

UNIVERSITE DE LIMOGES

FACULTE des Sciences de LIMOGES

ECOLE DOCTORALE Science – Technologie – Santé

Laboratoire de Génétique biochimique et de Cytogénétique

Thèse N°1916

Thèse

pour obtenir le grade de

Docteur de l'Université de Limoges

Discipline / Spécialité : Biologie, Sciences, Santé

présentée et soutenue publiquement par

Xavier MATA

le 03 Novembre 2003

Analyse structurale et fonctionnelle de gènes voisins du *locus*
de l' α -lactalbumine caprine : application à la recherche
d'éléments *cis*-régulateurs à effet dominant.

Thèse co-dirigée par

Hubert LEVEZIEL (Université/INRA)

et Jean-Luc VILOTTE (INRA)

Bourse co-financée par l'INRA et la région Limousine

JURY :

Dr Bruce A. WHITELOW, Roslin Institute, Rapporteur.

Dr Eve DEVINOY, INRA, Rapporteur.

Prof Raymond JULIEN, Université/INRA, Examineur.

Dr Edmond P. CRIBIU, INRA, Examineur.

Dr Hubert LEVEZIEL, Université/INRA, Examineur.

Dr Jean-Luc VILOTTE, INRA, Examineur.

*A vous, mes parents et à toi,
Sèverine pour votre soutien
de tous les instants.
Cette Thèse je vous la dédie.*

Remerciements

Je suis seul à signer cette Thèse et pourtant de nombreuses personnes ont contribué par leur aide, leur soutien et leurs conseils à la faire exister. Je tiens à remercier très sincèrement tous ceux qui ont participé de près ou de loin à ce travail.

Ce travail a été effectué au centre INRA de Jouy-en-Josas, dans le laboratoire de Génétique biochimique et de Cytogénétique. Je tiens à remercier son directeur, Monsieur Edmond Cribiu pour m'y avoir accueilli.

Je remercie très sincèrement Monsieur Jean-Luc Vilotte pour ces précieux conseils, sa patience, son enthousiasme et surtout pour la confiance qu'il m'a accordé tout au long de ce travail. Je tiens aussi à le remercier pour sa grande disponibilité, en un mot Jean-Luc, MERCI.

Je remercie l'ensemble des membres de mon jury et notamment les Dr Bruce Whitelaw et Eve Devinoy pour avoir accepté de juger mon travail.

Je n'oublie pas de remercier Messieurs Raymond Julien et Hubert Levéziel, ainsi que leurs collègues de l'institut des biotechnologies (Faculté des Sciences de Limoges) qui, malgré l'éloignement géographique, ont cependant contribué au bon déroulement de cette Thèse.

Je remercie tout particulièrement mes collègues de l'équipe et de l'atelier transgénèse, Nathalie Besnard, Fabienne Le Provost, José Costa, Gaëlle Tilly et Marthe Hudrisier qui m'ont permis par leurs conseils et leur soutien quotidien de toujours avancer.

Je tiens aussi à remercier ceux qui ont participé à certaines périodes de ma Thèse mais qui ne sont plus au laboratoire aujourd'hui. Merci à Solange Soulier pour ses conseils et l'aide précieuse qu'elle m'a offerte en début de stage. Merci à Daniel Vaiman pour ces discussions dynamiques toujours enrichissantes. Merci à Rachid Essalmani pour son amitié inconditionnelle et sa joie de vivre. Merci à Karine Le Roux pour avoir su me décharger d'une partie de mon travail dans les moments difficiles.

Je remercie l'ensemble des membres de mon laboratoire d'accueil pour avoir répondu aux questions qui sortaient de mon secteur de prédilection. Je remercie notamment Séad Taourit pour son excellent travail de séquençage et sa disponibilité pour mes problèmes de contigage...Je remercie Anne Roig pour sa bonne humeur et pour m'avoir offert de retrouver l'ambiance festive du Sud-Ouest en région parisienne. Je remercie aussi Françoise Troton et Patricia Pacher pour leur gestion efficace des différentes tâches administratives me concernant.

Je remercie surtout tous les thésards du laboratoire, Céline Chantry, Gian-Mario Cosseddu, Amandine Duchesne, Mathieu Gautier, Dragan Milenkovic et Maëlle Perrocheau pour les « break » bien mérités que l'on s'est accordés au cours de ces années, au cours desquels nous avons pu partager nos doutes et bien sûr nos espoirs pour l'avenir.

Je remercie les équipes des Dr Louis-Marie Houdebine (BDR-INRA) et Hubert Laude (VIM-INRA) et notamment Sylvie Rival, Caroline Maeder et Jérôme Chapuis pour leur aide et leur collaboration fructueuse. Je remercie également Olivier Andréoletti (ENVT-INRA) pour sa contribution à l'analyse des souris transgéniques.

Je remercie aussi mes anciens collègues de FAC et surtout, Thomas Zuliani, Jérôme Boutin, Sébastien Degot et Christophe Dick pour leur présence à plus ou moins longue distance tout au long de cette Thèse.

Je remercie très sincèrement mes parents pour leur soutien inconditionnel tout au long de ces années et pour leur confiance. Je tiens aussi à remercier ma sœur pour m'avoir supporté et surtout soutenu à chacune de mes visites. Sèverine Merci, même si je pense que ton calvaire n'est pas terminé ;-).

Il reste tout ceux que je n'ai pas la place de nommer individuellement mais qui m'ont soutenu moralement par leur affection. Je pense particulièrement à mes amis de Périgueux toujours prêts à me changer les idées afin de repartir du bon pied et qui ne m'ont jamais abandonné.

Merci à tous

Résumé

Dès les premières expériences de transgénèse, la perspective d'utiliser la glande mammaire comme bio-réacteur pour la production de protéines d'intérêt pharmaceutique mais aussi pour modifier la composition du lait à des fins nutritionnelles et/ou technologiques, a été envisagée. Elle s'appuie sur l'utilisation des séquences promotrices des gènes spécifiant les protéines majeures du lait. Malgré de nombreuses études, la mise au point d'un vecteur efficace pour l'expression ciblée, sans « effet de position », de gènes dans la glande mammaire n'a toujours pas abouti. Plusieurs promoteurs du gène de l' α -lactalbumine, un des gènes spécifiant les protéines majeures du lait, utilisés en transgénèse ont conduit à des expressions dépendantes du site d'intégration. L'utilisation récente de grands fragments d'ADN (BACs et YACs) a permis de s'affranchir de cet « effet de position ». Cela a été le cas pour le vecteur développé au laboratoire, qui correspond à l'insert d'un BAC caprin de 160 kb contenant le gène de l' α -lactalbumine (BAC 41), suggérant la présence d'éléments *cis*-régulateurs dominants. Mon sujet de thèse visait à analyser plus finement cet insert.

Des expériences de transgénèse utilisant un BAC raccourci dérivé (BAC 6) nous ont permis d'effectuer une primo-localisation de ces éléments en amont de ce BAC. Dans cette région deux *loci* ont été identifiés : celui de la cycline T1 et FLJ20436. La caractérisation fonctionnelle de ces deux *loci* nous a permis de montrer que ces gènes sont actifs au sein du BAC 41 et disposent d'un mode d'expression de type ubiquiste. De façon inattendue, l'utilisation du promoteur de la cycline T1 en transgénèse a conduit à une expression certes ubiquiste mais abondante dans la lignée germinale mâle. Le gène FLJ20436 présente quant à lui un épissage complexe, conservé au cours de l'évolution. Ces études nous ont amenés à suspecter la présence de deux domaines chromatinien putatifs séparant ces gènes à expression ubiquiste de celui de l' α -lactalbumine. L'analyse structurale de ces *loci* a permis de dresser une carte précise du BAC 41 et de délimiter une région frontière séparant les deux domaines chromatinien putatifs. Une recherche en son sein d'éléments *cis*-régulateurs dominants a été initiée. L'identification et l'association de tels éléments au promoteur du gène de l' α -lactalbumine devraient contribuer à la mise au point de vecteurs d'expression efficaces pour la transgénèse mammaire.

Summary

Since the beginning of transgenesis, the use of the mammary gland as a bio-reactor to produce pharmaceutical proteins or to modify the technological and/or nutritional properties of the milk was suggested. It was based on the use of the major milk protein genes regulatory sequences. Despite numerous studies, the creation of a vector for an efficient and specific expression of genes in the mammary gland without “position effect” has not been achieved. Several promoters of α -lactalbumin gene, a major milk protein-encoding gene, have been used but resulted in integration site dependant expressions. The recent use of large genomic fragment (BACs or YACs) has allowed to avoid this “position effect”. This has been observed with a vector that was developed in our laboratory that consists of a 160 kb goat BAC insert (BAC 41) encompassing the α -lactalbumin gene, suggesting the occurrence of dominant *cis*-regulatory elements. The aim of this thesis was to further analyse this insert.

Transgenic experiments using a derived shorter BAC of 60 kb allowed us to localise these regulatory elements in a 5' distal region of the α -lactalbumin *locus*. In this region two loci were identified: the cyclin T1 and FLJ20436. Characterisation of these genes revealed that they were functional within the BAC 41 and ubiquitously expressed. Surprisingly, the use of the cyclin T1 promoter in transgenics resulted in an ubiquitous expression unexpectedly high only in male germ cells. FLJ20436 pre-mRNA has a very complex splicing pattern that is conserved during evolution. These observations led us to suspect the occurrence of two chromatin domains separating these ubiquitously expressed genes from the α -lactalbumin one. Structural analysis of these genes has allowed to define a precise restriction map of the BAC 41 and to precise the location of the potential border region within the two chromatin domains. Search for *cis*-regulatory elements within this region was initiated. Their identification and association with the α -lactalbumin promoter should contribute to the creation of efficient mammary specific expression vectors.

Sommaire

Liste des Abréviations	1
Liste des Illustrations	2
Introduction	4
Organisation des génomes eucaryotes.....	5
Organisation générale des gènes	5
Notion de domaines chromatiniens	6
Description des isolateurs.....	7
Propriétés associées aux isolateurs.....	7
Hypothèses des mécanismes d'action des isolateurs.....	8
Présentation de quelques isolateurs caractérisés	11
Séquences scs et scs' (specialized chromatine structure) chez <i>D. melanogaster</i>	11
Rétrotransposon gypsy chez <i>D. melanogaster</i>	11
Caractéristiques communes aux isolateurs chez <i>D. melanogaster</i>	12
Séquence 5' HS4 du <i>locus</i> de la β -globine de poulet.....	13
Autres éléments régulateurs ayant des propriétés isolatrices	15
Les caractéristiques des séquences MARs	15
MAR du <i>locus</i> du lysozyme de poulet	15
MAR du <i>locus</i> de l'apolipoprotéine B humaine.....	16
Séquence répétée de type Alu du <i>locus</i> de la kératine 18 humaine (K18).....	16
Introduction à la transgénèse.....	18
Historique et définitions	18
Les principales techniques utilisées en transgénèse	19
La micro-injection	19
L'infection rétro-virale	20
Utilisation des cellules souches embryonnaires	21
Le clonage	22
Les autres techniques utilisées	22
Les applications potentielles de la transgénèse	23
Modèles d'étude <i>in vivo</i> des fonctions physiologiques.....	23
Modèles de maladies	26
Sources de matériel biologique	27
L'amélioration des productions animales	29

Paramètres influençant le niveau d'expression d'un transgène	31
Variabilité des niveaux d'expression intra lignées.....	31
Variabilité des niveaux d'expression inter lignées : l'effet de position	31
Interprétation de la variation des profils d'expression	32
Mécanismes d'extinction des gènes : cas de la méthylation	32
Utilisation de grands fragments.....	34
Conclusions	35
Démarche scientifique du laboratoire	36
La glande mammaire comme bio-réacteur.....	36
Le gène de l' α -lactalbumine	38
Orientation vers l'utilisation d'un BAC caprin	39
Objectifs de la thèse	40
Utilisation d'un BAC réduit en transgénèse.....	40
Mise en évidence des différents <i>loci</i> présents au sein du BAC 41.....	41
Analyse plus fine d'une région du BAC 41	41
CHAPITRE I : Primo-localisation de régions régulatrices par analyse de l'expression en transgénèse d'un BAC raccourci.....	42
DISCUSSION DU CHAPITRE I.....	45
CHAPITRE II : Caractérisation du gène de la cycline T1 présent au sein du BAC 41. Hypothèse de l'existence de deux domaines chromatiniens.....	47
DISCUSSION DU CHAPITRE II	50
CHAPITRE III : Caractérisation de la région promotrice du gène de la cycline T1 caprine... 52	
DISCUSSION DU CHAPITRE III.....	58
CHAPITRE IV : Caractérisation du gène caprin FLJ20436, troisième gène présent au sein du BAC 41. Assignation de ce locus à un des deux domaines chromatiniens putatifs.....	64
DISCUSSION DU CHAPITRE IV.....	66
CHAPITRE V : Caractérisation de la région de 10 kb, frontière putative entre deux domaines chromatiniens.	67
Expériences menées pour caractériser la région frontière de 10 kb.....	68
Récapitulatif des travaux qui ont permis la localisation de la région frontière.....	68
Caractérisation de la région frontière	69
Matériels et méthodes.....	72
Obtention et analyse de la séquence de la région frontière	72
Sous-clonage de la région	72

Séquençage et contigage de la région	73
Analyse de la séquence	73
Analyse de la région frontière par transfection transitoire	73
Description du vecteur de base Enh-TKluc.....	73
Le gène rapporteur	74
Un promoteur ubiquiste	74
Un activateur fort	74
Un terminateur de la transcription	74
Schéma global	75
Constructions de vecteurs d'expression et sous-clonage de la région frontière.....	75
Orientation.....	76
Cultures cellulaires et la transfection	77
Milieux et Conditions de culture	77
Transfection	77
Analyse du niveau d'expression du gène rapporteur	78
Prélèvement des cellules transfectées	78
Dosage de la luciférase par bioluminescence	78
Les souris transgéniques.....	78
Préparation des constructions pour la micro-injection.....	78
Identification et caractérisation des souris transgéniques	79
Analyse du niveau d'expression du gène de l' α -lactalbumine dans les souris transgéniques et du taux de traduction	81
Résultats et discussion.....	83
Analyse comparative de la séquence de la région frontière de 10 kb	83
Obtention de la séquence.....	83
Comparaison avec les régions homologues humaine et murine	83
Recherche de zones potentiellement impliquées dans la région frontière	84
Sites potentiels de fixation de la protéine CTCF.....	84
Sites potentiels de MAR.....	85
Analyse du profil de méthylation	86
Analyse de la région de 10 kb par transfection de culture cellulaire	87
Analyse de la région de 10 kb par transgénèse chez la souris.....	90
CONCLUSIONS ET PERSPECTIVES.....	96
Conclusions de ce travail de thèse.....	97

Choix et utilisation du BAC 41 en transgénèse.....	97
Caractérisation du BAC 41	98
Perspectives	99
Fin de l'analyse de la région de 10 kb.....	99
Analyse par transfection de culture cellulaire	99
Analyse par transgénèse chez la souris	99
Analyse de sensibilité à la DNase	100
Test de la fonctionnalité des MAR prédites	100
Remise en cause de cette région de 10 kb.....	101
Cas du <i>locus</i> du lysozyme du poulet	101
Approches différentes pour la transgénèse chez la souris.....	102
BIBLIOGRAPHIE	104
ANNEXES	126

Liste des Abréviations

aa : Acide aminé
ADN/ADNc : Acide désoxyribonucléique/Acide désoxyribonucléique complémentaire
ARN/ARNm : Acide ribonucléique/Acide ribonucléique messenger
BAC : « Bacterial artificial chromosome » : chromosome artificiel de bactérie
BEt : Bromure d'éthidium
CAT : Chloramphénicol acétyl transférase
Ci : Curie
CTCF : CCCTC-binding Factor
CHO : Chinese hamster ovary
°C : Degré Celsius
DEPC : Diéthyl pyrocarbonate
DNase I : Désoxyribonucléase I
DO : Densité optique
DTT : Dithiotréitol
EDTA : Acide éthylènedinitrotétra acétique
HS : Site hypersensible à la DNase I
IPTG : Isopropyl thiogalactoside
LB : Milieu « Luria Broth »
LCR : *Locus control region*
ng/µg/mg/g : Nanogramme/microgramme/milligramme/gramme
µl/ml/l : Microlitre/millilitre/litre
µM/mM/M : Micromolaire/millimolaire/molaire
MAR : « Matrix attached regions »
Min/Sec/ms : Minute/seconde/milliseconde
mmol : Millimole
MOPS : Acide morpholinopropanesulfonique
nt : Nucléotide
pb/kb : Paires de bases/kilo paires de bases
PCR/RT PCR : « Polymerisation chain reaction » : réaction en chaine de polymérisation/
« Reverse transcription polymerisation chain reaction » réaction en chaine de
polymérisation après rétro-transcription
RLU : Unité arbitraire de luciférase
rpm : « round per minute » : tour par minute
SDS : Sodium dodécyl sulfate
SSC : Citrate de sodium salin
TBE : Tampon Tris HCl, Borate, EDTA
TK : Thymidine Kinase
Tris : Tris (hydroxyméthyl)-aminoéthane
UTR : « Untranslated region » : région de l'ARNm ne codant pas la protéine
UV : Ultraviolet
V : Volt
X-Gal : 5-bromo-4-chloro-3-indoyl-β-D-galactoside
YAC : « Yeast artificial chromosome » : chromosome artificiel de levure
WAP : « Whey Acid Protein » : Protéine acide du lactosérum

Liste des Illustrations

Fig. I.1 : Description schématique d'un gène	5
Fig. I.2 : Illustrations des propriétés d'un isolateur (Kuhn and Geyer, 2003).	9
Fig. I.3 : Mécanismes potentiels d'action des isolateurs <i>in vivo</i> (Kuhn and Geyer, 2003).	10
Fig. I.4 : Représentation schématique du complexe isolateur gypsy (West <i>et al.</i> , 2002).	12
Fig. I.5 : Schéma du <i>locus</i> des β -globines chez l'homme, la souris et le poulet (Farrell <i>et al.</i> , 2002).	14
Fig. I.6 : Schéma de la région chromosomique du <i>locus</i> de la β -globine du poulet (Farrell <i>et al.</i> , 2002).	14
Fig. I.7 : Schéma de l'organisation d'une alvéole ou <i>acinus</i> de glande mammaire (Biology of the mammary gland : http://mammary.nih.gov/).	24
Fig. I.8 : Facteurs impliqués dans le développement et la différenciation de la glande mammaire (Biology of the mammary gland : http://mammary.nih.gov/).	25
Fig. I.9 : « Whole mount » de glande mammaire à différents stades de développement (Biology of the mammary gland : http://mammary.nih.gov/).	26
Table I.1 : Caractéristiques des espèces d'élevage utilisées comme bio-réacteur (Rudolph, 1999).	28
Table I.2 : Exemples d'estimation des besoins en protéine pour la thérapie humaine (Rudolph, 1999).	28
Fig. I.10 : Schéma de la régulation épigénétique du <i>locus</i> <i>Igf2/H19</i> chez la souris (Arney, 2003).	33
Table I.3 : Exemples de protéines recombinantes produites dans la glande mammaire par des animaux transgéniques.	37
Fig. I.11 : Schéma du BAC 41.	39
Fig. I.12 : Schéma du BAC 41 et du BAC 6.	40
Fig. 1.1 : Schéma du BAC 41 et du BAC 6.	44
Fig. 1.2 : Schéma de la région chromosomique du <i>locus</i> α -lactalbumine chez l'homme.	46
Fig. 2.1 : Schéma du BAC 41.	50
Fig. 3.1 : Illustrations des propriétés d'un isolateur (Kuhn and Geyer, 2003).	53
Fig. 3.2 : Représentation schématique des structures 3D des protéines PrPc et PrPres (Prusiner, 1998).	55
Fig. 3.3 : Schéma des constructions TgA et TgB.	56
Fig. 3.4 : Schéma de la protéine prion bovine et localisation des anticorps utilisés.	60

Fig. 3.5 : Analyse de la synthèse de protéine PrP par western blot .	60
Fig. 3.6 : Analyse de la synthèse de protéine PrP par western blot, (cf. légende Fig. 3.5).	61
Fig. 4.1 : Schéma du BAC 41.	66
Fig. 5.1 : Schéma du BAC 41.	69
Fig. 5.2 : Schéma des constructions Tg ISO et Tg ISO-C.	71
Fig. 5.3 : Schéma du vecteur de base Enh Tkluc (5211 bp).	75
Table 5.1 : Caractéristiques des oligonucléotides et des sondes utilisés.	76
Fig. 5.4 : Schéma du montage pour le transfert d'acides nucléiques par capillarité.	80
Fig. 5.5 : Courbe de prédiction de MARs sur la région de 10 kb.	85
Table 5.2 : Paramètres corrélés à la courbe de prédiction de MARs (MAR-Wiz).	86
Graphique 5.1 : Résultats de la première série de transfections transitoires.	88
Graphique 5.2 : Résultats de la deuxième série de transfections transitoires.	88
Graphique 5.3 : Résultats de la troisième série de transfections transitoires.	89
Table 5.3 : Récapitulatif des résultats relatifs à la micro-injection et aux souris transgéniques.	91
Fig. 5.6 : Analyse de l'expression du gène α -lactalbumine caprin.	92
Fig. 5.7 : Analyse de l'expression du gène β -caséine murin, (cf. légende Fig. 5.6).	93
Fig. 5.8 : Analyse par Southern blot des lignées Tg ISO 41, 51 et 112.	93
Table 5.4 : Récapitulatif des données obtenues sur les trois lignées transgéniques.	94
Fig. 5.9 : Analyse de la synthèse de protéine α -lactalbumine dans les lignées Tg ISO 41, 51 et 112 par Western blot.	94
Fig. C&P.1 : Schéma du <i>locus</i> du gène du lysozyme du poulet (Chong <i>et al.</i> , 2002).	101
Fig. C&P.2 : Schéma du <i>locus</i> du gène WAP de porc (Rival-Gervier ^{a et b} <i>et al.</i> , 2003).	102

Introduction

Organisation des génomes eucaryotes

L'analyse de la structure et de la régulation des gènes présents dans les génomes eucaryotes a permis de montrer que les éléments régulateurs intervenant lors de la transcription n'étaient pas localisés seulement dans les régions proximales encadrant ou dans les unités de transcription des gènes (pour revue Levine and Tjian, 2003). Les *loci* s'étendent souvent sur des distances de plusieurs dizaines de kb au sein desquels de nombreuses séquences distales jouent un rôle non négligeable dans la régulation de l'expression des gènes.

I. Organisation générale des gènes

Dans cette thèse, nous considérerons un gène eucaryote comme défini par sa région transcrite et son promoteur minimal auxquels vont s'associer les régions régulatrices proximales et distales.

La région transcrite correspond à l'unité de transcription du gène, c'est-à-dire à la structure comprenant un ou plusieurs exons séparés par des introns, caractéristiques des gènes eucaryotes. Les exons comprennent le cadre de lecture ainsi que les séquences 5' et 3' non traduites (**Fig. I.1**).

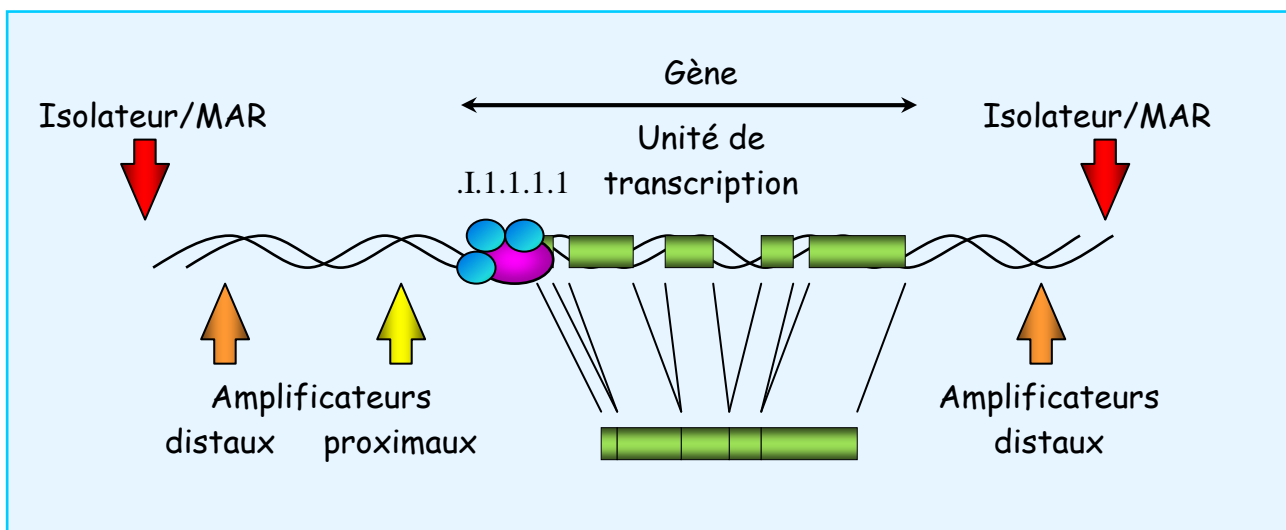


Fig. I.1 : Description schématique d'un gène

Cette structure « multi-exonique » est à l'origine de mécanismes de régulation complexe propres aux organismes supérieurs, qui permettent, à partir d'un même gène, la traduction de

plusieurs protéines grâce à des phénomènes d'épissage alternatif (Graveley, 2001). Cet épissage permet une combinatoire des exons dans une proportion encore inconnue. Ce mécanisme en association avec l'augmentation de la complexité des régions régulatrices pourrait expliquer la faible différence qui existe entre le nombre de gènes présents chez le nématode *Caenorhabditis elegans*, la mouche *Drosophila melanogaster* ou l'*Homo sapiens* (Graveley, 2001).

La région régulatrice proximale correspond au « core promoteur » où vient se fixer la machinerie responsable de la transcription des gènes. Il s'agit de la région directement en amont de l'unité de transcription (-150 à -200 pb) qui renferme les quelques séquences régulatrices nécessaires au recrutement de l'ARN polymérase (**Fig. I.1**).

La région régulatrice distale englobe un ensemble d'éléments pouvant agir en *cis* sur de grandes distances. Ces séquences sont reconnues par des protéines régulatrices qui interagissent entre-elles et avec le « core promoteur ». Il s'agit aussi de séquences d'attachement à la matrice nucléaire (MARs : Matrix Attachment Regions) permettant de délimiter différentes régions chromosomiques (**Fig. I.1**).

II. Notion de domaines chromatinien

Les dernières études indiquent que le génome humain renferme plus de 30000 gènes spécifiant des protéines, répartis de manière non homogène le long des chromosomes (Lander *et al.*, 2001). Comme indiqué ci-dessus, certaines régions régulatrices distales peuvent agir en *cis* sur de longues distances, pouvant dans certains cas dépasser les centaines de kb. Ces deux observations posent de nombreuses questions :

Comment des gènes localisés les uns à côté des autres pouvaient-ils garder des profils d'expressions spécifiques ?

Pourquoi les éléments régulateurs de promoteurs forts n'interagissaient-ils pas avec des promoteurs situés à proximité ?

Comment des gènes localisés proches de régions inactives (telle de la chromatine condensée) continuaient-ils à pouvoir s'exprimer ?

Des éléments de réponse ont été apportés avec les premières descriptions de l'organisation des génomes. Les génomes eucaryotes sont organisés en domaines chromatinien pouvant contenir aussi bien des gènes individuels que des clusters de gènes (Reitman and Felsenfeld,

1990 ; Kellum and Schedl, 1991). Ces domaines fonctionnels garantissent un profil d'expression indépendant que ce soit au cours du développement embryonnaire ou dans des cellules différenciées et ce quelque soit le degré de compaction des chromosomes. Pour diriger la régulation de l'expression des différents gènes présents au sein de ces domaines, il existe trois niveaux de contrôle :

- la présence ou non dans la cellule de facteurs de transcription venant se fixer sur des séquences d'ADN spécifiques (régions régulatrices distales) pour moduler l'activation ou la répression de la transcription de ces gènes
- le complexe multi-protéique impliqué dans le recrutement de la machinerie ARN polymérase requise pour la reconnaissance du « core promoteur » et l'initiation de la transcription
- les protéines capables de modifier et de remodeler la chromatine pour assister le déplacement du complexe transcriptionnel le long de l'ADN (Levine and Tjian, 2003).

Il a été observé que des séquences régulatrices reconnues par des effecteurs tels que des activateurs capables, une fois fixés sur l'ADN, de stimuler des gènes cibles en *cis* sur de très longues distances (plusieurs Mb), ne peuvent agir que sur des promoteurs appartenant à leur propre domaine sans interférer avec les unités de transcription des domaines chromatiniens voisins. De même, il a pu être démontré que la délétion de régions distales au niveau des frontières de ces domaines pouvaient conduire à la modification de la structure chromatinienne et du profil d'expression des gènes présents sur les domaines adjacents, impliquant la perte de séquences importantes pour l'expression indépendante des gènes. Un tel système suppose l'établissement de frontières capables d'empêcher physiquement l'interaction entre les activateurs et les promoteurs de domaines chromatiniens adjacents (pour revue West *et al.*, 2002 ; Kuhn and Geyer, 2003).

III. Description des isolateurs

1. Propriétés associées aux isolateurs

L'existence de séquences d'ADN spécifiques pouvant s'associer à des complexes protéiques et qui pourraient jouer le rôle de barrières entre ces différents domaines a été révélée au début des années 90 (Kellum and Schedl, 1991 ; Chung *et al.*, 1993). Ces séquences, nommées isolateurs, ont la propriété d'une part d'empêcher l'activation d'un promoteur si elles sont

placées entre ce dernier et un activateur, le protégeant ainsi de signaux d'activation non spécifiques et d'autre part, d'empêcher la propagation de la chromatine condensée et donc éviter l'extinction des gènes à proximité. Ces deux propriétés semblent indépendantes l'une de l'autre comme cela a pu être montré pour l'isolateur 5' HS4 (DNase I Hypersensible Site 4), frontière du *locus* de la β -globine chez le poulet (Recillas-Targa *et al.*, 2002). Ces séquences ont pu être identifiées dans de nombreuses espèces eucaryotes, de la levure à l'homme en passant par la drosophile, suggérant que ces éléments ont conservé leur rôle au cours de l'évolution.

2. Hypothèses des mécanismes d'action des isolateurs

Il semblerait que ces séquences soient capables de recruter différents complexes protéiques notamment :

- des enzymes qui vont cibler le remodelage de la chromatine telles que des HATs (Histones Acetylase Transferase)
- des protéines de type « anti-enhancer » telle que la protéine à douze doigts de zinc CTCF (CCCTC-binding Factor) chez les vertébrés ou la protéine Su(Hw) (Suppressor of Hairy-wing) chez *D. melanogaster* (pour revue West *et al.*, 2002 ; Kuhn and Geyer, 2003).

Les protéines HATs permettent l'acétylation des résidus lysines des histones H3 et H4 ce qui a pour effet de diminuer la charge positive globale des nucléosomes, et donc leur affinité pour l'ADN. La conséquence d'un tel remodelage de la chromatine est la décondensation de l'ADN au niveau des sites de fixation des HATs, permettant un accès plus facile des facteurs de transcription sur les séquences régulatrices impliquées dans l'expression des gènes. L'acétylation des histones va ainsi interrompre la propagation de la chromatine condensée responsable de l'extinction des gènes.

En ce qui concerne les protéines de type « anti-enhancer » des travaux effectués chez la drosophile montrent que la formation d'un complexe protéique est nécessaire pour pouvoir bloquer l'activation d'un promoteur par un activateur non spécifique (Cuvier *et al.*, 1998 ; Gerasimova and Corces, 1998). Les « anti-enhancers » en se fixant sur l'ADN au niveau des isolateurs permettraient la protection des gènes, au sein d'un domaine chromatinien, de l'effet des séquences régulatrices des domaines fonctionnels voisins. Ainsi, pour caractériser de telles séquences, des expériences ont été menées chez *D. melanogaster*, avec comme gène

rapporteur le « white mini gene », responsable en partie de la pigmentation des yeux chez cette espèce. Lors de son utilisation en transgénèse, il va subir les effets de la chromatine environnante avec pour conséquences une coloration des yeux des drosophiles transgéniques allant du jaune pâle au rouge. En encadrant ce gène rapporteur de séquences potentiellement isolatrices, il a été possible de suivre le pouvoir de protection de ces séquences contre les éléments chromatiniens voisins du site d'intégration de la construction comprenant le « white mini gene ». Si les séquences utilisées sont des isolateurs alors la coloration des yeux de l'ensemble des lignées de drosophiles obtenues sera jaune pâle, synonyme d'un faible niveau d'expression de ce « white mini gene » et surtout d'une régulation indépendante du site d'intégration de la construction (Kellum and Schedl, 1991 ; Chung *et al.*, 1993). Ce système a permis de caractériser plusieurs isolateurs aussi bien chez *D. melanogaster* que chez d'autres espèces (Kellum and Schedl, 1991 ; Chung *et al.*, 1993 ; Gdula *et al.*, 1996).

Une dernière hypothèse contribuant à expliquer les propriétés des isolateurs serait la capacité de ces séquences à former des boucles topologiques fonctionnelles. Cette organisation spatiale permettrait ou pas l'accès aux sites potentiels de fixation de facteurs de transcription avec l'activation ou la protection des gènes se situant respectivement sur une même boucle ou sur des boucles voisines (Fig. I.2 et I.3).

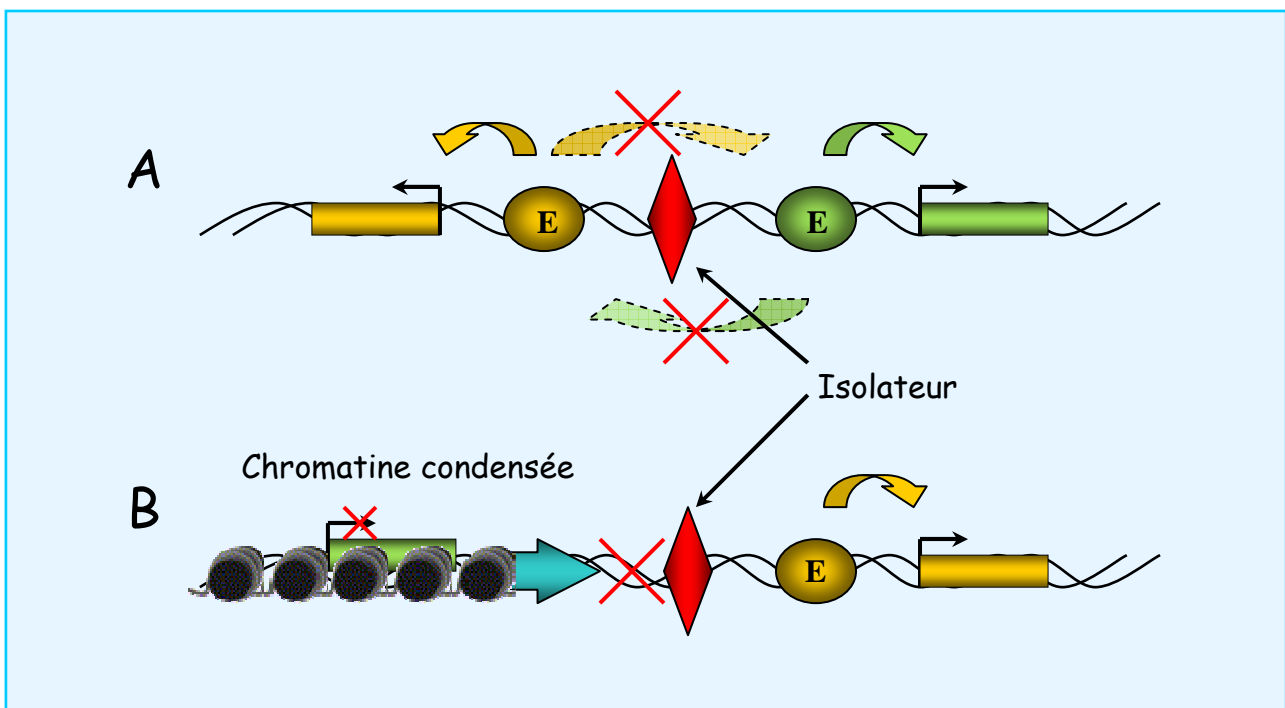


Fig. I.2 : Illustrations des propriétés d'un isolateur (Kuhn and Geyer, 2003).

A : blocage par l'isolateur de la stimulation du promoteur jaune par l'activateur vert et réciproquement du promoteur vert par l'activateur jaune.

B : l'isolateur peut empêcher la propagation de la chromatine condensée et ainsi éviter l'extinction des gènes voisins.

Les losanges rouges représentent les isolateurs, les ovales de verts et jaunes notés E les activateurs de gènes, les croix rouges l'absence d'activation et la chromatine condensée est illustrée par la présence de nucléosomes.

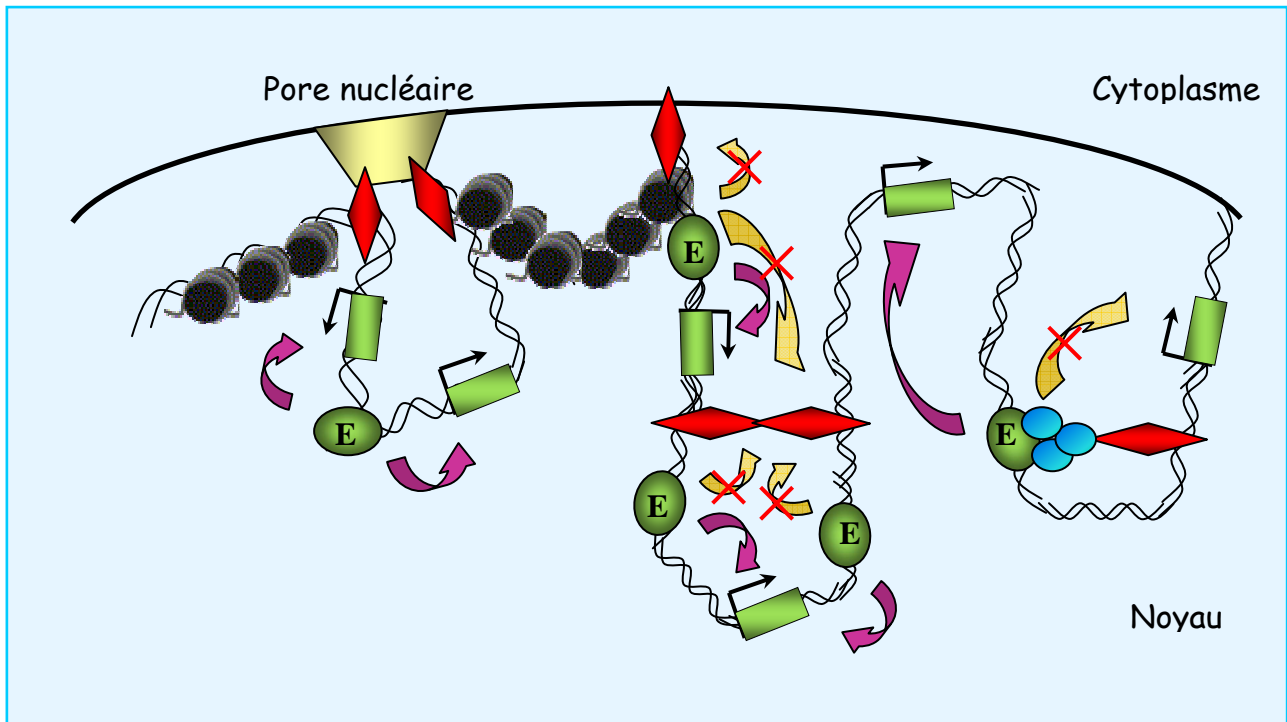


Fig. I.3 : Mécanismes potentiels d'action des isolateurs *in vivo* (Kuhn and Geyer, 2003).

Différentes boucles topologiques peuvent être mises en place afin de permettre une régulation indépendante de l'environnement chromatinien aux gènes présents sur ces boucles.

Les losanges rouges représentent les isolateurs, les ovales verts notés E les activateurs de gènes, les croix rouges l'absence d'activation et la chromatine condensée est illustrée par la présence de nucléosomes.

Au final, ces séquences isolatrices revêtent une telle importance, tant au niveau de la compréhension fondamentale du mécanisme qu'à celui des applications potentielles en transgénèse, que de nombreuses équipes s'intéressent à leur caractérisation et ce dans de nombreuses espèces eucaryotes. Cependant, malgré l'apport constant d'informations sur leurs caractéristiques, aucune structure commune ou séquence *consensus* n'a pu être mise en

évidence. Seules les propriétés bien spécifiques de ces séquences permettent de les identifier avec certitude.

IV. Présentation de quelques isolateurs caractérisés

1. Séquences scs et scs' (specialized chromatine structure) chez *D. melanogaster*

Ces séquences sont situées de part et d'autre de deux gènes *hsp70* impliqués dans la réponse aux chocs thermiques, délimitant ainsi le locus 87A7 (Udvardy *et al.*, 1985). Ces deux séquences de 1,7 kb correspondent à des sites d'hypersensibilité à la DNase I entourant une région résistante aux nucléases dont le profil est modifié lors de l'induction des gènes *hsp70* par choc thermique. Ces deux éléments combinés ont montré qu'ils pouvaient conférer une protection à un transgène contre l'effet de position et interférer entre un activateur et son promoteur (Kellum and Schedl, 1991 et 1992).

2. Rétrotransposon gypsy chez *D. melanogaster*

Cet élément régulateur est composé d'une succession plus ou moins dégénérée de 12 sites de fixation pour une protéine en doigts de zinc, la protéine Su(Hw) (Spana *et al.*, 1988), s'étendant sur à peine 340 pb. Parmi ces sites de fixation, 9 sur 12 sont nécessaires au bon fonctionnement de l'isolateur. De plus, ce système met en évidence la nécessité de la mise en place d'un complexe protéique spécifique pour activer les propriétés isolatrices de l'élément gypsy. En effet, au moins 4 protéines Su(Hw) sont nécessaires, de même que plusieurs co-facteurs spécifiques, dont mod(mdg4) qui vont venir se fixer sur Su(Hw), pour former une structure indispensable pour la propriété de blocage des activateurs (**Fig. I.4**) (Gerasimova *et al.*, 1995).

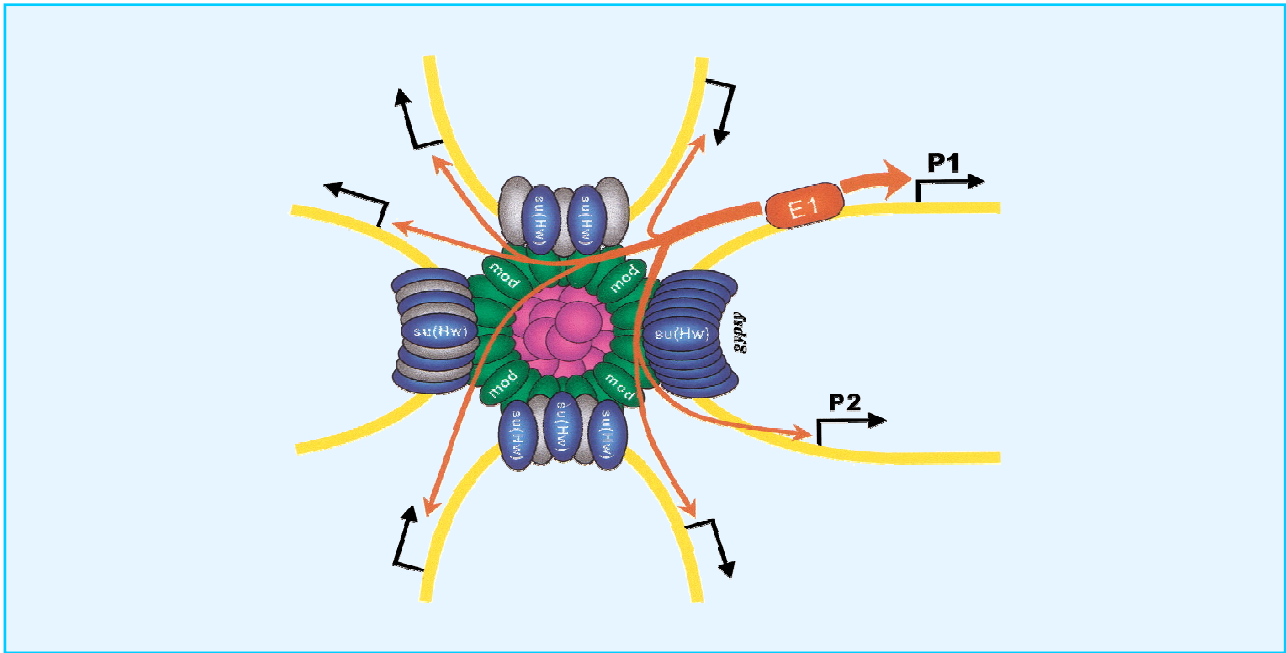


Fig. I.4 : Représentation schématique du complexe isolateur gypsy (West *et al.*, 2002).

