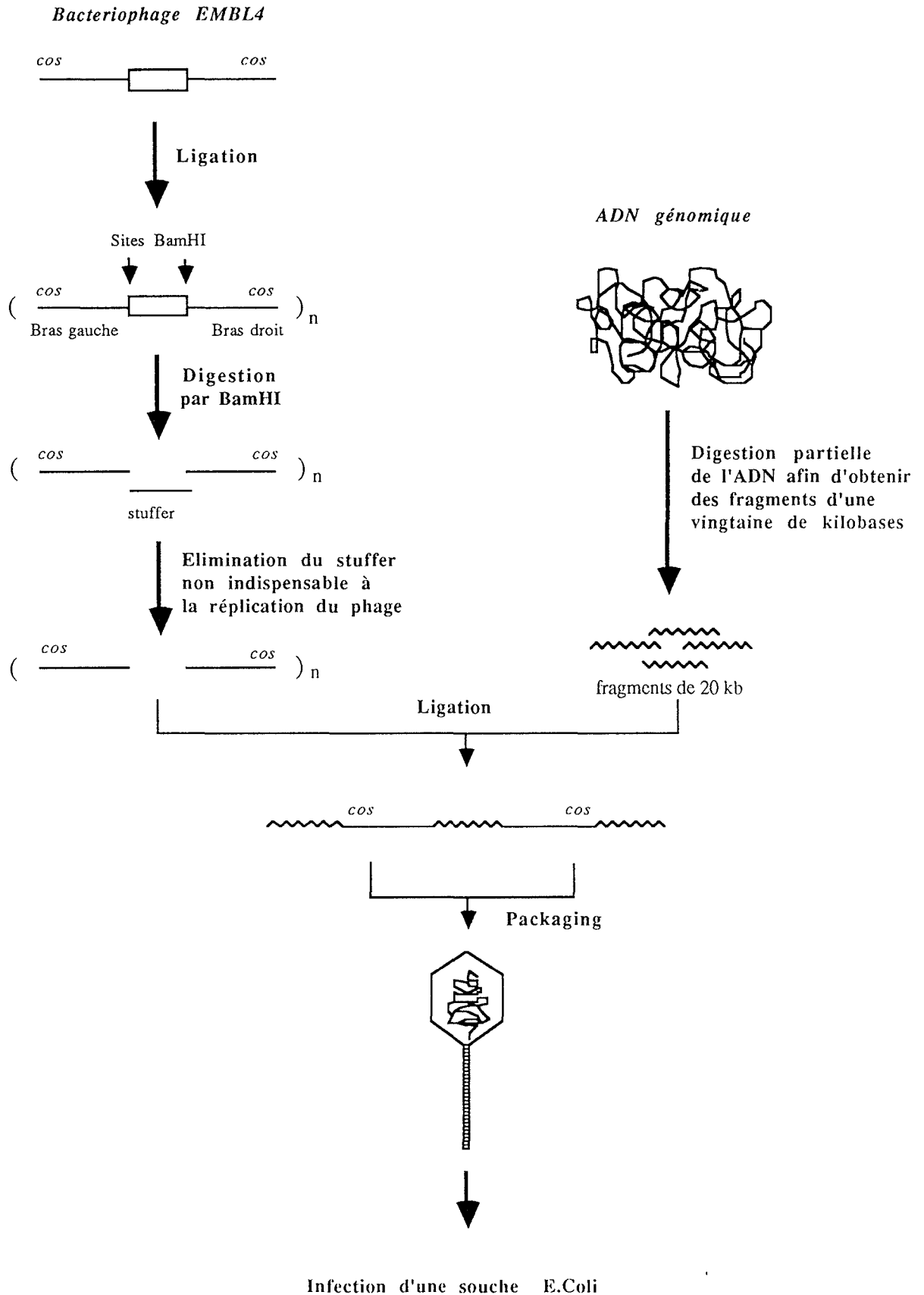


Figure 17 : Principe de la construction d'une banque génomique en vecteur phagique.

1) Fractionnement de l'ADN génomique. L'ADN de haut poids moléculaire est extrait dans des conditions particulières, qui minimisent les ruptures mécaniques et l'action des enzymes de dégradation. Par la suite, l'ADN à insérer est fractionné par hydrolyse partielle à l'aide d'une enzyme à site de restriction tétranucléotidique (Sau 3A), qui coupe par conséquent fréquemment. Il est indispensable de réaliser une hydrolyse partielle de l'ADN afin d'obtenir des clones chevauchants, condition nécessaire pour appliquer la technique de marche sur le chromosome. Les fragments sont ensuite sélectionnés en taille (de 15 à 20 kb) par ultracentrifugation en gradient de sucrose.

2) Préparation du vecteur. Ne connaissant pas la taille du ou des gènes, nous avons choisi une méthode classique en vecteur lambdaïde EMBL4, qui peut accepter des insertions d'environ 20kb. Après coupure par l'enzyme BamHI (compatible avec Sau 3A) ses bras droit et gauche, qui contiennent l'essentiel de l'information génétique nécessaire à la croissance lytique du virus, sont séparés du fragment central (stuffer) par ultracentrifugation en gradient de sucrose. Pour une meilleure séparation entre les bras et le stuffer, on réalise au préalable la concatémérisation des vecteurs par ligation des sites cos.

3) Ligation. Les fragments calibrés d'ADN génomique et les bras du vecteur peuvent donc être reliés entre eux sous forme de longues molécules concatémériques, qui sont par la suite encapsidées in vitro dans des pro-têtes protéiques de bactériophages λ . Les bactériophages recombinants sont amplifiés dans E. Coli. La banque amplifiée peut être ainsi maintenue à un titre élevé pendant plusieurs années et permettre de nombreux tours de criblage.

Avant de commencer le criblage, il fallait s'assurer que la banque était représentative, c'est à dire capable de renfermer au moins une fois l'ensemble des séquences de chacun des gènes. Pour cela, il suffit de calculer le nombre minimum de clones nécessaires, en appliquant la formule statistique de Clarke et Carbon (1976).

$$N = \frac{\log(1-P)}{\log(1-1/n)}$$

P = Probabilité de présence d'une séquence donnée

n = $\frac{\text{longueur du génome}}{\text{longueur moyenne des fragments insérés}}$

Les clones positifs ont tous été soumis au même protocole d'étude qui comporte les étapes suivantes :

- amplification du clone isolé ;
 - purification ;
 - réalisation de la carte de restriction, par simples et doubles hydrolyses totales ;
 - localisation des exons ;
 - séquençage de ces exons et comparaison des séquences avec les banques internationales de données ;
- la dernière étape de ce protocole consiste à effectuer une sonde distale à partir du clone de départ, pour à nouveau cribler la banque et isoler ainsi un clone chevauchant.

C. BANQUE GENOMIQUE EN COSMIDES

L'avantage des vecteurs cosmidiqes est qu'ils peuvent accueillir une quarantaine de kb d'ADN génomique. Les premiers résultats de mégacartographie nous ont incités à utiliser une banque cosmidiqie chevauchante (hydrolyse partielle de l'ADN génomique par Mbo I), de façon à obtenir chacun des gènes MUC isolés de la région 11p15 sous la forme d'un ou deux clones. Nous avons donc poursuivi notre étude à l'aide d'une banque génomique en vecteur cosmidiqie pWE15 (Figure 18), commercialisée par Stratagene.

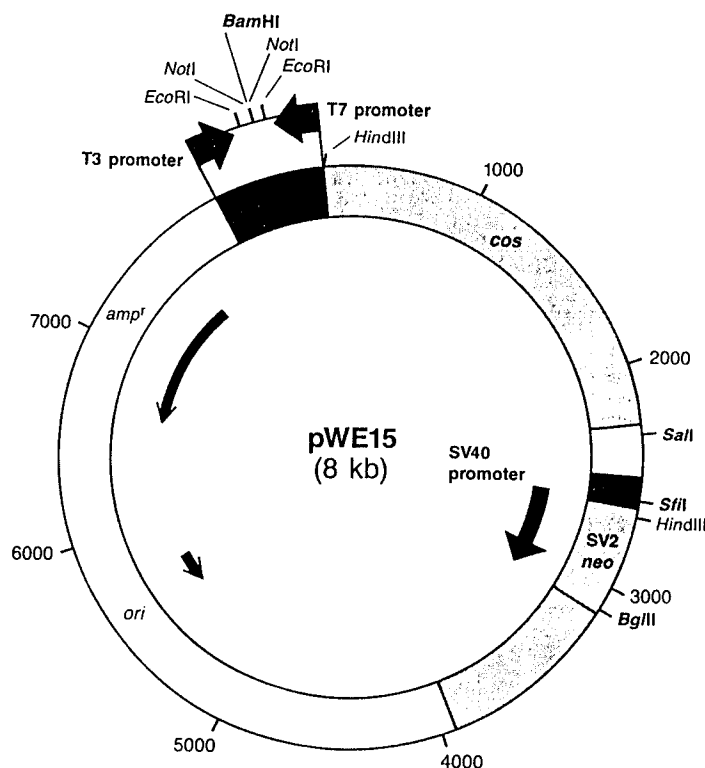


Figure 18 : Le cosmide pWE15.

La technique d'utilisation des cosmides est hybride entre celle des plasmides et celle des phages :

- le cosmide est ouvert par l'enzyme de restriction Bam HI;
- l'ADN génomique, comme pour la banque phagique, est hydrolysé partiellement par une enzyme de restriction à site tétranucléotidique (Mbo I) compatible avec Bam HI afin de générer des fragments partiels d'environ 40kb qui sont ensuite purifiés;
- après ligation, l'ADN recombinant est empaqueté dans les pro-têtes phagiques;
- les cosmides recombinants sont également amplifiés par croissance après infection dans E. Coli. Les bactéries recombinantes sont sélectionnées sur milieu antibiotique.

Les clones cosmidiques isolés sont soumis à un protocole d'étude du même type que celui des bactériophages mais avec quelques variantes techniques adaptées aux cosmides (Cf. appendice technique). La réalisation de la carte de restriction de ces clones est complexe, en raison de la taille importante de l'insert; il est impossible de placer les sites de restriction par la technique classique d'hydrolyse totale. On a donc recours à une technique particulière qui fait appel aux caractéristiques même du cosmide, qui sont, l'existence de part et d'autre de l'insert des promoteurs T3 et T7 et de deux sites NotI, site de restriction très rare dans l'ADN humain [184]. Le principe est le suivant (Figure 19) :

- double hydrolyse : NotI totale / enzyme "X" totale
- double hydrolyse : NotI totale / enzyme "X" partielle
- électrophorèse des produits de digestion en gel d'agarose.
- transfert sur membrane
- hybridation avec les sondes oligonucléotidiques T3, T7. Les sondes T3 et T7 permettent d'orienter les fragments de digestion partielle par rapport au vecteur de clonage.

Cette technique est assez laborieuse car les conditions d'hydrolyse partielle sont délicates à mettre au point et parcequ'elles varient en fonction de l'échantillon d'ADN et du lot d'enzyme.

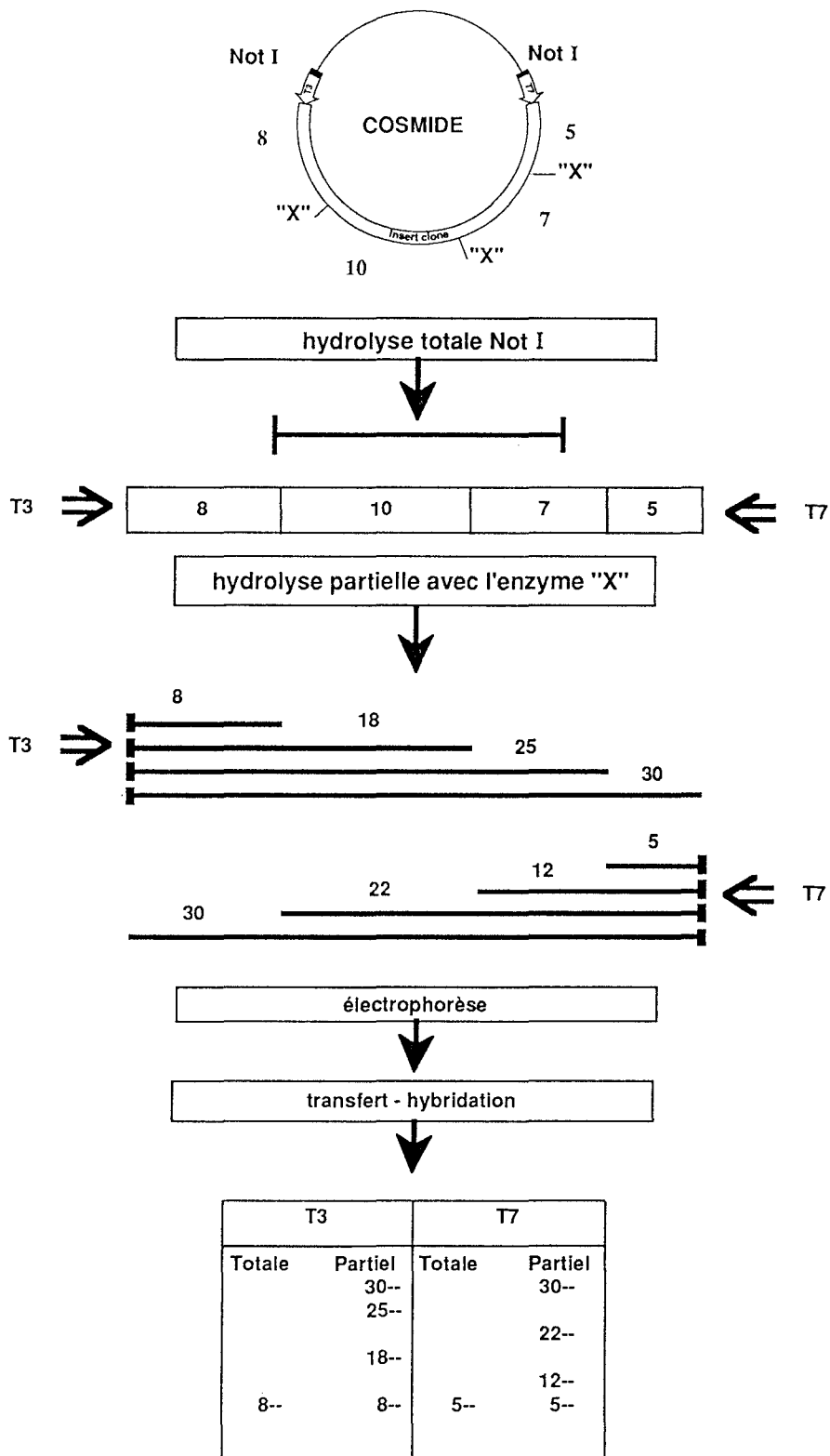


Figure 19 : Principe de cartographie d'un clone cosmidique pWE15.

D. LES SONDÉS

Le criblage de ces banques a été réalisé à l'aide des sondes ADNc suivantes, dont toutes les caractéristiques sont ici définies :

1) *JER57*

Pour étudier l'organisation génomique de MUC5B, nous avons quatre sondes à notre disposition :

Les 2 clones JER 28 (561 pb) et JER 57 (1830 pb) ont été isolés de la banque d'ADNc de muqueuse trachéobronchique humaine en vecteur λ gt11.

L'utilisation de JER 57 comme sonde pour le criblage d'une autre banque de la même muqueuse, construite en vecteur λ gt 10 a permis de sélectionner deux nouveaux clones JUL7 (1630 pb) et JUL 10 (991 pb). Ces quatre clones qui représentent 5 kb de séquence codante ont la même organisation nucléotidique en domaines répétitifs présentant entre eux une forte homologie (de 71% à 96%). Néanmoins ces domaines ont la particularité de varier en taille à la faveur de délétions ou d'insertions : si le domaine moyen, le plus fréquent, est de 87 pb, le plus petit est de 25 pb (JUL7), alors que le plus grand possède 102 pb (JUL10). En outre le clone JER 57 possède, à son extrémité 5', une séquence de 207 nucléotides non organisée selon la séquence consensus de 87 pb [83] (Figure 20).

Cette particularité structurale de JER 57 de même que sa longueur (1830 pb) nous ont décidés à choisir cet ADNc comme sonde.

Les autres caractéristiques de cette sonde sont les suivantes :

- ◆ Localisation chromosomique : la localisation en 11p15 a été vérifiée par deux techniques : l'hybridation in situ et l'utilisation d'un panel de lignées hybrides homme/rongeur [103].

- ◆ Northern blot : L'expression de MUC5B est très "ciblée" et concerne les glandes de la sous-muqueuse bronchique et trachéale ainsi que les glandes sous-maxillaires. Ceci est confirmé par les résultats d'hybridation in situ.

◆ Zoo blot : La sonde JER 57 hybride faiblement avec l'ADN de singe macaque. Aucun signal n'est perçu avec l'ADN d'espèces mammifères plus éloignées dans l'évolution : rat, souris, chien, vache, lapin, ni avec l'ADN de poulet ou de levure. Cette particularité autorisait le criblage d'une banque génomique restreinte construite à partir d'une lignée hybride homme/rongeur.

◆ Southern blot : La taille des fragments (indiquée dans le tableau ci-dessous en kb) obtenus avec quelques enzymes de restriction a permis d'avoir des repères dans le travail de marche sur le chromosome :

Eco RI	Xba I	Hind III	Bgl II	Bam HI	Pst I	Kpn I
> 30	> 30	25	16.5	1.7	1.4	8
				1.2	1.2	4
				0.3	0.9	2.3
					0.7	1.7

Certains enzymes, bien qu'aucune étude systématique n'ait encore été menée, ont démontré un polymorphisme de restriction :

Bam HI : Selon les individus, les trois bandes repérées montrent des différences d'intensité ou d'épaisseur laissant suspecter un polymorphisme lié à l'allélisme ou à un polymorphisme de restriction de type VNTR déjà décrit pour MUC1, MUC2, et MUC4 (Figure 21).

Kpn I : Lorsque l'ADN est coupé par Kpn I, JER57 hybride selon les individus avec des fragments de 1.7, 2.1, 2.3 et 8 kb (individu A) ou des fragments de 1.7, 2.1, 2.3, 4 et 8 kb (individu B) (Figure 22). L'absence de la bande à 4 kb chez l'individu A peut s'expliquer par l'existence d'un site Kpn I supplémentaire qui coupe ce fragment 4 kb en deux fragments de 1.7 et 2.3 kb qui serait alors plus intense. Or, il est difficile par Southern blot d'apprécier une telle différence d'intensité. Cette bande à 4 kb chez l'individu B peut également correspondre à la présence d'environ 45 tandems repeats supplémentaires. Dans ce cas la différence observée entre les individus A et B serait due à un polymorphisme de type VNTR.

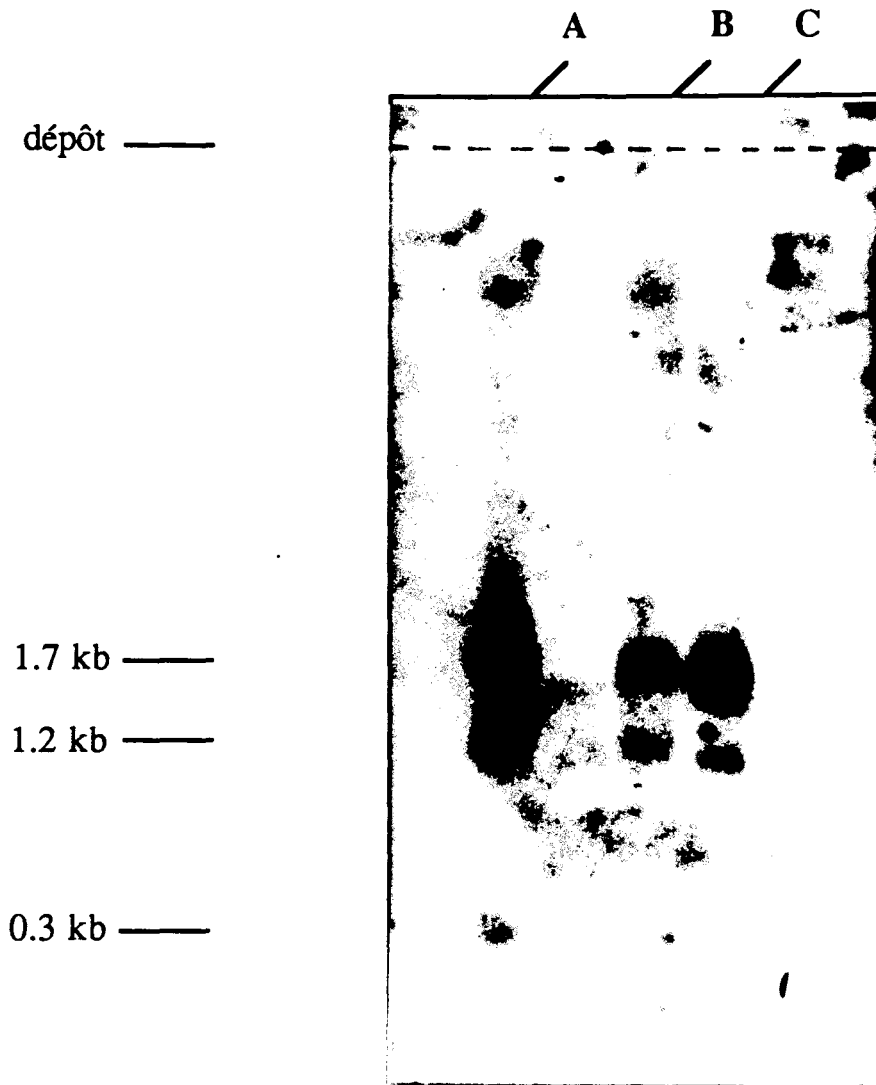


Figure 21 : Southern blot de l'ADN de trois individus A, B et C hydrolysé par Bam HI, analysé par la sonde JER57.

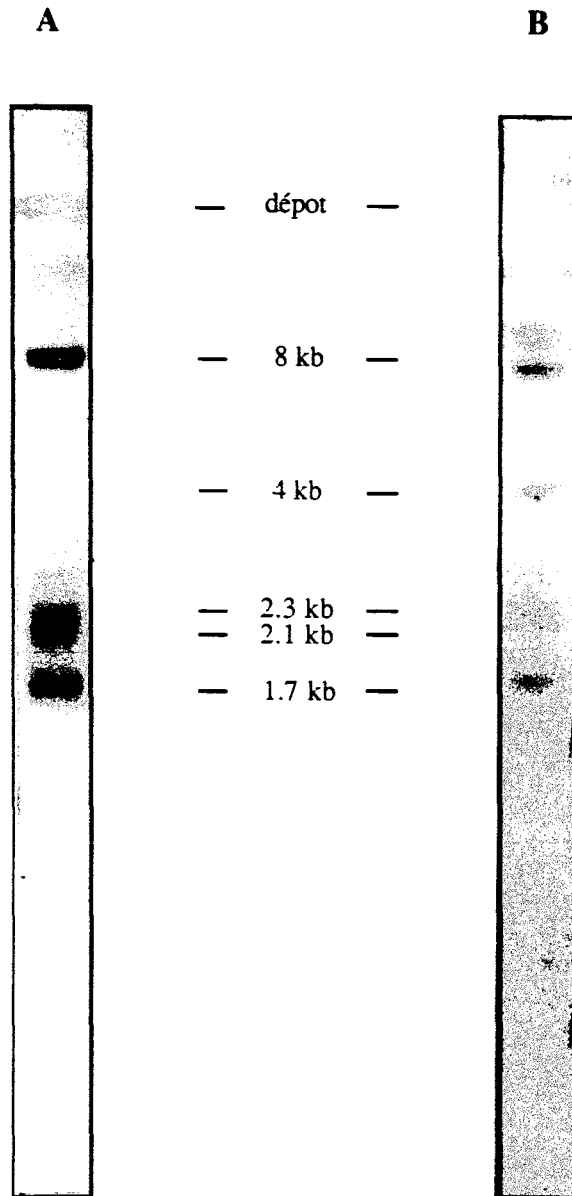


Figure 22 : Southern blot de l'ADN de deux individus A et B hydrolysé par Kpn I, analysé par la sonde JER57.

♦ Séquence peptidique déduite : L'organisation en domaines nucléotidiques, dont la longueur est modifiée par des insertions ou des délétions d'un nombre variable de nucléotides, a pour conséquence dans la plupart des cas, de modifier le cadre de lecture. Il en découle une organisation peptidique qui, elle, n'est pas en "tandem repeats". L'agencement observé alterne des régions irrégulières en taille, soit, riches en sérine et thréonine, soit, pauvres en acides aminés hydroxylés mais riches en proline et histidine (Figure 23).

2) JER58

Pour étudier l'organisation génomique de MUC5C, nous avons retenu, comme sonde pour le criblage des banques génomiques, JER58. Le choix était facile, tous les clones apparentés à cette famille :

JER 36: 264 pb ;	JER 51 : 437 pb ;	JER 54 : 201 pb ;
JER 55 : 180 pb ;	JER 61 : 400 pb ;	JER 58 : 830 pb ;
JER 63 : 420 pb.		

ont le même modèle d'agencement en domaines nucléotidiques répétitifs de 24 pb. Les domaines ne sont pas homologues à 100 % mais néanmoins, il peut être dégagé une séquence consensus :

ACC	ACC	AGC	ACA	ACC	TCT	GCT	CCT
T	T	S	T	T	S	A	P

JER 58 est la plus longue sonde de cette famille (Figure 24).

Les autres caractéristiques de cette sonde sont :

♦ Localisation chromosomique : La localisation chromosomique identique à celle de JER57, en 11p15 a été vérifiée par deux techniques différentes : l'hybridation en Southern blot, à l'ADN d'un panel de lignées hybrides homme-rongeur, et l'hybridation in situ (marquage au tritium) sur chromosomes en métaphase [103].

♦ Northern blot : L'expression tissulaire de MUC5C concerne la sphère trachéobronchique et digestive haute (muqueuse gastrique).

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
2	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
1	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
	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
2	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
	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
1	Ser Thr Pro Glu Thr Thr His Thr Ser Thr

Figure 23 : Séquence peptidique de JER57.

Alternance de domaines hydrophiles / hydrophobes.

A

		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						

B

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	Thr	Ser	Ile	Thr	Ser	Val	Pro
173	Thr	Thr	Ser	Thr	Asn	Ser	Ala	Pro
181	Ile	Ser	Ser	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	Thr	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 24 : Séquence nucléotidique (A) et peptidique (B) de JER58.

♦ Zoo blot : La sonde JER58 s'hybride avec l'ADN humain bien sûr, mais aussi avec l'ADN de singe macaque. L'ADN des autres espèces étudiées (rat, souris, chien, vache, lapin, poulet, levure) n'est pas reconnu.

♦ Southern blot : Quelques enzymes de restriction couramment utilisées en cartographie livrent les fragments suivants (indiqués en kb) avec la sonde JER 58 (Figure 25):

Eco RI	Bam HI	Hind III	Xba I	Pst I	Kpn I
29	21 ou 19	22 ou 20	22 ou 20	8	14

Un polymorphisme de restriction interindividuel a pu être décelé par les enzymes Bam HI, Xba I et Hind III.

Il est à noter que la sonde JER 47 (MUC5A) reconnaît les mêmes fragments que JER 58 avec toutes les enzymes. Ceci laisse supposer qu'il s'agit du même gène ou de gènes très proches (fragment Pst I commun de 8 kb).

♦ Séquence peptidique déduite : L'ensemble des séquences peptidiques déduites des séquences de type "24 pb" sont présentées sur la figure 26. Les motifs "TTSTTSAP", les plus représentés, sont en gras. Quelques variantes sur ce modèle : TTSTTSTP, TTRTTSAS, TASTTSGP, sont retrouvées plusieurs fois.

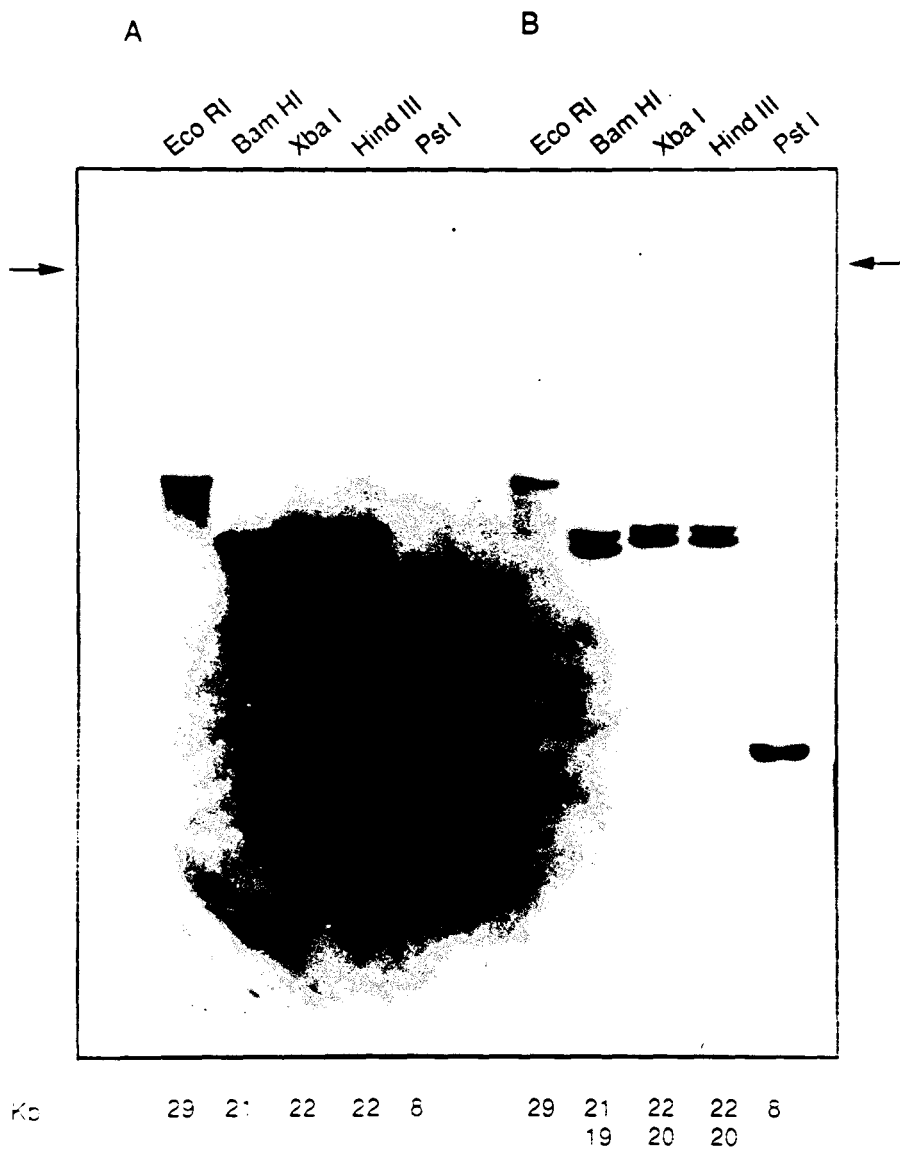


Figure 25 : Southern blot de l'ADN de deux individus A et B, analysé par la sonde JER58.

JER36	JER55	JER58	JER63
PQSTSSWQ	RTPSPVP	PSAP	RTTSTP
RSRTTTLV	TTSTTSAP	TTSTTSGP	QTTTSSAP
TTSTTSTP	TTRTTSAS	GTPSPVP	TSSTTSAP
QTSTTYAH	TASTTSGP	TTSTTSAP	TTSTISAP
TTSTTSAP	GSTPSPVP	TTRTTSAP	TTSTISAP
TARTTSAP	TTSTTSAP	KSSTTSAA	TTSTTSAP
TTRTTSAS	ITSMTSGP	TTSTTSGP	TASTTSAP
PASTTSGP	GTPR	ETTPRPVP	TSTTSSAP
GNTPSPVP		TTSTTSSP	TTNTTSAP
TTSTISAP	JER61-5	TTSTTSAP	TTSTTSAP
TTSITSAP	WVL	TTSTTSAS	ITSTISAP
	ELLPALSP	TTSTTSGA	TTSTTSTP
JER51	TTSTTSAP	GTPSPVP	QTSTISSP
RPTSGP	TTRTTSAS	TTSTTSAP	TTSTTSTP
GTPSPVP	TASTTSGP	TTRTTSAP	QTSTTSSP
TTSTTSAP	GSTPSPVP	ISSTTSAT	TTSTTSAP
ITSTTSGP	TTSTTSAP	TTSTTSGP	TTSTTSAP
GSTPSPVP	TTRTTPAS	GTPSPVP	TTSTTS
TTSTTSAP	TASTTSGP	TTSTTSAP	
TTSTTSCL	GTPA	TTSTTSGP	
		GTPSAVP	
JER54	JER61-7	TTSITSVP	
PSTTPAS	TPAQPLLL	TTSTNSAP	
IPSTTSAP	QTRTTSAS	ISSTTSAT	
TTSTTSAP	TASTTSGP	TTSRISGP	
TTSTTSAP	GSTPSPVP	ETTPSPVP	
TTSTTSTP	TTSTTSAP	TASTTSAS	
QTTTSSAP	TTRTTPAS	TTSTTSGP	
TSSTTSAP	TASTTSGP	GTPSPVP	
TTSTTLCP	GTPSPVP	TTSTISVP	
YNQV	TTSTTSAS	TTSTTSAS	
		TTSTTSAS	
		TTSTTSGP	
		GTPSPVP	
		TTSTTSAP	
		TTS	

Figure 26 : Séquences peptidiques déduites des séquences de type répétitif de 24 pb.

RESULTATS

RESULTATS

I. MEGACARTOGRAPHIE DE RESTRICTION

A. LES LIGNEES

Les lignées dont nous disposions, étaient des lignées établies depuis quelques années donc théoriquement fortement méthylées. Dans ces lignées, non seulement les régions inter-îlots étaient susceptibles d'être méthylées mais aussi les îlots eux-mêmes, ce qui risquait de compliquer la cartographie des îlots CpG. Nous nous attendions par conséquent à obtenir des profils d'hydrolyse partielle dans lesquels le polymorphisme pouvait encore ajouter à la complexité. Aussi, nous avons, dans un premier temps, préféré mener une étude cartographique comparative sur le plus grand nombre de lignées possibles (HBL100, Karpas, K562, 170, 194). Nous avons également, pour tenter de lever le problème de la méthylation, utilisé une lignée cultivée en présence de 5 aza-cytidine, (analogue non méthylable de la cytidine). Il s'agit de la lignée Karpas.

Si l'on prend l'exemple de l'analyse par Southern blot de l'hydrolyse Not I sur les quatre lignées K562, 170, Karpas et HBL100 testées avec les sondes MUC5A, MUC5B et MUC5C (Figure 27), on observe pour chacune des lignées un profil différent. Le fait qu'il existe plusieurs bandes de mêmes tailles mais qui varient en intensité d'une lignée à l'autre confirme l'existence d'un degré de méthylation variable. On peut également observer que l'hydrolyse des lignées K562 et 170 favorise l'apparition de plus petits fragments alors que l'hydrolyse des lignées Karpas et surtout HBL100 génère une majorité de grands fragments partiels.

L'ensemble de ces premières études réalisées avec de nombreuses enzymes, nous a permis de classer approximativement ces lignées, à partir des profils d'hydrolyse, en fonction de leur niveau de méthylation :

- HBL100 et Karpas sont très fortement méthylées,
- 170 et 194 sont moyennement méthylées
- K562 est peu méthylée.

Ce classement a ainsi facilité notre étude dans le choix ultérieur des lignées.

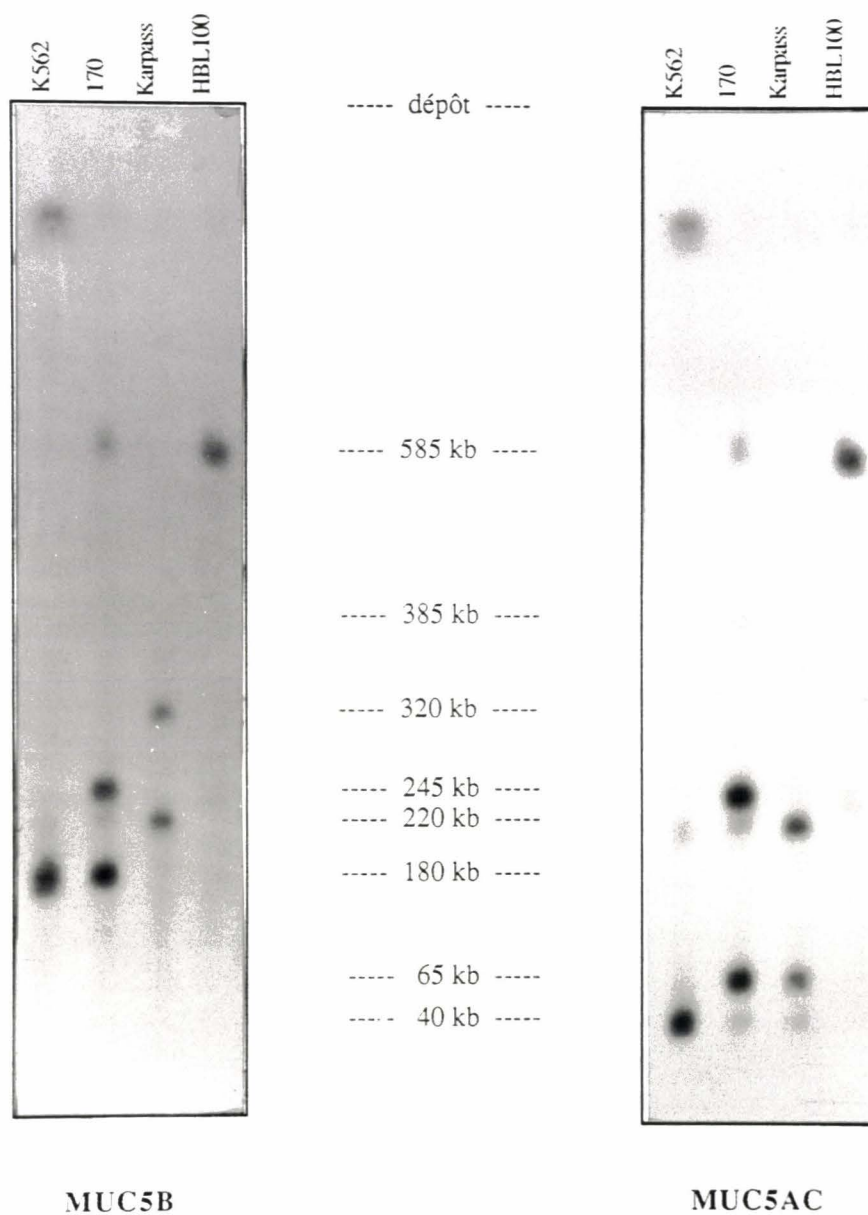


Figure 27 : Séparation par électrophorèse en champ pulsé de l'ADN des lignées (K562, 170, Karpas, HBL100) hydrolysé par Not I. Analyse en Southern blot à l'aide des sondes MUC5A, MUC5B et MUC5C.

Dans un deuxième temps, nous avons utilisé plus particulièrement deux lignées : 194 et K562. La lignée 194 a permis d'obtenir un profil d'hydrolyse partielle, permettant de relier des séquences d'ADNc sur un même fragment, alors que, après digestion complète, elles se trouvaient sur des fragments différents. La lignée K562 a, quant à elle, donné des profils d'hydrolyse quasi-totale pour certaines enzymes, permettant de visualiser les plus petits fragments hybridés avec chacune des sondes.

B. LES ENZYMES

1) *Groupe I : Nous avons utilisé les enzymes Asc I et Not I.*

D'après la littérature, la taille moyenne estimée d'un fragment Not I / Not I issu de l'hydrolyse totale de l'ADN génomique humain est de 625 kb [182]. Dans notre étude, la digestion totale par l'enzyme Not I a fourni des fragments de petites tailles 160, 180, 40 kb hybridant respectivement avec les sondes MUC2, MUC5B, MUC5A et MUC5C. Ceci laissait déjà entrevoir la présence de nombreux îlots CpG et par conséquent de nombreux gènes.

La forte sensibilité des sites Not I à la méthylation a favorisé la génération de grands fragments partiels lors de l'hydrolyse de lignées telles que Karpas et HBL100. Ainsi l'ensemble des sondes a pu être visualisé comme s'hybridant à un même fragment partiel de 585 kb.

L'hydrolyse Asc I a fourni un fragment partiel de 350 kb hybridant avec les sondes MUC2, MUC5A et MUC5C.

2) *Groupe II et III : Nous avons utilisé les enzymes Bss HII, Sac II, Ksp I et Nar I.*

La coïncidence de deux ou plusieurs sites de coupure par des enzymes du groupe II et III permet d'identifier un îlot CpG. Ceci se matérialise par l'observation que plusieurs enzymes de ces groupes livrent des fragments de tailles identiques. Le regroupement d'un certain nombre de ces sites a donc constitué un critère important à la localisation des îlots CpG sur notre cartographie.

Les enzymes Sac II et Ksp I qui possèdent le même site de reconnaissance, ont donné des profils d'hydrolyse identiques.

3) *Groupe IV : Nous avons utilisé les enzymes Mlu I et Nru I.*

Les sites de coupure de ces enzymes sont en majeure partie situés en dehors des îlots CpG et sont par conséquent souvent méthylés. Ces enzymes sensibles à la méthylation nous ont permis d'obtenir de grands fragments partiels s'hybridant à toutes les sondes : l'hydrolyse Nru I génère un fragment de 800 kb qui hybride avec chacune des sondes. L'hydrolyse Mlu I, sur les lignées les plus méthylées, génère un fragment partiel de 450 kb hybridant avec toutes les sondes ; par contre, sur la lignée K562, Mlu I divise ce fragment de 450 kb en trois fragments de 110, 140 et 25 kb qui hybrident respectivement avec MUC2, MUC5B, MUC5A et MUC5C, dissociant ainsi les fragments reconnus par les quatre sondes.

4) *Nous avons également utilisé les enzymes Swa I, Spe I et Ssp I.*

Toutes les enzymes décrites précédemment, de par leur site de reconnaissance riche en dinucléotides CG, sont sensibles à la méthylation. D'autre part, les lignées utilisées ont des taux de méthylation variables. Le concours de ces deux phénomènes a ainsi généré des profils d'hydrolyse divers et complexes en fonction des lignées. Il nous a alors paru nécessaire d'avoir recours à quelques sites de référence qui soient identiques d'une lignée à l'autre. Nous avons pour cela utilisé des enzymes plus récemment commercialisées, insensibles à la méthylation de l'ADN (leur site de reconnaissance ne comportant que A et T) et coupant rarement (Swa I, Spe I et Ssp I).

L'hydrolyse par Swa I produit un profil simple et identique quelle que soit la lignée (Figure 28). Les sites Swa I, trois sites identifiés dans la région étudiée, ont constitué le point de départ de la construction de la mégacarte.

Les hydrolyses Spe I et Ssp I réalisées sur plusieurs lignées ont donné des fragments multiples et de très petites tailles difficiles à localiser par électrophorèse en champ pulsé. Il ne nous a pas été possible de positionner précisément ces différents sites.

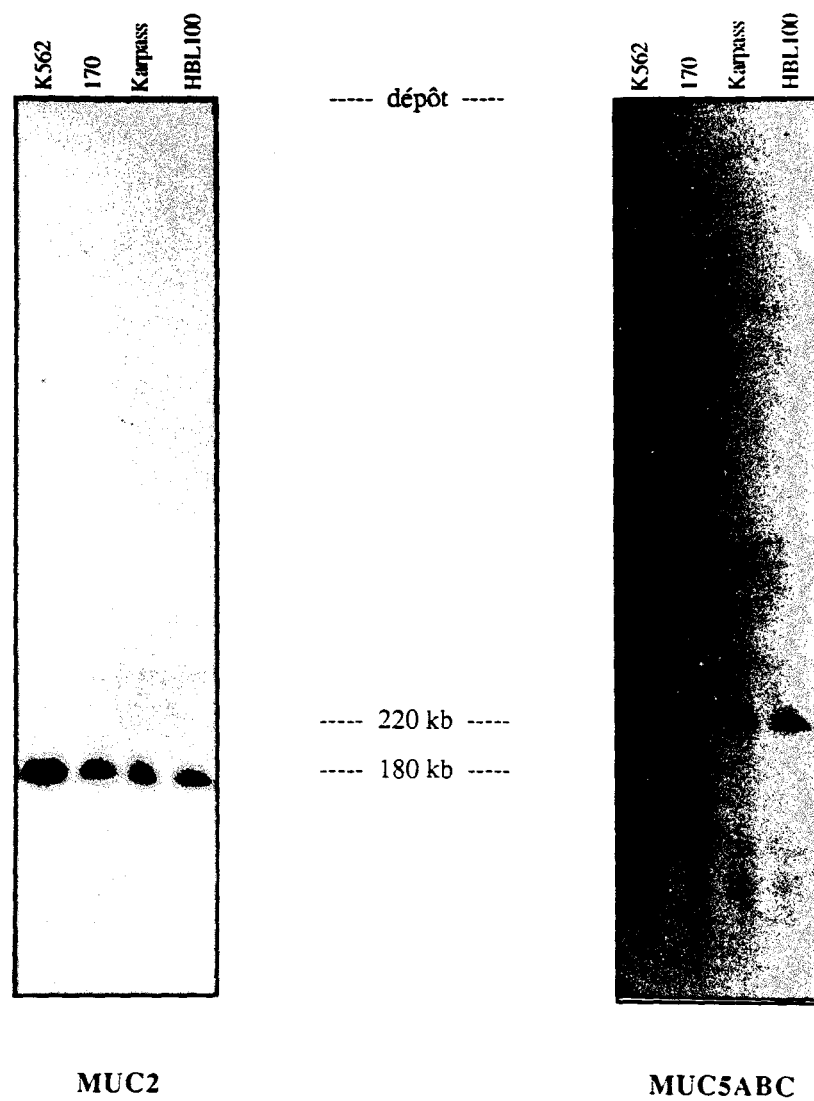


Figure 28 : Séparation par électrophorèse en champ pulsé de l'ADN des lignées K562, 194, HBL100 et Karpas hydrolysé par *Swa* I.

C. RESULTATS

Les tailles des fragments révélés par chacune des sondes étudiées par Southern blot, utilisant l'ADN des différentes lignées, hydrolysé par les enzymes citées précédemment, sont présentées dans les tableaux suivants. Ces nombreux tableaux constituent une compilation d'un grand nombre de résultats. La taille des fragments est indiquée en kilobases. Il faut noter que les conditions électrophorétiques utilisées dans ce cas autorisent une résolution de 1000 à 30 kb avec une erreur de 5 kb sur les petites tailles et une erreur de 10 kb sur les grandes tailles.

Les sondes MUC5A et MUC5C ont chaque fois révélé des fragments de même taille, c'est pourquoi nous avons noté "MUC5A-5C" dans les tableaux.

Not I					
	K562	170	194	Karpas	HBL100
MUC2	340 160	770 585 340 160		770 340 160	770 585 340
MUC5A-5C	220 40	585 385 245 220 65 40	245 220 65 40	385 220 65 40	585 245
MUC5B	220 180	585 385 245 220 180	245 220 180	385 320 220	585 245

Bss HII						
	K562	170	194	Karpas	HBL100	Karpas 5-aza Cytidine
MUC2	120		400	400	400 290	
MUC5A-5C	230 210 170 130 40 20	400 290 150	400 290	400 290	400 290	400 290 230 210 150 130 40
MUC5B	110	400 150 110	400 110 70	400	400 110	400 150 110 70

Le fragment de 290 kb généré par Bss HII et commun à MUC2, MUC5A et MUC5C n'a pas été visualisé avec la sonde MUC2 dans les lignées 170, 194 et Karpas alors qu'il aurait dû l'être comme dans la lignée HBL100. Ceci s'explique par le fait que le profil d'hybridation avec la sonde MUC2 a été analysé sur un blot endommagé par les déshybridations successives.

Bss HII / Not I			
	K562	194	HBL100
MUC2	160 120	250	400 290 250
MUC5A-5C	40 20	150 110 65 40	400 290 150
MUC5B	180 110	150 110 70	400 150 110

Swa I

	K562	170	Karpas	HBL100
MUC2	180	180	180	180
MUC5A-5C	220	220	220	220
MUC5B	220	220	220	220

Mlu I

	K562	194
MUC2	100	450
MUC5A-5C	25	450
MUC5B	140	450

Mlu I / Swa I

	K562
MUC2	100 80
MUC5A-5C	25
MUC5B	140

Mlu I / Not I

	194
MUC2	450 255
MUC5A-5C	450 195 170 65
MUC5B	450 195 170 130

Sac II		
	K562	194
MUC2	60	225
	25	160
		100
MUC5A-5C	65	290
	40	170
	25	110
	20	65
		25
MUC5B	110	110
	70	70
	25	

Ksp I		
	K562	194
MUC2	60	225
	25	160
		100
MUC5A-5C	25	170
	20	110
		65
		25
MUC5B	25	110
		70

Sac II / Bss III		
	K562	194
MUC2	60	225
	25	160
		100
MUC5A-5C	25	170
	20	110
		65
		25
MUC5B	25	110
		70

Nar I		
	K562	194
MUC2	60	
	25	
	20	
MUC5A-5C	25	40
		25
MUC5B	110	110
	25	70
	20	

Nar I / Bss HII	
	K562
MUC2	60
	25
	20
MUC5A-5C	25
	20
	15
MUC5B	110
	25
	20
	<15

Swa I / Bss HII			
	K562	Karpas	HBL100
MUC2	120	180	180
	80		
MUC5A-5C	220	180	180
	70	70	70
	40		
	20		
MUC5B	220	180	180
	110	110	110

Swa I / Not I		
	K562	HBL100
MUC2	120	180
MUC5A-5C	40	220
MUC5B	150	220

Asc I	
	K562
MUC2	350
MUC5A-5C	350

La séquence d'ADNc, MUC6, ayant été publiée en juin dernier, l'étude la concernant a été entreprise sur une autre série de digestions enzymatiques. Les premiers résultats sont les suivants :

	K562	HBL100
Swa I	180	180
Not I		585
Bss HII		400

D. INTERPRETATION DES RESULTATS

1) Le positionnement de la carte dans la région 11p15

L'hybridation in situ des sondes JER47, 57 et 58 avait permis d'affirmer la localisation du ou des gènes correspondants dans la région 11p15 et de suspecter plus précisément la sous-région 11p15.5. Ces résultats ont été confirmés grâce à l'étude d'un panel de lignées hybrides homme-souris établis à partir de l'ADN de sujets présentant des délétions ou translocations constitutionnelles variées des régions 11p13 et 11p15 [133].

