

UNIVERSITE DE SHERBROOKE

Isolement et caractérisation des gènes d'immunoglobuline Vk-1 de la souche
de souris BALB/c.

© par Daniel Ah-keong Chung Tze Chong

DEPARTEMENT DE BIOCHIMIE

Mémoire présenté à la Faculté de Médecine en vue d'obtenir un diplôme
de Maîtrise ès Sciences (M.S.c)

Juin 1989

TABLE DES MATIERES.

	Page
Table des matières.....	i
Liste des figures	iv
Liste des tableaux.....	viii
Liste des abréviations.....	x
Résumé	xii
1 Introduction.....	1
1.1 Structure d'une immunoglobuline.....	2
1.2 Conformation et structure spatiale.....	3
1.3 Immunogénétique.....	4
1.3.1 Organisation et réarrangement des gènes d'immunoglobuline.....	4
1.3.2 Organisation et structure des gènes d'immunoglobuline	5
a) Les gènes V κ	5
b) Les gènes V λ	6
c) Les gènes V H	6
1.4 Réarrangement des gènes.....	7
1.4.1 Réarrangement des gènes.....	7
1.4.2 Mécanisme de recombinaison.....	8
1.4.3 Modèle de recombinaison	9
1.5 L'expression du gène d'immunoglobuline	10
1.5.1 Le site du promoteur	11
1.6 Génération de la diversité des anticorps	14
1.6.1 L'association combinatoire	14
1.6.2 La diversité au site de jonction	14
1.6.3 La mutation somatique	15
1.6.4 La mobilité des modules.....	16

1.7 Estimation du nombre de segment V	16
1.8 Les pseudogènes	20
1.9 Intérêt de l'analyse des gènes du group Vk-1	22
1.10 Identification des autres gènes étroitement reliés à Vk-1	25
2 Matériel et méthodes	27
2.1 Extraction de l'ADN de rein	27
2.2 Digestions avec des enzymes de restriction	27
2.3 Electrophorèse	28
2.4 Construction d'une bibliothèque	29
2.4.1 Ligation	29
2.4.2 Encapsidation in vitro	30
2.4.3 Titration	30
2.4.4 Amplification de la bibliothèque	31
2.5 Purification de bacteriophage et construction d'un stock de lysat	31
2.6 L'extraction de l'ADN de phage	34
2.7 Prehybridation	36
2.8 Marquage de la sonde	36
2.9 Lavage des filtres	38
2.10 Preparation des cellules XL-1	38
2.11 Clonage dans M13+	40
2.12 La transformation	40
2.13 Sélection de grandeur d'insert par micro-preparation d'ADN	41
2.14 Séquençage	42
2.14.1 Reaction d'amorce sur plasmide M13+	42
2.14.2 Reaction pour la terminaison et le marquage	42
2.14.3 Gel de séquence	43
2.14.4 Lecture et compilation des données	44
2.15 La construction d'une carte de restriction	45
3 Résultats	47
3.1 Stratégie de séquençage	52

3.2 Région séquencée.....	54
3.3 Recherche du leader	57
3.4 Construction d'une carte de restriction.....	58
3.5 Identification des gènes correspondant aux DCDX et BA 43 dans l'ADN génomique.....	66
3.6 Séquences de DCDX et BA 43	72
3.7 Comparaison de BA 43 et DCDX avec d'autres groupes	75
4 Discussion	80
4.1 Identification des gènes obtenus	80
4.2 Sélection négative pour les gènes obtenus.....	82
4.3 La région codante, 5' et 3' du gène.....	84
4.4 Temps de divergence.....	86
4.5 Le rôle du pseudogène.....	88
4.6 Perspectives.....	90
Conclusion.....	92
Remerciements.....	93
Références	94

LISTE DES FIGURES

Figure 1	: Structure d'une immunoglobuline.	2
Figure 2a	: Organisation des gènes d'immunoglobulines chez la souris.....	6
Figure 2b	: Schéma démontrant l'organisation des signaux de reconnaissance.....	9
Figure 3	: Les sites de régulation d'expression dans un gène de Vk.	12
Figure 4	: Schéma démontrant le réarrangement et l'épissage d'un gène Vk.....	13
Figure 5a	: Autoradiogramme de l'ADN génomique de différentes souches de souris digérées Bam HI et Eco RI.(la membrane lavée dans des conditions strigentes.	24
Figure 5b	: Autoradiogramme de l'ADN génomique de différentes souches de souris digérées Bam HI et Eco RI (la membrane lavée dans des conditions peu stringentes).....	24
Figure 5c	: Carte de la sonde PPV101.	24
Figure 6	: Le vecteur Embl4.....	29
Figure 7	: Purification des clones positifs.....	47
Figure 8	: Photo du gel et de l'autoradiogramme des clones EB 14,16,22, 25 et 43 digérés	

	Bam HI.....	48
Figure 9	: Photo démontrant le nombre de sous-clones possédant le fragment 5.4 Kb.....	50
Figure 10	: Digestion des sous-clones avec Xba I et Bgl II.....	50
Figure 11	: Spots-tests pour détecter les bons sous-clones.	51
Figure 12	: Vecteur M13+ montrant les amorces et le fragment DCDX.	52
Figure 13	: Photo d'un gel du clone DCDX digéré avec Pst I, Hinc II, Hind III et Bgl II.....	53
Figure 14	: Carte montrant la stratégie de séquençage et les sous-clones utilisés.	54
Figure 15	: Photo d'un autoradiogramme d'un gel de séquence.	55
Figure 16a	: Gel de la digestion Bam HI et Eco RI des clones EB 14, 25, 43 et 45.	56
Figure 16b	: Autoradiogramme du gel précédent en utilisant une sonde A3.2	56
Figure 16c	: Carte de la sonde A3.2.....	56
Figure 17	: Gel montrant la différence entre la grandeur de DCDX et le 5.3 de BA 43 en les digérant simultanément avec Bam HI et Xba I.	57
Figure 18	: Gel et autoradiogramme de la digestion partielle de EB 14. Sonde utilisée ON-R.	59
Figure 19	: Carte de restriction du clone EB 14 avec les sites Hind III, Eco RI et Bam HI.	59

Figure 20	: Gel de la digestion du clone R26 avec Bam HI, Pst I, Xba I, Hind III, Eco RI + Bgl II, Kpn I, Bam HI1+ Eco RI et Sma I1.....	60
Figure 21	: Gels et autoradiogramme de la digestion Bam HI1, Eco RI, Pst I, Sma I, Bgl II et Hind III. Les sondes utilisées sont DCDX et DCDSHi.....	63
Figure 22	: Carte de restriction de EB 14 en détail.....	65
Figure 23	: La carte des sous-clones DCD et BA 43 montrant les régions utilisées comme sonde.....	66
Figure 24	: Sous-clones DCDX, DCDK, BA 43A et BA 43K utilisés comme sonde sur l'ADN génomique de BALB/c et c.C58.	67
Figure 25	: Autoradiogramme démontrant les fragments de l'ADN génomique des différentes souches de souris qui hybrident avec la sonde DCDX.	68
Figure 26	: Gel et autoradiogramme de la digestion de EB 14 et 43 avec les enzymes Apa I, Kpn I, Hinc II et Hind III.	69
Figure 27	: Autoradiogramme de l'ADN génomique de la souche (gauche) de souris Balbc digéré avec les enzymes Kpn I, Hinc II, Hind III, Apa I, Bam HI et Eco RI hybridé avec la sonde DCDX.....	69
Figure 27	: Autoradiogramme de l'ADN génomique de la souche de (droite) souris BALB/c digéré avec les enzymes Kpn I,	

	Hinc II, Hind III, Apa I, Bam HI et Eco RI hybridé avec la sonde DCDXK.....	71
Figure 28	: Alignement multiple des différents sous-groupes de Vk-1.....	73
Figure 29	: Comparaison de DCDX avec Vk-1C, BA 43 et K18.1 en utilisant le programme Diagonal.	77
Figure 30	: Comparaison des acides aminés des différents groupes de Vk et sous-groupes de Vk-1 avec DCDX et BA 43.....	81
Figure 31	: Représentation schématique du gène DCDX et BA 43 avec les délétions et un gène chimérique à partir de ces deux gènes.....	90

Liste des tableaux

Tableau 1	: Tampons pour différents enzymes.	27
Tableau 2	: Correspondance des différents clones.	49
Tableau 3	: Fragments de DCDX digéré avec Pst I, Hinc II, Hind III et Bgl II.....	54
Tableau 4	: Fragments de la digestion partielle de EB 14 hybridant avec la sonde ON-R.	58
Tableau 5	: Fragments du sous-clone R26 digéré avec Bam HI, Pst I, Xba I, Hind III, Eco RI+Bgl II, Kpn I, Bam HI+Eco RI et Sma I.	61
Tableau 6	: Fragments de EB 14 digéré avec Eco RI1, Bam HI, Pst I, Sma I, Bgl II et Hind III qui hybrident avec les sondes DCDX et DCDSHi.....	62
Tableau 7	: Fragments de l'ADN génomique des différentes souches de souris hybridant avec la sonde DCDX.....	68
Tableau 8	: Fragments de EB 14 et BA 43 digérés Hinc II, Kpn I et Apa I qui hybrident avec la sonde DCDX.	70
Tableau 9	: Fragments de l'ADN de la souche de souris de Balbc digéré avec Hinc II, kpn I et Apa I hybridant avec la sonde DCDX et DCDXK.....	70
Tableau 10	: Séquences des acides aminés du leader des différentes sous-groupes de Vk-1 et BA 43.	79

Tableau 11	: Séquences des acides aminés de la région variable des différentes sous-groupes de V κ -1,BA 43 et DCDX.....	79
Tableau 12	: Pourcentage d'homologie entre DCDX et les différents sous-groupes de V κ -1.....	87
Tableau 13	: Pourcentage de substitution neutre dans la région codante entre DCDX et les autres sous-groupes de V κ -1.....	88
Tableau 14	: Pourcentage d'acides aminés substitués entre DCDX et les autres sous-groupes de V κ -1.....	88

LISTE DES ABREVIATIONS

BPB	: Bleu de Bromophénol.
BSA	: Albumine bovine sérique.
C	: Région constante de la chaîne et du gène d'immunoglobuline.
CDR	: "Complementarity determining region".
cpm	: Comptes par minute.
D	: Régions de diversité des gènes et chaînes lourdes d'immunoglobuline.
DNP	: Dinitrophényl.
DO	: Densité optique
DTT	: Dithiothréitol.
Fc	: Fragment cristalisable.
FR	: "Frame-work".
FTL	: "Freeze-thaw lysate".
GAT	: Glutamoyl-Alanyl-tyrozone
H	: Gènes et chaînes lourdes.
IPTG	: Isopropyl β -D-thiogalactopyranoside
J	: Région 'joining' des gènes et chaînes d'immunoglobuline.
Kb	: Kilobase.
L	: leader.

VL	: Gènes et chaînes légères.
oct	: octamère.
pb	: paire de base.
pd	: Séquence pentadecanucléotidiques
PFU	: Plaque forming units.
REF	: Fragments d'enzymes de restriction
RPM	: Revolutions per minutes.
SDS	: Sodium dodécyl sulfate.
SE	: "Sonicated extract"
SM	: Milieu pour garder les phages viables
SSC	: Solution contenant 0.15 M NaCl et 0.15 M Na citrate.
U	: unité
V	: Région variable des gènes et chaînes d'immunoglobuline.
VL	: Région variable des gènes et chaînes V lambda.
VK	: Région variable des gènes et chaînes V kappa.
VH	: Région variable des gènes et chaînes lourdes.
WT	: Type sauvage.
X-gal	: 5-bromo 4-chloro 3-indolyl β -D galactoside

RESUME

Le but de ce projet consiste à isoler et caractériser des gènes germinaux reliés au groupe Vk-1 des régions variables des chaînes kappa dans la souche de souris BALB/c.

Dans ce projet on essaie de répondre à une question soulevée par la contradiction entre le nombre de gènes encodant le groupe Vk-1 suggéré par les études de "Southern blotting" d'une part et par la comparaison des séquences Vk-1 exprimées, d'autre part.

En effet, la comparaison des séquences des acides aminés de 53 chaînes légères du groupe Vk-1 exprimées dans des immunoglobulines de myélome et hybridome de diverses spécificités, a permis leur classification en 5 sous-groupes: Vk-1A, Vk-1B, Vk-1C, Vk-1D, Vk-1E dont au moins 4 sont présents dans la souche de souris BALB/c. Dans le cas de BALB/c, toutes les séquences décrites jusqu'à date peuvent être codées par seulement 5 gènes germinaux.

D'autre part, les résultats de "Southern blotting" avec les sondes dérivées des gènes Vk-1 montrent deux résultats différents dépendamment des conditions de lavage des filtres. Sur un blot de l'ADN de BALB/c digéré avec Bam HI ,dans des conditions de lavages sévères, la sonde détecte 2 bandes majeures correspondant aux gènes Vk-1C (8.5 Kb) et Vk-1A (6.3 Kb). Par contre avec les lavages dans des conditions plus douces, on remarque environ 8 bandes en plus des 2 bandes mentionnées ci-dessus. On se demande si ces bandes représentent d'autres familles de gènes Vk ou si elles représentent d'autres sous-groupes de Vk-1. Afin de répondre à ces questions, on a procédé au criblage d'une bibliothèque de BALB/c (dont l'ADN génomique a été digéré

partiellement avec Eco R1 et ligué au vecteur EmbI4) dans des conditions d'hybridation et de lavage légères. La sonde utilisée est un fragment d'un allèle de Vk-1C, Vk-1C^f (obtenu d'une souche de souris CE) et est nommée PPV101.

De 21 clones positifs isolés, 9 contenaient le gène Vk-1A, 8 le gène Vk-1C et 4 des nouveaux gènes reliés au groupe Vk-1. L'analyse de ces derniers clones digérés avec Bam HI a montré que tous les quatre contenaient des fragments de 5.4 Kb hybridant avec la sonde PPV101. Les fragments de 5.4 kb (BamHI) ont été sous-clonés dans le phagemide M13+. Une analyse détaillée des 4 sous-clones a révélé que 3 semblaient avoir des inserts identiques et que le quatrième était différent. Deux clones ont été choisis (DCDX et BA43) pour la caractérisation au niveau de la séquence.

La comparaison des séquences des régions V des clones DCDX a montré que les deux gènes étaient de véritables membres du groupe Vk-1 mais qu'ils codent pour des séquences d'acides aminés qui n'ont pas été reportées dans les chaînes kappa exprimées. Ils sont eux-mêmes étroitement reliés (>96% d'homologie de séquence). Ceci indique une duplication très récente comme origine des deux gènes. Malgré le fait que les deux gènes possèdent environ 85% d'homologie avec les autres membres du groupe Vk-1 dans la région codante, l'homologie de séquence s'arrête immédiatement avant et après les régions codantes des gènes.

De plus, l'analyse des séquences a démontré que les deux gènes sont des pseudogènes, chacun possédant une délétion différente et non-chevauchante dans la région 5'. Cette observation suggère que les deux gènes ont subi les délétions après leur duplication. Deux possibilités sont considérées pour

expliquer l'élimination rapide de ces deux gènes en evolution soit (i) qu'il y ait eu une sélection négative contre l'existence de ces gènes (supposant qu'ils codent pour un anticorps spécifique; dangereux pour la survie de l'animal) ou (ii) que la structure ou position des gènes sur le chromosome leur a confié une instabilité.

1 INTRODUCTION

Considérer le système immunitaire comme un mécanisme contribuant à l'homéostasie, c'est-à-dire la capacité de l'organisme de s'ajuster constamment aux modifications de l'environnement et de rétablir l'équilibre, nous aide à mieux le comprendre. Les mécanismes effecteurs de l'immunité sont en effet activés par la pénétration d'un antigène dans l'organisme ce qui déclenche une réaction d'alarme. La destruction de l'antigène assure le retour à l'équilibre. Deux systèmes sont responsables de cela: le système de l'immunité cellulaire et celui de l'immunité humorale. Etant donné que ce projet est axé sur l'immunité humorale et c'est sur cet aspect que nous allons insister dans l'introduction.

On savait déjà que le repertoire des anticorps est très grand (10^8 différents spécificités dans la souris BALB/c). Par ailleurs, on ne connaissait pas bien par quel mécanisme cela se produisait. Différentes hypothèses ont été proposées à ce sujet. On connaît notamment les hypothèses informative et sélective, et plus tard les théories germinale et somatique issues des modèles génétiques. Cette dernière connut un grand développement en 1965 après la formulation d'une hypothèse proposée par Dreyer et Bennet (1965) qui suggère que les chaînes légères d'immunoglobulines sont codées par deux séquences discontinues d'ADN, l'une codant pour la région variable V et l'autre pour la région constante C. A partir de là, de nombreuses recherches ont été entreprises en immunogénétique afin de vérifier comment un gène d'immunoglobuline peut se modifier par mutation somatique de façon à produire un très grand nombre de protéines différentes mais apparentées (Tonegawa 1983, Hood et al 1985) contribuant à la diversification des anticorps. De plus, les recherches se sont poursuivies sur l'association aléatoire par recombinaison des différents frag-

ments et sur l'association combinatoire d'une chaîne lourde avec n'importe quelle chaîne légère. Afin d'obtenir des informations plus précises sur les moyens de diversification, différents travaux ont été entrepris au laboratoire du Dr. Gibson. On chercha à caractériser les gènes $V\kappa$ -1 et les séquences d'acides aminés de $V\kappa$ des myélomes de différentes souches de souris consanguines.

1.1 STRUCTURE D'UNE IMMUNOGLOBULINE

On peut définir les anticorps ou les immunoglobulines comme des globulines plasmatiques synthétisées par des plasmocytes en réponse à une substance antigénique qui ont pour fonction de se combiner avec cette dernière de manière à la neutraliser. Les résultats obtenus à l'aide de différentes techniques

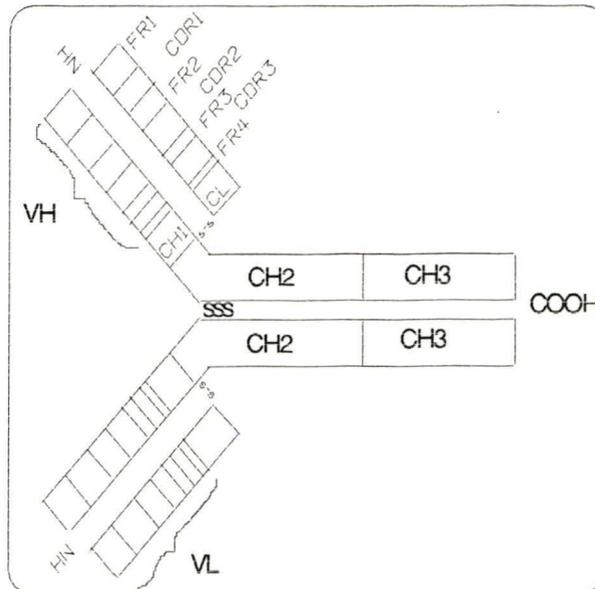


Figure 1: Représentation schématique d'une immunoglobuline (IgG).

de clivage ont permis à Porter de proposer un modèle structural linéaire pour l'IgG (Porter 1959). L'IgG est constituée de 4 chaînes polypeptidiques: deux chaînes lourdes et deux légères identiques. Les chaînes lourdes et légères sont

liées par des ponts disulfures. Chaque chaîne possède une partie variable et une partie constante. La partie variable est caractérisée par 3 régions hypervariables (CDR1, CDR2, CDR3), séparées par des régions moins variables FR (Wu et Kabat 1970). La partie constante de la chaîne lourde détermine la classe de l'anticorps soit: α , δ , $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$, μ et ϵ , par un mécanisme d'épissage alternatif. Cette région donne les propriétés effectrices, comme la fixation du complément et la fixation à certaines cellules.

1.2 CONFORMATION ET STRUCTURE SPATIALE

L'analyse de la séquence des acides aminés montre tant au niveau des chaînes lourdes que des chaînes légères, l'existence de séquences répétitives qu'on appelle domaines. Chaque domaine contient une centaine d'acides aminés dont une soixantaine forme une boucle maintenue dans l'espace par un pont disulfure. Les chaînes légères comprennent deux domaines: un domaine constant C et un domaine variable V; les chaînes lourdes γ et α comprennent 4 domaines: un variable et 3 constants (VH, CH1, CH2, CH3,). Les chaînes lourdes μ et ϵ contiennent 5 domaines: un variable et 4 constants (VH, CH1, CH2, CH3, CH4). Les domaines de la région constante présentent une forte homologie de séquence. La présence des ponts disulfures à l'intérieur des chaînes lourdes et des chaînes légères (ponts intrachaînes ou intracaténaux) est à l'origine d'une conformation particulière, importante pour le bon fonctionnement de l'immunoglobuline. Chaque domaine se replie et forme une masse globulaire, stabilisée par un pont disulfure intrachaîne. A l'intérieur de chaque domaine, le squelette carboné de la chaîne polypeptidique est relié en zig zag, déterminant une structure secondaire stabilisée par des liaisons aux ponts d'hydrogène. On sait toutefois que le site actif est formé par plusieurs

replissements des zones variables de chaînes lourdes et de chaînes légères. A cause de ces replissements les régions hypervariables se trouvent à proximité les unes des autres et forment une crevasse qui sera le site de liaison avec l'antigène. La spécificité de l'immunoglobuline dépend de la forme de la crevasse et des propriétés des acides aminés dans la crevasse (Tonegawa 1985).

1.3 IMMUNOGENETIQUE

Connaissant la structure protéinique de l'immunoglobuline, il importe de connaître l'organisation et l'arrangement des gènes codant pour l'immunoglobuline. Ceci nous permettra de mieux comprendre les mécanismes de leur évolution et de leur diversification. Cela est actuellement possible grâce aux techniques de clonage moléculaire.

1.3.1 Organisation et réarrangement des gènes d'immunoglobuline.

Dans la souris, les chaînes d'Ig sont codées par trois familles de gènes situées sur 3 chromosomes différents :VH sur le chromosome 12 (D'Eustachio et al 1979), V λ sur le chromosome 16 (D'Eustachio et al, 1981) et V κ sur le chromosome 6 (Gottlieb 1974, Gibson 1978, Swan et al 1979). Puisqu'une très petite partie seulement du génome animal est utilisée pour fabriquer les anticorps, Dreyer a proposé que l'ADN de toutes les cellules possède un grand nombre de segments variables (V) et un seul segment constant (C) et que dans chaque lymphocyte B un segment V s'associe au segment C au cours de leur développement pour former un gène d'anticorps complet et fonctionnel (Dreyer et Bennet 1965). Ce n'est qu'en 1976 que Hozumi et Tonegawa de l'Institut

d'immunologie de Basel, ont obtenu les premiers résultats démontrant qu'une réorganisation de l'ADN permettait la mise en place des gènes d'immunoglobuline fonctionnels (Hozumi et Tonegawa 1976). Après le clonage moléculaire des gènes des chaînes légères (Tonegawa et al 1977, Brack et Tonegawa 1977) et des gènes des chaînes lourdes (Sakano et al 1980, Maki et al 1980) on a eu une meilleure compréhension de l'organisation des gènes d'immunoglobulines.

1.3.2 Organisation et structure des gènes d'immunoglobuline.

Les gènes d'immunoglobuline étant constitués de différents isotypes, nous allons décrire l'étude de leur organisation en les traitant à tour de rôle.

a) Les gènes V κ :

Les chaînes légères kappa sont constituées de 3 régions différentes d'ADN : une région V, J et C. Dans la lignée germinale, il y a environ 90-320 gènes V κ (Cory et al 1981) Chez l'humain, la distance entre le dernier gène V κ et les segments J κ déterminée par la technique 'Chromosome Walking', est de 23 Kb (Zachau et al 1987). Mais chez la souris la distance n'a pas pu être déterminée encore. Il y a 5 segments dont J κ 3 est défectueux à cause d'une mutation dans le signal de reconnaissance, ceux-ci étant situés de 2.5 à 4.0 Kb en 5' de l'unique région C, fig 2a (Sakano et al 1979).

La région V κ compte 95 acides aminés. Le J κ en compte 12 et le C κ , 106. Il y a aussi le fragment L (leader) qui est constitué d'acides aminés majoritairement hydrophobes qui forment une partie du signal peptidique. Ce signal est responsable de la protection et la sécrétion de l'immunoglobuline (Nishuka et al 1980). Chaque région variable possède son propre leader dont les séquences

sont très conservées.

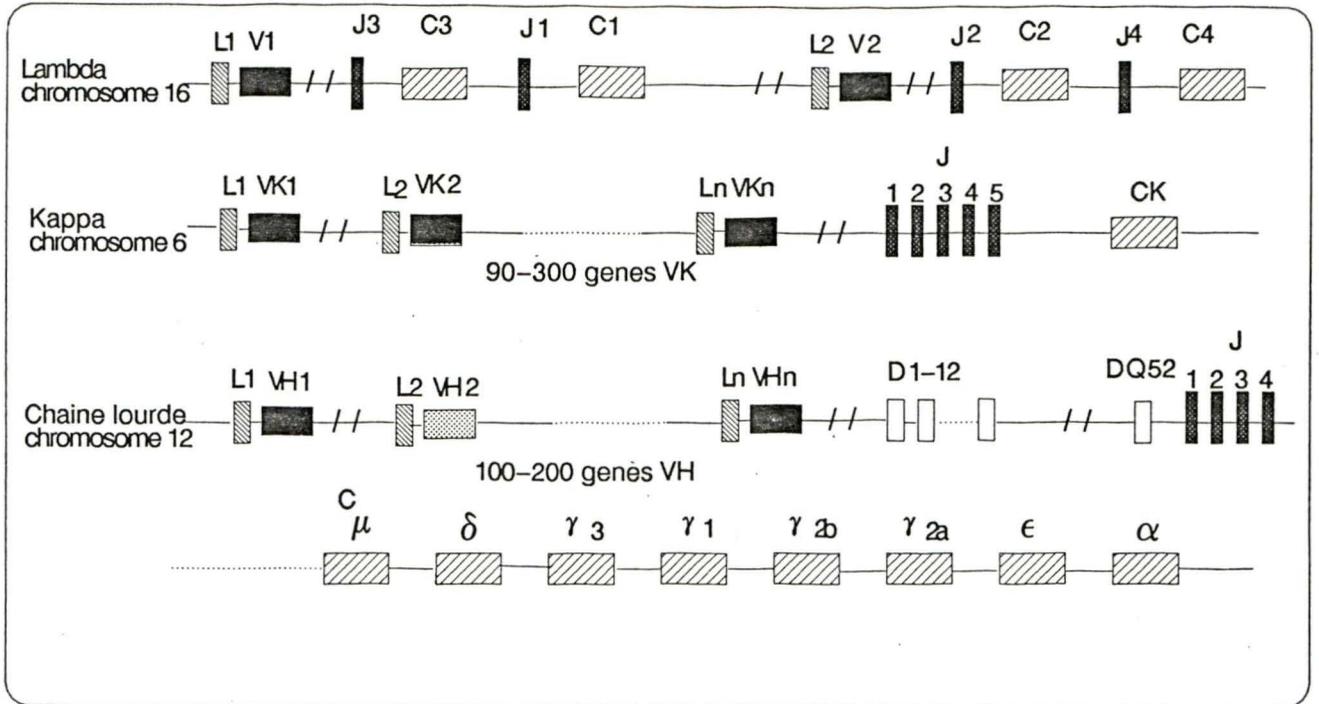


Figure 2a: Organisation des gènes d'immunoglobuline chez la souris

b) Gènes $V\lambda$

La famille du gène λ est constituée de 4 paires de segments J et C : J1C1, J2C2 et J3C3 et J4C4 qui est défectueux (Miller et Storb 1981). La famille comprend aussi deux paires de segments L et V : L1V1 et L2V2. Le segment V1 se lie préférentiellement à J3 ou J1 et V2 avec J2, fig 2a (Weigert et Cohn 1970).

c) Gènes VH

Le génome haploïde de la lignée germinale contient 100 à 200 différents segments LH-VH (Givol et al 1981), un groupe de 12 segments D (diversity)

un groupe de 4 segments JH (Schilling et al 1980, Sakano et al 1980) et un groupe de 8 copies de segments CH, une pour chaque classe d'immunoglobuline (réf fig 2a). Le segment D code pour la partie centrale de CDR3, composée des résidus 1-13. Pour chaque séquence codant pour une des huit copies de CH, il y a un exon correspondant. De plus, les portions intramembranaires et cytoplasmiques de chacune des huit chaînes sont codées par des exons indépendants qui se trouvent en 3' de chaque exon codant pour CH (Maki et al 1981).

1.4 REARRANGEMENT DES GENES

Pour que les gènes des immunoglobulines deviennent actifs, ils doivent être rassemblés. Un modèle a été proposé par Dreyer et Bennet (1965) et par la suite confirmé par Hozumi et Tonegawa(1976). En utilisant une sonde de mRNA d'un myélome sur l'ADN génomique de souris et les cellules sécrétant des anticorps digérés avec le même enzyme, ils ont montré que la région variable et la région constante sont séparées d'une bonne distance dans l'ADN génomique et qu'elles sont plus proches dans les cellules sécrétant des anticorps.

1.4.1 Réarrangement des chaînes légères

Ce réarrangement se produit entre la position 3' du gène V et le 5' du segment J. La jonction formée pendant la recombinaison est imprécise. Le 96ième codon dans le cas de κ et le 98ième codon dans le cas de λ peuvent être générés par la contribution de la deuxième ou troisième base du segment J et le reste par le segment V. Parfois il se peut qu'il y ait insertion d'un codon à ce point de jonction. Pour cette raison on l'appelle "le point chaud "(Sakano et al 1979, Weigert et al 1980). La jonction entre les segments $V\kappa$ et $J\kappa$ est

imprécise et peut donc varier d'un événement de recombinaison à l'autre par plusieurs nucléotides (Sakano et al 1980). Dans le cas du segment JH, le nombre de nucléotides peut varier jusqu'à 10. Pour la chaîne lourde on pense qu'il y a d'abord une liaison des segment J et D et ensuite liaison avec la région du V (Sakano et al 1981 ,Kurusowa et al 1982). Dans une étude des segments D-J et V-J on a observé des insertions de nucléotides pendant la recombinaison où la nouvelle région consistant de nouveaux nucléotides est appelée région N (Honjo et Habu 1985). Ceci contribue à la diversité mais peut aussi entraîner un changement de cadre de lecture qui résulte en un réarrangement aberrant (Altenburger et al 1980, Schiller et Lenhard 1978).