Les ovales bleus et verts représentent les protéines Su(Hw) et mod(mdg4) respectivement. Ces protéines localisées sur plusieurs domaines chromatiniens vont venir former le complexe isolateur gypsy à l'aide d'autres protéines capables de fixer l'ADN (ovales gris). L'activateur E1 ne pourra stimuler que les gènes de son domaine comme P1 mais sera inefficace sur les autres gènes tels que P2 en raison de la présence du complexe isolateur gypsy.

3. Caractéristiques communes aux isolateurs chez *D. melanogaster*

Les éléments scs et gypsy possèdent les deux propriétés fondamentales d'un isolateur que sont le blocage de la fonction d'un activateur sur un promoteur et l'élimination de l'effet de position pour aboutir à une expression indépendante du site d'intégration du transgène et proportionnelle au nombre de copies. Par ailleurs, ces éléments ont quelques caractéristiques communes. Ce sont des régions riches en adénosine (A) et thymidine (T) sur lesquelles peut se lier préférentiellement la topoisomérase II qui pourrait jouer un rôle dans la modification locale de la chromatine. De plus, ces séquences induisent la mise en place de complexes protéiques faisant intervenir certaines protéines non spécifiques comme BEAF-32A qui se fixent sur de nombreuses séquences au sein du génome de la drosophile et notamment aux frontières de domaines chromatiniens. Plusieurs équipes ont tenté d'utiliser ces séquences pour isoler des transgènes chez les vertébrés mais les résultats sont peu convaincants (Modin *et al.*, 2000). Une des explications serait l'absence chez les vertébrés des protéines nécessaires à la mise en place des complexes capables d'isoler les gènes chez *D. melanogaster* ou une modification des séquences cibles. Pourtant, il semblait probable que de tels systèmes aient leurs équivalents chez les vertébrés.

4. Séquence 5' HS4 du *locus* de la β -globine de poulet

L'isolateur 5' HS4 du *locus* de la β -globine de poulet est un des mieux caractérisés. C'est dans le cadre de l'étude de certaines formes de β -thalassémie humaine que l'identification de cette région a pu être réalisée (pour revue Grosveld, 1999). Malgré une structure normale du promoteur proximal du *locus* des globines, aucune expression du gène n'était détectée chez les patients. La région en amont du cluster présentait une délétion, provoquant l'extinction des gènes et une modification importante de la structure de la chromatine. Cette région chromosomique détectée a été très étudiée dans plusieurs espèces pour essayer de comprendre les mécanismes impliqués dans ce phénotype. Plusieurs sites hypersensibles à la DNase I ont pu être mis en évidence, cinq chez l'homme et la souris, quatre chez le poulet, le tout étendu sur environ un vingtaine de kb de région chromosomique (**Fig. I.5**). Cette séquence distale s'est avérée indispensable pour obtenir une expression forte et spécifique des gènes présents dans le cluster des globines *in vivo* et a été appelée LCR (*Locus Control Region*) (Grosveld *et al.*, 1987). La caractérisation de cette région a montré que sur l'ensemble des sites hypersensibles, le site le plus en amont (5' HS5 chez l'homme et la souris, 5' HS4 chez le poulet) est constitutivement présent tandis que les autres ne le sont que dans les cellules érythroïdes (cellules où s'expriment spécifiquement les gènes des globines).

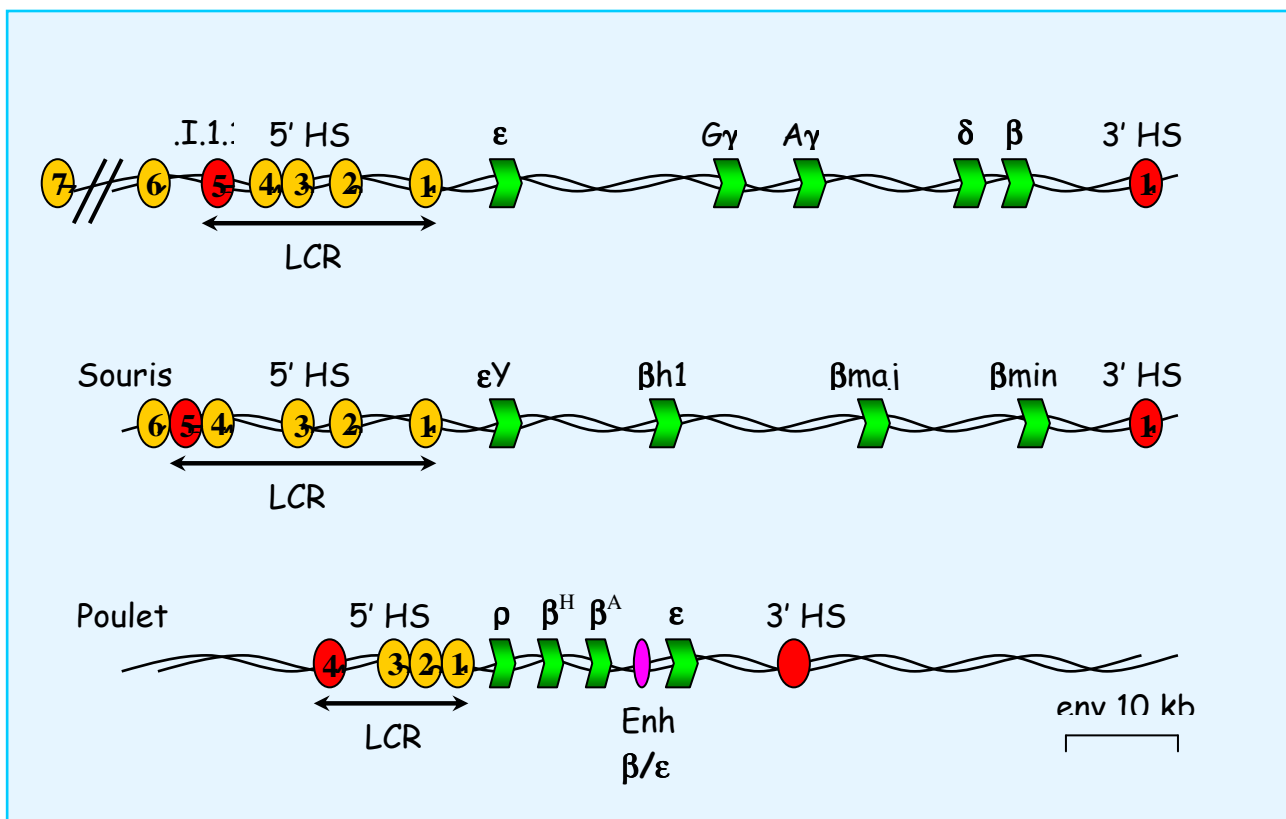


Fig. I.5 : Schéma du *locus* des β -globines chez l'homme, la souris et le poulet (Farrell *et al.*, 2002).

Les ovals rouges et oranges représentent les sites hyper sensibles à la DNase I, les ovals rouges symbolisent les frontières du domaine chromatinien des β -globines. Les flèches vertes représentent les gènes des globines. L'ovale mauve représente l'activateur des deux gènes β^H et β^A globines du poulet.

Ce site présent en permanence constitue la frontière entre deux domaines chromatiniens différents avec immédiatement en amont une région d'hétérochromatine (très dense) et la présence d'un autre gène (récepteur au folate) localisé à 16 kb du site 5' HS4 tandis qu'il est bordé en aval par une région d'euchromatine dite ouverte (Prioleau *et al.*, 1999 ; Saitoh *et al.*, 2000). Le gène du récepteur au folate présente une régulation spatio-temporelle de son expression très différente de celle des gènes des globines, induisant la nécessité d'un élément isolateur entre ces deux *loci* (**Fig. I.6**). L'étude de ce site hypersensible 5' HS4 a mis en évidence une séquence de 1,2 kb détenant toutes les propriétés définissant un isolateur : la capacité à interrompre la stimulation d'un activateur sur un promoteur et la protection d'un gène de l'effet de position (Chung *et al.*, 1997).

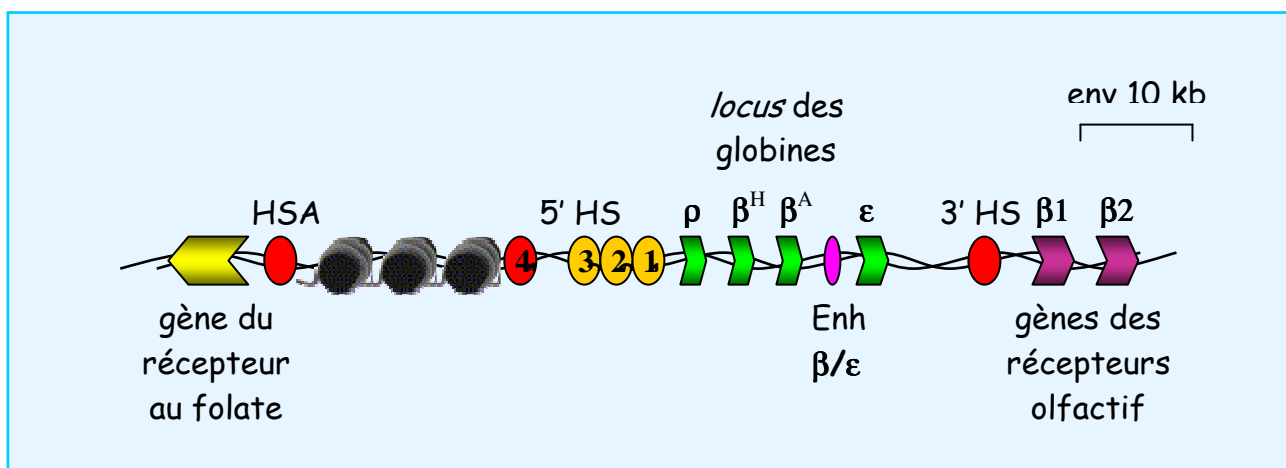


Fig. I.6 : Schéma de la région chromosomique du *locus* de la β -globine du poulet (Farrell *et al.*, 2002).

Les ovals rouges et oranges représentent les sites hyper sensibles à la DNase I, les ovals rouges symbolisent les frontières des domaines chromatiniens. Les flèches vertes, jaune et violettes représentent les gènes. L'ovale mauve représente l'activateur des deux gènes β^H et β^A globines du poulet.

V. Autres éléments régulateurs ayant des propriétés isolatrices

Un certain nombre d'éléments déjà connus se sont avérés avoir également une activité isolatrice comme certaines MARs (Forrester *et al.*, 1999 ; Porter *et al.*, 1999), l'élément répété de type Alu (Willoughby *et al.*, 2000) ou l'exemple de la structure en triple hélice (Palla *et al.*, 1997 ; Melfi *et al.*, 2000).

1. Les caractéristiques des séquences MARs

Les MARs permettent la liaison de l'ADN à la matrice nucléaire, ce qui a pour effet de favoriser le rapprochement des facteurs protéiques impliqués dans la régulation de l'expression des gènes. Elles furent les premières candidates pour la caractérisation des régions isolatrices chez les vertébrés car elles possèdent des propriétés communes avec les éléments scs et gypsy. En effet, ces régions sont riches en A et en T sans pour autant disposer des séquences consensus. Elles possèdent de nombreux sites potentiels de fixation pour la topoisomérase II. Elles correspondent souvent à des sites hypersensibles à la DNase I localisés en bordure de domaines chromatinien. Pourtant, la classe des MARs et celle des isolateurs sont deux classes différentes même si certaines MARs présentent des propriétés similaires aux éléments isolateurs (Forrester *et al.*, 1999 ; Porter *et al.*, 1999). Les MARs peuvent être étroitement associées à des régions activatrices, ce qui aura pour effet de favoriser fortement l'action de ces dernières. De telles séquences peuvent protéger un transgène mais seulement dans des cas très particuliers, spécifiques de l'activateur et du tissu (Forrester *et al.*, 1999 ; Porter *et al.*, 1999).

2. MAR du *locus* du lysozyme de poulet

La MAR la mieux caractérisée à l'heure actuelle est celle présente à l'extrémité 5' du domaine du *locus* du lysozyme du poulet (Loc and Stratling, 1988 ; Stief *et al.*, 1989). Cette MAR a la capacité de pouvoir protéger un transgène de l'effet de position aussi bien *in vitro* en culture cellulaire qu'*in vivo* dans des souris transgéniques (Bonifer *et al.*, 1990 ; Phi-Van *et al.*, 1990 ; McKnight *et al.*, 1992 et 1996, Zahn-Zabal *et al.*, 2001). Cependant, la caractérisation fine de cette séquence a permis de constater que les régions impliquées dans l'activité d'attachement à la matrice et d'isolement étaient des régions différentes, suggérant des fonctions dissociées (Phi-Van and Stratling, 1996).

3. MAR du *locus* de l'apolipoprotéine B humaine

Un autre exemple, décrit récemment, montre une association similaire de propriétés entre la MAR de l'extrémité 5' du domaine du *locus* de l'Apolipoprotéine B100 humaine et un isolateur. C'est un site chromatinien hypersensible à la DNase I dans les cellules intestinales. Cette région protège les transgènes contre l'effet de position chez la drosophile (Antes *et al.*, 2001). Une sous région de cette MAR possède une activité de blocage d'activateur polaire avec un site de fixation pour la protéine CTCF (Antes *et al.*, 2001). Ces exemples montrent que différents éléments interviennent dans la mise en place des complexes capables d'isoler efficacement les domaines chromatiniens, chacun ayant un rôle spécifique.

4. Séquence répétée de type Alu du *locus* de la kératine 18 humaine (K18)

D'autres séquences peuvent présenter, au moins partiellement, des caractéristiques d'isolateur. C'est le cas de la séquence répétée de type Alu localisée en amont du *locus* K18 (Willoughby *et al.*, 2000). La plupart de ces séquences possèdent un promoteur ARN polymérase III fonctionnel *in vitro* mais réprimé *in vivo*. L'activité transcriptionnelle de la séquence Alu du *locus* K18, par sa capacité à inhiber les interférences potentielles des unités de transcription voisines, protégerait de l'effet de position (Willoughby *et al.*, 2000).

L'ensemble de ces exemples montre que les éléments importants pour la régulation d'un gène ne se limitent pas à son promoteur et à sa région proximal. Au contraire il est de plus en plus clair que des régions parfois très éloignées de l'unité de transcription jouent un rôle essentiel sur la régulation de l'expression des différents gènes et sur leur indépendance les uns par rapport aux autres. L'incorporation de tels éléments dans des transgènes, et notamment de séquences isolatrices, serait importante afin d'optimiser leur expression et d'éviter l'influence éventuelle des activateurs et/ou inhibiteurs localisés à proximité de leur site d'insertion. Quelques expériences utilisant des fragments de taille relativement réduite (seulement quelques kb) ont permis en transgénèse une expression indépendante de l'environnement chromatinien. On citera les exemples d'expériences de transgénèse chez la souris utilisant le gène de la β -lactoglobuline du mouton (Whitelaw *et al.*, 1992) ou celui de la WAP (Whey Acid Protein) de rat (Krnacik *et al.*, 1995). Il semble cependant que ces constructions disposaient en fait d'une configuration optimale car l'ajout de séquences adjacentes, même de seulement 1 kb, conduit à une dérégulation de l'expression de ces transgènes, traduisant l'absence de séquences *cis*-régulatrices dominantes (Shani *et al.*, 1992 ; Krnacik *et al.*, 1995).

L'expression de transgènes de petite taille présente très souvent des problèmes de spécificité tissulaire et de corrélation aux nombres de copies, traduisant l'influence du site d'intégration et par là-même à nouveau l'absence de séquences *cis*-régulatrices dominantes. Ces résultats ne sont pas surprenants si on se réfère à l'exemple du *locus* de la β -globine du poulet, dans lequel la séquence LCR est étalée sur près de 25 kb et gère la régulation de l'expression d'un cluster lui-même s'étendant sur plus de 50 kb. Sans ces séquences *cis*-régulatrices dominantes, l'expression d'un transgène peut être influencée par plusieurs paramètres. Avant de mieux étayer ce point, nous reviendrons d'abord brièvement d'une part sur les méthodes d'obtention de tels animaux et d'autres part sur ses applications.

Introduction à la transgénèse

I. Historique et définitions

Au début des années 80, une révolution a éclaté dans le milieu de la recherche biologique avec l'émergence des techniques de transgénèse (Gordon *et al.*, 1980). En 1982, l'utilisation d'un gène hybride de rat codant pour une hormone de croissance a permis l'obtention de souris transgéniques « géantes » (Palmiter *et al.*, 1982). Pour la première fois en biologie, la preuve fut apportée qu'il était possible de s'affranchir de la barrière d'espèces et d'obtenir des phénotypes remarquables. Pourtant, ces expériences ont soulevé de nombreuses questions dans l'opinion publique. La croissance inhabituelle des souris était-elle due à l'expression du gène de l'hormone de croissance de rat ou s'agissait-il d'un phénomène résultant de l'utilisation d'ADN provenant d'un organisme tel que le rat ? L'analyse de l'expression du transgène dans les différents tissus de ces souris transgéniques démontrait que la forte croissance de ces animaux découlait de la sur-expression du gène de l'hormone de croissance de rat dans leur foie. A la suite de ces premiers résultats, les approches potentielles sont devenues nombreuses telles que des modèles d'études pour les effets des hormones de croissance ou le gigantisme, la possibilité d'accélérer la croissance des animaux, des solutions pour corriger les maladies génétiques, leur utilisation comme bio-réacteurs... Jamais une technique n'avait été aussi prometteuse au niveau de ces applications potentielles, même si pour certaines elle suscitait et suscite toujours des craintes et des questions éthiques (Whitelaw and Bruce, 2002).

Un animal transgénique peut être défini comme un animal ayant intégré au sein de son génome un fragment d'ADN étranger à la suite d'un transfert de gène. L'intérêt de créer de tels animaux est de pouvoir :

- passer outre la barrière d'espèces et ainsi étudier l'expression de gènes hétérologues dans des animaux modèles tels que la souris,
- intervenir sur la régulation d'un gène d'intérêt en le sur ou sous-exprimant, voire en l'invalisant, pour essayer d'y associer un phénotype et apporter ainsi des renseignements sur sa fonction,

- franchir la barrière tissulaire et analyser le phénotype créé par l'expression d'un gène donné hors de son contexte physiologique.

De plus, lors d'une intégration dans la lignée germinale, le transgène pourra être transmis aux générations suivantes pour aboutir à la création de lignées qui permettront de conserver ces animaux d'intérêt.

II. Les principales techniques utilisées en transgénèse

Dans cette présentation, nous traiterons le cas des mammifères, sachant que certaines des techniques de transgénèse ne sont pas applicables à d'autres espèces et/ou que leur utilisation nécessite parfois une adaptation en fonction de leur physiologie de la reproduction (cas des poissons et des oiseaux par exemple).

1. La micro-injection

Les techniques utilisées de nos jours pour l'obtention de tels animaux n'ont que peu évolué depuis l'époque des premières souris transgéniques. La plus utilisée d'entre elles, en transgénèse additive chez les mammifères, reste la micro-injection. Cette technique consiste à injecter dans des embryons au stade pronuclei (stade une cellule) plusieurs centaines de copies du fragment d'ADN qui doivent être intégrées dans le génome cible. Cette injection s'effectue directement dans l'un des pro-noyaux (souvent dans le pro-noyau mâle qui est plus gros, plus périphérique et donc plus accessible) à l'aide d'une micro-pipette. Cette technique, invasive et agressive pour les embryons, aboutit à la mort de beaucoup d'entre eux. Seuls 25% des embryons réimplantés dans les mères porteuses donneront naissance à des petits et seulement 15 à 25% de ces animaux seront transgéniques. Ces chiffres moyens obtenus chez la souris sont encore plus faibles dans d'autres espèces, notamment le porc, le lapin ou les ruminants (Pinkert and Murray, 1999).

Cette technique ne permet pas de maîtriser le site d'intégration ou le nombre de copies du transgène intégrées. La « non-maîtrise » du site d'intégration associée à la possible apparition de remaniements de l'ADN de l'hôte lors de l'intégration peut conduire à l'inactivation fortuite de gènes endogènes. Ce dernier phénomène affecte moins de 5% des lignées obtenues mais peut avoir des conséquences sur la viabilité et l'intérêt des animaux. D'autre part, l'expression du transgène peut être affectée par son environnement chromatinien. Ce

mécanisme appelé « effet de position » sera développé ultérieurement. De même, le nombre de copies du transgène intégrées peut influencer sur son expression. A nouveau ce mécanisme sera décrit dans la troisième section de cette introduction.

Bien que l'injection d'ADN soit effectuée à un stade très précoce du développement embryonnaire, près de 30% des fondateurs obtenus sont mosaïques (Whitelaw *et al.*, 1993 ; Wall, 2001). Ce mosaïcisme, caractérisé par une absence du transgène dans certaines lignées cellulaires, résulterait de son intégration après la première division cellulaire. Si ce phénomène affecte la lignée germinale, il en résulte un taux de transmission du transgène à la génération suivante inférieur à 50%. Si ce problème peut être considéré comme négligeable si on s'intéresse à un modèle animal comme la souris (petite taille avec une reproduction importante et rapide), il faut en tenir compte quand on décide d'utiliser des petits ou des gros ruminants car la multiplication nécessaire des lignées à partir du fondateur peut alors poser problème.

2. L'infection rétro-virale

Une seconde technique moins utilisée pour le transfert de gène est l'infection rétro-virale. Cette méthode présente l'avantage de pouvoir se faire du stade une cellule jusqu'à la moitié du développement embryonnaire et ceci sans aucune délétion ou réarrangement des chromosomes endogènes (Chan *et al.*, 2001). Par rapport à la micro-injection, cette technique permet une intégration plus efficace du transgène dans le génome cible en raison du fort taux infectieux associé aux préparations virales. Cependant, les vecteurs rétro-viraux ne peuvent incorporer que des inserts de relativement petite taille (en moyenne quelques kb) ce qui limite leur utilisation pour l'expression de transgènes. L'infection rétro-virale est par contre un outil de choix pour l'étude du devenir de lignées cellulaires lors du développement ou l'effet de mutagenèse aléatoire sur la régulation de l'expression et/ou la fonction de gènes. Si l'infection rétro-virale est réalisée au stade huit cellules par exemple, chaque cellule recevra un profil d'intégration spécifique qui pourra être suivi pendant le développement embryonnaire et la vie adulte de l'animal (pour revue Meisler, 1992 ; Rijkers *et al.*, 1994). Elle reste potentiellement une des méthodes les plus adaptées pour la mise en place de thérapies géniques (Kordower *et al.*, 2000) notamment grâce à son intégration efficace dans le génome d'un grand nombre de lignées cellulaires et pourrait être utilisée pour l'obtention de primates modèles de maladies humaines (Chan *et al.*, 2001).

3. Utilisation des cellules souches embryonnaires

Au début des années 80, deux équipes ont publié l'isolement et la culture pendant plusieurs passages de cellules souches embryonnaires de souris : les cellules ES (Embryonic Stem cells) (Evans and Kaufman, 1981 ; Martin 1981). Ces cellules, isolées à partir d'un embryon au stade blastocyste, sont cultivées sur une couche nourricière de fibroblastes primaires, afin de les maintenir indifférenciées. Après injection de ces cellules dans un blastocyste receveur, la naissance de souris chimériques (c'est-à-dire contenant une association de cellules provenant d'animaux différents) a pu être observée. Le caractère totipotent de ces cellules s'est traduit par la colonisation de l'ensemble des tissus de la souris chimère, y compris la lignée germinale (Gossler *et al.*, 1986). Ainsi la transmission du patrimoine génétique de la cellule ES à la génération suivante peut être obtenue.

Certains travaux obtenus à partir d'autres types cellulaires indiquaient que si la grande majorité des phénomènes d'intégration faisait appel à des mécanismes de recombinaison illégitime, il était possible d'observer dans certains clones une intégration par recombinaison homologue (Smithies *et al.*, 1985). Ce phénomène restant très rare, seule la transformation d'un grand nombre de cellules permet de l'observer. Ainsi, il n'y a pas d'exemple dans la littérature, à notre connaissance, de recombinaison homologue obtenue par micro-injection dans des œufs. Très vite, l'utilisation des cellules ES comme outil pour manipuler le génome de la souris par recombinaison homologue a été mise en œuvre. Des vecteurs spécialement conçus à cet effet, permettant de sélectionner les cellules transformées tout en contre-sélectionnant les intégrations issues de recombinaison illégitime ont été mis au point, notamment par l'équipe du Dr Capecchi (Thomas and Capecchi, 1987 ; Capecchi, 1989) et les premières souris obtenues à partir de cellules ES transformées ont rapidement vu le jour (Robertson *et al.*, 1986 ; et pour revue Müller, 1999).

Les possibilités offertes par cette technique sont multiples avec des modifications du génome pouvant aller de l'inactivation d'un gène donné à la simple mutation ponctuelle, avec la possibilité de suivre la régulation de l'expression d'un gène en substituant à son unité de transcription celle d'un gène rapporteur. Cependant, ce type de manipulation reste très lourd et délicat. L'étape de culture et de sélection des clones doit éviter la différenciation des cellules. L'état chimérique de l'animal rend la transmission du patrimoine génétique des cellules ES parfois difficile à obtenir. L'utilisation d'embryons receveurs tétraploïdes ou de

co-culture d'embryons au stade 8 cellules (Wang *et al.*, 1997 ; Wood *et al.*, 1993) ou l'utilisation des techniques de clonage (voir ci-dessous) pour éviter cette dernière étape semblent encore peu efficaces. Enfin, de véritables cellules ES n'ont été identifiées que chez la souris. Chez d'autres mammifères tels le rat, le lapin, le porc ou les bovins, seules des cellules pluripotentes ont été isolées (Du *et al.*, 1995 ; Wheeler, 1994 ; Notarianni *et al.*, 1997 ; First *et al.*, 1994 ; Cibelli *et al.*, 1998), il en est de même chez l'homme (Thomson *et al.*, 1998).

4. Le clonage

Développée en parallèle depuis de nombreuses années (pour revue Gurdon and Byrne, 2003), la technique de clonage est définie comme le transfert du noyau d'une cellule dans un ovocyte énucléé ou comme la fusion de ces deux cellules. En 1996, la première fusion utilisant des cellules embryonnaires fœtales de brebis a été réalisée avec succès (Campbell *et al.*, 1996). Cette technique a été appliquée depuis à de nombreuses espèces aussi bien chez les animaux d'élevage, domestiques (Baguisi *et al.*, 1999 ; Polejaeva *et al.*, 2000 ; Betthausen *et al.*, 2000 ; Onishi *et al.*, 2000 ; Chesné *et al.*, 2002) ou plus proches de l'Homme comme les primates (Chan *et al.*, 2000). D'autres types cellulaires ont pu également donner naissance à des animaux viables, notamment des cellules prélevées dans la glande mammaire adulte avec la naissance très médiatisée de la brebis Dolly (Wilmut *et al.*, 1997). Pendant la phase de culture des cellules embryonnaires fœtales, il s'est avéré possible de modifier leur génome, soit par recombinaison illégitime comme dans le cas de la brebis Polly (Schnieke *et al.*, 1997), soit même par recombinaison homologue (McCreath *et al.*, 2000 ; Keefer *et al.*, 2001 ; Denning *et al.*, 2001 ; Kuroiwa *et al.*, 2002). Le clonage s'avère donc une alternative possible à l'emploi des cellules ES pour de nombreuses espèces de mammifères. Il n'en demeure par moins que cette technique reste peu efficace (quelques % des fusions réalisées permettent d'obtenir des animaux clonés) et donne malheureusement souvent naissance à des animaux porteurs de pathologies parfois létales. Ces dernières s'expliquent notamment par une mauvaise régulation épigénétique du génome des animaux clonés (pour exemples : Rideout *et al.*, 2001 ; Dean *et al.*, 2001 ; Humpherys *et al.*, 2001 ; Jeanisch *et al.*, 2002 ; Inoue *et al.*, 2002)

5. Les autres techniques utilisées

D'autres techniques peuvent être utilisées pour créer des animaux transgéniques telles que l'utilisation du sperme comme vecteur pour transporter et faire pénétrer le transgène dans l'ovocyte receveur. Quelques résultats encourageants ont été obtenus (Lavitrano *et al.*, 1989 ;

Maione *et al.*, 1998 ; Perry *et al.*, 1999) mais la répétitivité de ce type de manipulations reste à être validée (Brinster *et al.*, 1989).

On citera également l'utilisation potentielle de spermatogonies génétiquement modifiées, qui après maturation pour former des spermatozoïdes seraient capables d'apporter le gène d'intérêt dans l'ovocyte receveur (Shinohara *et al.*, 2000 ; Nagano *et al.*, 2002 et pour revue Brinster, 2002).

Finalement, même si les différentes techniques utilisées ont peu évolué depuis les années 80, une meilleure compréhension des mécanismes intervenant lors de l'intégration du transgène (Bishop and Smith, 1989 ; Hamada *et al.*, 1993) comme la longueur des séquences d'ADN isogénique nécessaire à la recombinaison homologue ou la cassure des brins d'ADN au site d'intégration du transgène conduit à l'amélioration progressive de leurs rendements. De même, l'accroissement de nos connaissances sur la régulation de l'expression des gènes a permis de développer des vecteurs de plus en plus efficaces et spécifiques. Ces vecteurs permettent une expression des transgènes ciblée et contrôlée tant au niveau spatiale à l'échelle tissulaire, voire cellulaire qu'au niveau temporelle. Il n'en demeure pas moins que malgré ces améliorations, les techniques de transgénèse restent coûteuses, ce qui limite encore dans les faits son utilisation malgré les possibilités qu'elle offre.

III. Les applications potentielles de la transgénèse

Pour illustrer les propos précédents voici une liste non exhaustive des applications potentielles ou déjà mises en pratique utilisant la transgénèse.

1. Modèles d'étude *in vivo* des fonctions physiologiques

La création de modèles d'étude des fonctions physiologiques est nécessaire pour espérer pouvoir appréhender la complexité des interactions cellules/cellules ou cellules/matrice extra cellulaire présentes *in vivo*. L'utilisation de cultures cellulaires, aussi bien primaires que des lignées immortalisées, permet de répondre à un certain nombre de questions mais les conditions physiologiques ne sont que partiellement reproduites. Ces cellules en culture bien qu'issues d'un organe ou d'un tissu donné ne peuvent reproduire que partiellement les conditions *in vivo* en raison de l'absence des interactions entre les différents tissus que l'on retrouve dans tout organisme complexe. Ces cultures cellulaires ne constituent donc que des modèles imparfaits, souvent réducteurs. La transgénèse, en permettant d'étudier l'action d'un

gène dans toute la complexité d'un organisme entier, constitue une approche irremplaçable pour mettre en évidence les rôles qu'exercent les gènes dans la régulation des grandes fonctions physiologiques (pour revue Hanson, 2003).

Prenons comme exemple l'organe complexe que représente la glande mammaire, dont la fonction est la synthèse et la production du lait en réponse à des conditions physiologiques particulières. Les cellules épithéliales sécrétrices des protéines spécifiques du lait sont organisées en structure en trois dimensions appelées *acini* formant une lumière dans laquelle sera sécrétée le lait (**Fig. I.7**). Ces *acini* sont regroupés autour de canaux plus larges et de citernes où sera stocké avant l'utilisation par le nourrisson, le lait produit par les cellules épithéliales. La sécrétion lactée est facilitée par la présence de cellules myoépithéliales entourant chaque *acinus* ayant pour rôle une action contractile pour aider à évacuer le lait.

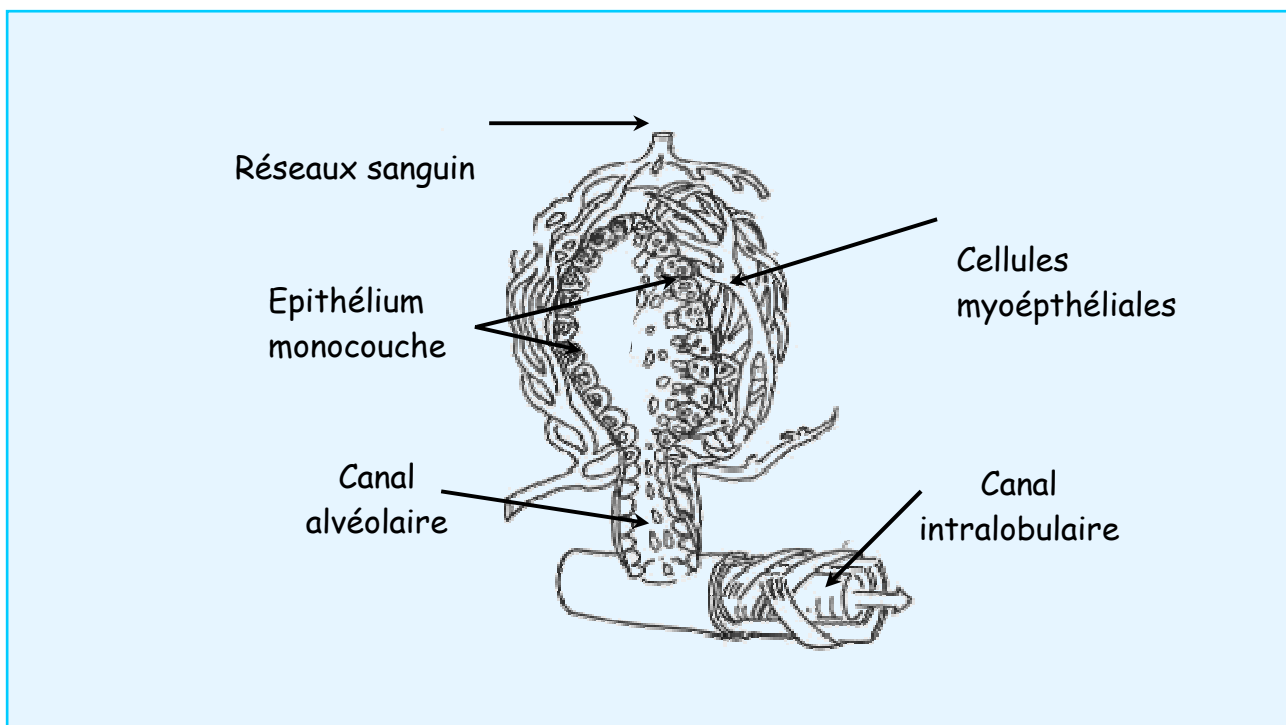


Fig. I.7 : Schéma de l'organisation d'une alvéole ou *acinus* de glande mammaire (Biology of the mammary gland : <http://mammary.nih.gov/>).

En plus d'une organisation complexe, la glande mammaire est soumise à des cycles hormonaux spécifiques intervenant lors des trois grandes phases qui sont la gestation, la lactation et l'involution (**Fig. I.8**).

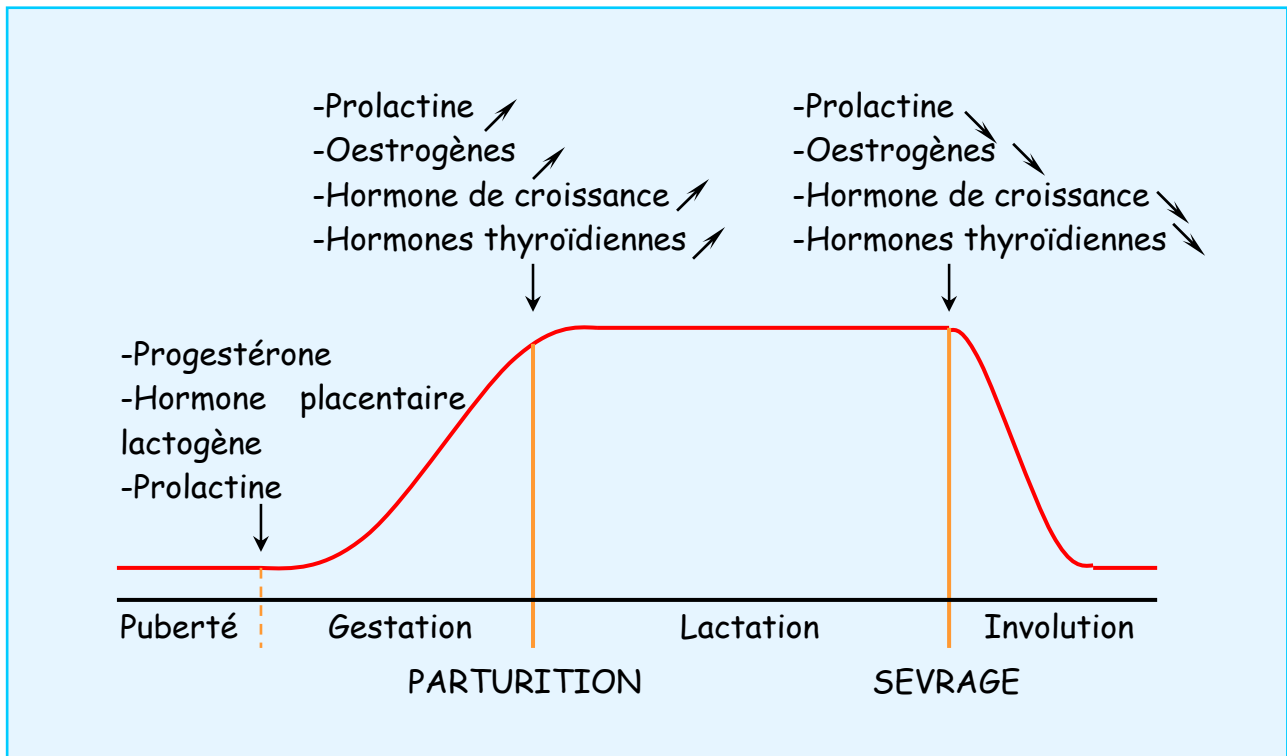


Fig. I.8 : Facteurs impliqués dans le développement et la différenciation de la glande mammaire (Biology of the mammary gland : <http://mammary.nih.gov/>).

Cette régulation multi-hormonale complexe, qui fait intervenir des hormones peptidiques et stéroïdiennes (pour revues Denamur, 1971 ; Topper and Freeman, 1980) est indispensable à la prolifération et à la différenciation des cellules de la glande mammaire (**Fig. I.9**).

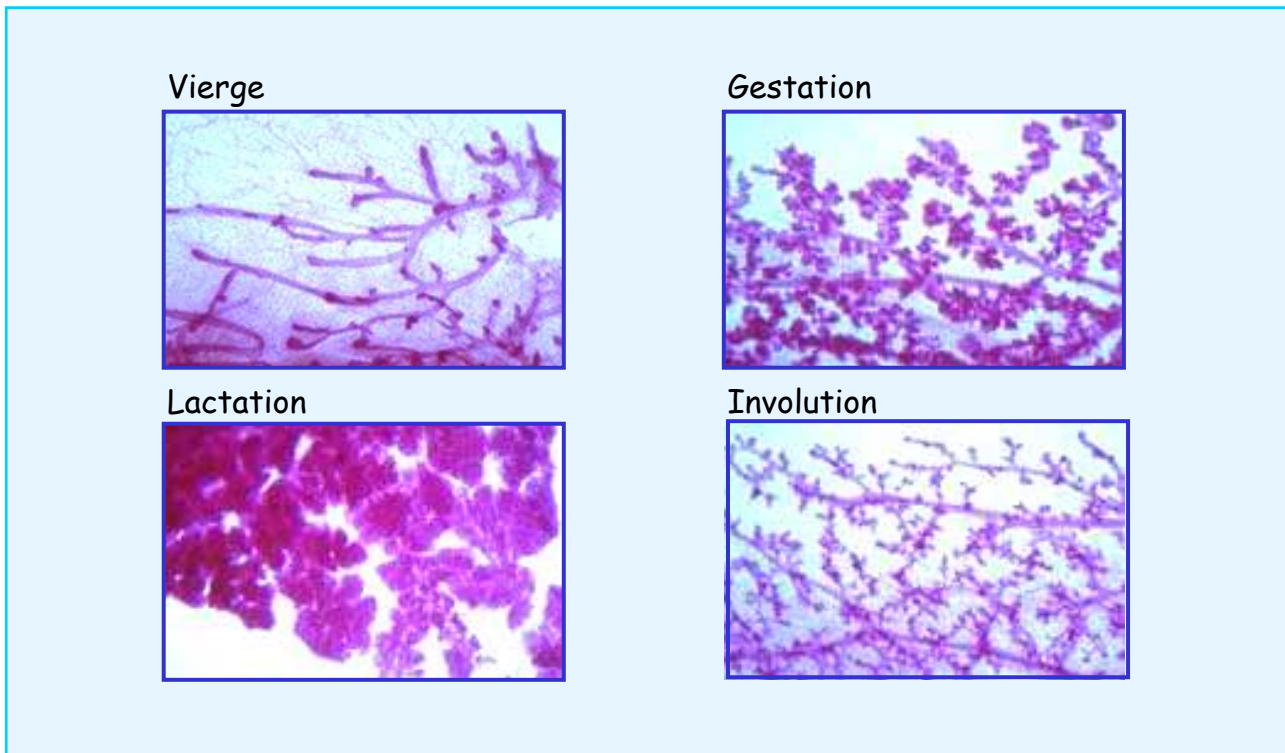


Fig. I.9 : « Whole mount » de glande mammaire à différents stades de développement (Biology of the mammary gland : <http://mammary.nih.gov/>).

Les interactions mises en œuvre lors de lactation, cellule-cellule mais aussi cellule-matrice extra-cellulaire, ne sont pas ou peu reproduites dans les modèles *in vitro* à notre disposition tels que les cellules épithéliales COMMA-1D (Danielson *et al.*, 1984), HC11 (Ball *et al.*, 1988), MCF-7 (Vickers *et al.*, 1988), BME-UV1 (Zavizion *et al.*, 1995), NISH (Ilan *et al.*, 1998).

2. Modèles de maladies

La mise au point de modèles animaux d'étude de maladies humaines est devenue un outil de choix en recherche médicale. L'addition, l'inactivation et le remplacement de gènes offrent la perspective de modèles adaptés pour valider leur implication et étudier leur rôle dans une maladie donnée mais aussi pour tester l'efficacité de traitements potentiels et pour déterminer leurs cibles cellulaires. L'utilisation de souris génétiquement modifiées pour éliminer certains types cellulaires comme les lymphocytes B et T a permis d'analyser le rôle de ces cellules dans le transport de l'agent infectieux prion lors de la neuro-invasion (Klein *et al.*, 1997 et 1998 ; Collinge and Hawke, 1998). Si le développement de modèles souris reste important (maladies cardio-vasculaires, athérosclérose, obésité, SIDA, maladies à prions...(pour revue Petters and Sommer, 2000)), ces modèles ne sont pas toujours aussi pertinents que l'on voudrait car ils ne reflètent parfois que partiellement les symptômes de la maladie en raison

des différences physiologiques entre espèces. L'utilisation d'espèces plus proches de l'homme reste souvent limitée par des difficultés méthodologiques, voire éthiques. Néanmoins, la demande de modèles animaux autres que la souris se fait de plus en plus pressante. Le rat, le lapin et le porc sont des espèces souvent citées (Bosze *et al.*, 2003).