Ce travail a été effectué en collaboration avec l'unité INSERM n° 178 (Directeur A. Zweibaum) qui nous a fourni l'ADN des lignées hybrides. Les diverses lignées utilisées et les résultats obtenus avec chacune des sondes sont représentés sur la figure 29.

Les sondes MUC2, MUC5A, MUC5B et MUC5C hybrident avec les lignées PEL37 et STO28.7 qui comportent la région 11p15.5. Elles n'hybrident pas avec la lignée PEL40 qui possède une translocation précisément dans cette région.

Malheureusement, cette étude ne nous a pas permis de préciser la localisation de MUC2, MUC5A, MUC5B et MUC5C par rapport à des loci connus. En revanche, nous avons pu vérifier

leur localisation dans la partie télomérique du chromosome c'est à dire dans la région p15.5, qui n'est pas facile de discerner par hybridation in situ.

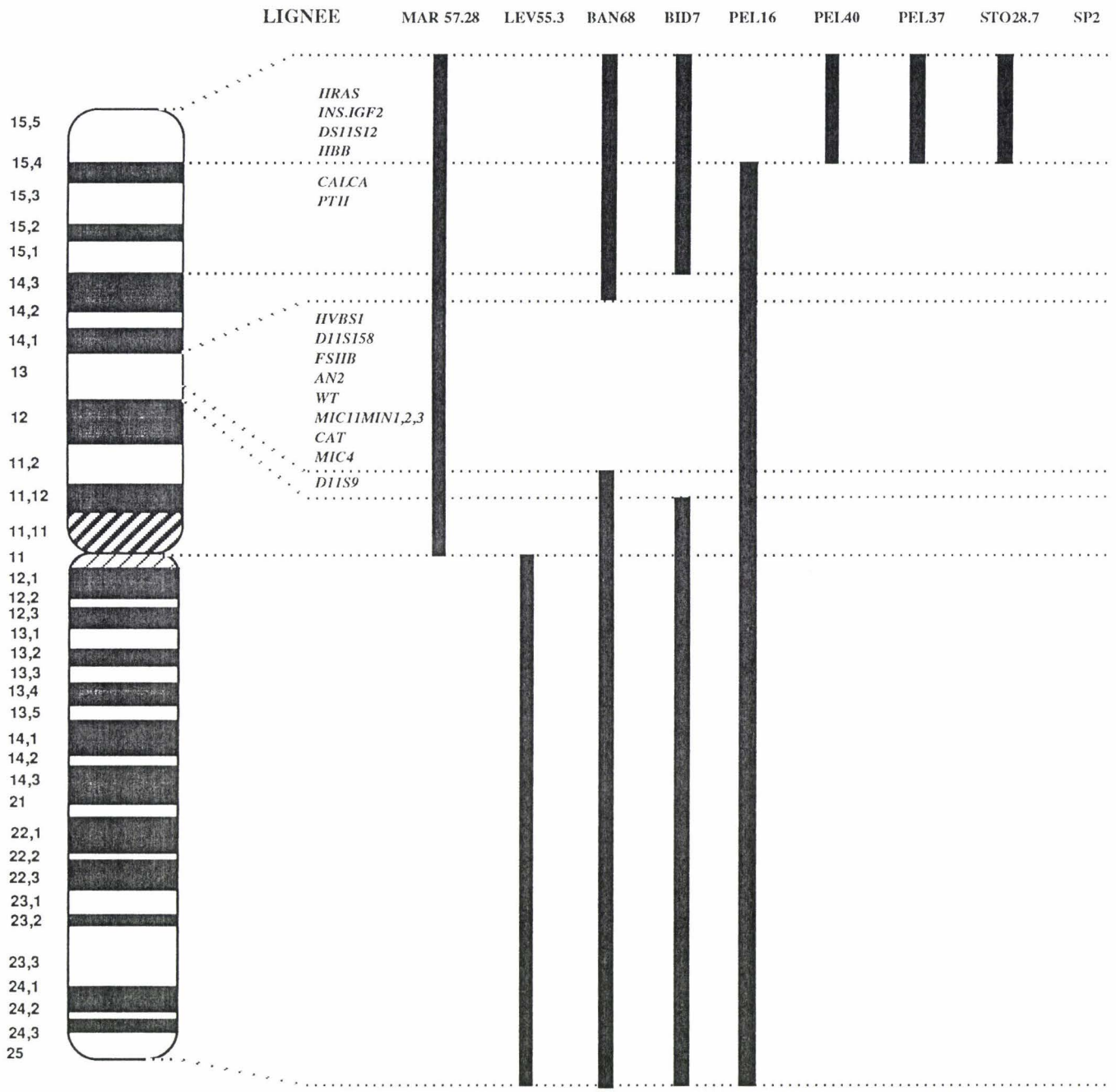
Le positionnement de la carte par d'éventuels "linkages" avec des marqueurs de référence est en cours de réalisation. Le choix de ces marqueurs de référence a été guidé par la lecture des travaux de liaison génétique que D. Swallow a réalisés à propos de MUC2. L'absence de recombinaison méiotique détectable sur 37 familles (417 sujets analysés) entre le locus MUC2 et les loci HRAS, INS, TH et HBBC indique une distance physique vraisemblablement faible entre MUC2 et ces marqueurs. L'orientation télomère-centromère de la carte n'a pas encore pu être déterminée.

2) *La proximité des séquences exoniques reconnues par les sondes*

	Taille des plus petits fragments (en kb) hybridant avec les 5 sondes		
	Not I partiel	Mlu I partiel	Bss HII partiel
MUC2	585	450	400
MUC5A-5C	585	450	400
MUC5B	585	450	400
MUC6	585	450	400

	Taille du plus petit fragment (en kb) hybridant avec 3 des sondes		
	Bss HII partiel	Swa I total	
MUC2	290	180	
MUC5A-5C	290	150	220
MUC5B		150	220

Le plus petit fragment qui hybride avec les 5 sondes est un fragment Bss HII/Bss HII partiel de 400 kb. La distance maximale séparant ces 5 séquences est donc de 400 kb. La distance maximale qui sépare MUC2, MUC5A et MUC5C est de 290kb. La distance maximale qui sépare MUC5A, MUC5B et MUC5C est de 150 kb.



11

MUC2	+	-	+	+	-	-	+	+	-
MUC5A-5C	+	-	+	+	-	-	+	+	-
MUC5B	+	-	+	+	-	-	+	+	-

Figure 29 : Représentation des fragments du chromosome 11 contenus dans chacune des lignées et des résultats obtenus avec les sondes MUC2, MUC5A, MUC5B et MUC5C sur Southern blot de l'ADN de ces lignées hybrides homme-souris. SP2 est la lignée parentale de souris.

MUC2, MUC5A et MUC5C sont localisés sur un même fragment de 290 kb, MUC5A, MUC5C, MUC5B peuvent également être regroupées sur un même fragment de 150 kb. Mais on ne retrouve pas de fragment hybridant à la fois avec MUC2 et MUC5B et excluant MUC5A, MUC5C. Ce qui permet de déduire l'ordre des sondes les unes par rapport aux autres, sans pouvoir toute fois, les orienter 5' - 3'. L'ordre est donc : MUC2 / MUC5A - MUC5C / MUC5B.

3) La construction de la cartographie à l'aide des sondes MUC2, MUC5A, MUC5B et MUC5C

L'ensemble des résultats compilés dans les nombreux tableaux précédents, nous ont permis de construire une cartographie de la région "MUC" en 11p15.5.

◆ **Swa I**

Swa I est la seule enzyme utilisée dans cette étude qui n'ait montré aucune variation d'une lignée à l'autre. Aucun profil d'hydrolyse partielle n'a été obtenu avec cette enzyme. Ce sont donc ces résultats qui ont constitué le point de départ de la cartographie.

Puisqu'il est maintenant certain que les 4 séquences d'ADNc sont localisées sur un fragment de 400 kb maximum, on peut donc, avancer l'hypothèse selon laquelle les fragments Swa I, hybridant avec les sondes MUC2 et MUC5 qui ont respectivement les tailles de 180 et 220 kb (soit un total de 400 kb), sont contigus dans le génome (Figure 30.a).

$$\text{Swa I}_1 / \text{Swa I}_2 = 220 \text{ kb}$$

$$\text{Swa I}_2 / \text{Swa I}_3 = 180 \text{ kb}$$

◆ **Not I**

Pour placer quelques sites Not I, nous nous sommes aidés des sites Swa I déjà positionnés. La double hydrolyse Swa I / Not I sur la lignée K562 (très peu méthylée) produit un fragment révélé par la sonde MUC5B de 150kb. Ce fragment est inférieur au fragment de 180 kb observé par la digestion Not I sur la même lignée. Ceci indique que le fragment 150 kb est donc un fragment Swa I / Not I (noté Swa I₁ / Not I₁ sur la figure 30.b). Un des 2 sites Not I du

fragment Not I / Not I de 180 kb étant déjà positionné, il est aisé de placer le second site (Not I₂) à 180 kb du premier (Not I₁) (Figure 30.c) le site Not I₁ permet de séparer les sondes MUC5A et MUC5C de MUC5B.

La lignée HBL100, beaucoup plus méthylée que la lignée K562 donne avec l'enzyme Not I un profil d'hydrolyse partielle. Le plus petit fragment hybridé par MUC5B l'est également par MUC5A et MUC5C, il fait 245 kb et correspond au fragment Not I₂ / Not I₃.

Le plus petit fragment hybridé par MUC2 fait 340 kb. Le plus petit fragment Not I / Not I hybridé par les 4 sondes mesure 585 kb c'est à dire qu'il représente la somme des 2 fragments précédents : $585 = 245 + 340$. Ces 2 fragments sont donc juxtaposés (Figure 30.d).

La lignée Karpas, également fortement méthylée mais en des sites différents donne un profil d'hydrolyse qui varie par rapport à la lignée HBL100 et permet de détecter d'autres sites Not I. Un fragment de 220 kb hybride avec les sondes MUC5A, MUC5B et MUC5C. Ce fragment inférieur au fragment Not I₂ / Not I₃ (245 kb) permet de localiser le nouveau site Not I₅. Not I₂ / Not I₅ = 220 kb (Figure 30.e).

Dans cette même lignée, on observe un fragment Not I / Not I de 320 kb qui hybride seulement avec la sonde MUC5B. Ce sixième site Not I peut donc être positionné au delà de Not I₂ à 320 kb du site Not I₁ (Not I₁ / Not I₆ = 320 kb). Avec les sondes MUC5A et MUC5C, on détecte deux petits fragments de 40 et 65 kb qui permettent de confirmer l'existence des sites Not I_{1,3} et 5.

Not I₁ / Not I₃ = 65 kb

Not I₁ / Not I₅ = 40 kb (Figure 30.f)

Les sondes MUC5A, MUC5B et MUC5C hybrident avec un même fragment de 385 kb visualisé dans les lignées Karpas et 170. Ce fragment correspond au fragment partiel Not I₃ / Not I₆.

◆ Mlu I

Afin de positionner le 1er site Mlu I, nous avons tout d'abord étudié la double digestion Not I / Mlu I.

Cette double hydrolyse sur la lignée 194 révèle avec la sonde MUC5B plusieurs fragments dont le plus petit a une taille de 130 kb, taille inférieure au fragment Not I₁ / Not I₂ (180 kb) hybridant avec MUC5B. Ce fragment de 130 kb correspond donc à un fragment Not I / Mlu I ; ce qui nous permet d'estimer deux positions possibles pour ce site Mlu I, soit à 130 kb de Not I₁, soit à 130 kb de Not I₂. Les fragments partiels Not I / Mlu I de 170 kb et 195 kb hybridant avec MUC5A, MUC5B et MUC5C permettent de préciser la localisation du site Mlu I₁ à 130 kb du site Not I₁. En effet, l'hydrolyse Not I / Mlu I sur la lignée 194 produit les fragments suivants (Figure 30.g) :

Mlu I₁ / Not I₁ = 130 kb hybride avec MUC5B

Mlu I₁ / Not I₅ = 170 kb hybride avec MUC5A, MUC5B, MUC5C

Mlu I₁ / Not I₃ = 195 kb hybride avec MUC5A, MUC5B, MUC5C



Seuls 2 sites Mlu I, dans cette lignée, sont sensibles à l'hydrolyse. Ces deux sites délimitent un fragment de 450 kb hybridant avec toutes les sondes. Le site Mlu I₂ peut donc être placé à 450 kb de Mlu I₁.

Sur cette lignée, après hydrolyse Not I / Mlu I, la sonde MUC2 révèle un fragment de 255 kb. Ce fragment est inférieur au fragment Not I₄ / Not I₃ de 340 kb, il correspond donc à un fragment Not I / Mlu I désigné Not I₃ / Mlu I₂ confirmant ainsi la position du site Mlu I₂ (Figure 30.h).

La simple hydrolyse Mlu I sur la lignée K562 révèle avec la sonde MUC5B un seul fragment de 140 kb supérieur au fragment Not I₁ / Mlu I₁ de 130 kb. On peut ainsi placer le site Mlu I₃ à 10 kb du site Not I₁.

Les sondes MUC5A et MUC5C révèlent, quant à elles, un seul fragment de 25 kb. Les sites de restriction Mlu I étant assez rares dans le génome et le segment comprenant MUC5A et MUC5C étant limité au fragment Not I₁ / Not I₅ de 40 kb, il est donc hautement probable que ce fragment Mlu I / Mlu I de 25 kb soit juxtaposé à celui de 140 kb hybridant avec MUC5B. De plus, l'hydrolyse Not I / Mlu I sur la lignée K562 révèle avec les sondes MUC5A et MUC5C un seul fragment dont la taille est identique à celle du fragment Mlu I / Mlu I de 25 kb. Ces deux sites sont donc localisés à l'intérieur du segment Not I₁ / Not I₅ et définissent le fragment Mlu I₃ / Mlu I₄ de 25 kb (Figure 30.i).

La sonde MUC2 hybride quant à elle sur cette même lignée avec un fragment Mlu I / Mlu I de 100 kb. La double digestion Mlu I / Swa I révèle avec MUC2, en plus du fragment 100 kb, un fragment inférieur de 80 kb. Ce fragment correspond donc à un fragment Swa I / Mlu I qui permet de positionner le site Mlu I₅ à 80 kb de Swa I₃. On peut alors placer Mlu I₆ à 100 kb de Mlu I₅ (Figure 30.i).

◆ Bss HII

Dans la lignée HBL100, fortement méthylée, l'enzyme Bss HII coupe rarement. Avec les 4 sondes utilisées, seuls 3 fragments sont visibles : l'un de 400 kb hybridant avec toutes les sondes, un autre de 290 kb hybridant avec MUC2, MUC5A et MUC5C, et un dernier de 110 kb hybridant avec MUC5B (Figure 30.j).

Afin de positionner ces 3 sites par rapport aux autres, nous avons analysé les produits de la double digestion Swa I / Bss HII sur cette même lignée.

Les sondes MUC5A, MUC5B et MUC5C hybrident avec un fragment partiel de 180 kb, inférieur au fragment Swa I₁ / Swa I₂ (220 kb). Ce fragment est donc un fragment Swa I / Bss HII. Le site Bss HII₁ est donc situé à 180 kb du site Swa I₂.

La sonde MUC5C hybride avec un fragment Swa I / Bss HII de 70 kb. Le site Bss HII₂ est ainsi localisé à 70 kb de Swa I₂, c'est à dire au niveau du site Not I₁.

La sonde MUC2 hybride avec un fragment de 180 kb identique au segment Swa I₂ / Swa I₃ (180 kb). Le troisième site Bss HII est donc situé au delà du site Swa I₃ et plus précisément à 400 kb du site Bss HII₁.

La digestion par Bss HII de la lignée 194 et de la lignée Karpas traitée à l'azacytidine produit avec la sonde MUC5B, en plus des fragments de 400 et 110 kb, un fragment de 70 kb. Le site Bss HII₄ délimitant une des extrémité de ce fragment doit se situer à 70 kb soit du site Bss HII₂ soit du site Bss HII₁, tout dépend de la localisation précise de MUC5B. Grâce à l'isolement d'un clone cosmétique (BEN1 détaillé plus loin), nous savons que MUC5B est localisé à une quinzaine de kilobases de Not I₁. Le site Bss HII₄ est donc localisé à 70 kb du site Bss HII₂ (Figure 30.k).

La lignée K562 hydrolysée par Bss HII, révèle avec MUC5A et MUC5C de nombreux fragments partiels qui laissent supposer l'existence de plusieurs sites Bss HII.

On observe un fragment Bss HII / Bss HII de 40 kb, identique à celui généré par l'hydrolyse Not I (Not I₁ / Not I₅) et par la double digestion Bss HII / Not I (Bss HII₂ / Not I₅). Le site Bss HII₅ est donc situé à 40 kb de Bss HII₂, soit au même endroit que le site Not I₅.

Les fragments partiels de 130 kb et 170 kb n'hybridant qu'avec MUC5A et MUC5C, permettent de localiser respectivement les sites Bss HII₆ et Bss HII₇ au delà du site Bss HII₅ vers MUC2 (Figure 30.l).

$$\text{Bss HII}_2 / \text{Bss HII}_6 = 130 \text{ kb}$$

$$\text{Bss HII}_2 / \text{Bss HII}_7 = 170 \text{ kb}$$

MUC5A et MUC5C hybrident également avec un petit fragment de 20 kb. Il existe donc un site Bss HII, Bss HII₈, qui coupe en deux le fragment Bss HII₂ / Bss HII₅ (40 kb) (Figure 30.l).

La sonde MUC2 sur la lignée K562 hybride avec un fragment de 120 kb qui correspond au segment Bss HII₃ / Bss HII₇ confirmant la localisation du site Bss HII₇. Sur cette même lignée, la double hydrolyse Bss HII / Swa I révèle avec la sonde MUC2 un fragment partiel de 120 kb et un fragment de 80 kb. Le fragment de 120 kb inférieur au fragment Swa I₂ / Swa I₃ (180 kb) correspond à un fragment partiel Swa I / Bss HII désigné Swa I₃ / Bss HII₆ confirmant l'existence du site Bss HII₆. Le fragment de 80 kb, inférieur au segment Bss HII₃ / Bss HII₇ et Swa I₂ / Swa I₃, est donc un fragment Bss HII / Swa I, Bss HII₇ / Swa I₃, qui confirme ainsi la position des sites Bss HII₇ et Swa I₃ et permet de localiser plus précisément MUC2 sur un fragment de 80 kb (Figure 30.l).

◆ Not I / Swa I

La digestion Not I sur la lignée K562, révèle avec la sonde MUC2 un fragment de 160 kb. L'hydrolyse Swa I / Not I révèle quant à elle un fragment inférieur, de 120 kb. Ce fragment est également inférieur au fragment Swa I₂ / Swa I₃. Le fragment de 120 kb est donc un fragment Not I / Swa I. Il permet de situer un nouveau site Not I₇ à 120 kb du site Swa I₃. Ce site Not I₇ se retrouve localisé au même endroit que Bss HII₆. On peut alors placer le site Not I₈ à 160 kb du site Not I₇, (Figure 30.m) c'est à dire au niveau du site Bss HII₃ : Not I₃ / Swa I₃ = 120 kb

$$\text{Not I}_7 / \text{Not I}_8 = 160 \text{ kb}$$

◆ Nar I

La sonde MUC5B sur la lignée 194 hydrolysée par Nar I hybride avec deux fragments de 110 kb et 70 kb. Sur la lignée K562, on retrouve le fragment 110 kb que ce soit avec la simple hydrolyse Nar I, la simple hydrolyse Bss HII ou la double digestion Nar I / Bss HII. Ce fragment de 110 kb Nar I₁ / Nar I₂ est identique au fragment Bss HII₂ / Bss HII₁.

Le fragment Nar I / Nar I de 70 kb, inexistant dans la lignée K562, laisse supposer que ce fragment est identique à Bss HII₂ / Bss HII₄ dont le site Bss HII₄ est insensible à l'hydrolyse dans la lignée K562. Nar I₁ / Nar I₃ = Bss HII₂ / Bss HII₄ (Figure 30.n).

Que ce soit avec l'hydrolyse par Bss HII ou par Nar I on n'obtient jamais pour la lignée K562 un fragment de 70 kb avec la sonde MUC5B, alors que celui-ci est visible sur la lignée 194. Les sites Nar I₃ et Bss HII₄ seraient donc soit méthylés soit polymorphiques dans la lignée K562.

L'hydrolyse Nar I sur la lignée 194 produit un fragment de 40 kb révélé par les sondes MUC5A et MUC5C et identique au fragment Bss HII₂ / Bss HII₅ (Nar I₁ / Nar I₄ = 40 kb). Les sondes MUC5A et MUC5C hybrident aussi avec un plus petit fragment de 25 kb, qui correspond au fragment Nar I₁ / Nar I₅, le site Nar I₅ étant situé aux environs du site Bss HII₈ (Figure 30.n).

◆ Sac II

Sur la lignée 194, les hydrolyses Sac II, Bss HII et la double digestion Sac II / Bss HII produisent avec la sonde MUC5B le même profil d'hybridation : deux fragments de 70 et 110 kb. Les sites Sac II et Bss HII, de part et d'autre de MUC5B, apparaissent donc localisés aux mêmes endroits (Figure 30.o).

$$\text{Sac I}_1 / \text{Sac I}_2 = \text{Bss HII}_1 / \text{Bss HII}_2 = 110 \text{ kb}$$

$$\text{Sac I}_1 / \text{Sac I}_3 = \text{Bss HII}_2 / \text{Bss HII}_4 = 70 \text{ kb}$$

Le même profil est également obtenu pour les deux hydrolyses Sac II d'une part et Sac II / Bss HII d'autre part, après hybridation avec les sondes MUC5A et MUC5C. Sachant que dans cette lignée 194, les sites Bss HII₅, Bss HII₆, Bss HII₇ et Bss HII₈ sont insensibles à l'hydrolyse, on peut expliquer la présence des fragments 25, 65, 110 kb et 170 kb de la façon suivante (Figure 30.o) :

$\text{Sac I}_1 / \text{Sac I}_4 = 65 \text{ kb}$
 $\text{Sac I}_5 / \text{Sac I}_6 = 110 \text{ kb}$
 $\text{Sac I}_5 / \text{Sac I}_7 = 25 \text{ kb}$
 $\text{Sac I}_1 / \text{Sac I}_8 = 170 \text{ kb}$

La sonde MUC2 sur la lignée 194 hydrolysée par Sac II hybride avec les fragments de 100, 160 et 225 kb. A partir des sites déjà placés on peut localiser les sites Sac I₉ et Sac I₁₀ :

$\text{Sac I}_4 / \text{Sac I}_9 = 225 \text{ kb}$
 $\text{Sac I}_9 / \text{Sac I}_6 = 160 \text{ kb}$
 $\text{Sac I}_6 / \text{Sac I}_{10} = 100 \text{ kb}$

Le fragment de 60 kb produit par l'hydrolyse Sac II de la lignée K562 hybridant avec la sonde MUC2, permet de vérifier la position du site Sac I₁₀ définissant le fragment : $\text{Sac I}_8 / \text{Sac I}_{10} = 60 \text{ kb}$ (Figure 30.p) et également de préciser la localisation de MUC2.

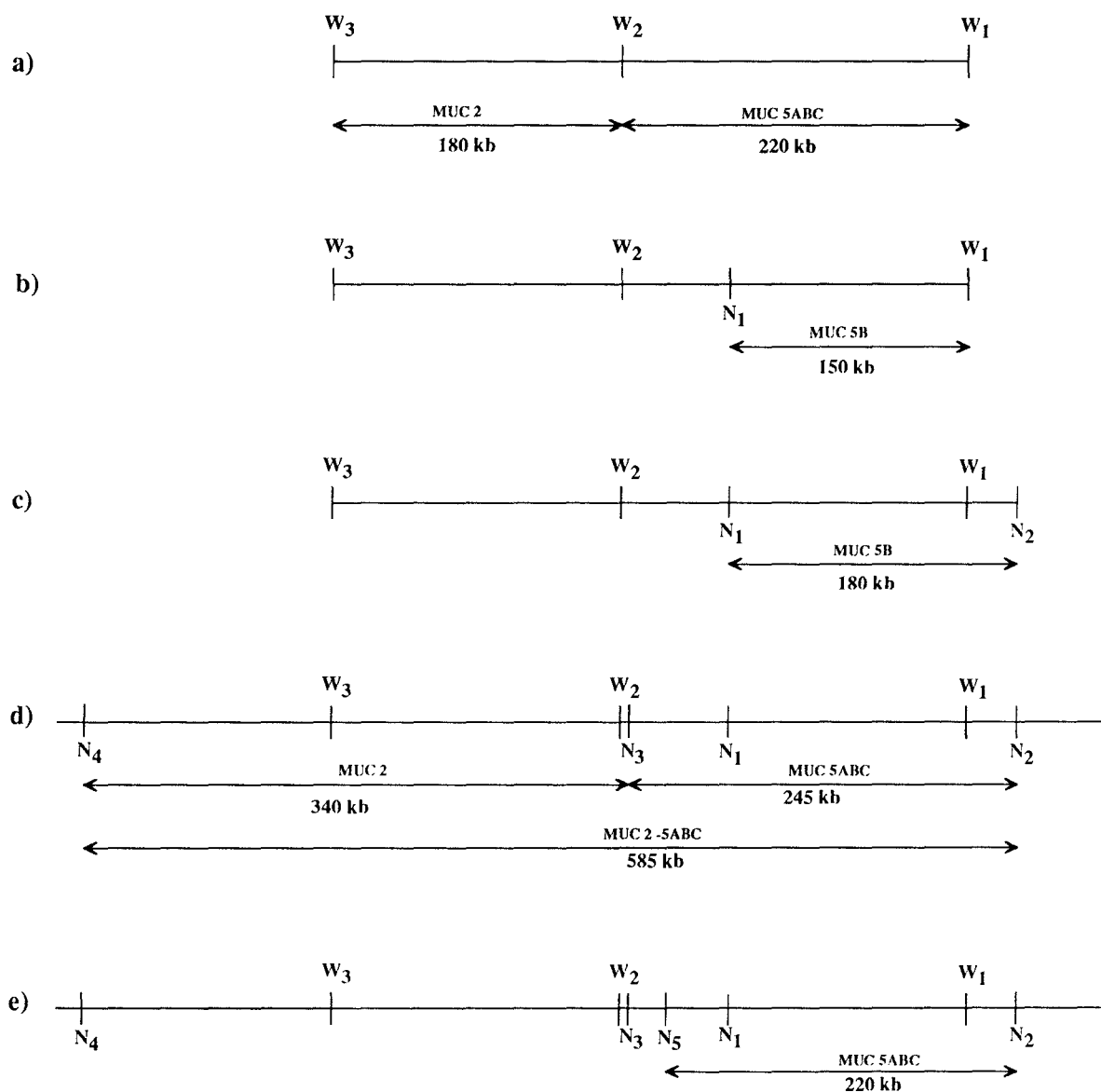
Il est difficile de positionner certains sites Sac II, Ksp I et Nar I qui donnent des petits fragments d'une vingtaine de kilobases ainsi que d'autres sites qui donnent de gros fragments partiels, souvent faiblement marqués et dont la présence dépend des conditions d'hydrolyse et de la lignée étudiée.

La cartographie finale obtenue est représentée sur la figure 31.

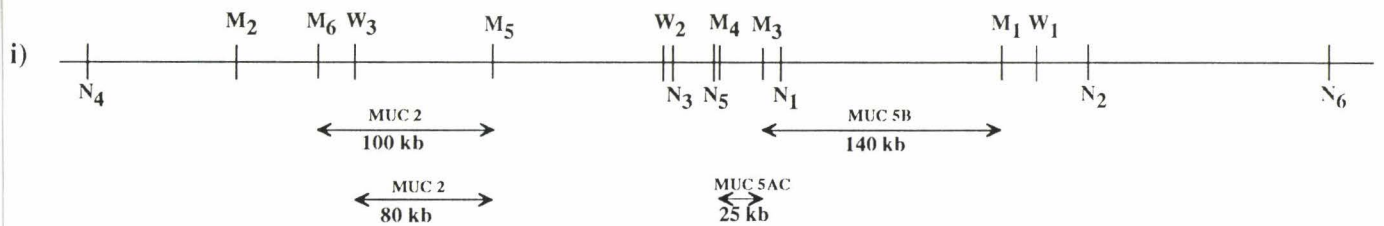
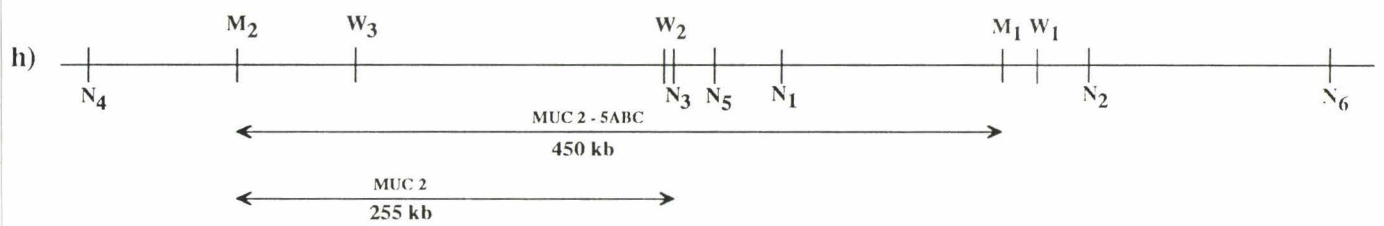
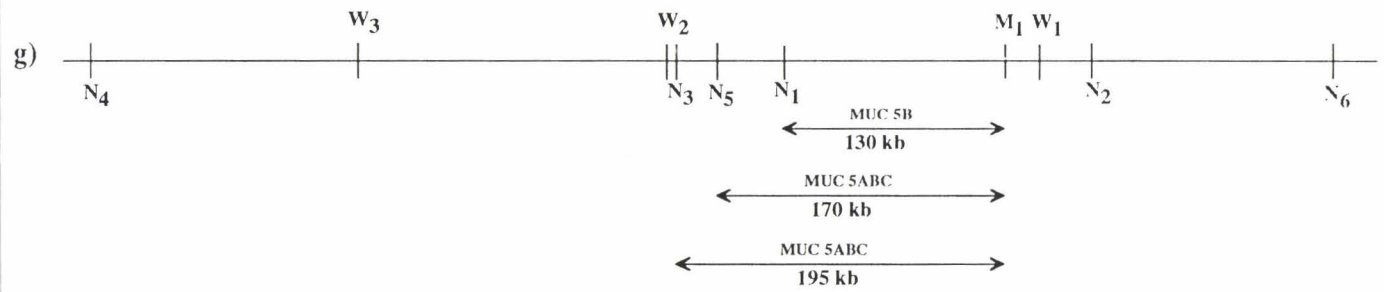
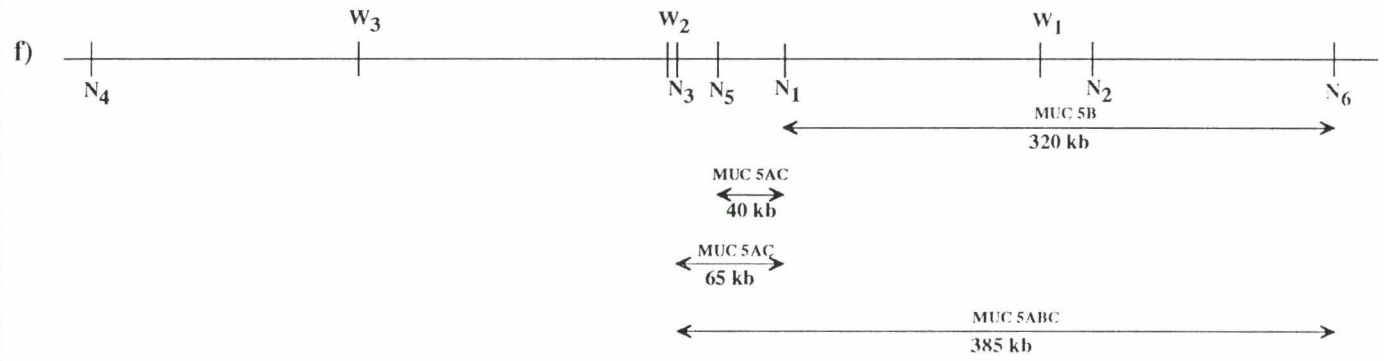
Les difficultés d'hybridation de l'oligonucléotide correspondant à MUC6 sur les blots de champ pulsé n'ont permis de fournir jusqu'à présent que des résultats préliminaires. Ces résultats permettent néanmoins de localiser MUC6 au sein de cette même région cartographiée, puisque la sonde correspondante a révélé des fragments Bss HII / Bss HII de 400 kb et Not I / Not I de 585 kb.

D'autres expériences sont en cours pour préciser sa localisation par rapport aux autres séquences MUC dans cette région.

Figure 30 (a à p) : Construction de la mégacartographie

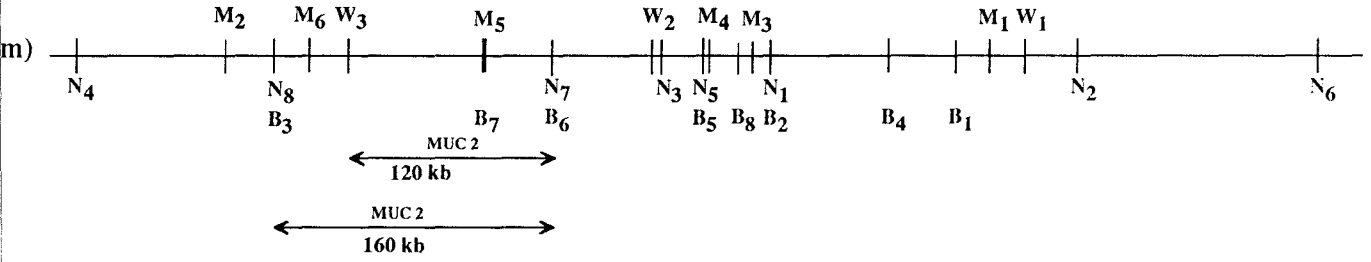
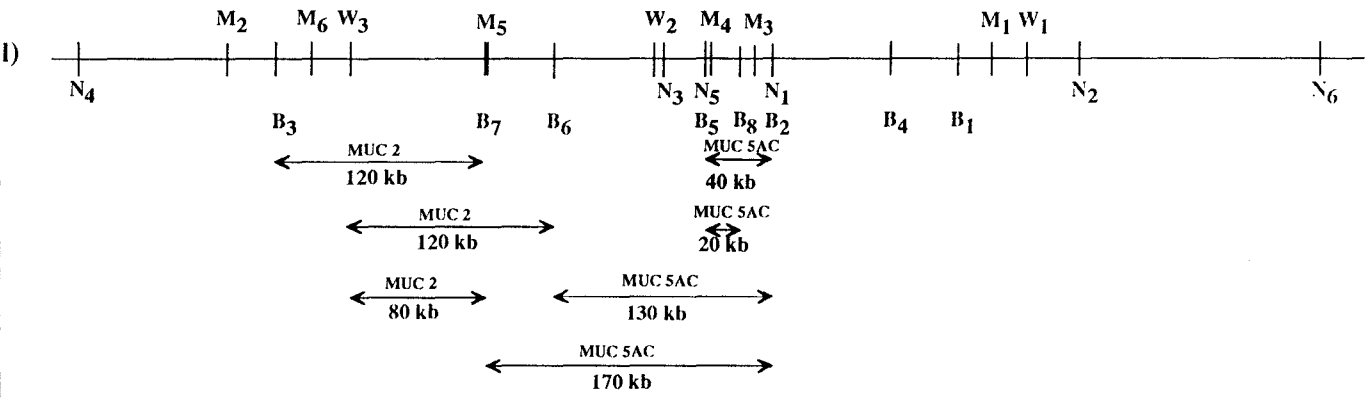
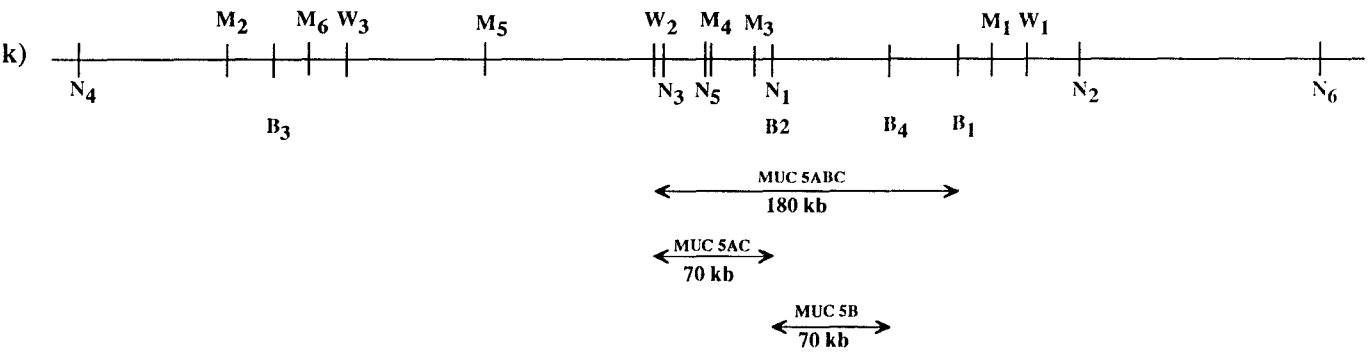
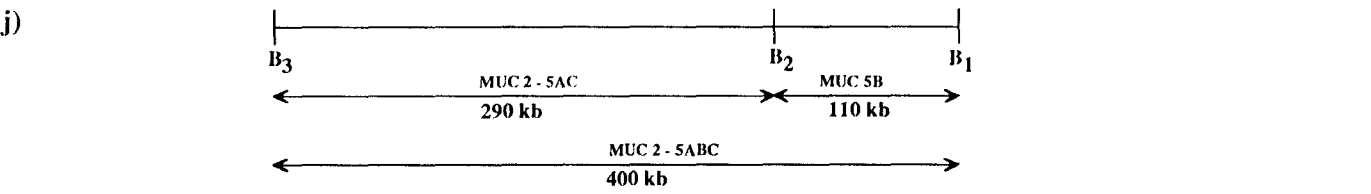

 20 Kb

A: Nar I B: BssH II M: Mlu I N: Not I S: Sac II W: Sva I



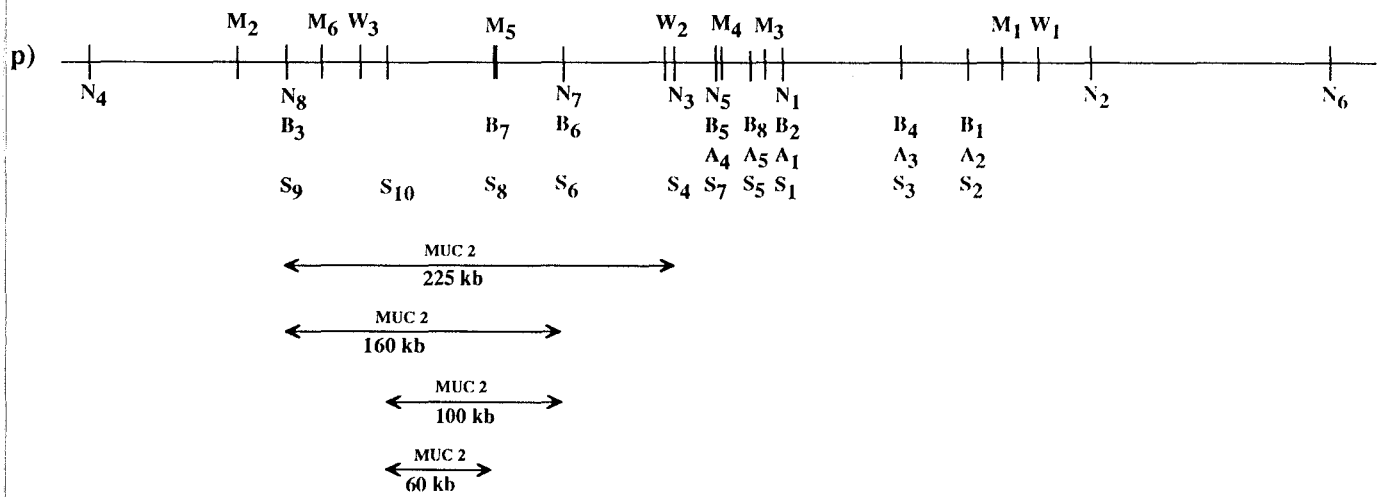
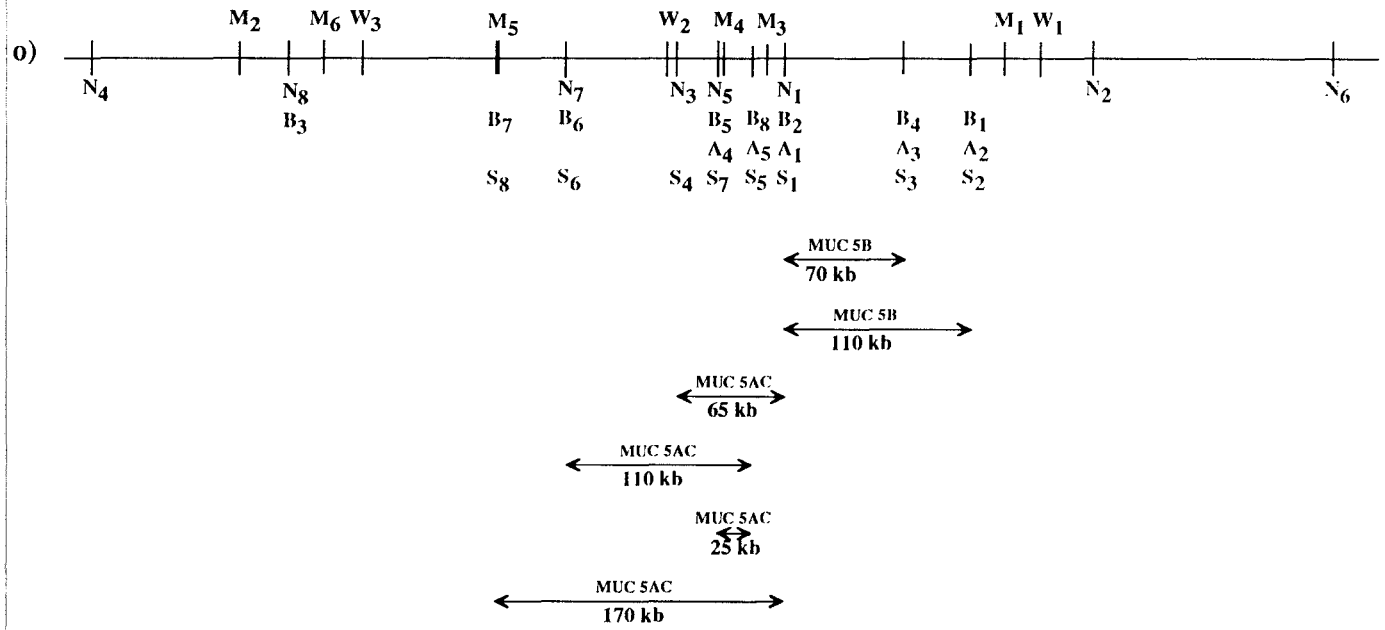
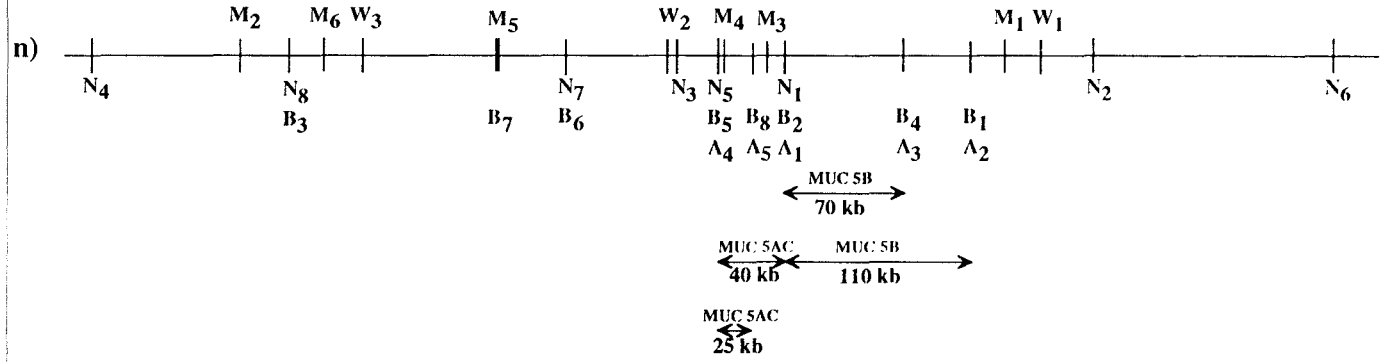
—|— 20 Kb

A: Nar I B: BssH II M: Mlu I N: Not I S: Sac II W: Swa I



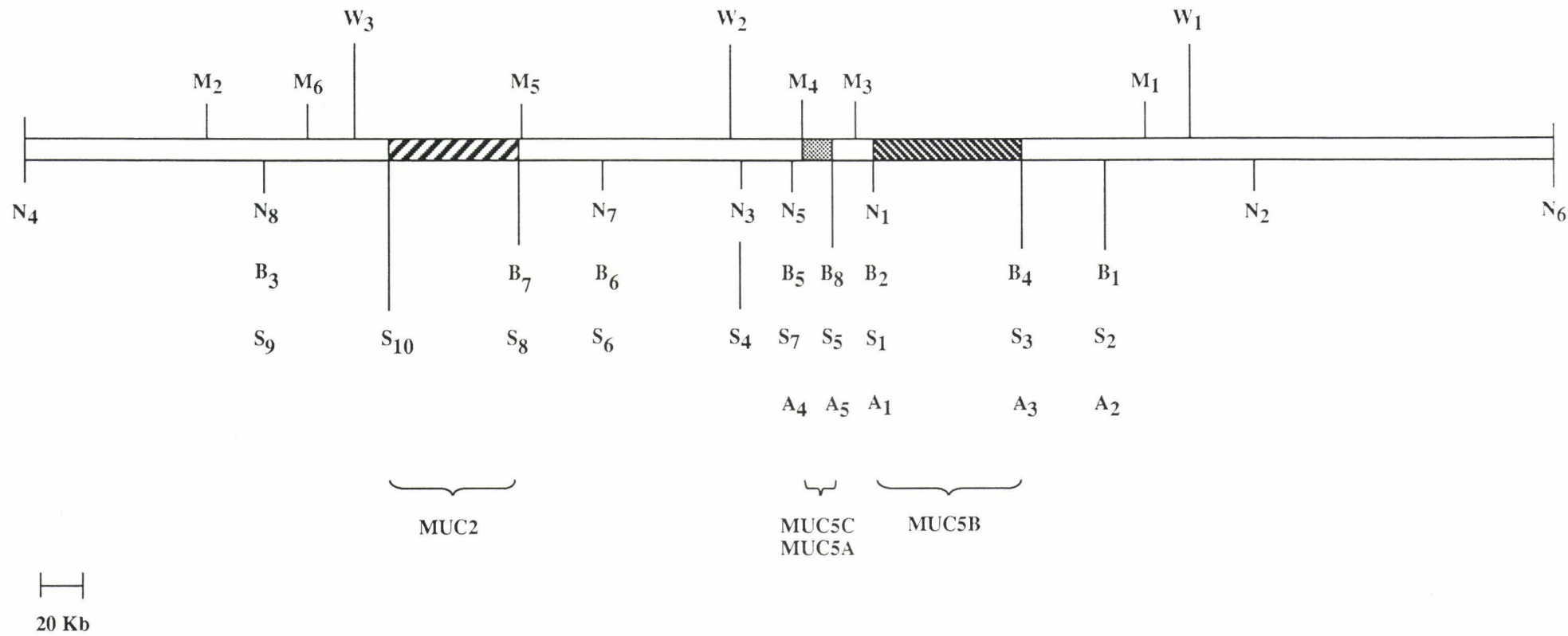
20 Kb

A: Nar I B: BssH II M: Mlu I N: Not I S: Sac II W: Swa I



20 Kb

A: Nar I B: BssH II M: Mlu I N: Not I S: Sac II W: Swa I



A : Nar I
 B : BssH II
 M : Mlu I
 N : Not I
 S : Sac II
 W : Swa I

Figure 31 : Mégacartographie de la région 11p15.5 contenant les gènes "MUC".

4) Les îlots CpG et la localisation des gènes MUC

Les îlots CpG constituent des balises annonciatrices de gènes tout le long de l'ADN génomique. La recherche de ces îlots nous a permis de supposer l'existence d'une succession de plusieurs gènes.

L'existence d'un site Not I est une preuve (à 95%) de la présence d'un îlot CpG. L'existence de 6 sites Not I implique donc la présence d'au moins 6 îlots CpG.