1.4.2 Mécanisme de recombinaison

Le mécanisme de recombinaison V-J ou VDJ n'est pas encore bien compris mais a été expliqué par certaines séquences spécifiques très bien conservées, qui semblent agir comme des séquences de réarrangement. Dans la lignée germinale les segments V et D sont suivis immédiatement soit par un heptamère CACAGTG ou son analogue et un nanomère ACAAAAACC ou son analogue, les deux séparés par une région de séquence non-conservée de 12 paires de bases. Les segments D et J à leur tour sont précédés par un consensus nanomérique GGTTTTTGT et un consensus heptamérique CACTGTG, séparé par 23 ± 1 paires de bases (Sakano et al 1979 , Early et al 1981). Il est à noter que les séquences héptamériques et nanomériques qui suivent les séquences V_k, V_H ou D sont complémentaires aux séquences précédant les segments J_k, D et J_H avec lesquels ils vont être réarrangés. On pense que ces structures répétées et inversées peuvent servir de site de

reconnaissance pour la 'recombinase' qui contient 2 protéines se liant à l'ADN, l'une reconnaissant l'heptamère et le nanomère séparés de 12 pb et l'autre reconnaissant les séquences nanomère et l'heptamère consensus séparés de 23 ± 1 paires de bases (fig.2b). Des expériences ont démontré qu'il y a des

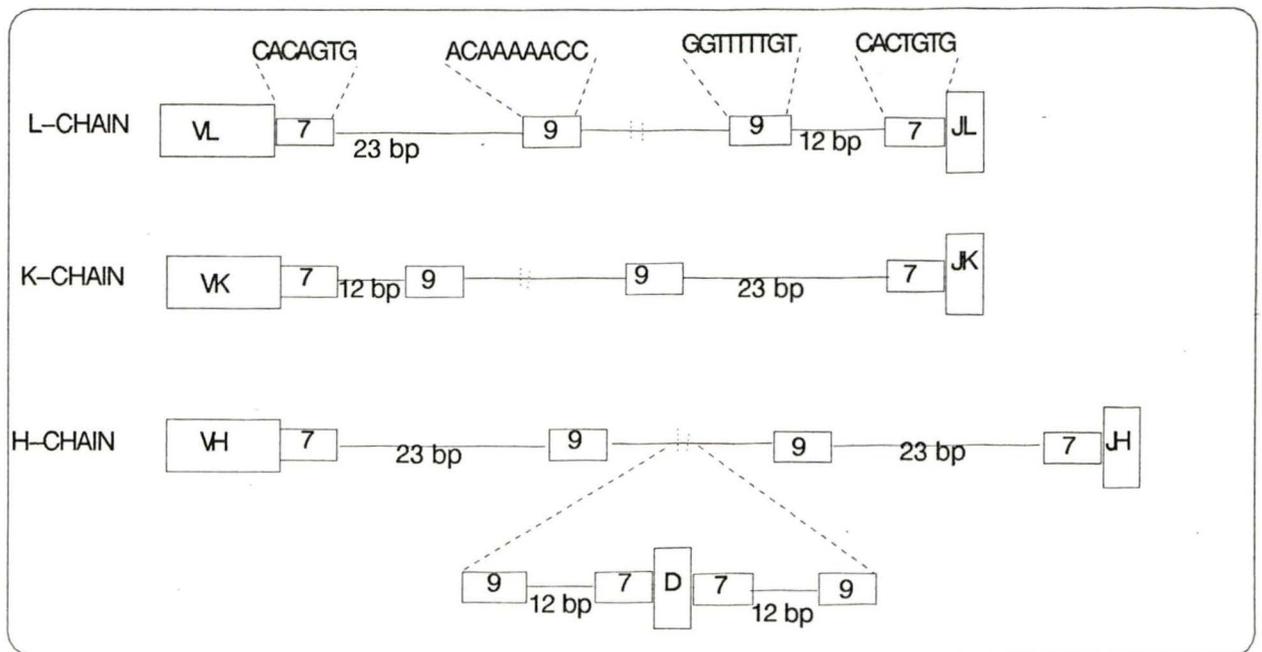


Figure 2b: Schéma donnant une idée sur l'organisation des séquences heptamériques, nanométriques et des spacers en relation avec différents segments de gènes

protéines nucléaires dans les cellules pré-B interagissant avec les signaux de séquences de réarrangement en utilisant des signaux de séquences synthétiques (Aguilera et al 1987).

1.4.3 Modèle de recombinaison

Plusieurs modèles basés sur la recombinaison intra-chromosomale ont été proposés. Dans le modèle "copy-insertion", le segment V est copié, puis inséré au site J. Alternativement, le segment V peut être excisé dans une structure ressemblant à un épisome, puis aller ensuite s'intégrer près d'un segment J

(Excision-insertion) (Hozumi et Tonegawa 1976). Dans le troisième modèle, l'intron entre le segment V et le segment J forme une boucle qui est excisée (looping out deletion). Ce mécanisme a été appuyé par une expérience démontrant l'absence de la région 5' flanquante du J et de la région 3' flanquante du segment V dans certaines myélomes produisant les chaînes kappa (Seidman et al 1980). Un autre modèle de réarrangement suggère que celui-ci se produirait par inversion, c'est-à-dire le segment V et C seraient dans une direction transcriptionnelle opposée au début et avec l'inversion de la séquence intervenante, le segment V serait réarrangé avec le segment C dans la bonne direction (Lewis et al 1983, 1985). Dans les gènes Vk de l'humain, les gènes Vk sont orientés dans la polarité inverse que celle du segment J. Ceci nous indique que le réarrangement dans certains cas peut être causé par inversion (Zachau et al 1987). Finalement il y a aussi une autre possibilité selon laquelle, il y aurait un échange inégal entre chromatides soeurs pendant la duplication mitotique. Le bout 3' d'un segment d'un gène V d'une chromatide soeur et le bout 5' du segment d'un gène J dans la seconde chromatide soeur se lient. Tandis qu'une des chromatides se retrouve avec le locus kappa réarrangé, l'autre porte le produit réciproque de recombinaison. Les deux chromatides sont ensuite séparées dans deux cellules filles (Van Ness et al 1979).

1.5 L'EXPRESSION DU GENE D'IMMUNOGLOBULINE

Dans les cellules pré-B, la région variable des gènes est rassemblée. Les gènes V de la chaîne lourde sont rassemblés avant ceux de la chaîne légère pendant la différenciation des cellules B (Alt et Yanocopoulos 1986). Il y a évidence à l'effet que le réarrangement des classes spécifiques de segments est

contrôlé par la modulation de l'accessibilité à la 'recombinase' (Yancopoulos et al 1985). Blackwell a obtenu des résultats qui suggèrent que la transcription peut jouer un rôle dans l'assemblage des segments du gène V dans les lignées pré-B (Blackwell et al 1986). On pense que pendant la transcription par la RNA polymérase, l'ADN se déroule ; ce qui est nécessaire pour l'initiation de recombinaison (Blackwell et al 1986). Le locus V réarrangé et fonctionnel est sensible à la DNase dans une région intronique constituée de 200-250 bp qui se trouve à 600 pb en 5' du segment C. Cette région est responsable de l'augmentation de l'activité transcriptionnelle de 10,000 fois plus que le promoteur seulement (Storb et al 1981). Il y a une certaine homologie d'environ 11 pb de cette région et l'élément d'enhancer de 72 pb de SV40. Cette région d'homologie (GGGGACTTCC) correspond à la région sensible aux nucléases. L'expérience avec les cellules 70 Z/3 (une lignée de cellule pré-B qui possède la chaîne légère V_k réarrangée mais presque pas transcrite) indique qu'il y a d'autres facteurs qui sont nécessaires à la transcription autre que le réarrangement. On pense que certaines protéines spécifiques à une certaine séquence vont reconnaître la région intronique régulatrice et augmenter la transcription à partir du promoteur V_k le plus proche en agissant en cis. Plusieurs modèles ont été proposés et selon lesquels l'interaction avec le promoteur et l'enhancer peut se faire-soit par le modèle 'scanning' ou 'looping'. Ce dernier suggère que l'enhancer et le promoteur interagissent l'un avec l'autre via les protéines par la formation d'une boucle (Müller et al 1988).

1.5.1 Le site du promoteur

En dehors de l'enhancer, pour la transcription des gènes d'immunoglobuline la séquence TATA en position 5' du leader est aussi importante et se trouve à 30 pb du 'cap site' qui est lui-même situé à 20 pb du codon d'initiation ATG. (Falkner et Zachau 1984). On trouve aussi des séquences entre -90 et -160 pb en amont de la région codante qui sont nécessaires pour la transcription correcte. Ces régions contiennent des séquences déca et pentadécanucléotidiques, TNATTTGCAT et TGCA[G/C]TGTGNCCAGT qu'on appelle dc et pd respectivement. Les séquences reliées au dc et pd se trouvent en 5' de toutes les chaînes légères humaines et murines. Dans l'enhancer de la chaîne lourde de souris, les séquences se trouvent dans une forme complémentaire et inversée de l'élément dc qu'on appelle cd (ATGCAAATNA), dans la région 5' de tous les gènes variables de la chaîne lourde. Ces séquences sont conservées dans leurs nucléotides et dans leur distance de la région codante. Parslow et al (1984) ont démontré par l'analyse délétionnelle que le dc décrit par Falkner,

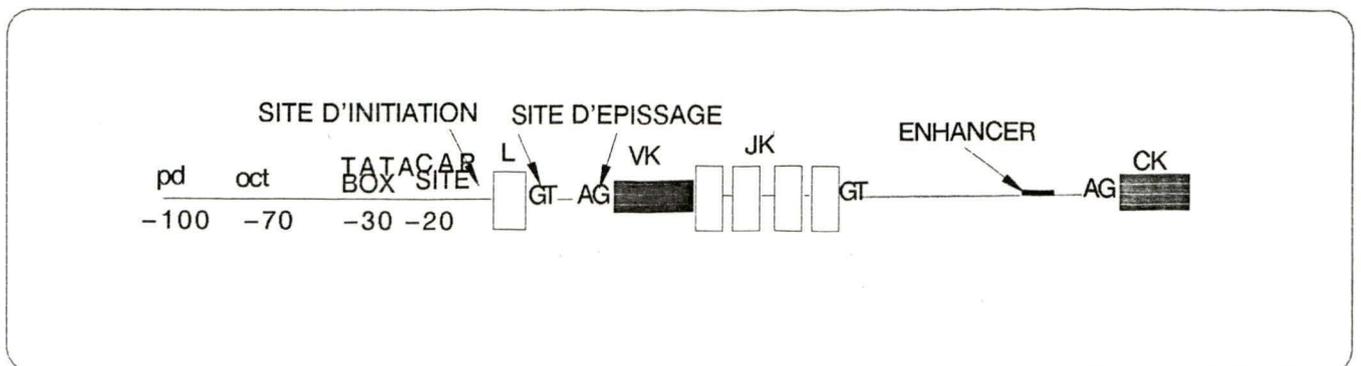


Figure 3: Schéma montrant les régions importantes pour l'expression d'un gène Vk réarrangé (d'après Parslow et al 1984). oct=octamère, pd=pentadécanucléotide

important pour la transcription, n'est qu'un motif octamérique (ATTTGCAT). L'octamère se situe à 70 pb en amont du site d'initiation de la transcription et le pentadecanucléotide se situe à 100 pb du site d'initiation de transcription (Falkner et Zachau 1984). (ref. fig 3). Ces séquences servent de site de reconnaissance pour le facteur NF-A2 régulant l'expression des gènes

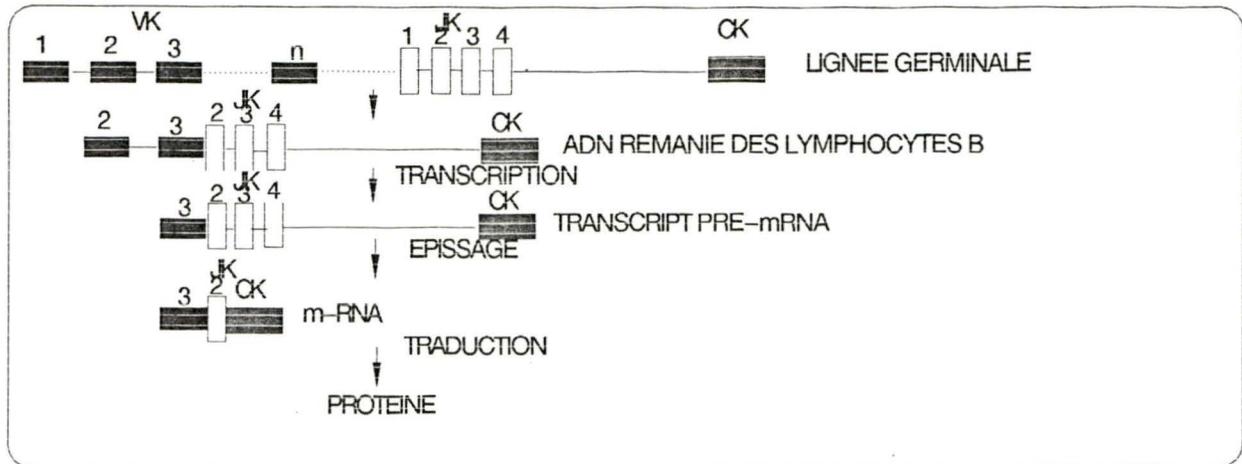


Figure 4: Organisation des gènes κ et réarrangement de l'ADN permettant la production d'une chaîne κ .

d'immunoglobuline d'une manière spécifique aux tissus. Ce facteur est codé par le gène *oct-2* (Staudt et al 1988).

Après la transcription, il y a production d'un pré-mRNA d'immunoglobuline. Il est constitué d'une région codant pour le signal peptidique séparé du segment V par un intron de 100-200 bp. Puis le segment V associé avec le segment J est suivi d'un intron d'environ 1,400 bp. Le segment C κ est à son tour suivi d'une séquence 3' non-traduite et d'une queue de poly A. Les introns sont enlevés par

épissage pour donner le mRNA mature(voir figure 4).

1.6 GENERATION DE LA DIVERSITE DES ANTICORPS

1.6.1 L'association combinatoire

C'est un moyen non-génétique pour augmenter la diversité des anticorps. Cette diversité s'explique par le fait que les régions $V\lambda$ et VH sont codées par des segments de gène différents respectivement. Ainsi, si le génome possède 2 segments $V\lambda$, 3 segments $J\lambda$, 300 segments $V\kappa$ et 4 segments $J\kappa$, le maximum de différentes régions $V\lambda$ générées somatiquement est $(2 \times 3) + (300 \times 4) = 1206$. Il est de même pour la chaîne lourde: 200 $VH \times 12 D \times 4 JH$ font 9600 VH (Tonegawa 1983). Si on suppose qu'au niveau de la protéine n'importe quelle chaîne légère peut associer avec n'importe quelle chaîne lourde, l'association combinatoire va donner $1206 V\lambda \times 9600 = 10^7$ molécules d'anticorps différentes.

1.6.2 Diversité au site de jonction.

La jonction des segments V-J, VH-D et D-JH permet une certaine flexibilité sur quelques nucléotides permettant ainsi beaucoup de variation à ces positions (Sakano et al 1980). On a vu dans la section 1.4.1 que la jonction V-J est un point chaud où on peut avoir insertion ou délétion suggérant que pour le même segment V et J l'événement de recombinaison peut varier de quelques nucléotides. Malgré tout ça et parce que tous les segments des gènes des immunoglobulines doivent être traduits dans un seul cadre de lecture pour donner naissance à une chaîne d'immunoglobuline fonctionnelle, il arrive souvent que l'imprécision de ces jonctions résulte en un segment qui est en

dehors du cadre de lecture . La diversité est donc accomplie avec un certain gaspillage (Weigert et al 1980).

1.6.3 La mutation somatique

L'idée de la mutation somatique a été proposée dans les années 1950 (Lederberg et al 1959). La première évidence de la mutation somatique a été obtenue quand Weigert et Cohn (1970) ont identifié la région $\nu\lambda$ de 19 différentes myélomes de BALB/c et de NZB. Ils ont trouvé que 12 d'entre eux sont identiques au prototype, tandis que 7 autres (séquence variable) étaient différents par 1-3 nucléotides du prototype. Comme il a été démontré qu'il n'y a qu'un seul segment $V\lambda$ par génome haploïdique (Brack et al 1977, Bernard et al 1978), ceci implique que les 7 autres sont formés par mutation somatique. Plusieurs autres exemples de séquences réarrangées comparées avec leurs contre-parties germinales démontrent plusieurs mutations. La mutation somatique se trouve exclusivement près des gènes réarrangés (Pech et al 1981; Kim et al 1981; Gearhart et Bogenhagen 1983). Les mutations semblent localisées dans une région de 1 Kb qui inclut 300 pb 5' de la région codant pour les segments V, D, et J, la région codante elle-même et la région 3' du segment J. Les mutations somatiques n'ont pas été trouvées en dehors de cette région de hypermutation (Kim et al 1981; Gearhart et Bogenhagen 1983). Le taux de mutation ponctuelle est de 10^3 à 10^5 mutations par paires de bases par division. (Clarke et al 1985; Wabl et al 1985), ce qui est 10^4 fois plus que le taux des mutations spontanées dans l'ADN eucaryotique (Drake et al 1970, Vogel 1970). Plusieurs sources de mutation existent: le mauvais appariement des tautomères et isomères de paire de bases normales, la mauvaise incorporation

causée par les erreurs de la polymérase, la déamination et la déalkylation causées par les électrophiles endogènes et la radiation (Drake et al 1970). En analysant les séquences dans la régions V, Golding et al ont remarqué la présence de la mutation dans des régions qui ont suffisamment d'homologie avec les séquences avoisinantes. Il a suggéré que la présence des séquences répétées directes et palindromiques peut servir de matrice pour ces mutations en permettant le mauvais alignement pendant la réplication ou la réparation de l'ADN. Ce mauvais alignement peut aboutir à une délétion, duplication, substitution de bases, changement de cadre de lecture ou une combinaison de tout cela (Golding et al 1987).

1.6.4 Mobilité des modules.

Avec les études tridimensionnelles par cristallographie, on a pu voir de façon plus détaillée les mouvements des domaines. (Bennet et Huber 1984). Jusqu'à dernièrement on croyait que les modules VH-V λ , CL-CH1, CH3-CH3 et le domaine CH2 étaient rigides. Il a été prouvé récemment qu'il y a une certaine variabilité dans l'association de VH-V λ et dans les autres modules aussi. Le mode d'association avec VH-V λ a une influence directe sur l'arrangement spatial du CDR de la chaîne lourde et légère déterminant ainsi la spécificité de liaison avec l'antigène. Si la variabilité dans l'arrangement VH-V λ d'un certain anticorps existe, ceci pourrait ajouter une autre dimension dans la diversité des anticorps en laissant un certain anticorps reconnaître une variété d'épitopes. (Huber et Bennet 1987).

1.7 ESTIMATION DU NOMBRE DE SEGMENTS V

Les études des acides aminés des immunoglobulines et des gènes

d'immunoglobuline démontrent que la diversité des anticorps est produite par: (1) la recombinaison somatique des segments V, J, D; (2) les mutations somatiques; (3) le nombre de gènes V dans la lignée germinale. Pour évaluer la contribution des différents gènes à la diversité des anticorps, il est important de connaître la quantité de gènes codant pour les chaînes lourdes et légères dans la lignée germinale. Le nombre de gènes est le point de départ pour tout le mécanisme de diversification combinatoire ou mutationnelle. Différentes techniques ont été utilisées afin de calculer le nombre de gènes VH et V λ . Chaque région V est constituée de sous-groupes dont les séquences sont étroitement reliées. Cela suggère que différents gènes structuraux contrôlent les différentes séquences de protéines dans un groupe (Weigert et al 1980, Mc Kean et al 1978). En gros le nombre de gènes V est égal au nombre de sous-groupes multiplié par le nombre de groupes de gènes dans la lignée germinale. Ceci n'est pas exact parce que certains groupes ont plus de membres que d'autres.

En étudiant 79 séquences d'acides aminés de Vk, Potter (1977) les a classées en 18 groupes-chaque groupe pouvant avoir moins que 12 substitutions d'acides aminés entre l'acide aminé 1 jusqu'au tryptophane 35. Chaque groupe est codé par au moins un gène germinal distinct (Potter 1977). Comme la région jusqu'au trp 35 contient la région FR1 et CDR1 et montre plus de variation que d'autres régions de la séquence, ce type de classification est un bon indicateur de la variation des gènes kappa. Dans l'étude de la distribution des groupes alléliques dans différents Igk haplotypes, D'Hoostelaere et al (1988) ont utilisé 14 groupes de Vk comme sonde dans des conditions rigoureuses de lavage pour détecter les hybridations des hétéroduplexes de plus de 80% d'homologie. Dans la souris BALB/c environ 96 REF (fragment

d'enzyme de restriction) au total hybrident avec les 14 sondes utilisées. Chaque REF étant égal à un exon de V κ , il y a en moyenne $96 \div 14$ V κ par sonde V. Le nombre de groupes V κ total d'après les séquences d'acides aminés jusqu'au Trp35 (18) et Cys23 (6) est égal à 24 groupes. Le nombre total d'exons V κ est égal à $24 \times 96 \div 14 = 165$ (D'Hoostelaere et al 1988).

D'autres chercheurs, en utilisant une sonde de cDNA faite à partir d'un mRNA du gène V κ 21 ont hybridé jusqu'à saturation une quantité fixe d'ADN. Le même procédé a été fait par une sonde contenant le C κ seulement. Puis le nombre de sous-groupes de V κ 21 a été évalué entre 4-6. D'après les séquences d'acides aminés de la région 1 à 23 de la souche de souris NZB, on a estimé environ 50 groupes. De ça on peut déduire que le nombre total de gènes dans la lignée germinale est environ 300 gènes (Valbuena et al 1978). En utilisant la technique d'hybridation d'une sonde V κ distincte et C κ avec le mRNA de la rate, on est arrivé à calculer le nombre de sous-groupes. Avec la sonde C κ , on obtient le nombre total de segments V κ et avec la sonde V κ distincte on obtient seulement ses sous-groupes. Le ratio de C κ /V κ distinct donne le nombre de groupes qui est environ 280. En assumant qu'il y a 7 gènes pour chaque groupe de V (d'après les études de Southern blot) au total dans la lignée germinale on obtient environ 2000 gènes V κ (Zeelon 1981). Malheureusement cette valeur est très approximative car 20% du mRNA détectable par hybridation sont incapables de produire des chaînes d'immunoglobulines fonctionnelles (Bothwell et al 1981). Cory et al (1981) ont préparé 10 sondes de V κ à partir de cDNA et ont hybridé avec l'ADN génomique de souris. En comparant le patron des bandes sur "Southern blot" de différentes digestions, ils ont trouvé 4 ensembles de groupe de gènes V κ chacun contenant environ

15-17 bandes au maximum. Ce qui fait environ 68 le nombre de gènes de Vk. Finalement, le Dr. Gibson (1984) a tenté de caractériser le nombre de groupes fonctionnels par focalisation isoélectrique. De 426 chaînes légères de myélome des souches de souris BALB/c et NZB caractérisées par focalisation isoélectrique et électrophorèse à pH 3, approximativement 50% de ces chaînes ont des propriétés isoélectriques uniques et les autres 50% se trouvent dans des groupes de 2 ou plus de protéines paraissant identiques. Dans ce dernier, en utilisant la distribution de fréquence par la méthode 'Poisson', il a estimé environ 65 groupes de gène V fonctionnels. Certains groupes peuvent avoir été dupliqués récemment et n'ont pas eu assez le temps de diverger. Ainsi le nombre de V pourrait être supérieur à 65(Gibson 1984).

D'autre part des travaux ont été effectués pour estimer le nombre de gènes VH. En utilisant 24 sondes des segments de VH, on a remarqué que chaque sonde hybridait à un des 7 patrons de restriction distincts de Eco R1 sur un "Southern blot" d'ADN de foie de BALB/c. Chaque patron peut donner des fragments 2-3 avec sonde 36-60 à 35 fragments avec la sonde J558. Ainsi, chaque famille étant distincte dans le nombre et le patron de restriction, on peut déduire 7 familles de gènes VH et en calculant le nombre de fragments dans chaque patron, on arrive à 108 gènes VH. L'incertitude réside cependant dans l'existence additionnelle de familles VH qui n'ont pas été encore identifiées (Brodeur et Riblet 1984). La supposition que chaque bande hybridant avec la sonde J558 correspond à un seul gène pour l'expérience mentionnée ci-dessus, s'est avérée fautive par le clonage de ces bandes. De plus, le nombre de gènes VH de J558 présents dans les bandes de blot génomique analysées par densitométrie révèle que chaque bande contient environ 10-20 segments de

VH. L. Hood et al (1986) ont essayé de mesurer quantitativement la grandeur du sous-groupe de la famille J558 dans la souris BALB/c par hybridation. Une quantité en excès de sonde simple brin J558 est utilisée pour hybrider une quantité progressive de l'ADN génomique de BALB/c qui a été préalablement brisé en fragment de 500-600 bp. La fraction de la sonde qui forme un duplex et la fraction de la masse de l'ADN réagissant avec la sonde sont calculées. Le nombre de gènes est obtenu par cette formule :

$$N = \frac{2(1/a)L_G}{L_P}$$

N= le nombre de gènes; 1/a= la fraction de la masse de l'ADN réagissant avec la sonde; L_G =la longueur du génome haploïde; L_P =la longueur de la sonde
v.

En supposant que les autres familles ont une hybridation croisée avec la sonde utilisée, ils ont déduit qu'il y a 500-1000 ou plus de gènes VH dans le génome, ce qui est une quantité 10 fois supérieure à la quantité estimée par Brodeur (Hood et al 1986). Les principales difficultés en comptant les gènes par hybridation avec le "Southern blot", consistent en la sous-estimation résultant de la co-migration des fragments de restriction, l'incapacité de transférer des gros fragments ou des très petits fragments ou la présence de plus d'un gène dans une bande de restriction. Il y a aussi possibilité de surestimation due à des sites de restriction dans la région couverte par la sonde et à la détection de gènes homologues à la sonde mais non-fonctionnels.

1.8 PSEUDOGENES

Les pseudogènes sont définis comme des gènes ayant des séquences d'ADN qui ont une homologie avec les gènes fonctionnels mais qui ont des

différences dans la région codante ou régulatrice ou dans le site d'épissage. Ceci peut causer une terminaison prématurée de la traduction, la formation d'un peptide ayant peu d'homologie avec le produit du gène fonctionnel ou un silence transcriptionnel. Le premier pseudogène fut décrit en 1977 par Jacq dans le gène 5S d'oocyte de *Xenopus Laevis* (Jacq et al 1977). Ensuite grâce aux techniques perfectionnées de clonage, la découverte des pseudogènes dans d'autres familles de gènes s'en est suivie. La famille de multigènes comme la famille des gènes d'immunoglobulines est formée par amplification et la duplication d'un gène primordial (Hood et al 1975). Les échanges inégaux de chromatides soeurs ont été postulés comme mécanisme pour l'expansion et la contraction de la famille. Cela expliquerait pourquoi certains groupes de gènes sont fixés dans le chromosome et sont en grand nombre tandis que d'autres sont très peu représentés ou pas du tout. La divergence du gène qui a été fixé, est causée par des mutations ponctuelles, des délétions, des insertions et conversions (Baltimore et al 1981, Hood et al 1978). Souvent le prix est coûteux. Pendant la divergence il y a introduction des changements qui ne sont pas compatibles au bon fonctionnement du gène qu'on appelle *pseudogène*. Malgré le fait qu'on décrit un pseudogène comme un gène qui n'est pas fonctionnel, on a vu que ce n'est pas toujours le cas. Pour augmenter la diversification du seul gène $V\lambda$ dans le poulet, les pseudogènes sont utilisés (Reynaud et al 1985). Quant au nombre de pseudogènes de $V\kappa$ et VH , on pense qu'il y a une grande proportion de gènes non-fonctionnels. Bien sûr, il y a aussi une catégorie de gènes qui sont apparemment fonctionnels (d'après les séquences) mais qui, pour une raison ou un autre ne sont pas recombinaisonnés pour former des gènes exprimés. Peut-on les considérer aussi comme des

pseudogènes ?

Crews et al (1981), Bothwell et al (1981) et Givol et al (1981) estiment que 30% des gènes VH sont non-fonctionnels dû à une ou plusieurs mutations aberrantes. Le même phénomène a été noté pour les gènes Vk (Zachau et al 1984). La plupart des pseudogènes décrits possèdent soit une délétion, soit une mutation dans le signal de reconnaissance (heptamère, nanomère), soit une mutation dans le site d'initiation ou d'épissage ou dans l'octamère, soit la formation d'un 'codon stop' dans le gène ou la substitution dans le frame-work. L'exemple le plus frappant, est le gène HK100Vk, le premier pseudogène d'immunoglobuline trouvé et décrit chez l'humain. Ce gène possède plusieurs mutations aberrantes: un 'codon stop' à la position Trp35 où TGG est remplacé par TGA, une délétion de 2 pb dans la région de l'acide aminé 17/18, une insertion de 4 bp, 3 bp et 1 bp aux positions 25/26, 85/86 et dans la région de la cystéine invariante à la position 88 respectivement. De plus dans le bout 5' de la région intervenante au lieu du site d'épissage GT hautement conservé on observe le dinucléotide GC. De plus, dans la séquence de reconnaissance, l'heptamère est correct mais dans le nanomère elle est substituée par ATAAGCTCC au lieu de ACAAAACC. Ce gène ne peut donc pas se joindre au segment J (Bently et Rabbits 1980).

1.9 INTERET DE L'ANALYSE DES GENES DU GROUPE Vk-1

D'après les études des patrons de focalisation isoélectrique, on voit que les chaînes d'immunoglobulines codées par le sous-groupe Vk-1A représentent 2% de toutes les chaînes légères d'immunoglobulines (Ng et al 1989). De plus il existe un polymorphisme entre différentes souches de souris au niveau du marqueur Ef2a, la présence duquel est associée au gène structural Vk-1A. Cela permet d'étudier l'évolution de ces gènes plus facilement et d'avoir une idée

générale sur l'évolution du complexe Vk.