3. Sources de matériel biologique

Les animaux transgéniques peuvent être utilisés comme source de matériel biologique avec la production de cellules immortalisées, de protéines recombinantes d'intérêt pharmaceutique ainsi que de cellules et d'organes pour la xéno-transplantation humaine.

Des animaux transgéniques exprimant des oncogènes immortalisants, tel que l'antigène T du virus SV40, peuvent être la source de lignées cellulaires immortalisées d'origine variée (Pincon-Raymond *et al.*, 1991).

Il existe plusieurs systèmes capables de produire des protéines recombinantes, les bactéries, les levures et les cellules d'insectes. Seulement ces systèmes ne peuvent produire certaines protéines à structure complexe devant subir des modifications post-traductionnelles telles que des glycosylations ramifiées, des ponts di-sulfures, des repliements particuliers et des clivages spécifiques des eucaryotes supérieurs. La culture de lignées cellulaires animales a donc été entreprise pour répondre aux besoins en protéines spécifiques mais ces cultures sont fragiles, coûteuses et de faible rendement. Une alternative est l'utilisation d'animaux transgéniques dans lesquels différentes protéines recombinantes ont pu être exprimées dans différents fluides physiologiques (sang, salive, urine ; Massoud *et al.*, 1990 ; Kerr *et al.*, 1998 ; Ryoo *et al.*, 2001 ; Golovan *et al.*, 2001) ou le blanc d'œuf (Harvey *et al.*, 2002). La glande mammaire s'est révélée être le meilleur bio-réacteur avec des quantités importantes de protéines sécrétées dans le lait (**Table I.1** et **I.2**, pour revue Rudolph, 1999), relativement faciles à collecter et à purifier en comparaison des autres bio-réacteurs (pour revues Houdebine, 2000 ; Larrick and Thomas, 2001). Plusieurs sociétés pharmaceutiques ont développé de tels modèles animaux et certaines protéines ainsi produites devraient prochainement être commercialisées. Au total, plusieurs dizaines voire des centaines de protéines, dont notamment des anticorps monoclonaux (pour revue Houdebine, 2002), devraient être produites, ainsi même si l'investissement initial nécessaire à de telles manipulations reste lourd.

	Lapin	Porc	Mouton	Chèvre	Vache
Temps de gestation (mois) :	1	4	5	5	9
Maturité sexuelle (mois) :	5	6	8	8	15
Temps entre transgénése et première lactation					
Fondateur femelle					
Lactation induite chez les femelles fondatrices prépubères :					
			9	9	16
Lactation naturelle chez les femelles Fondatrices :					
	7	16	18	18	33
Fondateur mâle					
Lactation induite chez les filles prépubères :					
			22	22	45
Lactation naturelle chez les filles :					
	15	28	31	31	57
Nombre de descendants par portée :	8	10	1 à 2	1 à 2	1
Production laitière annuelle par animal (Litre) :	4 à 5	300	500	800	8000
Quantité de protéine recombinante par femelle par an (kg) ¹ :	0,02	1,5	2,5	4	40

Table I.1 : Caractéristiques des espèces d'élevage utilisées comme bio-réacteur (Rudolph, 1999).

¹ : ces valeurs sont estimées pour un niveau d'expression moyen des protéines de 5 g/litre de lait et pour un taux de récupération de 60%.

Protéine	Besoin estimé (kg par an)	Espèce	Taille du troupeau ¹
Sérum albumine humaine	100000	vache	5400
□1-antitrypsine humaine	5000	mouton	4300
Anticorps monoclonaux	100	chèvre	58
Anti-thrombine III humaine	75	chèvre	43
Facteur IX humaine	2	porc	4

Table I.2 : Exemples d'estimation des besoins en protéine pour la thérapie humaine (Rudolph, 1999).

¹ : la taille du troupeau tient compte du niveau d'expression moyen des protéines (5 g/litre de lait), du taux de récupération de ces protéines (60%), de l'efficacité de purification de la protéine recombinante, du volume total de lait et du pourcentage d'animaux du troupeau qui ne sont pas en lactation à un moment donné (30%).

A l'heure actuelle, l'allogreffe (greffe entre deux individus de la même espèce) est le seul moyen de pouvoir remplacer un organe malade par un organe sain. La demande toujours croissante des besoins et la pénurie d'organes ont conduit les biologistes à chercher d'autres solutions parmi lesquelles la xéno greffe (greffe entre deux individus d'espèces différentes). Ces organes lorsqu'ils sont greffés chez une autre espèce sont soumis à un phénomène de rejet léthal principalement en raison de leur vascularisation. Les anticorps naturels de l'hôte vont reconnaître les antigènes des cellules du greffons et par l'activation du complément,

conduire à la destruction des cellules endothéliales ainsi qu'à des thromboses locales intenses. Des porcs transgéniques, issus de la technique de transfert de gène par le sperme, exprimant le gène DAF (Decay Activator Factor) humain, permettent d'obtenir des organes plus résistants contre l'action du complément humain responsable du phénomène de rejet lors des xéno-transplantations (Lavitrano *et al.*, 2002). D'autres expériences ont permis de créer des porcs transgéniques ayant à la surface de leurs cellules des antigènes mieux tolérés, notamment l'antigène humain Gal α 1-3Gal β 1-4Glc-NAcR. Les deux méthodes utilisées consistent en la sur-expression d'enzymes impliquées dans les phénomènes de glycosylation : la β -D-mannoside β -1,4-*N*-acetylglucosaminyltransferase III (Miyagawa *et al.*, 2001) et la α -1,3-Galactosyltransferase (Phelps *et al.*, 2003). L'ensemble de ces résultats obtenus chez des porcs transgéniques entretient l'espoir de pouvoir prochainement se servir de ces animaux comme donneurs d'organes pour l'homme.

4. L'amélioration des productions animales

L'amélioration des productions animales a toujours fait l'objet d'une recherche intensive en biologie et notamment en ce qui concerne la reproduction, les caractères d'intérêt agronomique, la nutrition et la lutte contre la maladie. L'amélioration du rendement des cheptels s'appuie sur les conditions d'élevage, les apports nutritionnels et la qualité intrinsèque des animaux par la sélection génétique. L'impact potentiel de cette recherche est limité par la variabilité génétique naturelle des individus. La transgénèse pourrait permettre d'élargir cette variabilité, ouvrant ainsi de nouvelles possibilités pour la sélection.

L'amélioration des cartes du génome des espèces d'élevage permet de se rapprocher de plus en plus de QTL (Quantitative Trait *Loci*) ou de gènes à effet majeurs agissant sur des caractères d'intérêt agronomique. L'avènement du génie génétique et de la transgénèse permettra de tester les gènes candidats ainsi identifiés. Ainsi même si plusieurs exemples démontrent l'identification de tels gènes sans faire appel à la transgénèse comme pour les gènes RN : Rendement Napole (Milan *et al.*, 2000), Booroola (Mulsant *et al.*, 2001) et DGAT1 : acylCoA:diacylglycerol acyltransferase 1 (Grisart *et al.*, 2002 ; Winter *et al.*, 2002), d'autres se sont appuyés sur (cas du gène de la Myostatine : Grobet *et al.*, 1997) ou ont utilisé des modèles transgéniques (cas du QTL de susceptibilité à l'asthme : Symula *et al.*, 1999 et du gène OCTN2 : organic cation/carnitine transporter (Zhu *et al.*, 2000)).

D'autres possibilités pourraient être la lutte contre les maladies à l'aide de gènes de résistance, l'augmentation du bien-être animal ou la diminution du pouvoir polluant de ces animaux. On citera pour exemple l'obtention de porcs transgéniques qui expriment une phytase dans leur salive capable de diminuer le taux de phosphore dans leurs excréments (Golovan *et al.*, 2001). De même la synthèse de lysostaphine dans la glande mammaire confère une résistance contre les infections staphylococcales responsables des mammites comme cela a été démontré sur un modèle souris (Kerr *et al.*, 2001).

Une autre possibilité serait la modification de la composition du lait dans le but de produire des alicaments (produit biologique modifié pour augmenter ses qualités pour la santé du consommateur) tels que du lait appauvri en lactose (sucre du lait qui est un facteur d'intolérance chez plus de 75% de la population adulte mondiale) par la synthèse d'une lactase dans le lait (Jost *et al.*, 1999) ou du lait enrichi en fer en sur-exprimant la lactoferrine ou le lysozyme humains pour augmenter la résistance naturelle des nouveaux-nés aux agents infectieux (Platenburg *et al.*, 1994 ; Maga *et al.*, 1994 ; Van Berkel *et al.*, 2002). La volonté de produire du lait toujours plus proche du lait humain pour alimenter les nourrissons a conduit à l'étude de l'influence de la taille des micelles sur les propriétés du lait. Les conclusions montrent que des micelles de petites tailles, comme dans le lait humain, ont une plus grande digestibilité et stabilité thermique. Pour reproduire cette structure micellaire dans le lait des ruminants, le taux relatif de caséine β et κ a été modifié par transgénèse, « humanisant » ainsi le lait de consommation courante (Brophy *et al.*, 2003 ; Karatzas, 2003). D'autres expériences montrent qu'il est possible d'acquérir une immunité passive contre certains agents pathogènes par la consommation de lait contenant des anticorps monoclonaux spécifiques (Castilla *et al.*, 1998 ; Sola *et al.*, 1998 ; Kolb *et al.*, 2001 ; Stowers *et al.*, 2002). Pour des raisons économiques, éthiques et bio-sanitaires, ces expériences bien qu'encourageantes ne semblent pas encore à conduire à la production de produits de consommation courante.

Paramètres influençant le niveau d'expression d'un transgène

Un des problèmes souvent rencontré en transgénèse additive et inhérent aux techniques utilisées, principalement la micro-injection, est la difficulté de prévoir et de contrôler l'expression des transgènes. L'origine de ce problème peut être multiple :

I. Variabilité des niveaux d'expression intra lignées

Il existe une variabilité des niveaux d'expression au sein d'une même lignée de souris transgéniques. Ce phénomène, appelé variégation, résulte d'un mosaïcisme cellulaire de l'expression du transgène. En effet, on a pu observer chez certains animaux une extinction du transgène dans certaines cellules d'un tissu donné. Cette extinction est dépendante du site d'insertion d'un transgène dépourvu de séquences *cis*-régulatrices dominantes et fait appel à des régulations épigénétiques pouvant agir à proximité de ce site d'intégration. Ainsi, pour un même transgène, seule une ou quelques lignées transgéniques seront sujettes à la variégation. Il en résulte des niveaux d'expression variable au sein d'une même lignée (Whitelaw *et al.*, 1993 ; Robertson *et al.*, 1995 ; Dobie *et al.*, 1996). La régulation de cette variabilité du niveau d'expression des transgènes est dépendante de la nature du fond génétique utilisé. Ainsi, pour une lignée donnée, l'introgession du transgène sur deux fonds génétiques différents peut conduire à une variation des profils d'expression.

II. Variabilité des niveaux d'expression inter lignées : l'effet de position

Le problème majeur lié aux expériences de transgénèse reste sans aucun doute « l'effet de position » qui est directement lié au site d'intégration du transgène dans le génome cible. Il a en effet été démontré au début des années 90, que le profil d'expression aberrant, observé lors d'expériences en transgénèse utilisant le gène MUP (Major Urinary Protein) chez la souris (Al Shawi *et al.*, 1990), n'était pas dû à un réarrangement lors de l'intégration du transgène en multi-copies mais bien à un effet de position dépendant du site d'intégration. Pour vérifier cette hypothèse, une banque d'ADN génomique a été réalisée à partir de la lignée de souris

transgénique présentant un profil d'expression atypique afin d'isoler le transgène présent au sein du génome. La ré-injection de ce fragment d'ADN d'intérêt dans des embryons au stade pronucléi a permis l'obtention de souris avec un profil transcriptionnel normal, soulignant l'influence des éléments régulateurs voisins du site d'intégration du transgène dans les premières souris transgéniques et l'absence de remaniement du transgène (Al Shawi *et al.*, 1990). Par la suite d'autres expériences ont été réalisées pour confirmer l'influence des éléments régulateurs présents au niveau du site d'insertion d'un transgène. Une analyse des séquences chromosomiques flanquantes d'un transgène ayant un bon niveau d'expression dans des souris transgéniques a permis de montrer que l'association de ces séquences au même transgène augmentait fortement le niveau d'expression de ce gène rapporteur dans de nouvelles souris transgéniques. Cette étude a aussi montré une spécificité de ces séquences vis-à-vis du transgène utilisé (Cranston *et al.*, 2001).

Cet effet de position va se traduire de plusieurs manières avec la plupart du temps une perturbation de la régulation de l'expression du gène d'intérêt. Ces perturbations vont engendrer divers phénomènes tels que des profils d'expression atypiques, des sur-expressions ou simplement une extinction de l'expression du transgène.

1. Interprétation de la variation des profils d'expression

L'intégration d'un transgène dépourvu de séquences isolatrices à proximité de séquences régulatrices telles que des activateurs va voir son profil d'expression influencé par ces dernières. Cela peut se traduire par l'expression de ce gène dans des tissus où il ne s'exprime pas de façon naturelle et/ou par l'augmentation de son activité promotrice conduisant à des niveaux d'expression anormalement élevés.

2. Mécanismes d'extinction des gènes : cas de la méthylation

En ce qui concerne le phénomène d'extinction de l'expression de gènes, plusieurs facteurs peuvent intervenir. Le site d'intégration est encore une fois déterminant puisque une intégration dans ou à proximité d'hétérochromatine (chromatine inactive) va entraîner l'extinction du transgène. Cette inactivation s'accompagne souvent d'une méthylation des régions promotrices. Il s'agit de l'addition d'un groupement méthyle sur une C (Cytosine) dans des régions riches en îlots CpG (Cytosine et Guanidine). Cette méthylation n'intervient pas seulement lors d'une intégration à proximité de chromatine condensée, c'est aussi une fonction biologique naturelle qui régule l'activité de gènes endogènes et agit dans le cadre de

fonctions épigénétiques telle l’empreinte parentale en inactivant un des allèles dans l’embryon (maternel ou paternel) au niveau de certains *loci* (pour revue Newell-Price *et al.*, 2000). L’exemple le mieux caractérisé pour illustrer ce phénomène est la régulation du *locus* H19/Igf2 chez la souris. En effet, les régions en amont de ces deux gènes vont subir une méthylation spécifique qui va aboutir à l’inactivation du gène H19 paternel et du gène Igf2 maternel. Cette modification est sous le contrôle d’une région ICR (Imprinting Control Region) ayant la capacité de bloquer l’activation de ces promoteurs par les activateurs environnants. Un site de fixation pour la protéine CTCF a pu être mis en évidence au niveau de cet élément régulateur (**Fig. I.10**, pour revue Arney, 2003).

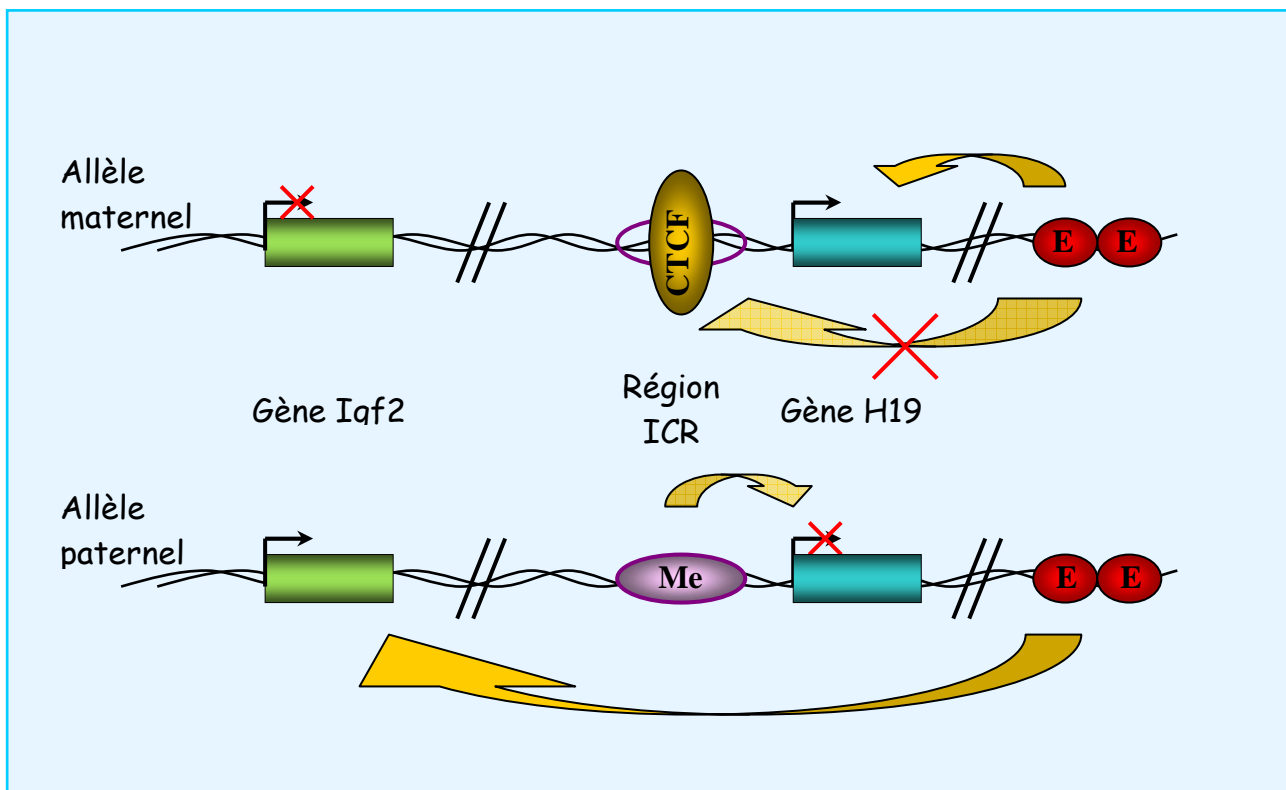


Fig. I.10 : Schéma de la régulation épigénétique du *locus* Igf2/H19 chez la souris (Arney, 2003).

Les ovales rouges représentent les activateurs du locus, la région ICR est méthylée sur l’allèle paternel et déméthylée sur l’allèle maternel permettant la fixation de la protéine CTCF en jaune.

***Allèle maternel** : Les activateurs stimulent l’expression du gène H19, tandis que la présence de la protéine CTCF sur la région ICR empêche la stimulation du gène Igf2 en amont.*

***Allèle paternel** : L’absence de la protéine CTCF permet la stimulation du gène Igf2 par les activateurs, tandis que la région ICR méthylée va éteindre le gène H19.*

Ce mécanisme d'inactivation est parfois dépendant de la structure du transgène sachant qu'un transgène sans intron est plus sujet à ce phénomène. Sa composition en nucléotides joue également un rôle car un transgène ayant un pourcentage G/C élevé, se rapprochant de celui d'un procaryote, serait plus facilement inactivé. Il s'agirait alors en fait de la mise en place d'un phénomène de protection cellulaire naturel qui se met en place lors de la réplication du matériel génétique de la cellule.

De même, il a pu être observé qu'un trop grand nombre de copies intégrées (au delà d'une centaine) au même endroit peut aboutir à une inactivation du transgène (Flavell, 1994 ; Garrick *et al.*, 1998). A nouveau cela refléterait la mise en place de mécanismes de défense de la cellule agissant contre les phénomènes de transposition et de duplication. Cette inactivation s'accompagne souvent de la méthylation de l'ADN.

3. Utilisation de grands fragments

L'addition à un transgène de séquences isolatrices a depuis longtemps été envisagée pour répondre à l'effet de position. Malheureusement, l'utilisation des isolateurs existants hors de leur contexte a donné souvent des résultats décevants *in vivo* (Modin *et al.*, 2000). Une approche alternative est apparue avec le développement de nouveaux vecteurs de clonage.

La mise au point des chromosomes artificiels a permis le sous clonage de l'ADN génomique de nombreuses espèces en fragments plus ou moins grands dont la taille peut aller de quelques dizaines de kb (PAC : Phage Artificiel Chromosome) à plusieurs centaines de kb voire un Mb (BAC : Bacterial Artificial Chromosome et YAC : Yeast Artificial Chromosome). D'autres chromosomes artificiels ont ensuite été développés afin de pouvoir contenir des fragments d'ADN génomique pouvant atteindre quelques Mb tels que MAC (Mammalian Artificial Chromosome). Ces chromosomes artificiels ont été mis au point dans plusieurs espèces de mammifères dont le porc et l'homme et ont été ensuite utilisés en transgénèse (Tomizuka *et al.*, 2000 ; Kuroiwa *et al.*, 2000 ; Poggiali *et al.*, 2002 ; Robl *et al.*, 2003).

Ces fragments d'ADN génomique offrent la possibilité d'appréhender l'environnement des gènes au sens large et non plus seulement au niveau du promoteur proximal. Ils permettent de pouvoir envisager la présence de l'ensemble des éléments *cis*-régulateurs d'un gène donné dans le fragment d'ADN injecté et ainsi de conserver le cadre endogène nécessaire à la

régulation du *locus* (pour revue Giraldo and Montoliu, 2001). L'importance de ces vecteurs s'est imposée dès que les premières expériences de micro-injection pro-nucléaire de ces fragments ont montré d'une part leur capacité à s'intégrer dans l'ADN du génome cible et d'autre part à conduire à l'expression parfaitement régulée des transgènes, permettant même de corriger des phénotypes résultant de l'inactivation de gènes endogènes (Brem *et al.*, 1996). Des techniques se sont ensuite développées autour de ces grands fragments d'ADN génomique afin de pouvoir les modifier par recombinaison homologue chez *Escherichia coli* ou *Saccharomyces cerevisiae* (Yang *et al.*, 1997 ; Narayanan *et al.*, 1999 ; Muyrers *et al.*, 1999 ; Lee *et al.*, 2001 ; Valenzuela *et al.*, 2003). Toutefois, malgré les nombreux avantages que procurent l'utilisation de ces grands fragments d'ADN génomique en transgénèse, il existe quelques inconvénients. L'utilisation de tels fragments pose le problème de la possible co-intégration d'autres gènes présents à proximité du gène d'intérêt. Le produit de l'expression de ces gènes pourrait interférer avec l'analyse du phénotype résultant. De plus, les techniques de recombinaison homologue sont lourdes en pratique et le pourcentage de réussite reste encore très faible même si on peut espérer une forte augmentation de leur efficacité dans les années à venir, les techniques évoluant rapidement au cours de ces dernières années avec la mise au point de moyens de recombinaison et de sélection toujours plus performants (Testa *et al.*, 2003 ; Valenzuela *et al.*, 2003).

III. Conclusions

Malgré ces inconvénients, la transgénèse est très utilisée de nos jours pour étudier toutes sortes de gènes intervenant à différents niveaux pour réguler des fonctions biologiques trop complexes pour être appréhendées avec des systèmes plus limités comme les cultures cellulaires. Le développement récent des techniques autour de la transgénèse laisse espérer beaucoup plus de facilité dans les années à venir en ce qui concerne l'efficacité de la micro-injection avec notamment l'utilisation d'enzymes recombinase pour augmenter la fréquence de et/ou cibler l'intégration du transgène dans le génome, la manipulation des grands fragments d'ADN génomique ou encore la compréhension des mécanismes impliqués dans la régulation des différents gènes les uns par rapport aux autres (Maga *et al.*, 2003 ; Testa *et al.*, 2003 ; Valenzuela *et al.*, 2003 ; Epinat *et al.*, 2003 ; Porteus and Baltimore, 2003). Ces améliorations laissent présager l'émergence de multiples applications aussi bien médicales qu'agronomiques dont notamment l'utilisation de la glande mammaire comme bio-réacteur ou la production de laits « à la carte » pour répondre à des besoins nutritionnels particuliers.

Démarche scientifique du laboratoire

I. La glande mammaire comme bio-réacteur

Dès les premières expériences de transgénèse, la glande mammaire a été désignée comme un organe cible potentiel pour la production de protéines hétérologues dans le lait (Palmiter *et al.*, 1982). L'importante activité sécrétoire de cet organe, sa spécificité par rapport au reste de l'organisme, sa capacité à effectuer des modifications co- et post-traductionnelles complexes et la facilité de collecte du lait ouvrait en effet la perspective d'utiliser la glande mammaire comme bio-réacteur pour la production de protéines d'intérêt pharmaceutique (Rudolph, 1999), mais aussi pour modifier la composition du lait à des fins nutritionnelles ou technologiques (Mercier and Vilotte, 1997, Jost *et al.*, 1999). La transgénèse pouvait de plus s'appuyer sur l'utilisation des séquences promotrices des gènes spécifiant les principales protéines du lait qui présentent une expression forte et limitée aux seules cellules épithéliales mammaires en cours de lactation (pour revue Rudolph, 1999). En effet, chez les ruminants, plus de 90% des protéines du lait résultent de l'expression, spécifique à ce tissu, de seulement six gènes (pour revue Mercier and Vilotte, 1993). Malheureusement, l'ensemble des promoteurs de ces gènes, utilisé à ce jour en transgénèse, souffre de « l'effet de position » (**Table I.3**).

Promoteur	Protéine recombinante	Gène/ADNc	Espèce	Quantité de protéine sécrétée	.I.1.1.1.1.1	Aut eurs
β -lactoglobuline ovine	α 1-antitrypsine humaine	Gène	Souris	7 mg/ml	Archibald AL <i>et al.</i> , 1990	
β -lactoglobuline ovine	α 1-antitrypsine humaine	Gène	Mouton	35 mg/ml	Wright G <i>et al.</i> , 1991	
β -lactoglobuline ovine	Fibrinogène	Gène	Souris	2 mg/ml	Prunkard D <i>et al.</i> , 1996	
β -lactoglobuline ovine	Fibrinogène	Gène	Mouton	0,5-5 mg/ml	Foster D <i>et al.</i> , 1997	
β -lactoglobuline ovine	Sérum albumine humaine	ADNc	Souris	2,5 mg/ml	Shani M <i>et al.</i> , 1992	
β -lactoglobuline ovine	Interféron γ humain	ADNc	Souris	20 ng/ml	Dobrovolsky <i>et al.</i> , 1993	
β -lactoglobuline ovine	Facteur IX humain	ADNc	Mouton	25 ng/ml	Simons JP <i>et al.</i> , 1988 ; Clark AJ <i>et al.</i> , 1989	
β -lactoglobuline ovine	α 1-antitrypsine humaine	ADNc	Souris/ mouton	18 mg/ml	McClenaghan <i>et al.</i> , 1991	
WAP murine	PS2 humain	Gène	Souris	1,5 μ g/ml	Tomasetto C <i>et al.</i> , 1989	
WAP murine	GH humaine	Gène	Souris	500 μ g/ml	Reddy <i>et al.</i> , 1991	
WAP murine	TPA humaine	ADNc	Souris	460 ng/ml	Gordon K <i>et al.</i> , 1987 ; Pittius CW <i>et al.</i> , 1988	
WAP murine	TPA humaine	ADNc	Chèvre	3 μ g/ml	Ebert KM <i>et al.</i> , 1991	
WAP murine	CD4 humain	ADNc	Souris	200 ng/ml	Yu S <i>et al.</i> , 1989	
WAP murine	Anticorps anti-coronavirus	ADNc	Souris	5-6 mg/ml	Castilla J <i>et al.</i> , 1998	
WAP murine	Protéine C humaine	ADNc	Porc	1 mg/ml	Velander WH <i>et al.</i> , 1992	
WAP de lapin	GH humaine	Gène	Souris	22 mg/ml	Devinoy E <i>et al.</i> , 1994 ; Stinnakre <i>et al.</i> , 1991	
WAP de lapin	GH bovine	Gène	Souris	6 mg/ml	Thepot D <i>et al.</i> , 1995	
WAP de lapin	κ -caséine	Gène	Souris	2,5 mg/ml	Hiripi L <i>et al.</i> , 2000	
WAP de lapin	HIP/HAP humaine	Gène	Souris	11,2 mg/ml	Christa <i>et al.</i> , 2000	
WAP de lapin	GH bovine	ADNc	Souris	280 μ g/ml	Petitclerc <i>et al.</i> , 1993	
β -caséine de rat	FSH humaine	ADNc	Souris	5 μ g/ml	Greenberg NM <i>et al.</i> , 1991	
β -caséine caprine	TPA humaine	ADNc	Chèvre	1-3 mg/ml	Ebert KM <i>et al.</i> , 1994	
β -caséine de lapin	Interleukine 2	ADNc	Lapin	430 ng/ml	Buhler TA <i>et al.</i> , 1990	
α -caséine bovine	TPA humaine	ADNc	Souris/lapin	50 μ g/ml	Riego E <i>et al.</i> , 1993	
α -caséine bovine	Urokinase humaine	ADNc	Souris	2 mg/ml	Meade H <i>et al.</i> , 1990	
MMTV	α s1-caséine bovine	ADNc	Souris	0,3 mg/ml	Yom HC <i>et al.</i> , 1993	

Table I.3 : Exemples de protéines recombinantes produites dans la glande mammaire par des animaux transgéniques.

La mise au point de vecteurs efficaces pour l'expression ciblée de gènes dans la glande mammaire est donc un objectif scientifique présentant un intérêt académique mais aussi des applications importantes tant sur un plan industriel que zootechnique.

II. Le gène de l' α -lactalbumine

C'est dans ce cadre que le laboratoire s'est intéressé depuis 1984 à un gène spécifiant une protéine majeure du lait : le gène de l' α -lactalbumine. Ce choix réside d'une part dans la fonction biologique de cette protéine qui est impliquée dans la synthèse du lactose dans la glande mammaire (Ebner and Brodbeck, 1968), et d'autre part du fait de son expression forte, limitée aux seules cellules épithéliales mammaires en fin de gestation et pendant la lactation. Son promoteur est donc un bon candidat pour cibler l'expression de protéines hétérologues dans le lait. L' α -lactalbumine est une métalloprotéine calcique qui modifie la spécificité de substrat de la β -1-4 UDP galactosyl-transférase (Takase and Ebner, 1984). Cette enzyme, localisée dans la membrane de l'appareil de Golgi, fixe un résidu galactose sur les résidus N-acétylglucosamine des protéines. En présence d' α -lactalbumine, l'association stœchiométrique de ces deux protéines forme le complexe binaire lactose synthétase (EC 2.4.1.22). Cette dimérisation crée virtuellement un site spécifique de fixation du glucose, conduisant à la synthèse du lactose (glucose + UDP-galactose \Rightarrow UTP + lactose). Ce sucre, prépondérant dans le lait, est un des principaux régulateurs osmotiques de la sécrétion lactée (Stinnakre *et al.*, 1994). L' α -lactalbumine dériverait de la duplication d'un gène ancestral du lysozyme (Qasba and Safaya, 1984 ; Hayssen and Blackburn, 1985). La spécialisation de cette protéine serait à l'origine de la lactation. Chez les ruminants, l'existence de pseudogènes α -lactalbumine a été mise en évidence (Soulier *et al.*, 1989 ; Vilotte *et al.*, 1991 ; Vilotte *et al.*, 1993). Ces pseudogènes seraient localisés en aval du *locus* du gène (Stinnakre *et al.*, 1999). Il est à noter que chez les Ruminants, et à nouveau contrairement à ce qui est observé chez les autres espèces, le lysozyme apparaît sous forme d'une famille multigénique (Dobson *et al.*, 1984 ; Jolles *et al.*, 1984 ; Irwin and Wilson, 1989 et 1990). Bien que l'ensemble de ces séquences soient localisées à proximité les unes des autres dans le génome, des expériences d'électrophorèse en champ pulsé ont démontré que les événements à l'origine de l'apparition de ces deux familles étaient indépendants (Gallagher *et al.*, 1993).

Plusieurs promoteurs courts du gène de l' α -lactalbumine, issus de différentes espèces, ont été utilisés en transgénèse (Vilotte *et al.*, 1989 ; Stinnakre *et al.*, 1991 ; Maschio *et al.*, 1991 ; Soulier *et al.*, 1992 ; Hochi *et al.*, 1992 ; Ninomiya *et al.*, 1994 ; Soulier *et al.*, 1999 ; Jost *et al.*, 1999). Il s'est avéré que, comme cela a été observé avec d'autres promoteurs de même taille de gènes des protéines du lait, tous étaient influencés par leur environnement chromatinien avec des niveaux d'expressions variables pouvant aller jusqu'à l'extinction du

transgène. Si la spécificité tissulaire s'est avérée respectée dans la plupart des lignées, avec une forte expression du gène dans le seul tissu épithéliale de la glande mammaire, il a pu parfois être observé de faibles expressions dans d'autres tissus tels que les glandes salivaires. Le niveau d'expression des transgènes basés sur le promoteur du gène de l' α -lactalbumine n'était pas corrélé au nombre de copies intégrées. L'ensemble de ces points sont caractéristiques de transgènes subissant l'effet de position.

III. Orientation vers l'utilisation d'un BAC caprin

L'émergence récente des grands fragments d'ADN génomique (BAC ou YAC) en transgénèse a ouvert de nouvelles perspectives pour l'emploi de ces promoteurs. Ainsi, l'utilisation d'un YAC humain de 210 kb renfermant le gène de l' α -lactalbumine a permis l'expression indépendante du site d'intégration de ce dernier dans des rats transgéniques (Fujiwara^a et b *et al.*, 1999). Un BAC caprin de 160 kb (BAC 41), contenant également le gène de l' α -lactalbumine qui s'étend sur environ 2 kb, a été isolé au laboratoire et utilisé pour la création de souris transgéniques. Ce BAC dispose de 150 kb de région 5' flanquant du gène et 10 kb seulement de région 3' flanquant ce qui permet de s'affranchir de l'ensemble des pseudogènes présents plus en aval (**Fig. I.11**).

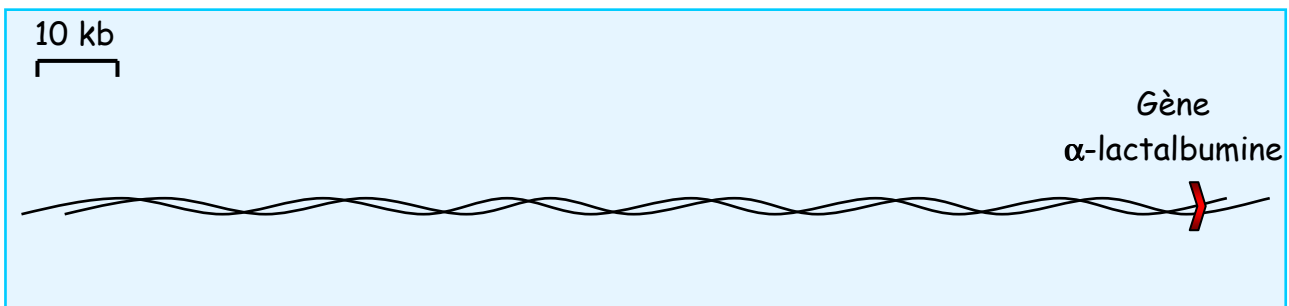


Fig. I.11 : Schéma du BAC 41.

L'absence de ces pseudogènes ne peut que faciliter l'interprétation des résultats qui suivront ces expériences, en évitant toute interférence avec des transcrits potentiellement mutés ou aberrants. L'utilisation de ce BAC 41 en transgénèse a permis d'obtenir des souris transgéniques dans lesquelles la régulation de l'expression du gène de l' α -lactalbumine caprin s'est avérée indépendante du site d'intégration et dépendante du nombre de copies intégrées (Stinnakre *et al.*, 1999). Ces résultats mettent ainsi en évidence la présence de séquences *cis*-régulatrices dominantes au sein de ce fragment d'ADN génomique.

Objectifs de la thèse

L'ensemble des expériences menées au cours de ce travail de thèse a eu pour but de caractériser finement le BAC caprin contenant le gène de l' α -lactalbumine (BAC 41), pour essayer par la suite d'identifier les régions régulatrices impliquées dans l'expression indépendante du site d'intégration de ce gène et proportionnelle au nombre de copies intégrées. Le clonage de telles séquences, permettrait la mise au point de vecteurs pour la transgénèse mammaire qui s'affranchiraient de l'effet de position.

I. Utilisation d'un BAC réduit en transgénèse

Pour réduire la région de recherche de ces éléments régulateurs, d'autres expériences de transgénèse chez la souris ont été menées en utilisant un BAC réduit (BAC 6) par rapport au BAC initial (BAC 41) (**Fig. I.12**).

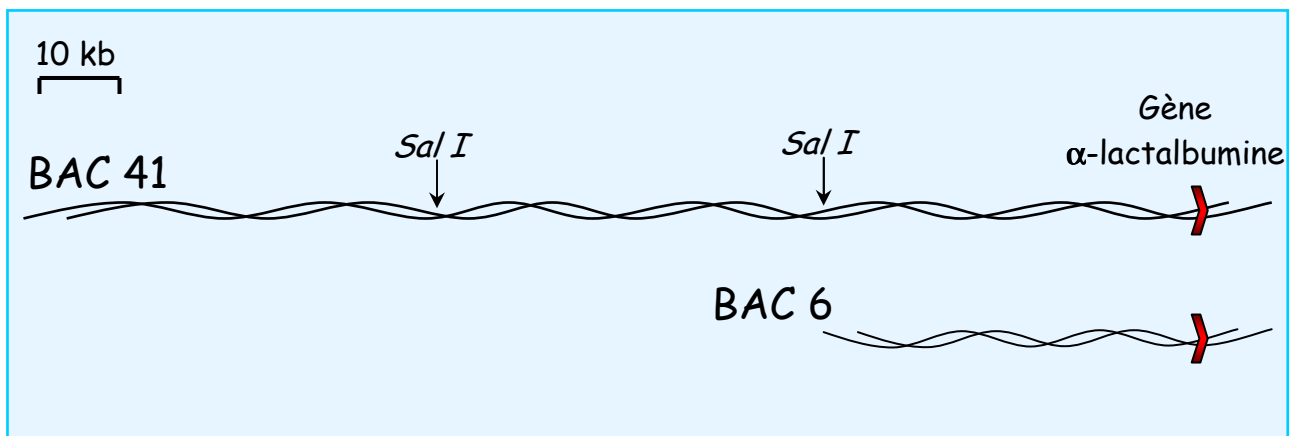


Fig. I.12 : Schéma du BAC 41 et du BAC 6.

Ce BAC 6 est issu de la digestion enzymatique par l'endonucléase de restriction *Sal I*, il contient 50 kb de région 5' flanquant et les 10 kb de région 3' flanquant ce qui permet d'envisager une première analyse voire même une localisation grossière des séquences importantes pour la régulation du gène de l' α -lactalbumine. J'ai participé à l'étude des souris transgéniques, obtenues à partir du BAC 6, qui avait débutée lors de mon arrivée.

II. Mise en évidence des différents *loci* présents au sein du BAC 41

Nous avons ensuite effectué une recherche au sein du BAC 41 d'autres gènes et approfondi l'analyse de leur expression, en s'appuyant notamment sur des éléments de cartographie comparée à la suite du séquençage des génomes de l'homme puis de la souris. En effet, la présence de gènes à expression très différente de celle de l' α -lactalbumine au sein du BAC pouvait, selon certains modèles avancés à l'époque (Prioleau *et al.*, 1999), suggérer l'existence de séquences isolatrices entre eux.

III. Analyse plus fine d'une région du BAC 41

Suite à ces expériences, une région d'environ 10 kb renfermant potentiellement un isolateur a pu être identifiée et je me suis attaché à sa caractérisation en fin de thèse. Différentes approches ont été utilisées telles qu'une analyse comparative de la séquence avec les bases de données existantes, un sous clonage de cette région pour des expériences de transfections en culture cellulaire et la mise au point de constructions pour analyse par transgénèse chez la souris.

La description des travaux réalisés au cours de cette thèse se fera sous forme d'articles pour les parties « **Utilisation d'un BAC réduit en transgénèse** » et « **Mise en évidence des différents *loci* présents au sein du BAC 41** », tandis que la dernière partie « **Analyse plus fine d'une région du BAC 41** » correspondra à des données encore préliminaires et non publiées.

CHAPITRE I : Primo-localisation de régions
régulatrices par analyse de l'expression en
transgénèse d'un BAC raccourci.

Une approche classique pour rechercher des régions renfermant des séquences *cis*-régulatrices agissant sur un gène consiste à étudier l'expression, soit en culture cellulaire, soit en transgénèse, de constructions présentant des délétions (Pierre *et al.*, 1992 ; Soulier *et al.*, 1992 ; Fujiwara^b *et al.*, 1999). Ces constructions sont souvent obtenues par l'établissement d'une carte de restriction de la zone concernée puis son utilisation pour l'isolement des fragments d'intérêt. Une telle stratégie est habituellement facile à mettre en œuvre pour l'étude de gènes de relativement petite taille mais beaucoup plus difficile pour de grands fragments d'ADN de part la complexité souvent observée de telles cartes. Dans le cas de longues régions d'ADN, une approche alternative consiste à introduire spécifiquement, par recombinaison homologe, des sites de restriction. Cela a notamment été utilisé dans l'étude de YACs, les techniques de recombinaison étant très développées chez la levure (pour revue Copeland *et al.*, 2001). Ainsi, l'équipe du Prof. Ueda a pu délimiter les régions renfermant des séquences *cis*-régulatrices au *locus* du gène α -lactalbumine humain (Fujiwara^{a et b} *et al.*, 1999), résultats que nous discuterons à la fin de ce chapitre. Bien que des méthodes de recombinaison homologe aient été développées chez *E. coli* afin de modifier des BACs (Yang *et al.*, 1997 ; Narayanan *et al.*, 1999 ; Muyrers *et al.*, 1999 ; Lee *et al.*, 2001 ; Valenzuela *et al.*, 2003 et *cf.* Introduction, Utilisation de grands fragments), leur mise en pratique reste plus délicate comme nous l'évoquerons par la suite.

Le BAC 41 présentait la particularité de ne posséder que peu de sites *Sal* I. Une première analyse avait laissé supposer qu'il n'en possédait qu'un localisé au centre de l'insert. Son étude plus détaillée a révélé en fait la présence de deux sites, libérant trois fragments de même taille (environ 60 kb). Malheureusement, l'identification de ce second site *Sal* I ne s'est produite qu'après la parution de l'article ci-dessous. Il convient donc dans cet article de modifier la longueur de la région 5' flanquant du BAC 6 qui n'est que de 50 kb. De même, la figure 4 devrait être comme suit : **(Fig. 1.1)**

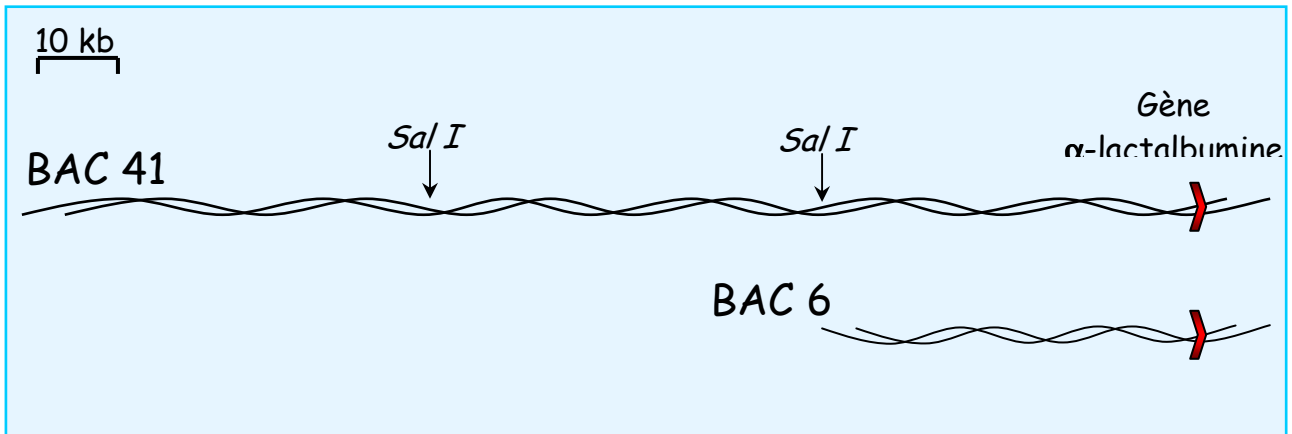


Fig. 1.1 : Schéma du BAC 41 et du BAC 6.