L'hybridation sur des Southern blots de l'ADN génomique digéré par des enzymes du groupe II telles que Bss HII et Sac II produit avec la sonde MUC5B des fragments qui apparaissent de même taille, dans la limite de résolution de l'électrophorèse en champ pulsé (± 5 à 10 kb). La double digestion avec ces deux enzymes ne réduit pas la taille des fragments hybridés (Figure 32). Les sites de coupure pour ces enzymes du groupe II sont donc regroupés sur quelques kb sur l'ADN. Le même profil d'hybridation est également observé sur l'ADN hydrolysé par Nar I, une enzyme du groupe III. Il existe donc un regroupement de sites Bss HII, Sac II et Nar I. L'accumulation de tels sites constitue une preuve solide de la présence d'un îlot CpG.

La concentration de sites pour des enzymes du groupe I, II ou III au même endroit a ainsi permis de déceler plusieurs îlots CpG. Sur notre carte établie sur 725 kb, nous avons pu cartographier au moins 12 îlots CpG (Figure 33), ce qui laisse suspecter l'existence d'au moins 12 gènes, peut-être même plus, puisqu'il a été démontré qu'un îlot CpG pouvait quelquefois être associé à deux gènes organisés en sens inverse (tête à tête) sur l'ADN [173, 185].

⇒ Il existe au moins une douzaine de gènes sur 725 kb : un cluster de gènes.

Quelles que soient les enzymes utilisées, le même profil d'hybridation est toujours observé avec les sondes MUC5A et MUC5C. Il est impossible, que ce soit avec les enzymes utilisées pour la technique en champ pulsé ou les enzymes utilisées (Eco RI, Bam HI, Hind III, Xba I, Pst I...) pour la technique classique de Southern blot, de différencier les résultats obtenus avec ces deux sondes. De plus aucun îlot CpG n'a pu être décelé entre MUC5A et MUC5C.

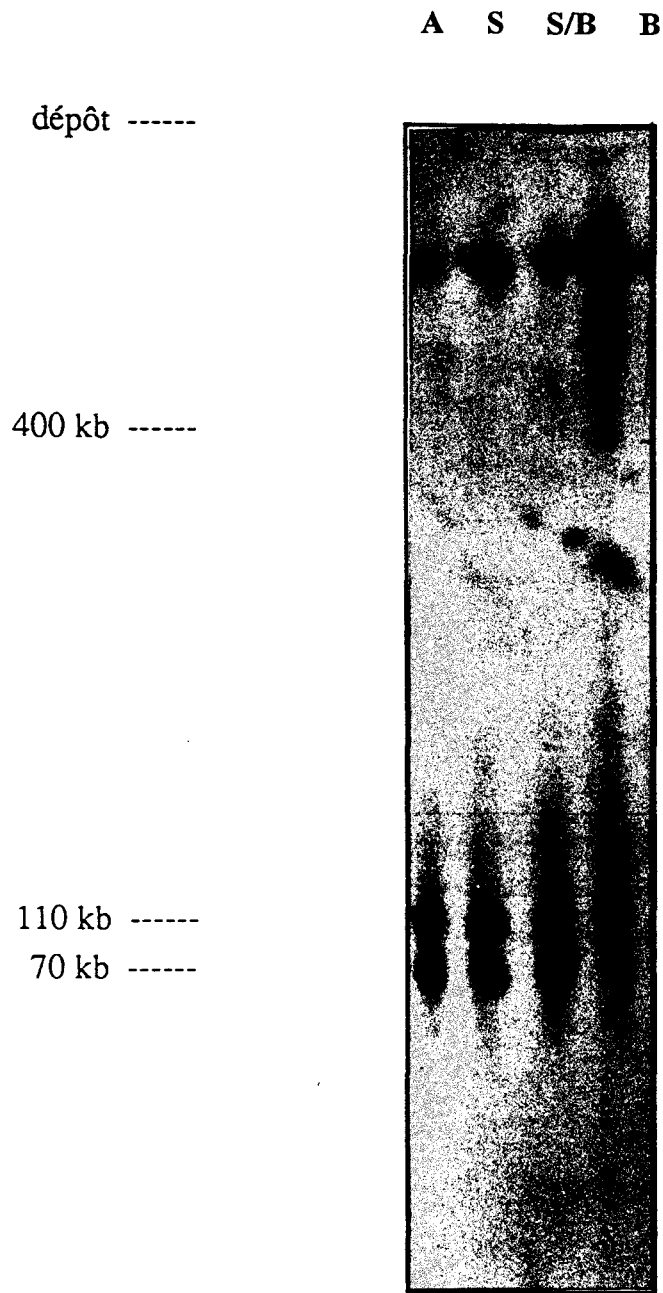
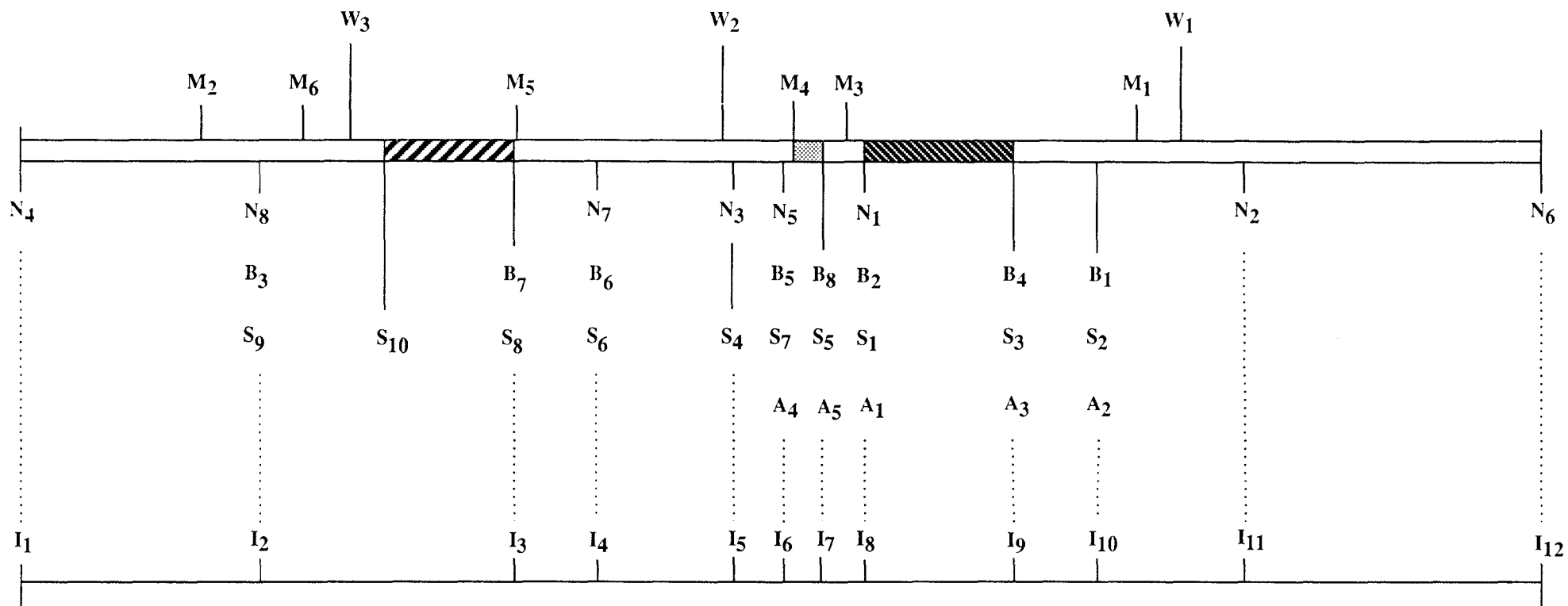


Figure 32 : Séparation par électrophorèse en champ pulsé de l'ADN de la lignée 194 hydrolysé par Nar I (A), Sac II (S), Sac II / Bss HII (S / B) et Bss HII (B). Analyse par la sonde MUC5B.



I : îlots CpG

- A : Nar I
- B : BssH II
- M : Mlu I
- N : Not I
- S : Sac II
- W : Swa I

Figure 33 : Cartographie des îlots CpG.

Identification des îlots CpG à partir des sites enzymatiques.

—|—|—
20 Kb

Malgré leurs séquences un peu différentes qui comportent quand même toutes les deux le motif répétitif "TTSTTSAP", on peut maintenant supposer que ces deux ADNc sont issus d'un seul et même gène. Cette hypothèse concorde avec les résultats d'expression tissulaire obtenus par hybridation in situ [82].

⇒ **MUC5A et MUC5C sont un seul et même gène dorénavant désigné MUC5AC**

D'après la cartographie des îlots CpG, on remarque que deux îlots séparent les sondes MUC5AC et MUC5B. On peut donc considérer que ces 2 sondes sont issues de 2 gènes différents : MUC5AC et MUC5B.

⇒ **Il existe au moins 2 gènes de mucines trachéobronchiques humaines en 11p15.5 : MUC5AC et MUC5B.**

L'îlot CpG "I_g" localisé en 5' de MUC5B est associé, d'après la carte, à ce gène. Il constituerait une région régulatrice de l'expression de ce gène de mucines trachéobronchiques humaines. La deuxième partie de ce travail, nous a permis d'isoler, par l'intermédiaire de la sonde JER57 (MUC5B) deux clones cosmidiqes dont l'un comprend à la fois cette région îlot CpG et une partie du gène MUC5B, confirmant ainsi la liaison entre le gène et l'îlot "I_g".

La sonde MUC2 est, quant à elle, séparée de MUC5AC par environ 130 kb et par trois îlots CpG. Elle caractérise donc à elle seule un gène de mucines intestinales.

D'après l'équipe de Y. Kim, MUC2 et MUC6 seraient deux gènes différents situés dans cette région.

⇒ **Il existe au moins un gène de mucines intestinales humaines et un gène de mucines gastriques en 11p15.5 : MUC2 et MUC6.**

⇒ **Sur les 725 kb explorés, ont pu être identifiés au moins 4 gènes de mucines différents.**

II. MARCHÉ SUR LE GÉNOME

A. LA BANQUE GÉNOMIQUE RESTREINTE DE LA LIGNÉE 1W1

1) *Les clones obtenus*

La lignée 1W1 est une lignée hybride homme/souris possédant uniquement les chromosomes humains 11 et X. L'utilisation de cette lignée présentait pour nous l'avantage d'exclure d'éventuels gènes apparentés aux mucines, suspectés sur le chromosome 13 par la technique d'hybridation in situ. Les deux chromosomes humains se sont avérés très peu représentés dans cette lignée et la plupart du temps masqués par l'ADN de souris de la banque réalisée à partir de cette lignée. En effet aucun clone humain, n'a pu être isolé après criblage par la sonde JER57.

Le criblage de cette banque, même à faible stringence de lavage, n'a permis d'isoler qu'une trentaine de clones. Lorsque nous avons testé ces clones par dot blot, réalisant l'hybridation sur ADN humain et ADN de souris, nous avons pu montrer que tous provenaient du génome de souris.

Le fait qu'un homologue de MUC1 ait été isolé chez la souris [77] et qu'une certaine homologie ait été observée entre MUC2 et un clone d'ADNc de mucine intestinale de rat [70], nous a incités à étudier plus précisément au moins l'un de ces clones. Nous nous sommes donc intéressés au clone **ROM 11X-6**, dans le but de caractériser la partie hybridant avec la sonde JER57 et de déterminer les homologies de séquences existant entre ce clone génomique de souris et l'ADNc de mucine humaine.

2) *Carte de restriction de ROM 11X-6*

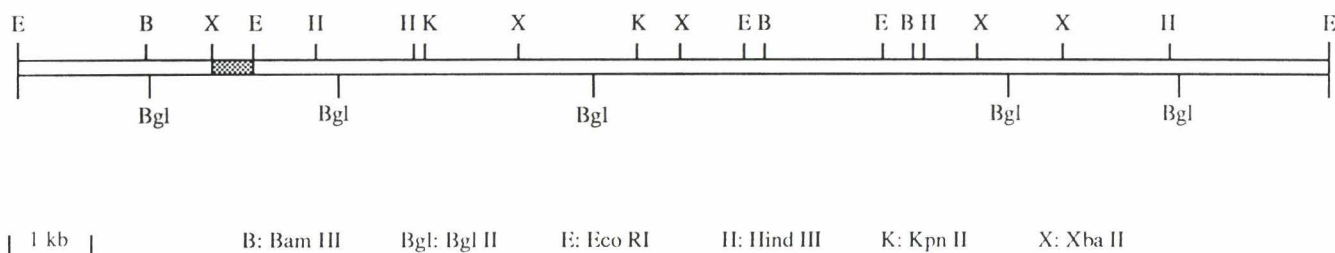


Figure 34 : Carte de restriction du clone ROM 11X-6.


```

                                     T
                                     |
                                     v
CAGCCATCCACGCTGCAGCCATCCACCCACGGCCACCTTCTCCACAGCTATCCACACCA
CAGCTATCCACACTGCAGCCATCCACCCACGGCCACCTTCTCCACAGCTATCCACACCA
CAGCTATCCACACTGCAGCCATCCACCCACGGCCACCTTCTCCACAGCTATCCACACCA
CAACTCAGCTATCCACACCTCGGCCACCCACACCACAGCCACCCACACTGCACCACAGCC
ACACATGGGCATGGGAAGAGCTGGAGCCACCTAGGACAGGGGCCAAGTCA

```

```

          10      20      30      40      50      60
          |      |      |      |      |      |
GCCACCACGCCCACCACACAGCCATCCACGCTGCAGCCATCCACCCACGGCCATC
AlaThrThrAlaHisHisThrThrAlaIleHisAlaAlaAlaIleHisProThrAlaIle
→ ProProProProThrThrProGlnProSerThrLeuGlnProSerThrProArgProSer
  HisHisArgProProHisHisSerHisProArgCysSerHisProProHisGlyHisPro

          70      80      90      100     110     120
          |      |      |      |      |      |
CTTCTCCACAGCTATCCACACCATCAGCTATCCACACTGCAGCCATCACCACGCACCTTC
LeuLeuHisSerTyrProHisHisGlnLeuSerThrLeuGlnProSerProArgThrPhe
PheSerThrAlaIleHisThrIleSerTyrProHisCysSerHisHisHisAlaProSer
SerProGlnLeuSerThrProSerAlaIleHisThrAlaAlaIleThrThrHisLeuLeu

          130     140     150     160     170     180
          |      |      |      |      |      |
TCACAGCTATCCACACCACAGCTATCCACACTGCAGCCATCCACCCACGGCCACCTTCT
SerGlnLeuSerThrProGlnLeuSerThrLeuGlnProSerThrProArgProProSer
HisSerTyrProHisHisSerTyrProHisCysSerHisProProHisGlyHisLeuLeu
ThrAlaIleHisThrThrAlaIleHisThrAlaAlaIleHisProThrAlaThrPheSer

          190     200     210     220     230     240
          |      |      |      |      |      |
CCACAGCTATCCACACCACAACCTCAGCTATCCACACCTCGGCCACCCACACCAGCCAC
ProGlnLeuSerThrProGlnLeuSerTyrProHisLeuGlyHisProHisHisSerHis
HisSerTyrProHisHisAsnSerAlaIleHisThrSerAlaThrHisThrThrAlaThr
ThrAlaIleHisThrThrThrGlnLeuSerThrProArgProProThrProGlnProPro

          250     260     270     280     290     300
          |      |      |      |      |      |
CCACACTGCACCACAGCCACCACATGGGCATGGGAAGAGCTGGAGCCACCTAGGACAGGG
ProHisCysThrThrAlaThrThrTrpAlaTrpGluGluLeuGluProProArgThrGly
HisThrAlaProGlnProProHisGlyHisGlyLysSerTrpSerHisLeuGlyGlnGly
ThrLeuHisHisSerHisHisMETGlyMETGlyArgAlaGlyAlaThr---AspArgGly

```

```

GCCAAGTCA
AlaLysSer
ProSer
GlnVal

```

Figure 35 : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de ROM 11X-6. Organisation en "tandem repeats".

3) Séquence du fragment (*Eco RI/Xba I*) hybridant avec JER57

Le seul fragment issu de ce clone ROM 11X-6 qui hybride avec la sonde JER57 est un fragment *Eco RI/Xba I* de 0.8 kb. Il a été sous-cloné en vecteur M13 et en partie séquencé par la technique de Sanger. La séquence partielle obtenue est représentée sur la figure 35.

Cette séquence présente une organisation en "tandem repeats" de 60 nucléotides soit 20 acides aminés et une richesse en résidus thréonine et serine, tout comme les ADNc de mucines caractérisés jusqu'à maintenant. Il n'existe pas entre les séquences JER57 et ROM11X-6 d'homologie en ligne. Par contre, on observe une homologie dispersée. Cette homologie repose sur la forte concentration en résidus serine, thréonine, proline et surtout en résidus histidine, observée dans le deuxième cadre de lecture, qui semble caractéristique des séquences des ADNc du gène MUC5B.

L'étude de ce fragment utilisé comme sonde sur des Northern blots de cellules caliciformes bronchiques de hamster en culture (collaboration avec K. C. Kim, Baltimore) montre une polydispersion des ARNm, semblable à celle observée pour les gènes de mucines humaines.

Ces deux caractéristiques, de séquence et d'expression, permettent de supposer que ROM11X-6 est un gène de mucine murin. Ce gène de mucine murin est-il l'homologue de MUC5B ou d'un autre gène de mucine ?

La région 11p15.5 est une région très conservée à travers les espèces. Elle correspond chez la souris à la région 90 du chromosome 7 où l'on retrouve les loci *INS*, *IGF2* et *HBB* (Tableau X) [117, 186]. Les gènes de mucines homologues à MUC2, MUC5B et MUC5C pourraient eux aussi se situer dans cette région 7⁹⁰.

Nous avons donc envisagé de déterminer la localisation chromosomique de ROM11X-6, afin de savoir s'il existe un locus "MUC" sur le chromosome 7 de souris et si ce cluster de gènes est lui aussi conservé entre espèces. Ce travail a pu être réalisé grâce à la collaboration avec M.G. MATTEI (U242, INSERM, Marseille). ROM 11X6 a ainsi été localisé sur le chromosome 8 murin, avec un signal suffisamment intense et spécifique pour valider le résultat :

150 cellules en métaphases ont été observées, 197 grains d'argent étaient associés aux chromosomes. 42 d'entre eux (21.3%) étaient localisés sur le chromosome 8. 29 grains sur 42

(c'est à dire 69%) étaient distribués sur la région [C2-D] du chromosome 8. La région du génome humain homologue à la région mucine 8C2-8D est localisée sur le chromosome 16.

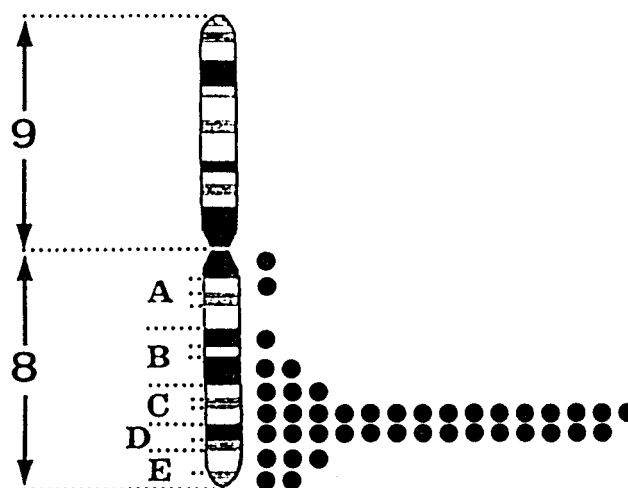


Figure 36 : Représentation des résultats d'hybridation in situ sur chromosome murin de cellules en métaphase du clone ROM 11X-6.

Tableau X : Conservation de la région 11p15.5 à travers les espèces.

Locus humain	Gene	Localisation humaine	Locus murin	Chromosome de souris
HRAS	Homologue de l'oncogène viral du sarcome de rat Harvey (V-Ha-ras)	11p15.5	Hras-1	742
INS	insuline	11p15.5	Ins-2	790
IGF2	"insulin-like growth factor 2"	11p15.5	Igf-2	790
HBB	hémoglobine chaîne, β	11p15.5	Hbb	790
PTH	hormone parathyroïdienne	11p15.2-p15.1	Pth	790
MYOD1	facteur myogénique 3	11p15	Myod-1	790

B. LA BANQUE GENOMIQUE HUMAINE EN VECTEUR EMBL4 CRIBLEE PAR LA SONDE JER58

1) Clones obtenus

Le criblage de cette génothèque par la sonde JER58 (MUC5A, MUC5C) a permis d'isoler 3 clones : CEL1, CEL2, CEL3. Ces trois clones ont posé de gros problèmes lors de la réalisation de la carte de restriction qui d'ailleurs s'est avérée identique dans les trois cas ; il a été impossible de localiser les bras droit et gauche du phage. Ces clones sont fortement recombines.

Les phénomènes de recombinaison sont assez fréquents lors de la fabrication de banques génomiques et affectent d'autant plus les séquences répétitives. Les gènes de mucines par leur caractère répétitif (VNTR) constituent un "point chaud" de recombinaison qui contribue à l'instabilité des banques. Le criblage de cette même banque au laboratoire, par la sonde MUC4 a également permis d'isoler un grand nombre de clones recombinants qui la plupart du temps ne survivent pas après les nécessaires étapes d'amplification dans une bactérie-hôte. De telles recombinaisons ont également été observées par l'équipe de San Francisco à propos de MUC2 : le clone génomique (GMUC) présente une délétion de 4 kb dans la partie répétitive de 7 kb [94].

2) Séquence de CEL2A

A cause de ces problèmes de recombinaison, seul un fragment Eco RI/Eco RI de 0.304 kb (CEL2A) hybridant avec la sonde JER58 a pu être isolé et étudié. Il a été sous-cloné en vecteur plasmidique Bluescript et séquencé entièrement.

Ce fragment génomique à cadre entièrement ouvert est caractérisé par une séquence peptidique déduite, répétée en tandem de 8 acides aminés de type TTSTTSAP plus ou moins conservés. Il présente, en dehors des deux sites Eco RI qui encadrent la séquence, 100% d'homologie avec la partie 3' de l'ADNc de muqueuse trachéobronchique JER58 et avec une partie de l'ADNc de muqueuse gastrique MAR2 (Figure 37). Ce petit fragment CEL2A caractérise donc une partie exonique du gène de MUC5AC.

JER 58 1 C C T C G G G C T C C T

JER 58 12 A C A A C C A G C A C A A C T T C T G G T C C T G G A A C T A C C C C A A G C C C T G T T C C C C
MAR 10 1 T C C A C T C C C G

JER 58 60 A C G A C C A G C A C A A C C T C T G C C C C T A C A A C A A G A A C A A C T T C T T G C T C C T
MAR 10 10 C A G A C C A G C A C A A C C T C T T C C C C T A C A A C T A G C A C A A C C T C A A G C T C C T

JER 58 108 A A A A G C A G C A C A A C C T C T G C C C G C T A C A A C C A G C A C A A C C T C T G G T C C C T
MAR 10 58 A C A A C C A G C A C A A C T T C T G C C C C T A C A A C C A G C A C A A C C T C C A C T C C A

JER 58 156 G A A A C T A C T C C T A G A C C T G T T C C T A C C A C C A G C A C A A C C T C T T C T C C C T
MAR 10 106 C A G A C C A G C A T A T C C T C T G C C C C C T A C A A G C A G C A C A A C C T C T G C T C C T

JER 58 204 A C A A C C A G C A C A A C C T C T G C T C C T A C A A C C A G C A C A A C C T C T G C T T C T
MAR 10 154 A C A G C C A G C A C A A T C T C T G C C C C T A C A A C C A G C

JER 58 252 A C A A C C A G C A C A A C C T C T G G T G C T G G A A C T A C T C C C A G C C C T G T T C C C

JER 58 300 A C C A C C A G C A C A A C C T C T G C T C C T A C A A C C A G A A C A A C C T C T G C C C C T

JER 58 348 A T A A G C A G C A C A A C C T C T G C C A C T A C A A C C A G C A C A A C C T C T G G T C C T

JER 58 396 G G A A C T A C T C C C A G C C C T G T T C C T A C C A C G A G C A C A A C C T C T G C T C C T

JER 58 444 A C A A C C A G C A C A A C C T C T G T G T C C T G G A A C T A C T C C C A G T G C T G T T C C C
MAR 11 1 C T T G T C C T G G A A C T A C T C C C A G T G C T G T T C C C

JER 58 492 A C C A C C A G T A T A A C C T C T G T A C C T A C A A C C A G T A C A A A C T C T G C C C C T
MAR 11 33 A C C A C C A G C A T A A C C T C T G C A C C T A C A A C C A G C A C A A A C T C T G C C C C T

JER 58 540 A T A A G C A G C A C A A C C T C T G C C A C T A C A A C C A G T A G A A T C T C T G G I C C T
MAR 2 1 C T G G T C C T
MAR 11 81 A T A A G C A G C A C A A C C T C T G C C A C T A C A A C C A G C A G A A T C T C T G G T C C T
CEL 2 1 C T

JER 58 588 G A A A C T A C T C C C A G C C C T G T T C C T A C C G C C A G C A C A A C C T C T G C T T C T
MAR 2 9 G A A A C T A C T C C C A G C C C T G T T C C T A C C G C C A G C A C A A C C T C T G C T T C T
MAR 11 129 G A A A C T A C T C C C A G C C C T G T T C C T A C C G C C A G C A C A A C C T C T G C T T C T
CEL 2 3 G A A A C T A C T C C C A G C C C T G T T C C T A C C G C C A G C A C A A C C T C T G C T T C T

JER 58 636 A C A A C T A G C A C A A C C T C T G G T C C T G G A A C T A C T C C C A G C C C T G T T C C T
MAR 2 57 A C A A C T A G C A C A A C C T C T G G T C C T G G A A C T A C T C C C A G C C C T G T T C C T
MAR 11 177 A C A A C T A G C A C A A C C T C T G G T C C T G G A A C T A C T C C C A G C C C T G T T C C T
CEL 2 51 A C A A C T A G C A C A A C C T C T G G T C C T G G A A C T A C T C C C A G C C C T G T T C C T

JER 58 684 A C C A C C A G C A C A A T C T C T G T T C C T A C C A C C A G C A C A A C T T C T G C T T C T
MAR 2 105 A C C A C C A G C A C A A T C T C T G T T C C T A C C A C C A G C A C A A C T T C T G C T T C T
MAR 11 225 A C C A C C A G C A C A A T C T C T G T T C C T A C C A C C A G C A C A A C T T C T G C T T C T
CEL 2 99 A C C A C C A G C A C A A T C T C T G T T C C T A C C A C C A G C A C A A C T T C T G C T T C T

JER 58 732 A C A A C C A G C A C A A C C T C T G C T T C T A C A A C C A G C A C A A C C T C T G G T C C T
MAR 2 153 A C A A C C A G C A C A A C C T C T G C T T C T A C A A C C A G C A C A A C C T C T G G T C C T
MAR 11 273 A C A A C C A G C A C A A T C T C T G C T T C T A G
CEL 2 147 A C A A C C A G C A C A A C C T C T G C T T C T A C A A C C A G C A C A A C C T C T G G T C C T

JER 58 780 G G A A C T A C T C C C A G C C C T G T T C C C A C C A C C A G C A C A A C C T C T G C T C C C
MAR 2 201 G G A A C T A C T C C C A G C C C T G T T C C C A C C A C C A G C A C A A C C T C T G C T C C C

CEL 2 195 G G A A C T A C T C C C A G C C C T G T T C C C A C C A C C A G C A C A A C C T C T G C T C C C

JER 58 828 A C A A C A A G C
MAR 2 249 A C A A C A A G C A C A A C C T C T G C C T C T A C A A C C A G C A C T A T C T C G

CEL 2 243 A C A A C A A G C A C A A C C T C T G C C T C T A C A A C C A G C A C T A T C T C G G

Figure 37 : Homologies de séquences entre les ADNc de muqueuse gastrique MAR2, MAR10, MAR11, l'ADNc de muqueuse trachéobronchique JER58 et le clone génomique CEL2.

L'existence d'une région 100% homologue existant dans l'ADNc de muqueuse gastrique, l'ADNc de muqueuse trachéobronchique et le clone CEL2A démontre la coexpression du gène MUC5AC dans les deux muqueuses respiratoire et gastrique. Cependant, cette coexpression du gène MUC5AC dans les deux muqueuses ne signifie pas pour autant que les molécules de mucines sécrétées soient identiques. La biosynthèse des mucines implique de nombreux mécanismes complexes post-traductionnels tels que, entre autres, la N-, la O-glycosylation et l'oligomérisation; aussi est-il possible d'imaginer qu'un peptide de mucine soit glycosylé ou mûri différemment selon la muqueuse qui le synthétise.

Les seules différences observées entre l'ADNc JER58, l'ADNc MAR2 et la partie exonique CEL2A se situent de part et d'autre de la séquence CEL2A, au niveau des deux sites de restriction Eco RI.

Quel que soit l'ADN humain étudié par Southern blot, aucun fragment Eco RI/Eco RI de 0.3 kb n'a été détecté avec la sonde JER58. Cet ADNc hybride avec un seul fragment d'une trentaine de kb.

Ces deux observations et le fait que ce clone soit fortement recombiné nous laissent supposer que les sites Eco RI de CEL2A n'existent pas naturellement dans le génome et que leur présence est due à un phénomène de recombinaison. Comme pour le clone génomique (GMUC) correspondant à MUC2, on peut supposer qu'une partie de la séquence répétée a été déléetée par recombinaison, ce qui expliquerait que l'on ne retrouve pas dans le clone CEL2 la partie 5' de JER58, MAR11 ou MAR10.

Ces problèmes de recombinaison n'ont malheureusement pas permis d'isoler d'exon entier, de partie intronique ou de site consensus d'épissage dans ce clone CEL2.

Afin de poursuivre l'étude génomique de MUC5AC, un deuxième criblage, cette fois ci de la banque cosmique a été effectué avec la sonde JER58, mais aucun clone positif n'a pu être isolé jusqu'à présent. Les clones du gène MUC5AC par leurs séquences en "tandem repeats" assez conservées semblent beaucoup plus instables que ceux du gène MUC5B possédant une séquence en "tandem repeats" plus dégénérée.

C. LA BANQUE GENOMIQUE HUMAINE EN VECTEUR λ EMBL4 CRIBLEE PAR LA SONDE JER57

1) Clones obtenus

Le criblage de la génothèque par la sonde JER57 (MUC5B) a permis d'isoler plusieurs clones, présentant tous la même carte de restriction. Nous avons étudié l'un d'entre eux : **CEL5**.

L'analyse en Southern blot de l'ADN humain grâce à la sonde JER57 montre un profil caractérisé par la présence de 3 fragments Bam HI de 1.7, 1.2 et 0.3 kb et une différence d'intensité selon les individus entre ces trois bandes. Cette différence d'intensité laisse supposer l'existence de plusieurs bandes de taille 1.7, 1.2 et 0.3 kb ainsi qu'une variation allélique des trois bandes.

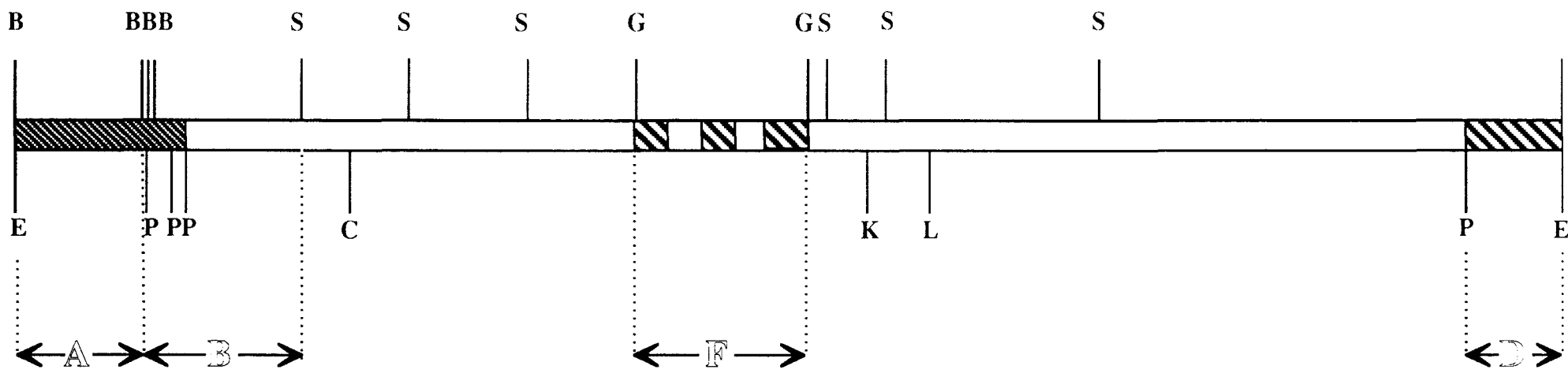
La sonde JER57 hybride avec le clone génomique CEL5, plus précisément, avec un fragment de 1.2 kb de ce clone coupé par Bam HI ainsi que de petits fragments de 40 et 50 pb qui ont été très difficiles à détecter. CEL5 contient donc la ou l'une des bandes Bam HI de 1.2 kb visualisées sur Southern blot.



2) Carte de restriction de CEL5

La carte de restriction du clone CEL5 est représentée sur la figure 38.

Afin d'identifier les parties exoniques du clone CEL5, nous l'avons divisé en petits fragments que nous avons ensuite testés sur Northern blot. La difficulté à se procurer des ARNm de tissus variés d'origine humaine et surtout de trachée n'a pas facilité ce travail de recherche des exons. De ce fait, cette étude n'a pas encore pu être effectuée sur la totalité du CEL5. Les fragments testés sur Northern blot (CEL5A, CEL5F, CEL5D) ont fourni les résultats présentés sur la figure 39. Quelques uns de ces fragments exoniques ont en partie été séquencés.

CEL5A montre le même profil d'hybridation sur Northern blot que JER 57, il est caractérisé par la même expression tissulaire. CEL5A est donc un fragment exonique du gène MUC5B. Les fragments CEL5F et CEL5D ont une expression tissulaire particulière, différente de celle de JER 57 mais qui laisse quand même supposer la présence d'une région exonique dans chacun des fragments.



 Séquence déterminée
 Séquence répétitive

B : Bam HI
 E : Eco RI
 K : Kpn I
 P : Pst I
 S : Sma I
 G : Bgl II
 C : Sac II
 L : Sal I

Figure 38 : Carte de restriction du clone génomique CEL5.

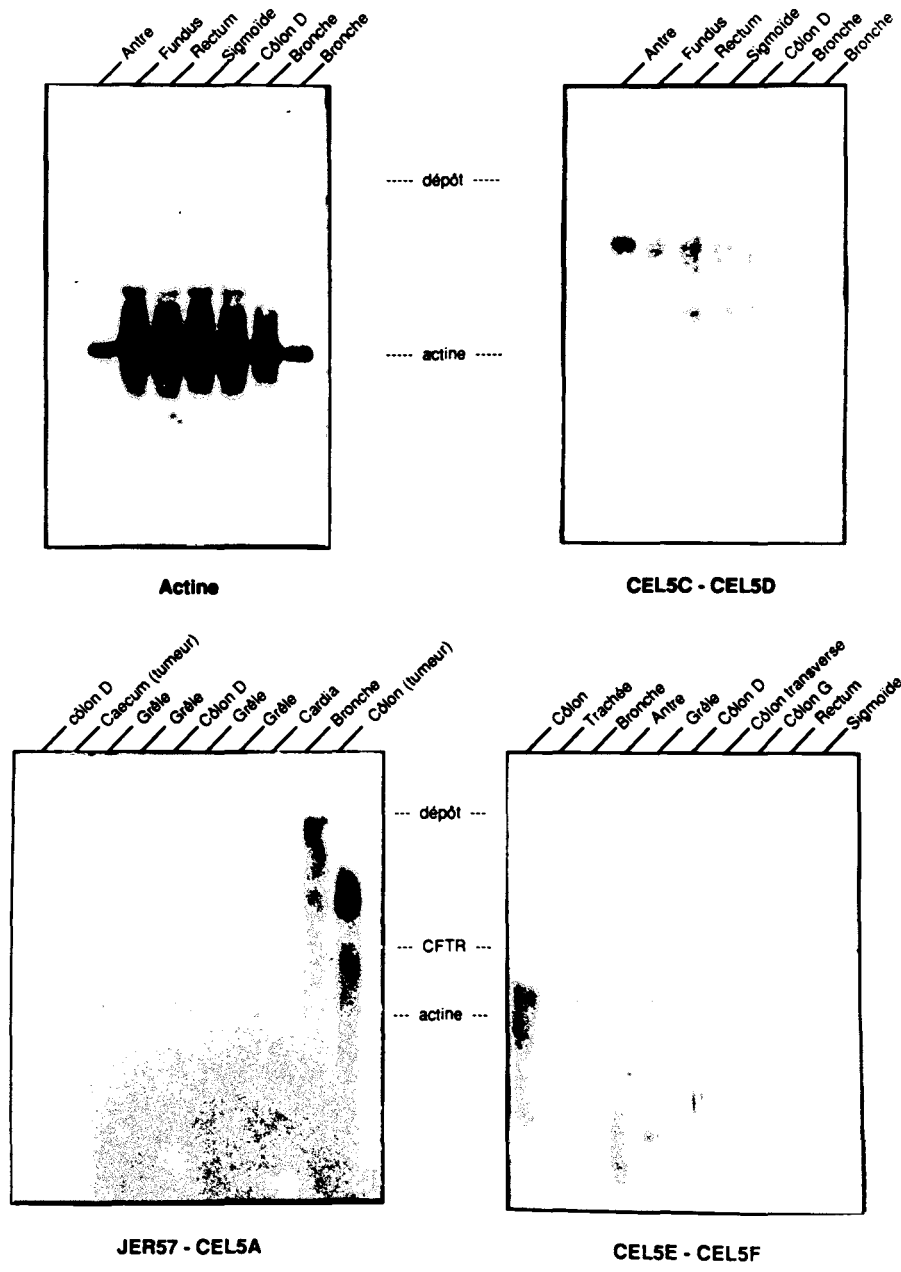
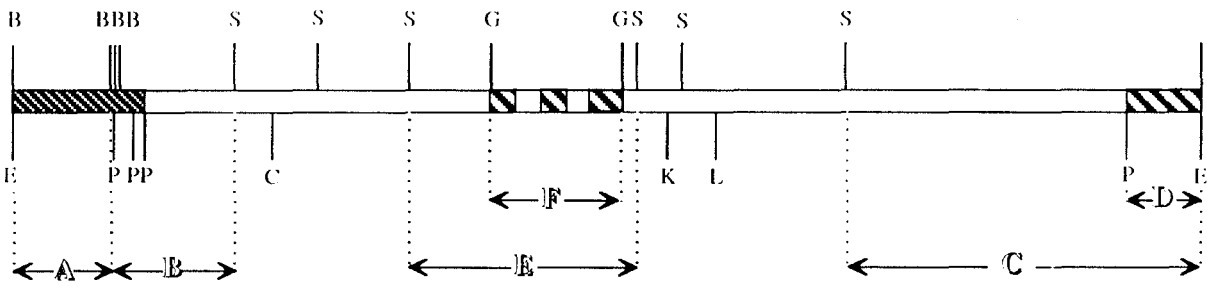


Figure 39 : Recherche des régions exoniques.

Etude en Northern blot de plusieurs fragments du clone CEL5.

3) *Séquence des fragments exoniques de CEL5*

Chacun des fragments séquencés ou partiellement séquencés a été comparé avec des banques de données. Aucune homologie significative n'a été détectée par rapport à des séquences publiées autres que celle de JER57.

- CEL5A-B :

Le fragment CEL5A et une partie du fragment CEL5B ont été séquencés (Figure 40). Cette séquence présente un cadre de lecture ouvert, comme le laissait supposer le signal observé sur Northern blot. Cette séquence est organisée en "tandem repeats" irréguliers de 87 pb en moyenne.

Le premier cadre de lecture, le seul ouvert d'un bout à l'autre de la séquence est caractérisé par une richesse en acides aminés sérine (15.3%), thréonine (30.7%), proline (10.6%) et surtout histidine (4%) tout comme les séquences de la famille MUC5B (Figure 41). On peut également mettre en évidence une alternance de domaines hydrophiles et de domaines hydrophobes, spécifique des clones ADNc JER57, JER28, JUL7 et JUL10 (Figure 42).

Néanmoins, on ne retrouve dans cette séquence, aucune homologie parfaite avec ces ADNc partiels. Il s'agit donc d'une nouvelle séquence appartenant aussi à cette famille qui comporte déjà 5 kb d'ADNc partiels.

- CEL5F :

Les parties du fragment CEL5F qui ont été séquencées sont indiquées sur la figure 43, leurs séquences sont représentées sur la figure 44 A, B, C. Les séquences partielles du fragment CEL5F (CEL5F₁, CEL5F₂, CEL5F₃) ne sont pas organisées en "tandem repeats" et ne possèdent aucune homologie avec les séquences précédentes.

Le signal observé en Northern blot laisse suspecter la présence d'exons. Chacune de ces séquences possède un cadre ouvert de lecture, mais sans ARNm correspondant, nous ne pouvons pas privilégier un cadre plutôt qu'un autre. Il sera donc nécessaire, afin d'exploiter ces résultats, d'isoler l'ARNm correspondant. Ces séquences vont ainsi nous permettre de choisir des amorces adéquates dans le but de synthétiser l'amplifiat ADNc correspondant à cette région exonique de MUC5B.

Figure 41 : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de CEL5A-B.

Bam III 10 20 30 40 50 60
 GGATCCACGGCCACCGCCTCCTCCACCCAGGCAACTGCTGGCACCCACATGTGAGCACCACGGCCACG
 GlySerThrAlaThrAlaSerSerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThr
 AspProArgProProProProProProArgGlnLeuLeuAlaProHisMET---AlaProArgProArg
 IleHisGlyHisArgLeuLeuHisProGlyAsnCysTrpHisProThrCysGluHisHisGlyHisAsp

70 80 90 100 110 120 130
 ACACCCACAGTCACCAGCTCCAAAGCCACTCCCTCCTCCAGTCCAGGGACTGCAACTGCCCTTCCAGCA
 ThrProThrValThrSerSerLysAlaThrProSerSerSerProGlyThrAlaThrAlaLeuProAla
 HisProGlnSerProAlaProLysProLeuProProProValGlnGlyLeuGlnLeuProPheGlnHis
 ThrHisSerHisGlnLeuGlnSerHisSerLeuLeuGlnSerArgAspCysAsnCysProSerSerThr

140 150 160 170 180 190 200
 CTGAGAAGCACAGCCACCACCCACAGCTACCAGCTTTACAGCCATCCCCTCCTCCTCCCTGGGCACC
 LeuArgSerThrAlaThrThrProThrAlaThrSerPheThrAlaIleProSerSerSerLeuGlyThr
 ---GluAlaGlnProProHisProGlnLeuProAlaLeuGlnProSerProProProProTrpAlaPro
 GluLysHisSerHisHisThrHisSerTyrGlnLeuTyrSerHisProLeuLeuLeuProGlyHisHis

210 220 230 240 250 260 270
 ACCTGGACCCGCCTATCACAGACCACCACCCACGGCCACCATGTCCACAGCCACACCCTCCTCCACT
 ThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMETSerThrAlaThrProSerSerThr
 ProGlyProAlaTyrHisArgProProHisProArgProProCysProGlnProHisProProProLeu
 LeuAspProProIleThrAspHisHisThrHisGlyHisHisValHisSerHisThrLeuLeuHisSer

280 290 300 310 320 330 340
 CCAGAGACTGTCCACACCTCCACAGTGTGCTTACCACCACGGCCACCACAACCGGGCACCAGGCTCTGTGG
 ProGluThrValHisThrSerThrValLeuThrThrThrAlaThrThrThrGlyAlaProAlaLeuTrp
 GlnArgLeuSerThrProProGlnCysLeuProProArgProProGlnProGlyHisArgLeuCysGly
 ArgAspCysProHisLeuHisSerAlaTyrHisHisGlyHisHisAsnArgGlyThrGlySerValAla

350 360 370 380 390 400 410
 CCACCCCTCCTCCACCCAGGAACAGCTCACACTACCAAAGTGGGACTACCACAACCACGGGCTTCAC
 ProProProProProProGlnGluGlnLeuThrLeuProLysCysAspTyrHisAsnHisGlyLeuHis
 HisProLeuLeuHisProArgAsnSerSerHisTyrGlnSerAlaThrThrThrThrThrGlyPheThr
 ThrProSerSerThrProGlyThrAlaHisThrThrLysValArgLeuProGlnProArgAlaSerGln

420 430 440 450 460 470 480
 AGCCACCCCTCCTCCAGCCAGGGACGGCACTCACGCCTCCAGTGTGGATCAGCACAACCACCACACC
 SerHisProLeuLeuGlnProArgAspGlyThrHisAlaSerSerValAspGlnHisAsnHisHisThr
 AlaThrProSerSerSerProGlyThrAlaLeuThrProProValTrpIleSerThrThrThrThrPro
 ProProProProProAlaGlnGlyArgHisSerArgLeuGlnCysGlySerAlaGlnProProHisPro

490 500 510 520 530 540 550
 CACAACCACCACCCACAACCAGTGGCTCCACGGTGACCCCTCCTCCATCCCAGGGACCACCCACACC
 HisAsnHisHisThrHisAsnGlnTrpLeuHisGlyAspProSerSerSerIleProGlyThrThrHisThr
 ThrThrThrThrProThrThrSerGlySerThrValThrProProProSerGlnGlyProProThrPro
 GlnProProHisProGlnProValAlaProArg---ProLeuLeuHisProArgAspHisProHisArg

560 570 580 590 600 610 620
 GCCAGAGTGTGACCACCACCACCAACTGTGGCCACTGGTCTATGGCAACACCCTCCTCTAGCACA
 AlaArgValLeuThrThrThrThrThrValAlaThrGlySerMETAlaThrProSerSerSerThr
 ProGluCys---ProProProProGlnLeuTrpProLeuValLeuTrpGlnHisProProLeuAlaHis
 GlnSerAlaAspHisHisHisHisAsnCysGlyHisTrpPheTyrGlyAsnThrLeuLeu---HisThr

630 640 650 660 670 680 690
 CAGACCAGTGGTACTCCCCATCACTGACCACCACGGCCACTACGATCACGGCCACCGGCTCCACCACC
 GlnThrSerGlyThrProProSerLeuThrThrAlaThrThrIleThrAlaThrGlySerThrThr
 ArgProValValLeuProHisHis---ProProArgProLeuArgSerArgProProAlaProPro
 AspGlnTrpTyrSerProIleThrAspHisHisGlyHisTyrAspHisGlyHisArgLeuHisHisGln

700 710 720 730 740 750
 AACCCCTCCTCAACTCCAGGGACAACACCCATCCCCCAGTGCTGACCAGCATGGCCACCACACCCGCA
 AsnProSerSerThrProGlyThrThrProIleProProValLeuThrSerMETAlaThrThrProAla
 ThrProProGlnLeuGlnGlyGlnHisProSerProGlnCys---ProAlaTrpProProHisProGln
 ProLeuLeuAsnSerArgAspAsnThrHisProProSerAlaAspGlnHisGlyHisHisThrArgSer

760 770 780 790 800 810 820
 GCCACCAGCTCCAAAGCCACTTCCTCCTCCAGTCCAAGGACTGCAACCACCCTTCCAGTGCTGACAAGC
 AlaThrSerSerLysAlaThrSerSerSerSerProArgThrAlaThrThrLeuProValLeuThrSer
 ProProAlaProLysProLeuProProProValGlnGlyLeuGlnProProPheGlnCys---GlnAla
 HisGlnLeuGlnSerHisPheLeuLeuGlnSerLysAspCysAsnHisProSerSerAlaAspLysHis

830 840 850 860 870 880 890
 ACAGCCACAAAATCCACAGCTACCAGCTTTACACCCATCCCCTCCTCCACCCTGTGGACCACGTGGACC
 ThrAlaThrLysSerThrAlaThrSerPheThrProIleProSerSerThrLeuTrpThrThrTrpThr
 GlnProGlnAsnProGlnLeuProAlaLeuHisProSerProProProProCysGlyProArgGlyPro
 SerHisLysIleHisSerTyrGlnLeuTyrThrHisProLeuLeuHisProValAspHisValAspArg

900 910 920 930 940 950 960
 GTCCAGCACAGACCACCACACCCATGTCCACCATGTCCACAATCCACACCTCCTCTACTCCAGAGACC
 ValProAlaGlnThrThrThrProMETSerThrMETSerThrIleHisThrSerSerThrProGluThr
 SerGlnHisArgProProHisProCysProProCysProGlnSerThrProProLeuLeuGlnArgPro
 ProSerThrAspHisHisThrHisValHisHisValHisAsnProHisLeuLeuTyrSerArgAspHis

970 980 990 1000 1010 1020 1030
 ACCCACACCTCCACAGTGCTGACCACCACAGCCACCATGACAAGGGCCACCAATTCCACGGCCACACC
 ThrHisThrSerThrValLeuThrThrThrAlaThrMETThrArgAlaThrAsnSerThrAlaThrPro
 ProThrProProGlnCys---ProProGlnProPro---GlnGlyProProIleProArgProHisPro
 ProHisLeuHisSerAlaAspHisHisSerHisHisAspLysGlyHisGlnPheHisGlyHisThrLeu

1040 1050 1060 1070 1080 1090 1100
 TCCTCCACTCTGGGGACGACCCGGATCCTCACTGAGCTGACCACAACAGCCACTACAACCTGCAGCCACT
 SerSerThrLeuGlyThrThrArgIleLeuThrGluLeuThrThrThrAlaThrThrAlaAlaThr
 ProProLeuTrpGlyArgProGlySerSerLeuSer---ProGlnGlnProLeuGlnLeuGlnProLeu
 LeuHisSerGlyAspAspProAspProHis---AlaAspHisAsnSerHisTyrAsnCysSerHisTrp

1110 1120 1130 1140 1150 1160 1170
Bam HI | | | | Bam HI | | |
 GGATCCACGGCCACCCTGTCCCTCCACCCCAGGGACCACCTGGATCCTCACAGAGCCGAGCACTATAGCC
 GlySerThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluProSerThrIleAla
 AspProArgProProCysProProProGlnGlyProProGlySerSerGlnSerArgAlaLeu---Pro
 IleHisGlyHisProValLeuHisProArgAspHisLeuAspProHisArgAlaGluHisTyrSerHis

1180 1190 1200 1210 1220 1230 1240
 ACCGTGATGGTGGCCACCGGTTCCACGGCCACCACCTCCTCCACTCTGGGAACAGCTCACACCCCAAA
 ThrValMETValProThrGlySerThrAlaThrThrSerSerThrLeuGlyThrAlaHisThrProLys
 Pro---TrpCysProProValProArgProProProProProLeuTrpGluGlnLeuThrProProLys
 ArgAspGlyAlaHisArgPheHisGlyHisHisLeuLeuHisSerGlyAsnSerSerHisProGlnSer

1250 1260 1270 1280 1290 1300 1310
 GTGGTGACCGCCATGGCCACTATGCCACAGCCACTGCCTCCACGGTCCCAGCTCGTCCACCGTGGGG
 ValValThrAlaMETAlaThrMETProThrAlaThrAlaSerThrValProSerSerSerThrValGly
 Trp---ProProTrpProLeuCysProGlnProLeuProProArgPheProAlaArgProProTrpGly
 GlyAspArgHisGlyHisTyrAlaHisSerHisCysLeuHisGlySerGlnLeuValHisArgGlyAsp

1320 1330 1340 1350 1360 1370 1380
 ACCACCGCACCCCTGCAGTGCTCCCCAGCAGCCTGCCAACCTTCAGCGTGTCCACTGTGTCCTCCTCA
 ThrThrArgThrProAlaValLeuProSerSerLeuProThrPheSerValSerThrValSerSerSer
 ProProAlaProLeuGlnCysSerProAlaAlaCysGlnProSerAlaCysProLeuCysProProGln
 HisProHisProCysSerAlaProGlnGlnProAlaAsnLeuGlnArgValHisCysValLeuLeuSer

1390 1400 1410 1420 1430 1440 Pst I
 GTCCTACACCCTGAGACCCACTGGCTTCCCAGCTCCCCTTCTCTACTCCCTGCTTCTGCAG
 ValLeuThrThrLeuArgProThrGlyPheProSerSerHisPheSerThrProCysPheCys
 SerSerProPro---AspProLeuAlaSerProAlaProThrSerLeuLeuProAlaSerAla
 ProHisHisProGluThrHisTrpLeuProGlnLeuProLeuLeuTyrSerLeuLeuLeuGln

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

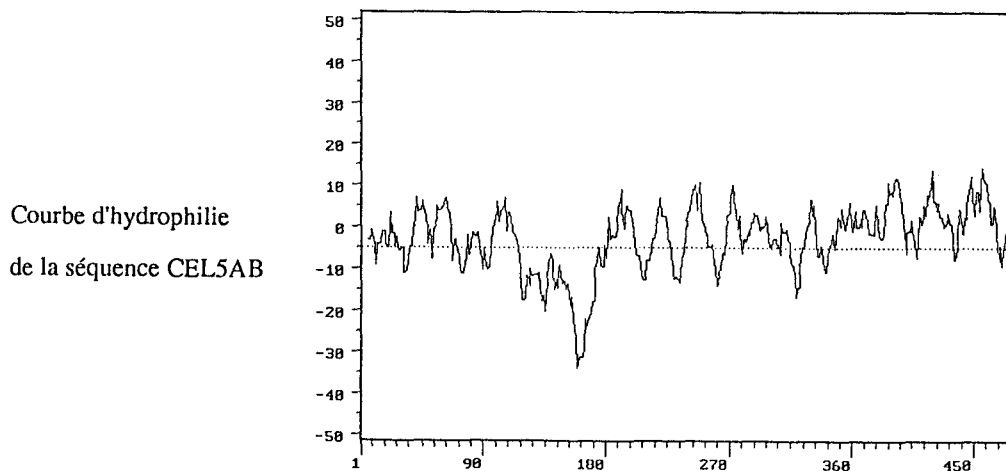


Figure 42 : Alternance de domaines hydrophiles / hydrophobes au niveau peptidique de CEL5A-B.