Les gènes Vk-1 sont aussi impliqués dans différentes réponses immunitaires contre plusieurs antigènes différents comme le DNP (Hum et al 1984), la flagelline (Smith et al 1977), le phenyloxazolone (Kaartinen et al 1983), la phosphorylcholine (Todd et al 1984), la GAT (Fougereau et al 1984), l'DNA (Köfler et al 1985) et la digoxine (Novotny et al 1983). L'étude de la réponse immunitaire à un antigène est rendue plus accessible grâce à l'injection de ces produits et l'analyse de ces anticorps qui proviennent des gènes structuraux Vk-1. De plus, le gène Vk-1 est responsable de plusieurs maladies autoimmunitaires. Par exemple, les études sur les auto-anticorps anti-IgG (facteur rhumatoïde) ont révélé l'implication du gène structural Vk-1C (Shlomchik et al 1986). De même, les études ont démontré que le gène (Vk-1A) codant pour des anticorps spécifiques contre les antigènes GAT et la digoxine est aussi responsable pour la formation de la chaîne légère de l'autoanticorps anti-ADN (Köfler et al 1985). Etant donné la haute fréquence des autoanticorps possédant des propriétés de liaisons multi-spécifiques, on a tenu souvent les gènes qui codent pour les anticorps contre des antigènes exogènes responsables pour des maladies autoimmunes. Malgré tout cela on ne peut pas s'empêcher de penser que le contrôle de la maladie elle-même se fait au niveau du gène. Par exemple : le gène de la chaîne VH de l'autoanticorps codant pour MRL-DNA 10 anti-DNA dans la souche de souris MRL/Mp-*lpr/lpr* ressemble au pseudogène 6 de la souche C57 BL/6 qui n'a pas l'auto-anticorps et qui n'est pas fonctionnel dû à une délétion d'une paire de base au codon 22 correspondant à un acide aminé essentiel, cystéine (Köfler et al 1985). Le groupe de Reininger et al (1987) a remarqué aussi que les gènes de la chaîne lourde et légère des autoanticorps qui reconnaissent des globules rouges traités au bromelin sont des gènes spécifiques (Reininger et al 1987). Ainsi les études sur le groupe

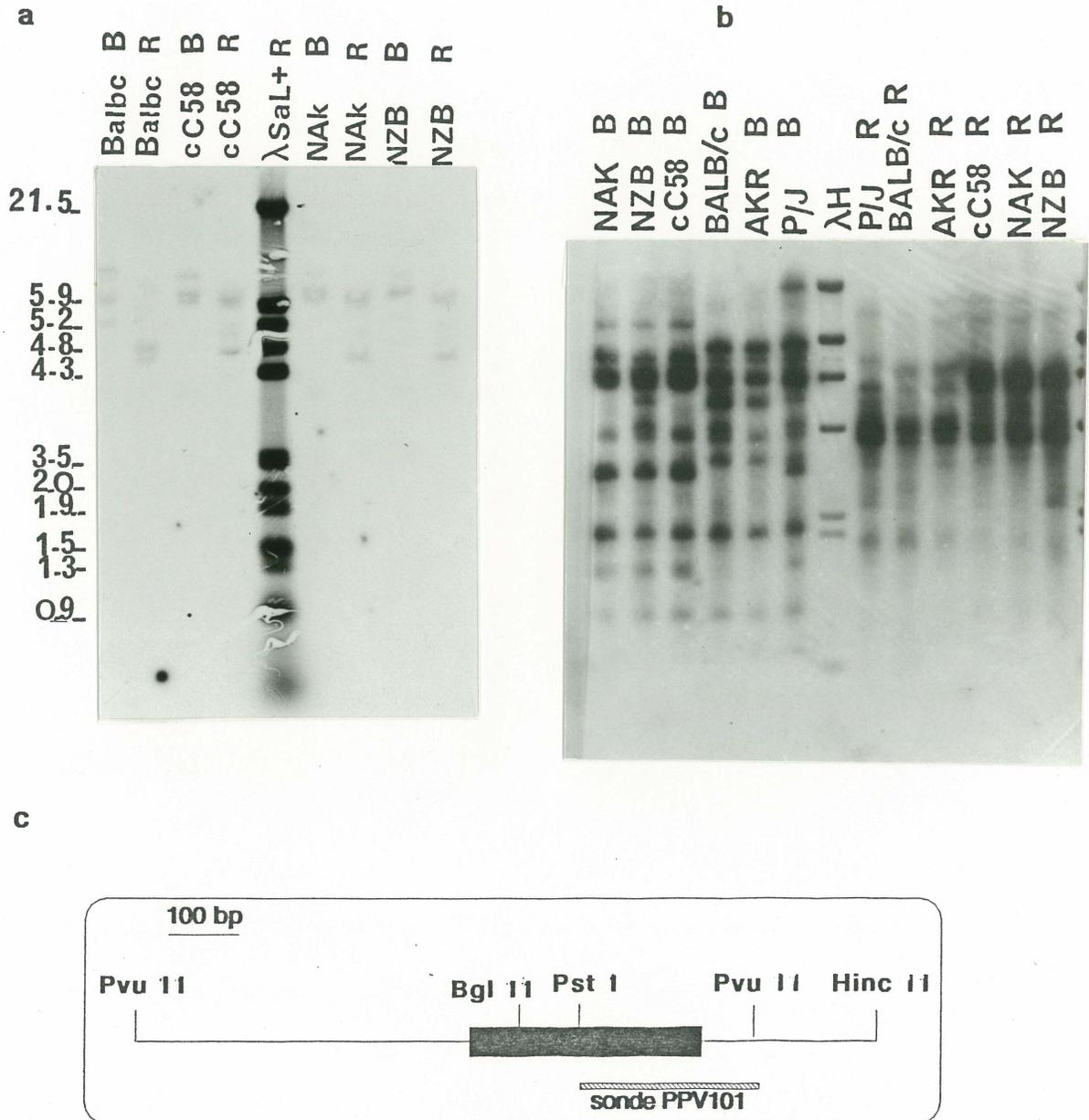


Figure 5: Autoradiogramme de différentes souches de souris dont l'ADN est coupé avec Bam H1 (B) et Eco R1 et hybridé avec la sonde PPV101. Dans la photo 5a, les lavages sont faits dans des conditions sévères (0.1x SSC 65°C 30 minutes). On peut voir les bandes de 8.5 Kb et 6.3 Kb pour la souche de Balbc digéré Bam H1(B). Par contre dans des conditions de lavages légères (0.5x SSC 65°C 30 minutes), en dehors de ces deux bandes on peut voir plusieurs autres bandes(5b).

Figure 5c: la sonde PPV 101 utilisée pour le criblage. Elle est une partie de la région V de la souche de souris CE 6.0 Kb (Vk1-cf) du site Pst I au site Pvu II dans la région 3'.

Vk-1 pourront nous aider à mieux comprendre le mécanisme responsable de l'apparition des maladies autoimmunes.

1.10 IDENTIFICATION DES AUTRES GENES ETROITEMENT RELIES A Vk-1

Avec une sonde flanquante 5' (1.1) un blot de l'ADN génomique de souches de souris consanguines (BALB/c, AKR, NZB, NAK, C58, CE, BDP et P/J) digéré avec Bam H1 et Eco R1 révèle 2 bandes majeures démontrant un certain polymorphisme entre les souches caractérisées Ef2a et Ef2b (Moynet et al 1985). En utilisant une sonde de la région V de Vk-1 (Pst I-Pvu II du gène Vk1-Cf) connue aussi sous le nom de PPV101 dans des conditions de lavages sévères pour l'ADN de la souris BALB/c digéré Bam H1, on remarque les deux bandes mentionnées plus haut. Celles-ci représentent les gènes Vk1-A et Vk-1C. Mais dans des conditions de lavages légères, en dehors de ces deux bandes on remarque plusieurs autres bandes supplémentaires (8 bandes) (fig 5). Certains de ces fragments sont présents seulement dans certaines souches de souris et correspondent au polymorphisme attendu entre les souches caractérisées comme Ef2a et Ef2b. Par contre, il y en a aussi qui apparaissent dans toutes les souches. Il se peut que ces gènes associés au Vk-1 se situent loin des autres membres du groupement principal sur le chromosome ou ces bandes peuvent représenter des fragments ayant la même grandeur mais contenant différents gènes Vk.

Si on suppose que ces bandes sont des bandes mineures de Vk-1, il semble y avoir des différences du point de vue du locus I μ K-Ef2 parmi les souches qu'on a caractérisées comme Ef2a. Les bandes du fragment 5.6 et 4.4 apparaissent seulement dans les souris BALB/c et SJL et non pas dans AKR, NZB ou C58 . Ceci pourrait indiquer que ces gènes se trouveraient sur le côté C kappa du

point de recombinaison dans la souche NAK parce que la plupart des différences entre BALB/c et AKR se situent dans cette région (D'Hoostelaere 1986). Les bandes sur le blot pourraient aussi représenter les autres groupes de Vk qui ont une certaine homologie avec la sonde. D'après l'intensité plus faible de certaines bandes, on pourrait aussi penser qu'elles représentent des gènes Vk-1 qui sont non-fonctionnels et qui au cours des années ont accumulé des mutations suffisantes pour donner seulement une faible homologie avec la sonde. Il y a aussi la possibilité que dans les conditions d'hybridation à faible stringence, les sondes Vk-1 (par exemple PPV101) hybrident avec d'autres groupes de Vk comme Vk-3.

D'après les séquences obtenues des chaînes légères de myélomes et des gènes de la lignée germinale, les gènes Vk-1 sont classés dans 5 sous-groupes (Ng et al sous-presse). Si on examine le nombre de bandes pour la souche de souris BALB/c en dehors de ces cinq sous-groupes il y en a encore huit autres qui n'ont pas encore été identifiés. Ces bandes mineures représentent-elles d'autres gènes Vk-1?

Afin de répondre à ces questions, l'identification de ces bandes est nécessaire. Pour ce faire, nous avons utilisé une sonde PPV101 pour cribler une bibliothèque de l'ADN de souris BALB/c clonée dans le vecteur Embl4, ce qui nous a permis d'isoler vingt et un clones. En les digérant, quatre d'entre eux donnent un fragment de 5.4 kb Bam H1 contenant la région Vk. Ils ont été sous-clonés dans le phagemide Bluescript M13+ et deux des quatre ont été séquencés. L'information obtenue nous a permis de les identifier comme étant différents. Ils ont été également comparés avec d'autres séquences d'acides nucléiques et peptidiques de différents groupes de Vk et de sous-groupes de Vk-1 pour les classer.

2 MATERIEL ET METHODES

2.1 EXTRACTION DE L'ADN DE REIN

C'est une technique du laboratoire Jackson (Bar Harbour, ME). Le rein est placé dans un homogénéisateur (Weaton scientific) froid. Le tampon LST (20 mM tris, 10 mM NaCl, 3 mM MgCl₂ pH 7.4) est ajouté à l'organe homogénéisé. On centrifuge le mélange. Après la centrifugation, on resuspend le culot nucléaire dans ACE (50 mM Na acétate, 10 mM EDTA pH 5.1). On ajoute du SDS pour avoir une concentration finale de 1%. On incube sur glace pendant une heure. Après l'incubation sur glace, on fait une extraction avec du phénol saturé avec ACE suivie d'une extraction chloroforme/alcool isoamylique (24:1). A la phase aqueuse obtenue après l'extraction, un volume égal d'éthanol 98% est ajouté très lentement. Puis l'ADN est fait adhérer à une tige de verre stérile puis resuspendu dans TE (10 mM tris, 1 mM EDTA pH 7.5).

2.2 DIGESTION AVEC DES ENZYMES DE RESTRICTION.

TABLEAU 1: Composants des différents tampons utilisées pour les digestions avec les enzymes de restriction.

Tampon L 10x :

(Sacl, SacII) 100 mM tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT et 500 mM NaCl

Tampon M 10x :

(HindIII, PstI, XbaI, XhoI) 100 mM tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT et 500 mM NaCl.

Tampon H 10x :

(EcoRI, BamHI, BglII) 500 mM tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT et 1.0 M NaCl.

Tampon KM 10x :

(KpnI, SmaI, HincII) 100 mM tris- HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT et 500 mM KCl

Pour la digestion on utilise 5 U d'enzyme par ug d'ADN. La solution de digestion contient 1/10 du volume final du tampon 10x, 100 µg/ml BSA et 20 mM DTT. Quand une double digestion doit être effectuée, dans différents tampons de différentes concentrations de sel, on fait une précipitation entre les étapes de digestion. Les tampons 10x de la compagnie Pharmacia sont utilisés pour les digestions même si dans certains cas l'enzyme Pharmacia n'est pas utilisé.

2.3 ELECTROPHORESE

Les échantillons d'ADN sont mis sur des gels d'agarose (Seakem) de 0.5 à 1.5% selon les fragments qu'on veut examiner. Plus le fragment est petit plus on a besoin d'un gel plus concentré. Les échantillons sont chauffés pendant 3 minutes à 65°C puis déposés avec 1/6 du volume d'une solution de bleu de bromophénol (0.1% bleu de Bromophénol, 50% de sucrose, EDTA 10 mM, tris-HCl 10 mM pH 7.8). La migration se fait entre 50-60 volts et pendant une période de 1½ pour un mini gel, 5-6 heures pour un gel moyen. Les marqueurs des poids moléculaires sont:

λH ₃	:23 kb, 9.4 kb, 6.7 kb, 4.3 kb, 2.5 kb, 1.98 kb, 0.58 kb, 0.125 kb.
φX 174 Hae III	:1353 pb, 1078 pb, 872 pb, 603 pb, 310 pb, 271-280 pb, 234 pb, 194 pb, 118 pb et 72 pb.
Ladder	:12.216 kb, 11.198 kb, 10.180 kb, 9.162 kb, 8.144 kb, 7.125 kb, 6.1 kb, 5.09 kb, 4.07 kb, 3.05 kb, 2.036 kb, 1.635 kb, 1.018 kb, 516 pb/506 pb, 394 pb, 298 pb et 211/210 pb.

Après la migration du gel, la coloration se fait dans une solution de bromure d'éthidium (1 µg/ml) pendant 30 minutes. Ensuite le gel est exposé au U.V (254 nm) et une photo est prise en exposant un film Polaroid type 57 (Kodak) pendant une minute avec un filtre orange sur la lentille.

2.4 CONSTRUCTION D'UNE BIBLIOTHEQUE.

La construction d'une bibliothèque de l'ADN de souris de BALB/c est réalisée à partir du vecteur Embl4. Ce dernier est dérivé de Lambda charon 1059 et consiste en une région 'polylinker' qui a des sites symétriquement placés comme suit : Eco RI, Bam HI et Sal I aux deux bouts d'un 'stuffer fragment' de 14.35 kb (Voir fig 6). Ce vecteur porte les gènes de gam et red dans le 'stuffer

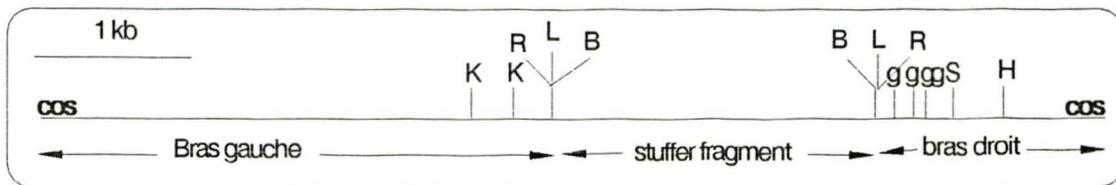


Figure 6: Carte du vecteur Embl4 d'après le Jour. Mol. biol. (1983). Les symboles utilisés sont B(Bam HI), R(Eco RI), K(Kpn I), L(Sal I), H(Hind III), g(Bgl II), S(Sma I). Les sites dans le 'stuffer fragment' n'ont pas été représentés pour simplifier les choses.

fragment'. Durant le clonage ce morceau est remplacé par les fragments d'ADN exogènes. Ainsi une sélection génétique est possible pour les phages recombinants qui sont red⁻ et gam⁻ et ont des phénotypes spi⁻. Ces phages vont donc pousser dans les bactéries comme NM 539 qui est lysogénique pour P2 (Spi⁺). L'ADN génomique de la souris BALB/c a été digéré partiellement avec RI afin d'augmenter la chance d'avoir le segment V et des régions flanquantes assez distantes du V. La digestion partielle a été vérifiée sur gel pour avoir une longueur environ 20 kb. Le produit de la digestion partielle est ensuite ligué avec Embl4 qui a été coupé avec Eco RI. Pour éviter que le fragment du milieu (14.35 kb) se ligue avec le vecteur, on a coupé aussi avec Bam HI. Les bras du vecteur obtenus sont de 22 kb et 9.5 kb.

2.4.1 Ligation

Différents rapports de molarité de l'insert et de bras ont été essayés. Les

rapports de 2:1, 1:1, 3:1 de bras à l'insert ont été utilisés. La ligation est faite dans un tampon 70 mM Tris, 7mM MgCl₂, 10 mM DTT, 1 mM ATP avec 1 unité Weiss de T4 DNA ligase (Biolabs) par 2 µg d'ADN total à liguer. La réaction s'effectue toute la nuit à 15 °C.

2.4.2 Encapsidation in vitro

La méthode de préparation des extraits pour l'encapsidation in vitro (le FTL et le SE) est comme celle décrite par Enquist et Sternberg (Enquist et Sternberg 1979). Les quantités optimales des extraits S.E. et F.T.L ont été préalablement déterminées. Une encapsidation a été faite avec les différents rapports de ligation (voir section 2.4.1). Pour l'encapsidation, on mélange dans l'ordre: 7 µl de tampon A (3 mM MgCl₂, 0.05% mercaptoéthanol, 1 mM EDTA, 20 mM Tris pH 8.0), 2 µl tampon M1 (0.2% β-mercaptoéthanol, 6 mM Tris, pH 8.0). L'encapsidation est faite en incubant pendant 1 heure à la température de la pièce avec agitation légère suivie par l'ajout d'un autre 20 µl d'extrait FTL et d'une autre heure d'incubation. On ajoute ensuite 1 ml de tampon SM (100 mM NaCl, 8 mM MgSO₄, 0.01% gélatine, 50 mM Tris pH 7.5). On titre le produit encapsidé et conserve le reste à 4°C. D'après le 'headfull hypothesis', le vecteur grâce à ses sites Cos et au produit du gène A empacte entre 75 et 105% de son ADN. Dans ce cas, les bras du vecteur qui se lient sur eux-mêmes ne sont pas empactables.

2.4.3 Titrations

Les bactéries NM538 sont utilisées pour l'étalement des phages. Celles-ci sont permissives, c'est à dire, qu'elles ne font pas la distinction entre les phages non-recombinants et recombinants. Par contre, les bactéries NM 539 qui sont

restrictives, permettent la croissance des phages recombinants seulement, à cause de leur phénotype Spi-. Les bactéries NM538 ou NM539 sont mises en culture dans le milieu T-broth contenant 5 mM MgSO₄ et 0.1% de maltose. Lorsque l'a densité optique à 600 nm est égale à 2.4, on centrifuge pour resuspendre le culot dans MgSO₄ (10 mM) de façon à avoir une O.D 600 = 2.0 ou on centrifuge une culture de nuit et le resuspend dans 10 ml de 10 mM MgSO₄ ce qui donne une densité optique 600 nm d'approximativement 2. Pour l'étalement, on fait une préadsorption en incubant les phages avec les bactéries pendant 20 minutes à 37° C. (Les phages recombinants avec NM 539 et le contrôle avec λ wt avec NM 538). On les étale sur des Pétris L-Mg (5mM MgSO₄) avec du T-soft (0.6% agar). Les recettes des milieux L et T broth ont été décrites (Maniatis et al 1982, p440).

2.4.4 Amplification de la bibliothèque

D'après les essais avec différents rapports (voir section 2.4.3), on a vu que la ligation avec rapport 2:1 vecteur//l'ADN a donné une meilleure efficacité que d'autres: un titre de 4.5×10^4 PFU/ μ g. On combine plusieurs matériaux d'emballage pour amplifier la bibliothèque. Trois cc du matériau d'emballage et 1 cc de NM 539 sont utilisés pour la préadsorption en incubant pendant 15 minutes à 37° C. On ajoute ensuite 35 ml de T-soft puis on étale sur une grande boîte de pétri (9"x9") et incube à 37°C pendant une nuit.

2.5 PURIFICATION DE BACTERIOPHAGE ET CONSTRUCTION DU STOCK DE LYSAT DE PHAGE

Trois grandes boîtes de pétri contenant environ chacune 4.07×10^4 PFU sont étalées. On met 50 ml de SM et laisse à 4°C pendant une nuit. Les phages diffusés dans le SM sont récoltés dans les tubes de propylène de 50 ml. Du

chloroforme est ajouté jusqu'à l'obtention d'une concentration finale de 4%. On brasse le tube pendant 15 minutes à 37°C, afin de lyser toutes les bactéries transfectées. Le surnageant est récolté et le volume mesuré. On ajoute NaCl et PEG afin d'avoir une concentration finale de 1 molaire et 10% respectivement. La suspension de phages a été mise sur glace pendant 1 heure et demie puis centrifugées à 7,000 rpm pendant 30 minutes pour précipiter les phages. Les récipients sont vidés et le culot est séché au maximum. Cinq ml de SM sont ajoutés pour dissoudre le culot. L'extraction au chloroforme est faite 3 fois afin d'enlever le PEG qui reste. A chaque fois, la phase aqueuse est récoltée. L'extraction peut être faite plus que 3 fois jusqu'à ce que la phase aqueuse soit devenue claire. Finalement, 0.5 gramme par ml de chlorure de césium est ajouté, dissout puis enfin gardé à 4°C. Les phages sont ensuite purifiés par un gradient de densité en chlorure de césium afin de se débarrasser des capsides, des phages endommagés et du PEG.

On procède en utilisant deux densités différentes:

1. $\rho = 1.35 \text{ g/ml}$ $n = 1.36720$ (environ 15 grammes de CsCl dans 30 ml de SM)

2. $\rho = 1.65 \text{ g/ml}$ $n = 1.39480$ (environ 21.2 grammes de CsCl dans 20 ml de SM)

$\rho =$ la densité, $n =$ indice de réfraction.

La densité de la solution contenant le phage est aussi ajustée à $\rho = 1.15 \text{ g/ml}$ $n = 1.3488$ afin d'avoir les phages à l'interface des deux solutions. Pour faire le gradient, d'abord on met dans un tube de cellulose de nitrate à l'aide d'une pipette pasteur stérile la solution la moins dense. Ensuite on ajoute la plus

dense en la pipétant très lentement en commençant par le fond afin de ne pas nuire à l'interface. La solution contenant des phages est déposée au-dessus très lentement puis centrifugée à 4° C \ 18,000 rpm pendant 2 heures. Après la centrifugation, une bande bleu pâle qui est à l'interface des deux solutions est observée. Celle-ci indique la présence des phages. La bande paraît mieux lorsqu'un objet foncé est placé à l'arrière du tube.

On place un morceau de ruban adhésif au niveau de la bande, et perce un orifice au moyen d'une aiguille de 18G1½ connectée à une seringue de 5 ml. Après l'insertion de l'aiguille dans l'interface, elle est dirigée d'une région à l'autre pendant la succion afin de récolter le maximum de phages. Après que la récolte des phages est effectuée, ceux-ci sont conservés à 4°C. Ces phages sont viables pendant quelques années dans ces conditions .

Le stock est ensuite titré . Après avoir obtenu le titre, 4 grandes boîtes de pétris sont étalées de phages afin d'avoir au moins des recombinants qui vont représenter chaque partie du génome de la souris. Les grandes boîtes sont étalées de phages afin d'avoir 3×10^4 PFU. Le nombre de phages ne doit pas trop excéder le nombre décrit en haut afin que les phages soient séparés l'un de l'autre pour rendre le criblage plus facile. Une sonde PPV 101 qui est dérivée du CE Vk-1C^f avec une partie de la région V est utilisée à cette fin. La sonde commence au site de Pst I dans le V et se termine au site de Pvu II 3' de la région du V. (voir section 1.10 figure 5c). Pour le criblage, un filtre de nylon Dupont (Genescreen) est coupé à la grandeur du pétri et placé sur la boîte. On prend la précaution de bien refroidir les boîtes à 4°C (au moins 2 heures) avant de commencer le transfert avec Genescreen. Pour éviter que l'agar mou ne se décolle pendant le transfert, il est aussi prudent de laisser bien sécher les

grandes boîtes avec les couvercles entrouverts avant de couler l'agar mou. Pendant le transfert, la feuille de 'Genescreen' a été laissée pendant 3 minutes et à l'aide d'une aiguille les coins de repère ont été marqués asymétriquement. Pour chaque boîte, on a fait un duplicata et on a laissé le filtre sur la boîte pendant 6 minutes.

Les filtres ont ensuite été placés dans des solutions, dénaturante (1.5 M NaOH, 0.4 M NaCl), neutralisante (3 M NaCl, 2 M Tris HCl pH 8.0) et stabilisante (2 x SSC). Pour chaque solution, un trempage de 2 minutes est effectué. Avec le genescreen, il n'est pas nécessaire de fixer l'ADN sur le filtre avec le séchage sous vide.

Les taches radioactives qui sont présentes en double sont repiqués et étalés de nouveau dans des pétris de 85 mm. Encore une fois, le transfert a été fait. Cette fois-ci, le filtre de 'Genescreen' a été exposé sur le pétri pour l'adhésion des phages pendant une période de 2 minutes pour le premier filtre et le duplicata pendant une période de 4 minutes.

2.6 EXTRACTION DE L'ADN DE PHAGE

Le processus de repiquage est continué jusqu'à ce que toutes les plages se trouvant sur le filtre donnent des signaux positifs avec la sonde. A ce moment, on peut dire que les clones criblés sont purifiés. Une plage du pétri purifié à 100% est sélectionnée puis prélevée avec un cure-dent stérile. Elle est ensuite mise dans 1 ml de SM pour faire diffuser les phages soit pendant une nuit à 4°C ou 1 heure à 37°C. Cent µl de NM 539 dans du MgSO₄ (D.O 600 = 2) sont ajoutés aux phages. La quantité de phages nécessaire pour avoir une lyse complète sur pétri est calculée. Si on prend pour acquis que 1 plage = 10⁴

phages pour une lyse complète on peut calculer la quantité de la solution contenant les phages qu'il faut ajouter. La préadsorption et l'étalement se sont faits comme décrit précédemment. Afin de ne pas inhiber l'action des enzymes de restriction, l'agar est substitué par l'agarose. Quant à l'agar mou en surface, au lieu de T-soft on a utilisé l'agarose à 0.7% pour la même raison que mentionnée plus haut. Après l'incubation sur pétri, 5 ml de SM sont ajoutés sur ces lyses complètes et laissé à 4°C pendant une nuit pour laisser diffuser les phages. Le milieu SM est ensuite récolté et transféré dans tube à centrifuger de 15 ml en propylène pour être centrifugé à 3,500 rpm pendant 15 minutes à 4°C. On ajoute de la RNase A et de la DNase 1 au surnageant afin d'avoir une concentration finale de 1 µl \ ml pour chacune. Le surnageant est alors incubé à 37°C pendant 30 minutes. Ensuite, un volume égal d'une solution composée de 20% WV PEG et 2 M NaCl dans SM est ajouté, mélangé par inversion et laissé pendant 1 heure à 0°C. Les phages sont ensuite récoltés par centrifugation à 8,000 rpm pendant 20 minutes à 4°C. Le surnageant est enlevé puis le tube est laissé inversé pour bien l'égoutter. On ajoute 0.5 ml de SM pour resuspendre les bactériophages: Le liquide est transféré dans des tubes Eppendorf de 1.5 ml et centrifugé à 4°C à 3,500 rpm. Le surnageant est transféré dans un autre tube Eppendorf. Cinq µl de 10% SDS et 5 µl de 0.5 M EDTA (pH 8) sont ajoutés et le tout est incubé à 68°C pendant 15 minutes. Après l'incubation, le tube est laissé à refroidir à la température de la pièce. Puis, une extraction au phénol suivie d'une extraction au phénol\chloroforme (1:1) et une extraction au chloroforme sont effectuées les unes après les autres. A la fin de l'extraction, un même volume d'isopropanol est ajouté à la phase aqueuse pour précipiter l'ADN. La précipitation est faite à -70°C pendant 20 minutes après

quoi le tube est laissé à dégeler puis centrifugé à 4°C pendant 15 minutes. Le culot est lavé avec de l'éthanol 70% puis séché sous vide pendant 3 minutes. Ce culot est dissout dans 50 µl de T.E.

2.7 PREHYBRIDATION

Le cocktail de préhybridation est constitué de 1% SDS, 1 M NaCl, 10% sulphate de dextran et est gardé à 65°C pour empêcher la précipitation. La concentration finale du sulphate de dextran doit être de 10% . La membrane 'Genescreen plus' est ensuite placée dans un sac de plastique (Kapak corporation) pour la préhybridation. Pour un sac de plastique de la grandeur 6 1/2" x 8", 15 - 20 ml de cocktail préhybridation (1% SDS, 1M NaCl, 10% sulphate de dextran chauffé et gardé à 65°C). Le sac est fermé hermétiquement et incubé pendant 15 minutes ou plus (jusqu'à 6 heures) à 65°C avec une agitation constante.

2.8 MARQUAGE DE LA SONDÉ

Une sonde à partir d'un sous-clone dans le phage M13 est faite d'après le protocole de Church et Gilbert (1984). L'insert dans le phage M13 est radioactivement polymérisé de la façon suivante. Après avoir hybridé l'amorce M13 (GTAAAACGACGACGGCCAT) avec la sonde M13, la polymérisation est faite dans une solution de 1 µl 100 mM DTT, 0.5 µl spermidine 80 mM, 0.5 µl de 2 mM dGTP, dATP, dTTP, 5 µl de dCTP radioactif (3000ci/mmol) et 1 µl (4U/µl) DNA polymerase E. Coli (klenow large fragment polymerase d'Amersham).

Pour faire une sonde à partir d'un sous-clone dans le plasmide M13, on doit dénaturer l'ADN afin d'obtenir un fragment simple brin qui servira de matrice. La dénaturation se fait comme suit: l'ADN est chauffé à 100°C pendant 4 minutes

puis placé dans l'azote liquide pendant 30 secondes. Ensuite, la réaction d'amorce et de polymérisation avec les nucléotides radioactifs est faite comme avec la sonde simple brin de M13 phage. Après avoir laissé la polymérisation se faire pendant 1 heure à la température de la pièce, la réaction est arrêtée avec 1/5 volume de 250 mM NaOH, 5 mM EDTA, 50% glycérol. La sonde est dénaturée par la suite à son tour sur un gel dénaturant consistant de 0.45 gm d'agar à basse température de fusion (BRL) dans 30 ml d'une solution 50 mM NaCl, 1 mM EDTA. Ce protocole est adapté pour l'appareil BRL avec des moules de 5 puits. Le gel est laissé équilibré dans un tampon d'électrophorèse (30 mM NaOH, 1 mM EDTA) pendant un minimum de 30 minutes. La réaction est bouillie dans un bain pendant 2 minutes puis refroidie sur glace avant de mettre sur gel pour l'électrophorèse à 40 volts pendant 2 heures. Un voltage plus grand que 40 volts risque de faire fondre le gel. La position de la sonde radioactive est localisée par autoradiographie après exposition de 3 minutes. La partie du gel contenant la bande correspondant à la sonde est coupée, fondue puis comptée.

Dans la préparation d'une sonde d'oligonucléotide, on dissout le contenu dans de l'eau stérile afin d'avoir une concentration finale de 10 pmoles / μ l. Huit μ l de cette solution sont ajoutés à 5 μ l de 10x tampon 1 de la kinase (0.5 M tris HCl pH 7.6, 0.1 M MgCl₂, 50 mM DTT, 1 mM spermidine, 1 mM EDTA, 13 μ l de (γ -³²P) ATP(3000 ci/mmmole), 22 μ l d'eau pour compléter le volume à 50 μ l et 4 μ l T4 de kinase (7.5 U/ μ l) de Pharmacia.

Pour l'ADN génomique, on utilise 1x 10⁶ cpm par ml de solution de préhybridation et environ 100 ng/ml d'ADN de sperme de saumon dénaturé par sonication. La sonde est dénaturée dans un bain à 100°C pendant 10 minutes

avec le sperme du saumon puis transférée dans le sac en plastique contenant la membrane. L'hybridation se fait à 65°C avec agitation constante pendant toute la nuit.

2.9 LAVAGES DES FILTRES

Après hybridation, on fait subir aux filtres de 'Genescreen' différents lavages. Les deux premiers lavages sont répétés 2 fois pendant 5 minutes dans 2 x SSC (0.3 M NaCl, 0.03 M Na citrate), suivis d'un lavage pendant de 30 minutes dans 2 x SSC, 1% SDS à 65°C. Pour les lavages dans des conditions sévères d'hybridation, le troisième lavage est fait dans 0.1 x SSC, 0.5% SDS à 65°C. Mais pour le lavage dans des conditions moins sévères, le lavage dure 30 minutes dans 0.5 x SSC, 0.5% SDS. Le dernier lavage est effectué dans 0.1 x SSC pendant 30 minutes et peut être répété encore une autre fois selon le besoin. Le filtre est laissé à sécher pour enlever seulement le surplus d'humidité mais pas complètement juste au cas où on voudrait réhybrider le filtre. Le filtre est ensuite enveloppé dans une pellicule de Saran wrap puis exposé à un film Kodak XAR-5 à -70°C pendant une nuit ou plus.