La carte de restriction du BAC 41 nous offrait donc l'opportunité d'étudier la régulation de l'expression du gène de l' α -lactalbumine lors de l'utilisation d'un BAC raccourci en transgénèse, sans avoir recourt à la recombinaison homologe. Notre but étant de mettre en évidence des séquences de type isolateur et travaillant sur un organe, la glande mammaire, pour lequel les modèles cellulaires restent peu nombreux (COMMA-1D : Danielson *et al.*, 1984, HC11 : Ball *et al.*, 1988, MCF-7 : Vickers *et al.*, 1988, BME-UV1 : Zavizion *et al.*, 1995, NISH : Ilan *et al.*, 1998) et souvent difficiles à utiliser comme les cultures d'explants, nous nous sommes tournés vers une analyse en transgénèse. Cette approche, bien que lourde et coûteuse, demeure la seule véritablement à même d'apporter une réponse quant à la présence ou non d'un isolateur au sein d'une construction.

Le premier article présenté au cours de cette thèse « *Distal(s) element(s) is(are) required for position-independent expression of the goat α -lactalbumin gene in transgenic mice. Potential relationship with the cyclin T1 locus* », rapporte les résultats de l'analyse de l'expression du gène de l' α -lactalbumine dans des souris transgéniques pour le BAC 6 et la mise en évidence de la présence d'un second gène au sein du BAC 41, celui de la cycline T1 (*cf.* fichier pdf ci-joint, **Article 1**).

DISCUSSION DU CHAPITRE I

Les résultats de ces expériences de transgénèse nous apprennent que les 50 kb de séquences caprine en amont du gène de l' α -lactalbumine ne sont pas suffisants pour obtenir une expression indépendante du site d'intégration et proportionnelle au nombre de copies intégrées. En 1999, Fujiwara *et al.* avait montré, par l'utilisation de YACs de tailles différentes contenant le gène humain de l' α -lactalbumine, que 50 kb de séquences en amont et en aval étaient suffisants pour éliminer l'effet de position en transgénèse. Nos résultats chez la chèvre semblent donc différents de ceux de l'équipe du Dr Fujiwara puisque nos séquences régulatrices, si elles existent, seraient localisées au-delà de la région humaine mise en évidence. Alternativement, on ne peut exclure soit la présence dans le génome humain de séquences *cis*-régulatrices dominantes situées à plus de 10 kb en aval du gène de l' α -lactalbumine, soit qu'au cours de l'évolution des phénomènes de délétions/insertions aient rapproché ces séquences de l'extrémité 5' du gène humain.

Dans un article récemment publié (Fujiwara *et al.*, 2003), cette même équipe a approfondi l'analyse de cette région pour localiser précisément les séquences empêchant l'effet de position en transgénèse. Les résultats de cette analyse, basée sur la comparaison de l'expression dans des rats de YACs humains plus ou moins grands, montrent qu'une région 5' de 2,5 kb (-3.5 kb à -0.8 kb), située en amont du gène humain de l' α -lactalbumine, entre deux sites hypersensibles à la DNase I, agirait comme une LCR. Les séquences potentiellement régulatrices chez la chèvre ne peuvent être localisées aussi près en amont car le BAC raccourci, disposant de 50 kb de séquences 5' flanquant, subit l'effet de position en transgénèse. Il semblerait donc que la localisation des éléments régulateurs chez l'homme soit différente de chez la chèvre. D'après le même article, un second gène a été identifié quelques kb en aval du gène de l' α -lactalbumine, en orientation inverse, gène similaire à un gène de récepteur olfactif (Olfactive Receptor MOR122-1). Pourtant lors de vérifications récentes des bases de données concernant les gènes mis en évidence dans cette région, il apparaît que si ces deux gènes sont bien orientés en sens inverse, celui du récepteur olfactif se situe à 2 kb en amont de celui de l' α -lactalbumine (**Fig. 1.2**). Les remaniements multiples observés au sein

des YACs utilisés par l'équipe japonaise remettent vraisemblablement en doute la validité de leurs résultats.

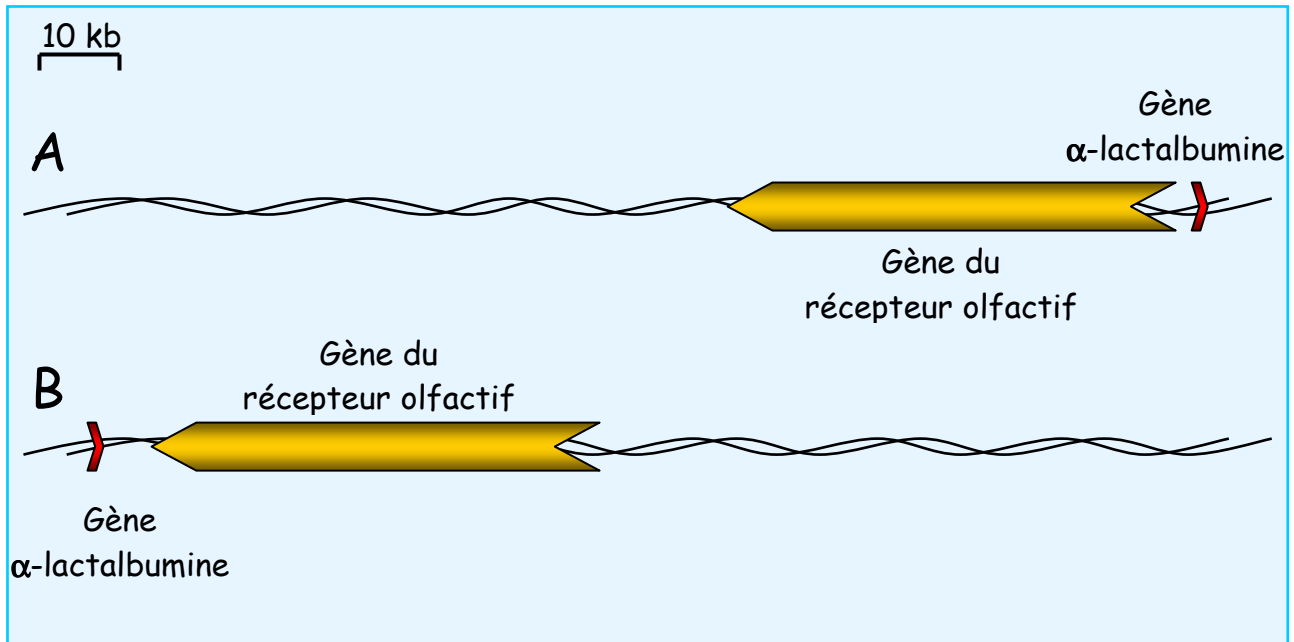


Fig. 1.2 : Schéma de la région chromosomique du locus α -lactalbumine chez l'homme.

A : région chromosomique telle qu'elle est décrite dans la base de donnée GenBank.

B : région chromosomique telle qu'elle est décrite par l'équipe japonaise du Dr. Fujiwara (Fujiwara et al., 2003).

Au final, ces expériences en transgénèse chez la souris, utilisant le BAC 6 caprin (BAC raccourci par rapport au BAC 41), nous ont permis d'effectuer une primo-localisation des séquences *cis*-régulatrices dominantes présentes au sein du BAC 41 par l'invalidation d'une région de 50 kb en amont du *locus* de l' α -lactalbumine. De plus, la mise en évidence d'un second gène, celui de la cycline T1 localisé en amont de cette région invalidée par les résultats en transgénèse, nous incitait à caractériser plus finement les différents éléments présents dans ce grand fragment d'ADN, pour essayer d'organiser une vue d'ensemble de la région chromosomique.

CHAPITRE II : Caractérisation du gène de la
cycline T1 présent au sein du BAC 41.
Hypothèse de l'existence de deux domaines
chromatiniens.

A la suite de ces expériences de transgénèse chez la souris utilisant le BAC 6, nous nous sommes attachés à vérifier si les séquences apparentées au gène de la cycline T1 nouvellement identifié au sein du BAC correspondaient au gène fonctionnel et si oui, quel était son mode d'expression. Des travaux récents suggèrent une organisation du génome en domaines regroupant des gènes présentant des modes d'expression comparables et séparés les uns des autres par des frontières physiques, renfermant notamment des séquences isolatrices (pour revue West *et al.*, 2002 ; Kuhn and Geyer, 2003). En s'appuyant sur ce modèle, devenu peut être un peu simpliste à la vue de données récentes, il nous a semblé important d'étudier le mode d'expression et la localisation des gènes présents sur le BAC 41. Dans la région où nous supposons se trouver les séquences isolatrices (*cf.* CHAPITRE I : primo-localisation de régions régulatrices par analyse de l'expression en transgénèse d'un BAC raccourci), le but était de définir de possibles domaines chromatiniens et ainsi de délimiter plus finement les zones à étudier.

Le gène de la cycline T1 est un gène récemment mis en évidence chez l'homme et la souris de part la fonction de la protéine qu'il code. Le produit de ce gène est impliqué dans la mise en place d'un complexe protéique, comprenant P-TEFb (Positive Transcription Elongation Factor b) et CDK9 (Cyclin-Dependent Kinase 9), nécessaire au fonctionnement de la machinerie ARN polymérase II. La protéine cycline T1 a une forte affinité pour la protéine Tat et la boucle de transactivation de la transcription (TAR) du VIH (Virus Immunodéficient Humain) (Wei *et al.*, 1998 ; Bieniasz *et al.*, 1998 ; Garber *et al.*, 1998 ; Peng *et al.*, 1998 et pour revue De Luca *et al.*, 2003). La compréhension de la régulation mais aussi de la fonction du gène de la cycline T1 est donc très importante pour empêcher cette interaction puis cette activation de la transcription virale. Ce gène a été localisé à proximité du *locus* de l' α -lactalbumine chez l'homme (chromosome 12q13.11 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=54934>) et la souris (chromosome 15F1 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=69612>) et présente dans ces deux espèces une expression ubiquiste. Néanmoins, lorsque nous avons entrepris cette étude :

- la structure du gène n'était que partiellement connue
- rien n'indiquait que le gène caprin, présent au sein du BAC 41, était fonctionnel et que son mode d'expression était identique à ses homologues humain et souris.

Les deux articles présentés ci-après « *Ubiquitous expression of goat cyclin T1 in transgenic mice* » et « *Sequence of goat cyclin T1 cDNA, gene organisation and expression analysis* » rapportent la caractérisation fonctionnelle et structurale du gène caprin de la cycline T1 (cf. fichiers pdf ci-joint, **Article 2 et 3**).

DISCUSSION DU CHAPITRE II

Dans l'article « *Ubiquitous expression of goat cyclin T1 in transgenic mice* » nous avons pu, en nous appuyant sur les souris transgéniques porteuses du BAC 41, montré que le gène caprin de la cycline T1 présent au sein de ce BAC était fonctionnel et disposait d'un mode d'expression ubiquiste, donc très différent de la régulation de l'expression du gène de l' α -lactalbumine. Sa localisation précise, son orientation et sa structure complète rapportée dans l'article « *Sequence of goat cyclin T1 cDNA, gene organisation and expression analysis* » nous ont permis de dresser une carte fine de cette région chromosomique. La comparaison avec les régions homologues humaine et murine montre une conservation de l'ordre et de l'orientation de ces gènes au cours de l'évolution.

Chez la chèvre, l'étude de l'expression de ce gène montre qu'il est faiblement exprimé dans l'ensemble des tissus analysés confirmant nos observations en transgénèse. Ces premiers résultats nous ont permis de définir différentes régions dans le BAC 41 avec une trentaine de kb de séquences régulatrices disponibles en amont du *locus* de la cycline T1, dont le gène s'étend lui-même sur plus de 40 kb (**Fig. 2.1**). Ils nous ont également conduit à suspecter l'existence de deux domaines chromatinien distincts au sein du BAC 41.

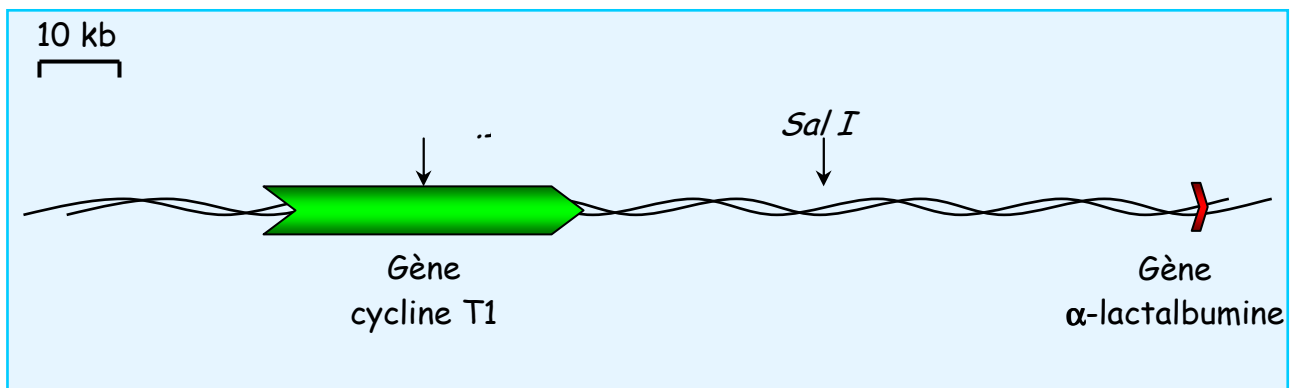


Fig. 2.1 : Schéma du BAC 41.

Avant cette caractérisation, une construction pour la transgénèse murine a été réalisée pour vérifier si le BAC 41 pouvait conférer une expression indépendante du site d'intégration à un gène hétérologue, substitué à l'unité de transcription du gène de l' α -lactalbumine. Le gène rapporteur choisi pour ce travail a été l'ADNc du gène CAT. Les résultats montrent une

extinction de l'expression du gène rapporteur dans les souris transgéniques obtenues malgré la présence d'éléments régulateurs mis en évidence précédemment au sein de ce BAC (Soulie *et al.*, 2003). Plusieurs exemples dans la littérature font état d'une extinction du gène CAT lors de son utilisation en transgénèse comme gène rapporteur (Clark *et al.*, 1997). Le phénomène aurait même la capacité de se propager le long de l'ADN pour éteindre les gènes voisins. Pour vérifier cette hypothèse, le niveau d'expression du gène de la cycline T1, toujours présent dans ce BAC modifié, a été étudié. Les résultats indiquent que l'extinction du gène CAT n'affecte pas celle du gène de la cycline T1. Cette observation renforce l'hypothèse de la présence d'éléments permettant des régulations transcriptionnelles indépendantes des gènes situés dans les deux putatifs domaines chromatiniens.

CHAPITRE III : Caractérisation de la région promotrice du gène de la cycline T1 caprine.

La caractérisation du gène de la cycline T1 caprine aussi bien fonctionnelle que structurale nous a permis de suspecter la présence de plusieurs domaines chromatiniens au sein du BAC 41, notamment en raison d'une régulation de l'expression très différente entre ce gène et celui de l' α -lactalbumine. Pour éviter aux éléments régulateurs de deux domaines chromatiniens voisins d'interagir les uns sur les autres, l'existence d'une barrière physique est nécessaire. Pour remplir ce rôle de protection, les séquences isolatrices doivent être localisées entre les régions promotrices des gènes concernés (**Fig. 3.1**).

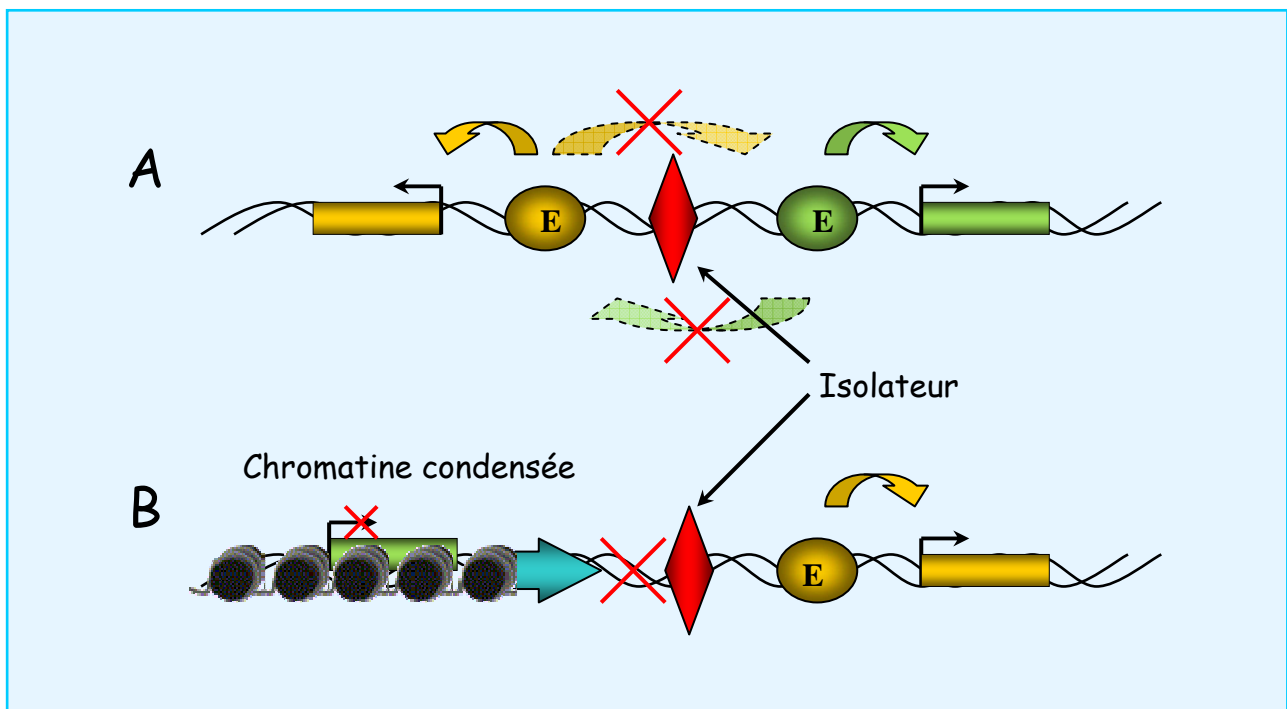


Fig. 3.1 : Illustrations des propriétés d'un isolateur (Kuhn and Geyer, 2003).

A : : blocage par l'isolateur de la stimulation du promoteur jaune par l'activateur vert et inversement du promoteur vert par l'activateur jaune.

B : l'isolateur peut empêcher la propagation de la chromatine condensée et ainsi éviter l'extinction des gènes.

Les losanges rouges représentent les isolateurs, les ovales de verts et jaunes notés E les activateurs de gènes, les croix rouges l'absence d'activation et la chromatine condensée est illustrée par la présence de nucléosomes.

Cependant, afin d'éliminer définitivement une région supplémentaire susceptible de contenir des séquences potentiellement isolatrices ou *cis*-dominantes, nous avons décidé de tester en transgénèse l'activité transcriptionnelle de la région promotrice du gène de la cycline T1 disponible au sein du BAC 41. A notre connaissance, aucune analyse *in vivo* n'avait été entreprise pour caractériser le promoteur de ce gène pourtant très étudié chez l'homme dans le cadre de son implication au niveau de l'activation de la transcription du VIH (Wei *et al.*,

1998 ; Bieniasz *et al.*, 1998 ; Garber *et al.*, 1998 ; Peng *et al.*, 1998 et pour revue De Luca *et al.*, 2003). Seules des analyses *in vitro* par transfection de lignées cellulaires ont été réalisées. Elles suggèreraient qu'un promoteur court (moins de 1 kb) est capable de diriger l'expression du gène de la cycline T1, aussi bien chez l'homme que chez la souris (Liu and Rice, 2000 ; Martin-Serrano *et al.*, 2002). En transgénèse, l'utilisation de promoteurs ubiquistes forts dirigeant l'expression du transgène dans la plupart des types cellulaires, si ce n'est tous, permettrait d'apporter quelques réponses à des problèmes biologiques comme par exemple, le gain de fonction, la thérapie génique ou le lignage cellulaire dans des animaux chimériques. Plusieurs promoteurs ubiquistes ont déjà été utilisés (Kisseberth *et al.*, 1999 ; Taboit-Dameron *et al.*, 1999 ; Zhan *et al.*, 2000) mais le problème majeur reste l'effet de position que subissent ces promoteurs qui rend les expériences difficiles et surtout beaucoup moins intéressantes du point de vue des applications potentielles.

Une des thématiques du laboratoire est l'étude des Encéphalopathies Spongiformes Subaigus Transmissibles (ESST), maladies neurodégénératives fatales chez l'homme et les animaux caractérisées par une longue phase d'incubation asymptomatique (plusieurs années) suivie d'une phase clinique subaigus. Les ESST humaines incluent la maladie de Creutzfeldt-Jacob (pour revue Knight, 2001), l'insomnie familiale fatale, le Kuru et le syndrome Gerstmann-Sträussler-Scheinker (pour revue Collins *et al.*, 2001). Chez les animaux, les ESST sont la scrapie chez le mouton et la chèvre, l'encéphalopathie spongiformes bovine (ESB), l'encéphalopathie spongiformes féline et l'encéphalopathie transmissible du vison (pour revue Dormont, 1999). Ces maladies sont causées par un agent infectieux non conventionnel appelé prion, semble-t-il dépourvu d'acide nucléique (pour revue Prusiner, 1998). Il s'agirait d'une forme anormale (PrPsc) d'une protéine cellulaire de l'hôte (PrPc). La PrPc est une protéine membranaire présente chez tous les mammifères, codée par le gène Prnp, gène à expression ubiquiste qui est fortement transcrit au niveau du système nerveux central. La fonction de la protéine PrPc reste largement inconnue. Les deux protéines PrPc et PrPsc ont une séquence en acide aminés identique et seule leur structure en trois dimensions diffère avec une prépondérance d'hélices α pour la forme cellulaire (PrPc), contrairement à la forme anormale (PrPsc) dont la structure prédite semble principalement composée de feuilletts β (**Fig. 3.2**). La conversion de la PrPc en PrPsc nécessite la présence des deux types de protéines et vraisemblablement l'intervention de protéines chaperonnes (pour revue Prusiner, 1998 ; Wadsworth *et al.*, 1999 ; Weissmann, 1999).

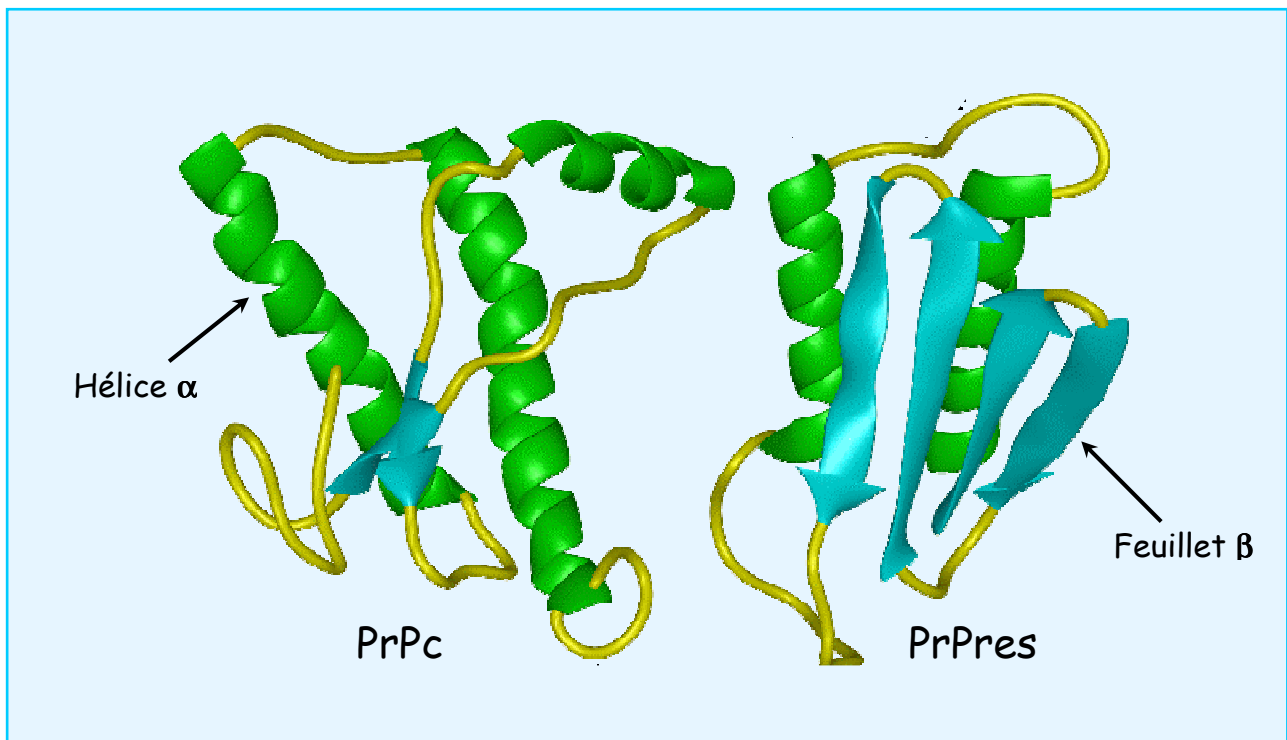


Fig. 3.2 : Représentation schématique des structures 3D des protéines PrPc et PrPres (Prusiner, 1998).

La présence de PrPc est indispensable à la propagation de la maladie car des souris dont le gène Prnp a été inactivé sont non susceptibles. Une sous fraction des protéines PrPsc (appelée PrPres) est partiellement résistante aux protéases induisant une accumulation dans le système nerveux central. Cette accumulation de PrPres ne s'accompagne pas d'une augmentation du niveau d'expression du gène Prnp, traduisant un mécanisme post-traductionnel (pour revue Prusiner, 1998). Pour étudier la propagation de la maladie, des modèles animaux comme la souris ont été utilisés. Seulement lors de l'injection intracérébrale de PrPres ovine, de hamster ou bovine à des souris conventionnelles, le délai d'apparition des symptômes est très longs (de l'ordre de 800 jours), voire dépasse la durée de vie de l'animal. Ce temps d'incubation est largement réduit lors de passages secondaires. Ce phénomène, appelé barrière d'espèce, traduit l'adaptation de l'agent infectieux à son nouvel hôte et ses difficultés sans doute à interagir avec les protéines endogènes avant de pouvoir se dupliquer. Parmi ces protéines, la PrPc joue un rôle essentiel. Ainsi, la synthèse d'une protéine PrPc issue de la même espèce que la source de prions dans des souris transgéniques au gène Prnp endogène invalidé s'est souvent traduit par l'obtention d'animaux présentant une plus forte sensibilité à l'agent infectieux que des souris conventionnelles. On observe alors un temps d'incubation

relativement court dès la primo-inoculation, sans adaptation lors des passages successifs (pour revue Vilotte and Laude, 2002).

Pour essayer de mettre au point un modèle animal plus sensible à l'agent infectieux de la BSE que ceux déjà présents au laboratoire, le gène rapporteur choisi en association avec le promoteur du gène de la cycline T1 a été le gène Prnp bovin. Nous espérons obtenir ainsi une expression de la protéine PrP bovine dans l'ensemble des tissus des souris transgéniques à des niveaux plus ou moins importants.

Afin de tester différentes hypothèses deux constructions ont été réalisées :

- une construction disposant d'un promoteur court, TgA, (environ 1 kb) pour vérifier les résultats disponibles chez l'homme qui suspectait que l'ensemble des séquences régulatrices régissant l'activité de ce gène se trouvent à moins de 1 kb du site d'initiation de la transcription (Liu and Rice, 2000 ; Martin-Serrano *et al.*, 2002),
- une seconde construction, obtenue par recombinaison homologue chez *E. coli*, utilisant les 30 kb de régions régulatrices située dans la partie 5' du BAC 41, TgB. Cette seconde construction visait à rechercher l'effet protecteur éventuel de séquences distales sur l'activité promotrice du gène de la cycline T1 (**Fig. 3.3**).

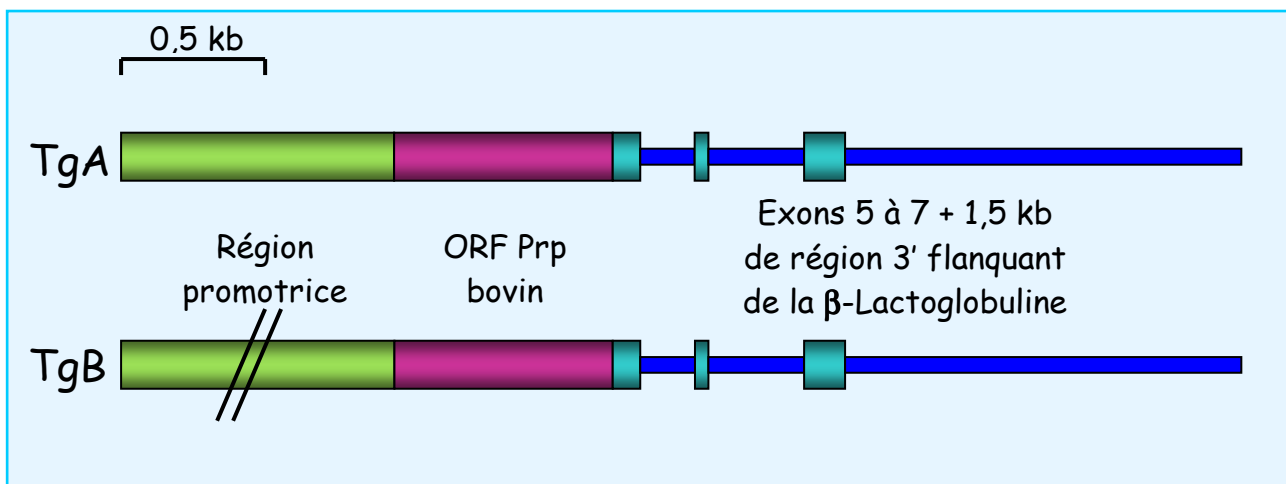


Fig. 3.3 : Schéma des constructions TgA et TgB.

Les objectifs de ce travail étaient de fait multiples :

- tester la présence ou l'absence de séquences isolatrices dans cette région promotrice de 30 kb en analysant l'effet de l'environnement chromosomique sur la régulation de l'expression du gène PrP bovin *in vivo*.

- vérifier si un promoteur court du gène de la cycline T1 caprine était capable de diriger l'expression d'un gène rapporteur *in vivo* comme cela a été suspecté *in vitro* chez l'homme, validant l'utilisation potentielle de cette séquence en thérapie génique.
- mettre au point un modèle animal plus sensible à la maladie de la « vache folle » que ceux déjà présents au laboratoire.

L'article présenté ci-après « *Unexpected high testis-specific transcriptional activity of the cyclin T1 promoter in transgenic mice* » illustre les résultats que nous avons obtenus au cours de ce travail (cf. fichier pdf ci-joint, **Article 4**).

DISCUSSION DU CHAPITRE III

La caractérisation de la région promotrice du gène de la cycline T1 caprine apporte plusieurs réponses aux questions posées au début de ce travail. L'analyse du profil d'expression du gène Prnp bovin, le gène rapporteur, montre une activité transcriptionnelle identique quelque soit la longueur du promoteur utilisée. La seule différence concerne le niveau d'expression qui est plus faible dans les souris transgéniques TgA (promoteur court). Ces résultats confirment partiellement les hypothèses émises à la suite des expériences *in vitro* concernant l'analyse de l'expression d'un promoteur court humain du gène de la cycline T1, à savoir qu'un tel promoteur est suffisant pour diriger l'expression de ce gène. Toutefois, une sur-expression testiculaire inattendue, car non observée avec le gène natif au sein du BAC 41, a pu être mise en évidence dans l'ensemble des lignées de souris transgéniques analysées, aussi bien les TgA que les TgB. Ces résultats suggèrent la perte d'éléments *cis*-régulateurs importants pour la régulation transcriptionnelle du gène de la cycline T1 caprine. D'après les constructions utilisées, ces éléments pourraient être localisés soit au niveau des régions non traduites (5' et 3' UTR) souvent impliquées au niveau de la régulation de l'expression des gènes (pour revue Kuersten and Goodwin, 2003) soit au niveau des séquences introniques et/ou 3' flanquantes du gène de la cycline T1. La nature des éléments éliminés ne peut être que supposée. Il pourrait s'agir de séquences inhibitrices agissant sur des activateurs impliqués dans l'initiation de la transcription au niveau des cellules germinales, ou bien de séquences activatrices permettant de recruter des facteurs de transcription nécessaires à un bon niveau d'expression dans l'ensemble des tissus. Quoiqu'il en soit, même si un promoteur cycline T1 court est suffisant pour diriger l'expression d'un gène rapporteur *in vivo*, l'ensemble des séquences associées à une régulation transcriptionnelle normale de ce gène ne sont pas toutes présentes dans la région promotrice.

L'ensemble des lignées de souris transgéniques TgB analysées présente une sur-expression du transgène plus ou moins importante dans les cellules germinales, cependant la régulation transcriptionnelle de ce gène est sous la dépendance de son site d'intégration car il n'existe pas de proportionnalité entre le niveau d'expression et le nombre de copies intégrées. Aucune

séquence capable de protéger le gène rapporteur ne semble donc être présente au sein des 30 kb de région promotrice. Cette observation a deux conséquences :

- elle renforce l'hypothèse de la présence de deux domaines chromatiniens distincts au sein du BAC.
- elle suggère que des séquences *cis*-régulatrices dominantes sont localisées en aval de la région promotrice du gène de la cycline T1.

Une caractérisation plus fine de la régulation de l'expression du gène *Prnp* sous dépendance du promoteur cycline T1 nous a permis de montrer que les transcrits n'étaient pas exprimés dans tous les types cellulaires du testicule mais seulement dans la lignée germinale, des spermatogonies aux spermatozoïdes. La mise au point d'un modèle souris hautement sensibles aux agents de la maladie de la « vache folle » reposait sur la présence de la protéine prion bovine dans de nombreux tissus des souris transgéniques à des niveaux relativement élevés et notamment au niveau du système nerveux central. A ce titre, les résultats obtenus n'ont pas répondu à notre attente. Ces souris ne présentent sans doute pas une sensibilité accrue à l'agent infectieux bovin en comparaison aux autres modèles déjà présents au laboratoire. Après concertation avec l'équipe du Dr H. Laude (VIM, INRA), il a d'ailleurs été décidé de ne pas analyser leur susceptibilité.

L'analyse de la protéine PrP produite dans les testicules des souris transgéniques montrent la présence d'une protéine partiellement tronquée par rapport à la protéine dite normale trouvée dans le cerveau de vache ou de souris transgéniques sur-exprimant le gène *Prnp* bovin. Afin de vérifier si cette différence de taille n'était pas due à une glycosylation différentielle, une N-déglycosylation des protéines totales a été réalisée sur la nuit selon les prescriptions du fournisseur (PNGase F, Biolabs) suivi d'un Western utilisant l'anticorps 18 (gentiment fourni par Dr Hawke S., Imperial college, London, UK). Cet anticorps est dirigé contre la région centrale de la protéine PrP bovine/ovine (**Fig. 3.4 et 3.5**).

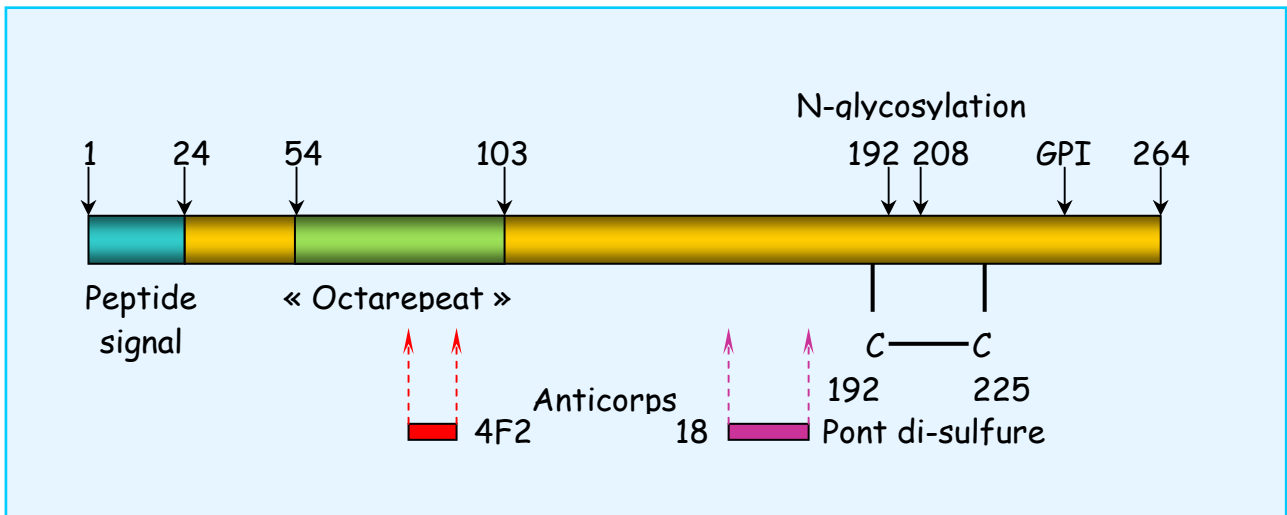


Fig. 3.4 : Schéma de la protéine prion bovine et localisation des anticorps utilisés.

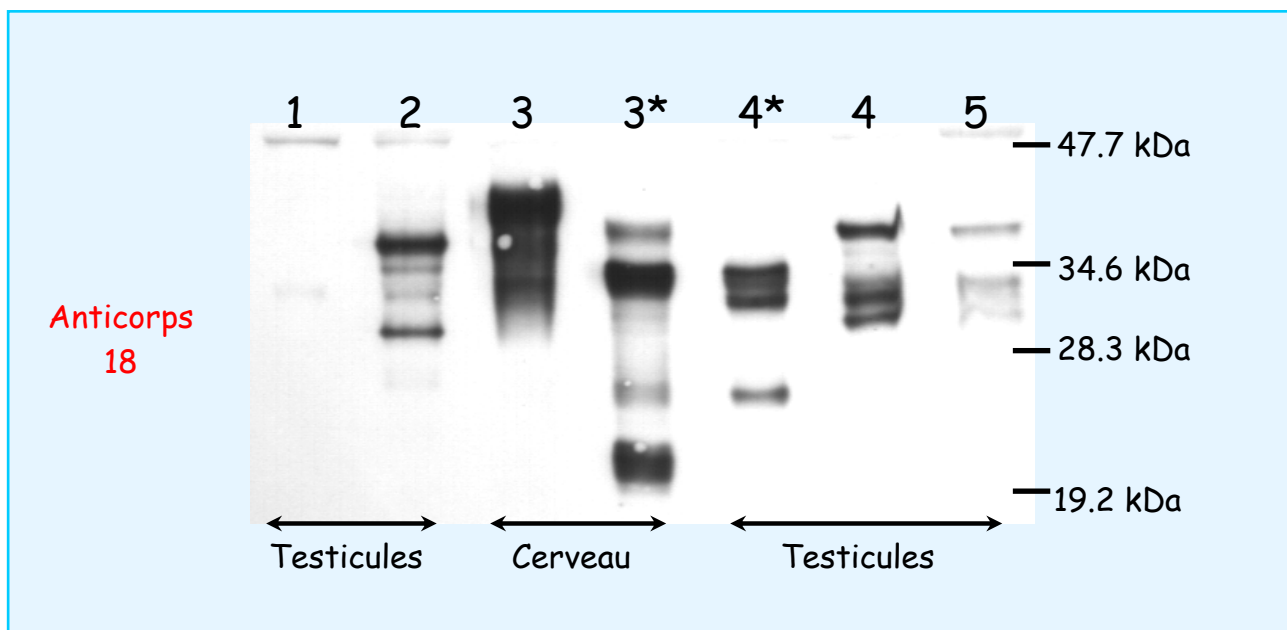


Fig. 3.5 : Analyse de la synthèse de protéine PrP par western blot .

Plusieurs lignées de souris transgéniques ont été utilisées pour cette analyse : 1-souris FVBN PrP0/0 dont le gène Prnp a été inactivé, 2-souris Tg3.38, qui sur-exprime la protéine PrP ovine, 3-souris Tg5.40 qui sur-exprime la protéine PrP bovine, 4-souris TgB.75, 5-souris TgA.12, * indique que les échantillons ont été traités à la PNGase F pour déglycosyler les protéines, cinquante µg de protéines totales ont été déposées dans chaque puits.

Pour caractériser plus finement cette troncation, l'anticorps 4F2 (gentiment fourni par DrGroschup, Institute for Novel and Emerging Infectious Diseases, Federal Research Centre for Virus Diseases, Insel Riems, Germany) dont l'épitope est localisé dans la portion N-terminale de la protéine, région 79-92 aa de la protéine bovine (Fig. 3.4), a été utilisé (Fig. 3.6).

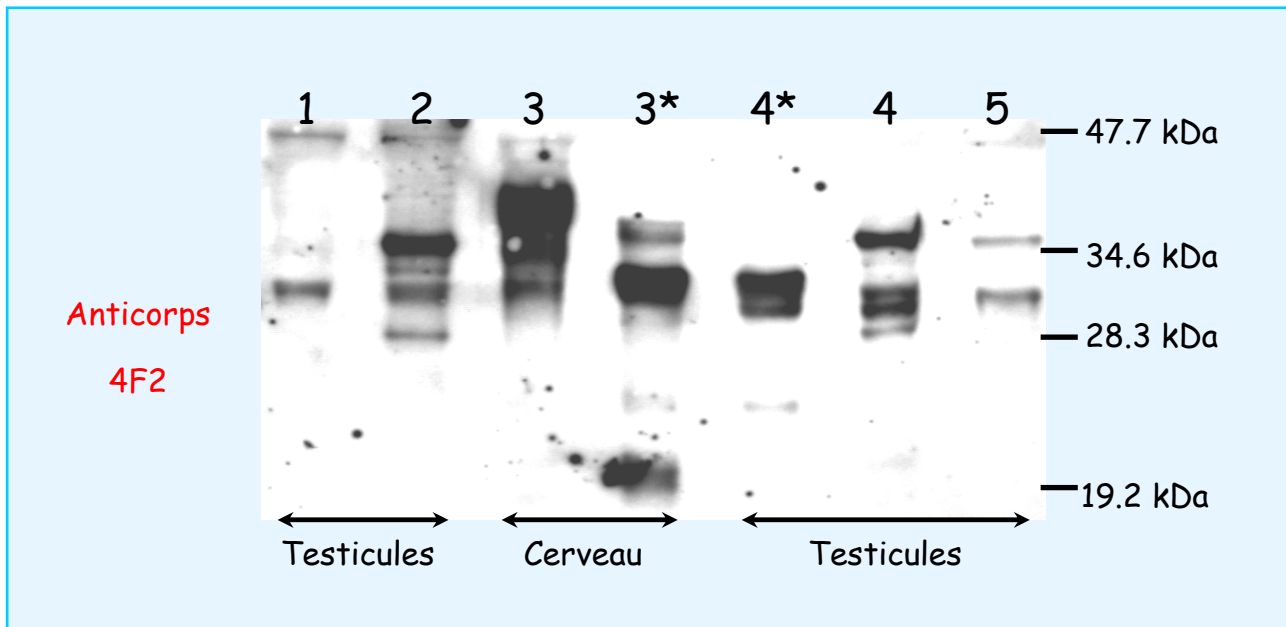


Fig. 3.6 : Analyse de la synthèse de protéine PrP par western blot, (cf. légende Fig. 3.5).