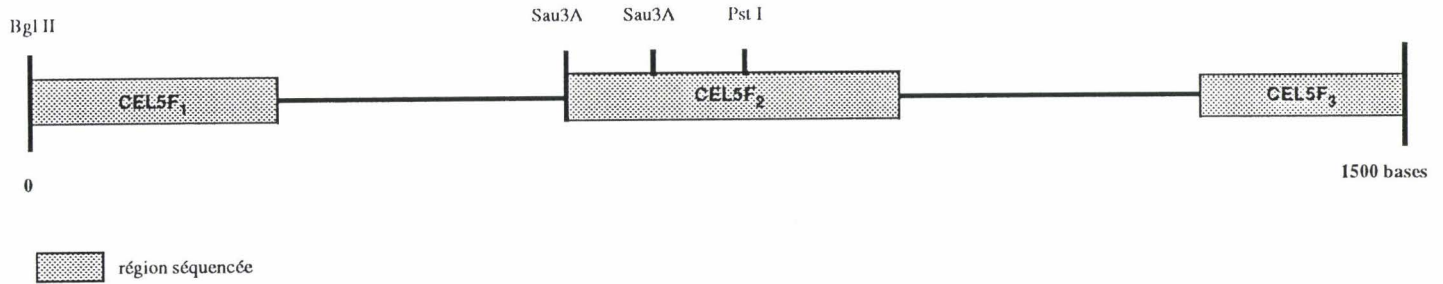


Figure 43 : Représentation des régions séquencées du fragment CEL5F.



Figure 44.A : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de CEL5F₁.

B

10 20 30 40 50 60
 Sau 3A | | | | | SPI
 GATCTGATGCTGAGCCAGTGAGTCTCTCCCTCGGGGGTTGCAGGCCCTGGGGTCCCCCGCCGCCGCAT
 AspLeuMETLeuSerGln---ValLeuProSerGlyValAlaGlyProGlyValProArgArgProHis
 Ile---Cys---AlaSerGluSerSerProArgGlyLeuGlnAlaLeuGlySerProAlaAlaArgMET
 SerAspAlaGluProValSerProProLeuGlyGlyCysArgProTrpGlyProProProProAlaCys

 70 80 90 100 110 120 130
 GCACGCACGCACGCAGCTCCCTGGGGCTGGGGCCTCCGTCCTGATCGCTTCCCCACAGGGTCTTTG
 AlaArgThrHisAlaAlaProTrpGlyTrpGlyProProSer---SerLeuSerProThrGlySerLeu
 HisAlaArgThrGlnLeuProGlyAlaGlyGlyLeuArgProAspArgPheProProGlnGlyLeuCys
 ThrHisAlaArgSerSerLeuGlyLeuGlyAlaSerValLeuIleAlaPheProHisArgValPheAla

 140 150 160 170 180 190 200
 CTGAGTGCCACAACCTTGTGCCCCCGGGCCATTCTCAACGCCTGCATCAGCGACCCTGCAGGGGCC
 LeuSerAlaThrThrLeuCysProArgAlaHisSerSerThrProAlaSerAlaThrThrAlaGlyAla
 ---ValProGlnProCysAlaProGlyProIleLeuGlnArgLeuHisGlnArgProLeuGlnGlyPro
 GluCysHisAsnLeuValProProGlyProPhePheAsnAlaCysIleSerAspHisCysArgGlyArg

 210 220 230 240 250 260 270
 GCCTTGAGGTGCCCTGCCAGAGCCTGGAGGCTTACGCAGAGCTCTGCCGCCCGGGGAGTGTGCAGTG
 AlaLeuArgCysProAlaArgAlaTrpArgLeuThrGlnSerSerAlaAlaProGlyGluCysAlaVal
 Pro---GlyAlaLeuProGluProGlyGlyLeuArgArgAlaLeuProProProGlySerValGln---
 LeuGluValProCysGlnSerLeuGluAlaTyrAlaGluLeuCysArgProArgGlyValCysSerAsp

 280 290 300 310 320 330 340
 ACTGGCGAGGTGCAACCGGTGGCCTGTTCGGTGAGTGGGGGGCGGCCCGGCCAGACCCCTCTTCC
 ThrGlyGluValGlnProValAlaCysSerValSerGlyGlyGlyProGlyProProAspProSerSer
 LeuAlaArgCysAsnArgTrpProValArg---ValGlyAlaAlaProGlyProGlnThrProLeuPro
 TrpArgGlyAlaThrGlyGlyLeuPheGlyGluTrpGlyArgProArgAlaProArgProLeuPheLeu

 350 360 370
 TCTCTGAGTGTCTGTGCGGTGAGTGGGG
 SerLeuSerValLeuCysGlyGluTrp
 Leu---ValSerCysAlaValSerGly
 SerGluCysProValArg---ValGly

C

10 20 30 40 50 60
 GTGTCCTGTGCGGTGATTGGGCGCCCTGGGCCCCCAACCCCTTGGCTTGTCTGACACCTCTCTGTGC
 ValSerCysAlaValIleGlyArgProGlyProProGlnProLeuGlyLeuSerAspThrSerLeuCys
 CysProValArg---LeuGlyAlaLeuGlyProProAsnProLeuAlaCysLeuThrProLeuCysAla
 ValLeuCysGlyAspTrpAlaProTrpAlaProProThrProTrpLeuVal---HisLeuSerValPro

 70 80 90 100 110 120 130
 CCACAGACCTCACCTGCCACCCACCAAGTGTACAACGCATGCGGCCCAACACAGCCTGCCACCTGCA
 ProGlnThrSerProAlaHisProProLysCysThrThrHisAlaAlaProTyrSerLeuProProAla
 HisArgProHisLeuProThrHisGlnSerValGlnArgMETArgProHisThrAlaCysHisLeuGln
 ThrAspLeuThrCysProProThrLysValTyrAsnAlaCysGlyProIleGlnProAlaThrCysAsn

 140 150 160 170 180 190 200
 ACTCTAGGTAATACAGGGATGGCTGGTGCCTTCCCTGCCTCCCAGGCCTGCCTGACCGGTCTGGGGGAG
 ThrLeuGlyAsnThrGlyMETAlaGlyAlaPheProAlaSerGlnAlaCysLeuThrGlyLeuGlyGlu
 Leu---ValIleGlnGlyTrpLeuValProSerLeuProProArgProAla---ProValTrpGlySer
 SerArg---TyrArgAspGlyTrpCysLeuProCysLeuProGlyLeuProAspArgSerGlyGlyAla

 210 220 230 240
 CAGGAGGAGGCCAGAGGGTGTCCCACTGTGGGCCACAGATCT
 GlnGluGluAlaArgGlyCysProThrValGlyHisArgSer
 ArgArgArgProGluGlyValProLeuTrpAlaThrAsp
 GlyGlyGlyGlnArgValSerHisCysGlyProGlnIle

Figure 44 : Séquences nucléotidiques et peptidiques déduites dans les trois cadres de lecture de CEL5F₂ (B) et de CEL5F₃ (C).

La séquence CEL5F₂ présente trois sites potentiels de reconnaissance de la protéine stimulatrice SP1 [187]. L'existence de tels sites permet de supposer que ce fragment est une séquence intronique qui aurait peut-être un rôle dans la stimulation de l'expression du gène MUC5B.

- CEL5D :

La séquence CEL5D (Figure 45) n'est pas organisée en "tandem repeats" et ne possède pas d'homologie avec des séquences connues (banques de données).

D'autre part, aucun des cadres de lecture n'est ouvert d'un bout à l'autre de la séquence, ce qui pourrait révéler la présence de la partie terminale non codante du gène. La présence à l'extrémité 3' d'un site potentiel de polyadénylation (AATAAA) (bases : 733-738) est un élément favorable à cette hypothèse et nous permet de supposer que CEL5D constitue la partie 3' terminale du gène [188]. De plus, la séquence non codante en aval du site (AATAAA) jusqu'à l'extrémité 3' est une séquence répétitive du génome humain de type "Alu family".

On peut déceler plusieurs sites potentiels d'épissage qui pourraient signaler la présence d'un intron [189] en particulier un site donneur (CAGGAAGT : position 95) et un site accepteur GAGG (position 202) précédé d'une séquence typique riche en bases pyrimidiques (position 180 à 199).

La partie 3' terminale du gène MUC2, en aval de la région répétitive, comprend au moins 7 introns dont certains sont assez grands (727 pb). L'ADNc de 2600 bases correspondant à cette extrémité est donc constitué de nombreux petits exons, au moins 8 qui peuvent avoir une taille moyenne de 300 pb. Si le gène MUC5B est construit de manière analogue à MUC2, on peut supposer qu'il puisse exister de nombreux petits exons dans la partie 3' du gène. Ces exons vont être difficiles à délimiter sans la connaissance de la séquence complète de l'ARNm correspondant. C'est pourquoi nous avons envisagé de réaliser une amplification par PCR à partir d'oligonucléotides localisés de part et d'autre des sites potentiels d'épissage. Malheureusement, la recherche informatique n'a pas permis de définir de couples d'oligonucléotides compatibles (problèmes de répétitions de séquences, de conformations).

Figure 45 : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de CEL5D.

Pst I 10 20 30 40 50 60
 CTGCAGAAGCCTGTGGTGTGGGTGCAGAGCGGGGCTTCCCCACACCGGGGCCTGCCTCTGCCCTGTGC
 LeuGlnLysProValValTrpValGlnSerGlyAlaPheProThrProGlyProAlaSerAlaLeuCys
 CysArgSerLeuTrpCysGlyCysArgAlaGlyProSerProHisArgGlyLeuProLeuProCysAla
 AlaGluAlaCysGlyValGlyAlaGluArgGlyLeuProHisThrGlyAlaCysLeuCysProValPro

 70 80 90 100 110 120 130
 CCTTGGGTGGGTCTGCCCTGCCAGGAAGTGATGCCATCACACTGTACATTTACAGAGCTCCCAGCTC
 ProTrpValGlyProAlaLeuProArgLys---CysHisHisThrValHisLeuGlnSerSerGlnLeu
 LeuGlyTrpValLeuProCysProGlySerAspAlaIleThrLeuTyrIleTyrArgAlaProSerSer
 LeuGlyGlySerCysProAlaGlnGluValMETProSerHisCysThrPheThrGluLeuProAlaPro

 140 150 160 170 180 190 200
 CTGGGGGCCAGCCAGGCCCTGCCAGCGTCCAGAGCTCCAGCCTCTCCTGCCCTCCTTGCCGAGGCCCA
 LeuGlyAlaGlnProGlyProAlaSerValGlnSerSerSerLeuSerCysProProCysArgGlyPro
 TrpGlyProSerGlnAlaLeuProAlaSerArgAlaProAlaSerProAlaLeuLeuAlaGluAlaGln
 GlyGlyProAlaArgProCysGlnArgProGluLeuGlnProLeuLeuProSerLeuProArgProSer

 210 220 230 240 250 260 270
 GCAGGCGAGCCACCCAGGACTGGGCTGGAGATAGGCAGGGGCCTGACAAGGTTGCAGGTGGGCTCTGG
 AlaGlyGluProProGlnAspTrpAlaGlyAspArgGlnGlyProAspLysValAlaGlyGlyLeuTrp
 GlnAlaSerHisProArgThrGlyLeuGluIleGlyArgGlyLeuThrArgLeuGlnValGlySerGly
 ArgArgAlaThrProGlyLeuGlyTrpArg---AlaGlyAla---GlnGlyCysArgTrpAlaLeuVal

 280 290 300 310 320 330 340
 TGGGCGGTGGGCTAGGTTGGTCACTGGGGTGAGCTTCTATAGGTTTACCCTGCGTCTCCGAGGCTG
 TrpAlaValGly---ValGlyHisHisTrpGlyGluLeuLeu---ValTyrProAlaSerProArgLeu
 GlyArgTrpAlaArgLeuValThrThrGlyValSerPheTyrArgPheThrLeuArgLeuArgGly---
 GlyGlyGlyLeuGlyTrpSerProLeuGly---AlaSerIleGlyLeuProCysValSerGluAlaAsp

 350 360 370 380 390 400 410
 ATTGTCTGGAGGTTGCGCCAGGTCACCTGCCTGACCTGGGTCTGCATTCCGCCACGTGTAAACAAATGGT
 IleValTrpArgLeuArgGlnValThrAla---ProGlySerAlaPheArgHisVal---ThrAsnGly
 LeuSerGlyGlyCysAlaArgSerLeuProAspLeuGlyLeuHisSerAlaThrCysLysGlnMETVal
 CysLeuGluValAlaProGlyHisCysLeuThrTrpValCysIleProProArgValAsnLysTrpPhe

 420 430 440 450 460 470 480
 TTCCGAAACCTCAGTCATTCCCATGGTGCATGGTTTCCCTTAGACCCATTTATGTGGTGTATGTTACA
 PheArgAsnLeuSerHisSerHisGlyAlaMETValSerLeuAspProPheMETSerValTyrGlyThr
 SerGluThrSerValIleProMETValArgTrpPhePro---ThrHisLeuCysArgCysMETValGln
 ProLysProGlnSerPheProTrpCysAspGlyPheLeuArgProIleTyrValGlyValTrpTyrLys

 490 500 510 520 530 540 550
 AGCACCCCTCTCAGAGGGCCCATGCTGTGCTGTCTCTGTGCCCCAGAACCCTATCATTCAATCCTGACAA
 SerThrProLeuArgGlyProCysCysAlaValSerValProGlnAsnArgIleIleGlnSer---Gln
 AlaProLeuSerGluAlaHisAlaValLeuSerLeuCysProArgThrValSerPheAsnProAspAsn
 HisProSerGlnArgProMETLeuCysCysLeuCysAlaProGluProTyrHisSerIleLeuThrIle

560 570 580 590 600 610 620
 | | | | | | |
 TACCCTGTGTGGTGTCTGCTATTGTGTCTCCCACTCCACAGATGGGAACACTGAGGCACAAAGCAATGA
 TyrProValTrpCysLeuLeuLeuCysLeuProLeuHisArgTrpGluHis---GlyThrLysGln---
 ThrLeuCysGlyValCysTyrCysValSerHisSerThrAspGlyAsnThrGluAlaGlnSerAsnGlu
 ProCysValValSerAlaIleValSerProThrProGlnMETGlyThrLeuArgHisLysAlaMETLys

630 640 650 660 670 680 690
 | | | | | | |
 AAAGGTGTGGTTCTGTTCAGGGGAAAGCTGGGCTCACATCCAGGCAGCCCAAGTCGAGGCCACACTCAA
 LysGlyValValLeuSerGlyGlyLysLeuGlySerHisProGlySerProSerArgGlyHisThrGln
 LysValTrpPheCysGlnGlyGluSerTrpAlaHisIleGlnAlaAlaGlnValGluAlaThrLeuLys
 ArgCysGlySerValArgGlyLysAlaGlyLeuThrSerArgGlnProLysSerArgProHisSerLys

700 710 720 730 740 750
 | | | | | | |
 AACCACAGGGCTCTGCAATGCCCCAGAGCAAGTAAGTATAAAAAATAAACACAGTGGGCCAGGCGCGGTG
 AsnHisArgAlaLeuGlnCysProArgAlaSerAsnTyrLysAsnLysHisSerGlyProGlyAlaVal
 ThrThrGlyLeuCysAsnAlaProGluGlnValThrIleLysIleAsnThrValGlyGlnAlaArgTrp
 ProGlnGlySerAlaMETProGlnSerLys---Leu---Lys---ThrGlnTrpAlaArgArgGlyGly

760 770 780 790 800 810 820
 | | | | | | |
 GCTCATGCCTGTAATCCCAGCATTGTTGGGAGGCTGAGACGGGCGGATTACAGGTCAGGAGATCCGGGGA
 AlaHisAlaCysAsnProSerIleLeuGlyGly---AspGlyArgIleThrGlyGlnGluIleArgGly
 LeuMETProValIleProAlaPheTrpGluAlaGluThrGlyGlyLeuGlnValArgArgSerGlyGlu
 SerCysLeu---SerGlnHisPheGlyArgLeuArgArgAlaAspTyrArgSerGlyAspProGlyAsn

830
 |
 ATTC
 Ile
 Phe

Considérant la séquence nucléotidique en amont du premier site donneur, on constate que celle-ci code pour une région potentiellement riche en cystéine (5 résidus de cystéine pour 31 résidus d'acides aminés soit 16%, lisible dans le 2^{ème} cadre). Une telle richesse a déjà été observée au niveau des domaines C-terminaux de mucines sécrétées tels que ceux de MUC2 (Cys : 9%) [96], d'une mucine sous-maxillaire de porc (Cys : 12%) [190] et d'une mucin-like intestinale de rat [191]. Cette séquence riche en cystéine est un argument de plus en faveur de la présence, au niveau de CEL5D, de l'extrémité 3' terminale du gène MUC5B.

Afin d'authentifier le site de poly-adénylation et d'établir l'agencement précis exon-intron de cette région, il sera indispensable d'isoler le clone d'ADNc correspondant à l'extrémité 3' terminale du gène MUC5B.

Afin d'isoler la partie du génome contenant les fragments Bam HI/Bam HI de 1.7, 0.3 kb et peut être d'autres fragments 1.2 kb correspondant à MUC5B, nous avons réalisé un nouveau criblage avec la sonde JER57 et des sondes des extrémités 5' et 3' du clone CEL5. Aucun clone différent de celui-ci n'a pu être isolé à partir de la banque en vecteur λ EMBL4. Nous avons donc été amenés à utiliser une nouvelle banque, cette fois-ci cosmétique pouvant contenir des inserts d'une quarantaine de kilobases.

D. LA BANQUE GENOMIQUE HUMAINE EN VECTEUR COSMIDIQUE CRIBLEE PAR LA SONDE JER57

Un premier criblage de cette banque a permis l'obtention d'un clone cosmétique BEN1, hybridant avec la sonde JER57.

1) Carte de restriction de BEN1

L'existence d'un site de restriction Not I dans l'insert n'a pas permis de construire les deux cartes partielles chevauchantes (T3, T7) et n'a donc, par conséquent, pas facilité le positionnement de chacun des sites. C'est pourquoi la cartographie de la région centrale est encore un peu imprécise (Figure 46).

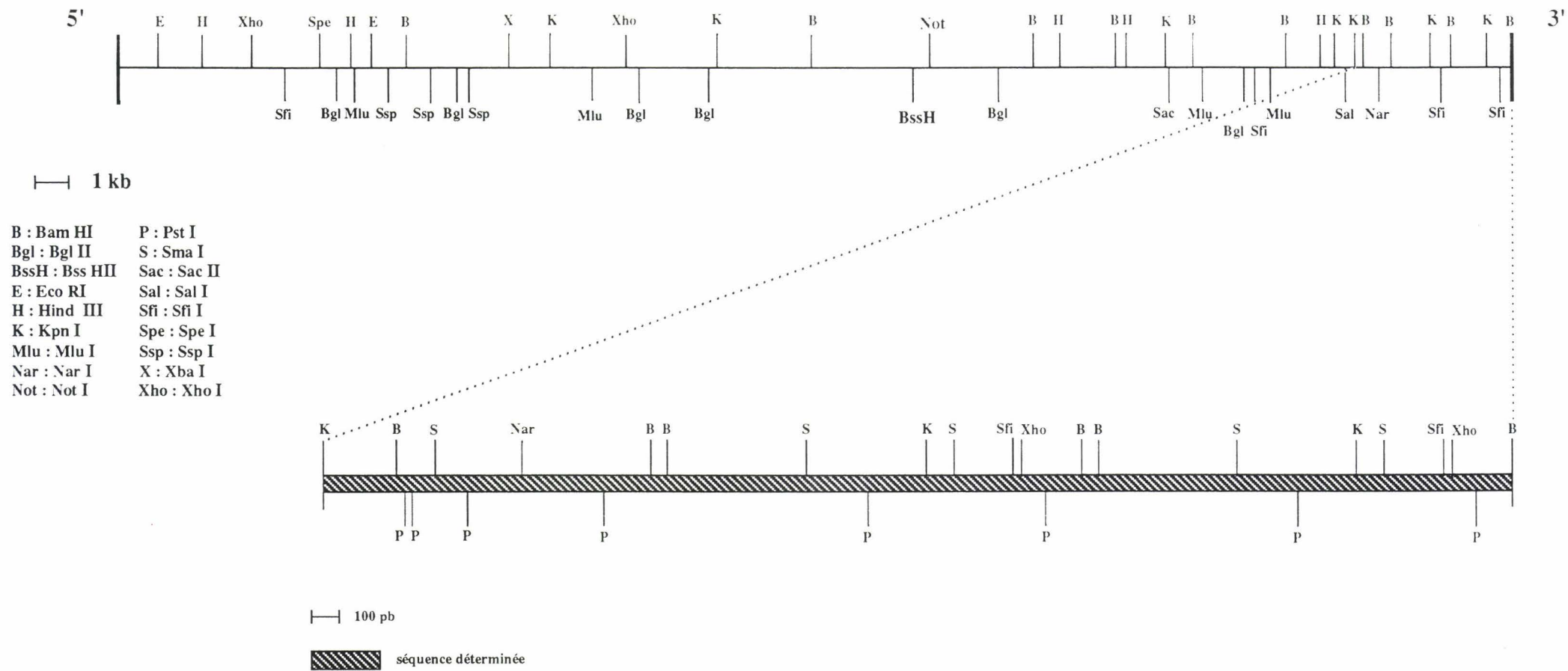


Figure 46 : Carte de restriction du clone cosmidique BEN1

La carte de restriction de BEN1 met en évidence la présence de deux sites rares Not I et Bss HII à 0,5 kb l'un de l'autre. Ce fragment de 0.5 kb Not I / Bss HII a été séquencé. Sa séquence présente un pourcentage élevé de résidus (G + C) (66%) caractéristique des îlots CpG. En effet, dans le reste du génome le taux de G + C est de l'ordre de 40% seulement. La proximité de ces deux sites rares et la forte concentration des nucléotides G et C sont en faveur de l'existence d'un îlot CpG. Ces deux sites correspondent respectivement aux sites Not I₁ et Bss HII₂ de la cartographie de la région 11p15.5 décrite précédemment et définissent l'îlot CpG "I_g" (Figure 47), annonciateur du gène MUC5B.

Dans la littérature [173], il est décrit que les îlots CpG ont une taille moyenne de 1 à 2 kb et recouvrent en général les deux premiers exons du gène et leurs séquences promotrices. La séquence de ce petit fragment possède deux sites potentiels de fixation de la protéine stimulatrice SP1, permettant de suspecter l'implication de cette région dans la régulation de l'expression du gène MUC5B.

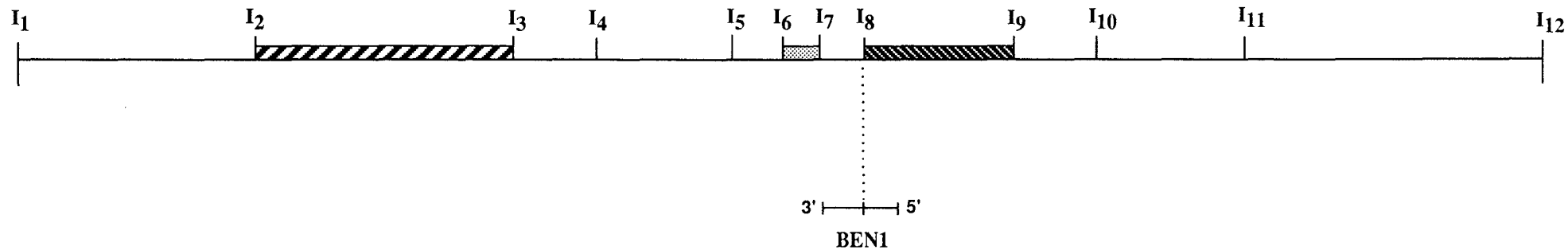
Cet îlot CpG "I_g" contenu dans le clone BEN1 permettrait de localiser le début du gène MUC5B et surtout la région contenant les séquences promotrices et régulatrices. La sonde JER57, hybride en aval de l'îlot CpG "I_g", définissant le gène MUC5B.

En amont de l'îlot "I_g", et en 5' du clone BEN1, nous avons pu identifier une nouvelle séquence organisée en "tandem repeats" (Figure 48). Cette séquence (BEN15') qui ne présente pas d'homologie avec des séquences connues possède un motif répétitif de 66 pb et montre un profil d'hybridation sur Southern blot de l'ADN humain, tout à fait spécifique.




Taille des bandes (kb)

Eco RI	Bam HI	Eco RI / Bam HI	Hind III	Xba I	Pst I
> 30	13	6	5	16	1.8
					1.3

La séquence n'est pas caractérisée par une richesse en acides aminés thréonine, sérine, proline spécifique des séquences de mucines isolées jusqu'à présent. Cependant, cette séquence caractérise peut être un nouveau gène de mucine. Pour nous en assurer, une étude de l'expression de ce nouveau fragment exonique est envisagée. L'existence d'un nouveau gène de mucine en amont de MUC5B renforcerait l'idée de la présence d'un cluster "MUC" en 11p15.5.



I : îlot CpG

-  Région inter-îlot contenant MUC2
-  Région inter-îlot contenant MUC5AC
-  Région inter-îlot contenant MUC5B


20 Kb

Figure 47 : Correspondance entre le clone BEN1 et la mégacartographie.



Figure 48 : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de l'extrémité 5' du clone BEN1 (BEN15').

T.R : "Tandem Repeats"

L'hybridation de ce fragment BEN15' sur les Southern blots de l'étude par électrophorèse en champ pulsé, a permis de vérifier sa localisation entre les îlots CpG "I7" et "I8".

La sonde JER57 hybride avec la partie 3' de BEN1 sur plus de 4,5 kb. Cette région représentée par un cadre "hachuré" sur la figure 46 a été entièrement séquencée.

2) *Séquence de BEN1A*

Cette séquence qui correspond à l'extrémité 3' de BEN1 est caractérisée par un motif répétitif de 87 pb (Figure 49). Elle comprend un fragment BamHI de 1.1 kb assimilable au fragment 1.2 kb observé en Southern blot ainsi que deux fragments de 1.6 kb (assimilable à la bande 1.7 kb). Elle présente une richesse en acides aminés hydroxylés (thréonine 23%, sérine 13.2%) caractéristiques des ADNc de la famille MUC5B, mais néanmoins, ne possède pas 100% d'homologie avec les clones ADNc déjà décrits dans cette famille. La séquence BEN1A est constituée d'un cadre ouvert de lecture d'un bout à l'autre (cadre n°1) (Figure 51). Aucun intron n'a été décelé. Cette séquence codante démontre l'existence d'un exon unique d'au moins 4.5 kb.

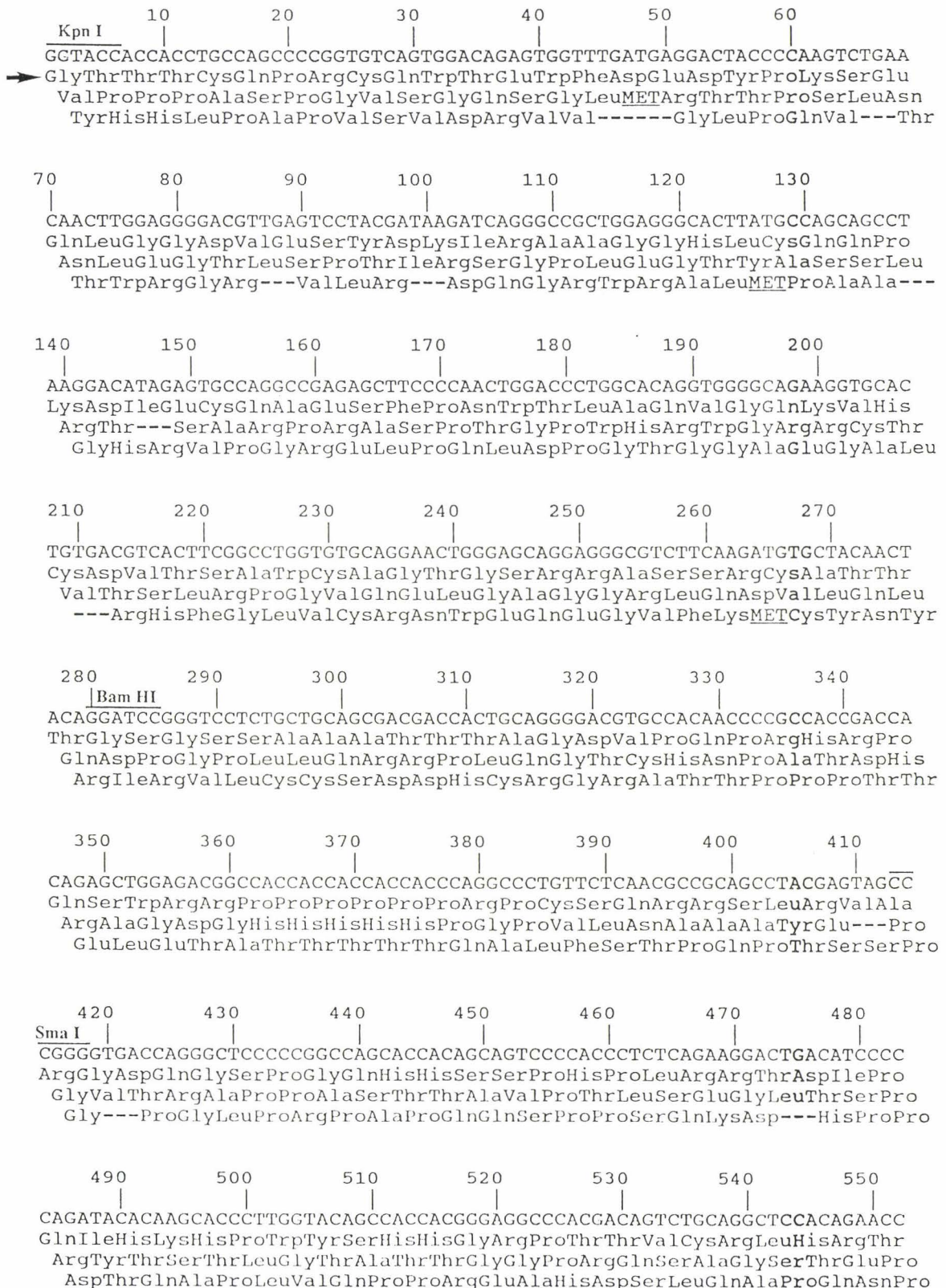
Si l'on considère, par comparaison avec les ADNc connus de la famille, le premier cadre comme étant le bon cadre de lecture, on peut remarquer tout au long de la séquence peptidique déduite, l'alternance de zones hydrophiles potentiellement glycosylables et de zones hydrophobes moins glycosylables dites nues (Figure 51)

La sonde JER 57 hybride avec le clone BEN1 par des fragments Kpn I de 2.1, 2.3 et 1.7 kb et des fragments Bam HI de 1.7 et 1.2 kb. Ce clone possède donc une partie des fragments visualisés sur Southern blot. En ce qui concerne l'hydrolyse Bam HI, il manque un fragment de 0.3 kb.

De plus les séquences obtenues ne montrent pas 100% d'homologie avec les séquences ADNc connues. Ce qui laisse supposer la présence d'autres régions organisées en tandem repeats de 87 pb.

L'étude par Southern blot de l'ADN humain après hydrolyse par l'enzyme Bgl II révèle avec JER57 un fragment de 16.5 kb qui permet de placer BEN1 en 5' de CEL5, les 2 clones étant séparés de 3.5 kb (Figure 52).

Figure 50 : Séquence nucléotidique et peptidique déduite dans les trois cadres de lecture de BEN1A.



560 570 580 590 600 610 620
 CACTGTCCAGGGGTGGCCACATCCACCCTTCCAACACGCTCAGCCCTTCCAGGGACGACGGGGAGCTT
 HisCysProArgGlyGlyHisIleHisProSerAsnThrLeuSerProSerArgAspAspGlyGluLeu
 ThrValProGlyValAlaThrSerThrLeuProThrArgSerAlaLeuProGlyThrThrGlySerLeu
 LeuSerGlnGlyTrpProHisProProPheGlnHisAlaGlnProPheGlnGlyArgArgGlyAlaTrp

630 640 650 660 670 680 690
 GGCACATGCCCCCTCACAGCCACCCACGCTGGCCCCAACACAATGGCAACCTCCAGAGCTCGCCC
 GlyHisMETAlaProLeuThrAlaThrHisAlaGlyProAsnAsnAsnGlyAsnLeuGlnSerSerPro
 GlyThrTrpProProSerGlnProProThrLeuAlaProThrThrMETAlaThrSerArgAlaArgPro
 AlaHisGlyProProHisSerHisProArgTrpProGlnGlnGlnTrpGlnProProGluLeuAlaArg

700 710 720 730 740 750
 GACAGGCACAGCCAGCACCGCTTCCAAAGAGCCGCTGACCACGAGCCTGGCGCCAACACTCACGAGCGA
 AspArgHisSerGlnHisArgPheGlnArgAlaAlaAspHisGluProGlyAlaAsnThrHisGluArg
 ThrGlyThrAlaSerThrAlaSerLysGluProLeuThrThrSerLeuAlaProThrLeuThrSerGlu
 GlnAlaGlnProAlaProLeuProLysSerArg---ProArgAlaTrpArgGlnHisSerArgAlaSer

760 770 780 790 800 810 820
 GCTGTCCACCTCTCAGGCCGAGACCAGCACGCCAGGACAGAGACGACAATGAGCCCCCTTGACTAACACC
 AlaValHisLeuSerGlyArgAspGlnHisAlaArgThrGluThrThrMETSerProLeuThrAsnThr
 LeuSerThrSerGlnAlaGluThrSerThrProGlyGlnArgArgGln---AlaPro---LeuThrPro
 CysProProLeuArgProArgProAlaArgGlnAspArgAspAsnGluProLeuAsp---HisHis

830 840 850 860 870 880 890
 ACCACCAGCCAGGGCAGCAGCCGCTGTCAACCGAAGTGTGAGTGGACAGAGTGGTTTGACGTGGACTTC
 ThrThrSerGlnGlyThrThrArgCysGlnProLysCysGluTrpThrGluTrpPheAspValAspPhe
 ProProAlaArgAlaArgProAlaValAsnArgSerValSerGlyGlnSerGlyLeuThrTrpThrSer
 HisGlnProGlyHisAspProLeuSerThrGluVal---ValAspArgValVal---ArgGlyLeuPro

900 910 920 930 940 950 960
 CCAACCTCAGGGGTTGCAAGCGGGGACATGGAACCTTTTGAACATCAGGGCTGCTGGGGGCAAGATG
 ProThrSerGlyValAlaSerGlyAspMETGluThrPheGluAsnIleArgAlaAlaGlyGlyLysMET
 GlnProGlnGlyLeuGlnAlaGlyThrTrpLysLeuLeuLysThrSerGlyLeuLeuGlyAlaArgCys
 AsnLeuArgGlyCysLysArgGlyHisGlyAsnPhe---LysHisGlnGlyCysTrpGlyGlnAspVal

970 980 990 1000 1010 1020 1030
 TGCTGGGCACCAAAGAGCATAGAGTGCCGGGGCGGAGAACTACCCCGAGGTAAGCATCGACCAGGTCCGGG
 CysTrpAlaProLysSerIleGluCysArgAlaGluAsnTyrProGluValSerIleAspGlnValGly
 AlaGlyHisGlnArgAla---SerAlaGlyArgArgThrThrProArg---AlaSerThrArgSerGly
 LeuGlyThrLysGluHisArgValProGlyGlyGluLeuProArgGlyLysHisArgProGlyArgAla

1040 1050 1060 1070 1080 1090 1100
 CAGGTGCTGACCTGCAGCCTGGAGACGGGGCTGACCTGCAAGAACGAAGACCAGACAGGCAGGTTCAAC
 GlnValLeuThrCysSerLeuGluThrGlyLeuThrCysLysAsnGluAspGlnThrGlyArgPheAsn
 ArgCys---ProAlaAlaTrpArgArgGly---ProAlaArgThrLysThrArgGlnAlaGlySerThr
 GlyAlaAspLeuGlnProGlyAspGlyAlaAspLeuGlnGluArgArgProAspArgGlnValGlnHis

1110 1120 1130 1140 1150 1160 1170
 ATGTGCTTCAACTACAACGTGCGTGTGCTTTGCTGTGACGACTACAGCCACTGCCCCAGTACCCTAGCC
 METCysPheAsnTyrAsnValArgValLeuCysCysAspAspTyrSerHisCysProSerThrLeuAla
 CysAlaSerThrThrThrCysValCysPheAlaValThrThrAlaThrAlaProValPro---Pro
 ValLeuGlnLeuGlnArgAlaCysAlaLeuLeu---ArgLeuGlnProLeuProGlnTyrProSerHis

1180 1190 1200 1210 1220 1230 1240
 | | | | | | |
 ACCAGCTCCACGGCCACGCCCTCCTCAACTCCGGGGACGACCTGGATCCTCACAAGCCGACCACAACA
 ThrSerSerThrAlaThrProSerSerThrProGlyThrThrTrpIleLeuThrLysProThrThrThr
 ProAlaProArgProArgProProGlnLeuArgGlyArgProGlySerSerGlnSerArgProGlnGln
 GlnLeuHisGlyHisAlaLeuLeuAsnSerGlyAspAspLeuAspProHisLysAlaAspHisAsnSer

 1250 1260 1270 1280 1290 1300 1310
 | | | | | | |
 GCCACTACGACTGCGTCCACTGGATCCACGGCCACCCGCTCCTCCACCCAGGCAACTGCTGGCACCCCA
 AlaThrThrThrAlaSerThrGlySerThrAlaThrAlaSerSerThrGlnAlaThrAlaGlyThrPro
 ProLeuArgLeuArgProLeuAspProArgProProProProProProArgGlnLeuLeuAlaProHis
 HisTyrAspCysValHisTrpIleHisGlyHisArgLeuLeuHisProGlyAsnCysTrpHisProThr

 1320 1330 1340 1350 1360 1370 1380
 | | | | | | |
 CATGTGAGCACCACGGCCACGACACCCACAGTCACCAGCTCCAAAGCCACTCCCTTCTCCAGTCCAGGG
 HisValSerThrThrAlaThrThrProThrValThrSerSerLysAlaThrProPheSerSerProGly
 MET---AlaProArgProArgHisProGlnSerProAlaProLysProLeuProSerProValGlnGly
 CysGluHisHisGlyHisAspThrHisSerHisGlnLeuGlnSerHisSerLeuLeuGlnSerArgAsp

 1390 1400 1410 1420 1430 1440
 | | | | | |
 ACTGCAACCGCCCTTCCAGCAGTGAAGCACAGCCACCACCCACAGCTACCAGCTTTACAGCCATC
 ThrAlaThrAlaLeuProAlaLeuArgSerThrAlaThrThrProThrAlaThrSerPheThrAlaIle
 LeuGlnProProPheGlnHis---GluAlaGlnProProHisProGlnLeuProAlaLeuGlnProSer
 CysAsnArgProSerSerThrGluLysHisSerHisHisThrHisSerTyrGlnLeuTyrSerHisPro

 1450 1460 1470 1480 1490 1500 1510
 | | | | | | |
 CCTCCTCCTCCCTGGGCACCACCTGGACCCGCCTATCACAGACCACACCCATGGCCACCATGTCC
 ProSerSerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThrThrProMETAlaThrMETSer
 ProProProProTrpAlaProProGlyProAlaTyrHisArgProProHisProTrpProProCysPro
 LeuLeuLeuProGlyHisHisLeuAspProProIleThrAspHisHisThrHisGlyHisHisValHis

 1520 1530 1540 1550 1560 1570 1580
 | | | | | | |
 ACAGCCACCCCTCCTCCACTCCAGAGACTGTCCACACCTCCACAGTGCTTACCACCACGGCCACCAC
 ThrAlaThrProLeuLeuHisSerArgAspCysProHisLeuHisSerAlaTyrHisHisGlyHisHis
 GlnProHisProSerSerThrProGluThrValHisThrSerThrValLeuThrThrThrAlaThrThr
 SerHisThrProProProLeuGlnArgLeuSerThrProProGlnCysLeuProProArgProProGln

 1590 1600 1610 1620 1630 1640 1650
 | | | | | | |
 AACGGGGCCACGGGCTCTGTGGCCACCCCTCCTCCACCCAGGAACAGCTCACACTACCAAAGTGCT
 AsnArgGlyHisArgLeuCysGlyHisProLeuLeuHisProArgAsnSerSerHisTyrGlnSerAla
 ThrGlyAlaThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThrThrLysValLeu
 ProGlyProProAlaLeuTrpProProProProProProGlnGluGlnLeuThrLeuProLysCys---

 1660 1670 1680 1690 1700 1710 1720
 | | | | | | |
 GACTACCACAACCACGGCTTACAGCCACCCCTCCTCCAGCCCAGGGAGGGCAGCAGCCTTCCAGTG
 AspTyrHisAsnHisGlyPheThrAlaThrProSerSerSerProGlyArgAlaArgThrLeuProVal
 ThrThrThrThrThrAlaSerGlnProProProProProAlaGlnGlyGlyHisAlaArgPheGlnCys
 LeuProGlnProArgLeuHisSerHisProLeuLeuGlnProArgGluGlyThrHisAlaSerSerVal

 1730 1740 1750 1760 1770 1780 1790
 | | | | | | |
 TGGATCAGCACAAACCACCACACCCACAACCAGAGGTTCCACGGTGACCCCTCCTCCATCCCGGGGACC
 TrpIleSerThrThrThrThrProThrThrArgGlySerThrValThrProSerSerIleProGlyThr
 GlySerAlaGlnProProHisProGlnProGluValProArg---ProProProProSerArgGlyPro
 AspGlnHisAsnHisHisThrHisAsnGlnArgPheHisGlyAspProLeuLeuHisProGlyAspHis

1800 1810 1820 1830 1840 1850 1860
 ACCCACACCCCCACAGTGGCTGACCACCACCACCACAACACTGTGGCCACTGGTTCTATGGCAACACCCCTCC
 ThrHisThrProThrValLeuThrThrThrThrThrThrValAlaThrGlySerMETAlaThrProSer
 ProThrProProGlnCys---ProProProProGlnLeuTrpProLeuValLeuTrpGlnHisProPro
 ProHisProHisSerAlaAspHisHisHisHisAsnCysGlyHisTrpPheTyrGlyAsnThrLeuLeu

1870 1880 1890 1900 1910 1920 1930
 TCTAGCACACAGACCAGTGGTACTCCCCATCACTGACCACCACGGCCACTACGATCACGGCCACCGGC
 SerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThrThrIleThrAlaThrGly
 LeuAlaHisArgProValValLeuProHisHis---ProProArgProLeuArgSerArgProProAla
 ---HisThrAspGlnTrpTyrSerProIleThrAspHisHisGlyHisTyrAspHisGlyHisArgLeu

1940 1950 1960 1970 1980 1990 2000
 TCCACCACCAACCCCTCCTCAACTCCAGGGACAACACCTATCCCCCAGTGGCTGACCACCACCGCCACC
 SerThrThrAsnProSerSerThrProGlyThrThrProIleProProValLeuThrThrThrAlaThr
 ProProProThrProProGlnLeuGlnGlyGlnHisLeuSerProGlnCys---ProProProProPro
 HisHisGlnProLeuLeuAsnSerArgAspAsnThrTyrProProSerAlaAspHisHisArgHisHis

2010 2020 2030 2040 2050 2060 2070
 ACACCTGCAGCCACCAGCAGCACAGTGACTCCCTCCTCTGCCCTAGGGACCACCCACACACCCCCAGTG
 ThrProAlaAlaThrSerSerThrValThrProSerSerAlaLeuGlyThrThrHisThrProProVal
 HisLeuGlnProProAlaAlaGln---LeuProProLeuPro---GlyProProThrHisProGlnCys
 ThrCysSerHisGlnGlnHisSerAspSerLeuLeuCysProArgAspHisProHisThrProSerAla

2080 2090 2100 2110 2120 2130
 CCGAACACCACGGCCACCACACACGGGCGATCCCTGTCCCCAGCAGTCCCCACACGGTGTGCACAGCC
 ProAsnThrThrAlaThrThrHisGlyArgSerLeuSerProSerSerProHisThrValCysThrAla
 ArgThrProArgProProHisThrGlyAspProCysProProAlaValProThrArgCysAlaGlnPro
 GluHisHisGlyHisHisThrArgAlaIleProValProGlnGlnSerProHisGlyValHisSerLeu

2140 2150 2160 2170 2180 2190 2200
 TGGACTTCGGCCACCTCAGGCACCTTGGGCACCACCCACATCACAGAGCCTTCCACGGGGACTTCCAC
 TrpThrSerAlaThrSerGlyThrLeuGlyThrThrHisIleThrGluProSerThrGlyThrSerHis
 GlyLeuArgProProGlnAlaProTrpAlaProProThrSerGlnSerLeuProArgGlyLeuProThr
 AspPheGlyHisLeuArgHisLeuGlyHisHisProHisHisArgAlaPheHisGlyAspPheProHis

2210 2220 2230 2240 2250 2260 2270
 ACCCCAGCAGCAACCACCGGTACCACCCAGCACTCGACTCCAGCCCTTCCAGCCCTCACCCCTAGCAGC
 ThrProAlaAlaThrThrGlyThrThrGlnHisSerThrProAlaLeuSerSerProHisProSerSer
 ProGlnGlnGlnProProValProProSerThrArgLeuGlnProPheProAlaLeuThrLeuAlaAla
 ProSerSerAsnHisArgTyrHisProAlaLeuAspSerSerProPheGlnProSerPro---GlnGln

2280 2290 2300 2310 2320 2330 2340
 AGAACCCAGAGTACCCCCCTTCTCCAGGGACGACCACCCCGGGCCACACCACGGCCACCTCCAGGACC
 ArgThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrThrAlaThrSerArgThr
 GluProProSerHisProLeuLeuGlnGlyArgProProArgAlaThrProArgProProProGlyPro
 AsnHisArgValThrProPheSerArgAspAspHisProGlyProHisHisGlyHisLeuGlnAspHis