2.10 PREPARATION DE CELLULES XL-1.

On a fait des stries à partir des cultures de cellules XL-1 congelées dans du glycérole sur des pétris tétracycline à fin de sélectionner les cellules qui ont leur épisome. En effet, l'épisome contient le gène pour la complémentation du gène lac et le marqueur de résistance à la Tétracycline. On ensemence dans 10 ml du milieu liquide SOB (voir texte ci-dessous pour la composition) une colonie qui a poussé dans le pétri tétracycline et on incube à 37°C pendant 16 à 18 heures. Cinq cent μ l de cette culture de nuit est utilisée pour réensemencer 120 ml de SOB contenant 2 molaires de Mg^{++} qu'on laisse pousser jusqu'à D.O550 =0.45.

supplémentaire afin de ne pas contaminer les autres cultures. Trente ml de culture dont la densité optique à 550 nm est 0.45 sont mis dans des tubes à capuchons polycarbonate et laissé sur glace pendant 10 minutes. Puis on centrifuge à 6,000 rpm dans une centrifugeuse Sorvall, rotor HB4 pendant 10 minutes à 4 °C. Le culot est resuspendu dans un tiers du volume original dans F.S.B froid. On laisse sur glace pendant 10 minutes puis on recentrifuge et resuspend avec 1\12.5 du volume initial de F.S.B. Du DMSO est ajouté pour avoir une concentration finale de 3.5%. On laisse sur glace pendant 5 minutes et on ajoute encore du DMSO pour avoir une concentration finale de 7%. Le tube est laissé sur la glace pendant 5 minutes. Par la suite, on les divise dans des tubes à culture avec capuchons (falcon 2059) en aliquots de 200 µl. On congèle par la suite les cellules dans un mélange éthanol et glace sèche ou faute de mieux dans l'azote liquide. Les tubes sont gardés à -80°C. Les cellules se conservent pendant au moins 6 mois dans ces conditions sans perdre leur efficacité.

Les compositions des milieux sont:

S.O.B. par litre:

Bactotryptone	20 gm
Extrait de levure	5 gm
NaCl	10 mM;
KCl	2.5 mM

(Stérilisation à l'autoclave)

Une solution de 1 M $MgCl_2$ + 1 M $MgSO_4$ stérilisée par filtration est ajoutée à 100 ml de cette solution S.O.B.

S.O.C.: Même préparation que S.O.B. mais seulement avant l'utilisation, on ajoute 1 ml de 2 M glucose stérilisé par filtration pour chaque 100 ml de S.O.C.

F.S.B. pour 500 ml:

Potassium Acétate 1 M	5 ml
Glycérol	50 ml
KCl	3.728 gm
MnCl ₂ .4H ₂ O	4.453 gm
CaCl ₂ .2H ₂ O	0.735 gm
HACoCl ₃ (chlorure de hexamine de cobalte III)	0.401 gm

(Filtrer pour stériliser et garder à 4°C. Le F.S.B. peut se conserver à cette température pendant 6 mois)

2.11 CLONAGE DANS LE PLASMIDE M13+

Le plasmide M13+ (Stratagene, San Diego, CA) contient des sites de multiclonaage contenant les sites de restriction permettant l'incorporation de fragments d'ADN dans le sens de synthèse désiré. Un choix de l'amorce nous permet aussi d'approcher certains sites de l'insert au site de l'amorce sans avoir besoins de resous-cloner pour séquencer ces régions. Ce plasmide peut recevoir des inserts jusqu'à approximativement 5 fois sa taille. En plus de cela, il contient l'opéron lac qui en présence de X-gal et IPTG donne des colonies bleues si on les transforme dans des cellules hôtes lac-(XL-1, JM109). Cependant si un insert est incorporé dans le site de multiclonaage dans ce cas le gène lac sera alors dissocié et deviendra non-fonctionnel et les colonies obtenues, seront blanches.

2.12 LA TRANSFORMATION.

Les tubes de cellules XL-1 sont dégelés sur glace pendant 10 minutes tout en les brassant doucement chaque 2 minutes. La quantité de l'ADN de la ligation qu'on va mettre, est la quantité calculée en terme de vecteur M13+ . On

utilise environ 1 nanogramme d'ADN pour 200 µl de cellules. Après avoir mis le produit de la ligation dans les tubes, on les laisse sur glace pendant 30 minutes. Ensuite, un choc thermique à 42°C est effectué pendant 45 secondes et on les laisse sur glace pendant 2 minutes. Neuf cent microlitres du milieu S.O.C est ajouté puis incubé à 37°C pendant 1 heure en brassant à 225 rpm dans un bain à agitation. Vingt µl des cellules sont étalées par pétri (85 mm) de Y.T ampicilline (8 gm Bacto tryptone; 5 gm NaCl; 5 gm d'extrait de levure par litre) On étale au préalable sur les pétris 80 µl X-gal 2% et 20 µl d'IPTG 500mM. Les pétris sont ensuite incubés à 37°C pendant 16 à 18 heures.

2.13 IA SELECTION DE GRANDEUR D'INSERT PAR MICRO-PREPARATION D'ADN.

Avec un cure-dent stérile, une colonie blanche est déposée en stries sur le pétri contenant un milieu d'ampicilline. Il est parfois utile de faire des stries sur pétri avec IPTG et X-gal pour éliminer les faux positifs (les colonies qui pour une raison ou une autre donne une coloration blanche sans que le plasmide ait un insert). Après l'incubation pendant une nuit à 37°C, les 2/3 de la hauteur de la strie de la colonie sont récoltés puis resuspendre dans 40 µl de tampon STE (100 mM NaCl, 20 mM Tris HCl pH 7.5, 10 mM EDTA) dans un tube Eppendorf. 40 µl de phénol\chloroforme est ajouté pour faire l'extraction. Au surnageant, 1µl de RNase A (1mg/ml) est ajouté et la solution est laissée pendant 2 minutes à la température de la pièce. Au 20 µl de cette solution, 4 µl de BPB sont ajoutés et le tout est mis sur un gel. Le restant est gardé à 4°C. La hauteur de l'insert est jugée d'après le gel en comparaison avec le plasmide seul sans insert. Le sous-clone qui nous intéresse est ensuite pris à partir de la strie qui reste sur le pétri original. Il est à nouveau étalé en stries sur des pétris contenant un milieu à

l'ampicilline et on choisit une colonie pour faire des préparations rapides de l'ADN de plasmide d'après le manuel de Maniatis . (Maniatis et al 1982, p 368-369).

2.14 SEQUENCAGE

Après avoir obtenu l'ADN de plasmide on séquence d'après la méthode de Sanger (Sanger et al 1977) en utilisant la trousse de 'Stratagene' et la méthode décrite dans le manuel de la compagnie (Stratagene , San Diego , CA). Pour avoir une séquence plus longue, on utilise la séquenase à la place de l'enzyme klenow parce que la processivité est plus haute et l'activité exonucléasique 3'->5' est moindre qu'avec la séquenase (United states Biochemical Corporation , Cleveland, Ohio). Ceci nous permet de déterminer une séquence jusqu'à 400 pb.

2.14.1 Réaction d'amorce sur plasmide M13+

Pour se faire, on prend le sous-clone dans M13+ (2.0 µg) qu'on veut séquencer et on ajoute 1 µl de l'amorce choisie (2.5 ng/µl) et 2 µl de tampon de séquence 5x(200 mM tris HCl pH 7.5, 50 mM MgCl₂, 250 mM NaCl). On chauffe le mélange pendant 4 minutes dans l'eau à 100°C et on met immédiatement dans l'azote liquide pendant 30 secondes. Ensuite, après la dénaturation de cette manière, on laisse le mélange à la température de la pièce pendant 20 minutes pour l'hybridation avec l'amorce.

2.14.2 Réaction pour le marquage et terminaison du brin qu'on veut séquencer

Après l'hybridation de l'amorce, on ajoute 1 µl de 0.1 M de DTT, 2 µl de 5X du mélange pour le marquage (7.5 µM dGTP, 7.5 µM de dCTP, 7.5 µM de

dTTP) dilué 1 dans 4 avec de l'eau, 1 μl $\alpha^{35}\text{S}$ dATP (1000 ci/mM) dilué 1 dans 2 avec de l'eau et le séquenase (12 U/ μl) diluée 1 dans 8 avec du T.E (10 mM Tris.Cl pH 8.0, 1 mM EDTA) préalablement refroidi sur glace. Le mélange est incubé à la température de la pièce pendant 1 minute. Pour la terminaison de la synthèse, on prélève 3.5 μl de cette solution et on les place dans tube A qui contient déjà 2.5 μl du mélange A (80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddATP) qui est préchauffé pendant au moins 1 minute à 45°C. On procède de la même manière pour les réaction C,G,T sauf que le 2.5 du mélange A est remplacé par le mélange C (80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, et 8 μM ddCTP), 2.5 μl du mélange G (80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddGTP) et 2.5 μl du mélange T (80 μM dGTP, 80 μM dATP, 80 μM dCTP, 8 μM ddTTP) respectivement. On laisse réagir à cette température pendant 5 minutes et on ajoute 4 μl de solution d'arrêt de la réaction (95% formamide, 20 mM EDTA, 0.05% BPB, 0.05% xylène cyanol FF).

2.14.3 Gel de séquence

Les gels de séquence contiennent 6% d'acrylamide avec gradient ou sans gradient de sel dépendamment si on veut compacter les bandes ou si on veut obtenir une lecture plus grande que 250 pb. Ces gels mesurent 45 cm x 32.5 cm et 0.35 mm d'épaisseur et sont placés sur un appareil Fotodyne (New Berlin , Wis). Deux à trois μl des réactions A,C,G et T sont déposés respectivement dans cet ordre dans les puits. Les puits sont nettoyés préalablement avec un tampon TBE (0.1 M Tris-Borate pH 8.3, 2 mM EDTA afin d'enlever les cristaux d'urée. On applique un courant de 2000 volts, 42 mA

et 80 watts pour faire migrer le gel de séquence dans un tampon TBE (2 mM EDTA, 0.1 M tris-borate, pH 8.5) pendant 2 1/2 heures pour des gels avec gradient et pendant 4 à 5 heures pour des gels sans gradient. Après la migration le gel est démoulé en soulevant tranquillement une des plaques en vitres. La plaque ayant le gel est mise dans une solution 10% d'acide acétique et 10% méthanol pendant 30 minutes. afin de fixer le gel . Le gel est ensuite déposé sur du papier Whatman 3M puis séché à 175°C sur un séchoir à photo pendant 90 minutes. Le gel est ensuite exposé à un film (35x43 cm) XRP1 de la compagnie Eastman Kodak (Rochester, N.Y) pendant 1 ou plusieurs jours.

Composition du tampon et du gel avec et sans gradient de sel.

Composition pour 2 litres du tampon 10x TBE				
Tris		242 gm		
Acide Borique		98 gm		
EDTA .2H ₂ O		11.5 gm		
	Volume	urée	10x TBE	Acrylamide 38%
				Bisacrylamide 2%
Gel 6% d'acrylamide	100 ml	46 gm	10 ml	15 ml
sans gradient				
Gel 6% d'acrylamide	100 ml	46 gm	50 ml	15 ml
avec gradient (avec 10 gm de sucrose et 5 mg de BPB)				

2.14.4 Lecture et compilation des données

L'autoradiogramme est lu de bas en haut (voir fig. 15). Les séquences obtenues sont compilées par dans l'ordinateur en utilisant le programme de M13 développé par le Dr. D. Gibson. Ce programme est conçu spécialement pour faire la comparaison avec différentes séquences des gènes Vk, établir l'alignement entre différents sous-clones, trouver des sites de restriction et chercher des séquences d'un certain intérêt parmi des séquences connues. Pour les alignements optimaux avec les séquences de 'GenBank', le programme NUCALN (de D.J. Lipman et W.J. Wilbur) est utilisé. Les séquences obtenues sont aussi comparées avec les séquences des autres gènes d'immunoglobuline des autres groupes de Vk avec une grande rapidité en utilisant le programme DIAGONAL de ATARI ST employant GFA-Basic où les régions homologues sont indiquées directement sur l'écran. Chaque point sur l'écran représente 4 nucléotides qui se concordent dans les séquences. Les régions qui sont homologues, forment un tracé diagonal et on peut imprimer les résultats. La comparaison et la recherche des séquences répétitives ou des séquences LINE-1(L1Md) dans les séquences entrées sont faites avec le même programme.

2.15 LA CONSTRUCTION D'UNE CARTE DE RESTRICTION.

L'ADN du phage EB14 est digéré partiellement avec l'enzyme Hind III (20 U/ μ l), Eco RI (20 U/ μ l) et Bam HI (10 U/ μ l) dans un volume final de 30 μ l par digestion. 0.5 μ g d'ADN est utilisé pour chaque clivage. Des aliquotes de 3 μ l sont prélevées aux temps suivants 0, 2, 5, 10, 15, 20, 30, 35, 40 et 60 minutes et déposées dans un tube contenant 1 μ l de 500 mM EDTA et refroidi sur glace pour arrêter la réaction. On combine ensuite les contenus de tous les tubes des différents temps dans un seul et les met sur gel. Le but de ce mode opératoire

est celui d'empêcher qu'un site qui se digère préférentiellement à un certain moment comparé à d'autre ne soit surreprésenté. Après avoir fait une réplique du gel, on hybride avec une sonde ON-R qui consiste en une sonde oligonucléotide de 12 pb (dGGGCGGCGACC) long et complémentaire à la terminaison Cos du bras droit du phage lambda. La taille des bandes qui hybrident avec la sonde est mesurée et les sites sont localisés.

3 RESULTATS

A partir de la bibliothèque de l'ADN de souris BALB/c coupée partiellement avec l'enzyme Eco RI et liguée dans le phage EmbI4, 1.2×10^5 phages ont été criblés et nous avons ainsi obtenu 21 clones positifs qui ont hybridé avec la

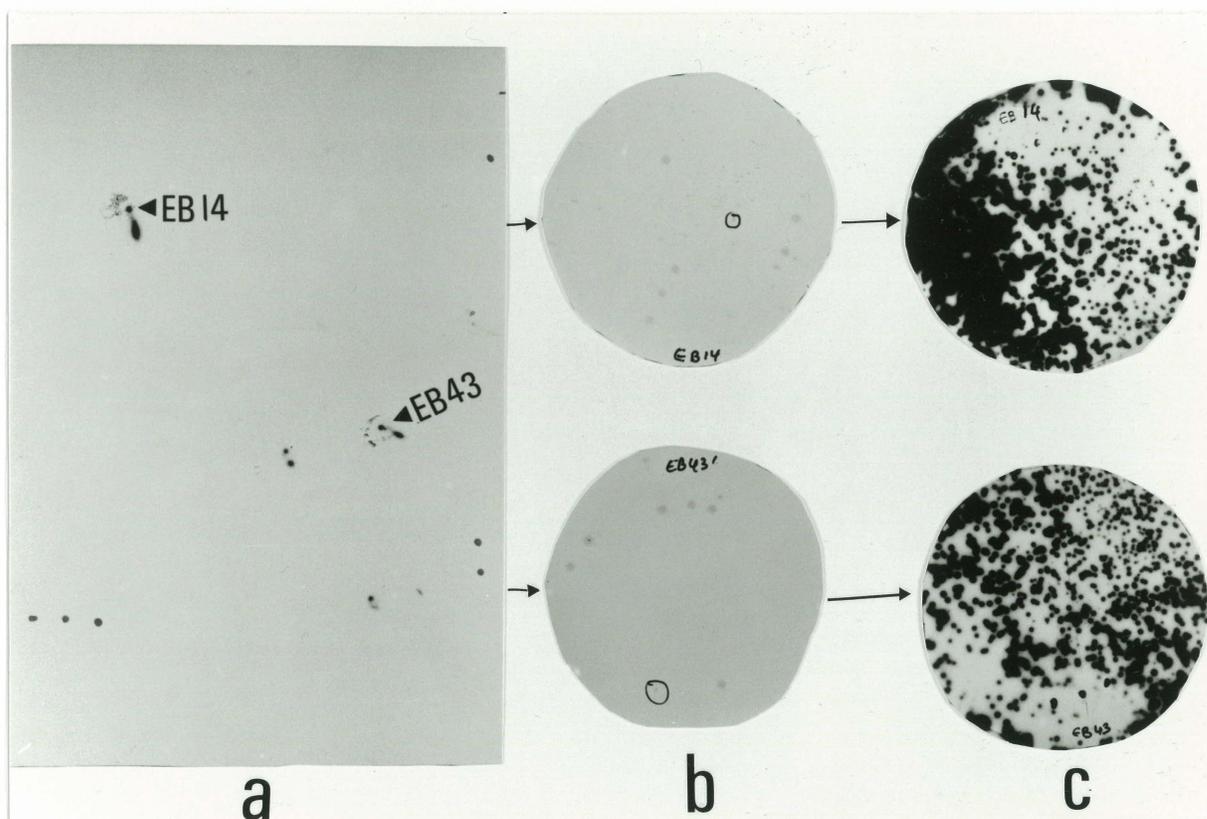
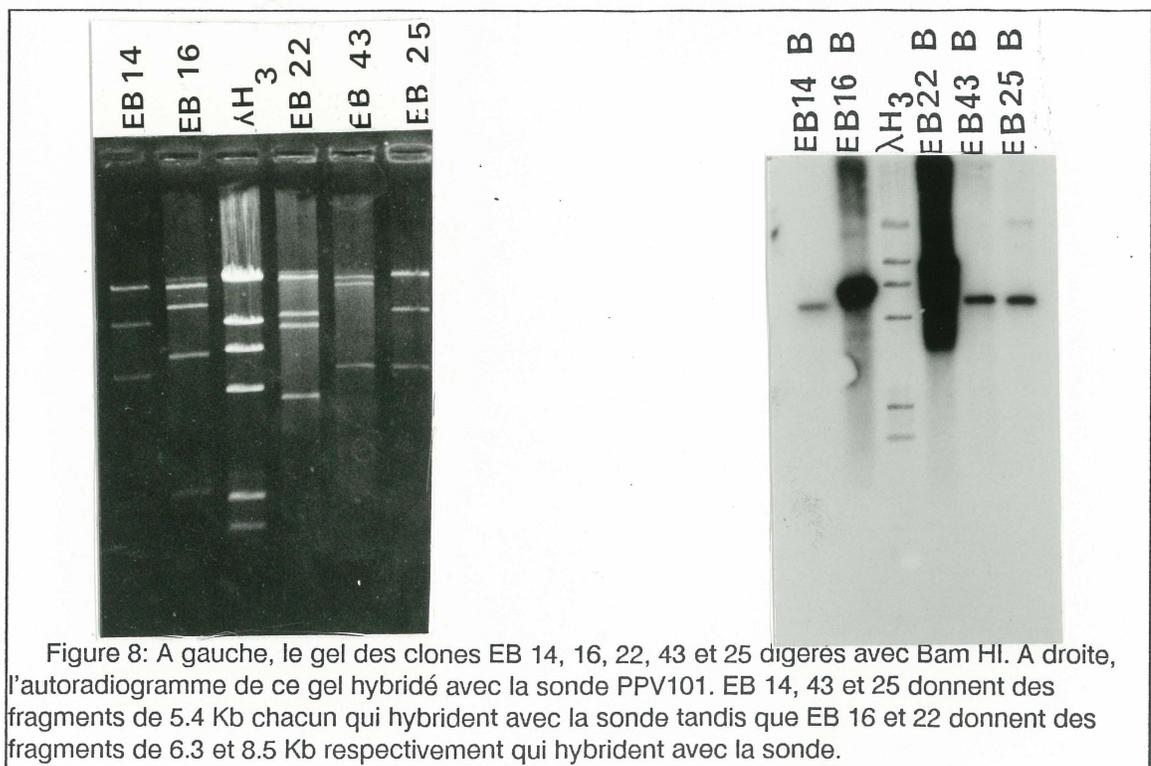


Figure 7: Différentes étapes de purification de deux clones.

- a) L'identification des clones sur des grandes boîtes de culture (9 pouces x 9 pouces).
 - b) Purification dans des pétris (85mm). Cette étape doit être répétée au moins deux autres 2 fois.
 - c) Les phages purifiés hybridant à 100% avec la sonde.
- Les transferts sur membrane de nylon pour les passages ont été faits en duplicata.

sonde de la région V (PPV101). Les clones ont été repiqués et étalés de nouveau jusqu'à ce qu'ils soient purifiés (100% des plages qui hybrident avec la sonde). Pour ce faire il a fallu procéder à 3 passages (Fig 7). Des préparations rapides de d'ADN de phage digérés avec les enzymes Bam H I et Eco R I, nous ont permis d'éliminer les clones ayant des fragments de 8.5 kb ou 6.3 kb Bam H I (déjà clonés) qui hybrident avec la sonde PPV101. Parmi les autres un seul identifié comme étant EB42 était un faux positif hybridant seulement avec une sonde de M13 phage sans insert. On a conclu que c'était dû à un contaminant. Les 4 autres clones qui restaient, soit EB25, EB14, EB43 et EB45 ont tous donné des fragments 5.4 kb BamH I qui hybrident avec la sonde PPV101 (Fig 8). Etant donné que EB45 n'a été isolé que plus tard, il n'apparaît pas dans la photo ci-dessous.



A partir des résultats obtenus, nous avons procédé ensuite à des opérations de sous-clonage. Ainsi les fragments 5.4 kb Bam H I des clones EB25, EB43 et EB14 ont été ligués dans le vecteur M13+ digéré avec l'enzyme Bam HI. La quantité de sous-clones recombinants ayant un insert de 5.4 kb pour les 3 clones était nombreuse. Le nombre de sous-clones suivant a été obtenu: EB14=3, EB25=5 et EB43=3 (tableau 2) .

Tableau 2: Ccorrespondance des différents clones.			
Les clones dans le phage Embl4 hybridant avec PPV101	Les sous-clones dans M13+ ayant 5.4 Kb	Les sous-clones qui contiennent le bon fragment de 5.4 Kb	Les bons sous-clones coupés avec Xbal puis ligués et retansformés
EB 14 (11 Kb)	EB 14-15 EB 14-19 EB 14-6	-----> DCD	-----> DCDX
EB 25 (12.5 Kb)	EB 25-1 EB 25-7 EB 25-8 EB 25-10 EB 25-14	-----> EB 25	
EB 43 (12.5 Kb)	EB 43-4 EB 43-8 EB43-25	-----> BA43	-----> 5.3
EB 45 (19 Kb)	semblable à EB 43		

Etant donné le grand nombre de sous-clones ayant la grandeur du fragment voulu (Fig 9), on peut croire que cela est dû à l'ADN bactérien présent pendant la préparation de l'ADN de phage ou à des sous-clones semblables. Le problème du nombre nous a obligé à éliminer les sous-clones semblables. Nous avons donc pour cela digéré les sous-clones avec les enzymes Xba I et Bgl II (Fig 10). Suite aux digestions, les sous-clones uniques ont été utilisés comme

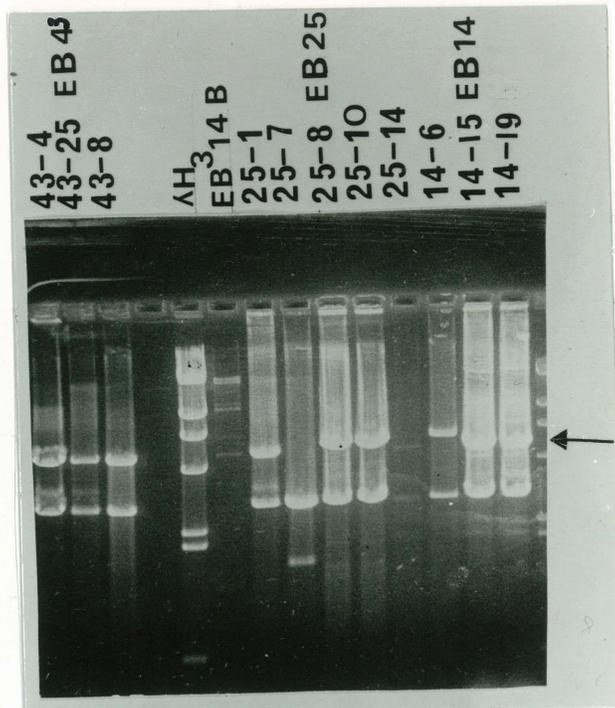
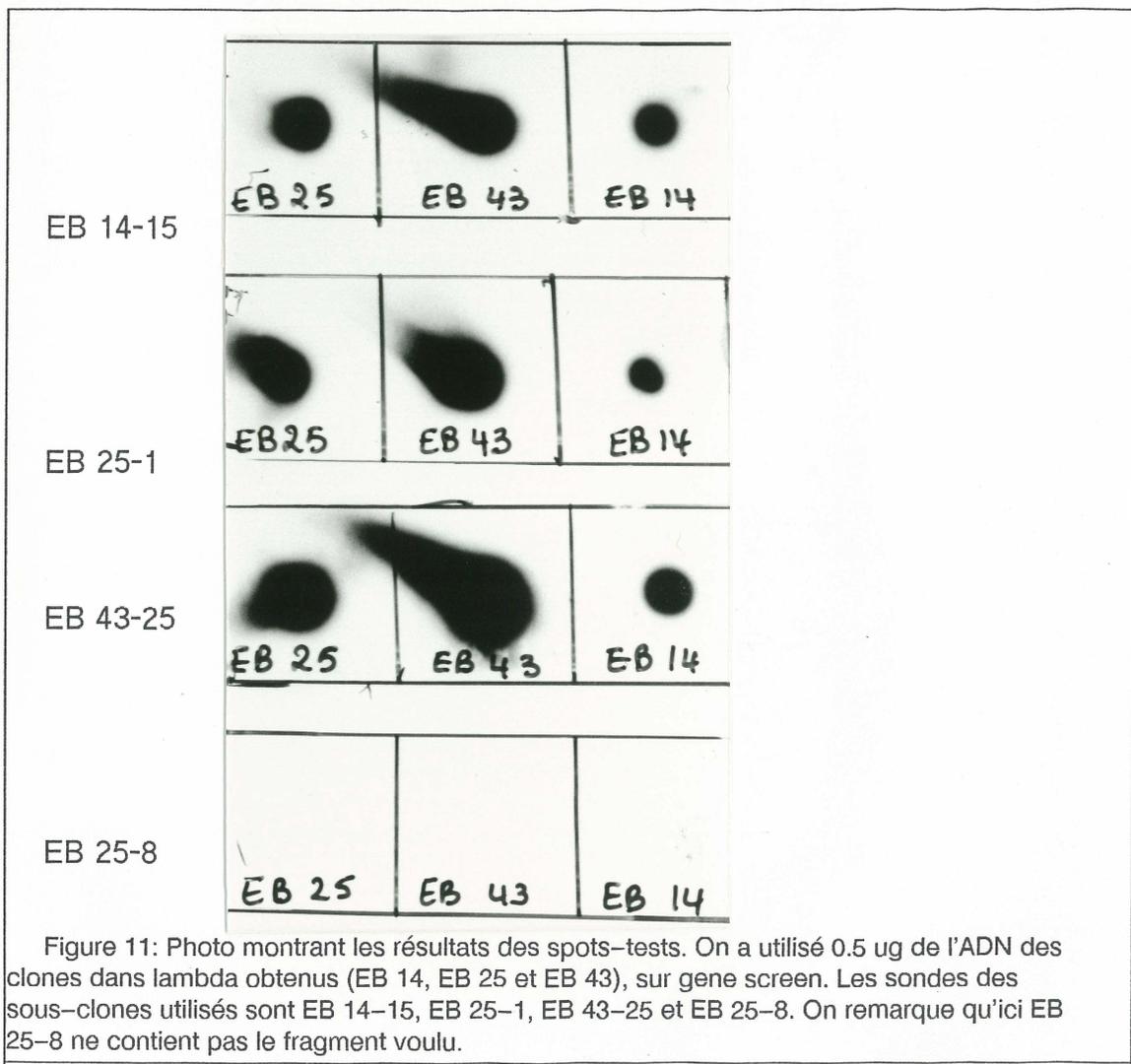


Figure 9: Photo démontrant les sous-clones obtenus ayant un fragment de 5.4 après avoir été digérés avec Bam H1(B) (indiqué par la flèche). Les nombres de clones obtenus sont EB 43=3, EB 25=5 et EB 14=3. Comme contrôle de grandeur le clone EB 14 digéré avec Bam H1(B).



Figure 10: La digestion des différents sous-clones obtenus dans M13+ avec Bgl II(Bg) et Xba I(X) afin d'identifier ceux qui sont différents. Seulement ceux qui ne se ressemblent pas ont été utilisés comme sonde. CONT: C'est le control d'un des sous-clones non-digérés. On remarque que EB 14-6, 14-15, 14-19 sont identiques tandis que EB 25-7 et EB 25-8 sont identiques.

sonde pour des spot-tests sur l'ADN des phages des clones en question. Nous n'avons pas utilisé la sonde PPV101 pour cribler les sous-clones parce qu'il était à prévoir que la sonde ayant été faite à partir du phage M13 allait hybrider fortement avec le plasmide M13+ et donc risquait de ne pas donner un signal très évident pour la sélection. Cela pourrait arriver malgré les précautions prises pour limiter la quantité des nucléotides radioactifs pendant la synthèse de la sonde. Cette difficulté étant surmontée, les résultats obtenus (Fig 11) indiquent que les clones EB14-15, EB 25-1 et EB 43-25 contenaient effectivement l'insert



désiré. Désormais on nommera le sous-clone EB 14-15, DCD et EB43-25, BA 43 . Etant donné qu'on a conservé par mesure de sécurité les sous-clones EB 25-1 et EB 43-25, il s'est avéré, que suite aux digestions ils se sont montrés semblables. Aussi , compte tenu de cela, nous ne retiendrons désormais que le sous-clone BA 43.