La présence d'un signal en western blot semble indiquer que la région N-terminale n'est pas tronquée, suggérant donc une troncation C-terminale. Une protéine tronquée en C-terminal (délétion de 70 à 100 aa) a été décrite dans le sperme mature et le liquide séminal du mouton (Gatti *et al.*, 2002). Chez l'homme la situation est plus complexe puisque plusieurs isoformes ont pu être mises en évidence dans le liquide séminal et au niveau du sperme mature avec deux isoformes tronquées en N-terminal : C1 (aa 1 à 112) et C2 (environ aa 1 à 90) (Peoc'h *et al.*, 2002) et une isoforme tronquée C-terminal (délétion d'une centaine d'aa) (Shaked *et al.*, 1999). Par contre, les analyses effectuées chez la souris ne permettent pas de mettre en évidence de forme tronquée de la protéine au niveau des testicules (Ford *et al.*, 2002). Même si l'utilisation d'autres anticorps serait nécessaire pour caractériser les différentes formes de PrPc présentes dans les cellules germinales des souris TgA et TgB et les testicules des souris Tg5.40 et Tg3.38 (Fig. 3.5), nos résultats suggèrent que l'absence de protéine PrP tronquée dans le testicule de souris conventionnelles tient plus à la nature de la PrP qu'à l'absence de

certaines protéases dans le tractus génital mâle de cet animal. Il est à noter qu'une forme tronquée de la PrPc est présente chez la souris dans le sperme mature mais pas dans l'épididyme.

La protéine prion bovine synthétisée dans les souris transgéniques TgA et TgB est localisée dans toute la lignée germinale, des spermatogonies au spermatozoïdes, mais pas dans les cellules de soutien telles que les cellules de Leydig et Sertoli. Chez l'homme, la protéine prion est présente dans les spermatocytes et spermatides et absente des spermatogonies et des Sertoli (Shaked *et al.*, 1999). Chez la souris, elle est abondante dans les cellules de Sertoli (Ford *et al.*, 2002) et les spermatozoïdes suivant un gradient inverse dans les cellules germinales de celui des ARNm qui sont présents dans les spermatocytes, les spermatides et les cellules de Sertoli. Bien que nous n'ayons aucune donnée sur les taux relatifs des transcrits TgA et TgB dans les spermatogonies par rapport aux spermatocytes et aux spermatozoïdes, nos observations suggèreraient que le contrôle traductionnel des transcrits Prnp observés chez la souris impliquerait des séquences localisées dans les régions 5' et 3' non traduites de l'ARNm Prnp, régions absentes dans les constructions utilisées pour créer les souris TgA et TgB. Toutefois, des expériences complémentaires seraient nécessaires pour étayer cette hypothèse.

L'analyse de l'influence de la sur-expression de la protéine prion dans les testicules sur la fertilité a été entreprise en raison du phénotype observé chez les souris dont le gène Dpl (Downstream Prion protein-Like ou Prnd) a été inactivé. Ces souris, dont le développement est normal, présentent une stérilité des mâles due à une déficience au niveau de la spermatogénèse (Behrens *et al.*, 2002). Le gène Dpl est un gène paralogue au gène Prnp, localisé à 16 kb en aval de celui-ci chez la souris. La protéine Dpl est une protéine N-glycosylée et attachée à la membrane cellulaire par une ancre GPI (Glycosyl Phosphatidyl Inositol) selon un schéma identique à celui de la protéine PrP. Par sa séquence et surtout sa structure, la protéine Dpl ressemble à une protéine PrP tronquée en N-terminal (Silverman *et al.*, 2000).

Les profils d'expression de ces deux gènes sont différents puisque le gène Dpl est fortement exprimé dans les testicules et le cœur et faiblement exprimé dans le cerveau tandis que le gène Prnp est fortement exprimé dans le cerveau et plus faiblement dans le cœur et les testicules. L'expression ectopique du gène Dpl dans le cerveau de souris, dont le gène Prnp a été

inactivé, conduit à des phénotypes sévères dont une ataxie et perte des cellules de Purkinje. Toutefois, ces phénotypes disparaissent lors du rétablissement du profil d'expression normal du gène *Prnp*, suggérant une certaine compétition entre ces deux protéines (Moore *et al.*, 1999 ; Rossi *et al.*, 2001). Aucune incidence sur la fertilité n'a pu être mise en évidence dans les lignées TgB sur-exprimant le plus le transgène indiquant que de hauts niveaux de protéine PrP dans les spermatozoïdes et au cours de la spermatogénèse n'interfèrent pas avec la reproduction de ces souris. Pourtant, il serait potentiellement intéressant de tester cette sur-expression du gène *Prnp* bovin testiculaire dans des souris dont le gène *Dpl* a été inactivé pour savoir s'il provoque alors une potentielle interférence avec la spermatogénèse. Des contacts ont été pris avec des équipes travaillant sur le gène *Dpl* pour discuter de l'intérêt d'une telle approche.

CHAPITRE IV : Caractérisation du gène caprin FLJ20436, troisième gène présent au sein du BAC 41. Assignation de ce locus à un des deux domaines chromatinien putatifs.

En parallèle à la caractérisation du gène de la cycline T1 caprine et de son promoteur en transgénèse, une analyse de la séquence humaine disponible dans les bases de données nous a permis de localiser un gène putatif à peine 10 kb en aval du gène de la cycline T1 (chromosome 12q13.11 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=54934>). Après une analyse comparative avec les banques d'EST, ce gène putatif FLJ20436 semblait disposer de plusieurs transcrits (*cf.* **Article 5**) révélant que ce gène pouvait être fonctionnel chez l'homme. Une recherche d'homologie au sein de cette région sur le génome de la souris nous a permis de confirmer la présence de ce gène putatif dans cet organisme (chromosome 15F1 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=69612>), de même que l'existence de plusieurs transcrits murins disponible dans la banque d'EST (*cf.* **Article 5**). Les différents transcrits ainsi mis en évidence proviennent de différents tissus suggérant la présence d'un gène dont le mode d'expression est potentiellement ubiquiste. L'ensemble de ces éléments conduirait à l'existence d'un gène fonctionnel appartenant au même domaine chromatinien que celui du gène de la cycline T1 puisque ces gènes semblent disposer du même type de régulation transcriptionnel.

Une analyse plus fine de cette région a été entreprise chez la chèvre afin de savoir si :

- le gène putatif FLJ20436 est présent au sein du BAC 6 et/ou du BAC 41
- dans la mesure où ce gène est présent, dans quel contexte génomique il se présente par rapport à ce qui est connu chez l'homme et la souris (orientation des gènes les uns par rapport aux autres et localisation précise)
- ce gène est fonctionnel dans cet organisme et au sein des souris transgéniques pour le BAC 41
- son mode d'expression correspond bien aux hypothèses émises à la suite de la recherche des ESTs, c'est-à-dire un mode d'expression ubiquiste comme c'est le cas pour le gène de la cycline T1.

L'article présenté ci-après « *Putative FLJ20436 gene characterisation in goat. Observed ubiquitous expression in goat and transgenic mice allowed to restrict the location of an hypothesised insulator element.* » rapporte la caractérisation fonctionnelle et structurelle du gène caprin FLJ20436 (*cf.* fichier pdf ci-joint, **Article 5**).

DISCUSSION DU CHAPITRE IV

La caractérisation de ce troisième *locus* nous a permis de localiser précisément le gène FLJ20436 en aval du gène de la cycline T1 et en amont du BAC 6 et de l'orienter par rapport aux autres gènes (**Fig. 4.1**).

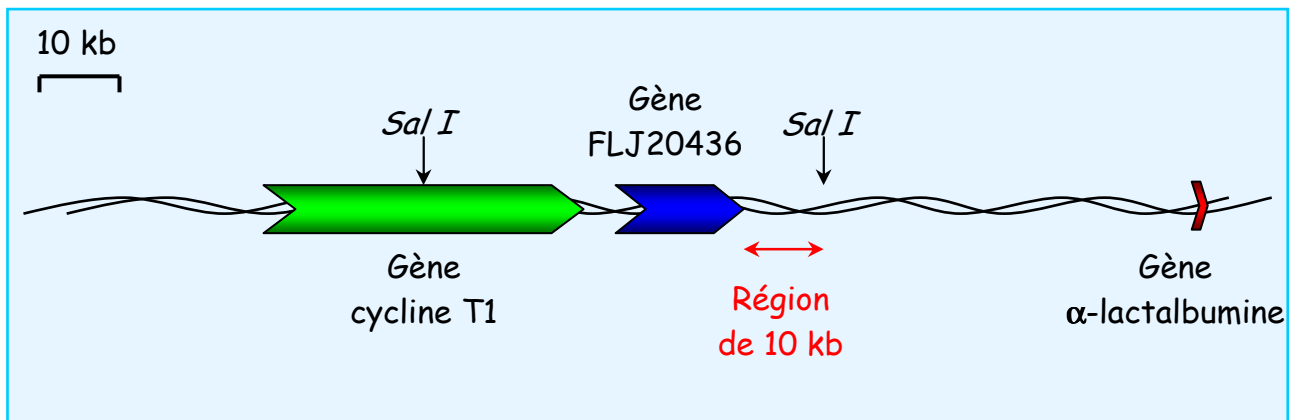


Fig. 4.1 : Schéma du BAC 41.

Une carte plus fine de la région a ainsi pu être dressée et la comparaison avec les régions homologues humaine et murine présente toujours cette conservation de l'ordre et de l'orientation entre ces différents gènes au cours de l'évolution. Nous avons pu montrer que le gène FLJ20436 est fonctionnel et de type ubiquiste chez la chèvre et en nous appuyant sur les souris transgéniques porteuses du BAC 41, nous avons pu confirmer qu'il s'agissait bien du gène présent au sein du BAC 41 qui était fonctionnel.

Ces expériences nous ont permis d'assigner le gène FLJ20436 au même domaine chromatinien que celui du gène de la cycline T1 conformément aux hypothèses émises précédemment et ainsi nous avons pu affiner la localisation de la région potentielle contenant les éléments *cis*-régulateurs dominants (**Fig. 4.1**).

L'analyse approfondie de la régulation de l'expression du gène FLJ20436 montre un gène ubiquiste tout comme le gène de la cycline T1 avec de nombreux transcrits issus de différents épissages alternatifs dont certaines formes sont encore inconnue chez l'homme ou la souris.

CHAPITRE V : Caractérisation de la région de
10 kb, frontière putative entre deux domaines
chromatiniens.

Expériences menées pour caractériser la région frontière de 10 kb

I. Récapitulatif des travaux qui ont permis la localisation de la région frontière

Les travaux réalisés sur le BAC 41 ont permis une caractérisation de différents éléments présents au sein de ce grand fragment d'ADN. Les expériences de transgénèse utilisant le BAC 41 puis le BAC 6 ont permis de mettre en évidence la présence de séquences *cis*-régulatrices dominantes nécessaires à la régulation de l'expression du gène de l' α -lactalbumine, indépendamment du site d'intégration.

La mise en évidence de la présence de deux autres gènes au sein du BAC 41, la cycline T1 et FLJ20436, nous a conduit à vérifier si ces gènes étaient fonctionnels et de quel mode d'expression ils disposaient. La caractérisation fonctionnelle de ces deux gènes, présents en amont du *locus* de l' α -lactalbumine et du BAC 6, a permis de montrer qu'ils étaient fonctionnels et exprimés de façon ubiquiste, donc très différemment de celui du gène de l' α -lactalbumine spécifique du tissu épithéliale mammaire en lactation. D'après nos hypothèses, des gènes ayant le même mode d'expression devraient être localisés sur le même domaine chromatinien afin de permettre une régulation indépendante par rapport aux gènes voisins.

Nos observations suggéraient la présence de deux domaines chromatiniens au sein du BAC 41 et l'existence d'une frontière capable de séparer efficacement les gènes de la cycline T1 et FLJ20436 d'un côté et de l' α -lactalbumine de l'autre (**Fig. 5.1**).

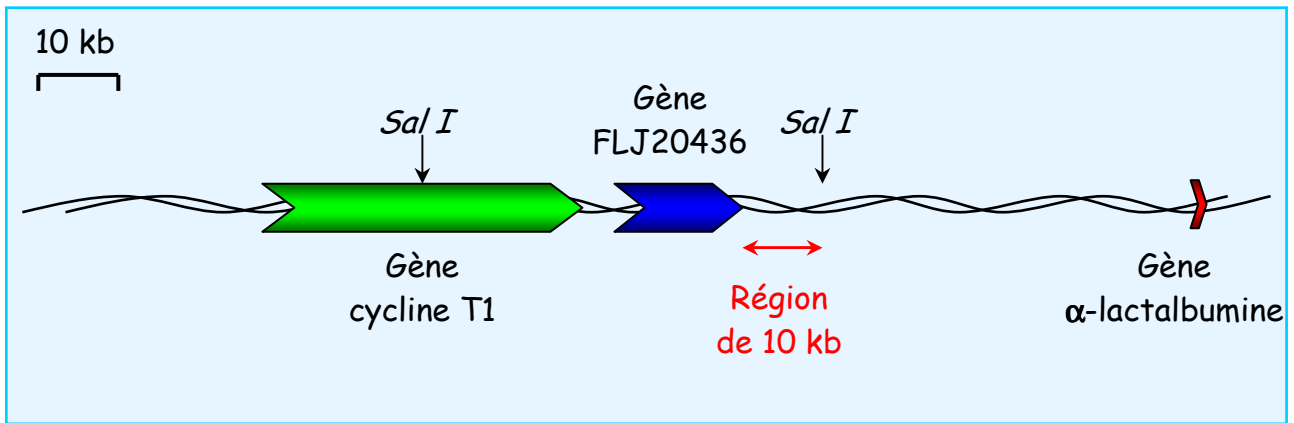


Fig. 5.1 : Schéma du BAC 41.

La caractérisation structurale de ces deux gènes ubiquistes nous a permis d'établir une carte précise du BAC 41 avec la localisation exacte des différents gènes, leur orientation et leur position les uns par rapport aux autres et de délimiter une région qui pourrait potentiellement jouer le rôle de frontière. Ainsi une séquence de 10 kb, située entre l'extrémité 3' du gène FLJ20436 et l'extrémité 5' du BAC 6, a été suspectée (**Fig. 5.1**).

II. Caractérisation de la région frontière

Afin de vérifier si cette région avait un rôle fonctionnel de frontière entre domaines chromatinien, différentes approches complémentaires ont été mises en œuvre pour essayer d'évaluer ses caractéristiques.

Le clonage et le séquençage de l'ensemble de cette région de 10 kb ont été réalisés. Malgré la divergence de séquence entre les différents isolateurs déjà caractérisés dans la littérature, on pouvait supposer qu'un même isolateur, de part son rôle fonctionnel important, présente un certain nombre de motifs conservés au cours de l'évolution. Les séquences obtenues chez la chèvre pouvaient être comparées aux bases de données existantes pour mettre en évidence des zones de forte homologie entre les différents génomes disponibles, principalement l'homme et la souris. Une recherche de sites potentiels de fixation pour des protéines de type « anti-enhancer » telle que CTCF, a aussi été réalisée dans l'espoir de trouver des éléments caractéristiques de certaines séquences *cis*-régulatrices du même type. De même, un logiciel dédié à l'identification de sites pouvant être associés à des MARs, souvent présentes au niveau des frontières chromatinien, a été utilisé pour déterminer si cette région en contenait.

Dans la littérature, les exemples d'isolateurs bien caractérisés ne sont pas nombreux (*cf.* Introduction, Présentation de quelques isolateurs caractérisés et Autres éléments régulateurs ayant des propriétés isolatrices). Le travail effectué sur le *locus* des globines du poulet a permis de mettre en évidence la frontière entre deux domaines chromatinien avec d'un côté le *locus* des globines et de l'autre le gène du récepteur au folate. Cette frontière est constituée de plusieurs éléments (*cf.* Introduction, Séquence 5' HS4 du *locus* de la β -globine de poulet) qui confèrent une régulation indépendante de l'expression des gènes présents sur ces domaines chromatinien voisins. Cette régulation indépendante se caractérise par des profils spécifiques de méthylation de la région « frontière » suivant que l'un ou l'autre *locus* est actif. Dans notre contexte chromatinien, l'analyse du profil de méthylation de cette région de 10 kb pouvait permettre de mettre en évidence une spécificité en fonction du tissu dans lequel est effectuée l'analyse. En effet, le gène de l' α -lactalbumine étant un gène spécifiquement exprimé dans le tissu épithéliale mammaire en lactation, son profil de méthylation devrait être différent suivant le tissu analysé, déméthylé dans la glande mammaire en lactation mais méthylé dans tous les autres tissus de même probablement que dans la glande mammaire à d'autres stades physiologiques. A l'inverse, les gènes de la cycline T1 et FLJ20436 étant exprimés de manière ubiquiste, la méthylation de leur ADN devrait être identique dans tous les tissus analysés. Une telle étude a été initiée.

Une approche plus fonctionnelle a également été entreprise avec une étude *in vitro* du potentiel de la frontière putative. Des analyses d'expression en culture cellulaire, qui vont permettre de tester la propriété de blocage de l'action des activateurs sur un promoteur, ont été réalisés pour essayer de mettre en évidence des zones isolatrices au sein de cette région de 10 kb. L'approche choisie utilise un vecteur grâce auquel le niveau d'expression du gène rapporteur, le gène de la luciférase, est mesurable en transfection transitoire. Il a été montré, notamment par l'équipe du Dr Felsenfeld (Recillas-Targa *et al.*, 1999) et plus récemment par une équipe japonaise (Ogata *et al.*, 2003), que des expériences *in vitro* de transfection transitoire sont suffisantes pour un premier criblage lors de la recherche de séquences potentiellement isolatrices. Une approche par transfection stable, beaucoup plus lourde, n'est habituellement envisagée que dans la mesure où quelques régions de petites tailles ont montré des résultats encourageants à la suite des transfusions transitoires. Le gène luciférase sous dépendance d'un promoteur minimal ubiquiste est fortement stimulé par un activateur viral placé en 5' (*cf.* CHAPITRE V, Description du vecteur de base Enh-TKluc). L'intégration

d'un isolateur bien caractérisé, tel que la séquence 5' HS4 de la β -globine de poulet, entre l'activateur viral et le promoteur montre une chute du niveau d'expression du gène de la luciférase. L'accès à la séquence de la région frontière a permis son sous-clonage précis dans le vecteur d'expression afin de tester si des fragments issus de cette zone étaient capables d'influencer le niveau d'expression du gène rapporteur.

En parallèle à ces expériences, une étude de la région a été entreprise *in vivo*. Deux constructions pour la transgénèse chez la souris ont été envisagées pour tester les propriétés de ces 10 kb :

- son association au BAC 6 afin de vérifier si la régulation de l'expression du gène de l' α -lactalbumine devenait alors équivalente à celle obtenue avec le BAC 41
- son association à un gène α -lactalbumine ne disposant que d'un promoteur court afin de vérifier si cette séquence est capable d'agir seule ou si elle nécessite l'intervention d'éléments distaux présents sur le BAC 6 (Fig. 5.2).

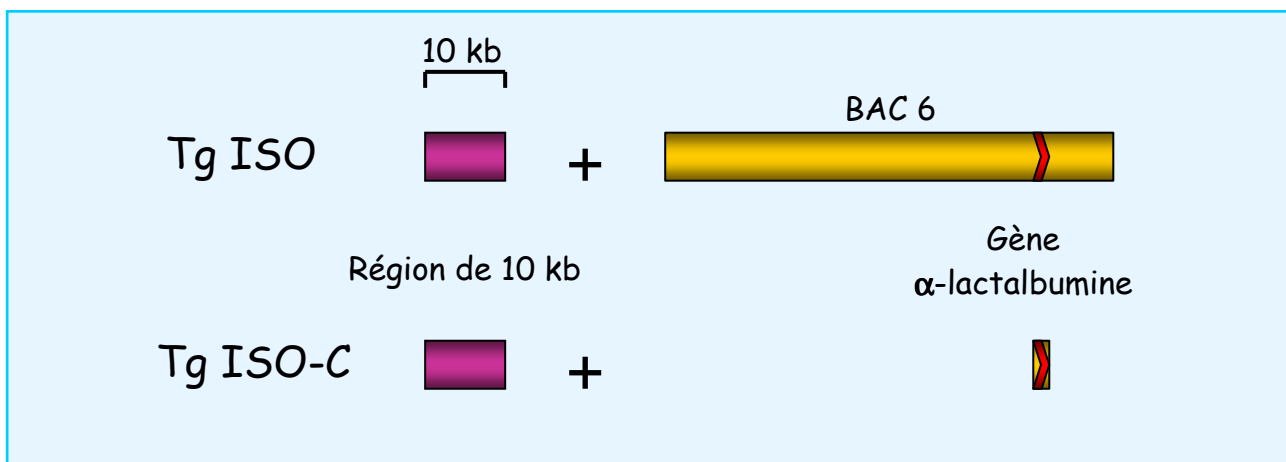


Fig. 5.2 : Schéma des constructions Tg ISO et Tg ISO-C.

Matériels et méthodes

L'ensemble des protocoles de base utilisés au cours de ce travail de thèse est issu d'un manuel dédié à la biologie moléculaire (Sambrook and Russel, 2001). Dans cette section, seuls les protocoles ne se rapportant pas à ce manuel seront donc référencés.

I. Obtention et analyse de la séquence de la région frontière

1. Sous-clonage de la région

La région frontière a été amplifiée par PCR en utilisant le kit « Expand long template PCR system » (Roche molecular biochemicals) qui contient un mélange de Taq polymérase très fidèles pour éviter toute erreur au niveau de la séquence. Une digestion partielle de la région de 10 kb a été réalisée à l'aide de l'endonucléase de restriction Sau3A. Après migration sur gel d'agarose 1%, les fragments de tailles comprises entre 0,7 et 1 kb ont été isolés en utilisant le kit « Wizard SV gel and PCR clean-up system » (Promega). La réaction de ligature a été effectuée dans un volume final de 20 µl pendant la nuit à 16°C. Lors de la préparation de la solution, un excès d'insert (1 à 5 fois en molarité) par rapport au vecteur (pUC 19 linéarisé par l'endonucléase de restriction BamH I, soit 50 ng, dont les extrémités sont compatibles avec Sau3A), a été ajouté au mélange pour une réaction optimale. 1 unité de ADN ligase T4 (Promega) a été utilisée dans ce volume réactionnel. La transformation des plasmides a été effectuée par électroporation dans des bactéries DH5α. Pour l'étalement des bactéries sur boîtes de Pétri après transformation, de l'Agar (15 g/l) a été ajouté au milieu LB (10 g/l bacto tryptone, 5 g/l d'extrait de levure, 10 g/l de NaCl) supplémenté en antibiotique (100 µg/ml d'ampicilline) ainsi qu'avec de l'IPTG et de l'X-Gal (10 µg/ml) pour la sélection blanc/bleu des clones résultants.

L'ensemble des cultures bactériennes a été réalisé en milieu LB supplémenté en antibiotique. Les ADN plasmidiques ont été extraits par la méthode de lyse alcaline pour les petits volumes (moins de 3ml de culture) et par l'utilisation du kit « Nucleobond AX » (Macherey-Nagel) pour des extractions de plus de 50 ml de culture bactérienne.

2. Séquençage et contigage de la région

A la suite du sous-clonage de cette région, une trentaine de clones, contenant des fragments de taille attendue (entre 0,7 et 1kb), ont été choisis pour être séquencés dans les deux sens. Les réactions de séquençage ont été effectuées sur les plasmides purifiés en utilisant le kit « Big Dye terminator » (Applied system) selon la technique de Sanger *et al.* (1987). L'analyse de la séquence a été réalisée par électrophorèse sur un séquenceur automatique ABI Prism 377 (Applied system).

La recherche de séquences chevauchantes entre les différents fragments a été réalisée in silico. Un premier alignement a été exécuté à l'aide des logiciels PHRED/PHRAP (Ewing and Green, 1998 ; Ewing *et al.*, 1998). Les lacunes entre les contigs ainsi mis en évidence ont été comblées par l'amplification de ces régions par PCR. Les réactions de séquençage ont alors été réalisées directement sur ces fragments amplifiés. Le programme CONSED a été utilisé pour éditer le résultat de l'assemblage (Gordon *et al.*, 1998).

3. Analyse de la séquence

Des recherches d'homologie ont été effectuées à l'aide de programmes spécialisés tels que BLAST du National Center for Biology Information (<http://www.ncbi.nlm.nih.gov/BLAST/>). D'autres sites ont été utilisés tels que le centre de ressources d'Infobiogen (<http://www.infobiogen.fr/>) pour des analyses telles que la réalisation de cartes de restriction, la conversion de séquences, les alignements locaux et globaux entre deux séquences d'acides nucléiques.

Une recherche plus spécifique de prédiction de MARs potentiellement présentes au sein de cette région frontière a été obtenue à l'aide du programme MAR-Wiz 1.5 (<http://www.futuresoft.org/MAR-Wiz/>).

II. Analyse de la région frontière par transfection transitoire

1. Description du vecteur de base Enh-TKluc

Le vecteur d'expression utilisé pour les transfections transitoires de culture cellulaire est basé sur le vecteur pGL3-basic (pGL3 luciferase reporter vectors, Promega) auquel plusieurs éléments ont été rajoutés par l'équipe du Dr L.-M. Houdebine (BDR, INRA) :

1) Le gène rapporteur

Déjà présent au sein du vecteur pGL3-basic, le gène rapporteur est le gène de la luciférase de luciole, dont l'expression est facilement mesurable. Ce gène est connu pour être peu sensible aux effets d'extinction chez les mammifères malgré une origine phylogénétique éloignée. Sa région codante a été optimisée afin de faciliter l'analyse de l'activité transcriptionnelle dans les cellules eucaryotes transfectées.

2) Un promoteur ubiquiste

Le promoteur du virus simien SV40 présent dans le vecteur pGL3-basic a été remplacé par le promoteur minimal du gène de la Thymidine Kinase (TK) du virus de l'Herpès simplex humain (position -109 à +55). Ce promoteur TK nécessite une forte stimulation par un activateur pour obtenir de bons niveaux d'expression du gène rapporteur.

3) Un activateur fort

L'amplificateur choisi pour être introduit dans ce vecteur pGL3-basic modifié a été celui du virus des polyomes (EnhPy) qui a fait ses preuves lors de travaux antérieurs. Cet activateur a été placé en amont du promoteur TK pour une stimulation efficace de la transcription.

4) Un terminateur de la transcription

Le terminateur déjà présent au sein du vecteur pGL3-basic est le terminateur des gènes tardifs du virus simien SV40. De nombreux travaux ont pu montrer son efficacité en transfection de culture cellulaire.

5) Schéma global

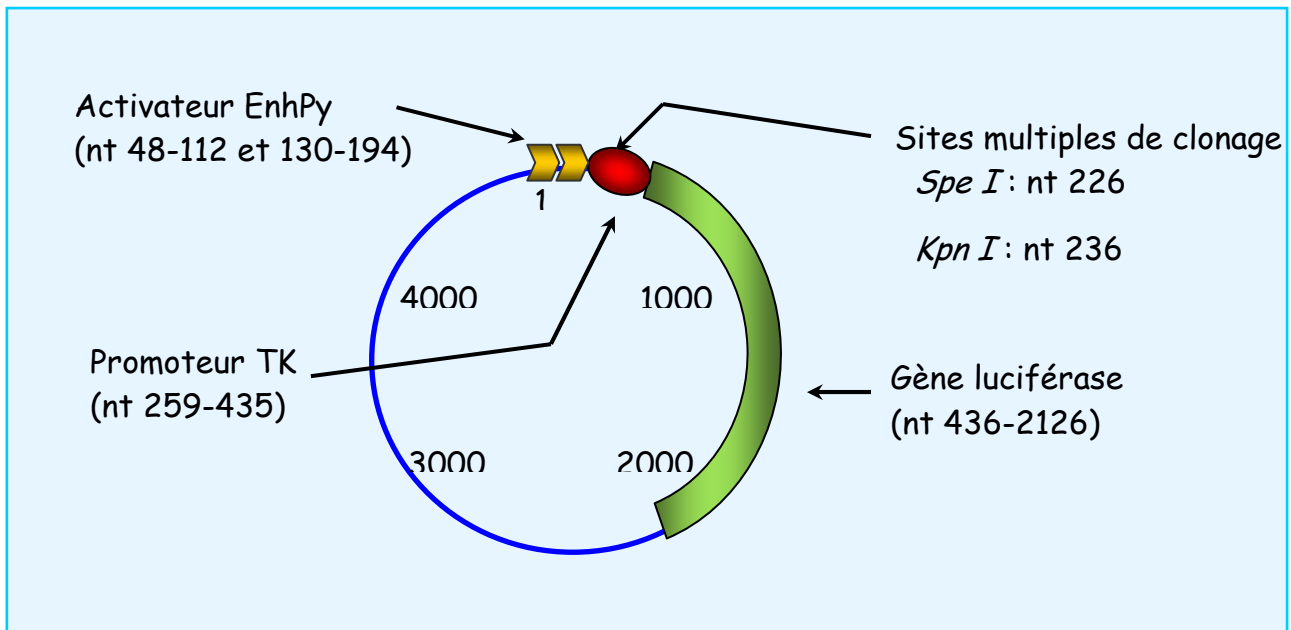


Fig. 5.3 : Schéma du vecteur de base Enh Tkluc (5211 bp).

Vecteur mis au point par l'équipe du Dr. L.-M. Houdebine (BDR-INRA).

2. Constructions de vecteurs d'expression et sous-clonage de la région frontière

Plusieurs couples d'oligonucléotides ont été choisis le long de cette séquence de 10 kb afin de générer par PCR des fragments chevauchants de 1,3 à 1,8 kb (**Table 5.1**). Pour obtenir un clonage orienté de ces fragments dans le vecteur d'expression modifié, un site de reconnaissance par une endonucléase de restriction a été incorporé au sein de la séquence oligonucléotidique (**Table 5.1**). Les fragments obtenus après amplification par PCR ont été préalablement clonés dans le plasmide pGEM-T Easy (Promega) selon le protocole décrit plus haut (*cf.* CHAPITRE V, Obtention et analyse de la séquence de la région frontière).

Les digestions enzymatiques d'ADN ont été réalisées suivant la méthode classique en respectant les prescriptions du fournisseur (Promega), avec un excès de 2 à 3 unités d'enzymes par μg d'ADN pour des coupures optimales. Les digestions ont été testées par électrophorèse d'un échantillon sur gel d'agarose 1% en présence de bromure d'éthidium et d'un marqueur de taille. La purification des fragments d'ADN digérés a été réalisée après migration sur gel d'agarose 1% en utilisant le kit « Wizard SV gel and PCR clean-up system » (Promega). Les fragments ainsi obtenus ont été clonés dans le vecteur d'expression EnhTKluc préalablement linéarisé avec deux endonucléases permettant d'orienter les fragments de la région frontière (**Fig. 5.3** et **Table 5.1**).

Nom de l'oligonucléotide ou de la sonde	Séquence	Position	III. Orientation	Taille du fragment (nt)
ctg1F_Spe I	5'-CGAGTGT ACTAGT CAGAGGTTGG-3'	nt 1-23	sens	1365
ctg1R_Kpn I	5'-GGATACAG GTACCA TGGCTTC-3'	nt 1345-1365	antisens	
ctg2F_Spe I	5'-TAAAACCT ACTAGT CTAGTTATA-3'	nt 615-637	sens	1525
ctg2R_Kpn I	5'-GGATCTTT GGTACCC AGAAATC-3'	nt 2119-2140	antisens	
ctg3F_Spe I	5'-GCTGGCAA ACTAGT AGCATCC-3'	nt 1252-1272	sens	1554
ctg3R_Kpn	5'-CAGCTGAT GGTACCA TATCCTTC-3'	nt 2785-2806	antisens	
ctg4F_Spe	5'-GATTTCTTA ACTAGT GCATGTTG-3'	nt 2033-2055	sens	1485
ctg4R_Kpn	5'-TCTACTGAG GTACCA TTATCCAT-3'	nt 3497-3518	antisens	
ctg5F_Spe	5'-GTATCTGT ACTAGT TCTAGGG-3'	nt 2757-2777	sens	1620
ctg5R_Kpn	5'-CATGTGCT GGTACCA TCCTTAG-3'	nt 4357-4377	antisens	
ctg6F_Kpn	5'-ATGGATAAG GTACCA CTCAGTAGAC-3'	nt 3497-3519	sens	1522
ctg6R_Sac	5'-TTAAGGGT GAGCTC TGCTATTC-3'	nt 4998-5019	antisens	
ctg7F_Kpn	5'-CTAAGGAG GTACCA GCACATG-3'	nt 4357-4377	sens	1509
ctg7R_Sac	5'-TGATGTT GAGCTC TTTTTCATG-3'	nt 5845-5866	antisens	
ctg8F_Spe	5'-GGACCCATA ACTAGT CCCACCG-3'	nt 4968-4988	sens	1774
ctg8R_Kpn	5'-ATTTCTCCT GGTACCA CTTGAG-3'	nt 6722-6742	antisens	
ctg9F_Spe	5'-CACAGAGGA ACTAGT ATAGACG-3'	nt 5787-5807	sens	1540
ctg9R_Kpn	5'-GCCCTTGAG GTACCA CTCTATG-3'	nt 7308-7327	antisens	
ctg10F_Spe	5'-TTACAA ACTAGT CTCAAGATTGC-3'	nt 6712-6732	sens	1757
ctg10R_Kpn	5'-ATTAAG GTACCA CATTTACAATCC-3'	nt 8447-8469	antisens	
ctg11F_Spe	5'-ATTCGCA ACTAGT TAGATGGAC-3'	nt 7297-7318	sens	1558
ctg11R_Kpn	5'-GGGCAGAG GTACCA ATCTACC-3'	nt 8833-8855	antisens	
ctg12F_Kpn	5'-AAG GTACCA CTTGACTACTGGATTG-3'	nt 8430-8452	sens	1501
ctg12R_Sac	5'-GAATATCT GAGCTC CGACAGCC-3'	nt 9910-9931	antisens	
ctg13F_Kpn	5'-GGTAGATT GGTACCA CTCTGCC-3'	nt 8833-8855	sens	1612
ctg13R_Sac	5'-CTCAG GAGCTC ATAGTGTGTG-3'	nt 104425-10445	antisens	
Sonde ISO 1		nt 1-331		331
Sonde ISO 2		nt 2405-2806		402
Sonde ISO 3		nt 10032-10445		414

Table 5.1 : Caractéristiques des oligonucléotides et des sondes utilisés.

Position et orientation des oligonucléotides, localisation des sondes et taille des fragments chevauchants obtenus par PCR.

1. Cultures cellulaires et la transfection

1) Milieux et Conditions de culture

Les milieux de culture, les réactifs, les antibiotiques et le transfectant ont été obtenus auprès de Life technology et de Sigma. Les cellules ont été cultivées en incubateur à 37°C, humidifié pour prévenir l'évaporation du milieu et à pression constante en CO₂ (5%) pour maintenir le pH entre 7,0 et 7,2 en présence de bicarbonate. Avant toute utilisation, le sérum de veau fœtal SVF (Bio West) a été préalablement décomplémenté par traitement thermique (30 minutes à 56°C).

Les cellules utilisées pour les transfections ont été des cellules ovariennes de hamster chinois (CHO). Elles ont été maintenues dans un milieu DMEM/F12 (50V/50V) additionné de 10% de SVF décomplémenté, de L-glutamine 2 mM, de pénicilline et de streptomycine à 100 µg/ml. Les cellules mises en culture ont été traitées à la trypsine tous les 2 à 3 jours afin de leur permettre une croissance et une multiplication optimales.

2) Transfection

Les cellules CHO entre 50 et 60% de confluence, en boîtes de 6 cm de diamètre, ont été transfectées avec 5 µg du plasmide d'intérêt en présence de la LipofectamineTM comme transfectant dans de l'OptiMEM (Life technology) selon les prescriptions du fournisseur. Un rapport ADN/LipofectamineTM de 1/4 a été retenu pour réaliser l'ensemble de ces transfections. Quatre heures après le début de la transfection, le mélange ADN/LipofectamineTM dans l'OptiMEM a été remplacé par le milieu de culture habituel, DMEM/F12 (50V/50V) additionné de 10% SVF décomplémenté, de L-glutamine 2 mM, de pénicilline et de streptomycine à 100 µg/ml. Lors de chaque transfection, plusieurs plasmides dits de référence ont été également utilisés afin d'évaluer l'efficacité de l'expérience. Chaque résultat présenté est la moyenne de 3 transfections indépendantes d'un même plasmide dans les mêmes conditions. Les cellules ont été récupérées 48 heures après la transfection afin de procéder à la mesure de l'expression du gène rapporteur.

2. Analyse du niveau d'expression du gène rapporteur

1) Prélèvement des cellules transfectées

Les cellules ont été rincées en Phosphate Buffer Saline : PBS (Life technology) puis grattées dans 1,5 ml de PBS. Elles ont été ensuite culottées par centrifugation pendant 5 minutes à 5000 g (centrifugeuse Sigma 2MK). Le culot a été lysé par 3 à 4 chocs thermiques successifs, congélation dans l'azote liquide puis décongélation rapide à 37°C, en présence de 300 µl de tampon de lyse (Glycylglycérine 25 mM, EDTA 1 mM, MgSO₄ 8 mM, Triton 1%, Glycérol 15%, et DTT 1 mM ajouté extemporanément).

2) Dosage de la luciférase par bioluminescence

Ce dosage a été préalablement décrit par l'équipe de Dr De Wet (De Wet *et al.*, 1987). Après avoir lysé les cellules, l'activité luciférase a été immédiatement mesurée en présence d'une solution contenant $\frac{1}{4}$ de luciférine 1 mM et $\frac{3}{4}$ de tampon B (Glycylglycérine 25 mM, MgSO₄ 15 mM, Adénosine triphosphate 5 mM). Les mesures ont été effectuées à l'aide d'un luminomètre (Berthold Lumat LB9501). Cinq µl d'une dilution au 1/10 du lysat cellulaire ont été suffisants pour obtenir une valeur incluse dans la partie linéaire de la courbe de l'activité de la luciférase en fonction de la concentration d'enzyme. L'activité a alors été mesurée en unité arbitraire (RLU). Cette activité a été rapportée à la quantité de protéine totale dosée selon la technique de Bradford.

IV. Les souris transgéniques

1. Préparation des constructions pour la micro-injection

Les constructions de petites tailles (1 à 15 kb) ont été préparées par digestion enzymatique (gène de l' α -lactalbumine avec promoteur court) ou par PCR (région de 10 kb) *cf.* **Annexe 1**. Les fragments obtenus ont été séparés sur gel d'agarose 1% et la bande correspondante à la construction d'intérêt a été isolée et purifiée en utilisant le kit « Wizard SV gel and PCR clean-up system » (Promega). Les transgènes ont été repris dans une solution Tris-HCl 10 mM, EDTA 0,1 mM pH 7,4, qui a été utilisée comme tampon pour la micro-injection.

La construction de grande taille (50 à 60 kb), issue de la digestion enzymatique du BAC 6 par l'endonucléase de restriction *Not I*, a été séparée de son vecteur sur gel d'agarose 1%, extraite par électro-élution puis purifiée sur colonne DEAE-Sephacel (DiEthylAminoEthyl) anion

exchanger (Sigma). La solution d'ADN a alors été dialysée sur membrane (Millipore type VS 0,0025 µm) contre une solution protégeant les grands fragments des cassures et utilisée pour la micro-injection (Tris-HCl 10 mM pH 7,4, EDTA 0,1 mM, NaCl 100 mM, spermine 30 µM et spermidine 70 µM).

La région de 10 kb n'ayant pu être clonée ou assemblée au gène de l' α -lactalbumine avec promoteur court d'une part et au BAC 6 par recombinaison homologue chez *E. coli* d'autre part, les différents fragments ont été co-injectés dans des embryons de souris FVB/N. Les solutions d'ADN ainsi préparées ont été micro-injectées à des concentrations comprises entre 1 et 5 ng/µl, suivant la technique classique décrite dans l'ouvrage de Hogan *et al.* en 1986 par le personnel de l'atelier transgénèse présent au sein de notre unité. Lors de la préparation des solutions pour la micro-injection, un rapport de 1 pour 4 en molarité a été utilisé entre le gène de l' α -lactalbumine avec promoteur court et la région de 10 kb, et un rapport de 1 pour 3 entre cette même région et le BAC 6.

Différents protocoles ont été utilisés pour tenter de réaliser la recombinaison homologue chez *E. coli* nécessaire à la grande construction initialement envisagée (Yang *et al.*, 1997 ; Narayanan *et al.*, 1999 ; Muyrers *et al.*, 1999 ; Lee *et al.*, 2001). Les expériences n'ayant pas abouti, aucun détail sur ces techniques ne sera rapporté ici.

2. Identification et caractérisation des souris transgéniques

Le criblage de l'ensemble des animaux a été réalisé par PCR en utilisant un couple d'oligonucléotides spécifiques du gène de l' α -lactalbumine caprin et 3 autres spécifiques de la région de 10 kb (**Table 5.1**). L'extraction des ADN génomiques a été effectuée, à partir de l'extrémité des queues de souris prélevée 10 à 20 jours après la naissance des animaux, selon la technique classique de digestion à la protéinase K et d'extraction phénol/chloroforme. Les culots d'ADN génomique ont été repris et conservés dans une solution de Tris-HCl 10 mM, EDTA 1 mM, pH 7,8. Le même protocole a été utilisé pour l'extraction d'ADN génomique à partir de tissus de chèvre.

Pour évaluer le nombre de copies intégrées au sein des lignées mises en évidence par PCR, la technique (dite de Southern blot) de transfert sur membrane d'ADN génomique coupé par digestion enzymatique à l'aide de l'endonucléase de restriction *Pst* I, a été réalisée. Après migration sur gel d'agarose 1 %, les fragments d'ADN ont été dépurinés par immersion du gel

dans une solution de HCl 0,15 M, 2 fois 15 minutes sous agitation puis dénaturés dans un bain de soude 0,5 M, NaCl 1,5 M dans les mêmes conditions. Le gel a finalement été placé 30 minutes sous agitation dans une solution neutralisante (Tris-HCl 0,5 M pH 7,4, NaCl 3 M). Le transfert de l'ADN sur membrane de nylon Hybond N (Amersham biosciences) a été effectué à l'aide d'un pont salin selon le protocole original de Southern durant la nuit (**Fig. 5.4**). L'ADN a ensuite été fixé sur la membrane par exposition aux UV pendant quelques minutes.

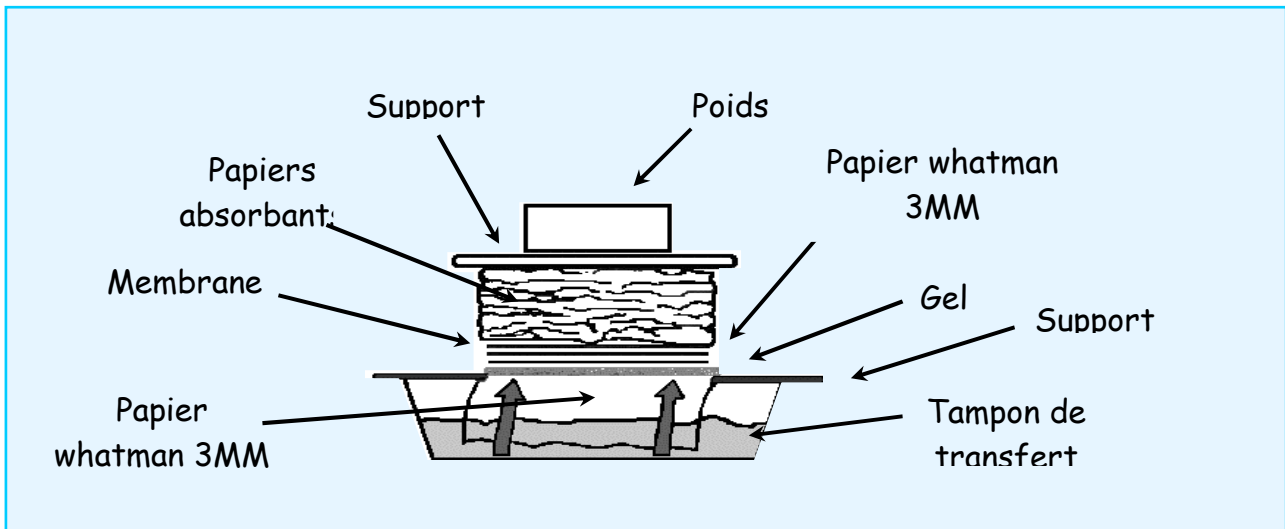


Fig. 5.4 : Schéma du montage pour le transfert d'acides nucléiques par capillarité.