2350 2360 2370 2380 2390 2400 2410
 ACAGCCACGGCCACACCAGCAAGACCCGCACCTCGACCTGCTGCCAGCAGCCCCACATCGGCCCCC
 ThrAlaThrAlaThrProSerLysThrArgThrSerThrLeuLeuProSerSerProThrSerAlaPro
 GlnProArgProHisProAlaArgProAlaProArgProCysCysProAlaAlaProHisArgProPro
 SerHisGlyHisThrGlnGlnAspProHisLeuAspProAlaAlaGlnGlnProHisIleGlyProHis

2420 2430 2440 2450 2460 2470 2480
 ATAACCACGGTGGTGACCATGGGGCTGTGAGCCCAGTGTGCCTGGTCAGAGTGGCTGGACTACAGCTACC
 IleThrThrValValThrMETGlyCysGluProSerValProGlyGlnSerGlyTrpThrThrAlaThr
 ---ProArgTrp---ProTrpAlaValSerProValCysLeuValArgValAlaGlyLeuGlnLeuPro
 AsnHisGlyGlyAspHisGlyLeu---AlaGlnCysAlaTrpSerGluTrpLeuAspTyrSerTyrPro

2490 2500 2510 2520 2530 2540 2550
 CCATGCCGGGGCCCTCTGGCGGGGACTTTGACACCTACTCCAACATCCGTGCGGGCCGGAGGGGCGCTCT
 ProCysArgGlyProLeuAlaGlyThrLeuThrProThrProThrSerValArgProGluGlyProSer
 HisAlaGlyAlaLeuTrpArgGlyLeu---HisLeuLeuGlnHisProCysGlyArgArgGlyArgLeu
 METProGlyProSerGlyGlyAspPheAspThrTyrSerAsnIleArgAlaAlaGlyGlyAlaValCys

2560 2570 2580 2590 2600 2610 2620
 GTGAGCAGCCCCCTGGGCCTCGAGTGCCTGCCAGGCCAGCCTGGTGTCCCCCTGCGGGAGTTGGGC
 ValSerSerProTrpAlaSerSerAlaValProArgProSerLeuValSerProLeuArgGluLeuGly
 ---AlaAlaProGlyProArgValProCysProGlyProAlaTrpCysProProCysGlySerTrpAla
 GluGlnProLeuGlyLeuGluCysArgAlaGlnAlaGlnProGlyValProProAlaGlyValGlyPro

2630 2640 2650 2660 2670 2680 2690
 CAGGTCGTGGAATGCAGCCTGGACTTTGGCCTGGTCTGCAGGAACCGTGAGCAGGTGGGGAAGTTCAAG
 GlnValValGluCysSerLeuAspPheGlyLeuValCysArgAsnArgGluGlnValGlyLysPheLys
 ArgSerTrpAsnAlaAlaTrpThrLeuAlaTrpSerAlaGlyThrValSerArgTrpGlySerSerArg
 GlyArgGlyMETGlnProGlyLeuTrpProGlyLeuGlnGluPro---AlaGlyGlyGluValGlnAsp

2700 2710 2720 2730 2740 2750 2760
 ATGTGCTTCAACTATGAAATCCCGTGTGTTCTGCTGCAACTACGGCCACTGCCCCAGCACCCCGGCCACC
 METCysPheAsnTyrGluIleArgValPheCysCysAsnTyrGlyHisCysProSerThrProAlaThr
 CysAlaSerThrMETLysSerValCysSerAlaAlaThrThrAlaThrAlaProAlaProArgProPro
 ValLeuGlnLeu---AsnProCysValLeuLeuGlnLeuArgProLeuProGlnHisProGlyHisGln

2770 2780 2790 2800 2810 2820
 AGCTCTACGGCCACGCCCTCCTCCACTCCAGGGACGACCTGGATCCTCACAGAGCTGACCACAACAGCC
 SerSerThrAlaThrProSerSerThrProGlyThrThrTrpIleLeuThrGluLeuThrThrThrAla
 AlaLeuArgProArgProProProLeuGlnGlyArgProGlySerSerGlnSer---ProGlnGlnPro
 LeuTyrGlyHisAlaLeuLeuHisSerArgAspAspLeuAspProHisArgAlaAspHisAsnSerHis

2830 2840 2850 2860 2870 2880 2890
 ACTACGACTGAGTCCACTGGATCCACGGCCACCCCGACCTCCACCCTGAGAACAGCTCCCCCTCCCAA
 ThrThrThrGluSerThrGlySerThrAlaThrProThrSerThrLeuArgThrAlaProProProLys
 LeuArgLeuSerProLeuAspProArgProProArgProProPro---GluGlnLeuProLeuProLys
 TyrAsp---ValHisTrpIleHisGlyHisProAspLeuHisProGluAsnSerSerProSerGlnSer

2900 2910 2920 2930 2940 2950 2960
 GTGCTGACCACCACGGCCACCACCCACAGTCACCAGCTCCAAAGCCACTCCCTCCTCCAGTCCAGGG
 ValLeuThrThrThrAlaThrThrProThrValThrSerSerLysAlaThrProSerSerSerProGly
 Cys---ProProArgProProHisProGlnSerProAlaProLysProLeuProProProValGlnGly
 AlaAspHisHisGlyHisHisThrHisSerHisGlnLeuGlnSerHisSerLeuLeuGlnSerArgAsp

2970 2980 2990 3000 3010 3020 3030
 ACTGCAACCGCCCTTCCAGCACTGAGAAGCACAGCCACCACCCACAGCTACCAGCGTTACCCCATC
 ThrAlaThrAlaLeuProAlaLeuArgSerThrAlaThrProThrAlaThrSerValThrProIle
 LeuGlnProProPheGlnHis---GluAlaGlnProProHisProGlnLeuProAlaLeuHisProSer
 CysAsnArgProSerSerThrGluLysHisSerHisHisThrHisSerTyrGlnArgTyrThrHisPro

3040 3050 3060 3070 3080 3090 3100
 CCCTCTTCCCTCCCTGGGCACCACCTGGACCCGCCCTATCACAGACCACCACACCCACGGCCACCATGTCC
 ProSerSerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMETSer
 ProLeuProProTrpAlaProProGlyProAlaTyrHisArgProProHisProArgProProCysPro
 LeuPheLeuProGlyHisHisLeuAspProProIleThrAspHisHisThrHisGlyHisHisValHis

3110 3120 3130 3140 3150 3160 3170
 ACAGCCACACCCTCCTCCACTCCAGAGACTGCCACACCTCCACAGTCTTACCGCACGGCCACCACAAC
 ThrAlaThrProSerSerThrProGluThrAlaHisThrSerThrValLeuProHisGlyHisHisAsn
 GlnProHisProProLeuGlnArgLeuProProGlnSerTyrArgThrAlaThrThrThr
 SerHisThrLeuLeuHisSerArgAspCysProHisLeuHisSerLeuThrAlaArgProProGlnLeu

3180 3190 3200 3210 3220 3230 3240
 TGGGGCACCGGCTCTGTGGCCACCCCTCCTCCACCCAGGAACAGCTCACACTACCAAAGTGCCAACT
 TrpGlyThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThrThrLysValProThr
 GlyAlaProAlaLeuTrpProProProProProGlnGluGlnLeuThrLeuProLysCysGlnLeu
 GlyHisArgLeuCysGlyHisProLeuLeuHisProArgAsnSerSerHisTyrGlnSerAlaAsnTyr

3250 3260 3270 3280 3290 3300 3310
 ACCACAACCACGGCTTCACAGCCACCCCTCCTCCAGCCCAGGGACGGCACTCACGCCTCCAGTGTGG
 ThrThrThrThrAlaSerGlnProProProSerSerSerProGlyThrAlaLeuThrProProValTrp
 ProGlnProArgLeuHisSerHisProProProProAlaGlnGlyArgHisSerArgLeuGlnCysGly
 HisAsnHisGlyPheThrAlaThrProLeuLeuGlnProArgAspGlyThrHisAlaSerSerValAsp

3320 3330 3340 3350 3360 3370 3380
 ATCAGCACAAACCACCACACCCACAACCAGAGGCTCCACGGTGACCCCTCCTCCATCCCGGGGACCACC
 IleSerThrThrThrThrProThrThrArgGlySerThrValThrProSerSerIleProGlyThrThr
 SerAlaGlnProProHisProGlnProGluAlaProArg---ProProProProSerArgGlyProPro
 GlnHisAsnHisHisThrHisAsnGlnArgLeuHisGlyAspProLeuLeuHisProGlyAspHisPro

3390 3400 3410 3420 3430 3440 3450
 CACACCGCCACAGTGCTGACCACCACCACCAACTGTGGCCACTGGTTCTATGGCAACACCCTCCTCT
 HisThrAlaThrValLeuThrThrThrThrThrValAlaThrGlySerMETAlaThrProSerSer
 ThrProProGlnCys---ProProProProGlnLeuTrpProLeuValLeuTrpGlnHisProProLeu
 HisArgHisSerAlaAspHisHisHisHisAsnCysGlyHisTrpPheTyrGlyAsnThrLeuLeu---

3460 3470 3480 3490 3500 3510
 AGCACACAGACCAGTGGTACTCCCCATCACTGACCACCACGGCCACTACGATCACGGCCACCGGCTCC
 SerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThrThrIleThrAlaThrGlySer
 AlaHisArgProValValLeuProHisHis---ProProArgProLeuArgSerArgProProAlaPro
 HisThrAspGlnTrpTyrSerProIleThrAspHisHisGlyHisTyrAspHisGlyHisArgLeuHis

3520 3530 3540 3550 3560 3570 3580
 ACCACCAACCCTCCTCAACTCCTGGGACAACCTCCCATCCCCCAGTGCTGACCACCACCGCCACCACA
 ThrThrAsnProSerSerThrProGlyThrThrProIleProProValLeuThrThrAlaThrThr
 ProProThrProProGlnLeuLeuGlyGlnLeuProSerProGlnCys---ProProProProHis
 HisGlnProLeuLeuAsnSerTrpAspAsnSerHisProProSerAlaAspHisHisArgHisHisThr

3590 3600 3610 3620 3630 3640 3650
 CCTGCAGCCACCAGCAACACAGTGACTCCCTCCTCTGCCCTAGGGACCACCCACACACCCCCAGTGCCG
 ProAlaAlaThrSerAsnThrValThrProSerSerAlaLeuGlyThrThrHisThrProProValPro
 LeuGlnProProAlaThrGln---LeuProProLeuPro---GlyProProThrHisProGlnCysArg
 CysSerHisGlnGlnHisSerAspSerLeuLeuCysProArgAspHisProHisThrProSerAlaGlu

3660 3670 3680 3690 3700 3710 3720
 AACACCACGGCCACCACACACGGGCGATCCCTGCCCCACAGCAGTCCCCACACGGTGCCACAGCCTGG
 AsnThrThrAlaThrThrHisGlyArgSerLeuProProSerSerProHisThrValProThrAlaTrp
 ThrProArgProProHisThrGlyAspProCysProProAlaValProThrArgCysProGlnProGly
 HisHisGlyHisHisThrArgAlaIleProAlaProGlnGlnSerProHisGlyAlaHisSerLeuAsp

3730 3740 3750 3760 3770 3780 3790
 ACTTCGGCCACCTCGGGCATCTTGGGCACCACCCACATCACAGAGCCTTCCACGGGGACTTCCACACC
 ThrSerAlaThrSerGlyIleLeuGlyThrThrHisIleThrGluProSerThrGlyThrSerHisThr
 LeuArgProProArgAlaSerTrpAlaProProThrSerGlnSerLeuProArgGlyLeuProThrPro
 PheGlyHisLeuGlyHisLeuGlyHisHisProHisHisArgAlaPheHisGlyAspPheProHisPro

3800 3810 3820 3830 3840 3850 3860
 CCAGCAGCAACCACCGGTACCACCCAGCACTCGACTCCAGCCCTTCCAGCCCTCACCCCTAGCAGCAGA
 ProAlaAlaThrThrGlyThrThrGlnHisSerThrProAlaLeuSerSerProHisProSerSerArg
 GlnGlnGlnProProValProProSerThrArgLeuGlnProPheProAlaLeuThrLeuAlaAlaGlu
 SerSerAsnHisArgTyrHisProAlaLeuAspSerSerProPheGlnProSerPro---GlnGlnAsn

3870 3880 3890 3900 3910 3920 3930
 ACCACCGAGTCACCCCTTCTCCAGGGACGACCACCCGGGCCACACCACGGCCACCTCCAGGACCACA
 ThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrThrAlaThrSerArgThrThr
 ProProSerHisProLeuLeuGlnGlyArgProProArgAlaThrProArgProProProGlyProGln
 HisArgValThrProPheSerArgAspAspHisProGlyProHisHisGlyHisLeuGlnAspHisSer

3940 3950 3960 3970 3980 3990 4000
 GCCACGGCCACACCAGCAAGACCCGACCTCGACCCTGCTGCCAGCAGCCCCACATCGGCCCCATA
 AlaThrAlaThrProSerLysThrArgThrSerThrLeuLeuProSerSerProThrSerAlaProIle
 ProArgProHisProAlaArgProAlaProArgProCysCysProAlaAlaProHisArgProPro---
 HisGlyHisThrGlnGlnAspProHisLeuAspProAlaAlaGlnGlnProHisIleGlyProHisAsn

4010 4020 4030 4040 4050 4060 4070
 ACCACGGTGGTGACCATGGGCTGTGAGCCAGTGTGCCTGGTCAGAGTGGCTGGACTACAGCTACCCCA
 ThrThrValValThrMETGlyCysGluProSerValProGlyGlnSerGlyTrpThrThrAlaThrPro
 ProArgTrp---ProTrpAlaValSerProValCysLeuValArgValAlaGlyLeuGlnLeuProHis
 HisGlyGlyAspHisGlyLeu---AlaGlnCysAlaTrpSerGluTrpLeuAspTyrSerTyrProMET

4080 4090 4100 4110 4120 4130 4140
 TGCCGGGGCCCTCTGGCGGGGACTTTGACACCTACTCCAACATCCGTGCGGGCGGAGGGGCGCTCTGTG
 CysArgGlyProLeuAlaGlyThrLeuThrProThrProThrSerValArgProGluGlyProSerVal
 AlaGlyAlaLeuTrpArgGlyLeu---HisLeuLeuGlnHisProCysGlyArgArgGlyArgLeu---
 ProGlyProSerGlyGlyAspPheAspThrTyrSerAsnIleArgAlaAlaGlyGlyAlaValCysGlu

4150 4160 4170 4180 4190 4200
 AGCAGCCCTGGGCCTCGAGTGCCGTGCCAGCCCAGCCTGGTGTCCCTGCGGGAGTTGGGCCAGGTCGTG
 SerSerProGlyProArgValProCysGlnProSerLeuValSerLeuArgGluLeuGlyGlnValVal
 AlaAlaLeuGlyLeuGluCysArgAlaSerProAlaTrpCysProCysGlySerTrpAlaArgSerTrp
 GlnProTrpAlaSerSerAlaValProAlaGlnProGlyValProAlaGlyValGlyProGlyArgGly

4210 4220 4230 4240 4250 4260 4270
 GAATGCAGCCTGACTTTGGCCTGGTCTGCAGGAACCGTGAGCAGGTGGGGAAGTTCAAGATGTGCTTCA
 GluCysSerLeuThrLeuAlaTrpSerAlaGlyThrValSerArgTrpGlySerSerArgCysAlaSer
 AsnAlaAla---LeuTrpProGlyLeuGlnGluPro---AlaGlyGlyGluValGlnAspValLeuGln
 METGlnProAspPheGlyLeuValCysArgAsnArgGluGlnValGlyLysPheLysMETCysPheAsn

4280 4290 4300 4310 4320 4330 4340
| | | | | |
ACTATGAAATCCGTGTGTTCTGCTGCAACTACGGCCACTGCCCCAGCACCCCGGCCACCAGCTCTACGG
ThrMETLysSerValCysSerAlaAlaThrThrAlaThrAlaProAlaProArgProProAlaLeuArg
Leu---AsnProCysValLeuLeuGlnLeuArgProLeuProGlnHisProGlyHisGlnLeuTyrGly
TyrGluIleArgValPheCysCysAsnTyrGlyHisCysProSerThrProAlaThrSerSerThrAla

4350 4360 4370 4380
| | | |
CCATGCCCTCCTCCACTCCGGGGACGACCTGGATCC
ProCysProProProLeuArgGlyArgProGlySer
HisAlaLeuLeuHisSerGlyAspAspLeuAsp
METProSerSerThrProGlyThrThrTrpIle

	5	10	15
1 Gly Thr Thr Thr Cys Gln Pro Arg Cys Gln Trp Thr Glu Trp Phe			
16 Asp Glu Asp Tyr Pro Lys Ser Glu Gln Leu Gly Gly Asp Val Glu			
31 Ser Tyr Asp Lys Ile Arg Ala Ala Gly Gly His Leu Cys Gln Gln			
46 Pro Lys Asp Ile Glu Cys Gln Ala Glu Ser Phe Pro Asn Trp Thr			
61 Leu Ala Gln Val Gly Gln Lys Val His Cys Asp Val Thr Ser Ala			
76 Trp Cys Ala Gly Thr Gly Ser Arg Arg Ala Ser Ser Arg Cys Ala			
91 Thr Thr Thr Gly Ser Gly Ser Ser Ala Ala Thr Thr Thr Ala			
106 Gly Asp Val Pro Gln Pro Arg His Arg Pro Gln Ser Trp Arg Arg			
121 Pro Pro Pro Pro Pro Pro Arg Pro Cys Ser Gln Arg Arg Ser Leu			
136 Arg Val Ala Arg Gly Asp Gln Gly Ser Pro Gly Gln His His Ser			
151 Ser Pro His Pro Leu Arg Arg Thr Asp Ile Pro Gln Ile His Lys			
166 His Pro Trp Tyr Ser His His Gly Arg Pro Thr Thr Val Cys Arg			
181 Leu His Arg Thr His Cys Pro Arg Gly Gly His Ile His Pro Ser			
196 Asn Thr Leu Ser Pro Ser Arg Asp Asp Gly Glu Leu Gly His Met			
211 Ala Pro Leu Thr Ala Thr His Ala Gly Pro Asn Asn Asn Gly Asn			
226 Leu Gln Ser Ser Pro Asp Arg His Ser Gln His Arg Phe Gln Arg			
241 Ala Ala Asp His Glu Pro Gly Ala Asn Thr His Glu Arg Ala Val			
256 His Leu Ser Gly Arg Asp Gln His Ala Arg Thr Glu Thr Thr Met			
271 Ser Pro Leu Thr Asn Thr Thr Thr Ser Gln Gly Thr Thr Arg Cys			
286 Gln Pro Lys Cys Glu Trp Thr Glu Trp Phe Asp Val Asp Phe Pro			
301 Thr Ser Gly Val Ala Ser Gly Asp Met Glu Thr Phe Glu Asn Ile			
316 Arg Ala Ala Gly Gly Lys Met Cys Trp Ala Pro Lys Ser Ile Glu			
331 Cys Arg Ala Glu Asn Tyr Pro Glu Val Ser Ile Asp Gln Val Gly			
346 Gln Val Leu Thr Cys Ser Leu Glu Thr Gly Leu Thr Cys Lys Asn			
361 Glu Asp Gln Thr Gly Arg Phe Asn Met Cys Phe Asn Tyr Asn Val			
376 Arg Val Leu Cys Cys Asp Asp Tyr Ser His Cys Pro Ser Thr Leu			
391 Ala Thr Ser Ser Thr Ala Thr Pro Ser Ser Thr Pro Gly Thr Thr			
406 Trp Ile Leu Thr Lys Pro Thr Thr Thr Ala Thr Thr Thr Ala Ser			
421 Thr Gly Ser Thr Ala Thr Ala Ser Ser Thr Gln Ala Thr Ala Gly			
436 Thr Pro His Val Ser Thr Thr Ala Thr Thr Pro Thr Val Thr Ser			
451 Ser Lys Ala Thr Pro Phe Ser Ser Pro Gly Thr Ala Thr Ala Leu			
466 Pro Ala Leu Arg Ser Thr Ala Thr Thr Pro Thr Ala Thr Ser Phe			
481 Thr Ala Ile Pro Ser Ser Ser Leu Gly Thr Thr Trp Thr Arg Leu			
496 Ser Gln Thr Thr Thr Pro Met Ala Thr Met Ser Thr Ala Thr Pro			
511 Leu Leu His Ser Arg Asp Cys Pro His Leu His Ser Ala Tyr His			
526 His Gly His His Asn Arg Gly His Arg Leu Cys Gly His Pro Leu			
541 Leu His Pro Arg Asn Ser Ser His Tyr Gln Ser Ala Asp Tyr His			
556 Asn His Gly Phe Thr Ala Thr Pro Ser Ser Ser Pro Gly Arg Ala			
571 Arg Thr Leu Pro Val Trp Ile Ser Thr Thr Thr Thr Thr Thr Arg			
586 Arg Gly Ser Thr Val Thr Pro Ser Ser Ile Pro Gly Thr Thr His			
601 Thr Pro Thr Val Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr			
616 Ser Met Ala Thr Pro Ser Ser Ser Thr Gln Thr Ser Gly Thr Pro			
631 Pro Ser Leu Thr Thr Thr Ala Thr Thr Ile Thr Ala Thr Gly Ser			
646 Thr Thr Asn Pro Ser Ser Thr Pro Gly Thr Thr Pro Ile Pro Pro			
661 Val Leu Thr Thr Thr Ala Thr Thr Pro Ala Ala Thr Ser Ser Thr			
676 Val Thr Pro Ser Ser Ala Leu Gly Thr Thr His Thr Pro Val			
691 Pro Asn Thr Thr Ala Thr Thr His Gly Arg Ser Leu Ser Pro Ser			
706 Ser Pro His Thr Val Cys Thr Ala Trp Thr Ser Ala Thr Ser Gly			
721 Thr Leu Gly Thr Thr His Ile Thr Glu Pro Ser Thr Gly Thr Ser			
736 His Thr Pro Ala Ala Thr Thr Gly Thr Thr Gln His Ser Thr Pro			
751 Ala Leu Ser Ser Pro His Pro Ser Ser Arg Thr Thr Glu Ser Pro			
766 Pro Ser Pro Gly Thr Thr Thr Pro Gly His Thr Thr Ala Thr Ser			
781 Arg Thr Thr Ala Thr Ala Thr Pro Ser Lys Thr Arg Thr Ser Thr			
796 Leu Leu Pro Ser Ser Pro Thr Ser Ala Pro Ile Thr Thr Val Val			
811 Thr Met Gly Cys Glu Pro Ser Val Pro Gly Gln Ser Gly Trp Thr			
826 Thr Ala Thr Pro Cys Arg Gly Pro Leu Ala Gly Thr Leu Thr Pro			
841 Thr Pro Thr Ser Val Arg Pro Glu Gly Pro Ser Val Ser Ser Pro			
856 Trp Ala Ser Ser Ala Val Pro Arg Pro Ser Leu Val Ser Pro Leu			
871 Arg Glu Leu Gly Gln Val Val Glu Cys Ser Leu Ser Phe Gly Leu			
886 Val Cys Arg Asn Arg Glu Gln Val Gly Lys Phe Lys Met Cys Phe			
901 Asn Tyr Glu Ile Arg Val Phe Cys Cys Asn Tyr Gly His Cys Pro			
916 Ser Thr Pro Ala Thr Ser Ser Thr Ala Thr Pro Ser Ser Thr Pro			
931 Gly Thr Thr Trp Ile Leu Thr Glu Leu Thr Thr Thr Ala Thr Thr			
946 Thr Glu Ser Thr Gly Ser Thr Ala Thr Pro Thr Ser Thr Leu Arg			
961 Thr Ala Pro Pro Pro Lys Val Leu Thr Thr Thr Thr Thr Thr Pro			
976 Thr Val Thr Ser Ser Lys Ala Thr Pro Ser Ser Ser Pro Gly Thr			
991 Ala Thr Ala Leu Pro Ala Leu Arg Ser Thr Ala Thr Thr Pro Thr			
1006 Ala Thr Ser Val Thr Pro Ile Pro Ser Ser Ser Leu Gly Thr Thr			
1021 Trp Thr Arg Leu Ser Gln Thr Thr Thr Pro Thr Ala Thr Met Ser			
1036 Thr Ala Thr Pro Ser Ser Thr Pro Glu Thr Ala His Thr Ser Thr			
1051 Val Leu Pro His Gly His His Asn Trp Gly Thr Gly Ser Val Ala			
1066 Thr Pro Ser Ser Thr Pro Gly Thr Ala His Thr Thr Lys Val Pro			
1081 Thr Thr Thr Thr Thr Ala Ser Gln Pro Pro Ser Ser Ser Ser Pro			
1096 Gly Thr Ala Leu Thr Pro Pro Val Trp Ile Ser Thr Thr Thr Thr			
1111 Pro Thr Thr Arg Gly Ser Thr Val Thr Pro Ser Ser Ile Pro Gly			
1126 Thr Thr His Thr Ala Thr Val Leu Thr Thr Thr Thr Thr Thr Val			
1141 Ala Thr Gly Ser Met Ala Thr Pro Ser Ser Ser Thr Gln Thr Ser			
1156 Gly Thr Pro Pro Ser Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr			
1171 Thr Gly Ser Thr Thr Asn Pro Ser Ser Thr Pro Gly Thr Thr Pro			
1186 Ile Pro Pro Val Leu Thr Thr Thr Ala Thr Thr Pro Ala Ala Thr			
1201 Ser Asn Thr Val Thr Pro Ser Ser Ala Leu Gly Thr Thr His Thr			
1216 Pro Pro Val Pro Asn Thr Thr Ala Thr Thr His Gly Arg Ser Leu			
1231 Pro Pro Ser Ser Pro His Thr Val Pro Thr Ala Trp Thr Ser Ala			
1246 Thr Ser Gly Ile Leu Gly Thr Thr His Ile Thr Glu Pro Ser Thr			
1261 Gly Thr Ser His Thr Pro Ala Ala Thr Thr Gly Thr Thr Gln His			
1276 Ser Thr Pro Ala Leu Ser Ser Pro His Pro Ser Ser Arg Thr Thr			
1291 Glu Ser Pro Pro Ser Pro Gly Thr Thr Thr Pro Gly His Thr Thr			
1306 Ala Thr Ser Arg Thr Thr Ala Thr Ala Thr Pro Ser Lys Thr Arg			
1321 Thr Ser Thr Leu Leu Pro Ser Ser Pro Thr Ser Ala Pro Ile Thr			
1336 Thr Val Val Thr Met Gly Cys Glu Pro Ser Val Pro Gly Gln Ser			
1351 Gly Trp Thr Thr Ala Thr Pro Cys Arg Gly Pro Leu Ala Gly Thr			
1366 Leu Thr Pro Thr Thr Ser Val Arg Pro Glu Gly Pro Ser Val			
1381 Ser Ser Pro Gly Pro Arg Val Pro Cys Gln Pro Ser Leu Val Ser			
1396 Leu Arg Glu Leu Gly Gln Val Val Glu Cys Ser Leu Thr Leu Ala			
1411 Trp Ser Ala Gly Thr Val Ser Arg Trp Gly Ser Ser Arg Cys Ala			
1426 Ser Thr Met Lys Ser Val Cys Ser Ala Ala Thr Thr Ala Thr Ala			
1441 Pro Ala Pro Arg Pro Pro Ala Leu Arg Pro Cys Pro Pro Pro Leu			
1456 Arg Gly Arg Pro Gly Ser			

hydrophobe

hydrophile

hydrophobe

hydrophile

hydrophobe

Courbe d'hydrophilie de la séquence BEN1A

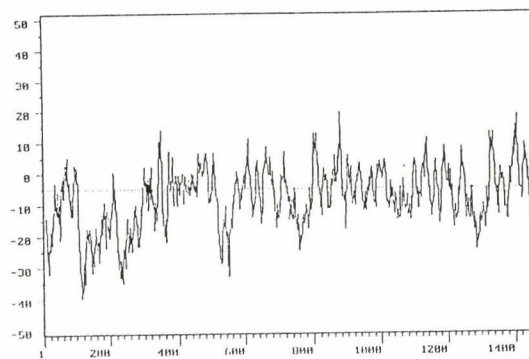


Figure 51 : Alternance de domaines hydrophiles / hydrophobes au niveau peptidique, de BEN1A.

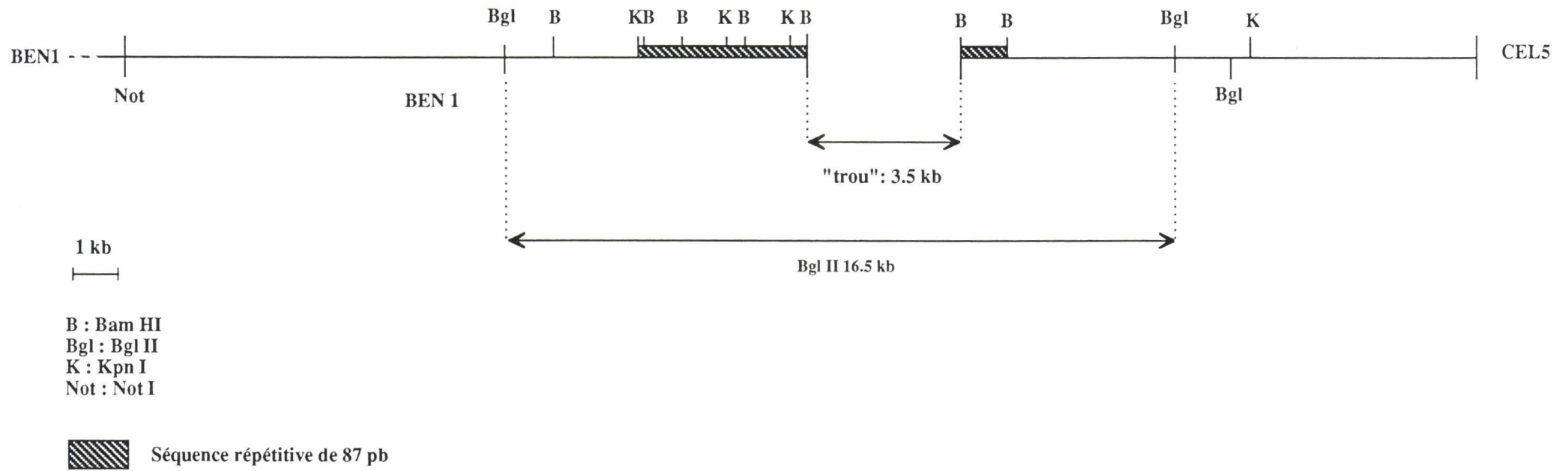


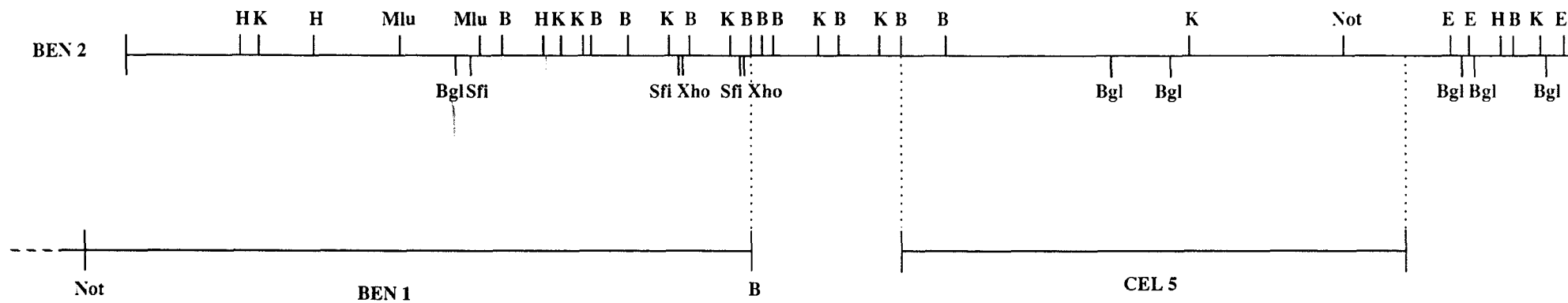
Figure 52 : Les clones BEN1 et CEL5 sont séparés par 3.5 kb dans le génome.

Les deux séquences de part et d'autre de ce "trou" (3.5 kb) sont caractérisées par une séquence codante organisée en "tandem repeats" de 87pb. On peut alors imaginer qu'il existe une continuité de cette séquence entre BEN1 et CEL5 et ainsi supposer l'existence de nombreux autres "tandem repeats".

3) *Carte de restriction de BEN2*

Un nouveau criblage avec la sonde JER57 a permis d'isoler un autre clone cosmidique, BEN2, qui chevauche une partie de BEN1 et la totalité de CEL5. Ce clone n'a été isolé que récemment. Cependant, grâce à la réalisation de sa carte de restriction nous avons pu établir une continuité entre les 3 clones (Figure 53).

La carte de restriction de BEN2 met en évidence deux nouveaux fragments Bam HI de 1.6 kb et deux fragments de 0.3 kb hybridant avec JER57. On peut donc supposer que comme les fragments Bam HI de 1.6 kb déjà séquencés, cette région est constituée d'une séquence répétée en tandems de 87pb. Il existerait ainsi une séquence codante répétitive sur plus de 9 kb, formant un énorme exon comme cela a déjà été décrit pour le gène MUC2. Les séquences répétitives non encore séquencées pourraient correspondre aux clones ADNc, JER28, JER57, JUL7 et JUL10. Certains de ces ADNc pourraient également provenir d'un polymorphisme inter-individuel. Le clone BEN2 contient tous les fragments visualisés sur Southern blot et susceptibles d'hybrider avec JER57, comme le montre la figure 54, il apparait contenir le gène MUC5B entier. Ce clone est donc d'un grand intérêt car il est susceptible d'être le matériel de départ pour une étude qui sera menée ultérieurement au laboratoire et qui étudiera l'expression de ce gène transfecté en cellule eucaryote.



1 kb

B : Bam HI Mlu : Mlu I
 Bgl : Bgl II Not : Not I
 E : Eco RI Sfi : Sfi I
 H : Hind III Xho : Xho I
 K : Kpn I

Figure 53 : Continuité entre les clones BEN1, CEL5 et BEN2.

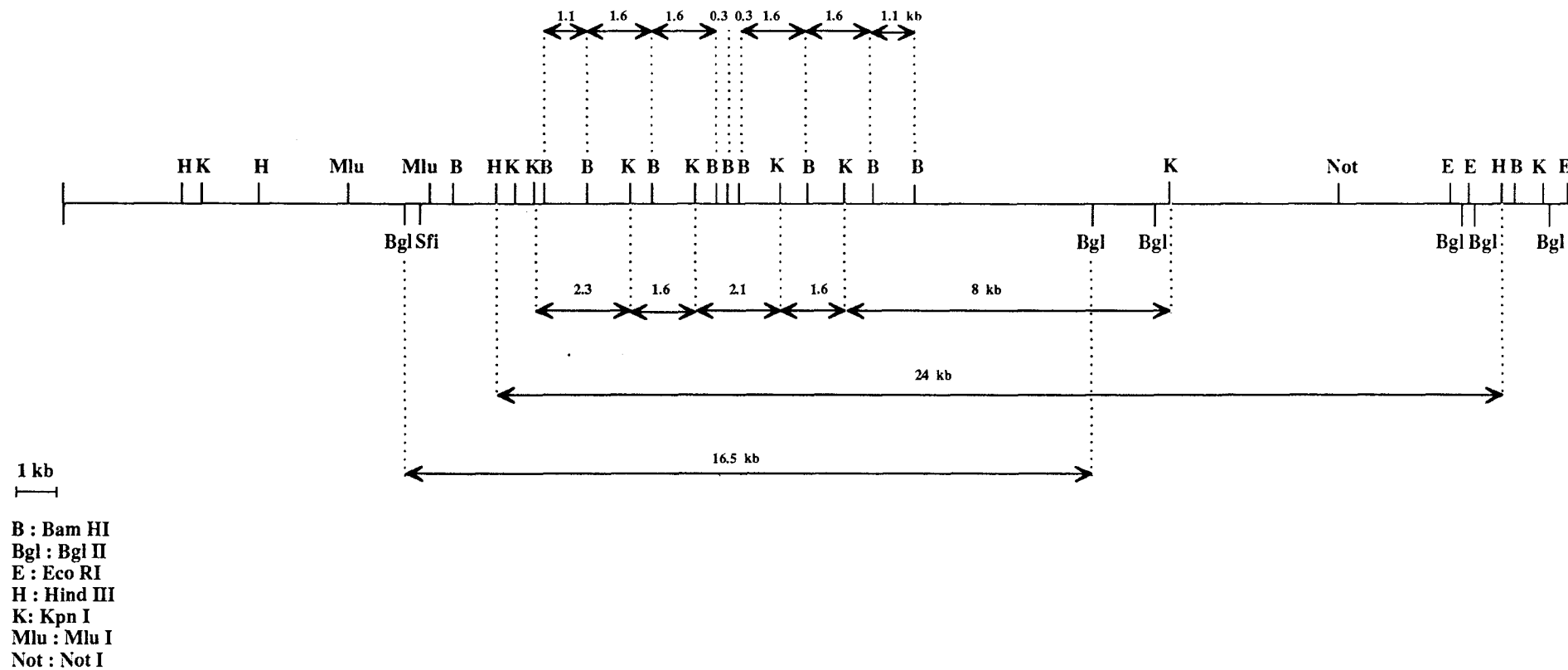


Figure 54 : Comparaison entre la carte de restriction de BEN2 et les fragments visualisés en Southern blot par la sonde JER57.

La synthèse de ces résultats fait apparaître que les clones BEN1, CEL5 et BEN2 nous permettront de définir le schéma complet d'organisation du gène MUC5B. Les résultats partiels actuellement en notre possession montrent que :

Le gène MUC5B semble être construit, comme le gène MUC2, d'un énorme exon central répétitif qui constitue la région O-glycosylable, avec peut-être de part et d'autre quelques petits exons. Ces petits exons par épissage alternatif formeraient les régions C- et N-terminales riches en résidus cystéine.

Les régions 5' et 3' terminales apparaissent conservées d'une espèce à l'autre contrairement aux régions répétitives dont le nombre de domaines élémentaires varie d'un individu à l'autre. Ce sont probablement des régions essentielles à la fonctionnalité des mucines. On peut alors imaginer grâce à un mécanisme d'épissage alternatif, la production, à partir d'un même motif répétitif, de mucines plus ou moins oligomérisées créant ainsi des propriétés de visco-élasticité différentes.

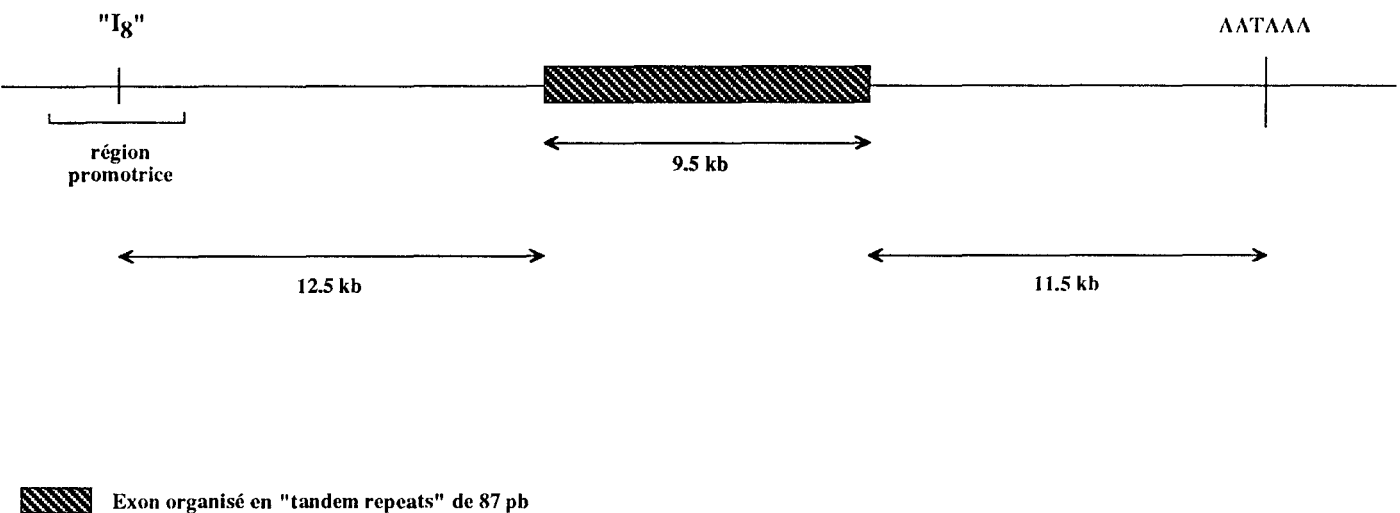


Figure 55 : Schéma hypothétique de l'organisation du gène MUC5B.

DISCUSSION

DISCUSSION

I. RESULTATS ACQUIS

A. LA CARTOGRAPHIE

L'existence de séquences partielles d'ADNc multiples possédant des expressions tissulaires caractéristiques ne constituait pas, malgré leur localisation commune en 11p15 un faisceau d'arguments définitifs pour affirmer qu'il s'agissait de gènes distincts. Les résultats de cartographie ont permis de définir sur plusieurs centaines de kilobases l'existence d'au moins 4 de ces gènes : un gène de mucines intestinales MUC2, un gène de mucines gastriques MUC6 et deux gènes de mucines trachéobronchiques MUC5B et MUC5AC.

En outre elle a montré que MUC5A et MUC5C sont regroupées derrière un même îlot CpG sur une vingtaine de kilobases. Parallèlement, les résultats d'hybridation in situ obtenus au laboratoire ont démontré une coexpression de ces deux ADNc. L'ensemble de ces résultats a permis d'affirmer que MUC5A et MUC5C constituent un seul et même gène appelé désormais MUC5AC [82].

Sur 725 kb, une concentration importante d'îlots CpG nous laisse suspecter la présence d'un cluster de gènes. La séquence en 5' du clone BEN1, caractérise un nouveau gène localisé entre MUC5AC et MUC5B. Il reste évidemment à démontrer si ce gène organisé lui aussi en structure répétitive est encore un gène de mucine. Il pourrait donc exister dans cette petite région de la carte (entre MUC5AC et MUC5B couvrant environ 70 kb) 3 gènes de mucines les uns à la suite des autres.

B. LES GENES DE MUCINES TRACHEOBRONCHIQUES

1) *MUC5AC*

L'isolement du clone génomique CEL2 possédant une partie exonique de MUC5AC permet d'authentifier les séquences ADNc en "tandem repeats" de 8 acides aminés.

L'homologie de séquence entre une partie exonique de MUC5AC, un ADNc de muqueuse gastrique et un ADNc de muqueuse trachéobronchique démontre l'expression multitissulaire du gène MUC5AC. Une coexpression dans les muqueuses intestinales et respiratoires a également été démontrée pour MUC2.

2) *MUC5B*

Le gène MUC5B est le premier gène entier de mucine que nous ayons isolé. Il a une taille d'une trentaine de kilobases.

Les séquences du domaine exonique, organisé en "tandem repeats" de 87 pb, permettent de vérifier l'authenticité des clones ADNc et d'écarter les hypothèses d'artéfacts de clonage quant à leur existence.

Ces séquences répétées semblent s'étendre sur plus de 9.5 kb, formant ainsi un énorme exon. Cet exon coderait pour un peptide d'environ 3000 a.a., caractérisé par l'alternance de domaines glycosylables et de domaines nus. En fonction du pourcentage d'acides aminés potentiellement O-glycosylables (35-40%) des régions séquencées, on peut estimer que le nombre maximum approximatif de chaînes glycaniques que pourrait supporter ce peptide est compris entre 1000 et 1200. L'apomucine produite aurait une taille d'au moins 300 kD.

On peut observer plusieurs niveaux de répétitions dans ce gène de mucine :

1) une répétition d'un motif élémentaire plus ou moins dégénéré de 87 pb engendrant une répétition complexe au niveau peptidique de zones glycosylables et de zones nues organisées de façon alternée;

2) une répétition des fragments de restriction générés par Bam HI, Kpn I, ou encore Sma I constitue une super-structure répétitive. Il est probable que ces nombreuses répétitions découlent de la duplication partielle ou totale d'un motif de base.

On observe également encore une répétition à plus grande échelle au niveau de la mégacartographie, où la distance entre les îlots, c'est à dire les gènes, définit des structures ayant un dénominateur commun de 20kb (20, 40, 65, 110...). Ceci nous permet donc de supposer que ces gènes appartiennent à la même famille multigénique et qu'ils viennent d'un gène ancestral

commun. La construction apparemment superposable de MUC2 et MUC5B étaye encore cette hypothèse.

II. LES INTERROGATIONS

A. UN CLUSTER "MUC" ?

La mégacartographie réalisée sur 725 kb révèle au moins 12 îlots CpG. Hors seulement 4 gènes de mucines non adjacents sont identifiés jusqu'à maintenant dans cette région. Pour vérifier l'hypothèse du "cluster MUC", il nous faudra étudier les gènes adjacents à MUC5B.

Pour valoriser cette cartographie, il est maintenant indispensable de préciser sa localisation dans la région 11p15.5 par des liaisons physiques à des loci environnants tels que INS, IGF2, HRAS, HBB, TH... Ce travail de "linkage" est en cours. Il permettrait également de déterminer l'orientation télomère-centromère de la carte et peut-être d'identifier des liaisons génétiques entre les gènes de mucines et les gènes proches, entre autres ceux qui sont impliqués dans les phénomènes tumoraux. Ceci ouvrirait de nouveaux horizons quant à l'étude de l'implication des gènes de mucines dans les phénomènes tumoraux. Les sondes de mucines pourraient constituer également des marqueurs hautement polymorphiques utilisables pour l'étude génétique des gènes adjacents déterminant des maladies mentales, métaboliques, tumorales à composante héréditaire.

B. LE GENE MUC5B

L'organisation du gène MUC5B reste encore à compléter. Combien d'exons contient-il ? Sont-ils soumis à un mécanisme d'épissage classique?

Plusieurs sites potentiels d'épissage ont été décelés dans les séquences en 3' du gène mais nous n'avons encore aucune preuve définitive de la réelle existence de tous ces sites.

Par contre aucun site d'épissage n'a été trouvé dans l'exon central. Le début et la fin de l'exon n'ont toujours pas été précisés ; or, cet exon s'étend déjà sur plus de 9.5 kb.

Un seul site typique de polyadénylation a été décelé. Existe-t-il d'autres sites comme celui-ci en amont ou en aval qui engendreraient des ARNm de tailles différentes?

L'étude de la séquence de fragments correspondant à l'extrémité 5' du gène est en cours. Nous en attendons les informations sur le ou les sites d'initiation, sur les séquences promotrices et donc sur les systèmes de régulation du gène MUC5B. Il reste néanmoins un certain volume de travail à fournir pour pouvoir réaliser une carte précise de "l'anatomie" de MUC5B.

C. LE GENE DE SOURIS HOMOLOGUE A MUC5B

Le fragment d'ADN génomique de souris ROM11X-6, isolé grâce à la sonde d'ADNc humaine correspondant à MUC5B n'a pas été localisé contrairement à nos espérances dans la région murine homologue à la région 11p15. Nous y voyons deux explications possibles.

1) Le gène murin de mucine homologue à MUC5B n'est pas associé à la région synténique correspondant à la région 11p15, chez la souris.

2) Le gène caractérisé par ROM11X-6 n'est pas en réalité l'homologue de MUC5B mais un autre gène de mucine apparenté. Pour l'instant, nous adhérons à la deuxième hypothèse dans la mesure où le criblage d'autres banques de muqueuses humaines, avec la sonde JER57, nous a permis, à plusieurs reprises d'identifier des ADNc localisés sur le chromosome 16 humain; or précisément, la région homologue au chromosome 16 humain à pour correspondance le chromosome 8 de souris. Il pourrait donc encore exister d'autres gènes de mucines ou d'autres gènes apparentés sur le chromosome 16 humain. En l'état actuel, nous n'avons pas suffisamment de preuves pour authentifier les clones du chromosome 16 à des clones de mucines.

D. L'EXPRESSION DES GENES DE MUCINES ?

La plus insolite des caractéristiques, commune aux gènes de mucines est leur extrême hétérogénéité d'expression. MUC1 et une mucine intestinale de rat semblent cependant y faire exception. Cette hétérogénéité pourrait laisser supposer l'existence de gènes complexes, composés de multiples exons, associés à de nombreux mécanismes de maturation. Or, que ce soit pour MUC2 ou MUC5B, il n'en est rien apparemment. Une organisation donc assez simple, même "rustique", certainement conservée au cours de l'évolution ne permet pas, en tout cas à elle seule,

d'expliquer une telle hétérogénéité de messagers. Par contre la coopération de plusieurs facteurs pourrait contribuer à cette hétérogénéité d'expression. Une alternative serait la dégradation artéfactuelle des ARNm de plus grande taille, en admettant que les ARNm de mucines soient plus fragiles que d'autres de taille équivalente. Les ARNm de grande taille constituent statistiquement de meilleures cibles aux ribonucléases que ceux de plus petite taille et les messagers de mucines doivent donc être spécialement sensibles à la dégradation. Il est également possible que la structure répétitive des ARNm les rendent plus instables et/ou plus sensibles aux RNases par l'intermédiaire de mécanismes encore inconnus.

Toutefois, les facteurs artéfactuels ne semblent pas être la seule cause d'une telle expression. En effet, lorsque l'on examine l'expression des gènes MUC1 à MUC5 sur modèles cellulaires en culture par exemple HT-29MTX, on remarque un profil d'expression en bandes discrètes souvent [192].

Un turn-over rapide des ARNm serait une autre hypothèse, qui pourrait être examinée expérimentalement par une technique de "pulse-chase" dans des cellules en culture.