3.1 STRATEGIE DE SEQUENCAGE

Après s'être assuré d'avoir obtenu les sous-clones désirés, on a pu commencer la stratégie de séquençage. Le but principal de ces opérations était d'obtenir les séquences de la région V, le leader et de la région 5' de la région du segment V contenant les séquences de régulation de l'expression du gène (pentadecanucléotides, octamère, la boîteTATA). Sachant que les sites de Pst I, Kpn I et Bgl II sont presque toujours présents dans la région du segment V, nous avons cherché à obtenir des sous-clones qui commencent directement à

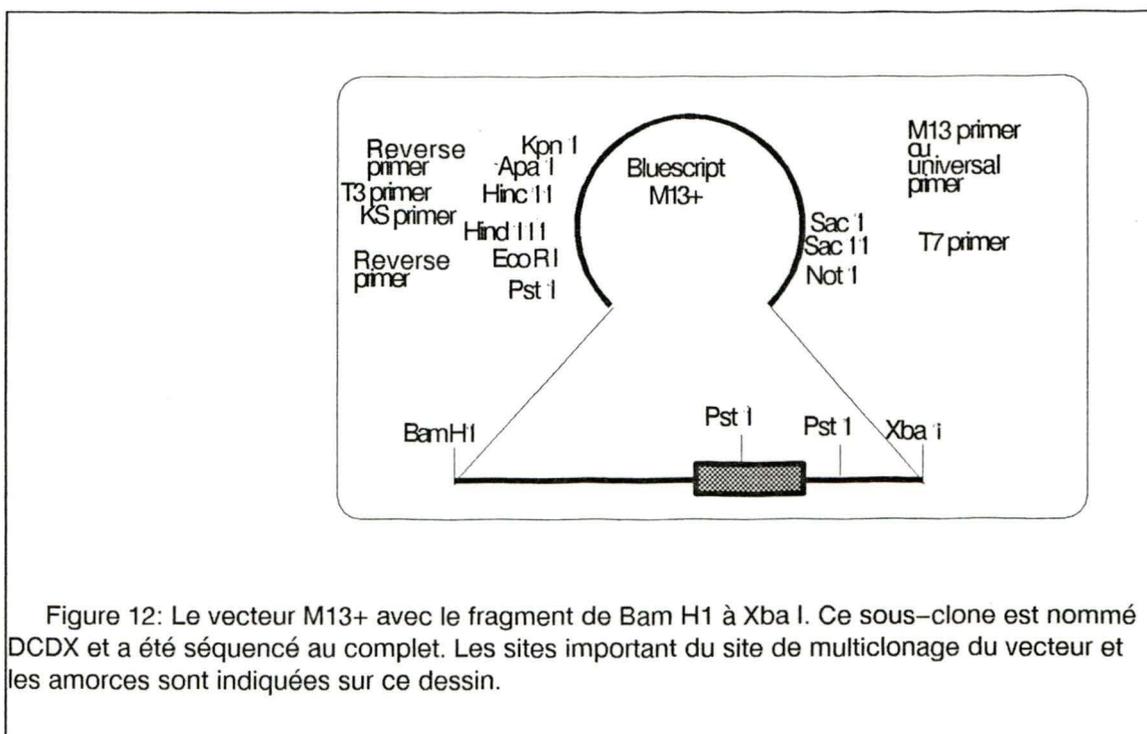
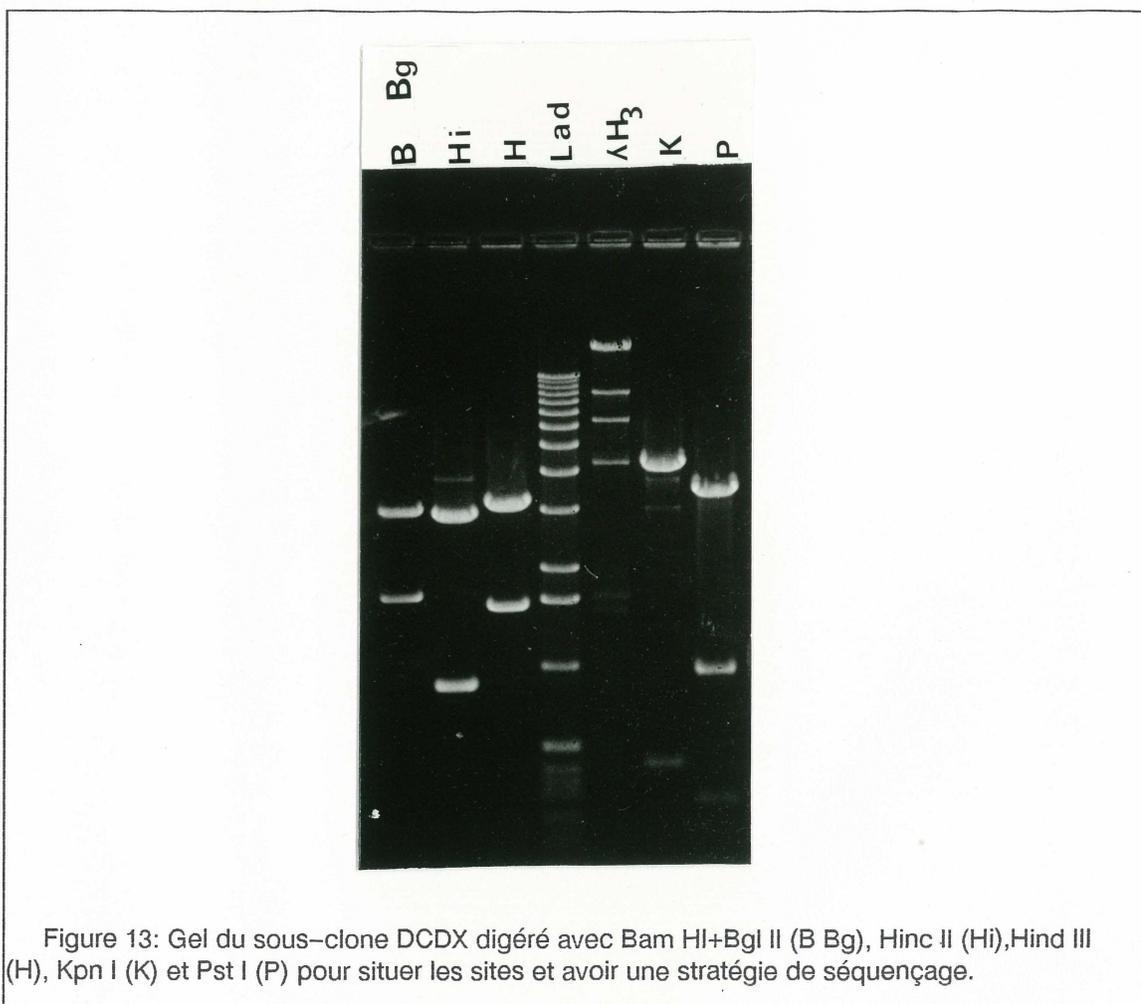


Figure 12: Le vecteur M13+ avec le fragment de Bam H1 à Xba I. Ce sous-clone est nommé DCDX et a été séquençé au complet. Les sites important du site de multiclonage du vecteur et les amorces sont indiquées sur ce dessin.

ces sites en procédant à la digestion des sous-clones avec Pst I, Bgl II et Kpn I. Par la suite, le produit de ces digestions a été ligué et transformé de nouveau. Cette stratégie nous a effectivement permis d'obtenir le clone EB 14 digéré partiellement avec Pst I (DCDP) dans la région V (Fig 14). L'étape suivante a consisté à chercher à obtenir des fragments contigus qui se chevauchent. Pour ce faire, nous avons digéré avec différents enzymes comme Xba I, Hind III, Eco RI, Bgl II puis reliqué afin d'avoir des clones qui contiennent la région V et la partie 5' flanquante du segment V. Le clone DCDX (Fig 12) donne la région qui



contient la région V et la partie 5'. On s'est alors concentré sur ce fragment. Les digestions Pst I, Hinc II, Hind III et Bgl II ont été faites afin de situer leurs sites. Les fragments obtenus sont indiqués sur le tableau 3 et sur le gel de la figure 13.

Enzyme	Fragments résiduels +vecteur M13+	Fragments libérés
Pst I	3.7 kb	600 bp, 300 bp
Kpn I	4.3 kb	500 bp
Hinc II	3.7 kb	872 bp, 100 bp
B + Bgl II	3.0 kb	1.7 kb

3.2 REGION SEQUENCEE

La partie qui nous intéresse, c'est-à-dire celle contenant la région V et la région 5' du V se trouve dans le clone DCDX (Fig 12 et 14). Profitant des sites de multiclonage de M13+, sans avoir besoin de sous-cloner à nouveau les fragments, nous avons fait digérer avec certains enzymes ayant des sites dans le site de multiclonage et dans l'insert. Ainsi en religuant et transformant ce plasmide, on a pu avoir des séquences contigues et qui se chevauchaient sans

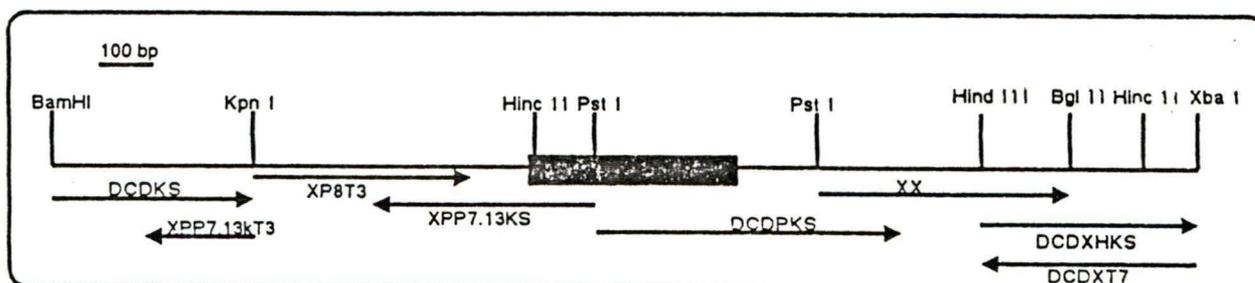


Figure 14: La carte de restriction du clone DCDX avec les différents sous-clones utilisés pour le séquençage. L'orientation et la région couverte de chaque clone par le séquençage sont indiquées par les flèches.

avoir besoin de chercher des fragments. Pour séquencer la partie 5' à partir du site de site de Pst I dans la région V, nous avons dû sous-cloner de nouveau le fragment Pst I-Pst I. Le clone a été nommé XPP7.13. Les clones utilisés pour le séquençage sont indiqués ci-dessous. On voit les sites, la direction et la longueur des séquences obtenues (Fig. 14). Après le clonage, il ne reste qu'à

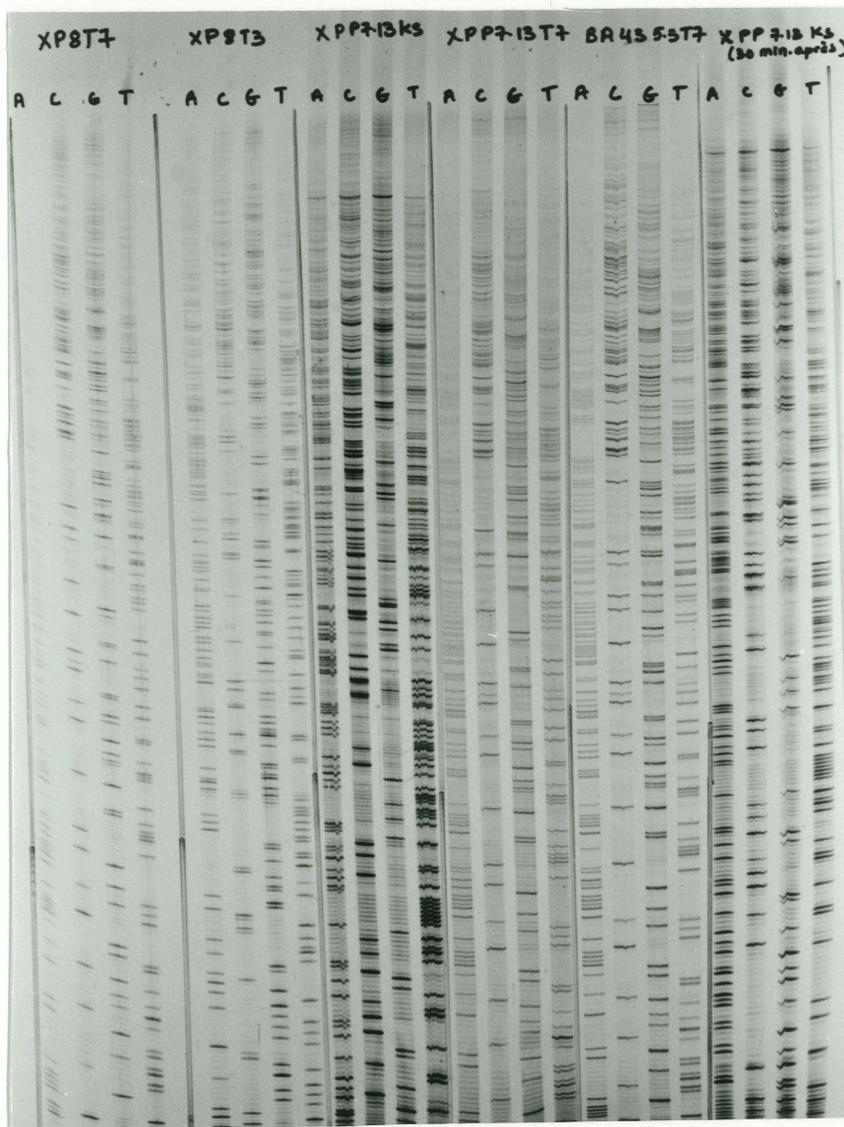


Figure 15: Un exemple d'un autoradiogramme d'un gel de séquence montrant les échantillons de différents sous-clones avec les 4 différentes réactions de terminaisons.

faire la réaction de synthèse en utilisant les amorces(primers) des deux côtés de l'insert pour avoir des séquences qui se chevauchent et on répartit les échantillons de réactions sur des gels d'acrylamide. Après avoir exposé le gel à un film 35x43 cm XRP1 pendant une nuit, on lit directement les bases à partir du bas de l'autoradiogramme. La figure 15 nous donne un exemple de l'autoradiogramme d'un gel de séquence.

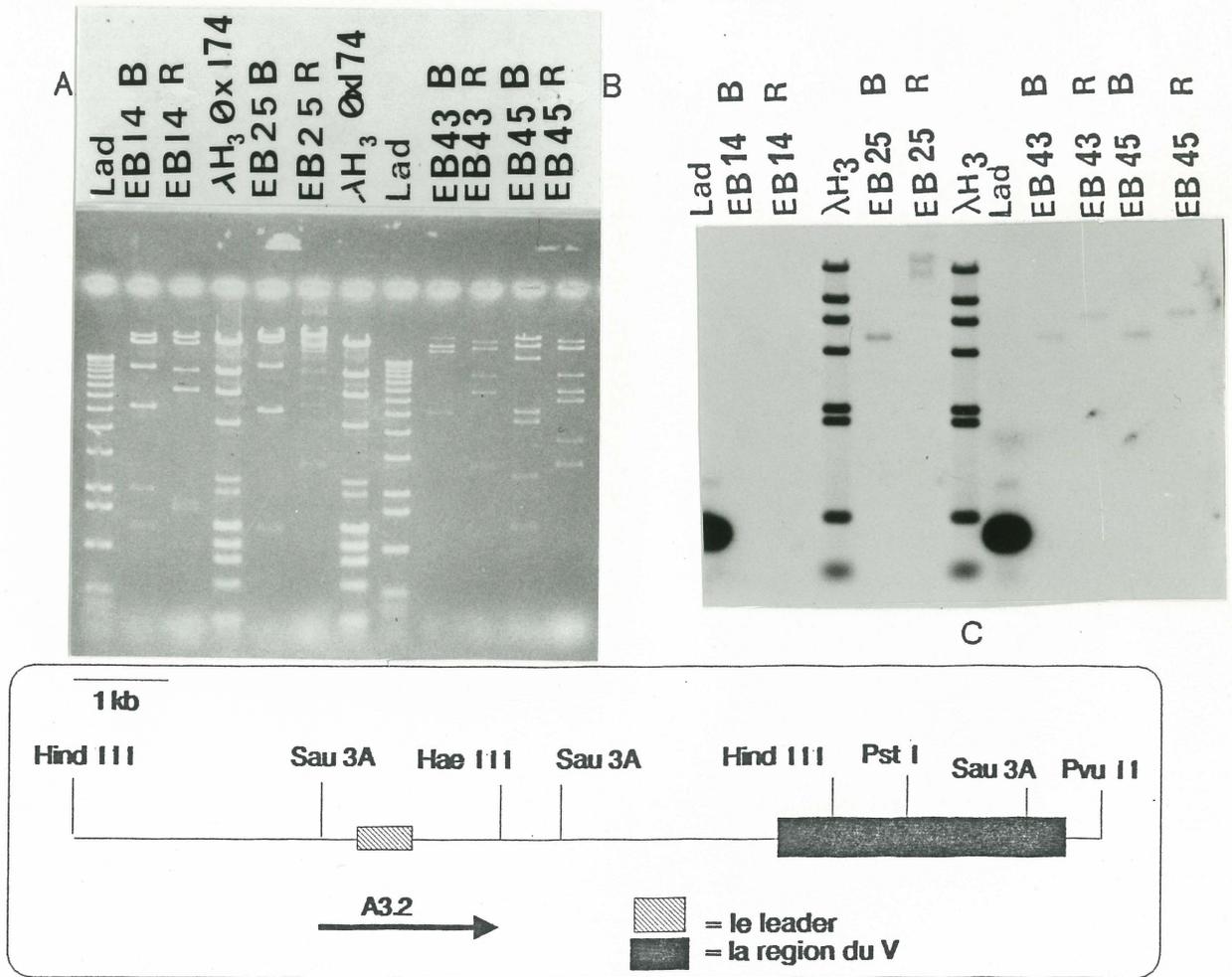
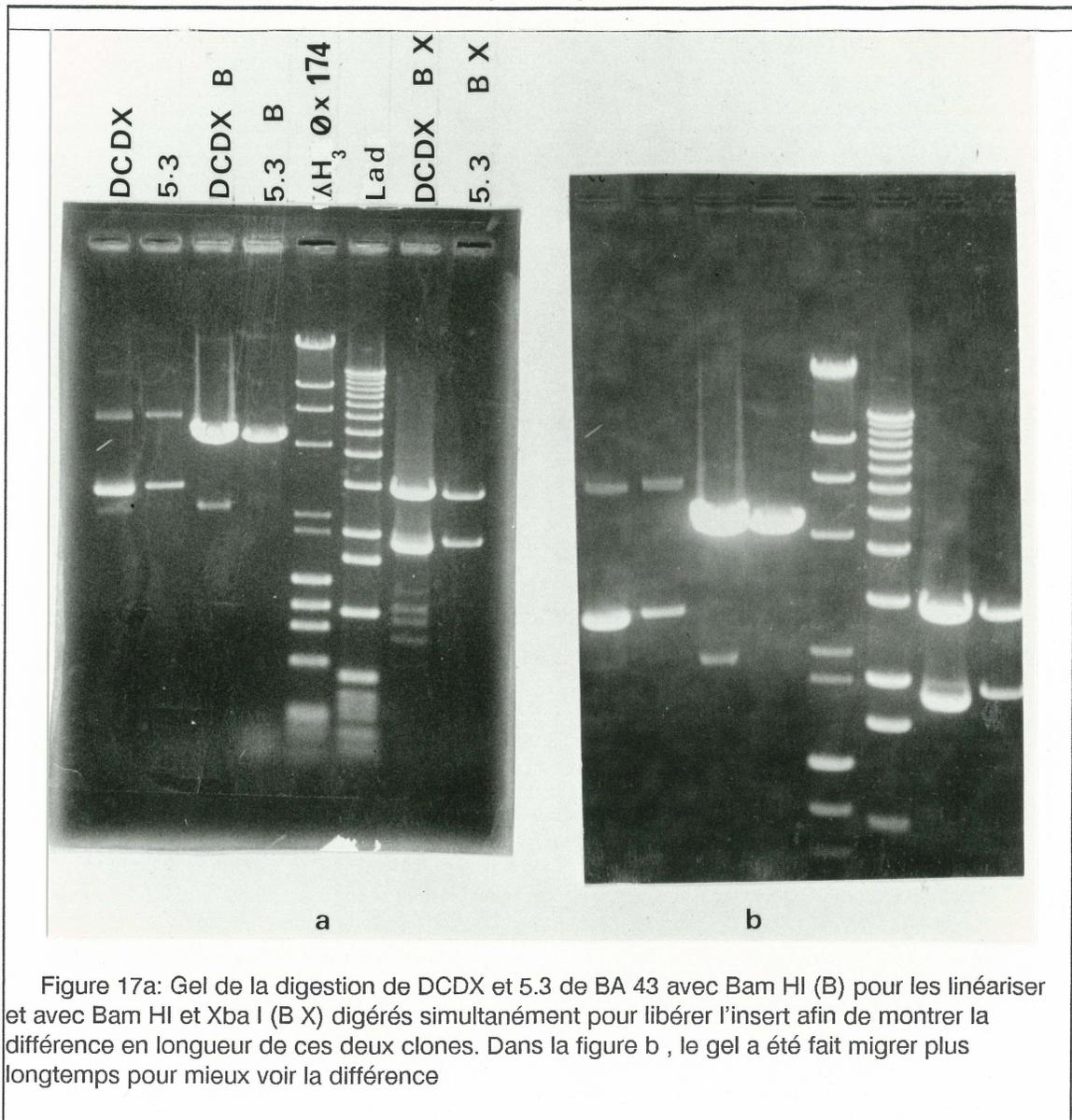


Figure 16a: Gel de la digestion des clones EB 14, EB 25, EB 43 et EB 45 avec Bam H1(B) et Eco R1(R). 16b: l'autoradiogramme du gel avec la sonde du leader(A3.2). On remarque ici que EB 14 ne donne aucune bande qui hybride tandis que les clones EB 25, EB 43 et EB 45 donnent les fragments 5.4 Kb Bam H1 et 6.8 Kb Eco R1 qui hybrident avec la sonde. La digestion EB 25 (R) est partielle. 16c: Carte démontrant la sonde A3.2 dérivée de la souche de souris NZB Vk1-B et commence de Sau 3a à Hae III.

3.3 LA RECHERCHE DU LEADER

La région du leader et celle de l'octamère n'ont pas pu être identifiées dans les séquences 5'. Nous avons donc dû chercher une autre voie pour y arriver. Nous avons voulu voir si dans l'insert, du clone EB14, le leader pourrait se trouver plus en amont du fragment 5.4 kb Bam HI sous-cloné dans M13+. Nous avons pris alors une sonde qui contient le leader (A3.2) dérivé de NZB Vk-1B (fig. 16c) . La sonde a été utilisée pour hybrider les clones EB 14, 25, 43, 45



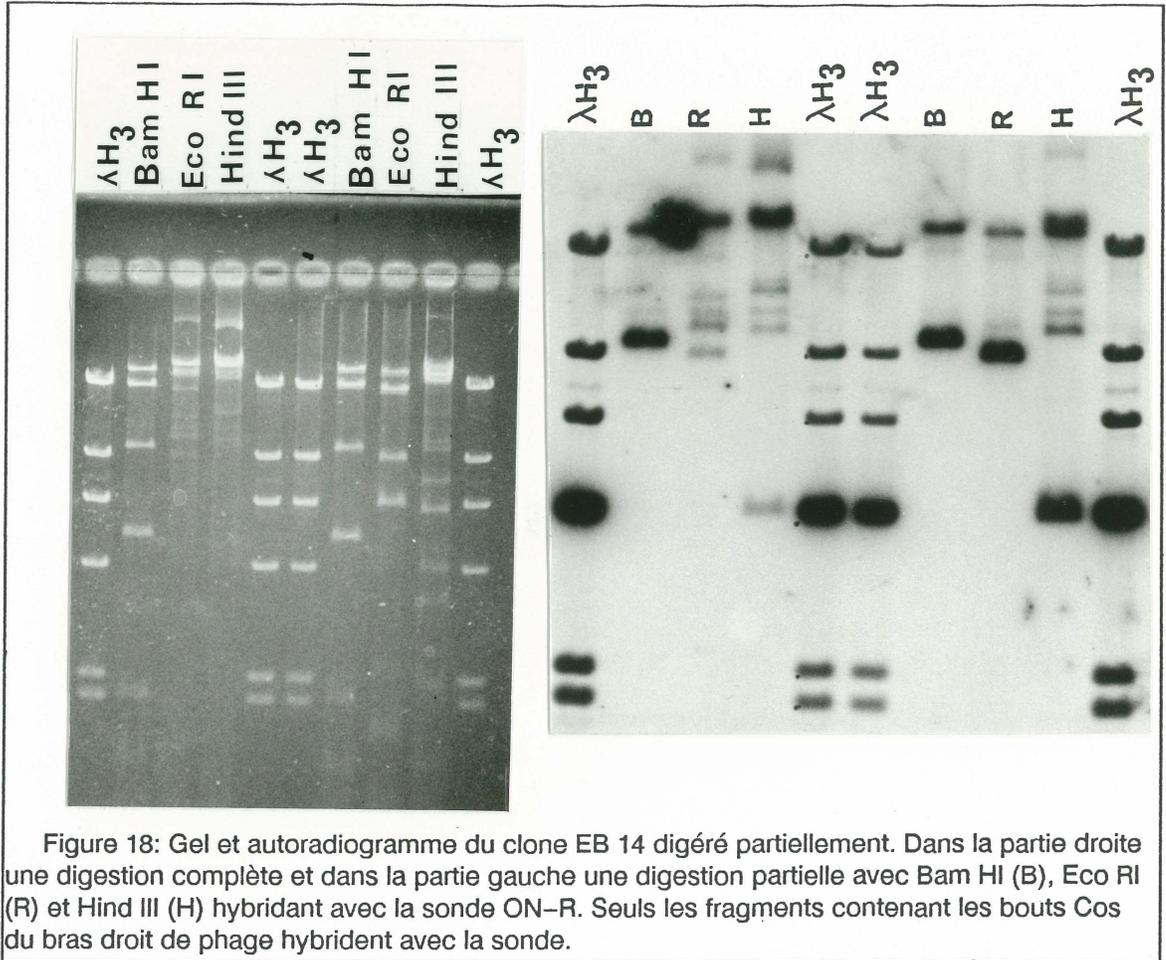
traités avec les enzymes Bam HI et Eco RI respectivement. Les trois autres clones à l'exception de EB14 possèdent le fragment de 5.4 kb qui hybride (fig. 16b). Nous avons également cherché à démontrer que les clones BA 43 et DCD n'étaient pas semblables malgré que les fragments BamHI de ces clones semblent être de la même taille. Les clones DCDX et 5.3 de BA 43 ont été retenus parce qu'ils représentent la région qui nous concerne le plus et c'est le fragment de la région de Bam HI à Xba I pour les deux clones. En les linéarisant ou en les digérant simultanément avec Bam HI et Xba I, nous avons pu constater une très petite différence de longueur (fig 17). Les analyses plus approfondies de ces clones expliquent cette observation. Cela nous donne quand même un indice qu'il pourrait y avoir une délétion dans DCDX dans la région comprenant le leader et les séquences régulatrices. Ainsi le gène n'est pas fonctionnel.

3.4 CONSTRUCTION D'UNE CARTE DE RESTRICTION.

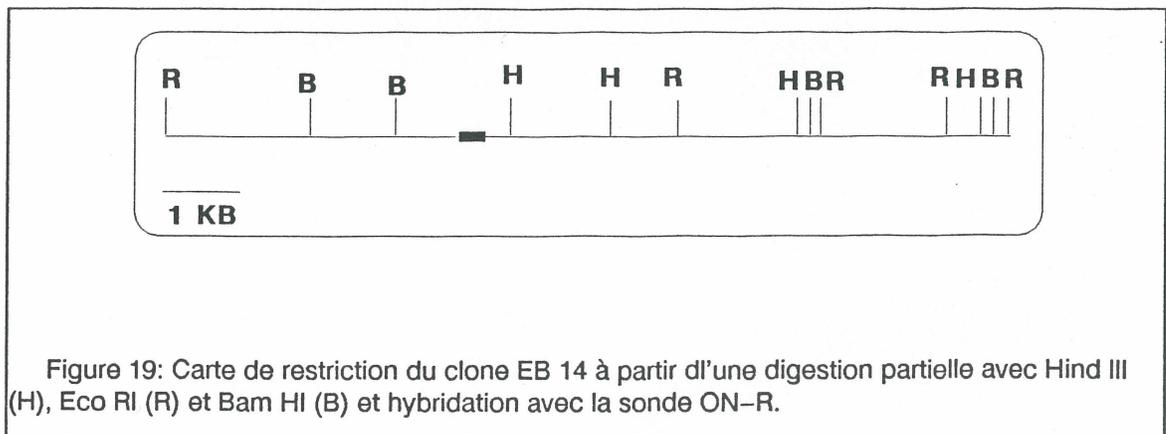
Après avoir fait une digestion partielle du phage EB 14 avec les enzymes BamHI, Eco RI et Hind III et l'hybrider avec la sonde ON-R (sonde oligonucléotide du bras droit du phage λ) (fig.18) nous avons pu situer ces sites sur le clone EB 14 (fig. 19). Les fragments qui hybrident avec la sonde sont indiqués dans le tableau 4.

Tableau 4: Les fragments du clone EB 14 qui hybrident avec la sonde ON-R.

Eco RI	Hind III	Bam HI
21.1 Kb	23.0 Kb	22.3 Kb
14.6 Kb	16.8 Kb	13.5 Kb
13.0 Kb	15.0 Kb	10.4 Kb
11.5 Kb	10.7 Kb	
9.5 Kb		



De plus en sous-clonant le fragment Eco RI qui englobe la région DCDX et s'étend en 3' et 5' de cette région, nous avons réussi à obtenir des sites en



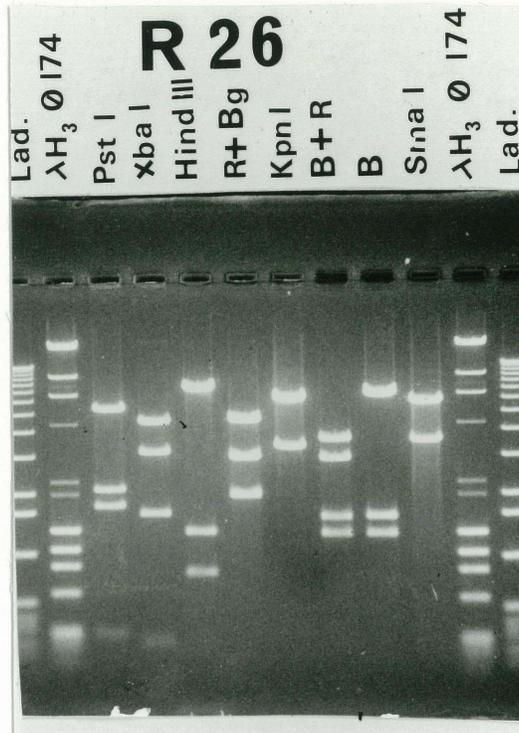


Figure 20: Le sous-clone R 26 obtenu par le clonage du fragment 6.8 Kb Eco RI dans le plasmide M13+. Ce sous-clone a été digéré par Pst I, Xba I, Eco RI+Bgl II (R+Bg), Kpn I, Bam HI(B)+Eco RI(R) et Sma I.

dehors de la région que nous avons séquencée. Le sous-clone du fragment 6.8 Kb Eco RI est nommé R26. Ce sous-clone a été digéré avec les enzymes Bam HI, Pst I, Xba I, Hind III, Eco RI +Bgl II, Kpn I, Bam HI + Eco RI et Sma I (ig.20). Les fragments obtenus sont décrits dans le tableau 5.

Tableau 5: Les fragments obtenus du sous-clone R26 et digéré avec Eco RI+Bgl II , Bam HI, Pst I, Xba I, Hind III, Kpn I, Bam HI+Eco RI et Sma I sont indiqués ci-dessous.

Enzymes	Vecteur M13+	Vecteur M13+ +fragment adjacent résiduel	Taille des fragments libérés
Xba I		3.07 kb	1.6 kb 4.5 kb 0.3 kb
Eco RI + Bgl II	2.8 kb		4.5 kb 2.2 kb
Bam HI		6.8 kb	1.6 kb 1.2 kb
Bam HI + Eco RI	2.8 kb		3.9 kb 1.6 kb 1.2 kb
Pst I		5.4 kb	2.2 kb 1.7 kb 0.3 kb
Hind III		7.2 kb	1.4 kb 0.9 kb
Sma I		6.0 kb	3.4 kb

En connaissant la localisation des sites de restriction pour ces enzymes dans le vecteur M13+ par rapport à l'insert et en utilisant les données obtenues lors de la stratégie de séquençage, on a pu situer ces sites. Pour avoir plus de précision sur les sites en dehors de la région de la sonde DCDX et DCDSHi (fig. 22), on a utilisé ces sondes pour hybrider la réplique du gel contenant l'ADN du phage digéré avec les enzymes EcoRI, Bam HI, Pst I, Sma I, Bgl II et Hind III (voir fig. 21a et 21b). Les fragments qui hybrident avec la sonde DCDX et DCDSHi sont indiqués dans le tableau 6:

Tableau 6: Fragments du clone EB 14 digéré avec Eco RI, Bam HI, Pst I, Sma I, Bgl II et Hind III qui hybrident avec les sondes DCDX et DCDSHi.