Les membranes ont ensuite été hybridées avec une sonde spécifique selon le protocole de Church et Gilbert (Church and Gilbert, 1984). Une solution de tampon phosphate sodique 500 mM pH 7,2/SDS 7%, à raison de 500 $\mu\text{l}/\text{cm}^2$ de membrane, a été chauffée à 65°C pour préhybrider les membranes en sac scellé au bain marie à 65°C sous agitation lente pendant au moins 1 heure. La sonde (50 ng) a été préparée en parallèle avec le kit « rediprime™ II random prime labelling system » (Amersham biosciences). L'utilisation de ce kit permet d'obtenir des activités spécifiques des sondes de l'ordre de 10^9 dpm/ μg . La sonde a été dénaturée par addition de NaOH (0,4 N finale) pendant 2 minutes à température ambiante puis ajoutée à une solution de tampon phosphate sodique 500 mM pH 7,2/SDS 7% (500 $\mu\text{l}/\text{cm}^2$ de membrane) chauffée à 65°C en remplacement de la solution de pré-hybridation. La membrane a ensuite été incubée pendant 16 heures sous agitation lente à 65°C. Trois lavages successifs ont été réalisés en tampon phosphate sodique 40 mM pH 7,2/SDS 0,1% chauffée à 65°C pour éliminer les interactions non spécifiques de la sonde avec l'ADN fixé sur la membrane. Un film autoradiographique (Kodak X) en contact avec cette membrane a permis,

après incubation à -80°C en présence d'écrans amplificateurs, de mettre en évidence les fragments d'ADN marqués spécifiquement par la sonde radioactive.

3. Analyse du niveau d'expression du gène de l' α -lactalbumine dans les souris transgéniques et du taux de traduction

L'analyse des transcrits du gène de l' α -lactalbumine dans les souris transgéniques a été effectuée à partir d'ARN totaux extraits de plusieurs organes (cerveau, rate, rein, foie, cœur et glande mammaire en lactation). Après broyage des tissus en présence de RNAnow (Biogentex, inc), une extraction phénol/chloroforme a été réalisée conformément aux prescriptions du fournisseur. Les ARNs totaux obtenus ont été quantifiés par absorbance à 260 nm et fractionnés sur gel d'agarose dénaturant 1,5% en tampon MOPS 1X (Sigma). Les échantillons ont été déposés sur gel à l'aide de tampon de charge dénaturant (formamide, formaldéhyde, MOPS et bleu de bromophénol) après avoir été incubés à 65°C pendant 10 minutes. Après électrophorèse en condition dénaturante, le transfert des ARNs a été réalisé à l'aide d'un pont salin, comme décrit ci-dessus pour le transfert de l'ADN (*cf.* CHAPITRE V, Identification et caractérisation des souris transgéniques), sur membrane de nylon Hybond N (Amersham biosciences). Les ARNs ont ensuite été fixés par exposition aux UV pendant quelques minutes. L'hybridation de ces membranes a été effectuée dans les mêmes conditions que celles décrites ci-dessus (*cf.* CHAPITRE V, Identification et caractérisation des souris transgéniques).

Afin d'analyser la protéine α -lactalbumine caprine recombinante produite dans le lait de souris transgéniques, 50 à 100 μl de lait ont été prélevés au quatrième jour de lactation. Quatre volumes d' H_2O milliQ ont été ajoutés pour diluer les protéines puis la solution a été centrifugée pendant 10 minutes à 3500 g (centrifugeuse Sigma 2MK) et seule la phase aqueuse sous la couche de matière grasse a été récupérée pour être analysée. Les protéines ont ensuite été séparées en fonction de leurs poids moléculaires sur gel SDS/PAGE 12,5% (Sodium Dodecyl Sulfate/PolyAcrylamde Gel Electrophoresis) en tampon Tris-HCl 250 mM pH 8,8, SDS 0,07%, persulfate d'ammonium 0,05% et TEMED (TétraMéthyl EthylèneDiamine) 0,033%, le gel de concentration, situé au dessus du gel de résolution, étant un gel d'acrylamide à 3,75% en tampon Tris-HCl 125 mM pH 6,8, SDS 0,1%, persulfate d'ammonium 0,075% et TEMED (TétraMéthyl EthylèneDiamine) 0,075%. La migration a été effectuée en cuve « mini protean II » (Biorad) en présence de tampon de migration Tris-HCl

25 mM pH 7,4, glycine 192 mM et SDS 0,1% pendant 2 heures à 100 V. Le transfert des protéines a été réalisé sur membrane de nitrocellulose Hybond-C (Amersham biosciences) en présence de tampon de transfert Tris-HCl 10 mM pH 7,4, glycine 96 mM, SDS 0,005% et méthanol 20% dans un appareil « MilliBlot-Graphite Electrobloetter » (Millipore) pendant 1 heure 30 minutes sous une tension de 0,8 mA/cm². La membrane a été saturée en présence d'albumine de sérum bovin (5%) dans du tampon PBS pH 7,0 (H₂PO₄Na 10 mM, NaCl 150 mM et Tween 20 0,1%) sur la nuit à 4°C, puis elle a été incubée en présence d'un anticorps anti- α -lactalbumine caprine puis d'un anticorps secondaire anti-IgG de lapin couplé à la peroxydase. La révélation a été effectuée à l'aide du kit « ECLTM western blotting detection reagents » (Amersham biosciences).

Résultats et discussion

I. Analyse comparative de la séquence de la région frontière de 10 kb

1. Obtention de la séquence

La caractérisation du BAC 41 nous a permis de définir :

- l'existence d'éléments *cis*-régulateurs dominants en amont du BAC 6
- la présence de deux domaines chromatiniques putatifs contenant les gènes de la cycline T1 et FLJ20436 d'une part et celui de l' α -lactalbumine d'autre part et d'une région frontière permettant une régulation indépendante de ces domaines.

Afin de déterminer si cette région frontière dispose d'éléments capables de protéger les gènes des activateurs et inhibiteurs voisins présents au sein du génome, son analyse a été entreprise. Son séquençage a été réalisé à partir d'une digestion partielle *Sau3A* et d'un sous-clonage des fragments dont la taille était comprise entre 0,7 et 1kb. Le séquençage dans les deux sens d'une trentaine de clones a permis d'établir un premier contigage à l'aide des logiciels Phred/Phrap (Ewing and Green, 1998 ; Ewing *et al.*, 1998) et Consed (Gordon *et al.*, 1998) disponibles au laboratoire. Afin de combler les différentes régions manquantes, des amplifications par PCR ont été effectuées et les fragments ainsi obtenus séquencés. Le résultat est une séquence de 10 kb comprise entre l'extrémité 3' du gène FLJ20436 et l'extrémité 5' du BAC 6 (**Fig. 5.1** et *cf. Annexe 1*).

2. Comparaison avec les régions homologues humaine et murine

Avec l'avancement du séquençage des génomes, nous avons pu comparer les régions homologues chez l'homme (chromosome 12q13.11 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=54934>) et chez la souris (chromosome 15F1 NCBI LocusLink <http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=69612>). Cette analyse a permis de montrer que le génome caprin présente une organisation similaire, avec la présence des mêmes gènes, dans la même orientation. Seules les distances entre ces gènes et la taille des introns présentent des variations qui peuvent s'expliquer par des insertions/délétions d'éléments répétés dans ces zones semble-t-il non-codantes. Une

recherche d'homologie entre les séquences humaine, murine et caprine a été réalisée (<http://www.infobiogen.fr/> : ALIGNn et LFASTAn) en espérant que si des séquences fonctionnelles existent entre les deux domaines chromatiniens putatifs, elles soient caractérisées par une certaine conservation entre ces espèces. Aucune homologie significative n'a pu être mise en évidence entre les séquences caprine et humaine d'un côté et caprine et murine de l'autre. Lors de la comparaison entre les séquences humaine et murine, seules les zones contenant des éléments répétés ont pu être révélées.

Ces résultats n'apportent pas d'information supplémentaire permettant de cibler plus précisément la recherche des éléments régulateurs. Il existe une divergence au cours de l'évolution qui ne permet pas de repérer facilement des régions homologues ayant des fonctions similaires d'une espèce à une autre par simple alignement de séquences. Prenons l'exemple de l'isolateur 5' HS4 du poulet (Chung *et al.*, 1993), une recherche d'homologie à partir de sa séquence ne permet pas de retrouver les séquences homologues humaine ou souris au niveau de leurs *loci* respectifs. Il a même récemment été montré lors de la caractérisation du *locus* des globines chez l'homme, que le site hypersensible à la DNase 5' HS5, à priori équivalent par sa position de l'isolateur 5' HS4 du poulet, ne possédait pas les mêmes propriétés ne jouant plus que le rôle de frontière spécifique lors la différenciation des cellules érythroïdes (Wai *et al.*, 2003). Cette divergence des éléments impliqués dans la régulation et la protection de l'expression des gènes pourrait suffire à expliquer que la recherche d'homologie entre les séquences d'espèces différentes dans des zones non-codantes ne donne aucun résultat concluant, même en analysant l'ADN d'espèces phylogénétiquement plus proches.

II. Recherche de zones potentiellement impliquées dans la région frontière

1. Sites potentiels de fixation de la protéine CTCF

La protéine CTCF a pu être associée à plusieurs isolateurs précédemment caractérisés comme la séquence 5' HS4 du poulet ou encore au niveau de la région de contrôle de l'empreinte parental du *locus* Igf2/H19 murin (Bell *et al.*, 1999 ; Arney, 2003). Cette protéine est impliquée dans la fonction de blocage de la stimulation d'un promoteur par un activateur. Une séquence consensus a été établie : 5'-CCGCNNGGNGGCAG-3' (Farrell *et al.*, 2002). Cette

séquence a été recherchée au sein de cette région de 10 kb mais aussi dans les régions homologues de l'homme et de la souris (<http://www.infobiogen.fr/> : LFASTAn).

L'analyse des séquences disponibles n'a pas permis de trouver de sites susceptibles de correspondre aux sites de fixation pour la protéine CTCF dans aucune des trois espèces étudiées. A nouveau, cette observation n'implique pas que ces régions ne renferment pas de zones isolatrices, puisque comme nous l'avons décrit au cours de l'introduction, il existe dans la littérature plusieurs exemples de tels éléments fonctionnels qui n'interagissent pas avec cette protéine.

2. Sites potentiels de MAR

Au niveau de certaines régions frontières, des séquences capables de lier l'ADN à la matrice nucléaire pour former des boucles topologiques ont pu être mises en évidence (cf. Introduction, 1.5. Autres éléments régulateurs ayant des propriétés isolatrices). Ces séquences présentent certaines propriétés structurales communes détectables avec des logiciels adaptés de type MAR-Wiz (<http://www.futuresoft.org/MAR-Wiz/>). L'utilisation d'un tel logiciel de prédiction de MARs nous a permis de mettre en évidence deux sites potentiels. (**Fig. 5.5** et **Table 5.2**).

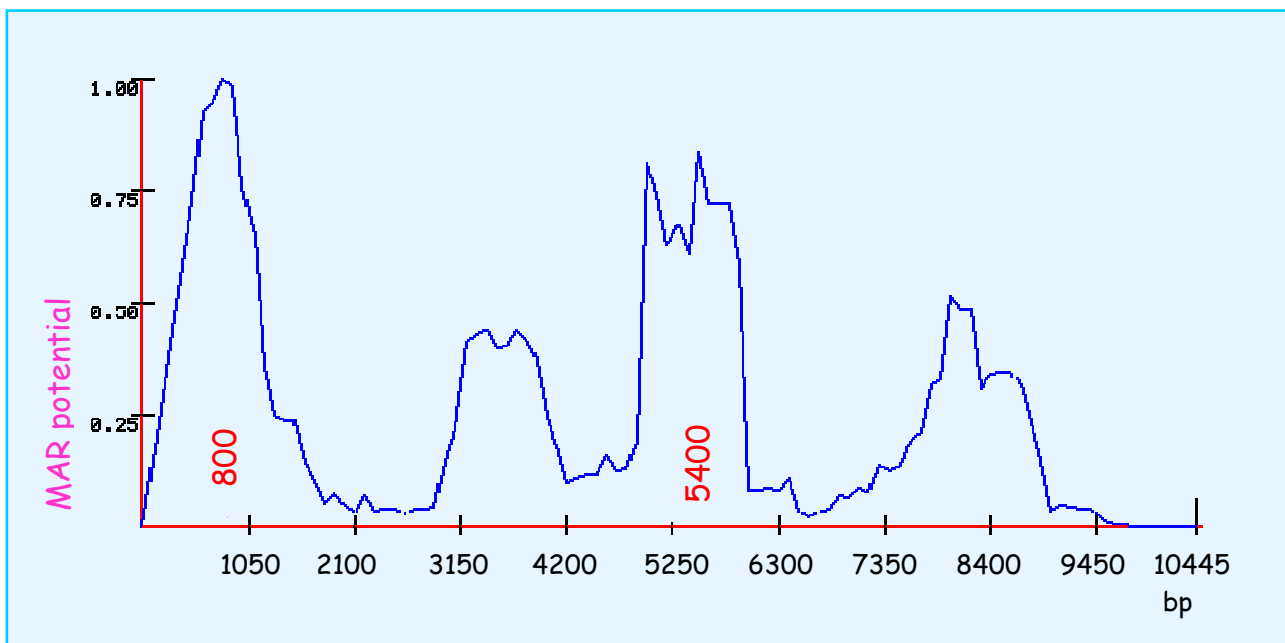


Fig. 5.5 : Courbe de prédiction de MARs sur la région de 10 kb. Cette courbe a été obtenue à partir du logiciel MAR-Wiz (<http://www.futuresoft.org/MAR-Wiz/>).

Région	Force intégrée	Force moyenne
nt 500-1000	0.834809	418.239
nt 5500-5800	0.671441	202.104

Table 5.2 : Paramètres corrélés à la courbe de prédiction de MARs (MAR-Wiz).

La prédiction de ces deux MARs renforce l'hypothèse d'une région frontière entre deux domaines chromatiniens putatifs. Cependant, seules des expériences complémentaires d'immuno-précipitation de la chromatine par exemple permettraient de confirmer ou d'infirmer les prédictions émises par ce logiciel. D'autre part la simple présence de séquences MAR n'implique pas que cette région chromosomique joue un rôle d'isolateur.

III. Analyse du profil de méthylation

L'obtention de la séquence de la région de 10 kb nous a permis de dresser une carte de restriction précise de cette zone concernant notamment des sites reconnus par des endonucléases sensibles au niveau de méthylation de l'ADN (*Hpa* II et *Msp* I). Ces deux enzymes ont la particularité de couper l'ADN au même site (C/CGG) mais seul *Hpa* II est sensible à la méthylation de l'ADN génomique des mammifères, c'est-à-dire qu'il ne peut cliver au niveau de ce site si les doublets CpG sont méthylés. A l'aide de deux sondes, une première en 5' et une seconde centrale (**Table 5.1**), nous avons initié l'analyse du degré de méthylation de cette zone dans différents tissus. Deux tissus de chèvre ont été utilisés pour ces expériences dont un à deux stades physiologiques différents :

- le cœur
- la glande mammaire à 110 jours de gestation
- la glande mammaire en lactation

Le degré de méthylation de l'ADN est corrélé à l'activité transcriptionnelle des gènes (Newell-Price *et al.*, 2000). Lorsque l'ADN est méthylé, les gènes présents dans la région sont éteints. Les régions actives du génome sont à l'inverse déméthylées. Ces profils de méthylation peuvent être spécifiques d'un tissu, voire de certaines cellules, en accord avec l'activation ou non du *locus* étudié. Les régions frontières entre domaines chromatiniens vont permettre de stopper l'action des méthylases et d'éviter la propagation de la méthylation de l'ADN aux *loci* avoisinants.

Dans notre cas particulier, nous recherchions la mise en évidence de deux profils spécifiques de méthylation en fonction du tissu analysé puisque le gène de l' α -lactalbumine n'est exprimé que dans la glande mammaire en lactation tandis que les gènes de la cycline T1 et FLJ20436 sont actifs dans tous les tissus. Les résultats des hybridations ont montré que les régions 5' et centrale de cette zone restaient méthylées quelque soit le tissu ou le stade physiologique analysé. Cette étude devrait être complétée d'une part par l'emploi d'une sonde couvrant l'extrémité 3' de la zone étudiée mais aussi le *locus* α -lactalbumine. Ces deux analyses n'ont malheureusement pas pu être réalisées au cours de cette thèse.

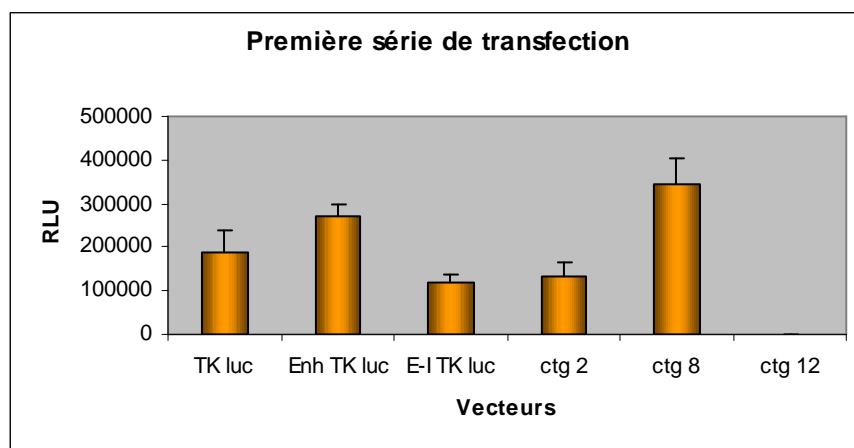
Ces résultats préliminaires, s'ils étaient corroborés par une déméthylation de l'ADN dans la glande mammaire en fin de gestation et lactation lors de l'utilisation des deux sondes susmentionnées serait un argument fort pour l'existence d'un isolateur. Ces analyses font indéniablement parties des priorités dans la poursuite de l'étude de ce *locus*.

IV. Analyse de la région de 10 kb par transfection de culture cellulaire

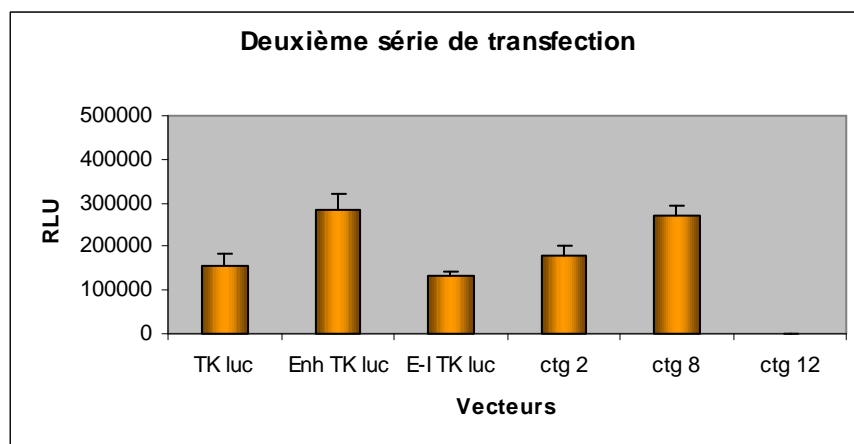
L'analyse d'une région d'intérêt pour la recherche d'éléments régulateurs de type isolateur peut être réalisée par transfection transitoire de cellules (Recillas-Targa *et al.*, 1999 ; Ogata *et al.*, 2003). Notre région d'intérêt a été sous-clonée dans un vecteur approprié à ce type d'analyse (*cf.* CHAPITRE V, Description du vecteur de base Enh-TKluc) en fragments de 1,3 à 1,8 kb se chevauchant les uns les autres afin de couvrir l'ensemble des 10 kb (**Table 5.1**). Le choix de la taille des fragments insérés tient compte de la taille moyenne des isolateurs décrits (Spana *et al.*, 1988 ; Kellum and Schedl, 1992 ; Chung *et al.*, 1997) et de l'inhibition artificielle que pourrait générer des inserts trop grands. Tous les fragments ont été orientés de manière identique. Plusieurs plasmides témoins ont été utilisés à chaque transfection :

- un vecteur pour tester le niveau basal d'expression du gène rapporteur luciférase, en absence d'activateur (TKluc)
- un vecteur contenant un activateur pour avoir un niveau d'expression de référence en absence de séquences isolatrices (Enh TKluc)
- un vecteur contenant l'activateur mais aussi un isolateur précédemment caractérisé tel que la séquence 5' HS4 du poulet, pour avoir un niveau de référence lorsqu'une telle région est présente entre l'activateur et son promoteur. (E-I TKluc).

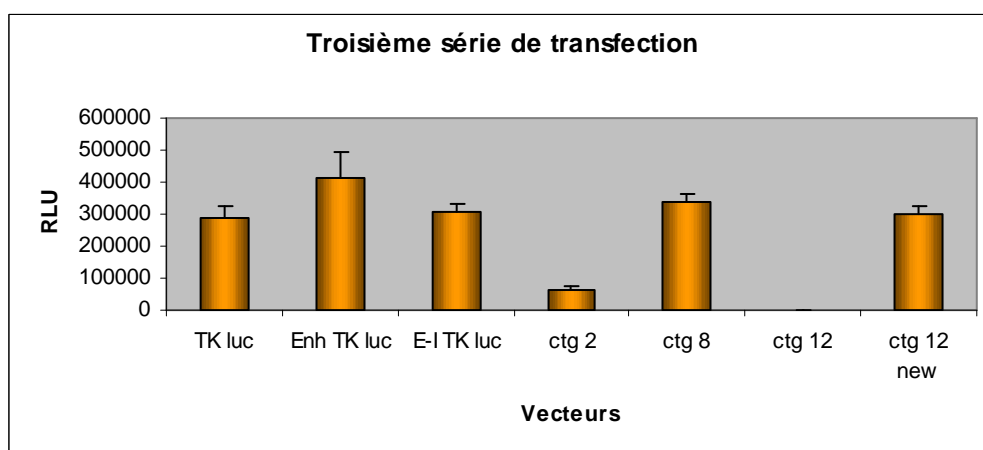
Le niveau d'expression du gène rapporteur est ensuite analysé pour l'ensemble des constructions disponibles et une comparaison des niveaux obtenus est effectuée par rapport aux différents niveaux des vecteurs de référence (**Graphiques 5.1, 5.2 et 5.3**).



Graphique 5.1 : Résultats de la première série de transfections transitoires.



Graphique 5.2 : Résultats de la deuxième série de transfections transitoires.



Graphique 5.3 : Résultats de la troisième série de transfections transitoires.

Certaines difficultés sont apparues au cours de ces expériences. Malgré l'emploi de plusieurs préparations plasmidiques différentes, un niveau d'expression basal relativement haut pour le vecteur TKluc, une faible stimulation par l'activateur dans le vecteur Enh TKluc et de fait une inhibition du niveau de l'expression seulement d'un facteur 2 pour le dernier vecteur témoin E-I TKluc ont été observés lors des trois séries de transfections réalisées (**Graphiques 5.1, 5.2 et 5.3**). Ces résultats très moyens ont rendus l'analyse de cette étude plus difficile.

Après avoir testé l'ensemble des 13 vecteurs recombinants dont les inserts couvrent l'ensemble de notre région d'intérêt et leurs témoins, seuls 2 fragments semblent potentiellement intéressants : ctg 2 (nt 615-2140, *cf. Annexe 1*) et ctg 12 (nt 8430-9931, *cf. Annexe 1*). Les autres vecteurs permettent un niveau d'expression du gène rapporteur qui n'est pas significativement différent de celui obtenu avec le vecteur Enh TKluc. A l'inverse, le vecteur contenant le fragment ctg 2 permet un niveau d'expression comparable au vecteur témoin E-I TK luc. Des analyses de variance qui prennent en compte les résultats obtenus au cours d'une même série de transfections et les différences entre séries démontrent que l'expression du vecteur renfermant ctg 2 est significativement inférieur à Enh TKluc et au vecteur ctg 8 mais identique à TK luc et E-I TKluc (Laloe D., SGQA-INRA et Le Guillou S., Génétique des Poissons-INRA, communication personnelle). Le vecteur ctg 12 quant à lui inhibe complètement l'expression du gène rapporteur. Cette dernière inhibition semblait étonnante car elle conduit à un niveau d'expression bien inférieur à celui observé avec le vecteur E-I TKluc, suggérant soit la présence d'un inhibiteur puissant, soit un défaut dans le vecteur utilisé. Afin de tester ces hypothèses, une seconde construction avec le fragment ctg 12 a été réalisée. Les résultats montrent un rétablissement du niveau d'expression de l'ordre

de ceux obtenus avec les autres vecteurs d'expression, suggérant une action neutre de l'insert ctg 12 sur l'expression du gène rapporteur et validant notre seconde hypothèse.

Le seul fragment potentiellement intéressant est donc ctg 2 qui est localisé en 5' de notre région frontière (nt 615-2140). Dans cette zone se trouve une des deux prédictions de MARs (**Fig. 5.5**) renforçant l'hypothèse d'une région pouvant jouer le rôle d'isolateur. La prédiction localise la MAR à cheval sur les fragments ctg 1 (nt 1-1365) et ctg 2 en position nt 500-1000 (*cf. Annexe 1*) avec une localisation relative au nt 800 en 5' du fragment ctg 2. Il est donc possible que seule une fraction de la séquence MAR soit présente dans notre insert ctg 2. Le vecteur recombinant renfermant le fragment ctg 1, qui devrait également contenir une partie de la MAR, ne montre aucune inhibition significative. Il faut toutefois rappeler que les fonctions d'isolateur et d'attachement à la matrice, souvent associées physiquement, correspondent à des séquences différentes. Si tel est le cas ici, notre région isolatrice putative serait donc située en 3' de la MAR, d'après l'orientation des différents *loci*. Cette localisation en 3' est intéressante si on se réfère au profil de méthylation. Si une transition existe, elle se situera aussi en 3' de la MAR, mais à une distance un peu plus grande que la zone isolatrice. Une telle situation ne semble pas illogique. La seconde MAR prédite, située en position 5500-5800 sur les fragments ctg 7 (nt 4357-5866) et ctg 8 (nt 4968-6742) (*cf. Annexe 1*), ne semble pas associée à une zone isolatrice. Cependant, des compléments d'analyse sont à envisager avec notamment :

- l'amélioration des niveaux des vecteurs témoins pour obtenir des résultats plus significatifs,
- la linéarisation des vecteurs avant de les transfecter pour tenter d'améliorer la répétitivité et l'amplitude des résultats
- un changement d'orientation du fragment ctg 2 afin de vérifier si son potentiel d'inhibition disparaît comme c'est souvent le cas pour des isolateurs dont l'effet est directionnel (Chung et al., 1993 ; Gerasimova et al., 1995 ; Dunaway et al., 1997).

V. Analyse de la région de 10 kb par transgénèse chez la souris

Les transfections de culture cellulaire ont donc permis de suspecter une zone d'avoir une activité capable de diminuer le niveau d'expression d'un gène rapporteur lorsqu'elle est placée entre un activateur et un promoteur. Pourtant, ces résultats sont insuffisants pour conclure aux propriétés potentielles de cette région frontière *in vivo*, seul test réellement

discriminatif. Afin de vérifier le caractère isolateur de notre séquence de 10 kb, deux constructions destinées à la transgénèse chez la souris ont été envisagées :

- Tg ISO : association de la région de 10 kb au BAC 6, réalisable par recombinaison homologue chez *E. coli*,
- Tg ISO-C : association de cette même région à un locus de l' α -lactalbumine ne disposant que d'un promoteur court.

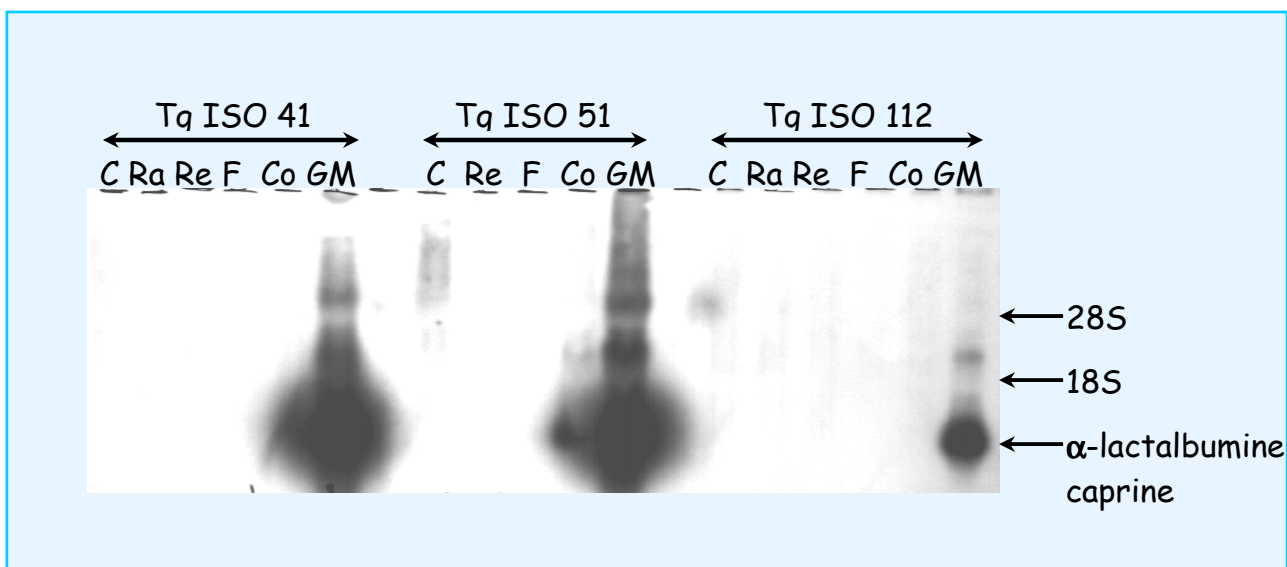
Rapidement, nous nous avons été confrontés à de nombreuses difficultés techniques. Après avoir amplifié cette région de 10 kb par PCR, nous avons été dans l'incapacité de cloner cette séquence dans un quelconque vecteur, aussi bien plasmidique que BAC. Ces problèmes ont empiré ensuite avec l'impossibilité de réaliser ces constructions par recombinaison homologue et ce malgré les différentes techniques utilisées (Yang *et al.*, 1997 ; Narayanan *et al.*, 1999 ; Muyrers *et al.*, 1999 ; Lee *et al.*, 2001). N'arrivant pas à obtenir ces constructions sous la forme souhaitée dans des délais raisonnables, nous avons décidé de co-injecter un mélange des deux fragments nécessaires (*cf.* CHAPITRE V, Préparation des constructions pour la micro-injection) et ensuite de cribler les animaux pour ne garder que ceux disposant des deux fragments au même site d'intégration. La micro-injection des ces deux « constructions » n'a généré que très peu de fondateurs Tg ISO et aucun Tg ISO-C. De plus, la majorité des souris transgéniques Tg ISO ne présentait qu'un seul des deux fragments ou, de façon plus surprenante, les deux mais intégrés à des sites différents rendant ces animaux inexploitable. En approfondissant l'analyse des souris transgéniques pour s'assurer de la présence des deux fragments (région de 10 kb + BAC 6), des remaniement de la région frontière ont pu être observés avec des délétions en 5' dont il faudrait affiner la localisation. Au final, seules trois lignées ont pu être obtenues : Tg ISO 41, 51 et 112, la lignée Tg ISO 51 présentant une délétion au niveau de son extrémité 5' d'au moins 0,5 kb (**Table 5.3**).

Constructions	Tg ISO	Tg ISO-C
nombre de souriceaux analysés	113	35
nombre de souris transgéniques	11	0
Souris Tg avec le BAC 6 seulement	6	0
Souris Tg avec le BAC 6 et la région de 10 kb mais sites intégration différents	2	0
Souris Tg avec le BAC 6 et la région de 10 kb	3	0

Table 5.3 : Récapitulatif des résultats relatifs à la micro-injection et aux souris transgéniques.

L'analyse de ces trois lignées par Northern blot révèle un niveau d'expression élevé du transgène dans la glande mammaire lactante pour les lignées Tg ISO 41 et 51 et bien plus faible pour la lignées Tg ISO 112. Cette analyse nous a permis de vérifier la spécificité tissulaire d'expression des transgènes (**Fig. 5.6**). Une faible expression du gène caprin était observable dans le cœur de la souris de la lignée Tg ISO 51 analysée. Cependant une hybridation de ces mêmes ARNs à l'aide d'une sonde β -caséine murine, obtenue auprès du Dr. J.M. Rosen (Baylor College, USA), nous a permis de conclure que ce signal résultait plus probablement d'une contamination de l'échantillon au moment du prélèvement (**Fig. 5.7**). Ces résultats bien qu'encourageant ne présuppose pas des propriétés de cette région de 10 kb puisque des observations identiques avaient été décrites lors de l'utilisation du BAC 6 en transgénèse, excepté pour la lignée 27 qui présentait une légère expression ectopique au niveau du cerveau (Soulie *et al.*, 2000).

Fig. 5.6 : Analyse de l'expression du gène α -lactalbumine caprin. Les ARNs totaux ont



été extraits à partir de différents organes de souris Tg ISO 41, 51 et 112. **C** : cerveau, **Ra** : rate, **Re** : rein, **F** : foie, **Co** : cœur, **GM** : glande mammaire lactante. Vingt μ g d'ARNs totaux ont été déposés pour chaque échantillon.

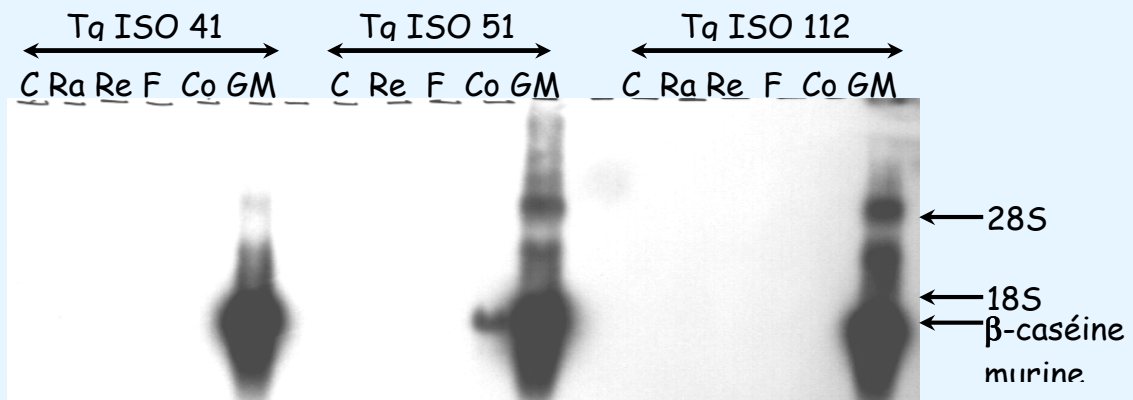


Fig. 5.7 : Analyse de l'expression du gène β -caséine murin, (cf. légende Fig. 5.6).

Une estimation relative du nombre de copies intégrées pour chaque lignées a été effectuée par analyse de Southern et à l'aide du logiciel Scion Imaging Software (<http://www.scioncorp.com/>), ainsi qu'une estimation de la quantité de protéine α -lactalbumine caprine sécrétée dans le lait de ces souris par analyse de Western (**Fig. 5.8 et Table 5.4**).

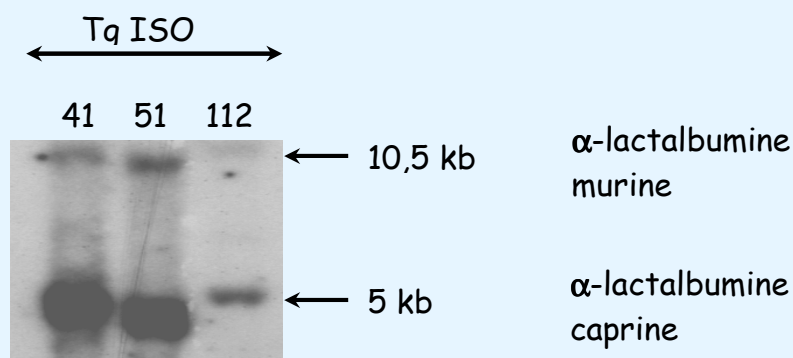


Fig. 5.8 : Analyse par Southern blot des lignées Tg ISO 41, 51 et 112.

Cette analyse a été réalisée sur des ADN génomiques, extraits à partir d'un morceau de la queue de ces souris transgéniques. Après digestion par l'endonucléase de restriction Pst I et transfert sur membrane des ADNs, l'hybridation a été effectuée à l'aide d'un mélange de sondes α -lactalbumine caprine et murine.

Lignées	Tg ISO 41	Tg ISO 51	Tg ISO 112
Estimation relative du nombre de copies	6	5	4
Estimation de la quantité de protéine α -lactalbumine (mg/ml/copie)	1,67	1	0,025

Table 5.4 : Récapitulatif des données obtenues sur les trois lignées transgéniques

Il a été précédemment montré, lors des expériences de transgénèse utilisant le BAC 41, que la quantité moyenne de protéine recombinante produite par copie intégrée du BAC 41 était de l'ordre de 0,7 à 0,8 mg/ml (**Fig. 5.9**) (Stinnakre *et al.*, 1999). Cet ordre de grandeur était cohérent avec une expression indépendante du site d'intégration du transgène puisque la quantité physiologique de protéine α -lactalbumine produite dans le lait est d'environ 1,2 mg/ml chez la chèvre (Mahé MF., LGBC-INRA, Communication personnelle). Notre analyse des lignées Tg ISO 41, 51 et 112, en ce qui concerne la quantité de protéine α -lactalbumine caprine produite dans le lait, montre comme pour les souris transgéniques pour le BAC 6 (Soulier *et al.*, 2000), un manque de proportionnalité entre le nombre de copies intégrées et la quantité de protéine synthétisée, ainsi qu'une activité transcriptionnelle par copie intégrée inférieure à ce qui était observé pour le BAC41 pour la lignée ISO 112 (**Table 5.4**). Les résultats obtenus pour les lignées Tg ISO 41 et Tg ISO 51 rappellent ceux obtenus avec la construction BAC 41.

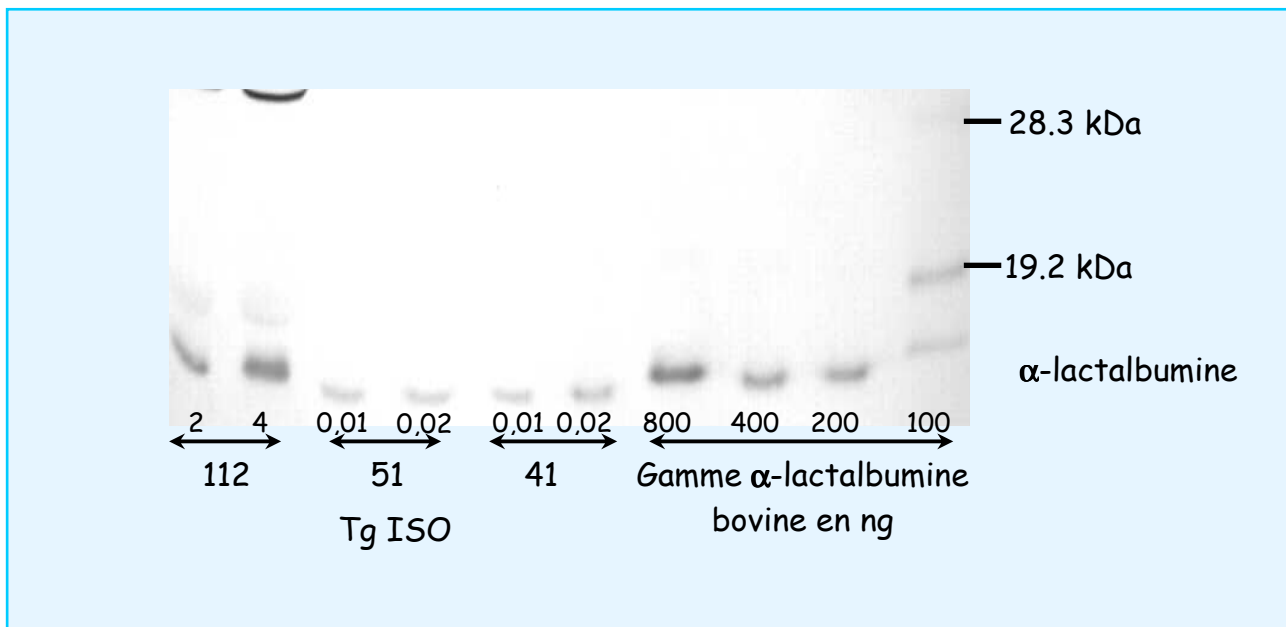


Fig. 5.9 : Analyse de la synthèse de protéine α -lactalbumine dans les lignées Tg ISO 41, 51 et 112 par Western blot.

Afin de rester dans la gamme de concentration, des volumes spécifiques de lait ont été déposés : Tg ISO 41 et 51, 0,01 et 0,02 μ l ; Tg ISO 112, 2 et 4 μ l.

Ces résultats suggèrent l'absence d'éléments *cis*-régulateurs dominants dans cette région pourtant un certains nombre de remarques sont à prendre en compte :

- trop peu de lignées de souris transgéniques exploitables ont pu être obtenus à partir de la micro-injection pour pouvoir construire une analyse correcte de l'ensemble des résultats,
- il existe une lignée de souris transgénique pour le BAC 41 (la lignée 17, Stinnakre et al., 1999) qui n'exprimait pas de niveau détectable d' α -lactalbumine sans présenter de réarrangement évident,
- la co-injection de nos deux fragments (région de 10 kb + BAC 6) ne permet de contrôler ni leur orientation ni leur localisation relative. Ainsi, une disposition différente de ces fragments par rapport au contexte chromosomique endogène pourrait interférer avec leur complémentarité.

En raison de ces résultats qui restent préliminaires et difficiles à interpréter, cette région de 10 kb ne peut pas être exclue en tant que région potentielle contenant les éléments *cis*-régulateurs dominants présents au sein du BAC 41. Des analyses complémentaires devraient être réalisées, dont certaines que nous souhaiterions suggérer dans la dernière section de cette thèse.

CONCLUSIONS
ET
PERSPECTIVES

Conclusions de ce travail de thèse

I. Choix et utilisation du BAC 41 en transgénèse

L'utilisation de promoteurs courts en transgénèse a souvent présenté des résultats non reproductibles et conduit à l'utilisation de grands fragments tels que les BAC ou les YAC. Ces vecteurs englobent fréquemment l'ensemble des éléments *cis*-régulateurs d'un gène donné et permettent ainsi une régulation similaire au gène endogène.

L'utilisation d'un BAC caprin contenant le gène de l' α -lactalbumine (BAC 41) en transgénèse a permis de s'affranchir de l'effet de position et d'obtenir une expression indépendante du site d'intégration et proportionnelle au nombre de copies intégrées (Stinnakre *et al.*, 1999). Ces expériences ont permis de mettre en évidence la présence d'éléments *cis*-régulateurs dominants capables de protéger le gène de l' α -lactalbumine de l'effet des activateurs et/ou inhibiteurs présents au voisinage du site d'insertion du BAC 41. Ces éléments présentant un fort potentiel pour la mise au point de vecteurs efficaces pour la transgénèse spécifique de la glande mammaire, une seconde série d'expériences utilisant un BAC réduit (BAC 6) a été effectuée. Les résultats montrent une dérégulation de l'expression du gène de l' α -lactalbumine avec une dépendance du site d'intégration et une perte de la proportionnalité de l'expression vis-à-vis du nombre de copies intégrées, caractéristique de la manifestation de l'effet de position. Ces expériences nous ont permis de réaliser une primo-localisation de séquences régulatrices en amont du BAC 6.