Il faut souligner que la plupart des ADNc isolés de mucines ne correspondent pas, comme on serait en droit de l'attendre, de par le procédé de fabrication des banques, aux extrémités 3' terminales des ARNm. Que sont donc ces ARNm tronqués ? Existe-t-il un système de régulation afin de synthétiser des mucines de toutes les tailles ? Le continuum est-il dû à un système de dégradation naturelle et contrôlée des ARNm et non à la synthèse d'une multitude de messagers différents ? Puisque les ARNm isolés semblent, généralement, issus d'un seul grand exon organisé en "tandem repeats" et qu'il est très difficile d'isoler les extrémités 3' terminales, on peut imaginer que ces extrémités ont été détruites par un système de dégradation progressif. Ceci permettrait d'expliquer le continuum observé en Northern blot. Ainsi, à partir d'un ARNm d'une vingtaine de kilobases, on pourrait obtenir un continuum d'ARNm de tailles de plus en plus petites. Seul l'ARNm initial posséderait une extrémité 3' terminale qu'il est si difficile d'isoler au milieu de cette multitude de molécules.

III. DISCUSSION

A. L'HETEROGENEITE GENIQUE

A ce jour, 7 gènes de mucines humaines ont été identifiés. Il en existe 4 en 11p15.5 et peut-être une douzaine ! Si chaque localisation chromosomique correspond à un cluster "MUC" comme celui observé en 11p15.5, il pourrait alors exister une vingtaine de gènes de mucines différents.

Pourquoi une telle diversité génique ? Pourquoi une telle hétérogénéité peptidique qui, de plus, est renforcée par une hétérogénéité en nombres et en structures des chaînes glycaniques ? Pourquoi la muqueuse doit-elle disposer de tant de molécules de mucines, à la fois semblables et néanmoins distinctes ? Y a-t-il une relation causale entre l'hétérogénéité peptidique et l'hétérogénéité glycanique ? L'hétérogénéité des mucines est-elle directement nécessitée par la diversité des agressions auxquelles la muqueuse doit répondre ?

Les gènes de mucines peuvent être considérés, par leur nombre élevé, comme une famille multigénique et par leur similitude d'organisation et de séquence comme des gènes répétés. Il existe deux catégories de gènes répétés, qui ne sont pas mutuellement exclusives : les gènes qui produisent une "répétition de dosage" c'est à dire des clusters de gènes constitués d'un grand nombre de copies de séquences identiques répétées en tandem, tels que les gènes des ARNr et des histones; et les gènes qui sont constitués de gènes répétés variables (variant repetition) [193]. Les gènes répétés variables ne sont pas identiques mais sont apparentés. Ils codent pour des produits qui peuvent avoir des fonctions différentes dans l'organisme. On pourrait donc classer les gènes de mucines dans cette catégorie de gènes répétés variables qui englobe déjà les gènes codant pour les immunoglobulines [194], les globines, l'actine, l'hormone de croissance humaine, les interférons et sûrement bien d'autres.

Ces gènes forment des familles plus ou moins complexes dont le nombre de gènes distincts diffère selon les familles. Ils peuvent être liés ou dispersés dans le génome.

	Nombre de gènes	liés ou dispersés	localisation
Collagène 6	plusieurs	l	21
Actine	20-30	d	
Tubuline	5 α , 5 β et 15 pseudo	d	
Globine	3 α et pseudo	l	11
	5 β et pseudo	l	16
IFNα	≥ 15 et > 3 pseudo	l	9p
IFNβ	≥ 2 et ?	d	9p, 2, autres
IFNγ	1		12

l : liés d : dispersés

Les multiples gènes d'une même famille, bien que quelquefois fort différents, dérivent la plupart du temps de la duplication d'un gène ancestral commun. Les gènes de mucines pourraient eux aussi constituer une famille de gènes apparentés et être issus d'un même gène ancestral. Bien que les gènes MUC5B, MUC5AC et MUC2 soient assez divergents pour ne pas s'hybrider entre eux, leur organisation apparaît suffisamment semblable pour qu'on imagine leur origine par répétition en tandem suivie d'une divergence dans le temps.

Si l'on observe attentivement la carte révélant les îlots CpG (Figure 33), et si l'on suppose que la plupart des îlots ont pu être décelés, on remarque qu'il existe une certaine symétrie, un ordre dans la taille et la disposition des différents fragments avec une répétition des fragments de 20, 40, 65-70, 110-120, 140, 160...kilobases. Ces tailles de fragments apparaissent à peu près toutes comme des multiples de 20, comme si les fragments de part et d'autre du centre de la carte étaient formés de la duplication ou de "n" duplications d'un fragment d'une vingtaine de kb. Les fragments situés au centre de la carte sont plus petits (20 kb) que ceux situés de part et d'autre en 5' et 3'. Est-ce dû au manque de sonde spécifique de ces régions terminales, qui empêcherait de mettre en évidence certains îlots CpG ? Ou, est-ce que plus l'on s'éloigne du centre plus les gènes sont réellement grands? Ces plus grands fragments sont-ils formés d'un gène balisé par l'îlot CpG et de un ou plusieurs pseudo-gènes. Ou, est-ce que, par exemple le fragment de 120 kb, est constitué de 6 gènes dont un seul est actif (1 seul îlot détectable) dans les lignées étudiées ? Quel que soit le cas de figure, il apparaît plausible que ces gènes dérivent de la duplication d'un gène ancestral d'une vingtaine de kilobases dont les séquences auraient petit à petit divergé au cours de

l'évolution. On peut alors imaginer un schéma d'évolution des gènes de mucines basé sur celui des gènes de la famille de l'hormone de croissance décrit par G. S. Barsh [195] (Figure 56).

L'extrême hétérogénéité des gènes de mucines est-elle la conséquence d'une pression de sélection dans l'acquisition d'une meilleure adaptabilité aux agressions de l'environnement ?

L'amplification des séquences d'ADN a un rôle dans l'évolution. Il est probable que la duplication en tandem de gènes se produise relativement fréquemment dans toutes les cellules, à la suite d'erreurs occasionnelles, au cours du processus normal de réplication et de recombinaison de l'ADN. Généralement ces duplications d'ADN ne confèrent aucun avantage sélectif à l'organisme dans lequel elles se produisent, et l'exemplaire supplémentaire du gène est rapidement perdu. Une copie altérée peut néanmoins devenir fonctionnelle, coder pour une nouvelle fonction, ou répondre à des signaux de régulation différents. Si les séquences altérées se révèlent utiles ou confèrent une plus grande adaptabilité, les nouveaux variants peuvent eux-mêmes être conservés par la pression de sélection.

Un tel modèle permettrait d'expliquer l'origine de l'hétérogénéité peptidique des mucines. L'évolution par amplification aurait ainsi généré des gènes différents mais apparentés permettant de synthétiser une multitude de molécules différentes plus adaptées à leurs nombreuses fonctions.

Les gènes de mucines ainsi amplifiés ont-ils évolués de façon concertée afin de conserver une certaine homogénéité de structure au sein de cette famille multigénique ?

Dans les familles multigéniques, la répétition en tandem n'est toutefois pas limitée aux gènes entiers, elle peut se manifester au niveau d'une région codante : les gènes de séralbumine, et de l' α foetoprotéine [196], le gène du récepteur des lipoprotéines de faible densité [197], le gène du collagène [198], le gène de la protéine glue chez *Drosophila* [199], les gènes des P.R.P [106-109] en sont une illustration. Le gène du collagène est constitué d'un grand nombre d'exons ayant tous des longueurs qui sont des multiples de 9 pb. On peut imaginer la structure de ce gène comme construite par des amplifications en tandem d'un motif initial de 9 pb. Le gène Sgs 4 de *Drosophila* est formé de répétitions directes en tandem d'un motif de 21 pb (soit 7 a. a.). Les répétitions ne sont pas parfaites et les heptapeptides répétés divergent donc également. Le nombre de répétitions de la séquence canonique varie selon la souche de *Drosophila* (19, 22 ou 27 répétitions). De tels changements sont supposés consécutifs à des crossing-over homologues inégaux.

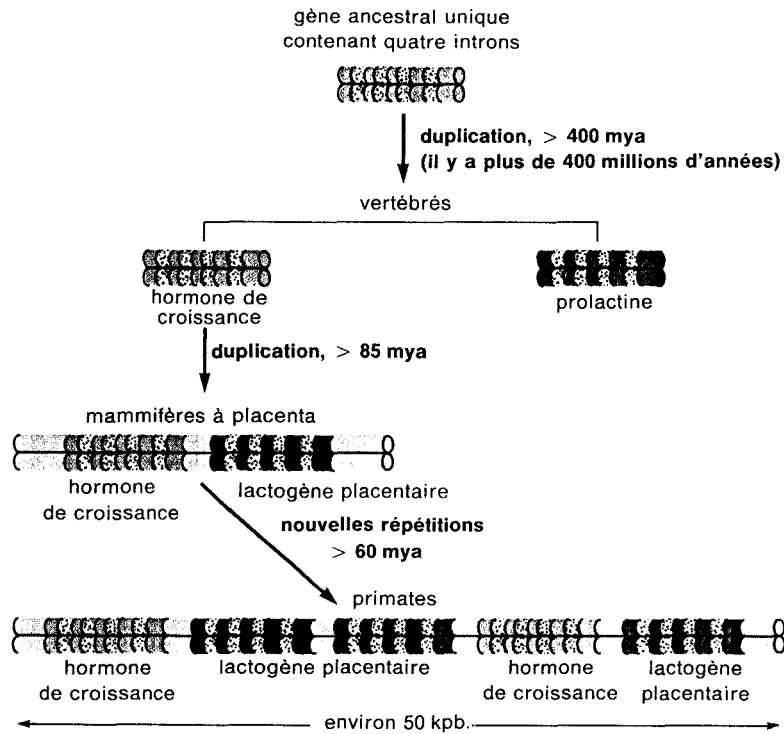


Figure 56 : Représentation schématique de l'évolution de la famille multigénique de l'hormone de croissance humaine. Les variations de tailles des introns ne sont pas représentées. D'après G. S. Brash.

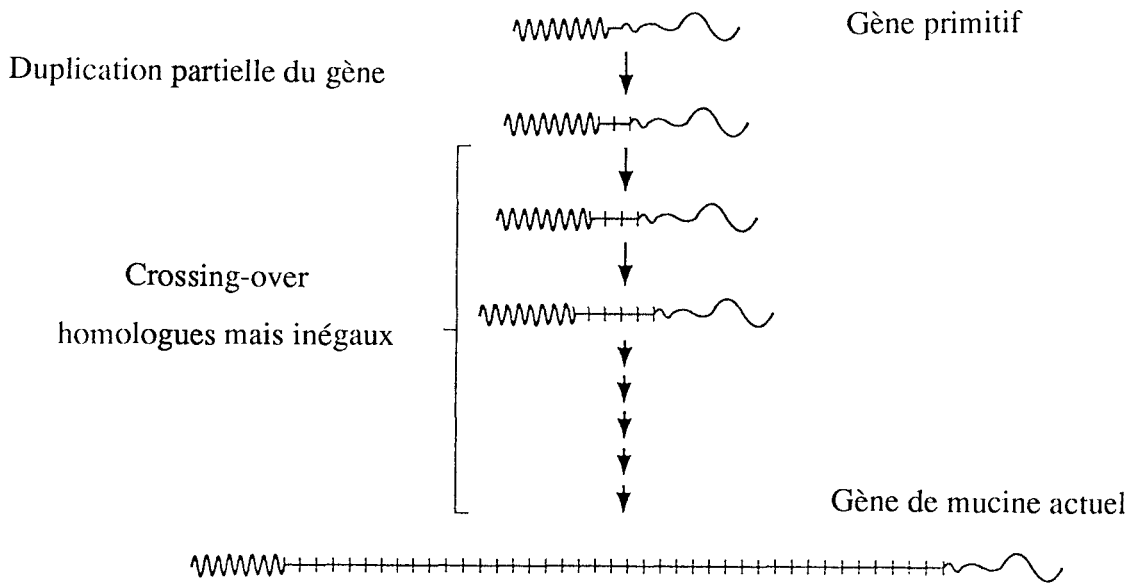


Figure 57 : Schéma hypothétique de l'évolution des séquences répétitives des gènes de mucines. D'après J. R. Gum.

A partir des modèles précédents, on peut imaginer pour les gènes de mucines un mécanisme commun d'évolution basé sur la réitération d'un motif canonique (Figure 57).

Modèle de Gum [200] : Un petit gène primordial subirait une duplication pour produire un simple fragment d'ADN répété. Si les crossing-over maintiennent le cadre de lecture ouvert et donnent des séquences riches en sites de O-glycosylation, alors ces altérations seront favorables à la structure des mucines. Des crossing-over homologues mais inégaux augmenteraient alors rapidement le nombre de "tandem repeats" dans des allèles différents, créant ainsi une mucine plus longue et donc par conséquent plus "gluante" et "visqueuse". Les allèles les plus grands seraient ainsi privilégiés par une sélection naturelle, produisant des mucines actuelles avec une séquence répétée très étendue.

Bien que ce modèle soit invérifiable dans l'immédiat, il permet d'expliquer plusieurs propriétés des mucines actuelles :

1) Les répétitions, une fois formées, devraient évoluer très rapidement. Par conséquent, les domaines répétitifs devraient être très divergents entre des espèces relativement éloignées; c'est effectivement ce qui est généralement observé [70, 77].

2) Le polymorphisme de type VNTR est compatible avec une évolution rapide.

3) Ce modèle prédit des séquences uniques N- et C-terminales vraisemblablement assez bien conservées, les mutations au niveau de ces extrémités devant être limitées par le rôle fondamental de ces domaines dans la formation des oligomères de mucines [77, 96, 190].

Ce modèle constituerait un deuxième niveau d'évolution après l'amplification génique. L'amplification génique aurait permis d'augmenter le nombre de gènes primaires et ainsi de diversifier les "petites mucines primordiales". L'amplification interne aurait, elle, favorisé l'élaboration de mucines plus longues et donc plus efficaces.

B. EXPRESSION ET REGULATION

Plusieurs facteurs modifiant l'expression des mucines ont été décrits. L'innervation cholinergique a été mentionnée par H. Florey comme étant impliquée dans la régulation de la sécrétion de mucus digestif [201]. Chez le chien et le rat, il a été observé que l'acétylcholine et le carbachol produisaient une augmentation de l'épaisseur de la couche de mucus au niveau de la muqueuse gastrique [202, 203]. Il a également été démontré que les cellules caliciformes des

cryptes intestinales, étaient sensibles à l'acétylcholine et à ses analogues alors qu'au niveau colique elles étaient insensibles à ce stimulus [204]. Au niveau de la trachée, la sécrétion de mucines peut être induite par l'AMPc [205]. Les prostaglandines ont aussi été décrites comme stimulant la sécrétion des mucines. In vitro, sur des épithélia de l'estomac, de la trachée et de la vésicule biliaire, il a été observé que la prostaglandine E2 accroissait la synthèse des mucines qui au contraire diminue après l'application d'indométhacine ou d'aspirine.

Toutes ces observations anciennes donneront les bases à l'étude des éléments régulateurs des gènes MUC transfectés dans des modèles cellulaires en culture ou dans les animaux transgéniques.

Une autre question d'importance doit être posée :

Les gènes de mucines sont-ils soumis à une régulation concertée, à une régulation transcriptionnelle multi-loci ?

Les gènes organisés dans un locus multigénique ont des fonctions apparentées et sont exprimés de façon coordonnée et régulés différemment suivant le stade de développement. Pour expliquer une telle régulation différentielle, il a été décrit pour les gènes des β globines [206] et les gènes du CMH, une région de contrôle en amont de chacun des loci appelée LCR (Locus Control Region) [207]. Ces éléments LCR ont un rôle important pour l'activation des loci au cours du développement. Il existe également d'autres niveaux de régulation plus complexes qui résultent de la compétition entre les gènes vis à vis de l'activation par le LCR et du fait que l'expression d'un gène peut affecter le moment d'activation et le taux d'expression d'un autre gène. A l'intérieur du locus HbB les gènes des β globines sont organisés dans l'ordre dans lequel ils sont activés. L'ordre des gènes semble avoir un rôle primordial dans la régulation. Ce système de régulation est associé à des loci multigéniques très conservés, c'est à dire que l'ordre des gènes est maintenu au cours de l'Evolution.

Les progrès considérables réalisés au cours de ces dernières années dans l'établissement de la carte génomique de l'homme et de la souris ont permis de reconnaître l'existence de nombreux groupes de liaisons conservés chez les deux espèces [186]. Ainsi de nombreux gènes de la région 11p15.5 se retrouvent liés dans la région 90 du chromosome 7 de la souris. S'il s'avère exact que le cluster MUC fait également partie de ce locus conservé, on pourra donc étudier l'hypothèse selon laquelle les gènes de mucines sont soumis à un système de régulation de type LCR.

C. ROLE DES MUCINES DANS LA TUMORIGENESE ?

Une association entre expression des mucines et tumorigénèse de divers types cellulaires a été décrite dans la littérature, surtout à propos de l'épisialine (MUC1) [208-210]. MUC1 est essentiellement associée aux tumeurs mammaires, pancréatiques et ovariennes alors que MUC2 est principalement associée aux tumeurs gastro-intestinales et trachéobronchiques [211, 212].

La fonction exacte des mucines "tumoraux" et le mécanisme permettant d'augmenter très nettement l'expression de ces glycoprotéines dans les tumeurs sont inconnus. Cependant, un certain nombre d'hypothèses émaillent la littérature :

Les cellules tumorales perdent leur caractère polarisé et une forte concentration de mucines membranaires s'étend sur toute la surface cellulaire. En raison de leur forte activité glycolytique, ces cellules produisent un taux élevé de lactate qui diminue le pH. L'excès de mucines protégerait les cellules du pH acide et favoriserait interactions et agrégations cellulaires. La O-glycosylation des mucines membranaires permet d'étendre leurs possibilités interactives bien au delà de la surface cellulaire. L'affinité des mucines "tumoraux" vis à vis des autres cellules, tumorales ou non, favoriserait la métastase [213, 214]. Les mucines membranaires protégeraient les cellules tumorales du système immunitaire en masquant, par encombrement stérique, les antigènes de surface impliqués dans la reconnaissance immunitaire. Les mucines acides protégeraient les cellules tumorales de la lyse par les lymphocytes T "natural killer" [215]. Les cellules cancéreuses libéreraient par protéolyse un taux élevé de mucines dans le sérum. Ces mucines circulantes formeraient des complexes immuns qui agiraient comme des facteurs bloquant le système immunitaire [216].

Le cluster MUC comprenant les gènes MUC2, MUC5B, MUC5AC et MUC6 est localisé dans une région du génome à la fois très conservée et impliquée dans de nombreux phénomènes de réarrangements chromosomiques. En effet des aberrations telles que des pertes de l'hétérozygotie, des trisomies, des translocations ont été trouvées au niveau de la région 11p15 chez des patients atteints de tumeur de Wilm's [217], du syndrome WAGR, du syndrome de Beckwith-Wiedemann (BWS) [218-220], de rhabdomyosarcome [221], de tumeur du sein [222], de leucémie lymphoblastique aigüe des cellules-T [223], de leucémie myéloïde aigüe [224]. Les gènes MUC sont-ils eux aussi touchés par ces anomalies qui déréguleraient leur expression ? Constituent-ils des gènes de susceptibilité aux cancers ?

La surexpression des mucines est-elle directement associée à l'expression d'oncogènes ou d'antioncogènes ? Existe-t-il des facteurs de régulation tels les facteurs de type LCR, communs à certains oncogènes, antioncogènes, gènes de mucines, de cette région p15.5 ?

Les connaissances récentes acquises sur les mucines grâce à la biologie moléculaire ont permis de mettre au jour l'existence d'un système biologique de grande complexité, dont le rôle est fondamental à la vie. L'étude de ces molécules et de leurs gènes devrait apporter un nouvel éclairage aux domaines de recherche très actuels de la biologie que sont l'étude de la différenciation cellulaire, de la communication intercellulaire, de la tumorigénèse.

APPENDICE TECHNIQUE

APPENDICE TECHNIQUE

I. ANALYSE DE L'ADN GENOMIQUE PAR LA TECHNIQUE DE SOUTHERN BLOT

La technique de Southern blot décrite en 1975 permet de visualiser une séquence dite unique à partir d'ADN génomique. Elle comporte les étapes suivantes:

- la préparation de l'ADN génomique ;
- l'hydrolyse de l'ADN par une ou plusieurs enzymes de restriction appropriées ;
- la séparation des fragments par électrophorèse en gel d'agarose ;
- la dénaturation in situ de l'ADN ;
- le transfert de l'ADN par capillarité sur un support solide (membrane de nylon) ;
- l'hybridation avec une sonde monobrin marquée (^{32}P) ;
- les lavages et la révélation du signal par autoradiographie.

A. PREPARATION DE L'ADN GENOMIQUE

L'extraction de l'ADN génomique à partir de sang, a été réalisée selon la méthode de M. Jeanpierre [225].

1) Lyse cellulaire

Le sang prélevé sur anticoagulant est congelé puis décongelé rapidement à 37°C et réparti en fractions de 10 ml dans des tubes de 50 ml.

La lyse des cellules est alors effectuée par l'adjonction de 40 ml d'une solution Tris 10 mM pH 7.4 , EDTA 10 mM pendant 10 minutes.

Une centrifugation de 10 minutes à 3 000 g permet d'éliminer dans le surnageant, l'hémoglobine et les autres constituants solubles. L'opération de lavage est répétée plusieurs fois (5 à 6 fois) jusqu'à l'obtention d'un culot de débris cellulaires et de noyaux pratiquement blanc.

2) Protéolyse

14 ml d'un dénaturant protéique, l'hydrochlorure de guanidinium 6 M filtré, sont ajoutés au culot qui est alors homogénéisé par une agitation ménagée dans un automate ("turbula") durant une trentaine de minutes. Puis 2 ml d'un détergent anionique, du sarcosyl 10%, et 150 µl de protéinase K à 10 mg/ml (préparée extemporanément) sont ajoutés à chaque préparation. Le tout est incubé une heure à 60°C ou une nuit à 37°C.

3) Précipitation de l'ADN

L'ADN est ensuite précipité par l'addition (qsp 50 ml) d'éthanol absolu froid auquel a préalablement été ajouté 1 ml d'acétate d'ammonium 7.5 M. Puis, il est recueilli par enroulement autour d'une baguette de verre, et rincé 4 fois à l'éthanol à 70% (v/v). L'ADN ainsi purifié est resuspendu dans du TE 1X (4 ml pour un tube de 10 ml de sang). Afin d'obtenir une bonne homogénéisation, l'ADN est laissé 48 h sous agitation.

4) Quantification et appréciation de la pureté

La concentration d'ADN extrait est calculée par mesure au spectrophotomètre de la densité optique (DO) à 260 nm. Une concentration de 50 µg/ml correspond à 1 de DO. La pureté de l'ADN est appréciée par le rapport de la DO à 260 nm sur la DO à 280 nm, qui doit être voisin de 2. Les solutions d'ADN sont stockées à 4°C.

B. HYDROLYSE ENZYMATIQUE

L'hydrolyse est réalisée sur 10 à 40 µg d'ADN génomique à raison de 5 à 10 U d'enzyme par µg d'ADN dans les conditions recommandées par le fabricant. Le temps d'incubation est de l'ordre de 24 heures. Une recharge d'enzyme peut être effectuée au bout de 6 heures si nécessaire.

C. ELECTROPHORESE

Les fragments d'ADN hydrolysés sont séparés par une électrophorèse en gel d'agarose (0.8%, en tampon phosphate) contenant du BEt (0.5 µg/ml), une nuit à 20 volts.

Sur chaque gel est déposé un témoin de masse moléculaire. Ce marqueur "13i" est un phage recombinant qui hydrolysé par l'endonucléase Eco RI libère 8 fragments dont les tailles sont 31, 22, 11, 4.3, 3.2, 2, 1.3 et 0.5 kb.

D. TRANSFERT ET FIXATION

Avant le transfert sur membrane, l'ADN doit être dénaturé. Pour cela, le gel est trempé 2 x 30 mn dans du tampon de dénaturation (NaCl 1.5 M; NaOH 0.5 N). Puis il est neutralisé par 2 bains de 30 mn dans du tampon de neutralisation (Tris - HCl 0.5 M pH 7.5 ; NaCl 3 M). L'ADN est alors transféré sur membrane de Nylon Hybond N+ (Amersham) par capillarité en tampon SSC 20X pendant une nuit. Le matériel nucléique est ensuite fixé sur la membrane par une exposition de 3 mn aux U.V. (312 nm).

E. HYBRIDATION A UNE SONDE NUCLEIQUE

1) Préparation de la sonde

Chaque sonde est réalisée à partir de 25 ng d'ADN. Deux techniques ont été utilisées :

- la "Nick translation" en présence d'un nucléotide marqué : α^{32} P dCTP (kit Amersham);

- le "multiprime" utilisant des amorces hexanucléotidiques et une extension par la Klenow en présence de α^{32} P dCTP (Random Primed DNA labelling kit Boehringer) qui nécessite au préalable la dénaturation de l'ADN (dénaturation par la chaleur : 10 mn à 95°C);

La sonde est purifiée par gel filtration sur colonne de Sephadex G50 éluee par du tampon STE-SDS 0.1%. Elle est ensuite dénaturée en présence de 1/10 (v/v) de NaOH 3 N, 10 mn à température ambiante, et par l'addition de HCl 3 N (5/6 du volume de NaOH).

2) Hybridation

La membrane avant d'être hybridée, doit passer par une étape de préhybridation qui consiste à saturer ses sites afin d'obtenir une fixation spécifique de la sonde sur l'ADN. Pour cela, la membrane est rincée dans du SSC 6X, puis incubée dans du tampon de préhybridation

SSC 6X - Denhart's 5X - SDS 0.5% (à raison de 50 μ l/cm² de membrane) au moins 1 heure à 65°C. Après cette première incubation, le tampon est remplacé par du tampon d'hybridation SSC 6X - Denhart's 5X - SDS 0.5% - sulfate de dextran 10% (p/v) - ssDNA (10 mg/ml) 2.5 % (v/v) auquel est ajoutée la sonde marquée au ³²P. L'hybridation est réalisée par une incubation d'une nuit à 65°C.

Pour éliminer certaines hybridations parasites dues à la présence de motifs de type "Alu family" ou autres séquences répétitives du génome humain, 100 μ g d'ADN humain dénaturé doivent, dans certains cas, être ajoutés pour 25 ng de sonde.

3) Lavages

L'excès de sonde est éliminé par plusieurs lavages réalisés comme suit :

- rinçage dans du SSC 2X ;
- lavage : 2 x 15 mn en tampon SSC 2X à 65°C ;
ou 2 x 15 mn en tampon SSC 2X - 0.1% SDS
- rinçage dans du SSC 2X.

4) Autoradiographie

La membrane ainsi lavée est mise en autoradiographie à -80°C pendant des temps variables pouvant aller d'une journée à 8 jours.

Avant de tester une nouvelle sonde, le blot est déshybridé par une solution de SDS 0.1% bouillant et un contrôle de deshybridation est effectué pour chaque autoradiographie.

F. HYBRIDATION A UNE SONDE OLIGONUCLEOTIDIQUE

1) Préparation de la sonde

Les oligonucléotides sont marqués en 5' OH par la T4 Polynucléotide Kinase selon le protocole du kit Amersham.

L'oligonucléotide est tout d'abord dénaturé par la chaleur, pendant 5 mn à ébullition, puis le mélange suivant est effectué :

5 picomoles de l'oligonucléotide (sous un volume $\leq 2 \mu\text{l}$)
1 μl de tampon de phosphorylation 10X
5 μl de $\gamma\text{ATP } ^{32}\text{P}$
1 μl de T4 Polynucléotide Kinase
qsp 10 μl avec H_2O stérile

Il est incubé à 37 °C au moins 30 mn. La réaction est arrêtée par un passage de 5 mn à 90 °C puis 5 mn dans la glace et par l'addition de 1/10^e de volume de STE 10X. L'oligonucléotide ainsi marqué est purifié par gel filtration sur "Quick Spin Columin G25" puis dénaturé par la chaleur. Un volume égal de formamide est alors ajouté.

2) *Hybridation*

La température d'hybridation ou de préhybridation dépend du T_m de l'oligonucléotide que l'on veut hybrider : $T \text{ d'hybridation} = (T_m - 10 \text{ à } 15^\circ\text{C})$.

Le tampon utilisé est le suivant : SSPE 6X, Denhardt's 10X, sulfate de dextran 10% (p/v), SDS 0.05%, ssDNA (10 mg/ml) 1.25% (v/v).

3) *Lavages*

Après une hybridation d'une nuit, l'excès de sonde est éliminé par un rinçage de la membrane dans du tampon SSPE 6X et un lavage de 15 mn dans du SSPE 4X - 0.1% SDS à une température légèrement supérieure à celle de l'hybridation ($T_m - 5^\circ\text{C}$).

II. ANALYSE DE L'EXPRESSION PAR LA TECHNIQUE DE NORTHERN BLOT

A. PREPARATION DES ARN TOTAUX

Le matériel biologique immergé dans l'azote liquide dès son recueil et conservé à -80°C est tout d'abord broyé dans l'azote liquide. Au tissu réduit en poudre fine sont ajoutés 40 ml de tampon GT (voir le lexique des formules). Le mélange est homogénéisé à l'aide du "Sorvall"

puis filtré sur gaze stérile afin d'éliminer le tissu conjonctif. Pour diminuer la viscosité, l'ADN est cassé par des passages successifs au travers d'aiguilles de diamètres décroissants.

16 g de chlorure de césium solide sont ajoutés à l'homogénat qui est ensuite déposé délicatement dans 6 tubes au dessus de 3.2 ml d'une solution de chlorure de césium 5.7 M, EDTA 0.1 M pH 7.5, DEPC 0.1%. Les tubes sont équilibrés deux à deux grâce à une solution constituée de 5 ml de tampon GT et de 2 g de CsCl. Ils sont ensuite centrifugés dans une ultracentrifugeuse Beckman L8 (rotor SW41) à 29 500 tours/mn pendant 15 heures à 18°C. Après élimination des surnageants, les culots sont repris par 1 ml de TE 1X - SDS 0.1% puis transférés dans 2 tubes Corex (30 ml).

Les ARN sont alors purifiés par 2 volumes de chloroforme - n butanol (4/1, v/v). Après agitation, les tubes sont centrifugés 5 mn à 3 000 tours/mn. La phase aqueuse est recueillie, alors que la phase organique est réextraite par 1 volume de TE 1X - SDS 0.1%. Les deux phases aqueuses sont réunies et réparties en 2 volumes égaux.

Les ARN sont précipités par 0.1 volume d'acétate de sodium 3 M pH 5.5 et 2.2 volumes d'éthanol absolu (-20°C) pendant au moins 2 heures à -80°C. Les ARN sont récupérés sous forme d'un culot après une centrifugation de 30 mn à 10 000 tours/mn. Le culot séché à l'azote est resolubilisé par 200 à 500 µl d'eau stérile. Les ARN extraits peuvent être ainsi conservés pendant quelques jours à -80°C. Pour une conservation à plus long terme, de quelques jours à un an, le culot de précipitation acétate/alcool doit être repris par un petit volume d'alcool absolu.

Afin de déterminer la quantité et la qualité des ARN extraits, 10 µl de chaque échantillon sont prélevés et dilués dans 1 ml d'eau. La concentration d'ARN est calculée par mesure de la densité optique au spectrophotomètre à 260 nm. Une DO à 260 nm égale à 0.025 correspond à 1 µg d'ARN par millilitre.

La pureté des ARN extraits est appréciée par deux critères :

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

- le rapport de la DO à 260 nm sur la DO à 280 nm (contamination par les protéines) doit être supérieur à 1.75.

B. ELECTROPHORESE DES ARN EN MILIEU DENATURANT

L'ARN purifié par la technique décrite précédemment est à nouveau précipité par 1/10^e de volume d'acétate de sodium 3 M pH 5.5 et 2.2 volumes d'alcool absolu. Le culot obtenu après une centrifugation de 30 mn à 10 000 tours/mn est cette fois-ci repris par du tampon de dénaturation de manière à avoir une concentration de 5 µg/µl. L'ARN est dénaturé par 5 mn de chauffage à 68°C en milieu dénaturant :

Tampon de dénaturation :

formamide désionisée sur résine AG 501 × 8	500 µl
formaldéhyde désionisée sur résine AG 501 ×	8178 µl
MOPS 10X	100 µl
eau DEPC stérile	qsp1000 µl

Avant usage, le matériel en contact avec la préparation : la cuve d'électrophorèse, le peigne et le barreau aimanté sont préalablement baignés plusieurs heures dans une solution de soude 1 N et rincés abondamment à l'eau stérile.

Le matériel dénaturé est maintenu dans la glace jusqu'au moment du dépôt sur gel. Des parties aliquotes de 10 à 50 µg d'ARN sont additionnées d'un tampon de dépôt contenant du glycérol et du bleu de bromophénol puis sont déposées sur un gel d'agarose à 0.8% préparé de la façon suivante :

agarose	0.8 g
MOPS 10X	10 ml
formol désionisé	17 ml
Eau DEPC stérile	qsp 100 ml

L'électrophorèse est effectuée en tampon MOPS 1X à 30 volts pendant une nuit sous agitation magnétique.

Le lendemain le gel est rincé par 6 bains successifs de tampon SSC 10X pendant 15 mn afin d'éliminer le formaldéhyde. L'intégrité des ARN est contrôlée par visualisation sous lampe U.V. des bandes d'ARN ribosomales (ARNr 18S : 2.06 kb; ARNr 28S : 5.30 kb) colorées au BEt. Pour cela, le gel est immergé 15 mn dans un bain de SSC 10X contenant 1/10 000 de BEt à 10mg/ml puis rincé 30 mn dans du SSC 10X.

C. TRANSFERT ET HYBRIDATION

Le matériel ribonucléique est transféré sur une membrane de type "Hybond C Extra" (Amersham) par capillarité, en tampon SSC 20X durant une nuit. La fixation covalente de l'ARN sur le support est assurée par la cuisson de la membrane à 80°C pendant 2 heures sous vide partiel.

Avant l'hybridation proprement dite, la membrane est tout d'abord incubée au moins 2 heures dans du tampon de préhybridation 3X. Ce tampon est ensuite remplacé par du tampon d'hybridation SSC 3X - dextran auquel est ajoutée la sonde marquée au ^{32}P . L'hybridation est réalisée par une incubation d'une nuit à 42°C.

L'excès de sonde est éliminé par 2 lavages de 15 mn à 55 ou 65°C en tampon SSC 2X-SDS 0.1%.

La membrane ainsi lavée est mise en autoradiographie à -80°C pendant des temps variables pouvant aller de quelques heures à une semaine selon les sondes.

III. MEGACARTOGRAPHIE DE RESTRICTION

A. PREPARATION DE L'ADN A PARTIR DE CELLULES EN CULTURE

Afin de maintenir l'intégrité de l'ADN, on doit éviter toutes les possibles cassures mécaniques (cisaillement par pipettage), chimiques, photochimiques (U.V.) et enzymatiques (nucléases); pour cela la préparation doit se faire rapidement avec des solutions stériles, en protégeant l'ADN par inclusion dans de l'agarose à faible point de fusion ("Low Melting Point" : LMP).

Les cellules sont tout d'abord lavées dans du PBS 1X à 37°C puis centrifugées 3 mn à 3 000 tours/mn. Le culot ainsi formé est repris par un volume de PBS 1X tel que la concentration soit de 3.5×10^7 cellules/ml. On ajoute un volume égal d'agarose LMP à 1% dans le PBS 1X et on coule le mélange dans un moule, composé de 20 trous ($0.6 \times 0.15 \times 0.9$ cm) d'une contenance de 80 μl . Chaque bloc contient 10^6 cellules.

Après solidification (30 mn sur la glace), les 20 blocs sont traités 2 fois par 5 ml d'EDTA 0.5 M pH 8, sarcosyl 1% et 100 $\mu\text{g/ml}$ de protéinase K à 50°C pendant 24 heures. Le

lendemain, les blocs sont lavés par du TE 1X, puis 2 fois par du TE 1X contenant 0.04 mg/ml de PMSF (Phenyl Methyl Sulfonyl Fluoride) dans l'isopropanol, à 50°C pendant 30 mn, afin d'éliminer la protéinase K et le sarcosyl.

Les blocs sont conservés dans l'EDTA 0.5 M à 4°C. Chaque bloc ainsi préparé contient 10 µg d'ADN.

B. DIGESTION DES BLOCS D'ADN PAR DES ENDONUCLEASES DE RESTRICTION

L'établissement d'une cartographie à grande échelle nécessite l'utilisation d'enzymes de restriction à site de coupure relativement rare. Les sites de reconnaissance de ces endonucléases sont des séquences de 6 ou 8 pb et/ou des séquences contenant 1 ou 2 motifs CpG.

Avant chaque hydrolyse, les blocs sont lavés 3 fois 20 mn dans du TE 1X pour éliminer l'EDTA puis sont équilibrés dans le tampon d'hydrolyse adéquat à l'enzyme pendant 15 mn. Chaque hydrolyse est réalisée sur un 1/2 bloc (40 µl) contenant 5 µg d'ADN, sous un volume réactionnel de 100 µl. Les conditions d'hydrolyse utilisées pour chaque enzyme sont :

- 1/2 bloc40 µl (5 µg d'ADN)
- Tampon 10 X 10 µl
- Enzyme 10 à 15 U/µg ex : Sfi I : 75 U
Nru I)
Mlu I) 50 U
Not I)
- SAB (1 mg/ml) 10 µl ou 0 µl
- Spermidine (100 mM)..... 5 µl ou 0 µl
- H₂O stérile qsp 100 µl

Enzyme	Fabricant	BSA (1 mg/ml)	Spermidine (100 mM)	Température d'hydrolyse
Asc I	Biolabs	-	-	37°C
Bss HII	Boehringer	-	-	50°C
Cla I	Boehringer	-	-	37°C
Ksp I	Boehringer	-	-	37°C
Mlu I	Boehringer	+	+	37°C
Nar I	Boehringer	-	-	37°C
Not I	Boehringer	+	+	37°C
Nru I	Gibco-BRL	+	+	37°C
Sac II	Biolabs	-	-	37°C
Sal I	Boehringer	-	-	37°C
Sfi I	Boehringer	+	-	50°C
Spe I	Boehringer	-	-	37°C
Ssp I	Boehringer	-	-	37°C
Swa I	Boehringer	-	-	37°C

Le mélange est incubé au bain-marie, à la température d'hydrolyse requise pendant 4 h. Pour obtenir une digestion totale, on effectue alors une recharge d'enzyme (à nouveau 5 U/ μ g d'ADN) et on laisse à incuber toute la nuit à température de digestion. Dans le cas d'une double digestion, les blocs sont ensuite équilibrés dans le tampon d'hydrolyse correspondant à la deuxième enzyme.

La réaction est arrêtée par lavage des blocs dans 1 ml de TE 1X. Les blocs sont équilibrés dans le tampon d'électrophorèse à +4°C, puis déposés directement sur le gel.

C. ELECTROPHORESE EN CHAMPS PULSES : TECHNIQUE CHEF (Contour-Clamped Homogeneous Electric Field)

1) Appareillage

L'appareil utilisé a été conçu à l'Unité INSERM n°124, il est composé de :

- un générateur Apex capable de débiter un courant stable d'une tension de 0 à 500 volts et d'une intensité de 0 à 500 mA;
- un pulseur (Multicomat) qui alterne régulièrement deux champs électriques de directions différentes;

- un microprocesseur (conçu à l'Unité INSERM n°124), qui permet la programmation des paramètres de migration;
- une cuve;
- un système de refroidissement associé à une pompe rotative, permettant la circulation permanente du tampon d'électrophorèse afin de maintenir une homogénéité ionique et une température constante (12°C).

2) Les paramètres d'électrophorèse

Paramètres d'électrophorèse		
GEL	Dimension	17 × 17 cm
	Pourcentage d'agarose	1%
	Tampon	0.25X TBE
TAMPON	Composition	0.25X TBE
	Température	12°C
	Débit de circulation	60 l/h
ELECTROPHORESE	Voltage	190 V
	Ampérage	150 mA
	Durée	39 à 48 h

Sachant que la séparation optimale des molécules d'ADN est fonction du temps de pulse et voulant séparer au mieux, d'une part des molécules allant de 200 à 600 kb et d'autre part des molécules plus petites comprises entre 20 et 200 kb, nous avons été amenés à utiliser deux conditions différentes d'électrophorèse :

200 à 600 Kb		20 à 200 Kb	
3 temps de pulse	50 s pendant 16 h	2 temps de pulse	50 s pendant 15 h
	30 s pendant 16 h		30 s pendant 24 h
	80 s pendant 16 h		

D. LES MARQUEURS DE TAILLE

Les marqueurs de taille utilisés sont le "13i" un clone phagique (cf. Technique du Southern blot) et des concatémères d'ADN du phage λ EMBL4. La concatémérisation du phage λ a été mise au point par J.P. Kerckaert (Unité INSERM n°124); elle s'effectue par ligation du phage λ EMBL4 sur lui-même, à 37°C dans de l'agarose LMP.

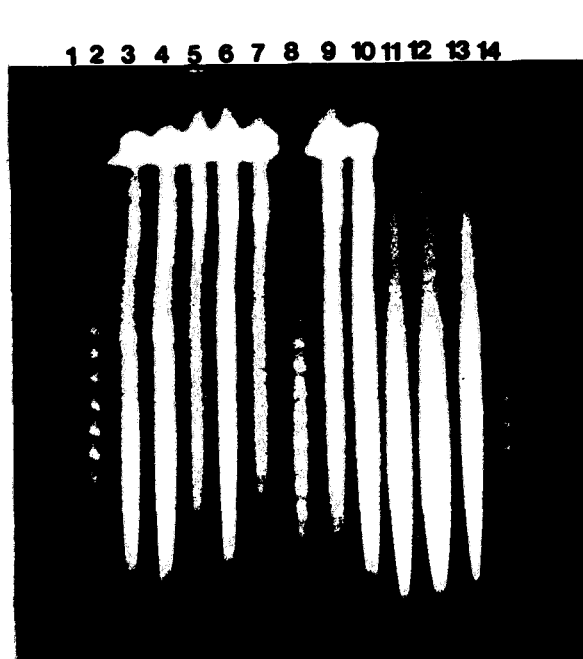


Figure 58 : Electrophorèse en champ pulsé d'ADN de la lignée 194 hydrolysé par différentes enzymes. 1 : marqueur (13i); 2, 8, 14 : marqueur, concatémère de λ EMBL4 (43.2 kb); 3 : Bss HII; 4 : Bss HII / Not I; 5 : Not I; 6 : Not I / Mlu I; 7 : Mlu I; 9 : Nru I; 10 : Not I / Nru I; 11 : Nar I; 12 : Cla I; 13 : Sal I. Conditions d'électrophorèse : 50 s - 16 h; 30 s - 16 h; 80 s - 16 h.

E. TRANSFERT, HYBRIDATION ET AUTORADIOGRAPHIE

Après électrophorèse le gel est coloré au BEt (4 μ g/ml) durant 15 mn, décoloré dans de l'eau une quarantaine de minutes puis photographié sous lampe UV. L'ADN est ensuite dépuriné par traitement du gel dans de l'HCl 0.25 N pendant 15 mn, puis dénaturé 2 \times 30 mn dans une solution dénaturante (NaCl 1.5 M, NaOH 0.5 N) et neutralisé 2 \times 30 mn dans une solution de neutralisation (Tris-HCl 0.5 M pH 7.5, NaCl 3 M).

L'ADN est transféré sur membrane Hybond N⁺ (Amersham) par la technique classique de Southern en tampon SSC 20X pendant 24 h. Le matériel est ensuite fixé sur la membrane de nylon par une exposition de 3 mn aux U.V..

La membrane est rincée dans du tampon SSC 6X, puis incubée dans du tampon de préhybridation SSC 6X - Denhardt's 5X - SDS 0.5% (à raison de 50 µl/cm² de membrane) au moins 1h à 65°C. Après cette première incubation, on remplace le tampon par du tampon d'hybridation SSC 6X - Denhardt's 5X - SDS 0.5% - sulfate de dextran 10% (p/v) - ssDNA (10 mg/ml) 7.5% (v/v) et on ajoute les sondes marquées au ³²P :

- 400 à 500 000 cpm/blot pour la sonde λ EMBL4
- 20 × 10⁶ cpm/blot pour la sonde étudiée.

L'hybridation est réalisée par une incubation d'une nuit à 65°C.

L'excès de sonde est éliminé par plusieurs lavages réalisés comme suit :

- rinçage dans du SSC 2X;
- lavage : 2 × 15 mn dans du tampon SSC 2X à 65°C
ou 2 × 15 mn dans du tampon SSC 2X - 0.1% SDS à 65°C
ou 1 × 15 mn dans du tampon SSC 0.1X à 65°C
- rinçage dans du SSC 2X.

La membrane ainsi lavée est mise en autoradiographie à -80°C pendant des temps variables pouvant aller de 6 h à 8 j.

IV. GENOTHEQUES EN VECTEUR λEMBL4

A. CONSTRUCTION DES GENOTHEQUES

Ne connaissant pas la taille du ou des gènes à cloner, nous avons tout d'abord choisi une méthode classique en vecteur lambdaïdoïde EMBL4 qui permet de cloner des fragments d'une vingtaine de kilobases. La construction de la banque s'effectue en plusieurs étapes dont la première consiste en la préparation du vecteur.

1) Préparation du phage λ EMBL4

On réalise tout d'abord une ligation des ADN phagiques formant ainsi des concatémères qui pourront être facilement séparés du stuffer. La ligation est effectuée sur 300 μ g d'ADN phagique sous un volume de 1ml avec 20 U de T4 DNA ligase (Boehringer) et 100 μ l de tampon de ligation 10X. Le mélange est incubé 2 heures à 37°C.

L'action de la ligase est stoppée par l'addition de 100 μ l de TpB (Boehringer) et par l'incubation du tube pendant 7 mn à 68°C. Afin de libérer le stuffer, partie non indispensable à la réplication du phage qui sera remplacée par l'insert génomique, nous réalisons une hydrolyse par Bam HI (1 000 U) 1 à 2h à 37°C.

L'enzyme est éliminée par extraction au phénol/chloroforme.

L'ADN phagique purifié est précipité par 1/10^e du volume d'acétate de sodium 2 M pH 5.5 et 2 volumes d'alcool absolu et enroulé autour d'une baguette de verre. Il est ensuite dissous dans 500 μ l de TE 1X et conservé à 4°C.

Par sécurité, on contrôle la coupure par électrophorèse en minigel d'agarose en tampon borate.

2) Hydrolyse partielle de l'ADN par Sau 3A

Les fragments d'ADN génomique d'une vingtaine de kb destinés à remplacer le stuffer dans le phage recombinant sont obtenus par une hydrolyse partielle de l'ADN génomique.

Pour cela, il est nécessaire d'utiliser une enzyme qui coupe à haute fréquence, c'est à dire une enzyme qui reconnaît 4 bases, dans des conditions permettant des recouvrements de séquences. Il ne faut pas 100% de coupure par site. Nous avons choisi Sau 3A (5'GATC3') complémentaire à Bam HI (5'GGATCC3').

Une cinétique d'hydrolyse a été réalisée au préalable pour définir avec précision les conditions optimales d'hydrolyse. Ces conditions sont obtenues pour une concentration enzymatique de 0.04 U par μ g d'ADN et des temps d'hydrolyse de 5 et 15 mn.

L'hydrolyse est réalisée sur 2 tubes ; chaque tube contenant :

75 µg d'ADN :	106,4 µl
TpA 10X (Boehringer) :	25 µl
Sau 3A :	3 U
TE 1X :	58,6 µl

L'un est incubé 5 mn à 37°C, l'autre 15 mn à 37°C. Les deux tubes sont mélangés et l'ADN hydrolysé est ensuite précipité par 1/10^e du volume d'acétate de sodium 2 M pH 5.5 et 2 volumes d'alcool absolu. Le précipité est dissous dans 500 µl de TE.

3) Purification des fragments de 20 kilobases d'ADN génomique et des bras du phage

Les fragments d'ADN génomique de 20 kilobases et les bras du phage sont, par la suite, séparés par centrifugation en gradient de sucrose 10-40% réalisé dans le tampon : Tris - HCl 20 mM pH 8.0, EDTA 5 mM, NaCl 1 M stérilisé par filtration.

La centrifugation est effectuée pendant 16 h à 24 000 trs/mn dans un rotor SW 28 sur une centrifugeuse L8-80 Beckman.

Le gradient est "démonté" par fraction de 1 ml. Puis on recherche par électrophorèse en gel d'agarose 0.5% (tampon borate) les fractions qui contiennent les fragments d'ADN de 20 kb et les fractions contenant les bras du phage.

Les fractions intéressantes sont diluées au 1/2 dans du TE puis précipitées par l'alcool absolu en milieu acétate de sodium 0.2 M et centrifugées 30 minutes à 10 000 trs/mn. Le culot d'ADN phagique et le culot d'ADN génomique sont repris par 20 µl de TE 1X.

4) Quantification de l'ADN phagique et des fragments d'ADN génomique d'une vingtaine de kilobases

Afin de réaliser la ligation dans un rapport phage-insert adéquat (4/1), il est nécessaire de doser au préalable les fragments d'ADN et les bras du phage. Cette quantification est effectuée par électrophorèse en gel d'agarose 1% par rapport aux fragments dosés d'un phage témoin (13i).

5) Ligation

Le rapport quantitatif optimum pour réaliser la ligation est : phage/ADN = 4/1. 480 ng de fragments d'ADN génomique sont donc "ligaturés" à 1920 ng d'ADN phagique par incubation à 16°C pendant une nuit avec de la T4 DNA ligase et du tampon de ligation 10X (Boehringer).