Enzymes	Bras du vecteur ou bras du vecteur + fragment adjacent	Fragments libérés
Eco RI	19.54 kb	6.8 kb (DCDX, DCDSHi)
	9.5 kb	1.7 kb (DCDSHi)
		1.6 kb
		0.9 kb
Bam HI	> 22.03 kb	5.4 kb (DCDX, DCDSHi)
		2.3 kb
		1.3 kb
Pst I	10.4 kb	4.2 kb (DCDX, DCDSHi)
	4.9 kb	1.6 kb (DCDX)
	2.8 kb	2.3 kb
	2.0 kb	
	1.0 kb	
	1.1 kb	
	0.9 kb	
Sma I	19.54 kb	6.8 kb
	6.2 kb (DCDX)	
	8.44 kb (DCDSHi)	
Bgl II	> 23.0 kb (DCDX)	
	7.5 kb	
	8.44 kb (DCDSHi)	
Hind III	> 22.3 kb (DCDX)	2.2 kb (DCDSHi)
	4.4 kb	2.2 kb
	6.6 kb	1.4 kb
		1.3 kb

De ces digestions, nous avons pu faire une carte de restriction plus détaillée du clone EB 14 (fig. 22).

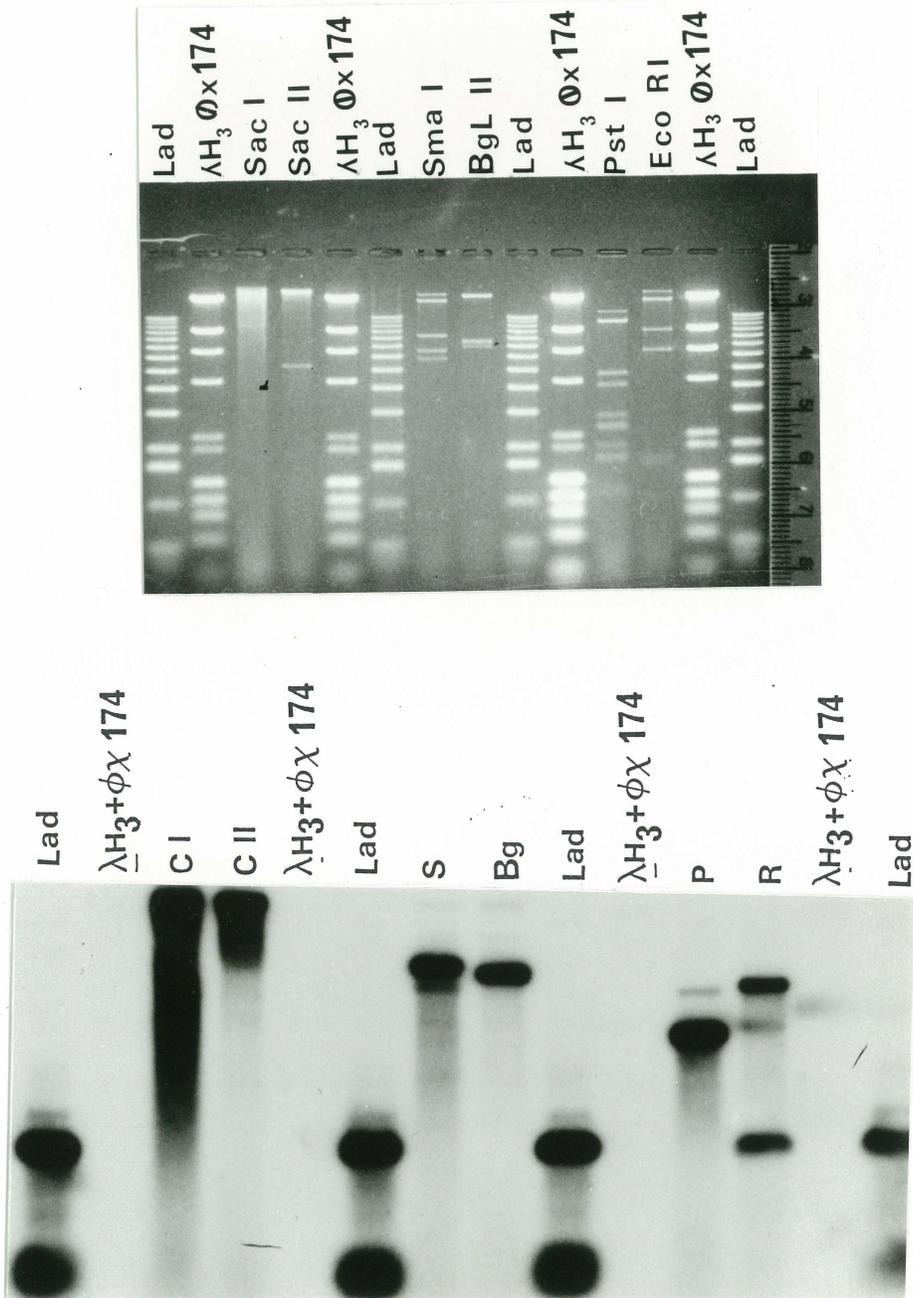


Figure 21a: Gel et autoradiogramme démontrant les fragments du clone EB 14 digérés avec Sac I (C I), Sac II (C II), Sma I (S), Bgl II (Bg), Pst I (P) et Eco RI (R) qui hybrident avec la sonde DCDSHi. λH_3 étant très faible n'apparaît pas sur la photo. Le ladder (lad) bien qu'il n'ait pas été marqué contient 2 fragments qui hybrident avec la sonde: le fragment 1.6 Kb et 516 pb.

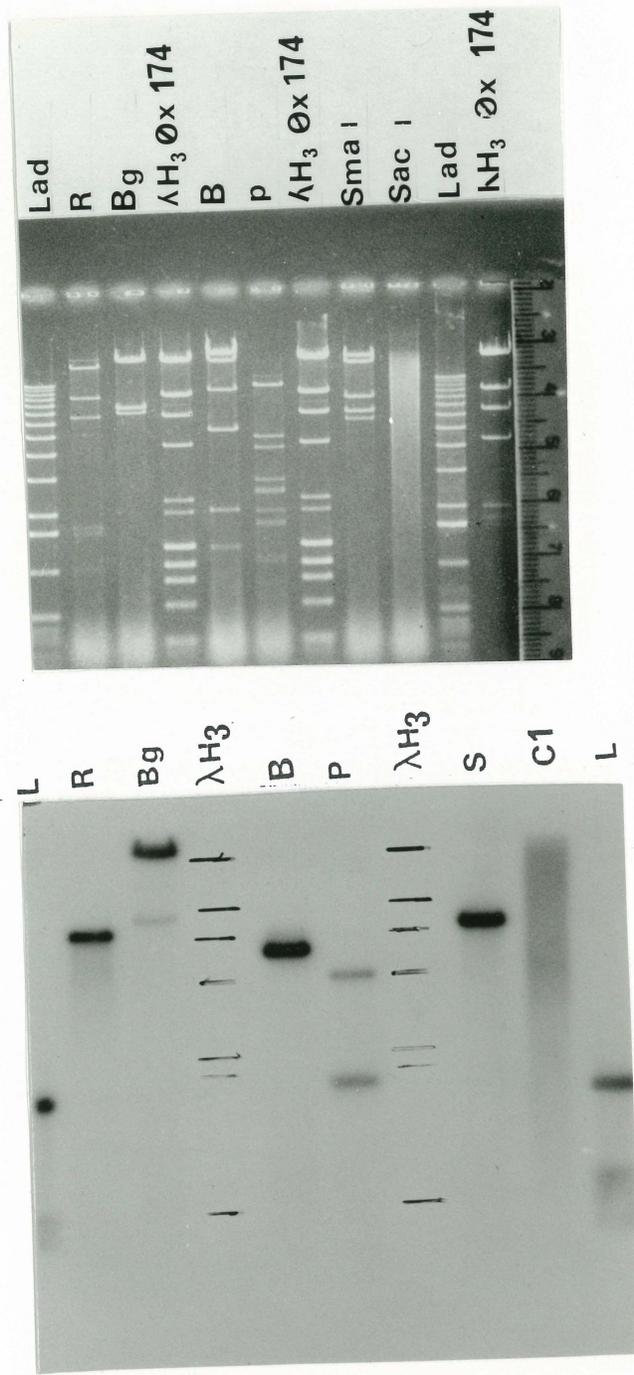


Figure 21b: Gel et autoradiogramme du clone EB 14 digéré avec Eco RI (R), Bgl II (Bg), Bam HI (B), Pst I (P) et Sma I (S) et hybridé avec la sonde DCDX. La digestion avec Sac I (C1) est dégradée. Les marqueurs de grandeur λH₃ sont marqués avec un stylo parce que les bandes étaient faibles

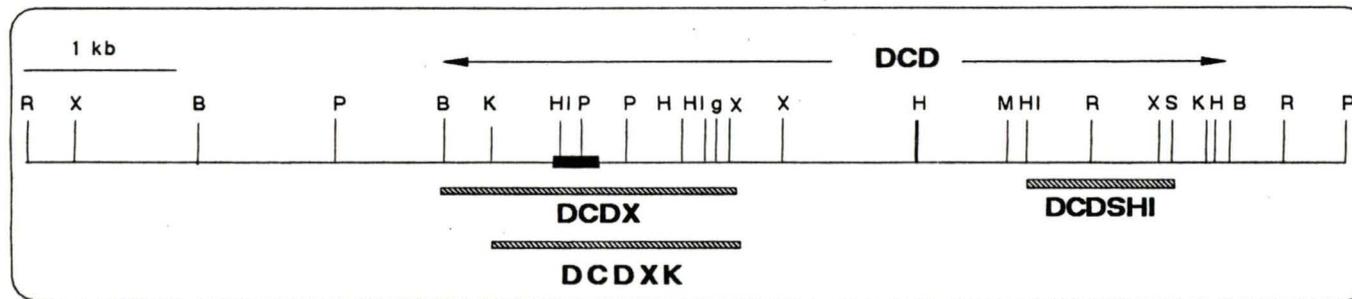


Figure 22: La carte en détail des sites Eco RI(R), Bam HI(B), Hinc II(Hi), Hind III(H), Pst I(P), Bgl II(g), Xba I(X), Sac I (S), Sma I(M), Kpn I (K). Les régions employées comme sonde sont indiquées sur cette carte. Le fragment de 5.4 Kb sous-clonée dans le plasmide M13+ connu sous le nom de DCD (tableau 2)est aussi représenté sur la carte.

3.5 IDENTIFICATION DES GENES CORRESPONDANT AUX DCD ET BA 43 DANS L'ADN GENOMIQUE

Nous avons aussi voulu vérifier la présence des gènes obtenus dans l'ADN génomique de souris. Pour se faire, nous avons essayé de construire une sonde avec différents sous-clones obtenus. Le sous-clone DCDK est obtenu en digérant le clone DCD avec l'enzyme Kpn I et est par la suite reliqué et transformé. Ce sous-clone donne une sonde qui devrait hybrider avec la région 3' de DCD tandis que le clone DCDX donne une sonde pour détecter la région 5' du clone DCD. De même avec BA 43, nous avons digéré avec Apa I et Kpn I puis reliqué et transformé pour avoir le sous-clone BA 43A et BA 43K (fig. 23).

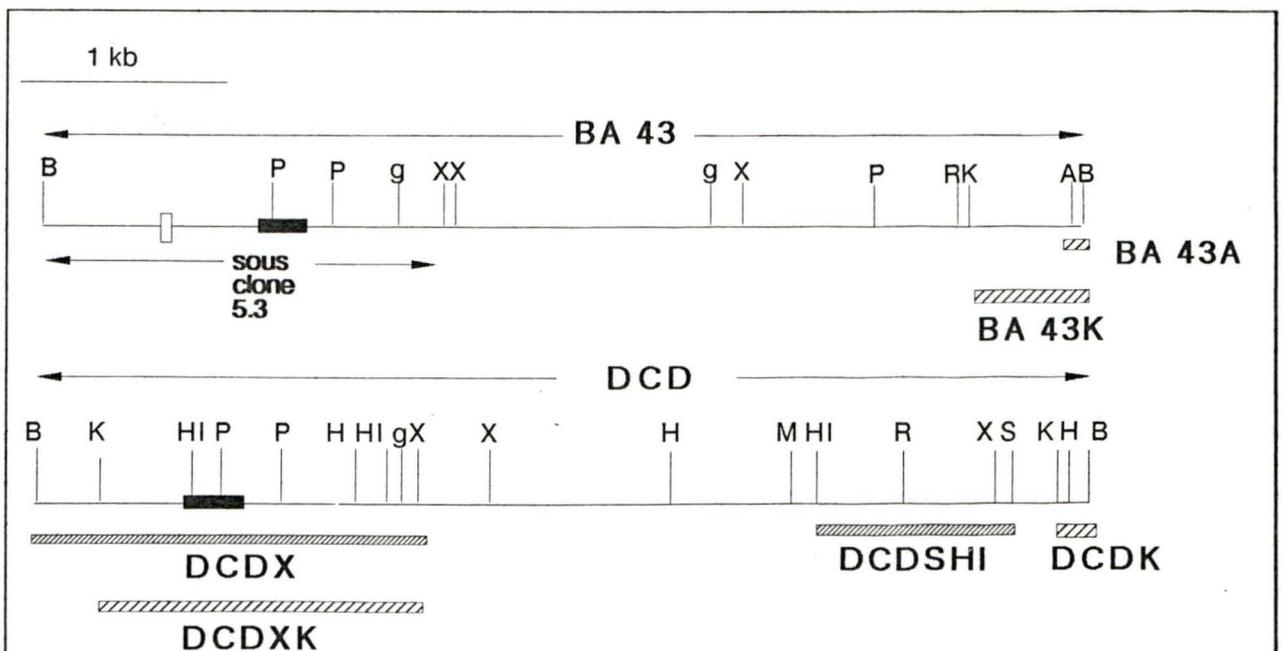
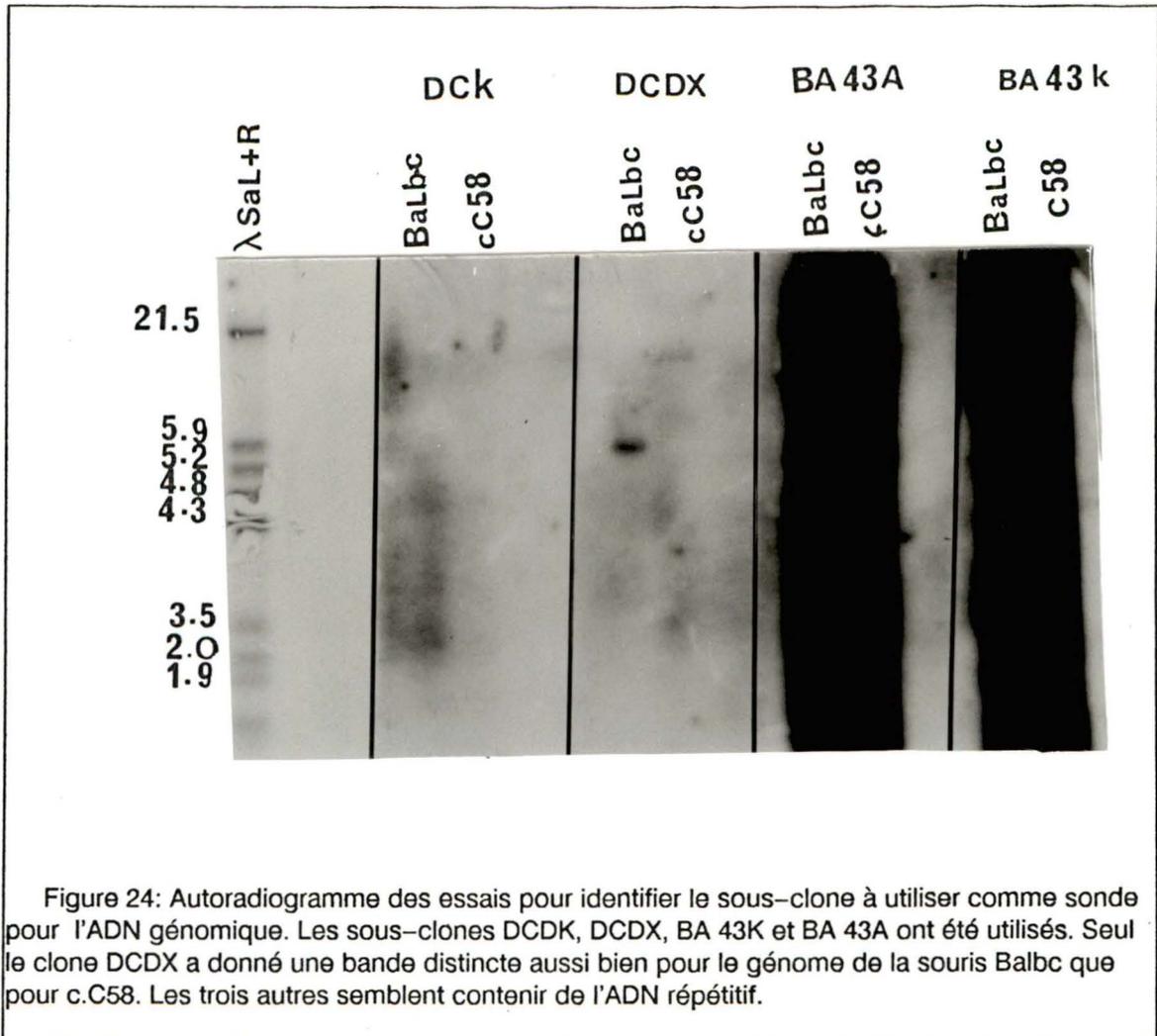


Figure 23: Cartes montrant les sites Eco RI(R), Bam HI(B), Hinc II(Hi), Hind III(H), Pst I(P), Bgl II(g), Xba I(X), sac I(s), sma I(M), kpn I(K), Apa I(A) dans les sous clones DCD et BA 43. Les régions employées comme sonde sur l'ADN génomique afin d'avoir une sonde spécifique aux clones obtenus sont: DCDX, DCDK, BA 43A et BA 43K. Les sondes DCDX et DCDSHI ont été utilisées pour la localisation de certains sites en 3' et 5' du sous-clone DCD, tandis que la sonde DCDXK a été utilisée sur l'ADN génomique pour vérifier la présence de certaines bandes. La région séquencée de BA 43, le sous-clone 5.3 est indiqué sur la carte tandis que la région séquencée du clone DCD, clone DCDX est aussi employé comme sonde. Le leader est représenté par un encadrement vide et le V par un encadrement plein.

En essayant les 4 sous-clones obtenus individuellement comme sonde sur un blot contenant l'ADN génomique de souris Balbc et c.C58 digéré avec l'enzyme Bam HI, seule la sonde DCDX donne une bande. Les autres sous-clones semblent contenir de l'ADN répétitif (Fig. 24) parcequ'ils hybrident partout sur



l'ADN génomique. Sachant que la sonde DCDX est assez spécifique, nous l'avons utilisée sur les répliques des gels contenant de l'ADN génomique de différentes souches de souris (Balbc, AKR, c.C58, P/J, NZB, NAK) digéré par Bam HI et Eco RI respectivement (Fig. 25). Les fragments qui hybrident, sont

indiqués dans le tableau 7.

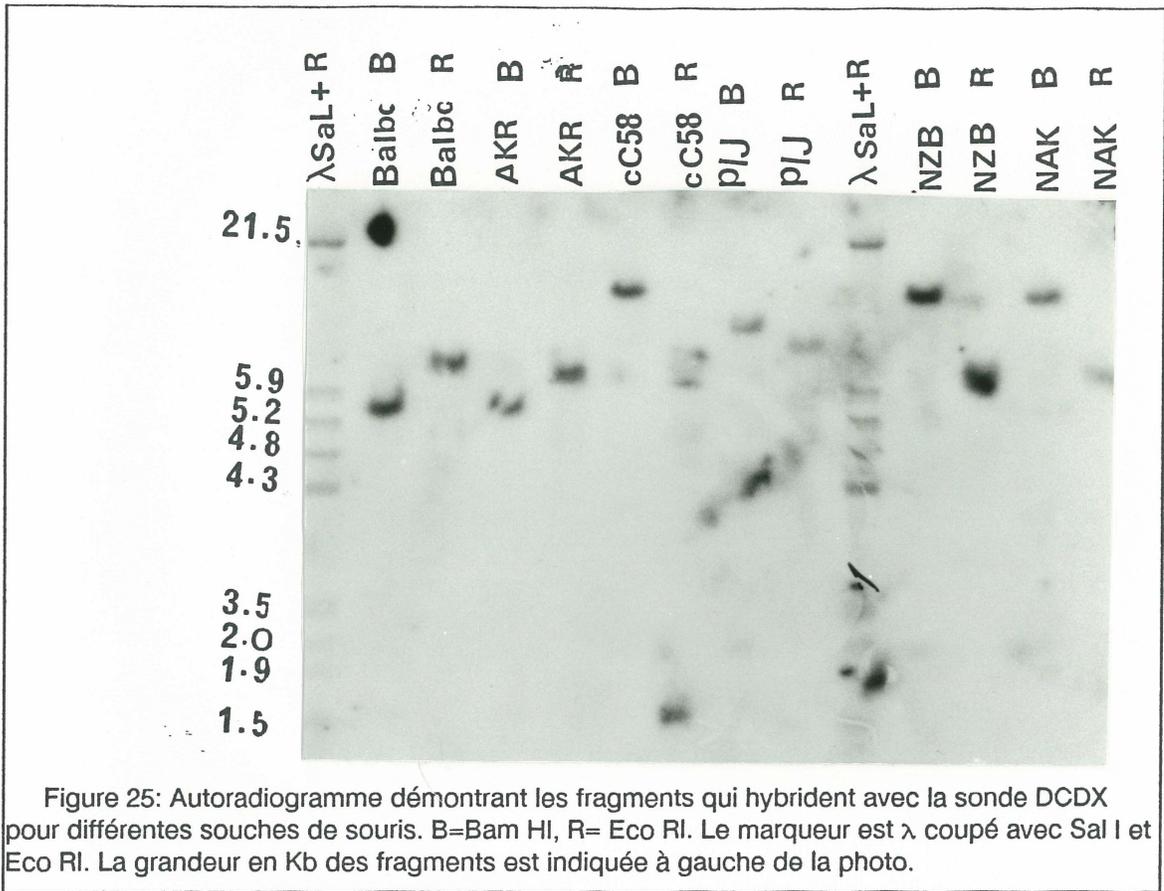


Figure 25: Autoradiogramme démontrant les fragments qui hybrident avec la sonde DCDX pour différentes souches de souris. B=Bam HI, R= Eco RI. Le marqueur est λ coupé avec Sal I et Eco RI. La grandeur en Kb des fragments est indiquée à gauche de la photo.

Tableau 7: Les fragments de l'ADN génomique des différentes souches de souris qui hybrident avec la sonde DCDX.

	Bam HI	Eco RI
BALB/c	5.4 kb	6.8 kb
AKR	5.4 kb	6.8 kb
cC58	12.0 kb	6.8 kb
P/J	9.0 kb	8.0 kb
NZB	12.0 kb	6.8 kb
NAK	12.0 kb	6.8 kb

Soupçonnant que BA 43 et DCD pourraient être 2 clones différents bien que les fragments Eco RI et Bam HI soient identiques pour l'ADN génomique, nous avons cherché à éliminer les facteurs d'erreurs qui pourraient surgir lors de

l'encapsidation et l'infection des bactéries. Pour cela, nous avons procédé à la digestion de l'ADN des clones EB 14 et EB 43 avec les enzymes dont les sites se trouvent dans un et non pas dans l'autre ou les sites qui ne se trouvent pas au même endroit dans les deux clones. Les enzymes utilisés dans ce but sont Hinc II, Hind III, Kpn I et Apa I. Dans le clone DCD, il n'y a pas de site Apa I et dans BA 43, il n'y a pas de site Hinc II et Hind III. Quant au site Kpn 1, il ne se trouve pas au même endroit dans DCD que dans BA 43. La grandeur des bandes qui hybrident avec la sonde DCDX n'est pas la même pour EB 14 et EB 43 coupés avec les enzymes mentionnés plus haut (fig 26 et tableau 8).

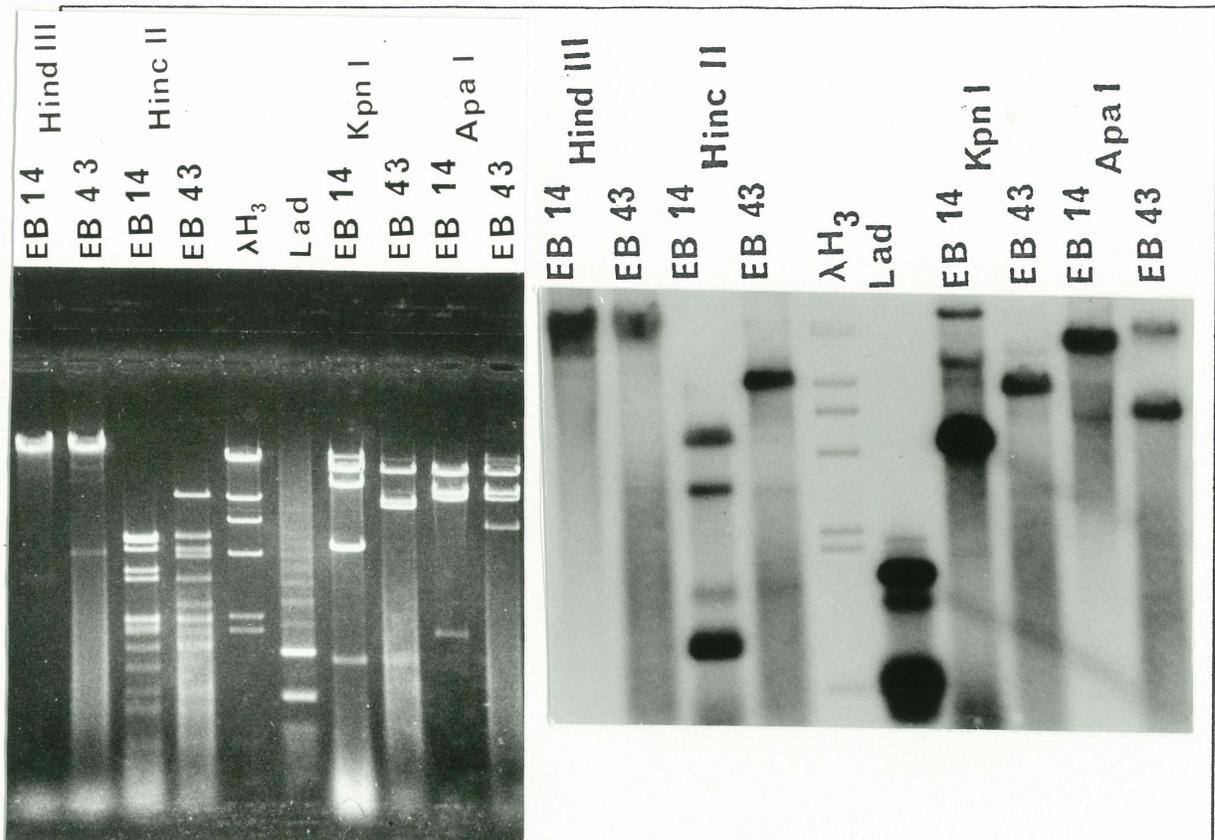


Figure 26: Gel de la digestion de EB 14 et EB 43 avec Hind III, Hinc II, Kpn I et Apa I. A droite, l'autoradiogramme du gel utilisant la sonde DCDX. La digestion Hind III est incomplète.

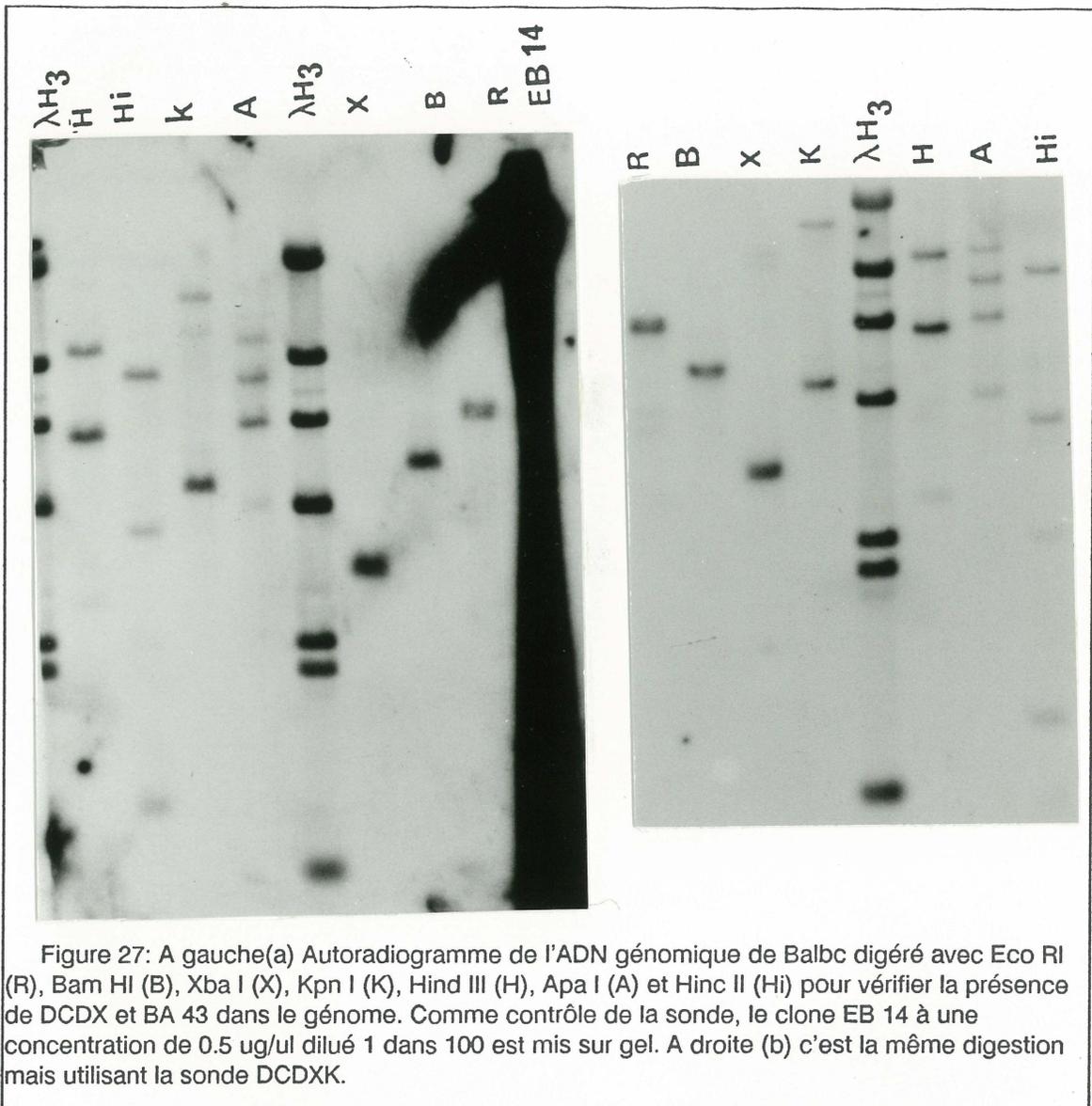
Tableau 8: Fragments des clones EB 14 et 43 qui hybrident avec la sonde DCDX.