Lors de cette thèse, je me suis attaché à caractériser les éléments présents au sein du BAC 41, en amont du BAC 6, le but ultime étant d'identifier si possible les séquences *cis*-régulatrices dominantes agissant sur le promoteur du gène de l' α -lactalbumine afin de construire de meilleurs vecteurs pour la transgénèse mammaire. Ces études m'ont conduit à l'identification, l'analyse structurale et fonctionnelle de deux *loci* à expression ubiquiste. A partir de cette observation et des travaux pionniers de l'équipe du Prof. Felsenfeld (Prioleau *et al.*, 1999), l'hypothèse de l'existence de deux domaines chromatiniens au sein du BAC 41 a été émise, ce

qui nous a conduit à analyser plus finement la région frontière entre l'extrémité 3' de ces *loci* et la région 5' du BAC 6.

II. Caractérisation du BAC 41

Les deux gènes mis en évidence sont des gènes faiblement exprimés qui possèdent des promoteurs typiques de gènes à expression ubiquiste, c'est-à-dire relativement riches en îlots CpG et dépourvus de TATA box. Le gène de la cycline T1 a été relativement étudié chez l'homme et la souris du fait de l'implication de la protéine pour laquelle il code dans la réplication du VIH. Comme nous l'avons cité, certaines études ont porté sur l'analyse de son promoteur. Nos travaux suggèrent que la régulation de ce gène est plus complexe que ne semblaient indiquer les études réalisées en culture cellulaire. Sa région promotrice, certes caprine, présente une activité transcriptionnelle *in vivo* inattendue :

- elle semble fortement active dans la seule lignée germinale mâle, ce qui d'ailleurs a suscité l'intérêt de certaines équipes qui nous ont sollicités pour l'utilisation de ce promoteur
- son analyse suggère la présence d'éléments *cis*-régulateurs en 3' qui restent à identifier

Le gène FLJ20436 demeure une énigme quant à sa fonction. Il présente une régulation très complexe qui fait appel à des épissages multiples, semble-t-il conservés entre espèces. Cette conservation suggère que chaque type de transcrit doit avoir une fonction spécifique. De fait, l'analyse du rôle des diverses protéines codées par ce gène serait extrêmement intéressante.

Ces différentes études, comme nous l'avons déjà indiqué devraient sans doute faire l'objet de projets spécifiques mais n'entrent pas dans les objectifs actuels du laboratoire, ni du département de Génétique Animal.

Dans cette dernière partie, nous nous efforcerons donc plutôt de décrire ce qui pourrait être envisagé pour poursuivre la recherche des éléments *cis*-régulateurs distaux agissant sur l'expression du gène de l' α -lactalbumine.

Perspectives

I. Fin de l'analyse de la région de 10 kb

L'ensemble de ces résultats n'étant que préliminaires, différentes expériences seront nécessaires pour s'assurer que cette région de 10 kb contient ou ne contient pas d'élément *cis*-régulateur dominants tel qu'un isolateur.

1. Analyse par transfection de culture cellulaire

Bien que les analyses par transfection transitoire n'aient pas été très significatives, différentes améliorations pourraient être apportées afin d'obtenir des résultats plus satisfaisants pour analyser cette région. Le vecteur d'expression mis au point pour ce travail demande quelques modifications afin d'acquérir une stimulation plus conséquente du gène rapporteur et de pouvoir observer des résultats plus tranchés lors de l'utilisation des autres vecteurs. En parallèle à ces transfusions transitoires, des transfusions stables devraient être envisagées. Pour des questions de délai, ce type d'expériences n'a pas pu être mis en place au cours de cette thèse. Ces expériences devront être réalisées sur un type cellulaire spécifique de la glande mammaire, comme les cellules HC11 déjà utilisées au laboratoire, au sein duquel pourra être reproduit différentes stimulations propres à cet organe. On ne peut en effet exclure que les phénomènes mis en place, au sein d'une région frontière active dans la glande mammaire à certains stades physiologiques, pourraient nécessiter la présence d'un complexe protéique dont certains constituants seraient spécifiques de l'épithélium mammaire, comme cela a été observé lors de la caractérisation de la séquence 5' HS5 humaine (Wai *et al.*, 2003). L'emploi d'autres types cellulaires, telles les cellules CHO utilisées lors de cette thèse, pourrait empêcher des interactions nécessaires à la fonction des isolateurs et nous induire en erreur lors de l'interprétation des résultats.

2. Analyse par transgénèse chez la souris

Les problèmes rencontrés lors de la réalisation des constructions pour la micro-injection, les remaniements observés dans les rares lignées obtenues ne nous ont pas permis d'analyser les résultats de ces expériences dans des conditions optimales. Afin de pallier à ce problème, la mise au point de constructions respectant le contexte chromosomique, c'est-à-dire aussi bien

les distances relatives que l'orientation de la(les) région(s) frontière(s) suspectée(s) par rapport au BAC 6, semble indispensable. Malgré les difficultés rencontrées pour le clonage de la région de 10 kb, l'amélioration permanente des techniques de recombinaison homologue (Yang *et al.*, 1997 ; Narayanan *et al.*, 1999 ; Muyrers *et al.*, 1999 ; Lee *et al.*, 2001) et les contacts établis avec les équipes spécialisées dans ce domaine devraient nous permettre de réaliser des constructions qui pourront apporter des renseignements vis-à-vis des éléments régulateurs présents dans cette région.

3. Analyse de sensibilité à la DNase

En parallèle à ces tests, la recherche de sites hypersensibles à la DNase I devrait être initiée car la présence de tels sites est souvent caractéristique de l'existence de régions impliquées dans régulation de l'expression des gènes. Une telle analyse permettrait de mieux délimiter les régions d'intérêt et de conforter ou à l'inverse d'infirmer les résultats obtenus *in silico* ou en culture cellulaire.

4. Test de la fonctionnalité des MAR prédites

A la suite de la prédiction de la présence de deux MAR au sein de cette région, différentes expériences devraient être effectuées afin de vérifier si ces prédictions se révèlent réelles. Une MAR étant une zone d'attachement où la matrice nucléaire peut se fixer sur l'ADN, des tests d'immuno-précipitation de la chromatine devraient nous permettre de montrer si des protéines sont liées au niveau de ces séquences.

La présence d'une MAR fonctionnelle au niveau de cette région potentiellement frontière nous placerait dans un contexte comparable à celui du domaine chromatinien contenant un autre gène majeur des protéines du lait, le gène de la WAP de lapin. En effet, une région d'attachement, très conservée chez la souris, a pu être prédite à 8 kb en amont de ce gène (Milot *et al.*, 2003). Même si aucune propriété relative aux isolateurs n'a été associée à cette MAR, il semblerait qu'il existe un premier niveau d'organisation de la chromatine qui n'interférerait pas avec une régulation par des éléments localisés encore plus en amont (Rival-Gervier^{a et b} *et al.*, 2003).

II. Remise en cause de cette région de 10 kb

1. Cas du *locus* du lysozyme du poulet

Les hypothèses de départ s'appuyaient sur la mise en évidence de domaines chromatiniens regroupant un ensemble de gènes dont la régulation transcriptionnelle est effectuée de la même façon (Prioleau *et al.*, 1999). Les gènes caractérisés au sein du BAC 41 ayant des modes d'expression différents de celui du gène de l' α -lactalbumine, l'hypothèse de l'existence de deux domaines putatifs avec une région frontière les séparant et permettant une régulation indépendante des différents gènes a été émise et plusieurs expériences ont été réalisées pour localiser et caractériser cette région. Cependant, il existe des exemples dans la littérature de domaines chromatiniens contenant des gènes avec des modes d'expression différents comme c'est le cas pour le domaine du gène du lysozyme du poulet (Chong *et al.*, 2002) (**Fig. C&P.1**). Le gène du lysozyme est spécifique de l'oviducte et des cellules myéloïdes chez le poulet avec un niveau d'expression en constante augmentation pendant la phase de différenciation des macrophages (Bonifer *et al.*, 1994). Son domaine chromatinien a été un des plus anciens caractérisé et est relativement court (21,5 kb), délimité par deux séquences MARS (Loc and Stratling, 1988 ; Stief *et al.*, 1989). Ce gène possède un mode de régulation très spécifique nécessitant un contrôle transcriptionnel important. Pourtant, seulement 200 pb en aval de ce gène a été identifié un second gène disposant d'un mode d'expression de type ubiquiste : le gène Gas41.

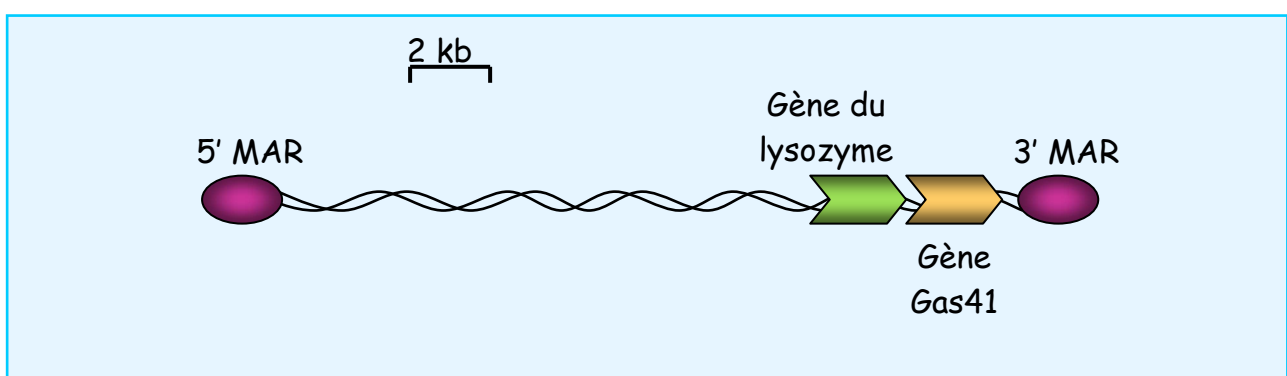


Fig. C&P.1 : Schéma du *locus* du gène du lysozyme du poulet (Chong *et al.*, 2002).

Ce gène fait donc parti du même domaine chromatinien et ce cas ne semble pas isolé. En effet, l'étude en transgénèse du *locus* WAP de porc, un gène spécifiquement exprimé dans la glande mammaire, suggère que des gènes, également à expression ubiquiste, sont présents dans le même domaine chromatinien (Rival-Gervier^{a et b} *et al.*, 2003).

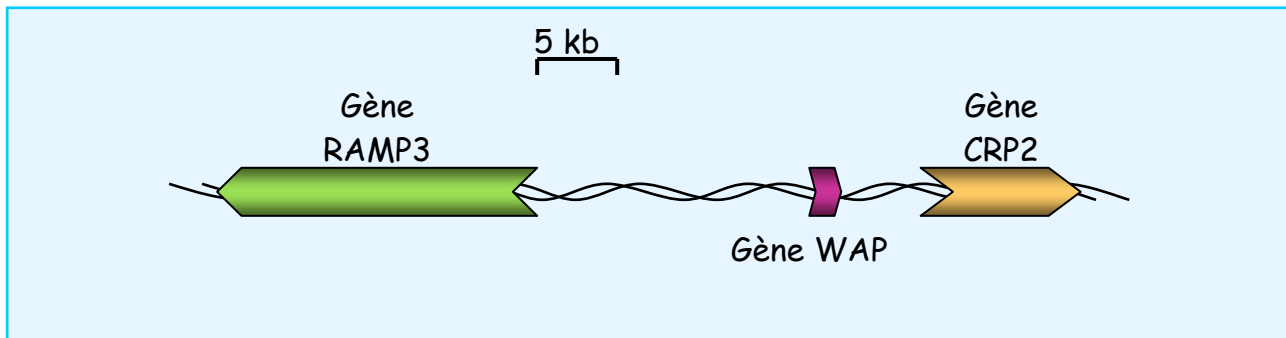


Fig. C&P.2 : Schéma du *locus* du gène WAP de porc (Rival-Gervier^{a et b} *et al.*, 2003). Il est à noter que dans ces deux cas, les gènes ubiquistes présentent un promoteur dépourvu de TATA box. En comparant ces exemples à la situation présente au sein du BAC 41 et en reprenant des hypothèses récemment émises sur la possible absence de compétition pour des activateurs entre des promoteurs avec ou sans TATA box (Levin and Tjian, 2003), il faudrait peut-être redéfinir les limites du domaine chromatinien du gène de l' α -lactalbumine pour y inclure un des deux voir les deux gènes ubiquistes en aval du BAC 6.

2. Approches différentes pour la transgénèse chez la souris

Si la région de 10 kb n'est pas suffisante pour contenir les éléments *cis*-régulateurs dominants, différentes nouvelles constructions devraient être envisagées. Les gènes ubiquistes en aval du BAC 6 étant potentiellement sur le même domaine chromatinien que le gène de l' α -lactalbumine, les constructions à réaliser devront associer au BAC 6 des régions en aval de plus en plus grandes. La région de 10 kb correspondant à l'extrémité 3' du *locus* FLJ20436, une deuxième construction devrait contenir l'ensemble de ce *locus* ainsi que la région de 5 kb entre les *loci* de la cycline T1 et FLJ 20436, c'est-à-dire BAC 6 + région de 30 kb en aval. Enfin une troisième construction pourrait être envisagée, BAC 6 + région de 60 à 70 kb en aval afin d'inclure l'unité de transcription du gène de la cycline T1 sachant que sa région promotrice a déjà été testée en transgénèse sans résultat concluant vis-à-vis de la mise en évidence des éléments *cis*-régulateurs (schéma constructions potentielles).

Une alternative à la réalisation de ces constructions, qui demanderaient l'utilisation des techniques de recombinaison homologe chez *E. coli* dont nous avons expérimenté au cours

de cette thèse la facilité de mise en œuvre..., serait le criblage de banques de BAC d'espèces phylogénétiquement proches de la chèvre telles que les bovins ou les ovins. Ce criblage pourrait permettre de trouver des BAC dont les inserts correspondraient aux spécifications des différentes constructions envisagées. Une telle approche a récemment permis l'analyse fine du *locus Gdf6* (Mortlock *et al.*, 2003).

BIBLIOGRAPHIE

- Al-Shawi R, Kinnaird J, Burke J and Bishop JO.** Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol Cell Biol.* 1990, 10 (3), 1192-8.
- Antes TJ, Namciu SJ, Fournier RE and Levy-Wilson B.** The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. *Biochemistry.* 2001, 40 (23), 6731-42.
- Arney KL.** H19 and Igf2--enhancing the confusion? *Trends Genet.* 2003, 19 (1), 17-23. Review.
- Baguisi A, Behboodi E, Melican DT, Pollock JS, Destrempe MM, Cammuso C, Williams JL, Nims SD, Porter CA, Midura P, Palacios MJ, Ayres SL, Denniston RS, Hayes ML, Ziomek CA, Meade HM, Godke RA, Gavin WG, Overstrom EW and Echelard Y.** Production of goats by somatic cell nuclear transfer. *Nat Biotechnol.* 1999, 17 (5), 456-61.
- Ball RK, Friis RR, Schoenenberger CA, Doppler W and Groner B.** Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* 1988, 7 (7), 2089-95.
- Behrens A, Genoud N, Naumann H, Rulicke T, Janett F, Heppner FL, Ledermann B and Aguzzi A.** Absence of the prion protein homologue Doppel causes male sterility. *EMBO J.* 2002 Jul 15;21(14):3652-8.
- Bell AC, West AG and Felsenfeld G.** The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell.* 1999, 98, (3), 387-96.
- Bethhauser J, Forsberg E, Augenstein M, Childs L, Eilertsen K, Enos J, Forsythe T, Golueke P, Jurgella G, Koppang R, Lesmeister T, Mallon K, Mell G, Misica P, Pace M, Pfister-Genskow M, Strelchenko N, Voelker G, Watt S, Thompson S and Bishop M.** Production of cloned pigs from in vitro systems. *Nat Biotechnol.* 2000, 18 (10), 1055-9.
- Bieniasz PD, Grdina TA, Bogerd HP and Cullen BR.** Recruitment of a protein complex containing Tat and cyclin T1 to TAR governs the species specificity of HIV-1 Tat. *EMBO J.* 1998, 17 (23), 7056-65.
- Bishop JO and Smith P.** Mechanism of chromosomal integration of microinjected DNA. *Mol Biol Med.* 1989, 6 (4), 283-98. Review.
- Bonifer C, Vidal M, Grosveld F and Sippel AE.** Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.* 1990, 9 (9), 2843-8.

- Bonifer C, Bosch FX, Faust N, Schuhmann A and Sippel AE.** Evolution of gene regulation as revealed by differential regulation of the chicken lysozyme transgene and the endogenous mouse lysozyme gene in mouse macrophages. *Eur J Biochem.* 1994, 226 (1), 227-35.
- Bosze Z, Hiripi L, Carnwath JW and Niemann H.** The transgenic rabbit as model for human diseases and as a source of biologically active recombinant proteins. *Transgenic Res.* 2003, 12 (5), 541-53.
- Brem G, Besenfelder U, Aigner B, Muller M, Liebl I, Schutz G and Montoliu L.** YAC transgenesis in farm animals: rescue of albinism in rabbits. *Mol Reprod Dev.* 1996, 44 (1), 56-62.
- Brinster RL, Sandgren EP, Behringer RR and Palmiter RD.** No simple solution for making transgenic mice. *Cell.* 1989, 59 (2), 239-41.
- Brinster RL.** Germline stem cell transplantation and transgenesis. *Science.* 2002, 296 (5576), 2174-6. Review.
- Brophy B, Smolenski G, Wheeler T, Wells D, L'Huillier P and Laible G.** Cloned transgenic cattle produce milk with higher levels of beta-casein and kappa-casein. *Nat Biotechnol.* 2003, 21 (2), 157-62.
- Campbell KH, McWhir J, Ritchie WA and Wilmut I.** Sheep cloned by nuclear transfer from a cultured cell line. *Nature.* 1996, 380 (6569), 64-6.
- Capecchi MR.** Altering the genome by homologous recombination. *Science.* 1989, 244 (4910), 1288-92. Review.
- Castilla J, Pintado B, Sola I, Sanchez-Morgado JM and Enjuanes L.** Engineering passive immunity in transgenic mice secreting virus-neutralizing antibodies in milk. *Nat Biotechnol.* 1998, 16 (4), 349-54.
- Chan AW, Dominko T, Luetjens CM, Neuber E, Martinovich C, Hewitson L, Simerly CR and Schatten GP.** Clonal propagation of primate offspring by embryo splitting. *Science.* 2000, 287 (5451), 317-9.
- Chan AW, Chong KY, Martinovich C, Simerly C and Schatten G.** Transgenic monkeys produced by retroviral gene transfer into mature oocytes. *Science.* 2001, 291 (5502), 309-12.
- Chesné P, Adenot PG, Viglietta C, Baratte M, Boulanger L and Renard JP.** Cloned rabbits produced by nuclear transfer from adult somatic cells. *Nat Biotechnol.* 2002, 20 (4), 366-9.

- Chong S, Riggs AD and Bonifer C.** A Functional chromatin domain does not resist X chromosome inactivation: silencing of cLys correlates with methylation of a dual promoter-replication origin. *Mol Cell Biol.* 2002, 22 (13), 4667-76.
- Chung JH, Whiteley M and Felsenfeld G.** A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell.* 1993, 74 (3), 505-14.
- Chung JH, Bell AC and Felsenfeld G.** Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A.* 1997, 94 (2), 575-80.
- Church GM and Gilbert W.** Genomic sequencing. *Proc Natl Acad Sci U S A.* 1984, 81 (7), 1991-5.
- Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce de Leon FA and Robl JM.** Transgenic bovine chimeric offspring produced from somatic cell-derived stem-like cells. *Nat Biotechnol.* 1998, 16 (7), 642-6.
- Clark AJ, Harold G and Yull FE.** Mammalian cDNA and prokaryotic reporter sequences silence adjacent transgenes in transgenic mice. *Nucleic Acids Res.* 1997, 25 (5), 1009-14.
- Collinge J and Hawke S.** B lymphocytes in prion neuroinvasion: central or peripheral players? *Nat Med.* 1998, 4 (12), 1369-70.
- Collins S, McLean CA and Masters CL.** Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, and kuru: a review of these less common human transmissible spongiform encephalopathies. *J Clin Neurosci.* 2001, 8 (5), 387-97. Review.
- Copeland NG, Jenkins NA and Court DL.** Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet.* 2001, 2 (10), 769-79. Review.
- Cranston A, Dong C, Howcroft J and Clark AJ.** Chromosomal sequences flanking an efficiently expressed transgene dramatically enhance its expression. *Gene.* 2001, 269 (1-2), 217-25.
- Cuvier O, Hart CM and Laemmli UK.** Identification of a class of chromatin boundary elements. *Mol Cell Biol.* 1998, 18 (12), 7478-86.
- Danielson KG, Oborn CJ, Durban EM, Butel JS and Medina D.** Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. *Proc Natl Acad Sci U S A.* 1984, 81 (12), 3756-60.
- De Luca A, De Falco M, Baldi A and Paggi MG.** Cyclin T: three forms for different roles in physiological and pathological functions. *J Cell Physiol.* 2003, 194 (2), 101-7. Review.

- De Wet JR, Wood KV, DeLuca M, Helinski DR and Subramani S.** Firefly luciferase gene: structure and expression in mammalian cells. *Mol Cell Biol.* 1987, 7 (2), 725-37.
- Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E and Reik W.** Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc Natl Acad Sci U S A.* 2001, 98 (24), 13734-8.
- Denamur R.** Reviews of the progress of dairy science. Section A. Physiology. Hormonal control of lactogenesis. *J Dairy Res.* 1971, 38 (2), 237-64. Review.
- Denning C, Burl S, Ainslie A, Bracken J, Dinnyes A, Fletcher J, King T, Ritchie M, Ritchie WA, Rollo M, de Sousa P, Travers A, Wilmot I and Clark AJ.** Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. *Nat Biotechnol.* 2001, 19 (6), 559-62.
- Dobie KW, Lee M, Fantes JA, Graham E, Clark AJ, Springbett A, Lathe R and McClenaghan M** Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc Natl Acad Sci U S A.* 1996, 93 (13), 6659-64.
- Dobson DE, Prager EM and Wilson AC.** Stomach lysozymes of ruminants. I. Distribution and catalytic properties. *J Biol Chem.* 1984, 259 (18), 11607-16.
- Dormont D.** Agents that cause transmissible subacute spongiform encephalopathies. *Biomed Pharmacother.* 1999, 53 (1), 3-8. Review.
- Du F, Giles JR, Foote RH, Graves KH, Yang X and Moreadith RW.** Nuclear transfer of putative rabbit embryonic stem cells leads to normal blastocyst development. *J Reprod Fertil.* 1995, 104 (2), 219-23.
- Dunaway M, Hwang JY, Xiong M and Yuen HL.** The activity of the scs and scs' insulator elements is not dependent on chromosomal context. *Mol Cell Biol.* 1997, 17 (1), 182-9.
- Ebner K and Brodbeck U.** Biological role of alpha-lactalbumin: a review. *J Dairy Sci.* 1968, 51 (3), 317-22. Review.
- Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A, Zanghellini A, Paques F and Lacroix E.** A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* 2003, 31 (11), 2952-62.
- Evans MJ and Kaufman MH.** Establishment in culture of pluripotential cells from mouse embryos. *Nature.* 1981, 292 (5819), 154-6.
- Ewing B, Hillier L, Wendl MC and Green P.** Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998, 8 (3), 175-85.

- Ewing B and Green P.** Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998, 8 (3), 186-94.
- Farrell CM, West AG and Felsenfeld G.** Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol.* 2002, 22 (11), 3820-31.
- First NL, Sims MM, Park SP and Kent-First MJ.** Systems for production of calves from cultured bovine embryonic cells. *Reprod Fertil Dev.* 1994, 6 (5), 553-62. Review.
- Flavell RB.** Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci U S A.* 1994, 91 (9), 3490-6. Review.
- Ford MJ, Burton LJ, Morris RJ and Hall SM.** Selective expression of prion protein in peripheral tissues of the adult mouse. *Neuroscience.* 2002, 113 (1), 177-92.
- Forrester WC, Fernandez LA and Grosschedl R.** Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev.* 1999, 13 (22), 3003-14.
- Fujiwara^a Y, Miwa M, Takahashi R, Kodaira K, Hirabayashi M, Suzuki T and Ueda M.** High-level expressing YAC vector for transgenic animal bioreactors. *Mol Reprod Dev.* 1999, 52 (4), 414-20.
- Fujiwara^b Y, Takahashi RI, Miwa M, Kameda M, Kodaira K, Hirabayashi M, Suzuki T and Ueda M.** Analysis of control elements for position-independent expression of human alpha-lactalbumin YAC. *Mol Reprod Dev.* 1999, 54 (1), 17-23.
- Fujiwara Y, Takahashi R, Hirabayashi M, Ueda M, Muramatsu T, Yamanaka H and Sekikawa K.** Analysis of the flanking regions of the human alpha-lactalbumin gene responsible for position-effect independent expression. *Gene.* 2003, 305 (1), 71-8.
- Gallagher DS Jr, Threadgill DW, Ryan AM, Womack JE and Irwin DM.** Physical mapping of the lysozyme gene family in cattle. *Mamm Genome.* 1993, 4 (7), 368-73.
- Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, Rice AP, Littman DR and Jones KA.** The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* 1998, 12 (22), 3512-27.
- Garrick D, Fiering S, Martin DI and Whitelaw E.** Repeat-induced gene silencing in mammals. *Nat Genet.* 1998, 18 (1), 56-9.
- Gatti JL, Metayer S, Moudjou M, Andreoletti O, Lantier F, Dacheux JL and Sarradin P.** Prion protein is secreted in soluble forms in the epididymal fluid and proteolytically processed and transported in seminal plasma. *Biol Reprod.* 2002, 67 (2), 393-400.

- Gdula DA, Gerasimova TI and Corces VG.** Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc Natl Acad Sci U S A*. 1996, 93 (18), 9378-83. Review.
- Gerasimova TI, Gdula DA, Gerasimov DV, Simonova O and Corces VG.** A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell*. 1995, 82 (4), 587-97.
- Gerasimova TI and Corces VG.** Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*. 1998, 92 (4), 511-21.
- Giraldo P and Montoliu L.** Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res*. 2001, 10 (2), 83-103. Review.
- Golovan SP, Meidinger RG, Ajakaiye A, Cottrill M, Wiederkehr MZ, Barney DJ, Plante C, Pollard JW, Fan MZ, Hayes MA, Laursen J, Hjorth JP, Hacker RR, Phillips JP and Forsberg CW.** Pigs expressing salivary phytase produce low-phosphorus manure. *Nat Biotechnol*. 2001, 19 (8), 741-5.
- Gordon D, Abajian C and Green P.** Consed: a graphical tool for sequence finishing. *Genome Res*. 1998, 8 (3), 195-202.
- Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA and Ruddle FH.** Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A*. 1980, 77 (12), 7380-4.
- Gossler A, Doetschman T, Korn R, Serfling E and Kemler R.** Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc Natl Acad Sci U S A*. 1986, 83 (23), 9065-9.
- Graveley BR.** Alternative splicing: increasing diversity in the proteomic world. *Trends Genet*. 2001, 17 (2), 100-7. Review.
- Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M and Snell R.** Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res*. 2002, 12 (2), 222-31.
- Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Menissier F, Massabanda J, Fries R, Hanset R and Georges M.** A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet*. 1997, 17 (1), 71-4.

- Grosveld F, van Assendelft GB, Greaves DR and Kollias G.** Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*. 1987, 51 (6), 975-85.
- Grosveld F.** Activation by locus control regions? *Curr Opin Genet Dev*. 1999, 9 (2), 152-7. Review.
- Gurdon JB and Byrne JA.** The first half-century of nuclear transplantation. *Proc Natl Acad Sci U S A*. 2003, 100 (14), 8048-52.
- Hamada T, Sasaki H, Seki R and Sakaki Y.** Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: sequence analysis of genome-transgene and transgene-transgene junctions at two loci. *Gene*. 1993, 128 (2), 197-202.
- Hanson RW.** New animal models for study of metabolism minireview series. *J Biol Chem*. 2003, 278 (31), 28357-8. Review.
- Harvey AJ, Speksnijder G, Baugh LR, Morris JA and Ivarie R.** Expression of exogenous protein in the egg white of transgenic chickens. *Nat Biotechnol*. 2002, 20 (4), 396-9.
- Hayssen V and Blackburn DG.** Alpha-lactalbumin and the origins of lactation. *Evolution*. 1985, 39, 1147-49.
- Hochi S, Ninomiya T, Waga-Homma M, Sagara J and Yuki A.** Secretion of bovine alpha-lactalbumin into the milk of transgenic rats. *Mol Reprod Dev*. 1992, 33 (2), 160-4.
- Hogan B, Constantini F and Lacy E.** In manipulating the mouse embryo: A laboratory manual, first edition. Cold Spring Harbor Laboratory Press, 1986.
- Houdebine LM.** Transgenic animal bioreactors. *Transgenic Res*. 2000, 9 (4-5), 305-20. Review.
- Houdebine LM.** Antibody manufacture in transgenic animals and comparisons with other systems. *Curr Opin Biotechnol*. 2002, 13 (6), 625-9. Review.
- Humpherys D, Eggan K, Akutsu H, Hochedlinger K, Rideout WM 3rd, Biniszkiewicz D, Yanagimachi R and Jaenisch R.** Epigenetic instability in ES cells and cloned mice. *Science*. 2001, 293 (5527), 95-7.
- Ilan N, Barash I, Gootwine E and Shani M.** Establishment and initial characterization of the ovine mammary epithelial cell line NISH. *In Vitro Cell Dev Biol Anim*. 1998, 34 (4), 326-32.
- Inoue K, Kohda T, Lee J, Ogonuki N, Mochida K, Noguchi Y, Tanemura K, Kaneko-Ishino T, Ishino F and Ogura A.** Faithful expression of imprinted genes in cloned mice. *Science*. 2002, 295 (5553), 297.

- Irwin DM and Wilson AC.** Multiple cDNA sequences and the evolution of bovine stomach lysozyme. *J Biol Chem.* 1989, 264 (19), 11387-93.
- Irwin DM and Wilson AC.** Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. *J Biol Chem.* 1990, 265 (9), 4944-52.
- Jeanisch R, Eggan K, Humpherys D, Rideout W and Hochedlinger K.** Nuclear cloning, stem cells, and genomic reprogramming. *Cloning Stem Cells.* 2002, 4 (4), 389-96. Review.
- Jolles P, Schoentgen F, Jolles J, Dobson DE, Prager EM and Wilson AC.** Stomach lysozymes of ruminants. II. Amino acid sequence of cow lysozyme 2 and immunological comparisons with other lysozymes. *J Biol Chem.* 1984, 259 (18), 11617-25.
- Jost B, Vilotte JL, Duluc I, Rodeau JL and Freund JN.** Production of low-lactose milk by ectopic expression of intestinal lactase in the mouse mammary gland. *Nat Biotechnol.* 1999, 17 (2), 160-4.
- Karatzas CN.** Designer milk from transgenic clones. *Nat Biotechnol.* 2003, 21 (2), 138-9.
- Keefer CL, Baldassarre H, Keyston R, Wang B, Bhatia B, Bilodeau AS, Zhou JF, Leduc M, Downey BR, Lazaris A and Karatzas CN.** Generation of dwarf goat (*Capra hircus*) clones following nuclear transfer with transfected and nontransfected fetal fibroblasts and in vitro-matured oocytes. *Biol Reprod.* 2001, 64 (3), 849-56.
- Kellum R and Schedl P.** A position-effect assay for boundaries of higher order chromosomal domains. *Cell.* 1991, 64 (5), 941-50.
- Kellum R and Schedl P.** A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol.* 1992, 12 (5), 2424-31.
- Kerr DE, Liang F, Bondioli KR, Zhao H, Kreibich G, Wall RJ and Sun TT.** The bladder as a bioreactor: urothelium production and secretion of growth hormone into urine. *Nat Biotechnol.* 1998, 16 (1), 75-9.
- Kerr DE, Plaut K, Bramley AJ, Williamson CM, Lax AJ, Moore K, Wells KD and Wall RJ.** Lysostaphin expression in mammary glands confers protection against staphylococcal infection in transgenic mice. *Nat Biotechnol.* 2001, 19 (1), 66-70.
- Kisseberth WC, Brettingen NT, Lohse JK and Sandgren EP.** Ubiquitous expression of marker transgenes in mice and rats. *Dev Biol.* 1999, 214 (1), 128-38.

- Klein MA, Frigg R, Flechsig E, Raeber AJ, Kalinke U, Bluethmann H, Bootz F, Suter M, Zinkernagel RM and Aguzzi A.** A crucial role for B cells in neuroinvasive scrapie. *Nature*. 1997, 390 (6661), 687-90.
- Klein MA, Frigg R, Raeber AJ, Flechsig E, Hegyi I, Zinkernagel RM, Weissmann C, Aguzzi A.** PrP expression in B lymphocytes is not required for prion neuroinvasion. *Nat Med*. 1998, 4 (12), 1429-33.
- Knight R.** Creutzfeldt-Jakob disease: a protein disease. *Proteomics*. 2001, 1 (6), 763-6. Review.
- Kolb AF, Pewe L, Webster J, Perlman S, Whitelaw CB and Siddell SG.** Virus-neutralizing monoclonal antibody expressed in milk of transgenic mice provides full protection against virus-induced encephalitis. *J Virol*. 2001, 75 (6), 2803-9.
- Kordower JH, Emborg ME, Bloch J, Ma SY, Chu Y, Leventhal L, McBride J, Chen EY, Palfi S, Roitberg BZ, Brown WD, Holden JE, Pyzalski R, Taylor MD, Carvey P, Ling Z, Trono D, Hantraye P, Deglon N and Aebischer P.** Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease. *Science*. 2000, 290 (5492), 767-73.
- Krnacik MJ, Li S, Liao J and Rosen JM.** Position-independent expression of whey acidic protein transgenes. *J Biol Chem*. 1995, 270 (19), 11119-29.
- Kuersten S and Goodwin EB.** The power of the 3' UTR: translational control and development. *Nat Rev Genet*. 2003, 4 (8), 626-37. Review.
- Kuhn EJ and Geyer PK.** Genomic insulators: connecting properties to mechanism. *Curr Opin Cell Biol*. 2003, 15 (3), 259-65. Review.
- Kuroiwa Y, Tomizuka K, Shinohara T, Kazuki Y, Yoshida H, Ohguma A, Yamamoto T, Tanaka S, Oshimura M and Ishida I.** Manipulation of human minichromosomes to carry greater than megabase-sized chromosome inserts. *Nat Biotechnol*. 2000, 18 (10), 1086-90.
- Kuroiwa Y, Kasinathan P, Choi YJ, Naeem R, Tomizuka K, Sullivan EJ, Knott JG, Duteau A, Goldsby RA, Osborne BA, Ishida I and Robl JM.** Cloned transchromosomal calves producing human immunoglobulin. *Nat Biotechnol*. 2002, 20 (9), 889-94.
- Lander ES, ... and Chen YJ; International Human Genome Sequencing Consortium.** Initial sequencing and analysis of the human genome. *Nature*. 2001, 409 (6822), 860-921.

- Larrick JW and Thomas DW.** Producing proteins in transgenic plants and animals. *Curr Opin Biotechnol.* 2001, 12 (4), 411-8. Review.
- Lavitrano M, Camaioni A, Fazio VM, Dolci S, Farace MG and Spadafora C.** Sperm cells as vectors for introducing foreign DNA into eggs: genetic transformation of mice. *Cell.* 1989, 57 (5), 717-23.
- Lavitrano M, Bacci ML, Forni M, Lazzereschi D, Di Stefano C, Fioretti D, Giancotti P, Marfé G, Pucci L, Renzi L, Wang H, Stoppacciaro A, Stassi G, Sargiacomo M, Sinibaldi P, Turchi V, Giovannoni R, Della Casa G, Seren E and Rossi G.** Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation. *Proc Natl Acad Sci U S A.* 2002, 99 (22), 14230-35.
- Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA and Copeland NG.** A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics.* 2001, 73 (1), 56-65.
- Levine M and Tjian R.** Transcription regulation and animal diversity. *Nature.* 2003, 424 (6945), 147-51. Review.
- Liu H and Rice AP.** Isolation and characterization of the human cyclin T1 promoter. *Gene.* 2000, 252 (1-2), 39-49.
- Loc PV and Stratling WH.** The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *EMBO J.* 1988, 7 (3), 655-64.
- Maga EA, Anderson GB, Huang MC and Murray JD.** Expression of human lysozyme mRNA in the mammary gland of transgenic mice. *Transgenic Res.* 1994, 3 (1), 36-42.
- Maga EA, Sargent RG, Zeng H, Pati S, Zarling DA, Oppenheim SM, Collette NM, Moyer AL, Conrad-Brink JS, Rowe JD, BonDurant RH, Anderson GB and Murray JD.** Increased efficiency of transgenic livestock production. *Transgenic Res.* 2003, 12 (4), 485-96.
- Maione B, Lavitrano M, Spadafora C and Kiessling AA.** Sperm-mediated gene transfer in mice. *Mol Reprod Dev.* 1998, 50 (4), 406-9.
- Martin GR.** Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A.* 1981, 78 (12), 7634-8.

- Martin-Serrano J, Li K and Bieniasz PD.** Cyclin T1 expression is mediated by a complex and constitutively active promoter and does not limit human immunodeficiency virus type 1 Tat function in unstimulated primary lymphocytes. *J Virol.* 2002, 76 (1), 208-19.
- Maschio A, Brickell PM, Kioussis D, Mellor AL, Katz D and Craig RK.** Transgenic mice carrying the guinea-pig alpha-lactalbumin gene transcribe milk protein genes in their sebaceous glands during lactation. *Biochem J.* 1991, 275 (Pt 2), 459-67.
- Massoud M, Bischoff R, Dalemans W, Pointu H, Attal J, Schultz H, Clesse D, Stinnakre MG, Pavirani A and Houdebine LM.** Production of human proteins in the blood of transgenic animals. *C R Acad Sci III.* 1990, 311 (8), 275-80.
- McCreath KJ, Howcroft J, Campbell KH, Colman A, Schnieke AE and Kind AJ.** Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature.* 2000, 405 (6790), 1066-9.
- McKnight RA, Shamay A, Sankaran L, Wall RJ and Hennighausen L.** Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proc Natl Acad Sci U S A.* 1992, 89 (15), 6943-7.
- McKnight RA, Spencer M, Wall RJ and Hennighausen L.** Severe position effects imposed on a 1 kb mouse whey acidic protein gene promoter are overcome by heterologous matrix attachment regions. *Mol Reprod Dev.* 1996, 44 (2), 179-84.
- Meisler MH.** Insertional mutation of 'classical' and novel genes in transgenic mice. *Trends Genet.* 1992, 8 (10), 341-4. Review.
- Melfi R, Palla F, Di Simone P, Alessandro C, Cali L, Anello L and Spinelli G.** Functional characterization of the enhancer blocking element of the sea urchin early histone gene cluster reveals insulator properties and three essential *cis*-acting sequences. *J Mol Biol.* 2000, 304 (5), 753-63.
- Mercier JC and Vilotte JL.** Structure and function of milk protein genes. *J Dairy Sci.* 1993 Oct;76(10):3079-98. Review.
- Mercier JC and Vilotte JL.** The modification of milk protein composition through transgenesis: progress and problems. In Houdebine LM (Ed), *Transgenic animals-generation and use.* Harwood academic publishers, Switzerland, 1997, 473-82.
- Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundstrom K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P and Andersson L.** A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science.* 2000, 288 (5469), 1248-51.

- Millot B, Montoliu L, Fontaine ML, Mata T and Devinoy E.** Hormone-induced modifications of the chromatin structure surrounding upstream regulatory regions conserved between the mouse and rabbit whey acidic protein genes. *Biochem J.* 2003, 15, (372), 41-52.
- Miyagawa S, Murakami H, Takahagi Y, Nakai R, Yamada M, Murase A, Koyota S, Koma M, Matsunami K, Fukuta D, Fujimura T, Shigehisa T, Okabe M, Nagashima H, Shirakura R and Taniguchi N.** Remodeling of the major pig xenoantigen by N-acetylglucosaminyltransferase III in transgenic pig. *J Biol Chem.* 2001, 276 (42), 39310-9.
- Modin C, Pedersen FS and Duch M.** Lack of shielding of primer binding site silencer-mediated repression of an internal promoter in a retrovirus vector by the putative insulators scs, BEAD-1, and HS4. *J Virol.* 2000, 74 (24), 11697-707.
- Moore RC, Lee IY, Silverman GL, Harrison PM, Strome R, Heinrich C, Karunaratne A, Pasternak SH, Chishti MA, Liang Y, Mastrangelo P, Wang K, Smit AF, Katamine S, Carlson GA, Cohen FE, Prusiner SB, Melton DW, Tremblay P, Hood LE and Westaway D.** Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. *J Mol Biol.* 1999, 292 (4), 797-817.
- Mortlock DP, Guenther C and Kingsley DM.** A general approach for identifying distant regulatory elements applied to the *gdf6* gene. *Genome Res.* 2003, 13 (9), 2069-81.
- Muller U.** Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev.* 1999, 82 (1-2), 3-21. Review.
- Mulsant P, Lecerf F, Fabre S, Schibler L, Monget P, Lanneluc I, Pisselet C, Riquet J, Monniaux D, Callebaut I, Crihiu E, Thimonier J, Teyssier J, Bodin L, Cognie Y, Chitour N and Elsen JM.** Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. *Proc Natl Acad Sci U S A.* 2001, 98 (9), 5104-9.
- Muyrers JP, Zhang Y, Testa G and Stewart AF.** Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* 1999, 27 (6), 1555-7.
- Nagano M, Watson DJ, Ryu BY, Wolfe JH and Brinster RL.** Lentiviral vector transduction of male germ line stem cells in mice. *FEBS Lett.* 2002, 524 (1-3), 111-5.
- Narayanan K, Williamson R, Zhang Y, Stewart AF and Ioannou PA.** Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in *E. coli* DH10B using an inducible homologous recombination system. *Gene Ther.* 1999, 6 (3), 442-7.

- Newell-Price J, Clark AJ and King P.** DNA methylation and silencing of gene expression. *Trends Endocrinol Metab.* 2000, 11 (4), 142-8. Review.
- Ninomiya T, Hirabayashi M, Sagara J and Yuki A.** Functions of milk protein gene 5' flanking regions on human growth hormone gene. *Mol Reprod Dev.* 1994, 37 (3), 276-83.
- Notarianni E, Laurie S, NG A and Sathasivam K.** Incorporation of cultured embryonic cells into transgenic and chimeric, porcine fetuses. *Int J Dev Biol.* 1997, 41 (3), 537-40.
- Ogata T, Kozuka T and Kanda T.** Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *J Virol.* 2003, 77 (16), 9000-7.
- Onishi A, Iwamoto M, Akita T, Mikawa S, Takeda K, Awata T, Hanada H and Perry AC.** Pig cloning by microinjection of fetal fibroblast nuclei. *Science.* 2000, 289 (5482), 1188-90.
- Palla F, Melfi R, Anello L, Di Bernardo M and Spinelli G.** Enhancer blocking activity located near the 3' end of the sea urchin early H2A histone gene. *Proc Natl Acad Sci U S A.* 1997, 94 (6), 2272-7.
- Palmiter RD, Brinster RL, Hammer RE, Trumbauer ME, Rosenfeld MG, Birnberg NC and Evans RM.** Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature.* 1982, 300 (5893), 611-5.
- Peng J, Zhu Y, Milton JT and Price DH.** Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.* 1998, 12 (5), 755-62.
- Peoc'h K, Serres C, Frobert Y, Martin C, Lehmann S, Chasseigneaux S, Sazdovitch V, Grassi J, Jouannet P, Launay JM and Laplanche JL.** The human "prion-like" protein Doppel is expressed in both Sertoli cells and spermatozoa. *J Biol Chem.* 2002, 277 (45), 43071-8.
- Perry AC, Wakayama T, Kishikawa H, Kasai T, Okabe M, Toyoda Y and Yanagimachi R.** Mammalian transgenesis by intracytoplasmic sperm injection. *Science.* 1999, 284 (5417), 1180-3.
- Petters RM and Sommer JR.** Transgenic animals as models for human disease. *Transgenic Res.* 2000, 9(4-5), 347-51. Review.
- Phelps CJ, Koike C, Vaught TD, Boone J, Wells KD, Chen SH, Ball S, Specht SM, Polejaeva IA, Monahan JA, Jobst PM, Sharma SB, Lamborn AE, Garst AS, Moore M, Demetris AJ, Rudert WA, Bottino R, Bertera S, Trucco M, Starzl TE, Dai Y and Ayares DL.** Production of alpha 1,3-galactosyltransferase-deficient pigs. *Science.* 2003, 299 (5605), 411-4.