6) Packaging

Cette opération consiste à introduire l'ADN recombinant dans un environnement protéique par l'utilisation du kit Packagène (Proméga) permettant ainsi de reformer des particules phagiques. Chaque tube de "packaging" est ensuite dilué par 500 µl de TMG et 60 µl de chloroforme.

7) Infection des bactéries DP50 P₂

La banque de gènes est ensuite amplifiée :

la souche bactérienne utilisée est E.coli DP50 P₂. Cette souche possède un prophage P₂ ce qui lui confère la propriété de sélectionner spécifiquement les phages recombinants.

L'infection est réalisée en plusieurs étapes :

Une colonie DP50 P₂ est mise en culture dans 5 ml de LB-DT pendant une nuit. Puis une goutte de cette pousse est remise à pousser dans 5 ml de LB-DT pour augmenter le titre bactérien pendant 4 h.

Une série de dilutions successives de cette solution phagique servira à établir le titre de la génothèque. Dans des tubes à hémolyse en verre, on réalise le contact de 100 µl de chaque dilution de la solution phagique avec 200 µl de culture bactérienne en présence de MgSO₄ 10 mM, 15 mn à 37°C.

Puis on ajoute 3 ml de Top Agar 0.5% , MgSO₄ 10 mM, DT 1 mM dans chaque tube. Le tout est étalé sur des boîtes de Pétri LB-DT qui sont placées à l'étuve, une nuit, à 37°C.

Le lendemain, on observe les plages de lyse qui contiennent chacune environ 10^6 phages recombinants. Ceux-ci sont alors récupérés en grattant la couche superficielle de Top Agar et resuspendus dans 1 à 2 ml de LB en présence de $1/10^6$ du volume de chloroforme qui tue les bactéries.

Ce mélange est centrifugé 10 mn à 7 000 tours/minute. Le surnageant qui constitue notre banque est conservé à 4°C avec quelques gouttes de chloroforme.

Les clones intéressants seront sélectionnés parmi l'ensemble des clones de la banque par une ou plusieurs sondes nucléiques, c'est ce que l'on appelle le criblage.

B. CRIBLAGE DES BANQUES EN VECTEUR λ EMBL4

Les phages recombinants ayant déjà été sélectionnés par la souche bactérienne DP50 P₂, on utilisera pour le criblage une souche DP50, facile à faire croître, ne possédant pas le prophage P₂. Avant l'infection, on effectue plusieurs relances de la souche afin d'obtenir un titre bactérien suffisant.

Le contact entre solution bactérienne et dilution phagique est réalisé comme précédemment décrit après avoir au préalable déterminé la dilution phagique permettant d'observer 1 000 à 2 000 plages de lyse par boîte.

L'ADN phagique est ensuite transféré sur une rondelle de nitrocellulose déposée à la surface de chaque boîte et laissée en contact 30 secondes. Les rondelles sont alors retirées délicatement et traitées comme suit :

- contact 2 mn avec du tampon de dénaturation phagique (NaOH 0.2 N - NaCl 1.5 M) ;
- contact 4 mn avec du tampon de neutralisation (Tris - HCl 0.5 M pH 7.5 - NaCl 3 M) ;
- cuisson 2 h à 80°C sous vide afin de fixer l'ADN sur la membrane ;
- incubation au moins 2 h dans du tampon de préhybridation 3X ;
- incubation au moins 4 h avec la sonde nucléique dans le même tampon ;
- lavages ;
- révélation par autoradiographie.

C. PREPARATION DES PHAGES RECOMBINANTS

Les clones révélés positifs par l'autoradiographie sont prélevés. Et leur titre phagique est augmenté en milieu liquide par différentes étapes :

1) *Préministock*

Un clone positif prélevé de la boîte de Petri à la micropipette sous forme d'une carotte de milieu est mélangé à 200 μ l de culture bactérienne DP50, en présence de $MgSO_4$ 10 mM.

Après 15 mn de contact à 37°C sans agitation, le mélange est versé dans un flacon contenant 5 ml de LB-DT- $MgSO_4$ et placé sous agitation à 37°C jusqu'à l'obtention de la lyse.

1/100^e du volume de chloroforme est ajouté et laissé en contact 15 mn. Le mélange est centrifugé 10 mn à 7 000 tours/mn. Le surnageant est récupéré et constitue notre préministock auquel sont ajoutées quelques gouttes de chloroforme.

2) *Ministock*

Dans un flacon de 125 ml sont ajoutés :

2 ml de LB-DT- $MgSO_4$

50 μ l de préministock

2 ml de la culture DP50 contenant du $MgSO_4$ 10 mM.

Après 15 mn de contact à 37°C, 16 ml de LB-DT- $MgSO_4$ sont ajoutés. Le mélange est placé à 37°C sous agitation jusqu'à l'obtention de la lyse.

Puis du chloroforme est ajouté au 1/100^e et laissé en contact 15 mn. Le surnageant obtenu après 10 mn de centrifugation à 7 000 tours/mn est conservé avec quelques gouttes de chloroforme et constitue notre ministock de phages.

Avant de réaliser l'amplification proprement dite et la purification de l'ADN phagique, on effectue ce que l'on appelle : "*une préparation rapide de l'ADN phagique*" afin de vérifier la qualité des phages recombinants.

3) Préparation rapide de l'ADN phagique

A 3 ml de ministock sont ajoutés 150 µl de DNase I (1 mg/ml dans NaCl 0.15 M, glycérol 50%) et 10 µl de RNase à 10 mg/ml (en tampon Tris - HCl pH 7.5, NaCl 0.015 M). Après chaque addition, il est nécessaire d'agiter légèrement. Puis, le tout est incubé à 37°C pendant 30 mn.

Au mélange précédent sont ajoutés :

65 µl de Tris - HCl 1 M pH 8.0

130 µl d'EDTA 0.5 M pH 8.0

17 µl de protéinase K (20 mg/ml)

le tout est incubé 5 mn à 37°C. Puis 330 µl de Sarcosyl à 5% sont ajoutés au mélange qui est alors incubé 30 mn à 37°C.

Une purification de l'ADN phagique est réalisée par extraction au phénol/chloroforme. Le surnageant contenant l'ADN phagique est récupéré puis précipité par 1/10^e du volume d'acétate de sodium 2 M pH 5.5 et 2 volumes d'alcool absolu à -20°C et enfin centrifugé 30 minutes à 10 000 tours/mn. Le culot est repris par 50 µl à 100 µl de TE 1X. La solution est dialysée 2 heures sur une membrane Millipore de type VS et de porosité 0.025 µm, contre du TE 1X.

L'ADN phagique ainsi purifié est hydrolysé par les enzymes : Eco RI, Bam HI et Eco RI / Bam HI. Les fragments sont ensuite séparés par électrophorèse en gel d'agarose 1%. A partir de ces gels sont réalisés des Southern blots qui permettent de vérifier la présence de l'insert dans le phage recombinant.

4) Amplification et extraction de l'ADN recombinant

Elle est réalisée dans les conditions déjà décrites précédemment, sous un volume de 500 ml à partir du ministock.

La purification des phages demande les étapes supplémentaires suivantes :

25 g de NaCl et 50 g de PEG sont ajoutés. Le mélange est laissé une nuit à 4°C pour permettre la précipitation des phages. Il est ensuite centrifugé 15 mn à 8 000 trs/mn.

Le culot est repris par 8 ml de TMN. Puis 8 ml de chloroforme sont ajoutés au mélange qui est centrifugé 10 mn à 5 000 trs/mn.

Le surnageant (à raison de 9,4 g par tube) est additionné de chlorure de césium (7 g par tube) puis est ultracentrifugé 2 heures à 4°C, à 63 000 tours/mn (VTi 80/L8-80 Beckman).

L'ADN phagique est prélevé à la seringue. Le chlorure de césium est éliminé par des dialyses successives :

- 2 fois 30 mn contre du Tris 20 mM, EDTA 10 mM, NaCl 100 mM
- 2 fois 30 mn contre du Tris 20 mM, EDTA 1 mM, NaCl 100 mM.

25 µl de protéinase K à 20 mg/ml sont ajoutés au dialysat ainsi que 63 µl de SDS 10% afin de libérer l'ADN. Après incubation de 2 h à 37°C l'ADN est purifié par extraction au phénol/chloroforme, puis précipité par 1/10^e du volume d'acétate de sodium 2 M pH 5.5 et 2 volumes d'alcool absolu à -20°C. Il est alors enroulé sur une baguette de verre et rincé à l'alcool à 70%. L'ADN est ensuite dissout dans du TE 1X, pendant au moins 24 heures et dosé au spectrophotomètre.

L'ADN recombinant ainsi obtenu, purifié en grande quantité peut alors être analysé.

D. ETABLISSEMENT DES CARTES DE RESTRICTION DES PHAGES RECOMBINANTS

La réalisation des cartes de restriction nécessite un choix d'enzymes :

- l'enzyme Eco RI qui permet de libérer l'insert ;
- des enzymes qui coupent dans le bras gauche et/ou dans le bras droit du phage permettant ainsi d'orienter l'insert (Bgl II, Kpn I, Hind III).
- des enzymes ne coupant pas dans le phage (Bam HI, Xba I, Xho I).

Avec ces enzymes nous avons réalisé des hydrolyses simples et doubles afin de localiser avec précision les sites de restriction de l'insert.

Après digestion, les fragments obtenus sont étudiés par électrophorèse en gel d'agarose à 1% en tampon phosphate révélés par le BEt.

Un Southern blot est ensuite réalisé et exposé à la sonde utilisée précédemment pour le criblage. Par comparaison de l'autoradiographie et du gel, on peut ainsi déterminer les fragments hybridant avec la sonde ADNc, c'est à dire les fragments possédant un ou plusieurs exons.

Ces fragments pourront être sous-clonés en vecteur Bluescript pKS pour l'étude de la séquence.

V. GENOTHEQUE EN VECTEUR COSMIDIQUE pWE 15

La banque utilisée est une banque d'ADN génomique placentaire humain en vecteur cosmique pWE 15, commercialisée par Stratagene (Ref. 95 1202). L'ADN génomique a été hydrolysé partiellement par l'enzyme Mbo I de manière à obtenir des fragments de 35-41 kb. Les inserts clonés dans un site Bam HI peuvent être excisés par l'enzyme Not I. Le titre estimé de la banque est de 2.0×10^9 cfu/ml. La bactérie hôte, NM554, a pour génotype : recA1, F⁻, araΔ139, Δ(ara,leu)7696, ΔLacI7A, gal, u⁻, galk⁻, hsr⁻, hsmt, strA, mcrA(-), mcrB(-).

A. CRIBLAGE DE LA BANQUE EN VECTEUR COSMIDIQUE

La génothèque diluée (1/10⁶) est étalée sur des boîtes contenant un milieu LB-Agar Ampicilline 50 µg/ml. Les boîtes sont placées durant une nuit à l'étuve à 37°C.

L'ADN cosmique, tout comme l'ADN phagique, est transféré sur une rondelle de nitrocellulose déposée à la surface de chaque boîte. Après un contact de quelques secondes, la membrane est retirée et traitée comme suit :

- 30 s de contact sur du papier Whatman 3 M imprégné de NaOH 0.5 M;
- 30 s de contact sur du papier Whatman 3 M imprégné de tampon Tris-HCl 1 M pH 7.6;
- 30 s de contact sur du papier Whatman 3 M imprégné de tampon NaCl 1.5 M, Tris-HCl 1 M pH 7.6;
- élimination du surplus bactérien en frottant la membrane dans du tampon NaCl 1.5M, Tris-HCl 1 M pH 7.6;
- fixation de l'ADN sur la nitrocellulose par 2 h de cuisson sous vide à 80°C;
- préhybridation;

- hybridation;
- lavages;
- révélation par autoradiographie.

B. PURIFICATION DE L'ADN DES COSMIDES RECOMBINANTS

Les clones positifs révélés par l'autoradiographie sont prélevés et mis en culture, soit dans 3 ml de LB Ampicilline 50 µg/ml, soit dans 500 ml de LB Ampicilline 50 µg/ml pour obtenir une grande quantité d'ADN cosmétique purifié.

1) Purification de l'ADN cosmétique à partir de 3 ml de culture

3 ml de LB Ampicilline 50 µg/ml sontensemencés avec un clone bactérien positif, et incubés une nuit à 37°C.

Le lendemain, la culture est centrifugée 2 mn à 10 000 tours/mn afin d'obtenir un culot bactérien. Celui-ci est remis en suspension dans une solution froide de glucose 50 mM, Tris-HCl 25 mM pH 8.0, EDTA 10 mM (GTE) et laissé incubé pendant 5 mn à température ambiante.

200 µl d'une solution de NaOH 0.2 N, SDS 1% (préparée extemporanément) sont alors ajoutés, mélangés délicatement par retournement, puis le mélange est laissé 5 mn dans la glace.

150 µl d'une solution froide d'acétate de potassium (pH 4.8) (KAc) sont ajoutés au mélange. Le tout est laissé 5 mn dans la glace.

KAc : Acétate de Potassium 5 M	60 ml
Acide Acétique glacial	11.5 ml
H ₂ O qsp	100 ml

Les débris bactériens sont éliminés par une centrifugation de 10 mn à 10 000 tours/mn. Le surnageant est récupéré dans un tube eppendorf.

2 µl de RNase A (10 mg/ml, préalablement bouillie) sont ajoutés au surnageant; le mélange est incubé 20 mn à 37°C.

L'ADN cosmétique est ensuite purifié par une extraction au phénol/chloroforme, puis précipité par 1/10^e du volume d'acétate de sodium 2 M pH 5.5 et 3 volumes d'alcool absolu.

Après 30 mn de centrifugation à 10 000 tours/mn, le culot d'ADN est rincé par de l'alcool à 70%, séché puis repris par 25-50 µl de TE 1X.

2) Préparation de l'ADN cosmique à partir de 500 ml de culture

Un clone est mis en culture dans 500 ml de LB Ampicilline 50 µg/ml pendant une nuit à 37°C. Le lendemain, la culture est centrifugée 10 mn à 6 000 tours/mn.

Après élimination du surnageant, le culot est repris par 18 ml de GTE froid et stérile. 2 ml de lysozyme (10 mg/ml de GTE) sont ajoutés. Le mélange bien homogénéisé est laissé 10 mn à température ambiante.

60 ml d'une solution de NaOH 0.2 N, SDS 1% (préparée extemporanément) sont ajoutés et mélangés délicatement puis le tout est laissé 10 mn dans la glace; 45 µl de KAc froid sont alors ajoutés au mélange qui est à nouveau placé dans la glace durant 30 mn.

Le mélange est ensuite centrifugé 10 mn à 10 000 tours/mn. Le surnageant est récupéré et filtré stérilement à travers de la laine de verre dans une seringue de 50 ml stérile.

0.6 volume de propane-2-ol est ajouté; le tout est laissé 30 mn dans la glace puis centrifugé 15 mn à 5 000 tours/mn à 4°C.

Après élimination du surnageant, le culot est repris par 6 ml de TE 1X, auxquels sont ajoutés 1/10^e de BEt (1/10 000) et 10 g de CsCl. Les tubes (Quick Seal Beckman) peuvent être éventuellement complétés par une solution à 1 g de CsCl pour 1 ml de TE 1X afin d'atteindre une masse finale de 9.5 à 10 g/tube. Une centrifugation à 68 K (rotor VTI 80) est effectuée pendant 4 h 30 à 16 °C pour isoler l'ADN cosmique.

L'ADN cosmique ainsi isolé est récupéré à l'aide d'une seringue de 1 ml, puis transféré dans un tube Falcon de 50 ml. Le tube est alors complété par du n-butanol saturé en eau (pour 125 ml de n-butanol ajouter 30 ml d'eau stérile, bien mélanger et laisser déphaser) qui va permettre d'éliminer le BEt. Après une agitation douce, la séparation des deux phases est obtenue et la phase organique supérieure contenant le BEt est éliminée. Cette opération est renouvelée afin d'enlever toute trace de BEt.

La phase aqueuse contenant l'ADN est transférée dans un tube Falcon de 15 ml. Un volume égal de propane-2-ol y est ajouté. Le mélange est laissé au moins 30 mn à -20°C puis

est centrifugé à 10 000 tours/mn pendant 30 mn à 4°C. Le surnageant est éliminé et après séchage le culot d'ADN est solubilisé dans du TE 1X.

L'ADN cosmique recombinant ainsi purifié peut alors être cartographié après dosage au spectrophotomètre.

C. ETABLISSEMENT DES CARTES DE RESTRICTION DES COSMIDES RECOMBINANTS

En raison de la taille importante des inserts, la technique classique de cartographie ne peut pas être appliquée dans le cas des cosmides. On utilise la caractéristique qu'ont les cosmides pWE 15 de posséder deux sites Not I qui permettent de libérer l'insert et d'être munis des "primers" T3, T7 de part et d'autre de l'insert.

La technique est la suivante :

- double hydrolyse : Not I totale / enzyme "X" totale
- double hydrolyse : Not I totale / enzyme "X" partielle
- électrophorèse des produits de digestion en gel d'agarose.
- transfert sur membrane
- hybridation avec les sondes oligonucléotidiques T3, T7. Les sondes T3 et T7 permettent d'orienter les fragments de digestion partielle par rapport au vecteur de clonage.

La cartographie de l'insert ainsi réalisée, permet de localiser avec précision les fragments exoniques correspondants à la sonde ADNc utilisée. Les fragments exoniques sont ensuite isolés, sous-clonés en vecteur de séquençage (plasmide Bluescript pKS) et séquencés par la technique de Sanger.

VI. SEQUENCE NUCLEOTIDIQUE

A. EXTRACTION ET PURIFICATION DU FRAGMENT A SEQUENCER

L'établissement de la carte de restriction des inserts génomiques, permet de déterminer le système enzymatique libérant le ou les fragments exoniques que l'on veut séquencer. Après

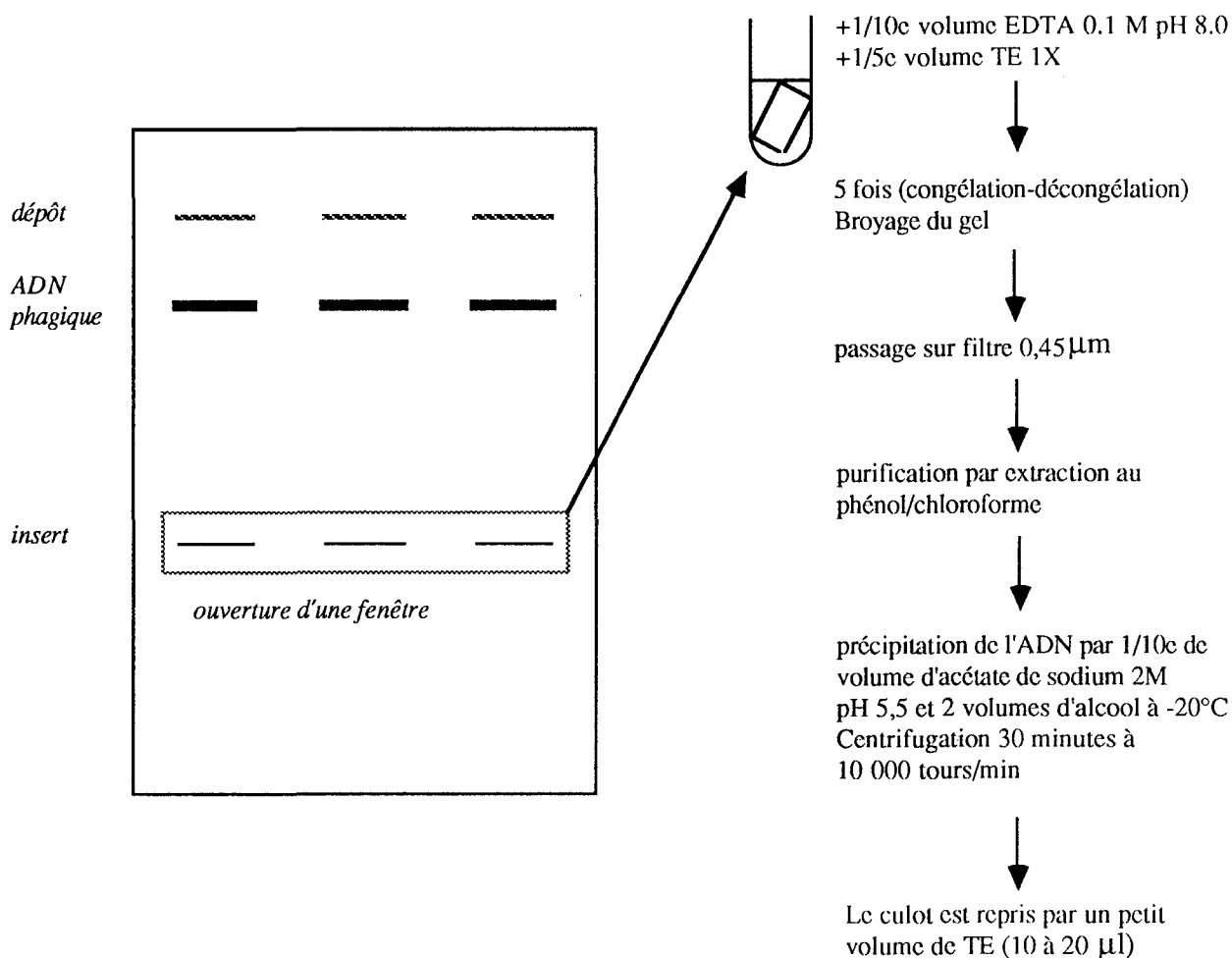
hydrolyse, le fragment est isolé par électrophorèse préparative en gel d'agarose dont le pourcentage dépend de la taille du fragment à purifier dans un tampon TEA et en présence de BEt.

La purification est effectuée suivant différentes techniques :

1) *Centrifugation sur filtre Spin X (Costar - acétate de cellulose 0.45 μm)*

2) *Utilisation du protocole : Prep-A-Gene DNA purification matrix (kit BIORAD)
GeneClean II kit (BIO 101 Inc.)*

3)



B. PREPARATION DU VECTEUR PLASMIDIQUE

Le vecteur plasmidique utilisé pour le clonage des inserts à séquencer est le plasmide Bluescript KS (pBS KS). Il a été choisi pour son bon rendement de préparation et sa simplicité d'utilisation par rapport au vecteur M13.

1) Hydrolyse

Le site de clonage du vecteur pBS est clivé par une ou 2 enzymes de restriction choisies pour générer des extrémités compatibles avec celles du fragment à séquencer. Afin d'éliminer le ou les enzymes utilisées pour la digestion, une extraction au phénol/chloroforme est effectuée, suivie de la précipitation de l'ADN du vecteur. Celui-ci est dosé en gel d'agarose, par comparaison au témoin 13i.

2) Déphosphorylation

Lorsque le vecteur est hydrolysé par une seule enzyme, le taux de ligation du vecteur sur lui-même est important et limite l'obtention de plasmides recombinants. Afin d'obtenir un meilleur rendement, une étape de déphosphorylation de l'extrémité 5'P a été quelquefois effectuée permettant ainsi d'éviter la ligation du vecteur sur lui-même.

Le protocole (Boehringer) utilisé est le suivant :

25 µg de plasmide sont hydrolysés par l'enzyme de clonage choisie, dans des conditions classiques. L'ADN est purifié par une extraction au phénol/chloroforme, une précipitation et un rinçage à l'alcool 70%. Le culot est alors repris par un tampon TE (10 mM pH 8.3) sous un volume de 90 µl.

1/10^e de tampon 10X (Tris-HCl 0.5 M, EDTA 1 mM pH 8.5) (Boehringer) et 1 unité de Phosphatase alcaline sont ajoutés (1 U pour 50 pmoles de 5'P, 2 µg d'ADN plasmidique de 5 kb linéarisé contiennent 1.4 pmoles de 5'P) et incubés 1 h à 37°C.

L'enzyme est inactivée par l'addition d'une solution (EDTA 5 mM, SDS 0.5%) et un passage du mélange réactionnel de 15 mn à 65°C. L'ADN du vecteur ainsi déphosphorylé est purifié classiquement par une extraction au phénol/chloroforme puis précipité par 1/10^e

d'acétate de sodium 2 M et 2.5 volumes d'alcool absolu. Après une centrifugation de 30 mn à 10 000 tours/mn, le culot d'ADN plasmidique est repris par un petit volume de TE 1X puis dosé en gel d'agarose par comparaison au témoin 13i.

C. LIGATION

Tenant compte des masses moléculaires respectives du vecteur (3.1 kb) et du fragment à cloner, de la nécessité de partir d'environ 50 à 100 ng de vecteur digéré, la ligation est réalisée dans un rapport molaire insert/vecteur de 3/1.

2 méthodes ont été utilisées :

- le mélange est incubé sous un volume de 10 μ l une nuit à 14°C :

- pKS
- insert
- Tp de ligation 10X (Boehringer)
- T₄ DNA ligase (0.5 U)
- TE qsp 10 μ l

- le mélange est incubé 2h à température ambiante sous un volume de 20 μ l :

- pKS
- insert
- Tp de ligation 10X (Boehringer)
- PEG 7.5%
- T₄ DNA ligase (0.5 U)
- TE 1X qsp 20 μ l

D. TRANSFORMATION DES BACTERIES E.COLI HB101

La veille de la manipulation, un clone de HB101 est relancé dans 5 ml de LB. Le lendemain 100 μ l de la relance sont remis en culture dans 10 ml de la LB pendant 3 heures à 37°C.

Après quelques minutes dans la glace, cette nouvelle pousse bactérienne est centrifugée 10 mn à 1 500 tours/mn, afin d'éliminer le milieu de culture. Les bactéries du culot obtenu sont rendues compétentes par un contact d'une trentaine de minutes dans la glace avec un tampon

TSS (LB contenant 10% de PEG, 5% de DMSO, un sel de Mg 20 à 50 mM) dans les conditions décrites par Chung et al [226].

La transformation bactérienne est réalisée dans la glace pendant 30 mn, utilisant 100 µl de bactéries compétentes et 2 µl de produit de ligation. 200 µl de SOC, un milieu de culture riche, sont alors ajoutés. Le mélange est incubé à 37°C pendant 1 h puis est mis sous agitation encore 1 h à 37°C. Le milieu de culture est étalé selon les dilutions désirées sur 1 à 3 boîtes de LB-agar ampicilline (25 µg/ml) et incubé une nuit à 37°C.

Les clones recombinants sont contrôlés par une hybridation avec une sonde nucléique, par amplification par PCR ou par boiling.

E. HYBRIDATION A UNE SONDE NUCLEIQUE

Les techniques de prise d'empreinte et d'hybridation sont les mêmes que celles décrites pour le criblage de la banque en vecteur λEMBL4. Seuls varient les temps et les tampons de dénaturation et de neutralisation du matériel nucléique :

Dénaturation : 3 mn dans du tampon (NaCl 1.5 M - NaOH 0.,5 M)

Neutralisation : 3 mn dans du tampon (Tris - HCl 0.5 M pH 7.5 - NaCl 3 M).

F. AMPLIFICATION DES CLONES PAR PCR

Cette technique d'amplification des inserts plasmidiques (pKS) par PCR à partir des primers T3 et T7, est une méthode rapide pour sélectionner les clones recombinants.

La réaction d'amplification est réalisée directement à partir d'une fraction de colonie bactérienne, dans les conditions suivantes :

	Tube expérimental	Tube témoin
Extrait bactérien	prélevé par simple contact	-
Tp Taq 10X	10 µl	10 µl
dNTP	300 µM	300 µM
Primer T3	15 pmoles	15 pmoles
Primer T7	15 pmoles	15 pmoles
H₂O stérile	qsp 100 µl	qsp 100 µl
Taq polymérase (Stratagene)	1.5 U	1.5 U
Huile minérale	80 µl	80 µl

Tampon Taq 10X : Tris-HCl 0.1 M pH 8.8
 KCl 0.5 M
 MgCl₂ 15 mM
 gélatine 0.001%.

Programmation de l'appareil (Techne PHC-2) :

	Température (°C)	Durée (mn)
1 cycle	94	5
	60	1
	72	1
30 cycles	94	1
	60	1
	72	1
1 cycle	94	1
	60	1
	72	5
1 cycle	4	24 heures

Chaque amplifiat est ensuite testé par électrophorèse sur gel d'agarose. On peut ainsi différencier rapidement en fonction de la taille de l'amplifiat, les clones sauvages des clones recombinants. Cette méthode constitue également un gain de temps précieux dans la sélection des clones recombinants issus de la technique de délétion-mutation à l'exonucléase III.

G. BOILING : PREPARATION DE L'ADN DES CLONES RECOMBINANTS

Les clones recombinants sont mis en culture 1 nuit à 37°C dans 5 ml de milieu LB Ampicilline (25 µg/ml).

Le lendemain, 1.4 ml de la culture est centrifugé 5 mn à 10 000 tours/mn. Le surnageant est éliminé par aspiration à la trompe à vide. Le culot est repris par 300 µl de tampon boiling (Sucrose 0.8%, Triton 0.5%, EDTA 50 mM pH 8.0, Tris 10 mM pH 8.0) et 50 µl de lysozyme (solution à 10 mg/ml dans du tampon boiling).

Le mélange est placé dans un bain-marie bouillant pendant 40 s puis centrifugé 15 mn à 10 000 tours/mn. Le culot est éliminé avec un "cure-dents".

Au surnageant est ajouté 1 µl de RNase A (10 mg/ml) préalablement bouillie, le mélange est incubé 20 mn à 37°C. Puis 5 µl de protéinase K sont ajoutés et le mélange est incubé 30 mn à 37°C.

Les protéines sont éliminées par au moins deux extractions au phénol/chloroforme (v/v) suivies de 15 mn de centrifugation et une extraction CHCl₃. L'ADN contenu dans la phase supérieure est précipité par l'addition de 1/10^e de volume en acétate de sodium 2 M pH 5.5 et de 2 volumes d'éthanol absolu. Après 30 mn de centrifugation à 10 000 tours/mn, le culot est rincé à l'éthanol 70 % (v/v) et à nouveau centrifugé 15 mn à 10 000 tours/mn. Le culot est séché puis repris par 25 µl de TE 1X; 1 µl est prélevé et hydrolysé par un couple d'enzymes permettant de sortir l'insert. L'électrophorèse en gel d'agarose permet de vérifier la présence, la taille et la pureté de l'insert. Le plasmide recombinant ainsi purifié, peut alors être séquencé.

H. METHODE DE DELETION-MUTATION A L'EXONUCLEASE III

La limite de résolution de la technique de séquençage est de l'ordre de 300 à 400 bases dans un sens. Les fragments dont la taille est supérieure à 600 ou 800 b, ne peuvent donc pas

être séquencés directement. On a alors recours à une technique de délétion à l'exonucléase III qui permet de diminuer la taille de l'insert par délétions unidirectionnelles (Figure 59).

Une cinétique d'hydrolyse par l'exonucléase III du clone plasmide recombinant permet d'obtenir des fragments plus petits dont la taille est vérifiée par électrophorèse en gel d'agarose. Les fragments de tailles intéressantes seront reliqués pour être séquencés. La compilation des diverses séquences permet de reconstituer la séquence entière de l'insert, même pour des fragments supérieurs à 5 kb. Le protocole décrit par Henikoff [227] en 1984 est constitué des étapes suivantes :

5 à 10 µg (en fonction du nombre de prélèvements) d'ADN recombinant (insert cloné dans un vecteur pKS) sont hydrolysés successivement par 2 enzymes de restriction, entre l'insert et le site d'hybridation du primer. Les deux enzymes utilisées ne doivent pas créer de sites de coupure dans l'insert à séquencer.

La première hydrolyse libère une extrémité 3' sortant, du côté du primer. La deuxième hydrolyse libère une extrémité 5' sortant, sensible à l'exonucléase III du côté de l'insert.

Après chaque hydrolyse l'ADN est purifié par un mélange phénol/chloroforme, puis précipité.

Le culot d'ADN est repris par du tampon Exo III 1X 2.5 µl par "tube temps" + 5 µl et préincubé 5 mn à 37 °C.

Tampon Exo III 10X: Tris-HCl 0.66 M pH 8.0
 MgCl₂ 66 mM

Avant la cinétique d'hydrolyse, 7.5 µl du mélange [tampon S1 + 2.5 U de nucléase S1] qui permettra de stopper la réaction sont déposés dans chaque "tube temps" (t₀ à t₅ ...).

Tampon S1 :

NaCl	500 mM
Acétate de Sodium pH 4.6	2 mM
Glycérol	10%
ZnSO ₄	2 mM

Juste avant d'ajouter l'exonucléase III, 2.5 µl du mélange [ADN + Tp Exo III] sont prélevés et ajoutés au tube t₀.

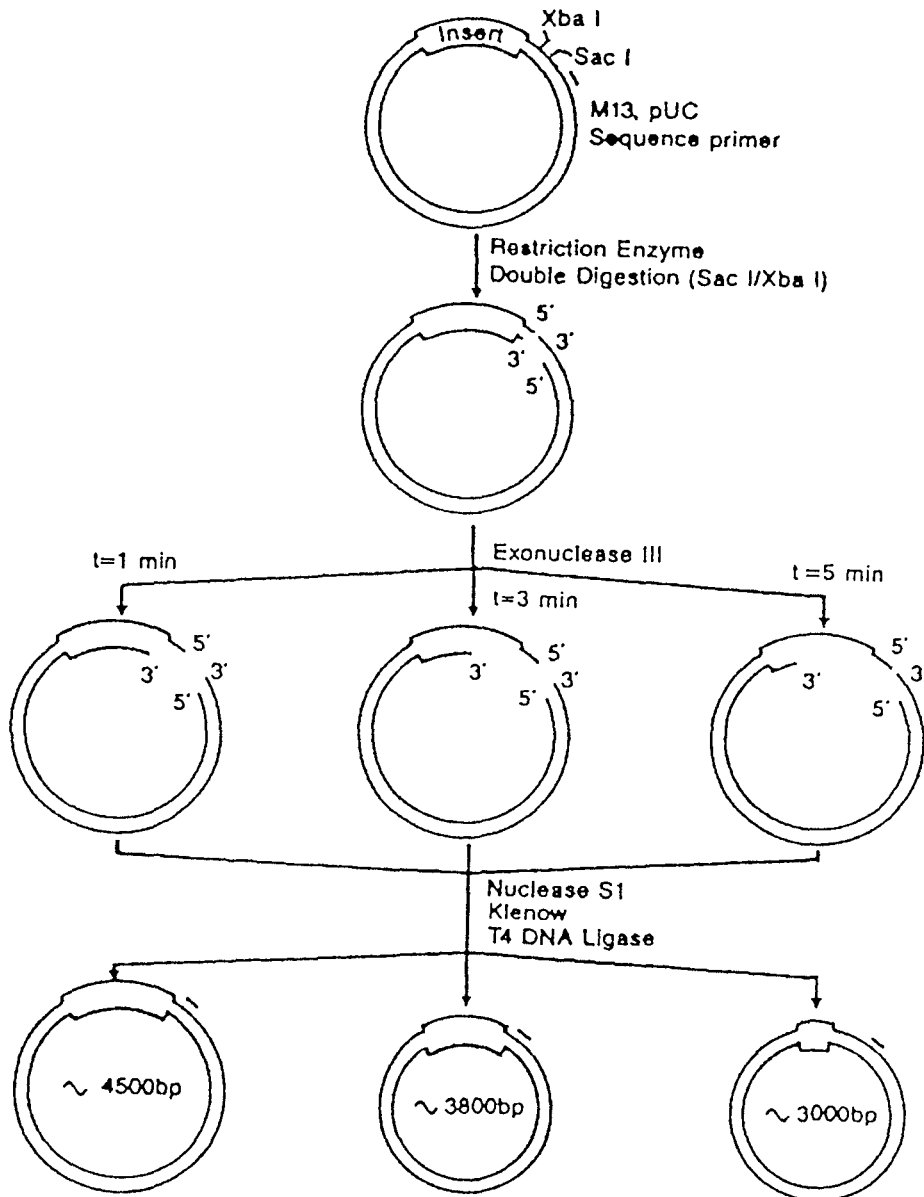


Figure 59 : Principe de la technique de délétion à l'exonucléase III.

La cinétique d'hydrolyse est réalisée par 150 U d'exonucléase III. Toutes les 30 s, 2.5µl du mélange sont prélevés et ajoutés respectivement aux tubes t_1 , t_2 , t_3 ... t_x .

Tous les tubes sont incubés 30 mn à 30°C pour favoriser la digestion de l'ADN monobrin par la nucléase S1. Celle-ci est ensuite inhibée par l'ajout de 1 µl d'une solution "stop S1" (Tris 0.8 M pH 8.0, EDTA 20 mM, MgCl₂ 80 mM) et une incubation de 10 mn à 70°C. Chaque tube est ensuite refroidi dans la glace.

2 µl sont alors prélevés de chaque "tube temps" afin de vérifier par électrophorèse en gel d'agarose la longueur de la délétion obtenue.

Pour compléter les extrémités 5' et 3', 1 µl du mélange [MgCl₂ 0.2 M, Tris 10 mM pH 7.6, 0.5 U de Klenow] est ajouté. Une préincubation de 5 mn à 37°C est effectuée avant d'ajouter à ce mélange 250 pmoles de chacun des 4 nucléotides (dATP, dCTP, dGTP, dTTP).

Après 15 mn à température ambiante, l'ADN est recircularisé par une incubation de 2 h à température ambiante en présence de 40 µl du mélange de ligation : tampon de ligation 10X (Boehringer) + PEG 7.5% + 0.2 U de T4 DNA ligase.

La transformation des bactéries HB101 est réalisée comme décrite précédemment à partir de chaque "tube temps".

I. SEQUENCAGE

L'ADN des plasmides recombinants purifié par la technique du boiling est alors prêt à être séquencé. La technique employée est celle de Sanger [228].

Une étape préalable de dénaturation de l'ADN du plasmide est nécessaire avant d'hybrider l'ADN alors monobrin à une amorce oligonucléotidique (T3, T7 ou primer -40). Puis quatre réactions de polymérisation en parallèle sont réalisées par l'ADN polymérase I (Kit Sequenase Version 2.0, U.S. Biochemical Corporation). Dans chacune d'elles sont ajoutés les déoxynucléotides dont un marqué au ³⁵S et un didéoxynucléotide différent dans chaque réaction qui arrête au hasard la polymérisation. Les fragments de tailles différentes sont séparés par électrophorèse sur gel de polyacrylamide (6%) en tampon urée. Une autoradiographie est alors réalisée, permettant une lecture directe de la séquence. La méthode de séparation électrophorétique est principalement responsable de la limite de séquençage d'un fragment.

LEXIQUE DES FORMULES

LEXIQUE DES FORMULES

TE 10X :	Tris-HCl 1 M pH 8.0 EDTA 0.5 M H ₂ O pH 7.0	10 ml 2 ml qsp 100 ml
STE 10X :	NaCl 3 M Tris-HCl 1 M pH 8.0 EDTA 0.5 M pH 8.0 H ₂ O	166.6 ml 50 ml 10 ml qsp 500 ml
SSC 20X :	NaCl Citrate trisodique H ₂ O pH 7.0	175 g 88.2 g qsp 1 l
SSPE 20X :	NaCl NaH ₂ PO ₄ , H ₂ O EDTA H ₂ O pH 7.4 par NaOH 10 N	174 g 27.6 g 7.4 g qsp 1 l
Eau DEPC :	DEPC H ₂ O	1 ml qsp 1 l
Tampon phosphate 100X :	Na ₂ HPO ₄ Na ₂ H ₂ PO ₄ H ₂ O pH 6.8	96.56 g 44.16 g qsp 1 l
Tampon borate 10X : (TBE)	Acide borique Tris EDTA H ₂ O pH 8.3	55 g 108 g 9.3 g qsp 1 l

Tampon TEA 10X :	Tris	148.4 g
	Acide acétique	11.42 ml
	EDTA 0.5 M	20 ml
	H ₂ O	qsp 1 l
	pH 8.0	
Tampon préhybridation 3X :	formamide désionisée	50 ml
	EDTA 0.5 M	0.3 ml
	Denhardt's 50X	2 ml
	Hépès 500 mM	10 ml
	SSC 20X	15 ml
	ssDNA	2.5 ml (25 mg)
	H ₂ O	qsp 100 ml
Tampon préhybridation 3X - dextran :	Tampon préhybridation 3X	
	Sulfate de dextran 50% (w/v)	22.5 ml
	H ₂ O	qsp 100 ml
Denhardt's 50X :	Ficoll 400	1 g
	Polyvinylpyrrolidone	1 g
	SAB	1 g
	H ₂ O	qsp 100 ml
Tampon de ligation 10X :	Tris-HCl pH 7.6	20 mM
	MgCl ₂	10 mM
	Dithioerythritol	10 mM
	ATP	0.6 mM
TMN 0.5 M:	Tris-HCl pH 7.5	20 mM
	NaCl	0.5 M
	MgSO ₄	10 mM
TMG :	Tris M pH 7.5	5 ml
	MgSO ₄ 10 mM	1 ml
	NaCl 100 mM	10 ml
	gélatine 0.01%	10 mg
LB :	Bactotryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	H ₂ O	qsp 1 l
	pH 7.2	

LB-Agar :	LB contenant 15 g d'Agar pour 1 l ajusté pH 7.2 avec NaOH 3 M	
Top-Agarose :	Agarose LB	800 mg 100 ml
2xTy :	Bactotryptone Yeast extract NaCl H ₂ O pH 7.5	16 g 10 g 5 g qsp 1 l
SOC medium BRL	Bactotryptone (w/v) Yeast extract(w/v) NaCl KCl MgCl ₂ MgSO ₄ Glucose	2% 0.5% 10 mM 25 mM 10 mM 10 mM 20 mM
Tampon GT :	Isothiocyanate de guanidinium Citrate trisodique Sarcosyl H ₂ O stérile DEPC	23.6 g 73.5 mg 250 mg qsp 50 ml 50 µl
MOPS 10X (photosensible) :	MOPS Acétate de sodium 3 M EDTA 0.5 M pH 8.0 Eau DEPC pH 7.0 ajusté par NaOH 3M (~12.8 ml)	16 g 6.72 ml 8 ml 400 ml

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- [1] R. PICOT, I. DAS, L. RIED. Pus, deoxyribonucleic acid and sputum viscosity. *Thorax*. (1978) 33 : 235-242.
- [2] N. HOUDRET, J. M. PERINI, C. GALABERT, A. SCHARFMAN, P. HUMBERT, G. LAMBLIN, P. ROUSSEL. The high lipid content of respiratory mucin in cystic fibrosis is related to infection. *Biochim. Biophys. Acta*. (1986) 880 : 54-61.
- [3] A. SLOMIANY, H. WITAS, M. AONO, B. L. SLOMIANY. Covalently linked fatty acids in gastric mucus glycoprotein of cystic fibrosis patients. *J. Biol. Chem.* (1983) 258 : 8535-8538.
- [4] Mucus et maladies respiratoires. publié par Experta Medica Medical Communications B. V. (An Elsevier Compagny) 1000 VC Amsterdam Pays-Bas.
- [5] G. LAMBLIN, J. J LAFITTE, M. LHERMITTE, P. DEGAND, P. ROUSSEL. Mucins from CF sputum. *Mod. Probl. Paediatr.* (1978) 19 : 153-164.
- [6] R. A. SCHOUMACHER, R. L. SHOEMAKER, D. R. HALM, E. A. TALLANT, R. W. WALLACE et R. A. FRIZELL. Phosphorylation fails to activate chloride channels from cystic fibrosis airways cells. *Nature* (1987) 330 : 752-754
- [7] M. LI, J. D. McCANN, C. M. LIEDTKE, A. C. NAIRN, P. GREENGARD et M. J. WELSH. Cyclic AMP-dependent protein kinase opens chloride channels in normal but not in cystic fibrosis airways epithelium. *Nature* (1988) 331 : 358-60.
- [8] M. P. ANDERSON, D. P. RICH, R. J. GREGORY, A. E. SMITH, M. J. WELSH. Generation of cAMP-activated chloride currents by expression of CFTR. *Science* (1991) 251 : 679.
- [9] M. P. ANDERSON, H. A. BERGER, D. P. RICH, R. J. GREGORY, A. E. SMITH, M. J. WELSH. Nucleoside triphosphates are required to open the CFTR chloride channel. *Cell*. (1991) 67 : 775-784.
- [10] J. BARASH, B. KISS, A. PRINCE, L. SAIMAN, D. GRUENERT, Q. AL-AWQATJ. Deffective acidification of intracellular organelles in cystic fibrosis. *Nature*. (1991) 352 : 70-75.
- [11] G. L. LUKACS, X. B. CHANG, N. KARTNER, O. D. ROTSTEIN, J. R. RIORDAN, GRINSTEINS. The cystic fibrosis transmembrane regulator is present and functional in endosomes. *J. Biol. Chem.* (1992) 267 : 14568-14572.

- [12] **W. V. BOGOMOLETZ, M. I. FILIPE, F. POTET.** Intérêt de l'histochimie des mucines dans le tube digestif normal et pathologique. *Gastroenterol. Clin. Biol.* (1984) 8 : 364.
- [13] **F. YOUNAN, J. P. PEARSON, A. ALLEN, C. W. VENABLES.** Changes in the structure of the mucus gel on the mucosal surface of the stomach in association with peptic ulcer disease. *Gastroenterology* (1982) 82 : 827-831.
- [14] **M. I. FILIPE.** Value of histochemical reactions for muco-substances in the diagnosis of certain pathological conditions of the colon and rectum. *Gut* (1969) 10 : 577-586.
- [15] **M. I. FILIPE.** Mucins in the human gastrointestinal epithelium: a review. *Invest. Cell. Pathol.* (1979) 2 : 195-216.
- [16] **M. I. FILIPE, KB. COOKE.** Changes in the composition of mucin in the mucosa adjacent to carcinoma of the investigation. *J. Clin. Pathol.* (1974) 27 : 315-318.
- [17] **P. E. REID, C. F. A. CULLING, W. L. DUNN, C. W. RAMEY, M. G. CLAY.** Chemical and histochemical studies of normal and diseased human gastrointestinal tract. I-A Comparison between histologically normal colon, colonic tumours, ulcerative colitis and diverticular disease of the colon. *Histochem. J.* (1984), 16 : 235-251.
- [18] **D. K. PODOLSKI, J. K. ISSELBACHER.** Composition of human colonic mucin. Selective alteration in inflammatory bowel disease. *J. Clin. Invest.* (1983), 72 : 142-153.
- [19] **M. EHSANULLAH, M. I. FILIPE, B. GAZZARD.** Mucin secretion in inflammatory bowel disease ; correlation with disease activity and dysplasia. *Gut* (1982), 23 : 485-489.
- [20] **L. R. JACOBS, P. W. HUBER.** Regional distribution and alteration of lectin binding to colorectal mucin in mucosal biopsies from controls and subjects with inflammatory bowel disease. *J. Clin. Invest.* (1985), 75 : 112-118.
- [21] **S. P. LEE, M. C. CAREY, et J. T. LAMONT.** Aspirin prevention of cholesterol gallstone formation in prairie dogs. *Science.* (1981) 211 : 1429.
- [22] **S. P. LEE, J. T. LAMONT, et M. C. CAREY.** The role of gallbladder mucus hypersecretion in the evolution of cholesterol gallstones: studies in the prairie dog. *J. Clin. Invest.* (1981) 67 : 1712.
- [23] **M. ELSTEIN.** Function and physical properties of mucus in the female genital tract. *Brit. Med. Bull.* (1978) 34 : 83-88.
- [24] **M. ELSTEIN.** The cervix and its mucus. *Clin. Obstet. Gynaecol.* (1974) 1 : 345-368.
- [25] **P. L. DEVINE ET I. F. C. MCKENZIE.** Mucins : Structure, Fonction, and Associations with Malignancy. *BioEssays.* (1992) 14(9) : 619-625.
- [26] **D. GAILLARD, C. PLOTKOWSKI et E. PUCHELLE.** Mucus et protection de la muqueuse respiratoire. Edition et communication médicales.

- [27] G. LAMBLIN, M. LHERMITTE, A. KLEIN, N. HOUDRET, A. SCHARFMAN, R. RAMPHAL et P. ROUSSEL. The carbohydrate diversity of human respiratory mucins: a protection of the underlying mucosa? *Am.Rev.Resp.Dis.* (1991) 144 : S19-24.
- [28] S. E. WILLIAMS, L. A. TURNBERG. Retardation of acid diffusion by pig gastric mucus: a potential role in mucosal protection. *Gastroenterology* (1980) 79 : 299-304.
- [29] D. F. HAYES, D. S. SILBERSTEIN, S. W. RODRIQUE et D. W. KUFÉ. DF3 antigen, a human epithelial cell mucin, inhibits adhesion of eosinophils to antibody-coated targets. *J. Immunol.* (1990) 145 : 962.
- [30] J. M. CREETH, K. R. BHASKAR, J. R. HORTON, T. DAS, M. T. LOPEZ-VIDRIER, L. REID. The separation and characterisation of bronchial glycoproteins by density-gradient methods. *Biochem. J.* 167 : 557-569.
- [31] T. MARSCHALL, A. ALLEN. Isolation and characterisation of the high molecular weight glycoproteins from pig colonic mucus. *Biochem. J.* (1978) 173 : 569.
- [32] J. K. SHEEHAN, I. CARLSTEDT. Size heterogeneity of human cervical-mucus glycoproteins. Studies performed with rate-zonal centrifugation and laserlight scattering. *Biochem. J.* (1987) 245 : 757-762.
- [33] D. J. THORNTON, J. R. DAVIES, M. KRAAYENBRINK, P. S. RICHARDSON, J. K. SHEEHAN, I. CARLSTEDT. Mucus glycoproteins from "normal" tracheobronchial secretion. *Biochem. J.* 265 : 179-186.
- [34] A. K. VIRMANI, V. SHANKAR, M. S. GILMORE, D. C. GRAVES, G. P. SACHDEV. Translation of messenger RNA from canine tracheal epithelial cells: identification of mucin core protein. *Am. J. Respir. Cell. Mol. Biol.* (1991) 5 : 149-154.
- [35] J. K. SHEEHAN, R. P. BOOT-HANDFORT, E. CHANTLER, I. CARLSTEDT, D. J. THORNTON. Evidence for shared epitopes within the "naked" protein domains of human mucus glycoproteins. *Biochem. J.* (1991) 274 : 293-296.
- [36] P. ROUSSEL, G. LAMBLIN, M. LHERMITTE, N. HOUDRET, J. J. LAFITTE, J. M. PERINI, A. KLEIN, A. SCHARFMAN. The complexity of mucins. *Biochimie.* (1988) 70 : 1471-1482.
- [37] J. P. AUBERT, N. PORCHET, M. CREPIN, M. DUTERQUE-COQUILLAUD, G. VERGNES, M. MAZZUCA, B. DEBUIRE, D. PETTITPREZ, P. DEGAND. Evidence for different human tracheobronchial mucin peptides deduced from nucleotide cDNA sequences. *Am. J. Resp. Cell. Mol. Biol.* (1991) 5 : 178-85.
- [38] V. L. N. MURTY, F. J. DOWNS, et W. PIGMAN. Rat colonic mucus glycoprotein. *Carbohydr. Res.* (1978) 61 : 139-145.
- [39] H. D. HILL, J. A. REYNOLDS et R. L. HILL. Purification, composition, molecular weight and subunit structure of ovine submaxillary mucin. *J. Biol. Chem.* (1977a) 252 : 3791-3798.