	EB 14	EB 43
Hinc II	4.9kb(digestion partielle) 4.0 kb 1.6 kb 0.8 kb	9.6 kb
Kpn I	4.8 kb	7.5 kb
Apa I	18.0 kb	6.7 kb

Les fragments de l'ADN des clones EB 14 et 43 digérés avec Hinc II, Kpn 1 et Apa 1 qui hybrident avec la sonde DCDX sont indiqués dans le tableau 8. La digestion avec Hind III n'a pas bien marché. D'autre part l'ADN génomique de Balbc a été digéré aussi avec Hind III, Hinc II, Kpn 1, Apa 1, Xba 1, Bam H1 et Eco R1. En utilisant la sonde DCDX, on

Tableau 9: Fragments obtenus avec l'ADN génomique de la souris Balbc qui hybrident avec les sondes DCDX et DCDXK. De plus, les fragments des clones EB 14 et 43 sont écrits à côté des fragments auxquels ils correspondent.

	DCDX	EB 14	EB 43	DCDXK
Hind III	10.9 kb 6.4 kb			10.5kb 6.5kb
Hinc II	9.8 kb 4.0 kb 2.3 kb 0.8 kb	(4.0) (1.6) (0.8)	(9.6)	9.4kb 4.0kb 2.3kb 0.86 kb
Kpn I	18.9 kb 13.4 kb		(7.5)	19.0kb
Apa I	4.9 kb 12.2 kb 8.8 kb 6.6 kb 4.3 kb	(4.8) (18.0)		4.9kb 13.6kb 8.5kb 6.6kb 4.7kb
Bam HI	5.4 kb			5.3kb
Eco RI	6.8 kb			6.8 kb



remarque que les fragments qui hybrident, correspondent quoique pas exactement avec les fragments dans le phage, mais ils concordent avec l'idée que BA 43 et DCDX sont 2 clones différents (fig. 27 et tableau 9). La sonde DCDXK (fig. 23) a été utilisée pour essayer de voir si la bande de 13.4 kb Kpn I provient d'un autre sous-groupe lié au DCD qui montre un certain polymorphisme à ce site ou si c'est dû au fragment Kpn I qui chevauche une partie de

DCD. En effet, la bande de 12 kb n'apparaît pas sur l'autoradiogramme; ce qui confirme la deuxième hypothèse. La grandeur du fragment 2.3 Kb obtenue par digestion avec Hinc II dans l'ADN génomique ne correspond pas au fragment de 1.6 Kb obtenu pour la même digestion de l'ADN du clone EB 14. Cette différence pourrait provenir du fait que les temps de migration du gel et l'ADN dans le deux cas n'étaient pas les mêmes. On observe souvent que les fragments de 2.2 et le 2.0 Kb du standard de poids moléculaire λ H₃ ne migrent pas toujours à la même distance. Le fragment de 18 Kb de la digestion Apa I du clone EB 14 correspond soit au fragment de 12.2 Kb ou 8.8 Kb de la digestion de l'ADN génomique parce que la région couverte par DCD ne contient pas de site Apa I et la grandeur du fragment qui peut hybrider avec la sonde DCDX devrait être au moins 8.5 Kb (calculé à partir du site de clonage dans Embl4 et jusqu'à la fin du clone DCD). Ainsi les bandes de 12 Kb ou 8.8 Kb et 4.3 Kb de cette digestion qui apparaissent dans le patron de digestion de l'ADN génomique et qui correspondent à aucune bande dans les clones digérés nous mènent à penser qu'il y a d'autres sous-groupes étroitement liés avec ceux qu'on a étudié. Dans la séquence du clone DCD, il n'y a pas de site Apa I et dans celle de BA 43 il n'y a qu'un site Apa I en dehors de la région couverte par la sonde; ce qui écarte la possibilité de surestimer le nombre de gènes qui hybride avec la sonde DCDX.

3.6 LES SEQUENCES DES CLONES DCDX ET BA43.

La séquence du clone DCDX comprend 1864 bp et commence au site Bam HI jusqu'au site Xba I. La région V ne semble pas avoir de codon 'stop' ou un changement de cadre de lecture. Les signaux de reconnaissance (heptamère-

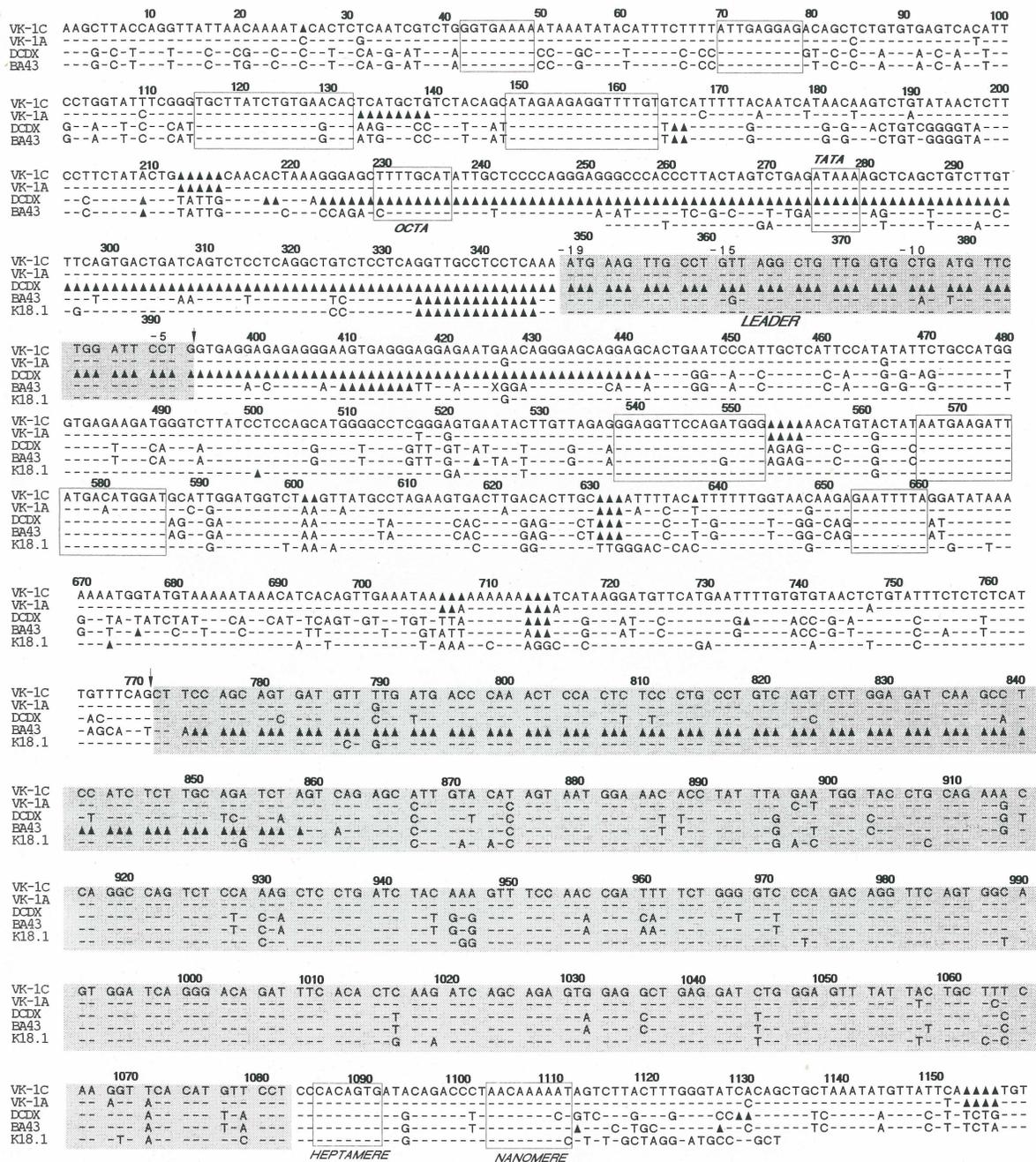


Figure 28a: Alignements multiples entre Vk-1C, Vk-1A, DCDX, BA 43 et K 18.1. Les séquences conservées sont encadrées. Les nucléotides incertains sont écrits en petits caractères ou remplacés par x. Les délétions ou insertions sont représentés par des triangles. Les régions codantes sont ombragées tandis que les deux premières séquences de Vk-1C et Vk-1A dans les régions homologues aux séquences Lines sont ombragées plus faiblement (voir fig. 28b). Les flèches pointant vers le bas (∇) sont les régions d'épissage.

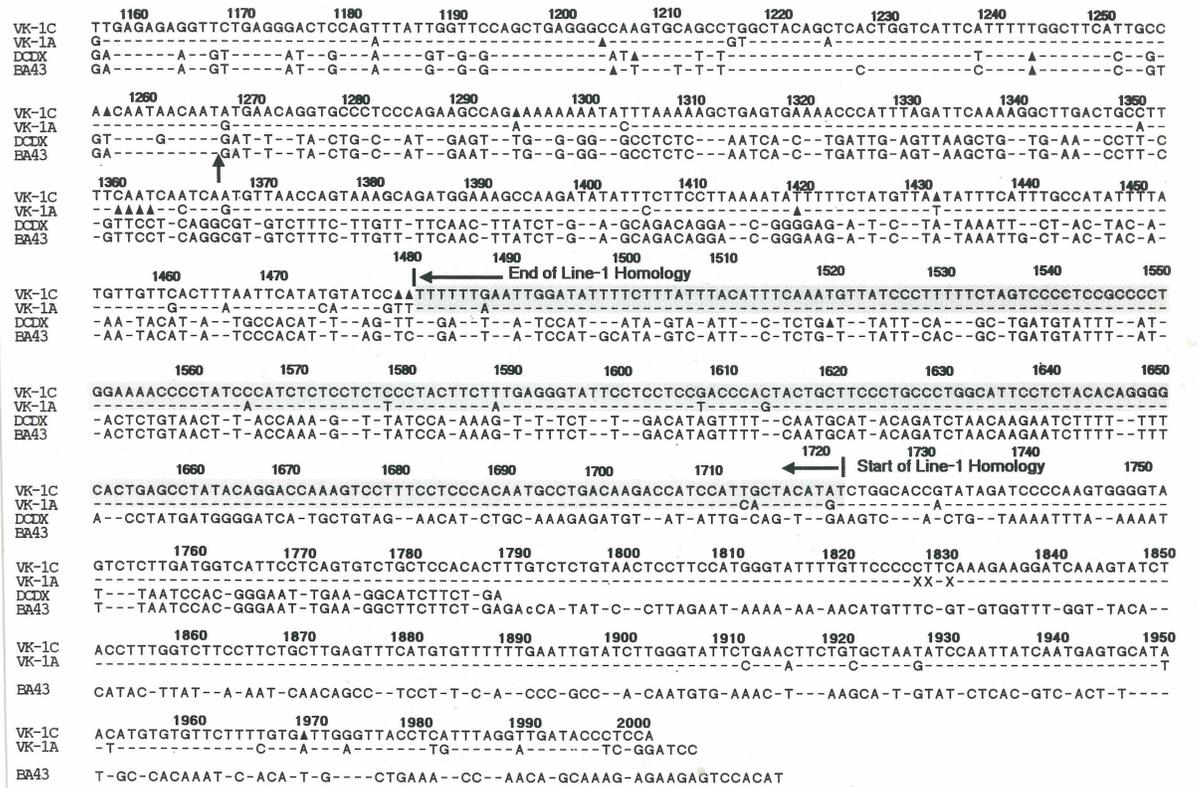


Figure 28b: La suite des alignements multiples entre Vk-1C, Vk-1A, DCDX, BA 43 et K18.1. Les régions homologues aux séquences Lines sont ombragées faiblement et se trouvent seulement dans Vk-1A et Vk1-C. La flèche pointant vers le haut (↑) indique la région où l'homologie entre DCDX et BA 43 avec les autres sous-groupes s'arrête.

nanomère) sont intacts. La séquence consensus retrouvée au site donneur du signal d'épissage AG 5' de la région V est aussi intacte (voir alignement multiple Fig. 28). La région variable est ombragée. Le nanomère et l'heptamère sont encadrés. La région V est constituée de 100 acides aminés. Le clone BA 43 a été aussi séquencé à partir du site de Bam HI au site Xba I et consiste en 2084 pb. En plus d'une délétion dans la région V, il y a aussi une mutation dans le site d'épissage à la position 771 (AG->AT). Les sites d'épissage sont indiqués dans figure 28 par des flèches pointées vers le bas. La comparaison de la séquence du clone DCDX avec le clone BA43 montre qu'il y a une homologie très grande autant que pour la région codante que pour la région non-codante. Cette homologie est de 96.7%. Il y a une délétion de 200 bp à la position 224 dans DCDX tandis que dans BA 43 la délétion est de 84 pb et se trouve dans la position 774. Les délétions sont indiquées par des triangles noirs (fig. 28) La délétion dans DCDX comprend le leader, la région octamérique et la boîte TATA. D'autre part, la délétion dans BA 43 se trouve dans la région V en commençant à la position -4 et se poursuit jusqu'au début du deuxième codon de CDR1.

3.7 COMPARAISON DES CLONES BA43 ET DCDX AVEC D'AUTRES GROUPES

Les séquences des clones BA 43 et DCDX ont été comparées avec le programme d'analyse de séquence DIAGONAL. Les clones DCDX et BA 43 sont assez identiques tout au long de la séquence à l'exception de l'endroit où il y a une délétion de 200 bp dans la région du leader dans DCDX et 84 pb dans la région du V dans le clone BA 43 (fig. 29). Les deux gènes ne sont ainsi pas fonctionnels à cause de la délétion. En utilisant le même programme , la

comparaison a été faite facilement avec plusieurs familles de $V\kappa$. Dans la majorité des familles, seulement la région V montre une certaine homologie avec DCDX. Par contre, la famille $V\kappa$ -1 montre une grande homologie dans la région V (fig.29) On constate que le sous-groupe $V\kappa$ -1D (K18.1) (Corbet S. et al 1987) est aussi étroitement lié à DCDX. La comparaison avec K18.1 en dehors de la région 5' et 3' n'est pas possible parce que la séquence n'est pas assez longue.

La comparaison d'homologie a été aussi faite par l'alignement multiple-un programme procuré par ATARI ST employant GFA-Basic. Les gènes de la famille $V\kappa$ -1 ($V\kappa$ -1C, $V\kappa$ -1A, K18.1) sont comparés avec DCDX et BA 43. Les insertions sont intercalées pour optimiser l'alignement (fig. 28). Etant donné que l'homologie du début de la séquence de DCDX et BA 43 sont faibles avec d'autres gènes $V\kappa$ -1, l'alignement est fait à partir de la position 283 et cela correspond à la position 1 de la séquence maîtresse ($V\kappa$ -1C). L'homologie de la région codante est de 85.5% autant avec BA 43 que DCDX. Avec les autres sous-groupes $V\kappa$ -1A, $V\kappa$ -1C, $V\kappa$ -1B, l'homologie s'atténue subitement dans la région 3' du segment V (à partir de la position 1267). L'homologie en dehors de ces régions est de 20 à 25%. Par contre l'homologie dans la région 5' jusqu'au leader varie entre 70 à 72% pour les deux clones en question. (la région comparée couvre 353 pb en amont du leader). Pour la région intronique, l'homologie est de 72-75%. Le pourcentage d'homologie dans la région V et du leader est de 87% et 94% respectivement. Pour la région du segment V et du leader, les clones DCDX et BA 43 sont utilisés respectivement pour la comparaison avec les autres sous-groupes à cause des délétions qui se trouvent dans chacun d'eux. La comparaison des acides aminés de BA 43 avec

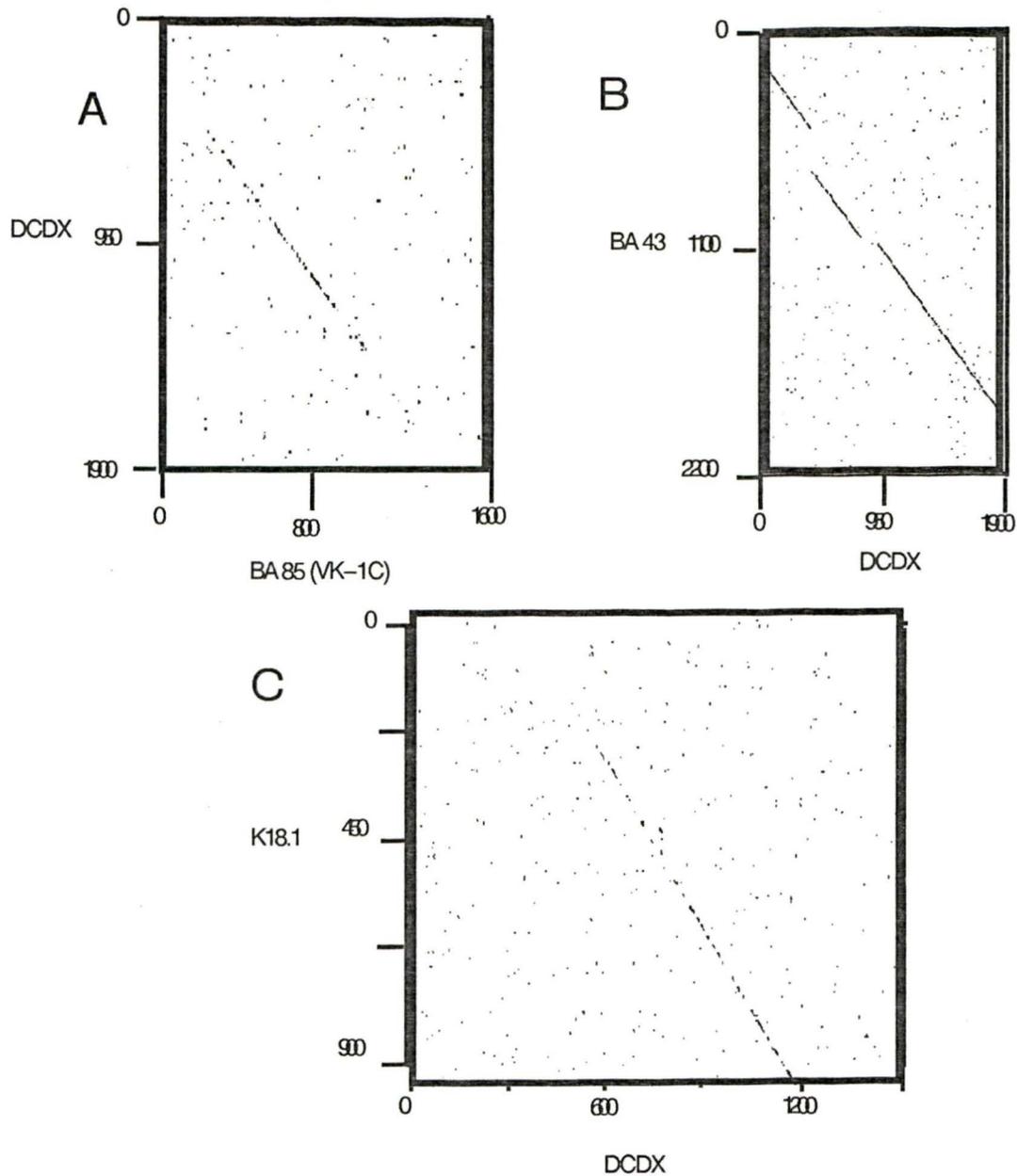


Figure 29: Comparaison des différents sous-groupes de V_{k-1} avec DCDX en utilisant le programme DIAGONAL .

A: DCDX vs $V_{k-1}C$. L'homologie commence à la position 769 et s'étend sur la région V mais s'arrête à la position 1267 dans $V_{k-1}C$ et à la position 1345 dans DCDX de la séquence originale.

B: DCDX vs BA 43. L'homologie est très grande à l'exception de l'endroit où il y a des délétions: positions 504 et 852 de la séquence originale de DCDX.

C: DCDX vs K18.1. L'homologie commence à partir de la position 525 de DCDX et termine juste après le segment V étant donné que la région 5' du V de K18.1 n'a pas été complètement séquencée.

(Les positions indiquées ici correspondent aux positions dans les séquences originales).

ceux des autres groupes révèle trois substitutions d'acides aminés dans la région du leader; dans la position -9, la méthionine est remplacée par la leucine,-10 la leucine par la glutamine et à la position ;-15 la valine par la glycine (tableau 10). Dans la région V, il y a 13 substitutions si on compare la séquence de DCDX avec Vk-1C. Les plus hauts taux de substitution se trouve dans la région de CDR2 et CDR3 qui sont de 42.9% et de 28.6% respectivement et le taux le plus bas se trouve dans FR3 qui est de 3% (tableau 11).

Tableau 10: Séquence des acides aminés du leader des différents sous-groupes de VN-1. (*=délétion, -=séquence identique à la séquence maîtresse et les chiffres indiquent les positions à partir de la région V).

	-19	-5	-4
VN-1A	MKLPVRLLMFWIP/..	(INTRON 382 pb)	.../SSSS
VN-1C	-----/..	(INTRON 378 pb)	.../SSSS
K18.1	-----/..	(INTRON 387pb)	.../SSSS
BA 43	----G--QL----	(INTRON 396 pb)	.../SSSS
DCDX	*****		

Tableau 11: les séquences des acides aminés de la région variable des différents sous-groupes VN-1. Les régions FR et CDR sont indiquées.

	FR1	CDR1	FR2
VN1C	DVLMTQTPLSLPVS	LGDQASISC	RSSQSIVHSNGNTYLE WYLQKPGQSPKLLIY
VN1A	--V-----	-----L-----	H-----
K18.1	-AV-----	-----LEN-----	N-----Q-----
DCDX	--LL-----F-----	S-----S--E	-H---S---LQ---
BA 43	***** * -K-----S--D		-H-----LQ---

	CDR2	FR3	CDR3
VN1C	KVSNRFS	GVPDRFSGSGSGTDFTLKISRVEAEDLGVYYC	FQGSHVP
VN1A	-----	-----	-F S-ST---
K18.1	R-----	--L-----	-F L-VT---
DCDX	E--K-N-	-----P-----	---T-L-
BA 43	E--K-N-	-----P-----	---T-L-

4 DISCUSSION

Les analyses des protéines exprimées du sous-groupe Vk-1 des myélomes et des anticorps du même sous-groupe stimulés par divers antigènes démontrent que 90% de ces séquences protéiques exprimées sont codées par seulement deux gènes Vk-1 (Vk-1A et Vk-1C). Par contre, en utilisant une sonde de Vk-1, plus précisément Vk-1C^f ou PPV101 sur un 'Southern blot' de BALB/c coupé avec l'enzyme Bam HI en dehors des deux bandes représentant Vk-1A et Vk-1C, 8 autres bandes sont révélées. Alors, il semble avoir une certaine incohérence entre les deux résultats. Afin d'éclaircir ce problème, nous avons essayé de cloner les autres bandes. Quatre clones indépendants ayant un fragment Bam HI 5.4 Kb chacun ont été isolés. Ils ont été nommés DCDX, BA 43, EB25 et EB 45.

4.1 IDENTIFICATION DES GENES OBTENUS.

Bien qu'on ne sache pas encore si les membres du sous-groupe des gènes DCDX et BA 43 ont tous été déjà identifiés, il est peu probable qu'il y en ait d'autres dû au nombre de bandes obtenues et leur intensités avec la sonde spécifique DCDX. Actuellement on a pu constater une certaine ressemblance entre EB 45, EB 25 et BA 43 suite aux analyses de sites de restriction.

Ayant admis l'identité propre des gènes DCDX et BA 43, nous avons cherché à identifier l'ancêtre commun de ces gènes. La comparaison avec différents groupes nous a permis de les placer dans le groupe Vk-1 (fig. 30). Cette information seule ne nous permet pas de localiser ces deux gènes de façon certaine dans le locus V du chromosome 6. Cependant, grâce aux résultats du patron de restriction de Bam HI et de Eco RI avec la sonde DCDX, on arrive à

situer ces deux gènes dans le locus Vk-1. Cela est possible parce qu'il existe un certain polymorphisme entre les souches de souris caractérisées Ef2a et Ef2b. Plus précisément, l'étude de patron de restriction est faite dans la souche c.C58, une souche congénique à BALB/c mais ayant un locus Vk provenant de la souche C58 (Ef2b) (Gottlieb et al 1980). Le résultat a révélé une distribution de bande avec la sonde DCDX plutôt identique aux souches caractérisées Ef2b qu'aux souches caractérisées Ef2a (par exemple BALB/c).

4.2 SELECTION NEGATIVE POUR LES GENES DCDX ET BA 43.

Constatant que les deux gènes DCDX et BA 43 sont non-alléliques et homologues à 96% et qu'ils ont une délétion importante chacun, nous nous sommes demandés pourquoi ils se sont dupliqués. Cela formerait-il une partie du processus normal de l'amplification du locus Vk-1 ou se pourrait-il que ces deux gènes jadis fonctionnels suite à un mécanisme quelconque et pour certaines raisons aient été délévés dans une certaine région? Sont-ils simplement les étapes finales des variants du gène Vk-1 dans l'évolution ou une relique de l'évolution sur leur voies vers l'oblitération? Les observations déjà recueillies (Jouvin et al 1986) nous portent à croire que pendant l'amplification d'une famille de gènes, au cours du processus normal, certains des gènes subissent des délétions ou des mutations aberrantes. Par contre, on a aussi observé que certaines séquences peuvent être impliquées dans la promotion d'une délétion, insertion ou recombinaison. Un exemple pertinent d'un tel phénomène a été observé dans la souche de souris Mus. Pahari. A l'endroit d'une grande délétion, on a remarqué la séquence CCAAAG soit au début d'une délétion comme dans le cas du gène PaVk24C (position 350) soit à la fin comme dans le cas du gène PaVk24A (Jouvin et al 1986). Dans le cas qui

nous préoccupe, nous n'avons trouvé aucune séquence semblable qui expliquerait la délétion. Par contre nous avons observé une séquence assez conservée (CTGTATT) au début de la délétion (10 nucléotides avant la délétion, position 210 dans la figure 28a) dans le gène DCDX et une séquence semblable (CTGTACTT) également au début de la délétion (21 nucléotides avant la délétion, position 748 dans la figure 28a) dans le gène BA 43 qui pourraient avoir les mêmes fonctions que la séquence observée dans *Mus. Pahari*. Indépendamment de la façon dont ces deux gènes sont rendus non-fonctionnels, on constate tout de même une sélection défavorable à ces gènes. En effet le groupe de Köfler a démontré la présence des gènes structuraux responsables de l'expression des autoanticorps dans les souches de souris MRL/Mp-lpr/lpr qui développent spontanément un syndrome comme le lupus erythematosus systémique avec la formation des anticorps anti-DNA (Köfler et al 1985). D'autre part on a noté aussi l'utilisation préférentielle de certaines familles de gènes (V κ -1, V κ -4, V κ -10 ou V κ -19) dans la formation des autoanticorps dans la souche de souris C57 BL/6JSmn-me^v/me^v comparée à leur contrôle la souche C57 BL/6JSmn (Painter C et al 1988). On peut dès lors déduire qu'il existe des gènes dans certaines familles qui codent pour des autoanticorps et qu'il existe des processus sélectifs procédant soit à l'exclusion de certains gènes, soit à leur mise hors de fonction pour la survie et le bien être de l'organisme. Cette sélection négative n'est pas à l'encontre de l'évolution neutre. Ainsi on connaît un exemple où le pseudogène 6 de l'haplotype Igh^b de la souche de souris C57 BL/6 rendu non-fonctionnel par une délétion dans le codon 22 est l'allèle fonctionnel qui code pour MRL-DNA 10 (un auto-anticorps anti-DNA) dans la souche MRL (Köfler et al 1985). On peut donc croire que

dans le cas des deux gènes clonés, il y a eu une sélection en faveur de l'inactivation de ces deux gènes plutôt qu'en un effet du hasard. Les larges blocs de délétion dans DCDX et BA 43 pourraient être un mécanisme primaire pour les enlever. Etant donné que ces deux gènes ne partagent pas la même délétion au même endroit et qu'en dehors de la délétion dans ces deux gènes, il ne semble pas y avoir d'autre mutation aberrante, le précurseur doit être un gène fonctionnel. On pourrait spéculer qu'un tel mécanisme de délétion agissant sur les deux gènes (DCDX et BA 43) qui étaient jadis fonctionnels juste après leur duplication, pourrait aussi agir sur le gène précurseur. Ainsi tous les membres de ce sous-groupe seraient rendus non-fonctionnels avant qu'ils puissent avoir le temps de se diverger. La grande homologie entre les séquences de DCDX et BA 43 proviendrait du fait que la duplication de ces deux gènes est très récente , ce qui n'aurait pas permis l'accumulation de mutations ou bien qu'il existerait un certain mécanisme de correction par conversion ou recombinaison qui les empêcherait de diverger (Cohen et Givol 1983, Schiff et al 1985). Cependant il est peu probable que cette dernière hypothèse soit vraie parce que les pourcentages de substitution de nucléotides dans ces deux gènes sont identiques soit environ 96% autant pour la moyenne de la région codante que pour la région non-codante (tableau 11). Ce qui est en accord avec l'évolution neutre des pseudogènes.

4.3 LA REGION CODANTE, 5' ET 3' DU GENE.

L'homologie entre les gènes DCDX et BA 43 s'étend tout au long de la région séquencée . Il est ainsi difficile de dire jusqu'à quelle région du gène il y a eu duplication. BA 43 est délété dans la région du segment V de la position 774 jusqu'au deuxième codon du CDR1. On remarque que la région CDR3 est

100% homologue entre DCDX et BA 43. Cette grande homologie dans cette région hypervariable comparativement aux autres régions nous laisse spéculer que ces gènes ne seront pas utilisés. Dans les régions du frame-work (FR), le rapport de substitution des acides aminés aux remplacements (silencieux) neutres des nucléotides doit normalement être moins que un pour les gènes fonctionnels dû à des contraintes structurelles (Givol et al 1981, Loh et al 1983, Blankenstein et al 1984) . Pour les régions FR2 et FR3, ces rapports concernant le gène Vk1-C et K18.1 (Vk-1D) sont 0.3 et 0.2 respectivement (calculé par le rapport du pourcentage du tableau 14 à celui du tableau 13). Quant au FR1, il n'y a pas de substitution neutre; ce qui fait que le rapport de substitution n'est pas calculable. En effet, une haute variabilité dans la région du FR1 a été rencontrée dans toutes les séquences de Vk (Kabat et al 1979). Cela peut impliquer que FR1 est moins important pour garder l'intégrité structurale de la région variable que FR2 et FR3. Dans le cas présent, ce rapport de substitution pour FR2 est un mais pour FR3, étant donné qu'il n'y a pas eu de substitution d'acide aminé le rapport n'a pas pu être calculé. Ainsi, à cause de la délétion dans BA 43, il n'a pas été possible aussi de calculer le rapport de substitution pour FR1. Un rapport plus élevé que un n'a pas été obtenu pour FR2 et n'est pas significatif pour FR3 probablement parce que le temps de divergence n'est pas assez long pour permettre à DCDX et BA 43 d'accumuler assez de mutation chacun de leur côté pour que le rapport soit significatif. La comparaison des différents sous-groupes de Vk-1 avec DCDX démontre une grande homologie de la région codante. Cependant, l'homologie dans la région 3' du segment V est atténuée considérablement à partir de la position 1267 (fig 28b). On voit aussi des séquences conservées dans la partie 5' autant que pour

DCDX et BA 43 que pour les autres sous-groupes (Voir les encadrements dans la figure 28b). On ne sait pas encore si ce sont des séquences qui jouent un rôle régulateur pour l'expression d'un gène d'immunoglobuline ou s'il s'agit d'une simple coïncidence. De même, la séquence répétitive (L1Md) qui se trouve à 400 pb en 3' de la région codante des autres sous-groupes (Ng et al sous-presse) n'est pas présente dans DCDX et BA 43. Cela pourrait être possible si le point de brisure (break point) était très proche du segment V pendant la première duplication séparant le précurseur de ces deux gènes au précurseur d'autres sous-groupes (Vk-1A, Vk-1B, Vk-1C, Vk-1D et Vk-1E). De plus, la région 3' aurait pu être modifiée par un mécanisme tel que la conversion ou la recombinaison. Dans les régions séquencées (600 pb 3' du segment V) de DCDX et BA 43 aucune séquence répétitive qui pourrait expliquer une insertion dans ces endroits, n'a été remarquée. Les séquences de BA 43 et DCDX à partir de l'endroit où la brisure pourrait arriver ne montre aucune homologie significative avec les séquences de 'genbank' de rongeur.