- Phi-Van L, Von Kries JP, Ostertag W and Stratling WH.** The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. *Mol Cell Biol.* 1990, 10 (5), 2302-7.
- Phi-Van L and Stratling WH.** Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry.* 1996, 35 (33), 10735-42.
- Pierre S, Jolivet G, Devinoy E, Theron MC, Malienou-N'Gassa R, Puissant C and Houdebine LM.** A distal region enhances the prolactin induced promoter activity of the rabbit alpha s1-casein gene. *Mol Cell Endocrinol.* 1992, 87 (1-3), 147-56.
- Pincon-Raymond M, Vicart P, Bois P, Chassande O, Romey G, Varadi G, Li ZL, Lazdunski M, Rieger F and Paulin D.** Conditional immortalization of normal and dysgenic mouse muscle cells by the SV40 large T antigen under the vimentin promoter control. *Dev Biol.* 1991, 148 (2), 517-28.
- Pinkert CA and Murray JD.** Transgenic farm animals. In: Murray JD, Anderson GB, Oberbauer AM and McGloughlin MM (Eds), transgenic animals in agriculture. CAB1 publishing, New York. 1999, 1-18.
- Platenburg GJ, Kootwijk EP, Kooiman PM, Woloshuk SL, Nuijens JH, Krimpenfort PJ, Pieper FR, de Boer HA and Strijker R.** Expression of human lactoferrin in milk of transgenic mice. *Transgenic Res.* 1994, 3 (2), 99-108.
- Poggiali P, Scoarughi GL, Lavitrano M, Donini P and Cimmino C.** Construction of a swine artificial chromosome: a novel vector for transgenesis in the pig. *Biochimie.* 2002, 84 (11), 1143-50.
- Polejaeva IA, Chen SH, Vaught TD, Page RL, Mullins J, Ball S, Dai Y, Boone J, Walker S, Ayares DL, Colman A and Campbell KH.** Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature.* 2000, 407 (6800), 86-90.
- Porter SD, Hu J and Gilks CB.** Distal upstream tyrosinase S/MAR-containing sequence has regulatory properties specific to subsets of melanocytes. *Dev Genet.* 1999, 25 (1), 40-8.
- Porteus MH and Baltimore D.** Chimeric nucleases stimulate gene targeting in human cells. *Science.* 2003, 300 (5620), 763.
- Prioleau MN, Nony P, Simpson M and Felsenfeld G.** An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J.* 1999, 18 (14), 4035-48.
- Prusiner SB.** Prions. *Proc Natl Acad Sci U S A.* 1998, 95 (23), 13363-83. Review.

- Qasba PK and Safaya SK.** Similarity of the nucleotide sequences of rat alpha-lactalbumin and chicken lysozyme genes. *Nature*. 1984, 308 (5957), 377-80.
- Recillas-Targa F, Bell AC and Felsenfeld G.** Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc Natl Acad Sci U S A*. 1999, 96 (25), 14354-9.
- Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M and Felsenfeld G.** Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A*. 2002, 99 (10), 6883-8.
- Reitman M and Felsenfeld G.** Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. *Mol Cell Biol*. 1990, 10 (6), 2774-86.
- Rideout WM 3rd, Eggan K and Jaenisch R.** Nuclear cloning and epigenetic reprogramming of the genome. *Science*. 2001, 293 (5532), 1093-8. Review.
- Rijkers T, Peetz A and Ruther U.** Insertional mutagenesis in transgenic mice. *Transgenic Res*. 1994, 3 (4), 203-15. Review.
- Rival-Gervier^a S, Thepot D, Jolivet G and Houdebine LM.** Pig whey acidic protein gene is surrounded by two ubiquitously expressed genes. *Biochim Biophys Acta*. 2003, 1627 (1), 7-14.
- Rival-Gervier^b S, Maeder C, Jolivet G, Thépot D, Prince S, Viglietta C, Leroux-Coyau M, Attal J and Houdebine LM.** Pig WAP gene insulator is located further than 80 kb upstream of the transcription start point. Poster L 48. In Transgenic animal research conference IV, 2003, California, USA.
- Robertson E, Bradley A, Kuehn M and Evans M.** Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature*. 1986, 323 (6087), 445-8.
- Robertson G, Garrick D, Wu W, Kearns M, Martin D and Whitelaw E.** Position-dependent variegation of globin transgene expression in mice. *Proc Natl Acad Sci U S A*. 1995, 92 (12), 5371-5.
- Robl JM, Kasinathan P, Sullivan E, Kuroiwa Y, Tomizuka K and Ishida I.** Artificial chromosome vectors and expression of complex proteins in transgenic animals. *Theriogenology*. 2003, 59 (1), 107-13. Review.

- Rossi D, Cozzio A, Flechsig E, Klein MA, Rulicke T, Aguzzi A and Weissmann C.** Onset of ataxia and Purkinje cell loss in PrP null mice inversely correlated with Dpl level in brain. *EMBO J.* 2001, 20 (4), 694-702.
- Rudolph NS.** Biopharmaceutical production in transgenic livestock. *Trends Biotechnol.* 1999, 17 (9), 367-74. Review.
- Ryoo ZY, Kim MO, Kim KE, Bahk YY, Lee JW, Park SH, Kim JH, Byun SJ, Hwang HY, Youn J and Kim TY.** Expression of recombinant human granulocyte macrophage-colony stimulating factor (hGM-CSF) in mouse urine. *Transgenic Res.* 2001, 10 (3), 193-200.
- Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M and Felsenfeld G.** Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO J.* 2000, 19 (10), 2315-22.
- Sambrook J and Russel DW.** Molecular cloning: A laboratory manual, third edition. Cold Spring Harbor Laboratory Press, 2001.
- Sanger F, Nicklen S and Coulson AR.** DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology.* 1992, 24, 104-8.
- Schnieke AE, Kind AJ, Ritchie WA, Mycock K, Scott AR, Ritchie M, Wilmut I, Colman A and Campbell KH.** Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science.* 1997, 278 (5346), 2130-3.
- Shaked Y, Rosenmann H, Talmor G and Gabizon R.** A C-terminal-truncated PrP isoform is present in mature sperm. *J Biol Chem.* 1999, 274 (45), 32153-8.
- Shani M, Barash I, Nathan M, Ricca G, Searfoss GH, Dekel I, Faerman A, Givol D and Hurwitz DR.** Expression of human serum albumin in the milk of transgenic mice. *Transgenic Res.* 1992, 1 (5), 195-208.
- Shinohara T, Orwig KE, Avarbock MR and Brinster RL.** Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. *Proc Natl Acad Sci U S A.* 2000, 97 (15), 8346-51.
- Silverman GL, Qin K, Moore RC, Yang Y, Mastrangelo P, Tremblay P, Prusiner SB, Cohen FE and Westaway D.** Doppel is an N-glycosylated, glycosylphosphatidylinositol-anchored protein. Expression in testis and ectopic production in the brains of Prnp(0/0) mice predisposed to Purkinje cell loss. *J Biol Chem.* 2000, 275 (35), 26834-41.

- Smithies O, Gregg RG, Boggs SS, Koralewski MA and Kucherlapati RS.** Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*. 1985, 317 (6034), 230-4.
- Sola I, Castilla J, Pintado B, Sanchez-Morgado JM, Whitelaw CB, Clark AJ and Enjuanes L.** Transgenic mice secreting coronavirus neutralizing antibodies into the milk. *J Virol*. 1998, 72 (5), 3762-72.
- Soulier S, Mercier JC, Vilotte JL, Anderson J, Clark AJ and Provot C.** The bovine and ovine genomes contain multiple sequences homologous to the alpha-lactalbumin-encoding gene. *Gene*. 1989, 83 (2), 331-8.
- Soulier S, Vilotte JL, Stinnakre MG and Mercier JC.** Expression analysis of ruminant alpha-lactalbumin in transgenic mice: developmental regulation and general location of important cis-regulatory elements. *FEBS Lett*. 1992, 297 (1-2), 13-8.
- Soulier S, Lepourry L, Stinnakre MG, Langley B, L'Huillier PJ, Paly J, Djiane J, Mercier JC and Vilotte JL.** Introduction of a proximal Stat5 site in the murine alpha-lactalbumin promoter induces prolactin dependency in vitro and improves expression frequency in vivo. *Transgenic Res*. 1999, 8 (1), 23-31.
- Soulier S, Hudrisier M, Da Silva JC, Maeder C, Viglietta C, Besnard N and Vilotte JL.** Substitution of the alpha-lactalbumin transcription unit by a CAT cDNA within a BAC clone silenced the locus in transgenic mice without affecting the physically linked Cyclin T1 gene. *Genet Sel Evol*. 2003, 35 (2), 239-47.
- Spana C, Harrison DA and Corces VG.** The *Drosophila melanogaster* suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes Dev*. 1988, 2 (11), 1414-23.
- Stief A, Winter DM, Stratling WH and Sippel AE.** A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature*. 1989, 341 (6240), 343-5.
- Stinnakre MG, Vilotte JL, Soulier S, L'Haridon R, Charlier M, Gaye P and Mercier JC.** The bovine alpha-lactalbumin promoter directs expression of ovine trophoblast interferon in the mammary gland of transgenic mice. *FEBS Lett*. 1991, 284 (1), 19-22.
- Stinnakre MG, Vilotte JL, Soulier S and Mercier JC.** Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proc Natl Acad Sci U S A*. 1994, 91 (14), 6544-8.
- Stinnakre MG, Soulier S, Schibler L, Lepourry L, Mercier JC and Vilotte JL.** Position-independent and copy-number-related expression of a goat bacterial artificial

- chromosome alpha-lactalbumin gene in transgenic mice. *Biochem J.* 1999, 339 (Pt 1), 33-6.
- Stowers AW, Chen Lh LH, Zhang Y, Kennedy MC, Zou L, Lambert L, Rice TJ, Kaslow DC, Saul A, Long CA, Meade H and Miller LH.** A recombinant vaccine expressed in the milk of transgenic mice protects Aotus monkeys from a lethal challenge with *Plasmodium falciparum*. *Proc Natl Acad Sci U S A.* 2002, 99 (1), 339-44.
- Symula DJ, Frazer KA, Ueda Y, Deneffe P, Stevens ME, Wang ZE, Locksley R and Rubin EM.** Functional screening of an asthma QTL in YAC transgenic mice. *Nat Genet.* 1999, 23 (2), 241-4.
- Taboit-Dameron F, Malassagne B, Viglietta C, Puissant C, Leroux-Coyau M, Chereau C, Attal J, Weill B and Houdebine LM.** Association of the 5'HS4 sequence of the chicken beta-globin locus control region with human EF1 alpha gene promoter induces ubiquitous and high expression of human CD55 and CD59 cDNAs in transgenic rabbits. *Transgenic Res.* 1999, 8 (3), 223-35.
- Takase K and Ebner KE.** Interaction of galactosyltransferase with alpha-lactalbumin and substrates. *Curr Top Cell Regul.* 1984, 24, 51-62.
- Testa G, Zhang Y, Vintersten K, Benes V, Pijnappel WW, Chambers I, Smith AJ, Smith AG and Stewart AF.** Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat Biotechnol.* 2003, 21 (4), 443-7.
- Thomas KR and Capecchi MR.** Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell.* 1987, 51 (3), 503-12.
- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS and Jones JM.** Embryonic stem cell lines derived from human blastocysts. *Science.* 1998, 282 (5391), 1145-7.
- Tomizuka K, Shinohara T, Yoshida H, Uejima H, Ohguma A, Tanaka S, Sato K, Oshimura M and Ishida I.** Double trans-chromosomal mice: maintenance of two individual human chromosome fragments containing Ig heavy and kappa loci and expression of fully human antibodies. *Proc Natl Acad Sci U S A.* 2000, 97 (2), 722-7.
- Topper YJ and Freeman CS.** Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol Rev.* 1980, 60 (4), 1049-106. Review.
- Udvardy A, Maine E and Schedl P.** The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol.* 1985, 185 (2), 341-58.

- Valenzuela DM, Murphy AJ, Frenthewey D, Gale NW, Economides AN, Auerbach W, Poueymirou WT, Adams NC, Rojas J, Yasenchak J, Chernomorsky R, Boucher M, Elsasser AL, Esau L, Zheng J, Griffiths JA, Wang X, Su H, Xue Y, Dominguez MG, Noguera I, Torres R, Macdonald LE, Stewart AF, DeChiara TM and Yancopoulos GD.** High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol.* 2003, 21 (6), 652-9.
- Van Berkel PH, Welling MM, Geerts M, van Veen HA, Ravensbergen B, Salaheddine M, Pauwels EK, Pieper F, Nuijens JH and Nibbering PH.** Large scale production of recombinant human lactoferrin in the milk of transgenic cows. *Nat Biotechnol.* 2002, 20 (5), 484-7.
- Vickers PJ, Dickson RB, Shoemaker R and Cowan KH.** A multidrug-resistant MCF-7 human breast cancer cell line which exhibits cross-resistance to antiestrogens and hormone-independent tumor growth in vivo. *Mol Endocrinol.* 1988, 2 (10), 886-92.
- Vilotte JL, Soulier S, Stinnakre MG, Massoud M and Mercier JC.** Efficient tissue-specific expression of bovine alpha-lactalbumin in transgenic mice. *Eur J Biochem.* 1989, 186 (1-2), 43-8.
- Vilotte JL, Soulier S, Printz C and Mercier JC.** Sequence of the goat alpha-lactalbumin-encoding gene: comparison with the bovine gene and evidence of related sequences in the goat genome. *Gene.* 1991, 98 (2), 271-6.
- Vilotte JL, Soulier S and Mercier JC.** Complete sequence of a bovine alpha-lactalbumin pseudogene: the region homologous to the gene is flanked by two directly repeated LINE sequences. *Genomics.* 1993, 16 (2), 529-32.
- Vilotte JL, Soulier S, Essalmani R, Stinnakre MG, Vaiman D, Lepourry L, Da Silva JC, Besnard N, Dawson M, Buschmann A, Groschup M, Petit S, Madelaine MF, Rakatobe S, Le Dur A, Vilette D and Laude H.** Markedly increased susceptibility to natural sheep scrapie of transgenic mice expressing ovine prp. *J Virol.* 2001, 75 (13), 5977-84.
- Vilotte JL and Laude H.** Transgenesis applied to transmissible spongiform encephalopathies. *Transgenic Res.* 2002, 11 (6), 547-64. Review.
- Wadsworth JD, Jackson GS, Hill AF and Collinge J.** Molecular biology of prion propagation. *Curr Opin Genet Dev.* 1999, 9 (3), 338-45. Review.
- Wai AW, Gillemans N, Raguz-Bolognesi S, Pruzina S, Zafarana G, Meijer D, Philipsen S and Grosveld F.** HS5 of the human beta-globin locus control region: a developmental stage-specific border in erythroid cells. *EMBO J.* 2003, 22 (17), 4489-4500.

- Wall RJ.** Pronuclear microinjection. *Cloning Stem cells*. 2001, 3 (4), 209-20. Review.
- Wang ZQ, Kiefer F, Urbanek P and Wagner EF.** Generation of completely embryonic stem cell-derived mutant mice using tetraploid blastocyst injection. *Mech Dev*. 1997, 62 (2), 137-45.
- Wei P, Garber ME, Fang SM, Fischer WH and Jones KA.** A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*. 1998, 92 (4), 451-62.
- Weissmann C.** Molecular genetics of transmissible spongiform encephalopathies. *J Biol Chem*. 1999, 274 (1), 3-6. Review.
- West AG, Gaszner M and Felsenfeld G.** Insulators: many functions, many mechanisms. *Genes Dev*. 2002, 16 (3), 271-88. Review.
- Wheeler MB.** Development and validation of swine embryonic stem cells: a review. *Reprod Fertil Dev*. 1994, 6 (5), 563-8. Review.
- Whitelaw CB, Harris S, McClenaghan M, Simons JP and Clark AJ.** Position-independent expression of the ovine beta-lactoglobulin gene in transgenic mice. *Biochem J*. 1992, 286 (Pt 1), 31-9.
- Whitelaw CB, Springbett AJ, Webster J and Clark J.** The majority of G0 transgenic mice are derived from mosaic embryos. *Transgenic Res*. 1993, 2 (1), 29-32.
- Whitelaw CB and Bruce DM.** Does genetic modification violate intrinsic value? *Trends Biotechnol*. 2002, 20 (12), 488-9.
- Willoughby DA, Vilalta A and Oshima RG.** An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J Biol Chem*. 2000, 275 (2), 759-68.
- Wilmot I, Schnieke AE, McWhir J, Kind AJ and Campbell KH.** Viable offspring derived from fetal and adult mammalian cells. *Nature*. 1997, 385 (6619), 810-3.
- Winter A, Kramer W, Werner FA, Kollers S, Kata S, Durstewitz G, Buitkamp J, Womack JE, Thaller G and Fries R.** Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc Natl Acad Sci U S A*. 2002, 99 (14), 9300-5.
- Wood SA, Pascoe WS, Schmidt C, Kemler R, Evans MJ and Allen ND.** Simple and efficient production of embryonic stem cell-embryo chimeras by coculture. *Proc Natl Acad Sci U S A*. 1993, 90 (10), 4582-5.

- Yang XW, Model P and Heintz N.** Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol.* 1997, 15 (9), 859-65.
- Zahn-Zabal M, Kobr M, Girod PA, Imhof M, Chatellard P, de Jesus M, Wurm F and Mermod N.** Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol.* 2001, 87 (1), 29-42.
- Zavizion B, van Duffelen M, Schaeffer W and Politis I.** Establishment and characterization of a bovine mammary epithelial cell line with unique properties. *In Vitro Cell Dev Biol Anim.* 1996, 32 (3), 138-48.
- Zhan Y, Brady JL, Johnston AM and Lew AM.** Predominant transgene expression in exocrine pancreas directed by the CMV promoter. *DNA Cell Biol.* 2000, 19 (11), 639-45.
- Zhu Y, Jong MC, Frazer KA, Gong E, Krauss RM, Cheng JF, Boffelli D and Rubin EM.** Genomic interval engineering of mice identifies a novel modulator of triglyceride production. *Proc Natl Acad Sci U S A.* 2000, 97 (3), 1137-42.

ANNEXES

1 CGAGTGTCT AGTCAGAGGT TGGGGGCAGG GACCAATAGT GTCTACAAGC
51 TGTGGATGTC TTCCCACTCC ATGCAAGGGG GAGAGGTAGG TGATTCCCAT
101 CAGAGCTGCT GGGATGGAGA GTCCTGGAGA CTGTGCCTAA GGAAGAGAG
151 GTTTTTGTA TGCTGTTTCC CAAGTCTTGG GATGTTGAGG GACCCAGTGG
201 TTGGTTATCT CTGTCTAGCA GCCTGTATTT TAGCCTTTCC TCACCTCTCA
251 TCCTTGGGTA CTCTGAAGCT CCTCTCCTCA ATGAACTCTG ACCCATAGTC
301 TTTGAGAAAA TAGCCCACAG AACTGGGCCT TCCTAAACCA TCCATCTTGC
351 TTCTAACCTG TGTAACCTTC TGTAGTGGCT ATCTTATGTA AGATACGATA
401 AAAAATTATT GTACATAGTT TCTATTAAAT GTTAAAGTGT TATGAAAATA
451 TAAAACAAGG AGTTGAATGC ATGTTTTGTT TGTGTTTGAG AGTGTGGTGT
501 CTATATGTTG GTGGGGGTGG GTTGGGAGCT GTCTGATGGG AAAGGGAAAA
551 GAAGACAGAT ATCACTTAAA AGATTTAAAA ACCTGTAATA AGTAAATGGA
601 AAGGGCTCCC TTATTTAAAC CTACTTGTCT AGTTATATTT CTGCATATTC
651 CCCCCAATG GGAGGAATTT TTCATCCTCT GTGCTCCCTC CTCCCACTCA
701 CATTTTTTTC TCACAATTAA TGGCTGTTGT CTCAAAATCC TCTAAGGCTA
751 GGAGTGGGTG TGGGGGGGTG TGGGTGTGTG TGTAAAGCTA GGAGTGTGTG
801 ATCTCACCGT AATGTAATAT GACAGGGAGG CCAAAAAGTT ACCCTTTGTA
851 AAAACTGGAA TGGAGAAAAG TTACATAGTT ACATTATCAC CTTGGCTTTA
901 GTTTATTTCC TGGGAATGGT CAGGAAGTGA CTCAGAAGAC CTGTCAGATG
951 AATCCTGAGA TCACCTTGAA TAGGGCTTGC TGTATACTCC AGTTTATTTG
1001 TTCCCACTGC TTTCTTGTCT AGTACTTTTT ATACTATTCA TGAACATCCC
1051 TTATGCAGAT TGAAAACTT CTGCAACAAG ATTAGGGATT AAAATATCTT
1101 CTTCCCATTT TTTTTTTTTT TTTCTGCTCC ACCACACAGC ATGTAAGATC
1151 TTAGTTTCCC AACAAAGACC CCACTCTCCT GGGGTGGTAG TGCACAGTGC
1201 TAACCACTAA ACTGCCAGGG AAGTCCCGTC TTTTATTGT TTTAAAAAAT
1251 TGCTGGCAA CTGCTAGCAT CCACGTCACT CAGTCCGGTG GTTTTATAAA
1301 TTAACAGGT CAGATACCTT TGGCAGGGGC CGGAGGGTGG AGTAG**AAGCC**
1351 **ATGGAACCTG** **TATCCTAGAA** CCCAGAGATT CTGATGGCCA CTTGTCTGCA
1401 AATGCAAGGT GAAGCGTTTG GAAAAGCCCA CAAATAGCTA CTACTCTCCT
1451 CCAAGTTTT ATCCTCTGTA ATTAAGATTT GGTAGGAAGA GACTTATTTG
1501 GGCTGCATTT ATAGCTCAGG AAATGCTGTT GCATGGTAGT CCCTAAGCAA
1551 AGGTACCTG TCAGAAGACA TGGCCTGCCT TTAGTGGCAG CATCCTGAGG
1601 GAGCTGTGGT CTCAGCGTAG ATCTACAGAT AGAGCTGGAG CCCTACTTTG
1651 TTCAGCTGAT CAGTCAGTTA TCCTGATACG GAGCTCCTTA TCCTGGACAG

1701 AGTTTGAAGT TTCCTGGTGG TACTTCAGAG GAGGAGATGG TTGGGGGTGT
1751 GATTGACTCA GTGAAGAGGA GTTTGAGCCA ACTCCAGGAA ATAGTGTAGG
1801 ACAGGGAAGC CTAGCGTGGG GTCGCAGAGT TGGACACACT GACTGAACAA
1851 CGAATAGATC CTATCTTTCC TGGCAAGAAT GCCTTAAGGA TCAGGACCTA
1901 TAATCTTGCA GAGCCTAAAG TCAGGTGGAA ATACCCAGCA GGTCATGGTT
1951 GTAATAAGCA GGGCTATTCT GCTTCTGTCT CCTGGTCCCA AGCCCATTTT
2001 TTAATTCCCA GAGGAACATA ACAGTTTATA GAGATTTCTT ATTTAGTGCA
2051 TGTTGTGCGT GCTGAAGAAG AGGGCCTCCA CAGAAGCTCA GCACCTGCAA
2101 TGCAGGAGAC CCGGGTTAGA **TTTCTGGGTT** **GCAAAGATCC** CCTGGAGATG
2151 GAAATGGCAA CCCACTCAAG TAATCTTGCC TGAAAAATCC CATGGACAGA
2201 GCCTGGTGGG CTACAGTGCA TGGGGTCACA ATAAGAATGC GACACAACCT
2251 AGCGACTAAA ATAATGAACA ACTGTGTCCT GAAGGATAGC ACCCATCCTT
2301 GAGAAGTATT TCCTCCATGC TGGTGTTTTT CAACTGTACC TTCAGTAGGC
2351 TGTTCCAGGG CTGTCTAAGG CCAGCTGCTC CTGGATGGTG CAGTGTGTGA
2401 TATG**GTCAGT** **TGATTCTCCG** **CTGTGCAGTG** **GATTCCTTAG** **GTGATAGGTC**
2451 **ATGTGGAATT** **TCATGCCTGC** **AGTGCTGCCT** **GAAGTGCTGC** **CTGAAGCCCT**
2501 **GTGAACAGGA** **AAGGCAAAAC** **CTGTACCTGA** **AATAGGGTAC** **ATCCCTGTGA**
2551 **GGATGTATCA** **CTTGCTCTTG** **AAGGACTGAA** **AGGGCCTGGT** **ATATTCTGCT**
2601 **TACTACCAAG** **AAGGTGGTGT** **GGTTGAGGGG** **TAGTGCATAA** **CAGGGGCTCA**
2651 **GTGTTGCCTC** **TGTTGGTCGC** **TGGTTAGACG** **TTTGGAGGGG** **ACAGTAGCCA**
2701 **GGTGATAACT** **AGGAGTCCAT** **TTCAGTATAA** **TCTCCATCTT** **TGTCTGCTTC**
2751 **GTTTATGTAT** **CTGTCCTAGT** **TCTAGGGTGA** **TGATGAAGGA** **TAGTTAACAT**
2801 **CAGCTGATCA** AGCCACTTTG TCTACTTGGT TTTCCAGTGC CATTTTTAAG
2851 TGGAGAATGC TTTCTGGTGG GTGTTAACAA GTGACAGAGA AACAGTTTCA
2901 TGTTTGATTC ATATTCTCAT GTCTTGTCAT GTATCTTTAC CTCACCCTCT
2951 TTGTGTCTTC TAGTCCCTTT TCTTCCAATT TCCCGACCAG TCAGCCAGAT
3001 CCTTGTTGTC TGCTTAAGAA TCTGTGTATA TTCTCACCTC CAGCCATGTC
3051 TCCTTCCACA AAAAGTAGAT GGCCAAGTGC ACTGCTCACG TCTTCATCTA
3101 CTGAGATGTT TTTCTCTTTT CACTGAAGGC TGGATTAAAG CATATAGGAA
3151 GGTGTCCTTA GGTTATATGC AAATACCATG CAGATTTTGG TATGCCCAAG
3201 TGTCCTTGAA TCAACTCTTC ACAGATACGA AGGGATAACT GTTAAACGCT
3251 CCAATAAAAC ATGGGGATAA ATCTTCATAA TTTTGGATTT AGCCACAGTT
3301 TTGTAGATCT GACAGCAAAG CACAGGCAAT GAAAAAATAA ATTGGATTTT
3351 CTCAAAATTA AAAATCTGTT TGCATCAAAA GATACAATTC AGAAAGTGAA

3401 AGTATAGCCC ATAATGAAAT AAAATATTTG CCCATTAAAT ATATGATGAA
3451 ACGTTATCCA GAGTATGCCA ACATAGACAA AACCCCCAGT TTTAAA**ATGG**
3501 **ATAAATGACC TCAGTAGACC** ATTCTTCACA CAAGAGATAT AAATGGTCAT
3551 TGCAGAAAAA AAGATGCCAA ATATCGTTAG TCATGATGAA AGTGCAAATT
3601 AAAATCAATA TTCAACAACCT CCACATATTC CAAAATGGTT GTAATCGAGA
3651 AAATAAAAAA TAACAAATGT TAGCAACAAT GTAGAGAAAT TGGAAACTTT
3701 AAAAATGGA AATGGGAAAA GGCACAACCT ATGTTTACAA CATAAATGTA
3751 CAATTAGCAT GTGGTTCAAC AACTCTGCTC TCAGATACAT ATCCCAAAGA
3801 ACTGAAAACA GCTGTACACA TGTATATTGA TAGCACCCT GTTCACAAGA
3851 GCCCAGAGGC AGAAACAGTG CAGATGCCCA TCAGTGGGTA AACAGATCAG
3901 CAAAATGGAG CAGATCCAGA GGTTAAGTGA ATCCTGATGG TTCAGTGGTT
3951 AAGACTCCAC ACTTCAGTGC AGGGGTGCAG GTTTACTCCT TGGTGGGGAG
4001 CGAAGATGCC CCGTGCTGCA CAGTGTGGCC AAAAGATAAC AAAAATGGGA
4051 TAGATTCCAT ATACAGCATT ATTATCTGTC ATAGGAGGAA CAGAGTGAAA
4101 TGGTTCAGCC TCTGTGGAAA ACAGTACGGT AGGTCTTTAA AAATTA AAC
4151 ACAGAATTAA TGTATGATCA ATTTACATG AGGATATATA CAAAAAGAC
4201 TGAAAGCAAA GACTAGAACA GATATGGGTA CAGTCATATT CATAGCAGCA
4251 TTAATCATAA GAGCAAACGG TAGAAGCAAC CCAAGTGTCC ACTGATGAAT
4301 GAATGCGCAA ACTAAATGTG GCATATATAC AAAGTAATGT AGTTCAGTCT
4351 TCAAAA**CTAA GGAAATACCA GCACATGCTG** CAATGTATAT GAGCCTTGG
4401 GTCATGATGA TGAGTGAAAT AACTCTCACA AAAAGACAAG AATGTGATTT
4451 CACATGTATG TGGTATCCAG AAAAGTCAAA TTCAGAACAC AGAATGCTGT
4501 TGGTTGCCAG GAGCTGAGGA AAAGTAGAGC TTAATGGCTA CAGAGTTTCA
4551 CTAGTGAAAA TTGAAAAGTT CTGAACATGG ATGGTGAGGA TATTTGTATA
4601 ACAATGTGAA TGTGTGTTAAT ACCATTGAAT TTAAAGCATA GTTATAACGG
4651 TTTGAATAGT AAATATCTAT GTTTAAGCAC AATGAAAAAA ATGAAGTACT
4701 GATACATGTT ACCACGTGGA TGAACATCCC AAGTGAAAGA AGCCACACAG
4751 ATAAAGAGTA GCTTGATCAT ACTCATCTAG TGGACTTTAA ATCCATCAAG
4801 ATTGTA AAAA TGGGGTAAAG TGTCTGTAAT GGTTCAGAA AAAGAACACT
4851 GGAGCAGTTT TAGTCTGGCC AACCAGCCCT GGTGACAAAG TGAAGAAGAC
4901 TCAGAGGCCA CGGTAGATCC CCTCAATGGA AGAGTGAATA GACTCCTC
4951 TGAAGGAACA GTGTGCAGGA **CCCATAGAAG TCCCACCGCA** CTTTTT**GAA**
5001 **TAGCAGTGCT GACCCTTAA** TCTGGCCCA CTAAAATTCT TGCTCAATTT
5051 TTCAAAGATA CATCAGAGTA CTGGTTGAAA TAGTAGGAAA TTGGCAATGA

5101 CACTTATGTT CATGGAGATG GGGATGGGAT CCATCAGTTT TAGTCCACAT
5151 GGATGGACAA ATCCTGTGCA TCTAAGAAAG TGGGTGTAGG GTCTCTGTCT
5201 GACGACAAGG AGACCTCAGG GTTCTCTTTG TGTATGTACA TGGACCAATT
5251 GTCTGCACCC AGGATCCTAT CTATAAGACC ACCTTTATGT GACATCAACA
5301 TAGATATGAA TGAATCTGCT TGTGTGAGAG AGAGAGAGAG AGGAGAGAAG
5351 CAGAGCCTGA GAGATATGCA CAAAAACAGT CATAGCGGTC ACTTCTGGGG
5401 ATGGGACTTA CCAATCTCTT CTCCTTTAGG TGAAATTTCA TGTTC AATA
5451 ATGTGTGTGT AATATTTTAA TAATTAAAAA ATAAAATCGT GTGTTAAATG
5501 TGGCCCATTC ATCACTCATC TATGGTGGCA CTAATGGTAA AGAGTCTGCC
5551 TGCTAATGCA GGAGATACTA GAGACGCTGA TTCGATTCTT GTGTTGGGAG
5601 GATCCCCTGG AGAAGGAAAT GGCAGCCAC TCCAGCATTC TTGCCTGGAA
5651 AACTCCACGG AGGGAGGAGC CTGGCGGGCT GCAGTCCATG GGGTCGCAAA
5701 AAGTCGCACA TGAAAGCATA GCAGCAGCAC AGCATCAGGA ACTACA ACTT
5751 AATGGCAAAA AAGAAAAAAA ATCTGTTTCA AAATTGCACA GAGGAACTGA
5801 ATAGACGTTT TTTCCAAAGA AAACATACAG AGGGGCAACA **GACACATGAA**
5851 **AAGATGCTCA** **ACATCACTCA** TCATCAAGGA ACTGCAAATC CAAACCACAG
5901 CGAGGTATTG CCTCACCCCT GTCAGGATGA CTGTCATCAA AAATACAAGA
5951 GAAAGTGTTG GTGAGGATGT GGGGAAAAGG GAAGCTCATT GTTGGTGGGA
6001 ATGCAAATTT GTGCAGTAGC TGTGGAAAAT AGTATGAAAC TTCTTCAAAA
6051 AATTAGAATG ACCATGTGAT CCAGCATTTT CACTCCTGGG AATACATCCA
6101 AAGGAAGTGA AATCACTATG TCAAAAGATG ATGCCGTTAA AGTGCTGCAC
6151 TCAATATGTC AGCAAATTTG GAAAATTCAG TAGTGGCCAC AGGACTGGAA
6201 AAGGTCAGTC GTCATTTCAA TACCAAAGAA AGATAATCCC AAAGATTGTT
6251 CAAACTATCA TACAATTGCC CTCATTTTAC ATGCTAGCAA GGTAATGCTC
6301 AAAATCCTTT AAGGTAGGCT TCAGCAGTAT ATGAACTGAG AACTTTTCAGA
6351 TGTACAAACT GGATTTAGAA AAGGCAGAGG AACAAGTGAT CAAACTGCCG
6401 ACATCTGTTG GATCATAGGA AAAGCAAGGG GATCACAAAA CAAACTTCTG
6451 CTTTATTGAC TACGCTAAAA CCTTTGAGTG TGTGGATCAC TACAAACTGT
6501 GGAAA ACTCT TACAGAGATG GGAATACCAG ACCACCTGAC CTGCCTCCTG
6551 AGAAATCTGT ATGCAGGTCA AGAAGCAACA GTTAGAACTG GACATGGAAC
6601 AACAGACTGG TTCCAAATTG GGAAAGGAGT ATGTCAAGGC TGTATGTTGT
6651 CACCTTGTTT ATTTAACTTA TATGCAAAGT GCATCATATG AAATGCTGGG
6701 CTAGATGAAT TACAAACTGG ACTCAAGATT GCCAGGAGAA **ATATCAACAA**
6751 CCTCAGATAT GCAGATGATA CCACTCTAAT GGCAGAAAGT GAAGAGGAAC

6801 TAAAGAGCCT CTTGATGAGG GTGAAAGAGG AGAGCAAAAA AACTGGCTTA
6851 AAAACTCAGC TTTCAAAAAA CTAAGATCAT GGCATCTGGT CCCATCACTT
6901 CGTGACAAAT AGAAGGGGAA AAAGTGGAAA CAGTGACAGA TTTTGTTTTC
6951 TTGGGCTCCA AAATTACTGT GGACGGTGAC TGCAGCCATG AAATTAAAAG
7001 ATGCTTGCTA TTTGGAAGGA CAGCATATTA AAAAGCAAAG ACTTCCCTTG
7051 TTGAAGTGAA GGACAGGGAA GCCTGGTATC CTGCAGTTCC CAGGGTCGAA
7101 AAAAGTTGGA CACAACCTAG TGACAACAAA AACAATGTCA AAGAAACCTG
7151 CACCATATTC ACTGCAGCAT TGTTTATAAT AACCAAGATA TGTAACACGC
7201 CTAAGTGTCC ATCGACAGAT GAATGAAGAA AGAAAATGCA GTATATAGAT
7251 ACAATGGAAT CTCACTGAGT CATAAAGAAA GAAAGCAAAT CTTGTTATTC
7301 GCAACAACAT AGATGGACCT **CAAGGGCATT** ATGATAAATT AAATAAGTCA
7351 GAGAAAGACA AATATTTATG ATCTTAATTA TGTGTGGAAT CTACAAAGTA
7401 AACTGAATTC ACAGAAACAA AGAACAAGTT GATGGTTGTC AGAGGCAGGG
7451 TGTTGAGGTG AATGGGTGAC TCGCGTGAAG ATGGTCAAAA GGCACAACT
7501 TCCAGCTATA TGATAAATCC TGGGCATGTA ATGTACATCA GGGTACTGT
7551 AGCTAACAAAC CCAGTGTTGT GTGTTTGAAA GTTGCTGGGG TTGATCCCTG
7601 GGTCAGGAAG ATCCCCTGGA GAAGGGAATG GCTACCCACT CCAGTATTCT
7651 TGCCTGGAGA ATTCCATGGA CAGAGGAGAT TGGCAGGTGA TTGCAAAGAG
7701 TTGGATATGA CTGAGCGACT TTCACTTTCA ATTCTCTTCC CCACCCTCAG
7751 AGGGGGCTTC CCAGGTGGGG CCAGCGGTAA AGAACCCGCC CGCCAACGCA
7801 GGAGACCTAA GAGACGGTGC ATGGGGTCTC AGAGAGACAT GACTGAAGTG
7851 ACTTAGCACA CACACCCTGA GAGGAGTCTC AGTGCTCCCG TTTTACACAC
7901 GAGAAAACAG GCTCATTTTG CTTTAAACAG CCGTCCAAGT GGTGGAGCAG
7951 CTTTGAACCC GTGTTCTTAT TCCAAAGCCC CGATTTCTAG ACGGTGGAGT
8001 CCAGGCACTG GAGACTATCC CTCTTCCACA CTGATGGTTT GCTCTAGGGC
8051 TGCTGGATTA ACCATTTAAT CCACTTACTC TGCACCCCT CAGCCCTTTG
8101 GCGGCTGCC TCCGGCTACT GCCTGACACT TCAGCCGGCT GAAGTTTTAC
8151 AGGTGAATCA TCACCATCCC CCTGCCCTCC CTGCCACGTG GGGCAGGCC
8201 CTCCTATCCT CATTATTCTA AATCTTTAGC TCACAAATCT GCAGAACTGT
8251 GTTAATAAAG CCCTGTTCAC CGTGTGTGTG CAGGCTTTGA AAAGGGAATC
8301 TGAGGGGGAG TTTGGGGTAT GAAAGGGACA GTATTTATGT CCCTAATGGG
8351 ACTTAATCTG AAGCTAAAAC CTAGGTCTCA GAAGAAAAAA AAAAAAAAAA
8401 GCTGAGGCTG AGGGCTACAA GCTGCTGAAA AGGGGGCTTG ACTACTGGAT
8451 TGTAATGTG **TAATTTAATT** TAAAAAATCA ACACTTGAT GATAGCAGTT

8501 TGCCTTTTAA GGCCAAATTG GCCACCTGAC ACTTGGGCAT TCCTCAATTC
8551 TGGGATGTGA TTGTCTCTCA GCTTTGGGGG GCCTTGACTG GGGGTGCTTT
8601 CCTCTCAGTT ATAGGCACAC CCACCCACAT GTGGACCCAT GACTGCAGTC
8651 CCTACAGGGC AGGGTGCTCC TCGTGCCTGG CCTCAGGTCA TAAGGAAGGA
8701 ATGGTATGTC TGC GCGGTTA CACGTATGTC CTCTGCAGGA ATGAGCATCA
8751 TGTGGAGCAG AAGCCTCCCC ATTCAGCAAG CAGCCTGGCT GGGGGTAGAG
8801 GAGCTGCCCC GTTCTGAAGT CATAAACTGG **ATGGTAGATT GGGATCCCTC**
8851 **TGCCC**TAGAC TCACCAGCAA GCCCCTCACC CACATCCACC TTCTGCTCTC
8901 TCCTCTGTAA AGTGTTGAC CACTGTGTAA ATTATAATAC AAAGTACTGG
8951 ATGTTTTGGG CACCTTGGGT TGTGCAGTCC CACAGCTGGG CAGAGGGCAG
9001 ATGGAAGAGG AAAGATGAGG GTGTAGCAGG CATGACAGCA GAGCTAAGAA
9051 CTGGCCAAGC ATGTGGATTT CAATATCCAG AACAGGTAG TTGCCACAAG
9101 CAAATCCATT AAGGCCATGT TTGCTCTCTC ATCAGTTACA ATCATCTGTT
9151 TGAACAGAAG CCTCTGGAGT TTCTTAATGA CATTATTGTC CATCACAGAA
9201 TGA CTCTGCT CGGCTCCTAG GCACTGGGGC CTCCACACAC CACGATTCTC
9251 CATGTGCTGT GCACTCACTT TCTTGGTTAG TGCCTCTCAG TAGCTCCCTG
9301 TTGCCTAGGG CCTGTGCTGC TCCAGGCAGA AACTGCCATA TGAGGTCCCTG
9351 CCCGCCTTCC CCACCTGTGC TGTGTGCCTG CAGCAGTGAA TCAATCACCT
9401 GCTGACTTAA AGAACAGTCA CAACCTAAAA CCTGCGAGTT ATGCTTTAGT
9451 TGGTGGGAAT TTTTAGGCCT TCAAGCCCGG GAGACAGCAT CTCAAGGAGG
9501 TGAGGGGAGA AGCCAGATTA TATAGAAGTT TCATAACAAA GGGCAAGTCA
9551 TCTTAATATT AAAAGATCAT TGTTAATCAA AGAAA ACTAG TTATCCAAAG
9601 TTAGGGCGTT TAATGGTTTT CTATGTATGG GAAGATAACA ACAGTCTGGG
9651 CTC ACTGAAA TTCATTTTGT ATGCATGTCA GCTATCCTGG AGAAGGAAAT
9701 GGCAACCCAC TCCAGTG TTC CTGTCTGGAG AATCCCAGGG ACGGCGGAGC
9751 CTGGTGGGCT GCCGTCTATG GGGTCGCACA GAGTCGGACA TGA CTGAAGT
9801 GACTCAGCAG CAGCAGCAGC AGCAGCAGCT ATCCTGGGCC AGTATCCTCT
9851 GTTCCTTCAC TTCTTGAGCG CCCTTGGGGC CCACCGTAGG GAGTGGCTGC
9901 AGTCCTGATG **GCTGTCGGAT CACAGATATT CTTCTTCCCG** AGTGCACTTA
9951 GGGCTTGCCA GCTCCCATTG GAGGGCTGAA ATTGCTAATG ACTGTGACAT
10001 CCTTGTTTAC TGACATGGCA GGAAGTGCTC **CGTTTCTCAC GCCAGTGATC**
10051 **CCCAAGCACT CAGGTCAGGT CCTTGCCTGC GCCTCTGCAC CTTTTCGCGG**
10101 **GCTTCTCCTC CCGCGCTCCT GCTCCCTTCT CTTTGAACCC GCAGCTCCCT**
10151 **TGTCATGTCC CCTTTATAGC AAGACGGTGT TGGCTGCTCA CCTGGAAATC**

10201 TTTGCCCAAC CCATCCTGCC TTCTTCCTGT GGATTGGTCC AAGCCAGCCA
10251 TGACACCCAC TTCCCCCACT GATGATTGGT CAGGACAAGG ACTCATGATC
10301 CAGTTCTGGC AAATGAGATA TGAGAGGAAG TCTACAGGGG CTTCTGTGAG
10351 AGCTTTTCCT TCCTGAAAAG AAAGAGATGG GCAAGGAAGA AGGTGCCTCC
10401 TTGTTTTTGG CTTGAGCACT CTGCCACACA **CTATGTGGTC CTAGG**

Annexe 1 : Séquence de la région de 10 kb.

Les oligonucléotides, dont la séquence et la position ont été reportées dans la **Table 5.1**, sont représentés en *italique* et souligné pour les oligonucléotides « sens » et en **gras** pour les « antisens ». Les régions **grisées** représentent les sondes utilisées dans cette région.