- [40] P. ROUSSEL, G. LAMBLIN, P. DEGAND, E. WALKER-NASIR, R. W. JEANLOZ. Heterogeneity of the carbohydrate chains of sulfated bronchial glycoproteins isolated from a patient suffering from cystic fibrosis. *J. Biol. Chem.* (1975) 250 : 2114-2122.
- [41] C. E. SNYDER, C. E. NADZIEJKO, A. HERP. Isolation of bronchial mucins from cystic fibrosis sputum by use of citraconic anhydride. *Carbohydr. Res.* (1982 b) 105 : 87-93.
- [42] J. J. LAFITTE, G. LAMBLIN, M. LHERMITTE, P. HUMBERT, P. DEGAND, P. ROUSSEL. Etude des glycoprotéines bronchiques humaines de type mucine obtenues par lavage de bronches macroscopiquement saines. *Carbohydr. Res.* (1977) 56 : 383-389.
- [43] G. LAMBLIN, M. LHERMITTE, P. DEGAND, P. ROUSSEL, H. S. SLAYTER. Chemical and physical properties of human mucus glycoproteins. *Biochimie.* (1979) 61 : 23-43.
- [44] A. PRAKOBPHOL, M. J. LEVINE, L. A. TABAK, M. S. REDDY. Purification of low-molecular-weight, mucin-type glycoprotein from human submandibular-sublingual saliva. *Carbohydr. Res.* (1982) 108 : 111-122.
- [45] D. V. GOLD, D. SHOCHAT, F. MILLER. Protease digestion of colonic mucin. Evidence for the existence of two immunochemically distinct mucins. *J. Biol. Chem.* (1981) 256 : 6354-6358.
- [46] E. C. YUREWICZ, K. S. MOGHISSI. Purification of human midcycle cervical mucin and characterization of its oligosaccharides with respect to size composition and microheterogeneity. *J. Biol. Chem.* (1981) 256 : 11895-11904.
- [47] J. T. LAMONT et A. S. VENTOLA. Purification and composition of colonic epithelial mucin. *Biochim. Biophys. Acta* (1980) 626 : 234-243.
- [48] H. D. HILL, M. SCHWYZER, H. M. STEINMAN, R. L. HILL. Ovine submaxillary mucin primary structure and peptides substrates of UDP-N-Acetylgalactosamine-mucine transferase. *J. Biol. Chem.* (1977b) 252 : 3799-3804.
- [49] A. E. ECKHARDT, C. S. TIMPTE, J. L. ABERNETHY, A. TOUMADJE, N. C. Jr JOHNSON, R. L. HILL. Structural properties of porcine submaxillary gland apomucin. *J. Biol. Chem.* (1987) 262 : 11339-11344.
- [50] P. DEGAND, R. HAVEZ, G. BISERTE. Séquence en acides aminés au voisinage du point d'attache d'un glycopeptide de mucine bronchique humaine. *C. R. Acad. Sci.* (1972) 275 : 491-493.
- [51] M. C. ROSE, B. KAUFMAN, B. MARTIN. Proteolytic fragmentation and peptide mapping of human carboxyamino methylated tracheobronchial mucin. *J. Biol. Chem.* (1989) 264 : 8193-8199.
- [52] H. S. SLAYTER, G. LAMBLIN, A. LETREUT, C. GALABERT, N. HOUDRET, P. DEGAND, P. ROUSSEL. Complex structure of human bronchial mucus glycoprotein. *Eur. J. Biochem.* (1984) 142 : 209-218.

- [53] **M. C. ROSE, W. A. VOTER, C. F. BROWN, B. HAUFMAN.** Structural features of tracheobronchial mucin glycoproteins. *Biochem. J.* (1984) 222 : 371-377.
- [54] **A. MIKKELSEN, B. T. STOKKE, B. E. CHRISTENSEN, A. EKGSAETER.** Flexibility and length of human bronchial mucin studied using low-shear viscometry, birefringence, relaxation analysis and electron microscopy. *Biopolymers.* (1985) 24 : 1683-1704.
- [55] **J. K. SHEEHAN, K. OATES, I. CARLSTEDT.** Electron microscopy of cervical, gastric and bronchial mucus glycoproteins. *Biochem. J.* (1986) 239 : 147-53.
- [56] **H. WOODWARD, B. HORSEY, V. P. BHAVANANDAN, E. A. DAVIDSON.** Isolation, purification and properties of respiratory mucus glycoproteins. *Biochemistry* (1982) 21 : 694-701.
- [57] **J. K. SHEEHAN, I. CARLSTEDT.** Electron microscopy of cervical mucus glycoproteins and fragments therefrom. *Biochem. J.* (1990) 265 : 169-178.
- [58] **I. CARLSTEDT, H. LINDGREN, J. K. SHEEHAN.** The macromolecular structure of human cervical mucus glycoproteins. *Biochem. J.* (1987) 213 : 427-435.
- [59] **A. ALLEN, R. H. PAIN, T. R. ROBSON.** Model for structure of the gastric mucous gel. *Nature.* (1976) 264 : 88-89.
- [60] **A. ALLEN.** Structure of gastrointestinal mucus glycoproteins and the viscous and gel-forming properties of mucus. *Brit. Med. Bull.* (1978) 34 : 28-33.
- [61] **A. ALLEN.** Structure and function of gastrointestinal mucus, in physiology of the Gastrointestinal Tract, *Johnson, L.R., Ed., Raven Press, New York* (1981), 617.
- [62] **J. DEKKER, A. VAN DER ENDE, P. H. AELMANS et G. J. STROUS.** Rat gastric mucin is synthesized and secreted exclusively as filamentous oligomers. *Biochem. J.* (1991) 279 : 251-256.
- [63] **J. DEKKER, P. H. AELMANS, G. J. STROUS.** The oligomeric structure of rat and human gastric mucins. *Biochem. J.* (1991) 277 : 423.
- [64] **K. L. CARRAWAY, S. R. HULL.** O-glycosylation pathway for mucin-type glycoproteins. *Bio. Essays.* (1989) 10 : 117-121.
- [65] **J. DEKKER et G. J. STROUS.** Covalent oligomerisation of rat gastric mucin precedes initial O-glycosylation. *J. Biol. Chem.* (1990) 265 : 18116-18122.
- [66] **M. R. NEUTRA, and J. F. FORSTER.** Gastrointestinal mucus: synthesis, secretion and function, in Physiology of Gastrointestinal Tract, *Johnson, L.R., Ed., Raven Press New York*, (1987) 975.
- [67] **N. SHARON, et H. LIS.** Sucres et reconnaissance cellulaire. *Pour la Science.* (1993) 185 : 58-65.

- [68] C. S. TIMPTE, A. E. ECKHARDT, J. L. ABERNETHY, et R. L. HILL. Porcine submaxillary gland apomucin contains tandemly repeated, identical sequences of 81 residues. *J. Biol. Chem.* (1988) 263 : 1081.
- [69] W. PIGMAN, J. MOSCHERA, M. WEIS, et G. TETTAMANTI. The occurrence of repetitive glycopeptide sequences in bovine submaxillary glycoprotein. *Eur. J. Biochem.* (1973) 32 : 148.
- [70] J. R. GUM, J. W. HICKS, R. E. LAGACE, J. C. BYRD, N. W. TORIBARA, B. SIDDIKI, F. J. FEARNEY, D. T. A. LAMPORT, et Y. S. KIM. Molecular cloning of rat intestinal mucin. Lack of conservation between mammalian species. *J. Biol. Chem.* (1991) 266 : 22733-22738.
- [71] S. J. GENDLER, J. M. BURCHELL, T. DUHIG, D. LAMPORT, R. WHITE, M. PARKER, J. TAYLOR-PAPADIMITRIOU. Cloning of partial cDNA encoding differentiation and tumor associated mucin glycoproteins expressed by human mammary epithelium. *Proc. Natl. Sci. USA.* (1988) 84 : 6060-6066.
- [72] D. H. WRESCHNER, M. HAREUVENI, I. TSARFATY, N. SMORODINSKY, J. HOREV, J. ZARETSKY, P. KOTKES, M. WEISS, R. LATHE, A. DION, J. KEYDAR. Human epithelial tumor antigen cDNA sequences : differential splicing may generate multiple protein forms. *Eur. J. Biochem.* (1990) 189 : 463-473.
- [73] J. SIDDIQUI, M. ABE, D. HAYES, E. SHANI, E. YUNIS, D. KUFÉ. Isolation and sequencing of a cDNA coding for the human DF₃ breast carcinoma associated antigen. *Proc. Natl. Acad. Sci. USA.* (1988) 85 : 2320-2323.
- [74] R. L. HILL, A. E. ECKHARDT, Y. WANG, J. E. ABERNETHY, et N. SWAMY. Structure and glycosylation of porcine submaxillary mucin. *2nd International Workshop on Carcinoma-Associated Mucins.* (Aug 1992) Cambridge.
- [75] J. C. PROBST, E. M. GERTZEN, et W. HOFMANN. An integumentary mucin (fimb.1) from *Xenopus laevis* homologous with Von Willebrand factor. *Biochemistry.* (1990) 29 : 6240.
- [76] H. SORIMACHI, Y. EMORI, H. KAWASAKI, K. KITAJIMA, S. INOUE, K. SUZUKI, et Y. INOUE. Molecular cloning and characterisation of cDNAs coding for apopolysialoglycoprotein of rainbow trout eggs. Multiple mRNA species transcribed from multiple genes contain diverged numbers of exact 39-base (13 amino acid) repeats. *J. Biol. Chem.* (1988) 263 : 17678.
- [77] A. O. SPICER, G. PARRY, S. PATTON, et S. J. GENDLER. Molecular cloning and analysis of mouse homologue of tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism. *J. Biol. Chem.* (1991) 266 : 15099-15109.
- [78] N. KILLEEN, A. N. BARCLAY, A. C. WILLIS, et A. F. WILLIAMS. The sequence of rat leucosialin (W3/13 antigen) reveals a molecule with O-linked glycosylation of one third of its extracellular amino acids. *EMBO. J.* (1987) 6 : 4029.

- [79] J. R. GUM, J. C. BYRD, J. W. HICKS, N. W. TORIBARA, D. T. A. LAMPORT, Y. S. KIM. Molecular cloning of human mucin cDNAs. *J. Biol. Chem.* (1989) 264 : 6480-6487.
- [80] J. R. GUM, J. W. HICKS, D. M. SWALLOW, R. L. LAGACE, J. C. BYRD, D. T. A. LAMPORT, B. SIDDIKI, Y. S. KIM. Molecular cloning of cDNA. Derived from a novel human intestinal mucin gene. *Biochem. Biophys. Res. Commu.* (1990) 171 : 407-415.
- [81] N. PORCHET, V. C. NGUYEN, J. DUFOSSE, J. P. AUDIE, V. GUYONNET DUPERAT, M. S. GROSS, C. DENIS, P. DEGAND, A. BERNHEIM, J. P. AUBERT. Molecular cloning and chromosomal localisation of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs. *Biochem. Biophys. Res. Commu.* (1991) 175 : 414-422.
- [82] V. GUYONNET DUPERAT, J. P. AUDIE, V. DEBAILLEUL, A. LAINE, M. P. BUISINE, S. ZOUITINA, A. JANIN, P. DEGAND, J. P. AUBERT, et N. PORCHET. Human mucin genes mapped to 11p15 : mucin cDNAs MUC5A and MUC5C belong to the same gene (MUC5AC) expressed in both human tracheobronchial and gastric mucosae. *Soumis*.
- [83] J. DUFOSSE, N. PORCHET, J. P. AUDIE, V. GUYONNET DUPERAT, A. LAINE, I. VAN SEUNINGEN, S. MARRAKCHI, P. DEGAND, J. P. AUBERT. Degenerate 87-base pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15. *Biochem. J.* (1993) 293 : 329-337.
- [84] N. W. TORIBARA, A. M. ROBERTON, S. B. HO, W. L. KUO, J. W. HICKS, J. R. GUM, J. C. BYRD, B. SIDDIKI, et Y. S. KIM. Human gastric mucin. *J. Biol. Chem.* (1992) 268 : 5879-5885.
- [85] S. J. GENDLER, C. A. LANCASTER, J. TAYLOR-PAPADIMITRIOU, T. DUHIG, N. BURCHELL, L. PEMBERTON, E. LALANI, et D. WILSON. Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* (1990) 265 : 15286-15293.
- [86] M. J. L. LIGTENBERG, H. L. VOS, A. M. C. GENNISSEN, et J. HILKENS. Episialine, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. *J. Biol. Chem.* (1990) 265 : 5573-5578.
- [87] M. S. LAN, S. K. BATRA, W. QI, R. S. METZGAR, et M. A. HOLLINGSWORTH. Cloning and sequencing of human pancreatic tumor mucin cDNA. *J. Biol. Chem.* (1989) 264 : 6480-6487.
- [88] D. M. SWALLOW, S. GENDLER, B. GRIFFITHS, G. CORNEY, J. TAYLOR-PAPADIMITRIOU, M. E. BRAMWELL. The human tumor-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. *Nature.* (1987) 328 : 82-84.
- [89] D. M. SWALLOW, S. GENDLER, B. GRIFFITH, A. KEARNEY, S. POVEY, D. SHEER, J. TAYLOR-PAPADIMITRIOU. The hypervariable gene locus PUM, which codes for the tumor associated epithelial mucins, is located on chromosome 1, within the region 1q21-24. *Ann. Hum. Genet.* (1987) 51 : 289.

- [90] S. J. GENDLER, A. P. SPICER, E. N. LALANI, T. DUHIG, N. PEAT, J. BURCHELL, L. PEMBERTON, M. BOSHELL, et J. TAYLOR-PAPADIMITRIOU. Structure and biology of carcinoma-associated mucin, MUC1. *Am. Rev. Respir. Dis.* (1991) 144 : S42-S47.
- [91] B. GRIFFITHS, D. J. MATTHEWS, L. WEST, J. ATTWOOD, S. M. POVEY, D. SWALLOW, J. R. GUM, et Y. S. KIM. Assignment of polymorphic intestinal mucin gene (MUC2) to chromosome 11p15. *Ann.Hum.Genet.* (1990), 54 : 277-285.
- [92] C. GERARD, R. L. EDDY, T. B. SHOWS. The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat structure. Evidence for common mucin in airway and gastrointestinal tissue. *J. Clin. Invest.* (1990) 86 : 1921-1927.
- [93] B. H. JANY, M. W. GALLUP, P. YAN, J. R. GUM, Y. S. KIM, et C. B. BASBAUM. Human bronchus and intestine express the same mucin gene. *J. Clin. Invest.* (1991) 87 : 77-82.
- [94] N. W. TORIBARA, J. R. GUM, P. J. CULHANE, R. E. LAGACE, J. W. HICKS, G. M. PETERSEN, et Y. S. KIM. MUC2 human small intestinal mucin gene structure. Repeated arrays and polymorphism. *J. Clin. Invest.* (1991) 88 : 1005-1013.
- [95] J. C. BYRD, J. NARDELLI, B. SIDDIQUI, et Y. S. KIM. Isolation and characterisation of a colon cancer mucin from xenografts of LS174T cells. *Cancer Res.* (1998) 48 : 6678-6685.
- [96] J. R. GUM, J. W. HICKS, N. W. TORIBARA, E. M. ROTHE, R. E. LAGACE, et Y. S. KIM. The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. *J. Biol. Chem.*
- [97] M. F. FOX, F. LAHBIB, W. PRATT, J. ATTWOOD, J. GUM, Y. KIM, D.M. SWALLOW. Regional localisation of intestinal mucin gene MUC3 to chromosome 7q22. *Ann.Hum.Genet.* (1992) 281-287.
- [98] G. LAMBLIN, J. P. AUBERT, J. M. PERINI, A. KLEIN, N. PORCHET, P. DEGAND, P. ROUSSEL. Human respiratory mucins. *Eur. Resp. J.* (1992) 5 : 247-256.
- [99] N. PORCHET, J. DUFOSSE, J. P. AUDIE, V. GUYONNET DUPERAT, N. GUYEN VAN CONG, P. DEGAND. et J. P. AUBERT. Structure features of the core proteins of human airway mucins ascertained by cDNA cloning. *Am. Rev. Respir. Dis.* (1991) 144 : S15-S18.
- [100] N. PORCHET, J. DUFOSSE, P. DEGAND, J. P. AUBERT. Les mucines humaines : pourquoi une telle hétérogénéité peptidique? *m/s* (1991) 7 : 1024-30.
- [101] M. CREPIN, N. PORCHET, J. P. AUBERT, P. DEGAND. Diversity of the peptide moiety of human mucins. *Biorheology.* (1990) 27 : 471-484.
- [102] M. S. GROSS, V. GUYONNET DUPERAT, N. PORCHET, A. BERNHEIM, J. P. AUBERT, V. C. NGUYEN. Mucin 4 (MUC4) gene : regional assignment (3q29) and RFLP analysis. *Ann. Genet.* (1992) 35 : 21-26.

- [103] N. VAN CONG, J. P. AUBERT, M. S. GROSS, N. PORCHET, P. DEGAND, J. FREZAL. Assignment of human tracheobronchial mucin gene(s) to 11p15 and a tracheobronchial mucin related sequence to chromosome 13. *Hum. Genet.* (1990) 86 : 167-172.
- [104] J. P. AUDIE, A. JANIN, N. PORCHET, M. C. COPIN, B. GOSSELIN, J. P. AUBERT. Expression of human mucin genes in respiratory, digestive and reproductive tracts ascertained by in situ hybridization. *J. Histochem. Cytochem.* (1993) sous presse.
- [105] B. JOHN, et L. G. MIKLOS. Functional aspects of satellite DNA and heterochromatin. *Int. Rev. Cytol.* (1979) 58 : 1.
- [106] H. S. KIM, et N. MAEDA. Structure of two ,Hae III- type genes in the human salivary proline-rich protein multigene family. *J. Biol. Chem.* (1986) 261 : 6712.
- [107] E. A. AZEN, K. M. LYONS, T. MCGONIGAL, N. L. BARRETT, L. CLEMENTS, N. MAEDA, E. F. VANIN, et O. SMITHIES. Clones from the human gene complex coding for salivary proline-rich proteins. *Proc. Natl. Acad. Sci. USA* (1984) 81 : 5561.
- [108] E. A. AZEN, P. A. GOODMAN, et P. A. LALLEY. Human salivary prolin-rich protein genes on chromosome 12. *Am. J. Hum. Genet.* (1985) 37 : 418.
- [109] N. de CLERCQ, K. HEMSCHOOTE, A. DEVOS, B. PEETERS, W. HEYNS, et W. ROMBAUTS. The 4.4 kilodalton proline-rich polypeptides of the rat ventral prostate are the proteolytic products of a 637 kilodalton displaying highly repetitive sequences and encoded in single exon. *J. Biol. Chem.* (1992) 267 : 9884-9894.
- [110] T. HORI, E. TAKAHASHI, A. TANIGAMI, T. TOKINO, Y. NAKAMURA. A high-resolution cytogenetic map of 168 cosmid DNA markers for human chromosome 11. *Genomics.* (1992) 13 : 129-133.
- [111] D. S. GERHARD, E. LAURENCE, J. WU, N. CHUA, N. MA, S. BLAND, C. JONES. Isolation of 1001 markers from human chromosome 11, excluding the region 11p13-p15.5 and their sublocalisation by new series of radiation-reduced somatic cell hybrids. *Genomics.* (1992) 13 : 1133-1142.
- [112] J. WEISSENBACH, G. GYAPAY, C. DIB, A. VIGNAL, J. MORISSETTE, P. MILLASSEAU, G. VAYSSEIX, et M. LATHROP. A second-generation linkage map of the human genome. *Nature.* (1992) 359 : 794-801.
- [113] C. BELLANNE-CHANTELOT, B. LACROIX, P. OUGEN, A. BILLAULT, S. BEAUFILS, S. BERTRAND, I. GEORGES, F. GILBERT, I. GROS, G. LUCOTTE, L. SUSINI, J. J. CODANI, P. GESNOUIN, S. POOK, G. VAYSSEIX, J. LU-KUO, T. REID, D. WARD, I. CHUMAKOV, D. LE PASLIER, E. BARILLOT, et D. COHEN. Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell.* (1992) 70 : 1059-1068.
- [114] NIH/CEPH Collaborative mapping group. A comprehensive genetic linkage map of the human genome. *Science.* (1992) 258 : 67-86.

- [115] J. C. KAPLAN, et M. DELPECH. Biologie moléculaire et médecine. *m/s Flammarion*.
- [116] S. SACCONI, A. de SARIO, G. DELLA VALLE, et G. BERNADI. The highest gene concentrations in the human genomes are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA*. (1992) 89 : 4913-4917.
- [117] C. JUNIEN et V. Van HEYNINGEN. Report of the committee on the genetic constitution of chromosome 11. *Cytogenet. Cell. Genet.* (1991) 58 : 459-554.
- [118] K. L. O'MALLEY, P. ROTWEIN. Human tyrosine hydroxylase and insulin genes are contiguous on chromosome 11. *Nucleic Acids Res.* (1988) 16 : 4437-4446.
- [119] R. SCHNEIDER, M. J. HIGGINS, D. KIENINGER, E. SCHNEIDER-SCHNERZER, M. HIRSCH-KAUFMANN, M. SCHWEIGER, R. L. EDDY, T. B. SHOWS, B. U. ZABEL. The human ribonuclease/angiogenin inhibitor (RNH) is encoded by a gene mapped to the chromosome 11p15.5 region within 90 kb of the HRAS protooncogene. *Cytogenet. Cell. Genet.* (1991).
- [120] S. M. ZNEIMER, D. CRAWFORD, N. R. SCHNEIDER, et B. BEUTLER. Mapping of human ribonuclease inhibitor gene (RNH) to chromosome 11p15 by in situ hybridization. *Genomics* (1990) 8 (1) : 175-8.
- [121] S. WEREMOWICZ, E. A. FOX, C. C. MORTON, B. L. VALLEE. The placental ribonuclease inhibitor (RNH) gene is located on chromosome subband 11p15.5. *Genomics* (1990) 8 (4) 717-721.
- [122] J. GELERTER, J. L. KENNEDY, H. H. VAN TOL, H. B. NIZNIK, O. CIVELLI, K. K. KIDD. The D4 dopamine receptor (DRD4) maps to distal 11p close to HRAS. *Cytogenet. Cell. Genet.* (1991) [A26951]
- [123] P. J. MacKINNON, B. C. POWELL, G. E. ROGERS, E. G. BAKER, V. J. HYLAND, D. F. CALLEN, G. R. SUTHERLAND. An ultrahigh-sulphur keratine gene of human hair cuticle is located at 11q13 and cross-hybridizes with sequences at 11p15. *Mammalian Genome*. (1991) 1 : 53-56.
- [124] M. DEAN, S. LUCAS-DERSE, S. J. O'BRIEN, D. GOLDMAN. Linkage mapping of six polymorphic lymphocyte proteins detected by two-dimensional gel electrophoresis. *Cytogenet. Cell. Genet.* (1991) [A27181].
- [125] L. DA-VEIGA-PEREIRA, R. J. DESNICK, D. A. ADLER, C. M. DISTECHE, E. H. SCHUCHMAN. Regional assignment of the human acid sphingomyelinase gene (SMPD₁) by PCR analysis of somatic cells hybrids and in situ hybridization to 11p15.1-p15.4. *Genomics* (1991) 9 : 229-234.
- [126] A. PUECH, C. ANTIGNAC, P. COUILLIN, M. JEANPIERRE, L. AHNINE, F. BARICHARD, T. BOEHM, P. AUGEREAU, H. SCRABLE, TH. RABBITS, H. ROCHEFORT, W. CAVENEE, C. JUNIEN. Subregional mapping of BWS, CTSD, MYOD1, and a T-ALL breakpoint in 11p15. *Cytogenet. Cell. Genet.* (1989) 51 : 1013.

- [127] M. STEN-LINDER, A. WEDELL, L. ISELIUS, S. EFENDIC, R. LUFT, et H. LUTHMAN. DNA polymorphisms in the human tyrosine hydroxylase / insulin / insulin-like growth factor II chromosomal region in relation to glucose and insulin responses. *Diabetologia*. (1993) 36 : 25-32.
- [128] NEISWANGER K, SLAUGENHAUPT S.A., HUGUES H.B., FRANK E., FRANKEL D.R., Mc CARTY M.J., CHAKRAVARTI A., ZUBENKO G.S., KUPFER D.J., KAPLAN B.B. Evidence against close linkage of unipolar affective illness to human chromosome 11p markers HRAs1 and INS and chromosome Xq marker DXS52. *Biol. Psychiatry* (1990) 28 : 63-72.
- [129] R. D. CIARANELLO, A. L. CIARANELLO. Genetics of major psychiatric disorders. *Annu. Rev. Med.* (1991) 42 : 151-158.
- [130] C. JUNIEN, I. HENRY. Bras court du chromosome 11 : empreinte parentale différentielle, tumorigenèse et pertes d'allèles. *m/s* (1989) 5 : 480-488.
- [131] S.F. DOWDY, C.L. FASCHING, D. ARAUJO, M. LAIK, E. LIVANOS, B.E. WEISSMAN, E.J. STANBRIDG. Suppression of tumorigenicity in Wilms tumor by the p15.5 p14 region of chromosome 11. *Science* (1991) 254 : 293-295.
- [132] H. R. WEIDEMANN. Tumors and hemihypertrophy associated with Wiedemann-Beckwith syndrome. *Eur. J. Pediatr.* (1983) 141 : 129.
- [133] P. COUILLIN, M. AZOULAY, I. HENRY, N. RAVISE, M. C. GRISARD, C. JEANPIERRE, F. BARICHARD, P. MATEZEAU, J. J. CANDELIER, W. LEWIS. Characterization of a panel of somatic cell hybrids for subregional mapping along 11p et within band 11p13. Subdivision of the WAGR complex region. *Hum. Genet.* (1989) 82 : 171-178.
- [134] A. J. PING, A. E. REEVE, D. J. LAW. Genetic linkage of Beckwith-Weidemann syndrome to 11p15. *Am. J. Hum. Genet.* (1989) 44 : 720-3.
- [135] M. A. MAW, P. E. GRUNDY, L. J. MILLOW, M. R. ECCLES, R. S. DUNN, P. J. SMITH, A. P. FEINBERG, D. J. LAW, M. C. PATERSON, P. E. TELZEROW, D. F. CALLEM, A. D. THOMPSON, R. I. RICHARDS et A. E. REEVE. A third Wilms' tumor locus on chromosome 16q. *Cancer. Res.* (1992) 52 : 3094-3098.
- [136] C.E. SCHWARTZ, D.A. HABER, V.P. STANTON, L.C. STRONG, M.H. SKOLNICK, D.E. HOUSMAN. Familial predisposition to Wilms' tumor does not segregate with the WT1 gene. *Genomics* (1991) 10 : 927-930.
- [137] W. E. Jr LOH, H. J. SCRABLE, E. LIVANOS, M. J. ARBOLEDA, W. K. CAVENEE, M. OSHIMURA, B. E. WEISSMAN. Human chromosome 11 contains two different growth suppressor genes for embryonal rhabdomyosarcoma. *Proc. Natl. Acad. Sci. USA.* (1992) 89 : 1755-1759.
- [138] I. HENRY, M. JEANPIERRE, P. COUILLIN, F. BARICHARD, J. L. SERRE, H. JOURNAL, A. LAMOUREUX, C. TURLEAU, J. de GROUCHY, C. JUNIEN. Molecular definition of the 11p15.5 region involved in Beckwith-Wiedemann syndrome and probably in predisposition to adenocortical carcinoma. *Hum. Genet.* (1989) 81 : 273-277.

- [139] E. R. FEARON, A. P. FEINBERG, S. H. HAMILTON, B. VOGELSTEIN. Loss of genes on the short arm of chromosome 11 in bladder cancer. *Nature*. (1985) 318 : 377-380.
- [140] J.M. HALL, B. HUEZ, J. MORROW , B. NEWMAN, M. LEE, E. JONES, C. CARTER, G. C. BUCHRING, M. C. KING. Rare HRAS alleles and susceptibility to human breast cancer. *Genomics* (1990) 6 : 188-191.
- [141] E. VIEGAS-PEQUIGNOT, A. FLURY-HERARD , H. DE CREMOUX, C. CHLECQ , J. BIGNON, B. DUTRILLAUX. Recurrent chromosome aberrations in human lung squamous cell carcinomas. *Cancer Genet. Cytogenet.* (1991) 49 : 37-49.
- [142] J. WHANG-PENG, T. KNUTSEN, A. GAZDAR, S.M. STEINBERG, H. OIE, A. LINNOIL, J. MUSHINE, M. NAU, J.D. MINNA. Non random structural and numerical chromosome changes in non-small-cell lung cancer. *Genes. chromosom. Cancer* (1991) 3 : 168-188.
- [143] E. A. McGUIRE, R. D. HOCKETT, K. M. POLLOCK, M. F. BARTHOLDI, S. J. O'BRIEN, et S. J. KORSMEYER. The t(11; 14) (p15; q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. *Mol. Cell. Biol.* (1989) 9 : 2124-2132.
- [144] C. L. ROSENBERG, H. G. KIM, T. B. SHOWS, H. M. KRONENBERG, A. ARNOLD. Rearrangement and overexpression of D11S 287E, a candidate oncogene on chromosome 11q13 in benign parathyroid tumors. *Oncogene*. (1991) 6(3) : 449-453.
- [145] A. ARNOLD, H. G. KIM, R. D.GAZ, R. L. EDDY, Y. FUKUSHIMA, M. G. BYERS, T. B. SHOWS, H. M. J. KRONENBERG. Molecular cloning and chromosomal mapping of DNA rearranged with the parathyroid hormone gene in parathyroid adenoma. *Clin. invest.* (1989) 83 : 2034-2040.
- [146] B. ROYER-POKORA, U. LOOS, W. D. LUDWIG. TTG2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t (11; 14) (p13 ; q11). *Oncogene*. (1991) 6(10) : 1887-1893.
- [147] T. BOEHM, L. FORONI, Y. KANEK , M.F. PERUTZ, T. H. RABBITT. The rhombotin family of cysteine-rich LIM-domain oncogenes : distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc. Natl. Acad. Sci. USA* (1991) 88 : 4367-4371.
- [148] J. H. LICHY, W. S. MODI, H. N. SEUANEZ, P. M. HOWLEY. Identification of a human chromosome 11 gene which is differentially regulated in tumorigenic and nontumorigenic somatic cell hybrids of HeLa cells. *Cell. Growth. Differ.* (1992) 3 : 541-548.
- [149] R. A. LOTHE, N. HASTIE, G. A. GEITVIK, S. D. FOSSA, A. L. BERRESEN. Allele losses in testicular cancer detected with 11p markers. *Cytogenet. Cell. Genet* (1991) 10 : 1-9.

- [150] M. P. LEIBOVITCH, V. C. NGUYEN, M. S. GROSS, B. SOLHONNE, S. A. LEIBOVITCH, A. BERNHEIM. The human ASM (adult skeletal muscle) gene : expression and chromosomal assignment to 11p15. *Biochem. Biophys. Res. Commun.* (1991) 180 : 124-150.
- [151] A. BARUCHEL, F. SIGAUX. Hypermethylation of calcitonin gene and leukemia. *Nouv.Rev.Fr.Hematol.* (1991) 33(6) : 551-553.
- [152] B. ROYER-POKORA, S. RAGG, B. HECKLOSTREICHER, M. HELD, U. LOOS, K. CALL, T. GLASER, D. HOUSMAN, G. SAUNDERS, B. ZABEL, B. WILLIAMS, et A. POUSTKA. Direct pulsed field gel electrophoresis of Wilms' tumours shows that DNA deletions in 11p13 are rare. *Genes, Chromosomes Cancer* (1991).3 : 89-100.
- [153] T. SAKAI, J. TOGUCHIDA, N. OHTANI, D. W. YANDELL, J. M. RAPAPORT, T. P. DRYJA. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am. J. Hum. Genet.* (1991) 48 : 880-888.
- [154] C. SAPIENZA. L'empreinte génomique. *Pour la science.* (1990) 158 : 50-56.
- [155] D. A. HARDY, J. I. BELL, E. O. LONG, T. LINDSTEN, et H. O. McDEVITT. Mapping of the class II region of the human major histocompatibility complex by pulsed-field gel electrophoresis. *Nature.* (1986) 323 : 453-455.
- [156] S. K. LAWRENCE, C. L. SMITH, R. SRIVASTAVA, C. R. CANTOR, S. M. WEISSMAN. Megabase-scale mapping of the HLA gene complex by pulsed-field gel electrophoresis. *Science* (1987) 235 : 1387-1390.
- [157] G. J. B. Van OMMEN, J. M. H. VERKERK, M. H. HOFKER, A. P. MONACO, L. M. KUNKEL, P. RAY, R. WORTON, B. WIERINGA, E. BAKKER, et P. L. PEARSON. Aphysical map of 4 million bp around Duchenne muscular dystrophy gene on the human X-chromosome. *Cell.* (1986) 47 : 499-504.
- [158] S. KENWRICK, M. PATTERSON, A. SPEER, K. FISCHBECK, et K. DAVIES. Molecular analysis of the Duchenne muscular dystrophy region using pulsed field gel electrophoresis. *Cell.* (1987) 48 : 351-357.
- [159] C. L. SMITH, J. G. ECONOME, A. SCHUTT, S. KLCO, C. R. CANTOR. A physical map of the *Escherichia coli* K 12 genome. *Science.* (1987) 236 : 1448-1453.
- [160] G. F. CARLE, M. V. OLSON. An electrophoretic karyotype for yeast. *Proc. Natl. Acad. Sci. USA.* (1985) 82 : 3756-3760.
- [161] L. S. LERMAN, M. L. FRISCH. Why does the electrophoretic mobility of DNA in gels vary with the length of molecule ? *Biopolymers.* (1982) 21: 995-997.
- [162] D. C. SCHWARTZ, C. R. CANTOR. Separation of yeast chromosome-sized DNAs by pulsed field gel electrophoresis. *Cell.* (1984) 37 : 67-75.

- [163] **G. F. CARLE, M. V. OLSON.** Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation-gel-electrophoresis. *Nucl. Acids Res.* (1986) 12 : 5647-5664.
- [164] **G. F. CARLE, M. FRANK, M. V. OLSON.** Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science.* (1986) 232 : 65-68.
- [165] **G. CHU, D. VOLLRATH, R. W. DAVIS.** Separation of large DNA molecules by contour clamped homogeneous electric fields. *Science.* (1986) 234 : 1582-1585.
- [166] **M. J. S. DYER, P. FISCHER, E. NACHEVA, W. LABASTIDE, A. KARPAS.** A new human B-cell non-Hodgkin's lymphoma cell line (Karpas 422) exhibiting both t(14; 18) and t(4; 11) chromosomal translocations. *Blood.* (1990) 75 : 709.
- [167] **F. ANTEQUERA, J. BOYES, et A. BIRD.** High levels of de novo methylation and altered chromatin structure at CpG islands. *Cell.* (1990) 62 : 503-514.
- [168] **M. NELSON, et M. McCLELLAND.** Effect of site specific methylation on restriction-modification enzymes. *Nucl. Acids Res.* (1987) 15 : r219-r230.
- [169] **R. ANAND.** Pulsed field gel electrophoresis : a technique for fractionating large DNA molecules. *Trends Genet.* 2 : 278.
- [170] **M. McCLELLAND, et R. IVARIE.** Asymmetric distribution of CpG in an "average" mammalian gene. *Nucl. Acids Res.* (1982) 10 : 7865-7877.
- [171] **A. P. BIRD, M. AGGART, M. FROMMER, O. J. MILLER, D. A. MCLEOD.** Fraction of mouse genome that is derived from island of non-methylated, CpG-rich DNA. *Cell.* (1985) 40 : 91-9.
- [172] **W. R. A. BROWN, A. P. BIRD.** Long-range restriction site mapping of mammalian genomic DNA. *Nature* (1986) 322 : 477-81.
- [173] **W. A. BICKMORE et A. P. BIRD.** Use of restriction endonuclease to detect and isolate genes from mammalian cells. *Meth. in Enzymology.* (1992) 216 : 224-245.
- [174] **B. R. JORDAN.** Ilots HTF : le gène annoncé. *m/s* (1991) 7 : 153-160.
- [175] **S. LINDSAY, A. P. BIRD.** Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* (1987) 327 : 336-8.
- [176] **A. P. BIRD.** CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* (1987) 3 : 342-347.
- [177] **A. P. BIRD.** CpG-rich islands and the function of DNA methylation. *Nature.* (1986) 321 : 206-213.
- [178] **K. GARDINER.** Pulsed field gel electrophoresis and investigations into mammalian genome organisation. *J. Cell. Science.* (1990) 96 : 5-8.

- [179] K. GARDINER, M. HORISBERGER, J. KRAUS, U. TANTRAVAHU, J. KORENBERG, V. RAO, S. REDDY, et D. PATTERSON. Analysis of human chromosome 21 : correlation of physical and cytogenetic maps, gene and CpG island distributions. *EMBO. J.* (1990) 9 : 25-34.
- [180] J. M. ROMMENS, M. C. IANNUZZI, B. S. KEREN, M. L. DRUMM, G. MELMER, M. DEAN, R. ROZMAHEL, J. L. COLE, D. KENNEDY, N. HIDAKA, M. ZSIGA, M. BUCHWALD, J. R. RIORDAN, L. C. TSUI, et F. S. COLLINS. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* (1989) 245 : 1059-65.
- [181] C. A. SARGENT, I. DUNHAM, et R. D. CAMPBELL. Identification of multiple HTF-island associated genes in the human major histocompatibility complex class II region. *EMBO. J.* (1989) 8 : 2305-2312.
- [182] F. LARSEN, G. GUNDERSEN, et H. PRYDZ. Choice of enzymes for mapping based on CpG island in the human genome. *GATA.* (1992) 9 : 80-85.
- [183] W. BICKMORE. Analysis of genomic DNAs by pulsed-field gel electrophoresis. *Techniques for analysis of complex genomes.*
- [184] G. M. WAHL, K. A. LEWIS, J. C. RUIZ, B. ROTHENBERG, J. ZHAO et G. A. EVANS. Cosmid vector for rapid genomic walking, restriction mapping and gene transfert. *Proc. Natl. Acad. Sci. USA.* (1987) 84 : 2160-2064.
- [185] P. REIGMAN, R. J. VLIETSTRA, L. SUURMEIJER, C. B. CLEUTJENS, et J. TRAPMAN. Characterization of kallikrein locus. *Genomics.* (1992) 14(1) : 6-11.
- [186] S. J. O'BRIEN, J. E. WOMACK, L. A. LYONS, K. J. MOORE, N. A. JENKINS, N. G. COPELAND. Anchored reference loci for comparative genome mapping in mammals. *Nature genetics.* (1993) 3 : 103-112.
- [187] P. BUCHER, E. N. TRIFONOV. Compilation and analysis of eucaryotic POL II promoteur sequences. *Nucl. Acids Res.* (1986) 14 : 10009-10026.
- [188] S.M. BERGER. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? *Nature* (1984) 309 : 179-182.
- [189] P. SEAPATHY, M. B. SHAPIRO, et N. L. HARRIS. Splice junctions, branch point sites, and exons : sequence statistics, identification and applications to genome project. *Meth. in Enzymology.* 183 : 253-278.
- [190] A. E. ECKHARDT, C. S. TIMPTE, J. L. ABERNETHY, Y. ZHAO, et R. L. HILL. Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in addition to a highly repetitive, glycosylated domain. *J. Biol. Chem.* (1991) 266 : 9678-9688.
- [191] G. XU, L. J. HUAN, I. A. KHATRI, D. WANG, A. BENNIC, R. E. F. FAHLM, G. G. FORSTNER et J. F. FORSTNER. cDNA for the carboxyl-terminal region of a rat intestinal mucin-like peptide. *J. Biol. Chem.* (1992) 267 : 5401-5407.

- [192] **T. LESUFFLEUR, N. PORCHET, J. P. AUBERT, D. SWALLOW, J. R. GUM, Y. S. KIM, F. X. REAL, et A. ZWEIBAUM.** Differential expression of the human mucin genes MUC1 to MUC5 in relation to growth and differentiation of different mucus-secreting HT-29 all subpopulations. *J. Cell. Science.* (1993) 106 : sous-presse.
- [193] **E. O. LONG, I. B. DAWID.** Repeated genes in eucaryotes. *Annu. Rev. Biochem.* (1980) 49 : 727-764.
- [194] **A. R. WILLIAMSON.** The biological origin of antibody diversity. *Annu. Rev. Biochem.* (1976) 45 : 467-500.
- [195] **G. S. BARSCH, P. H. SEEBURG, R. E. GELINAS.** The human growth hormone gene family : structure and evolution of the chromosomal locus. *Nucleic Acids Res.* (1983) 11 : 3939-3958.
- [196] **M. B. GORIN, D. L. COOPER, F. EIFERMAN, P. Van de RIJN, S. M. TILGHMAN.** The evolution of alpha-fetoprotein and albumin. I. A comparison of primary amino acid sequences of mammalian alpha-fetoprotein and albumin. *J. Biol. Chem.* (1981) 256 : 1954.
- [197] **M. BROWN et J. GOLDSTEIN.** A receptor-mediated pathway for cholesterol homeostasis. *Science.* (1986) 232 : 34.
- [198] **J. WOZNEY, D. HANAHAN, V. TATE, H. BOEDTKER, P. DOTY.** Structure of pro alpha 2 (I) collagen gene. *Nature* (1981) 294 : 129.
- [199] **M. A. T. MUSKAVETCH et D. S. HOGNESS.** An expandable gene that encodes a Drosophila glue protein is not expressed in variants lacking remote upstream sequences. *Cell.* (1982) 29 : 1041.
- [200] **J. R. GUM.** Mucin genes and the proteins they encode : structure, diversity and regulation. *Am. J. Respir. Cell. Mol. Biol.* (1992) 7 : 557-564.
- [201] **H. FLOREY.** The secretion and function of intestinal mucus. *Gastroenterology.* (1962) 43 : 326.
- [202] **S. McQUEEN, D. HUTTON, A. ALLEN, et A. GARNER.** Gastric and duodenal surface mucous gel thickness in rat : effects of prostaglandins and damaging agents. *Am. J. Physiol.* (1983) 245 : G388.
- [203] **A. ZALEWSKY, F. G. MOODY, M. ALLEN, et E. K. DAVIS.** Stimulation of canine gastric mucus secretion with intraarterial acetylcholine chloride. *Gastroenterology.* (1983) 85 : 1067.
- [204] **G. J. STROUS, J. DEKKER.** Mucin-type glycoproteins. *Critical Rev. Biochem. Mol. Biol.* (1992) 27 : 57-96.
- [205] **J. I. REARICK, M. DEAS, et A. M. JETTEN.** Synthesis of mucous glycoproteins by rabbit tracheal cells in vitro. Modulation by substratum, retinoids and cyclic AMP. *Biochem. J.* (1987) 242 : 19.

- [206] F. GROSVELD, G. B. VAN ASSENDELFT, D. GREAVES, G. KOLLIAS. Position-independent, high-level expression of human beta-globin gene transgenic mice. *Cell*. (1987) 51 : 975-985.
- [207] N. DILLON, et F. GROSVELD. Transcriptional regulation of multigene loci : multilevel control. *Trends in Genet.* (1993) Vol 9, 4 : 134-137.
- [208] S. ZOTTER, P. C. HAGEMAN, A. LOSSNITZER, W. J. MOOI, et J. HIGERS. Tissue and tumor distribution of polymorphic epithelial mucin. *Cancer Rev.* (1988) 11-12 : 55-101.
- [209] J. HILKENS, F. BUYS, et M. LIGTENBERG. Complexity of MAM-6, an epithelial sialomucin associated with carcinomas. *Cancer Res.* (1989) 49 : 786.
- [210] A. B. GRIFFITHS, J. BURCHELL, S. GENDLER, A. LEWIS, K. BLIGHT, R. TILLY, et J. TAYLOR-PAPADIMITRIOU. Immunological analysis of mucin molecules expressed by normal and malignant mammary epithelial cells. *Int. J. Cancer.* (1987) 40 : 319.
- [211] J. C. BYRD, J. J. L. HO, D. T. A. LAMPORT, B. SIDDIKI, J. HUANG, P. S. YAN, et Y. S. KIM. Relationship of pancreatic cancer apomucin to mammary and intestinal apomucins. *Cancer Res.* (1991) 51 : 1026-1033.
- [212] P. L. DEVINE, G. T. LAYTON, B. A. CLARK, G. W. BIRREL, B. G. WARD, P. X. XING, et I. F. C. MCKENZIE. Production of MUC1 and MUC2 mucins by human tumor cell lines. *Biochem. Biophys. Res. Commun.* (1991) 178 : 593-599.
- [213] A. ALLEN et S. M. MINNIKIN. The binding of the mucoprotein from gastric mucus cells in tissue culture and the inhibition of cell adhesion. *J. Cell. Sci.* (1975) 17 : 617.
- [214] E. ROOS. Cellular adhesion, invasion and metastasis. *Biochim. Biophys. Acta.* (1984) 738 : 263.
- [215] A. P. SHERBLOM, et C. E. MOODY. Cell surface sialomucin and resistance to natural cell-mediated cytotoxicity of rat mammary tumor ascites cells. *Cancer Res.* (1986) 46 : 4543.
- [216] K. L. CARRAWAY et J. SPIELMAN. Structural and functional aspects of tumor cell sialomucins. *Mol. Cell. Biochem.* (1986) 72 : 109.
- [217] A. E. REEVE, S. A. SIH, A. M. RAIZIS, et A. P. FEINBERG. Loss of allelic heterozygosity at a second locus on chromosome 11 in sporadic Wilm's tumor cells. *Mol. Cell. Biol.* (1989) 9 : 1799-1803.
- [218] M. WAZIRI, S. R. PATIL, J. W. HANSON, et J. A. BARTLEY. Abnormality of chromosome 11 in patients with features of Beckwith-Weidmann syndrome. *J. Pediatr.* (1983) 102 : 873-876.
- [219] C. TURLEAU, J. DE GROUCHY, F. CHAVIN-COLIN, H. MARTELLI, M. VOYER, et R. CHARLES. Trisomy 11p15 and Beckwith-Weidmann syndrome : A report of two cases. *Hum. Genet.* (1984) 67 : 219-221.

- [220] A. KOUFOS, P. GRUNDY, K. MORGAN, A. A. KIRK, T. HADRO, B. C. LAMPKIN, et W. K. CAVENEE. Familial Wiedemann-Beckwith syndrome and second Wilm's tumor locus both map to 11p15.5. *Amer. J. Hum. Genet.* (1989) 44 : 711-719.
- [221] H. J. SCRABLE, D. WITTE, H. SHIMADA, T. SEEMAYER, S. WANG-WUU, S. SOUKUP, A. KOUFOS, P. HOUGHTON, B. LAMPKIN, et W. CAVENEE. Molecular differential pathology of rhabdomyosarcoma. *Genes, Chromosomes Cancer.* (1989) 1 : 23-25.
- [222] I. U. ALI, R. LIDEREAU, C. THEILLET, et R. CALLAHAN. Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia. *Science.* (1987) 238 : 185-188.
- [223] M. M. LEBEAU, T. W. MCKEITHAN, E. A. SHIMA, R. E. GOLDMAN-LEIKIN, S. J. CHAN, I. G. BELL, J. D. ROWLEY, et M. D. DIAZ. T-cell receptor α chain gene is split in human T-cell leukemia cell line with a t(11;14)(p15;q11). *Proc. Natl. Acad. Sci. USA* (1986) 83 : 9744-9748.
- [224] C. M. MORRIS, P. H. FITZGERALD, M. A. KENNEDY, P. E. HOLLINGS, M. GARRY, et G. M. CORBETT. HRAS1 and INS genes are relocated but not structurally altered as a result of the t(7;11)(p15;p15) in a clone from a patient with acute myeloid leukemia (M4). *Brit. J. Hematol.* (1989) 9 : 481-486.
- [225] M. JEANPIERRE. Rapid method for the purification of DNA from blood. *Nucl. Acids Res.* (1987) 15 : 9611.
- [226] C.T. CHUNG, S.L. NIEMELA, R.H. MILLER. One-step preparation of competent Escherichia coli : Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* (1989) 86 : 2172-2175.
- [227] S. HENIKOFF. Unidirectional digestion with exonuclease III. Creates targeted breakpoints for DNA sequencing. *Gene* 28 : 351-359.
- [228] F. SANGER, S. NICKLEN, A. R. COULSON. Sequencing with chain-terminating inhibitors terminating. *Proc. Natl. Sci. USA* (1977) 74 : 5463-5467.