4.4 TEMPS DE DIVERGENCE

Pour situer dans le temps la duplication des deux gènes clonés du gène précurseur, nous avons utilisé la divergence entre DCDX et BA 43 comme horloge évolutive. La divergence est calculée en terme de pourcentage de substitution de nucléotides au lieu du pourcentage d'acides aminés substitués, ceci afin d'éviter une surestimation du temps de duplication. Etant donné que ces deux gènes ne sont pas fonctionnels, il n'y a pas une pression sélective sur la région codante, donnant ainsi un taux plus élevé de remplacement d'acides aminés comparé à celui d'un gène fonctionnel même si on a vu plus haut que la divergence n'a pas été assez longue pour donner une différence significative

dans la région du FR. Comme il y a une délétion dans la région V de BA 43, le taux de substitution de nucléotides pour la région codante est calculé à partir de l'endroit où il y a une délétion comme s'il y avait un mauvais alignement d'une paire de base. Etant donné que la moyenne du pourcentage de substitution est de 3.3% (tableau 12) et qu'on peut postuler qu'il faut 0.4 million d'années pour accumuler 1% de divergence (Catzeflis et al 1987), on peut conclure que la duplication a eu lieu il y a 1.3 million d'années, soit 3.3×0.4 million d'année. La duplication a donc eu lieu avant la division des différentes souches de souris consanguines puisque l'évolution de différentes espèces de souris est apparue il y a un million d'années de cela (Yonekava et al 1980). Cela est confirmé dans le 'Southern blot' avec la sonde DCDX sur l'ADN de différentes souches de souris (voir autoradiogramme fig. 25).

Par ailleurs, pour estimer le temps de divergence entre d'autres sous-groupes et celui de DCDX et BA 43, nous avons analysé le nombre de substitutions de nucléotides entre Vk-1C et DCDX dans la région codante. Le pourcentage de nucléotides substitués est de 15 % (tableau 12). De ce pourcentage, on estime que le temps de séparation entre les autres sous-groupes et DCDX est de 6 millions d'années. Cette estimation est très approximative car Vk-1C n'est pas un pseudogène et est soumis à une pression de sélection contrairement à DCDX.

Tableau 12: Pourcentage d'homologie entre DCDX et les différents sous-groupes de Vk1.

	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR+CDR
DCDX/Vk1C	88.4	83.33	82.22	80.95	93.81	82.6	85.2
DCDX/Vk1A	88.4	83.33	84.44	80.95	91.75	82.6	85.2
DCDX/BA43	-	92.85	97.77	95.23	97.93	100.0	96.7
DCDX/K18.1	86.95	79.16	82.22	80.95	89.69	86.95	84.5

Tableau 13: Pourcentage de substitution neutre dans la région codante

	FR1	CDR1	FR2	CDR2	FR3	CDR3
DCDX/BA43	-	12.5	6.6	14.2	6.0	0.0
DCDX/Vk1C	17.4	31.3	13.3	0.0	15.2	14.3
DCDX/K18.1	21.7	18.7	26.6	28.6	12.2	0.0
Vk1C/K18.1	0.0	12.5	20.0	0.0	15.2	42.8

Tableau 14: Pourcentage de substitution d'acides aminés.

	FR1	CDR1	FR2	CDR2	FR3	CDR3
DCDX/BA43	-	18.7	6.6	14.3	0.0	0.0
DCDX/Vk1C	8.7	6.3	25.0	42.9	3.0	28.6
DCDX/K18.1	17.4	31.3	20.0	42.9	12.1	28.6
Vk1-C/K18.1	8.7	25.0	6.3	14.3	3.0	14.3

4.5 LE ROLE DU PSEUDOGENE

La formation d'un pseudogène est un phénomène ubiquitaire dans toutes les familles de gènes d'eucaryotes. Même si plusieurs études ont été faites pour obtenir plus d'informations sur la formation et le rôle probable de ces gènes dans le génome, les pseudogènes restent encore un sujet controversé. Certains chercheurs ont proposé que les pseudogènes pourraient jouer un rôle d'inhibiteur intracellulaire, soit par la formation d'un RNA antisens, soit par la formation d'une protéine à partir d'un pseudogène muté dans la région codant pour le site actif et non pas dans la région codant pour le site de liaison (McCarrey et Riggs 1986). Dans le cas qui nous préoccupe il est peu probable qu'une telle fonction puisse être attribuée aux gènes clonés à cause de la délétion importante dans la région du promoteur et du leader pour DCDX et dans la région du V pour BA 43. On sait déjà que les pseudogènes V λ jouent un

rôle important dans la diversité des anticorps chez le poulet (Reynaud et al 1985) mais jusqu'à présent cela n'a pas été prouvé chez la souris. Par contre, un certain mécanisme de correction a été observé pour les pseudogènes du groupe VH11 de la souche de souris de BALB/c, ce qui a mené les auteurs à suggérer la contribution probable des pseudogènes dans la diversité des anticorps (Schiff et al 1985). Un tel mécanisme de correction n'a pas été observé dans notre cas.

Cleary et al (1981) ont suggéré que les pseudogènes dans une famille de multigènes pourraient jouer un rôle indispensable de 'spacer gene' pour le bon fonctionnement du complexe. Ainsi, ils peuvent fonctionner comme un régulateur agissant en cis et permettant l'expression d'un seul gène Vk. Malgré tout, il ne faut pas oublier que pour générer une grande diversité d'anticorps, le système immunitaire a développé un répertoire très vaste de gènes V par l'amplification des gènes qui se divergent par des mutations. Parfois certains gènes subissent des mutations aberrantes qui les rendent non-fonctionnels. Comme la sélection agit sur le répertoire V en entier, il n'y a pas élimination sélective des pseudogènes. Ainsi, ils vont persister dans le génome jusqu'à ce qu'ils aient perdu toute homologie ou jusqu'à ce qu'ils soient réactivés par conversion. De même, le fait qu'il y ait une grande quantité d'anticorps utilisant les chaînes légères codées par les gènes structuraux Vk-1, on peut penser que ces gènes ont réussi à être fixés dans le génome de la souris. L'amplification et la duplication récentes en sont la preuve. A la suite de ce phénomène, les gènes sont soumis à l'évolution moléculaire qui permet non seulement la formation de différents sous-groupes Vk-1 mais aussi la formation de pseudogènes comme DCDX et BA 43. Cela est en accord avec la théorie

neutraliste de l'évolution moléculaire développée par le théoricien japonais M. Kimura (Kimura 1968) qui selon lui le nombre de gènes polymorphes qu'on peut mettre en évidence dans une population est si élevé qu'il serait absurde de vouloir rendre la sélection naturelle responsable.

4.6 PERSPECTIVES...

On peut concevoir qu'on arrivera dans l'avenir à construire un gène fonctionnel ou chimérique à partir de ces deux gènes clonés, l'un étant délété dans le leader et l'autre en partie dans le V. On sait en effet que pour être fonctionnel un gène d'immunoglobuline doit contenir un leader, une région V, les bons sites d'épissage, les sites de reconnaissance et les séquences régulatrices. Nous avons réussi à obtenir deux clones ayant respectivement une partie de ces caractéristiques. En effet, l'un des gènes obtenus comprenait une région V sans d'autres mutations aberrantes et un site d'épissage toujours intact. L'autre

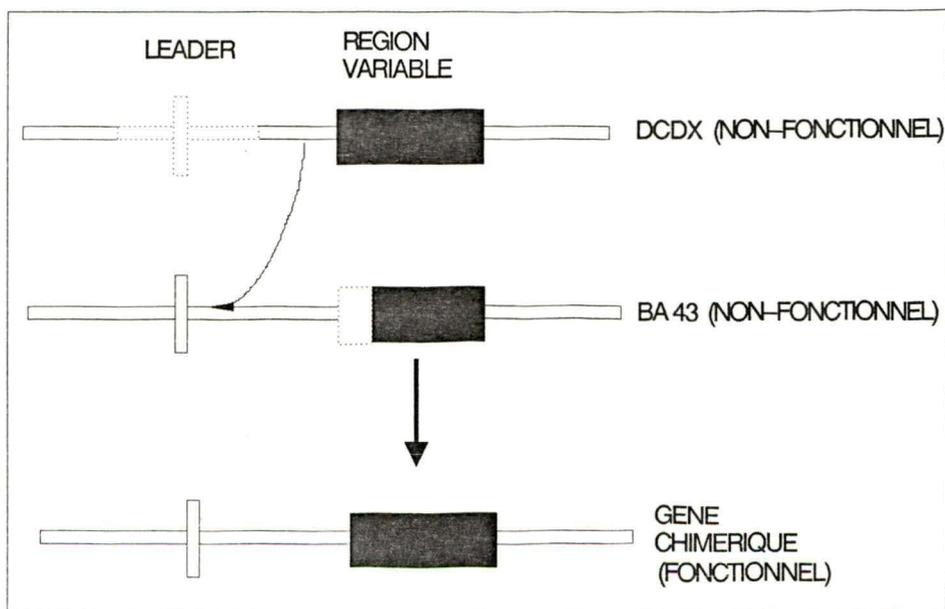


Figure 31: Représentation schématique de la construction d'un gène chimérique à partir des pseudogènes DCDX et BA 43.

gène contenait le leader avec le bon site d'épissage et les séquences régulatrices mais manquait une partie de la région V. Donc on peut imaginer qu'on puisse procéder à une recombinaison du leader du gène BA 43 et DCDX (fig 31). Ainsi , on aurait un gène fonctionnel pour faire des études sur des souris transgéniques, ce qui permettrait d'étudier leurs effets sur l'organisme.

De même, les résultats obtenus pourraient permettre dans l'avenir de vérifier si les différentes espèces et sous-espèces dans le genre *Mus* contiennent ces deux gènes, ceci en utilisant la sonde spécifique DCDX. Ce faisant, on devra vérifier s'ils sont fonctionnels. On pourra alors mieux comprendre l'évolution des gènes de la famille $V\kappa-1$, ce qui donnera une indication de l'évolution globale des gènes d'immunoglobuline. Aussi, les deux pseudogènes clonés pourraient être utilisés comme marqueur génétique pour étudier les relations qui existent entre les populations de souris et établir un arbre phylogénétique.

CONCLUSION

Depuis quelque temps un des travaux entrepris par le laboratoire du Dr. Gibson consiste à caractériser les gènes de la famille Vk1 afin de mieux comprendre l'évolution des gènes d'immunoglobuline. Dans cette perspective, notre projet avait pour objet de cloner les fragments contenant des gènes Vk-1 afin de les identifier et de démontrer leurs contributions spécifiques dans la diversité des anticorps. Nous avons réussi à identifier deux de ces gènes : DCDX et BA 43 et à les classer dans la famille Vk-1. Par la même occasion, les travaux nous ont permis de constater la duplication récente de ces deux gènes et d'observer une pression de sélection négative sur eux: les deux ayant une délétion importante dans différentes régions. Cela n'étant pas contradictoire avec l'évolution neutre, on a pu comprendre comment certaines familles des gènes d'immunoglobulines ont réussi à être fixées dans le génome tout en entraînant de ce fait l'apparition des mutations aberrantes mais potentiellement utiles à l'organisme.

De plus, l'information obtenue de ces deux gènes nous permet de penser à un mécanisme évolutif d'élimination des gènes nocifs et de leur duplication en les rendant non-fonctionnels. D'autre part, étant donné que l'identification et le séquençage de l'ensemble de tous les gènes hybridant avec la sonde PPV101 demeurent inachevés, certaines questions restent sans réponse. Ainsi, on ne peut pas savoir actuellement le nombre de ces gènes qui ne sont pas fonctionnels et s'ils sont effectivement de la famille Vk-1. Cependant les résultats que nous avons obtenus contribuent à l'avancement des travaux nécessaires pour arriver à de telles connaissances.

REMERCIEMENTS

Au terme de la rédaction de ce mémoire de maîtrise, je tiens à adresser mes remerciements à tous ceux et celles qui ont de près ou de loin, contribué à mes travaux. Plus spécialement, mes remerciements s'adressent au professeur Gibson qui m'a accepté dans son laboratoire, initié à ses recherches et supporté dans mes travaux.

Je ne voudrais pas passer sous silence le travail de l'équipe du laboratoire qui a facilité la réalisation de mon projet.

Je désire enfin remercier les professeurs Dupuis et Weber qui ont accepté de faire partie du jury.

LISTE DES REFERENCES

Aguilera R., Okazaki K., Akira S. et Sakano H. (1987). A pre-B cell nuclear protein that specifically interacts with the immunoglobulin V-J recombination sequences. *Cell* 51, 909-917.

Alt W. et Yancopoulos G.D. (1986). Regulation of the assembly and expression of variable region genes. *Annu. Rev. Immunol.* 4: 338-367

Baltimore D (1981). Gene conversion: Some implications for immunoglobulin genes. *Cell* 24, 592-594.

Bennet W.S. et Huber R. (1984) Structural and functional aspects of domain motions in proteins. *CRC Crit. Rev. Biochem.* 15 291-384.

Brack C., Hirama M., Lenhard-Schuller K. et Tonegawa S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell* 15, 1-14.

Brenner S. et Milstein C. (1966). Origin of antibody variation. *Nature* 211, 242-243

Bentley D.L. et Rabbits T.H. (1980). Human immunoglobulin variable region genes - DNA sequences of two VK genes and a pseudogene. *Nature* 288, 730-733.

Bernard N., Hozumi H. et Tonegawa S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* 15, 1133-1136.

Blackwell K.T., Moore M.W., Yancopoulos G.D., Selsing E. et Alt F. (1986). Recombination between variable region gene segments is enhanced by transcription. *Nature* 324, 585-588.

Bothwell A.L.M., Paskind M., Reth M., Imarishi-Kari., Rajewsky K., Baltimore D. (1981) Heavy chain variable region contribution to the NP^b family of antibodies. Somatic mutation evident in a 2a variable region. *Cell* 24, 625.

Brodeur P.H. et Riblet R. (1984). The immunoglobulin heavy chain variable region (IghV) locus in the mouse-one hundred IghV genes comprise seven families of homologous genes. *Eur. J. Immunol.* 14, 922-930.

Catzefflis F.M., Sheldon F.H., Ahlquist J.E. et Sibley C.G.(1987). DNA-DNA hybridisation evidence of the rapid rate of murid rodent DNA evolution. *Mol. Biol. Evol.* 4(3): 242-253.

Chung S.Y, Folsom V. et Wooley J. (1983). DNASE 1 hypersensitive sites in the chromatin of immunoglobulin K light chain genes. *Proc. Natl. Acad. Sci. USA* 80, 2427-2431.

Church G. M. et Gilbert W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci USA* . 81 1991-1995

Clarke SH, Huppi K, Ruezinsky D, Staudt L, Geerhard W et Weigert M (1985). Inter and intraclonal diversity in the antibody response to influenza hemagglutinin. *J. Exp. Med.* 161, 687-704.

Cleary M.L., Schoon E.A. et Lingrel J.B. (1981). Two related pseudogenes are the result of a gene duplication in the goat β -globin locus. *Cell* 26, 181-187.

Cohen et Givol (1983). Conservation and divergence of immunoglobulin VH pseudogenes. *Eur. Mol. Bio. Organ. J.* 2: 1795-1800.

Corbet S., Milili M., Fourgureau M. et Schiff C. (1987). Two VK germ-line genes related to GAT idiotypic network (Ab1 and Ab3/Ab1) account for the

major subfamilies of the mouse VK-1 variability subgroup. *J. Immunol.* 138,(3), 932-939.

Cory S., Tyler B.M. et Adams J.M. (1981). Sets of immunoglobulin VK genes homologous to ten cloned VK sequences: implications for the number of VK genes. *J. Mol. Appl. Genet.* 1, 103-116.

D'Eustachio P., Leinwand L., Seidman J., Keithley D. et Ruddle F.H. (1979). Chromosomal assignment of the mouse K light chain genes. *Proc. Natl. Acad. Sci. USA* 76, 2735-2739.

D'Eustachio P., Bothwell A.L.M., Takaro T.K., Baltimore D. et Ruddle F.H. (1981). Chromosomal localisation of structural genes encoding murine immunoglobulin / light chain. *J. Exp. Med.* 153 793-800.

D'Hoostelaere L.A., Huppi K., Beveerly M., Mallet C. et Potter M. (1988). The Igk L chain allelic groups among the IgK haplotypes and Igk cross-over populations suggest a gene order. *J. Immunol.* 141, (2) 652-661.

Drake J.W. (1970). *The molecular basis of mutation.* Holden-day, Oakland, California.

Dreyer W.J. et Bennet J.C. (1965). The molecular basis of antibody formation. *Proc. Natl. Acad. Sci. USA* 54, 864-869.

Early P., Huang H., Davis M., Lecalame K. et Hood L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA. *Cell* 19, 981-992.

Efstradis A., Posakony J.W., Maniatis T., Lawn R.M., O'Connell C., Spritz

R.A, Deriel J.K., Forget B.G., Weisman S.M., Slightom J.L., Blechl A.E, Smithies O., Baralle F.E, Shoulders C.C et Proudfoot N.J (1980). The structure and evolution of human β -globin gene family. *Cell* 21, 653-668.

Enquist L. et Sternberg N., (1979). In vitro packaging of Dam vectors and their use in cloning DNA fragments, *Methods Enzymol.* 68: 281-298.

Falkner F.G. et Zachau H.G. (1984). Correct transcription of an immunoglobulin K gene requires an upstream fragment containing conserved sequence elements. *Nature* 310, 71-74.

Fourgereaue M., Rocca-Serra J., Schiff C. et Tonnellet C (1984). Molecular analysis of the anti-GAT repertoire in 3 strains of mice. *Ann. Inst. Pasteur* 135, 69-75.

Frischauf A.M., Lenrach H., Poustka A.M. et Murray N. (1983). Lambda replacement vector carrying polylinkers sequences. *J. Mol. Biol.* 170, 827-842.

Gearhart P.J., Johnson N.D., Douglas R. et Hood I. (1981). IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature* 291, 29-34.

Gearhart P.J. and Bogenhagen D.F. (1983). Clusters of point mutations are found exclusively aroundrearranged antibody variable genes. *PNAS USA vol.* 80, 3439-3443.

Gibson D. (1984). Evidence for 65 electrophoretically distinct groups of light chain in Balbc and NZB myelomas. *J. Immunol.* 21 421-432.

Givol D., Zakut R., Efron K., Rechavi G., Ram D. et Cohen B. (1981). Diversity of germline immunoglobulin VH genes. *Nature* 293, 426-430.

Golding B., Gearhart P.J. et Glickman B.W. (1987). Patterns of somatic mutations in immunoglobulin variable genes. *Genetics* 115, 169-176.

Grews S., Griffin J., Huang H., Calame K. et Hood L. (1981) A single VH gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of antibody. *Cell* 25, 59.

Hood L., Perlmutter R.M., Berson B. et Griffin J.A. (1985) Diversity in the germline antibody repertoire. *J. Exp. Med.* 162 1998-2016.

Hood L., Weigert M., Gattmaitan L., Loh E. et Schilling J. (1978). Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* 276, 21-28.

Hood L., Livant D. et Blatt. (1986). One heavy chain variable region gene segment subfamily in Balbc mouse contains 500-1000 or more members. *Cell* 47, 461-470.

Hozumi N. et Tonegawa S. (1976). Evidence of somatic rearrangement of immunoglobulin genes coding for the variable and constant regions. *Proc. Natl . Acad. Sci. USA* 73, 3628-3632.

Huber R. et Bennet W. (1987). Antibody and antigen flexibility. *Nature* 326, 334-335.

Hum W.T., Lazure C., Lavigneur A., Maclean S.J. et Gibson D. (1984) Structure and genetic control of VK-1 light chains. *Annu. Rev. Immunol.* 135C, 163-170.

Jacq C., Miller J.R. et Brownlee G.G. (1977). A pseudogene structure in 5S DNA of *Xenopus Laevis*. *Cell* 12, 109-120.

Jouvin E. et Rudikoff H. (1986). Evolution of a VK gene family. *Immunogenetics* **24**, 191-201.

Joho R., Gershenfeld H. et Weissman I.L. (1984). Evolution of a multigene family of VK germline genes. *Eur. Mol. Bio. Organ. J.* **3**, 185-191.

Kaartinen M., Griffiths G.M., Hamlyn P.H., Marklam A.F., Karjalainen K., Pelkone J., Malela O. et Milstein C. (1983). Anti-oxazolone hybridomas and the structure of the oxazolone idiotype. *J. Immunol.* **130**, 937-945.

Kabat E.A, Wu T.T, and Bilfsky H. (1979). Sequences of immunoglobulin chains, National Institute of Health, Bethesda, MD.

Kimura M. (1968) The theory of neutral evolution . *Nature* **217** 624-626.

Köfler R., Noonan D.J., Levy D., Wilson M.C., Moller H., Dixon F.J. et Argyrios N (1987). Genetic elements used for a murine lupus anti-DNA autoantibody are closely related to those for antibodies to exogenous antigens. *J. Exp. Med.* **161**, 805-815.

Köfler R., Duchosal M. A. et Dixon F. (1989). Complexity, polymorphism and connectivity of mouse Vk gene families. *Immunogenetics* **29**; 65-74.

Kurusowa Y. and Tonegawa S. (1982). Organisation, Structure and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* **155**, 201-218.

Lederberg (1959). Genes and antibodies. *Science* **129**, 1649-1653.

Lewis S., Gifford A. and Baltimore D. (1985). DNA elements are asymmetrically joined during the site-specific recombination of Kappa immunoglobulin genes. *Science* **228**, 677-685.

Mackean D.J., Bell M. et Potter M. (1978). Mechanism of antibody diversity: multiple genes encode structurally related mouse K variable regions. Proc. Natl. Acad. Sci. USA 75, 3913.

Maki R., Traunecker A., Sakano H., Roeder W. et Tonegawa S. (1980). Exon shuffling generates an immunoglobulin heavy chain gene. Proc. Natl. Acad. Sci USA 77 (4) 2138-2142.

Maniatis T, Fritsch E.F. and Sambrook J (1982) Molecular cloning (a laboratory manual , Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y)

McCarrey J.R. et Riggs A. (1986). Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. Proc. Natl. Sci. USA 83, 679-683.

Miller J., Selsing E., Ronald W. et Storb U. (1982). Evolution of mouse immunoglobulin λ genes. Proc. Natl. Acad. Sci USA 79, 4681-4685.

Moynet D., Maclean S.J., Anctil D. et Gibson D. (1985). Polymorphism of K variable region (VK-1) genes in inbred mice: relationship to the IgK-Ef2 serum light chain marker. J. Immunol. 134, 3455-3460.

Müller M., Gersler T. et Schaffner W. (1988). Enhancer sequences and the regulation of gene transcription. Eur. J. Biochem. 176, 485-495.

Nishuka Y. et Leder P. (1980). Organisation and complete gene sequence of identical embryonic and plasmocytoma K V region gene. J. Biol. Chem. 255, 3691-3694.

Ng K., Ricard L., Lavigueur A., Bouvrette M., Maclean S., Cloutier D. et

Gibson D. (1989). Characterisation of allelic kappa variable region (VK-1) genes in inbred strains of mice. (sous-presse)

Painter C., Monestier M., Chew A, Bona-Dimitri A., Kasturi k., Bailey C., Scott E.U., Sidman C. et Constantin B. (1988). Specificities and V genes encoding monoclonal auto-antibodies from viable motheaten mice. *J. Exp. Med.* 167, 1137-1153.

Parslow T. et Granner D.K. (1983). Structure of a nuclease-sensitive region inside the immunoglobulin kappa gene ; evidence for a role in gene regulation. *Nucleic. Acids Res.* 11, 4775-4792.

Parslow T.G., Blair D.L., Murphy W.J. et Granner D.K. (1984). Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl Acad Sci USA* 81,2650-2654.

Pech M., Hochtl J., Schnell H et Zachau H.G. (1981). Differences between germ-line and rearranged immunoglobulin VK coding sequences suggest a localised mutation mechanism. *Nature* 291, 668-670.

Porter R.R. (1959). The hydrolysis of rabbit gamma-globin and antibodies with crystalline papain. *Biochem. J.* 73, 119-127.

Potter M. (1977). Antigen binding myeloma proteins in mice. *Adv. Immunol.* 25, 141-211.

Potter M., Newell J.B., Rudikoff S. et Haber E. (1982). Classification of mouse VK groups based on partial amino acid sequence to the first invariant tryptophan. *Mol. Immunol.* vol 19, 1619-1630.

Reininger L., Ollier P., Poncet P., Kaushik A. et Jaton J.C. (1987). Novel V

genes encode virtually identical variable regions of six murine monoclonal anti-bromelain treated red blood cell antibodies. *J. Immunol.* 138, 316-323.

Reynaud C. A., Anquez V., Dahan A. et Weill J.C. (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* 40, 283-291.

Sakano H., HÜppi K., Heinrich G. et Tonegawa S. (1979). Sequences at the somatic recombination sites of immunoglobulin light chain genes. *Nature* vol 280 288-294.

Sakano H., Maki R., Kurosowa Y., Roeder W. et Tonegawa S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy chain genes. *Nature* 286, 676-683.

Sakano H., Kurosowa Y., Weigert M.S. et Tonegawa S. (1981). Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy chain genes. *Nature* 290, 562-565.

Sanger F., Nicklen S. et Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 5463-5467.

Schiff C., Milili M. et Fougereau M. (1985). Functional and pseudogenes are similarly organised and may equally contribute to the extensive antibody diversity of the IgVH 11 family. *Eur. Mol. Bio. Organ. J* 14 1225-1230.

Schilling J., Clevinger D., Davie J.M. and Hood L. (1980). An amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V region gene segments. *Nature* 283, 35.

Shlomchik M., Nemoree D., Sato V., Van S., Carson D. et Weigert M. (1986).

Variable region sequences of murine IgM anti-IgG monoclonal auto-antibodies (rheumatoid factor). *J. Exp. Med.* 164, 407-427.

Seidman J.G., Leder A., Edgell M.M., Polsky F., Tilghman S., Tilemeir D.C. et Leder P. (1978). Multiple related immunoglobulin variable region genes identified by cloning and sequence analysis. *Proc. Natl. Acad. Sci. USA* 75, 3881-3885.

Seidman J.G., Nau M.M., Norman B., Kwan S.P., Scharff M. et Leder P. (1980). Immunoglobulin V/J recombination is accompanied by deletion of joining site and variable region segments. *Proc. Natl. Acad. Sci. USA* 77, 6022-6026.

Selsing R. et Storb V. (1981). Somatic mutation of immunoglobulins light chain variable region genes. *Cell* 25, 47-58.

Staudt L.M., Singh H., LeBowitz J.M., Sharp P.A. et Baltimore D. (1988) A lymphoid-specific protein binding to the octamer motif. *Science* 241 577-580.

Storb V., Wilson R., Selsing R. et Walfield E. (1981). A rearranged and germline immunoglobulin K genes. Different states of DNASE 1 sensitivity of constant K genes in immunocompetent and non-immune cells. *Biochemistry* 20, 990-996.

Swan D., D'Eustachio P., Leinward I. et Ruddle F.H. (1979). Chromosomal assignment of the mouse K light chain genes. *Proc. Natl. Acad. Sci. USA* 76, 2735-2739.

Todd I., Chang S.P., Perimutter, Aebersold R., Heusser C., Hood L. et

Rittenberg M.B. (1984). Immunologic memory to phosphorylcholine V hybridomas representative of group 1 antibodies utilise VK1-3 genes. *J. Immunol.* 132, 1556-1560.

Tonegawa S., Brack C., Hozumi N. et Schuller R. (1977). Cloning of an immunoglobulin variable region gene from mouse embryo. *Proc. Natl Acad. Sci. USA* 74, 3518- 3522.

Tonegawa S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.

Tonegawa S. (1985). The molecules of the immune system. *Sci. Am.* 254 122-131.

Valbuena O, Kenneth B. Marcu, Weigert M. et Perry R.P. (1978). Multiplicity of germline genes specifying a group of related mouse K chains with implications for the generation of immunoglobulin diversity. *Nature* 276, 21-28.

Van Ness B.G., Weigert M., Coleclough C., Mather E.L., Kelly D.E. et Perry R.P. (1981). Transcription of the unrearranged mouse CK locus sequence of the initiation region and comparison of activity with a rearranged VK-CK gene. *Cell* 27, 593-602.

Van Ness B.G., Coleclough C. et Weigert M. (1982). DNA between variable and joining gene segments of immunoglobulin K chain is frequently retained in cells that rearrange the K locus. *Proc. Natl Acad. Sci. USA* 79 262-266.

Vogel F. (1970). Spontaneous mutation in man. Fugel et Ronrborn, Springer-Verlag, Berlin.

Wabl M., Burrows P.D., Gabain A.V. et Steinberg C. 1985. Hypermutation at

the immunoglobulin heavy chain locus in a pre-B cell line. Proc. Natl. Acad. Sci. USA 82 479-482.

Weigert M., Cesari I.M., Yonkovich S.J. et Cohn M. (1970). Variability in the lambda light chain sequences of mouse antibody. Science 228, 1045-1048.

Weigert M. et Cohn M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature 228 1045-1047.

Weigert M., Perry R.P., Kelley D., Hunkapiller T., Schilling J. et Hood L. (1980). The joining of V and J gene segments creates antibody diversity. Nature 283 497-499.

William R.P et Lipman D. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA vol 85, 2444-2448

Wu T.W. et Kabat E.A. (1970). An analysis of the variable regions of Bence Jones proteins and myeloma light chain and their implications for antibody complementarity. J. Exp. Med. 132, 211-250.

Yancopoulos G. D, Clarke S.H. et Alt F.W. (1985). Regulation of the assembly and expression of variable region genes. Ann. Rev Immunol. 4 339-368.

Yonekawa H., Moriwaki K., Gotoh O., Hayashi Y.I, Watanabe J., Miyachita N., Petras M.L. et Tagashira Y. (1981). Evolutionary relationship among five subspecies of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA. Genetics 98: 801-816, 1981.

Zachau H., Jaenichen H., Pech M. et Lindenmaier W. (1984). Composite human VK genes and a model in evolution. Nucleic Acids Res. 12, 5249-5263.

Zachau H., Mocikat R., Falkner F. et Mertz R. (1986). Upstream regulatory sequences of immunoglobulin genes are recognised by nuclear proteins which also bind to other gene regions. *Nucleic Acids Res.* 14, 8829-8844.

Zeelon E.P., Bothwell A.L.M., Kantor F. et Schechter I. (1981). An experimental approach to enumerate the genes coding for immunoglobulin variable regions. *Nucleic Acids Res.* 9 3809-3812.