

Université de Montréal

Régulation transcriptionnelle du gène  $\beta$ -globine  
et des effets de position chez les souris transgéniques

par  
Louis-Georges Guy

Programme de biologie moléculaire  
Faculté des études supérieures

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Philosophiae Doctor (Ph.D.)  
en biologie moléculaire

Octobre 1997

Louis-Georges Guy, 1997



QH  
506  
U54  
1998  
V.003

Document de référence

Régulation transcriptionnelle du gène B-220  
et des autres de fonction des B-cells transgéniques

par  
Luis-Guillermo Cruz

Thèse présentée en vue de l'obtention du grade de  
Docteur en Médecine

Cette thèse a été présentée et soutenue publiquement  
le 15 Mars 1998 à 14 heures devant le jury  
composé de :



Président du jury : M. [Name]  
Membres du jury : M. [Name], M. [Name]

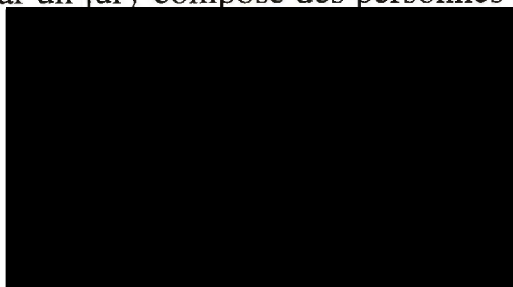
Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée:

Régulation transcriptionnelle du gène  $\beta$ -globine  
et des effets de position chez les souris transgéniques

Présentée par:  
Louis-Georges Guy

a été évaluée par un jury composé des personnes suivantes:



Thèse acceptée le: 29.01.1998

## Sommaire:

Les gènes de type  $\beta$ -globine sont régulés durant le développement de telle sorte que  $\epsilon$ -globine est exprimée chez l'embryon, les  $\gamma$ -globines chez le fœtus et  $\beta$ -globine chez l'adulte. Tous ces gènes sont regroupés dans le même locus et leur spécificité développementale est obtenue en partie par la compétition des gènes pour l'activation par une région régulatrice située en amont du locus et appelée région de contrôle du locus (LCR). Le LCR a une activité activatrice (enhancer) très puissante. Contrairement aux enhancers classiques, le LCR a également un effet de décompaction de la chromatine du locus et peut de ce fait conférer une expression indépendante de la position d'intégration dans le génome à un gène de globine chez les souris transgéniques. Dans le but de mieux comprendre les interactions entre le LCR et les régions régulatrices du gène de  $\beta$ -globine, nous avons étudié des souris portant des transgènes où le LCR et divers éléments du gène de  $\beta$ -globine dirigent l'expression d'un gène rapporteur *lacZ*. Nous démontrons que l'insertion des séquences *lacZ* dans le transgène interfère avec l'expression indépendante de la position et que l'activité enhancer du LCR est toujours présente, même si elle est influencée par la position. La structure de la chromatine du LCR, mesurée par l'hypersensibilité à l'ADNase I du transgène, est elle aussi influencée par la position en présence du rapporteur *lacZ*. L'ajout de toutes les séquences du gène  $\beta$ -globine au transgène ne peut rétablir l'expression indépendante de la position, indiquant que ce n'est pas un manque d'élément en *cis* qui interfère, mais la présence même de *lacZ*. L'étude de constructions plus complexes a pu montrer qu'un gène de  $\beta$ -globine complet était soumis à des effets de position malgré la présence du LCR si une unité transcriptionnelle indépendante codant pour *lacZ* compétitionne pour l'interaction avec le LCR. Nous avons développé un modèle où le LCR doit entrer en interaction mutuelle avec un gène de globine pour décondenser la chromatine du locus et prévenir les effets de position. En présence de *lacZ*, une interaction non-productive, c'est-à-dire incapable de prévenir les effets de position, se produit entre le LCR et le promoteur dirigeant *lacZ*. Pour appuyer ce modèle, des constructions où les séquences *lacZ* sont remplacées par l'ADNc TIMP sont elles aussi soumises à des effets de position, indiquant que des séquences exogènes introduites dans le locus de globine empêche la formation d'une structure favorable à l'ouverture de la chromatine.

Récemment, des rapports font états de variéation de l'expression (expression hétérocellulaire) chez des souris transgéniques soumises à des effets de position. Afin de mieux comprendre les effets de position et le rôle du LCR, nous avons étudié la possibilité qu'il y ait variéation de l'expression du transgène dans nos lignées. Nous avons noté, par un coloration *in situ* détectant l'expression du transgène, que certaines cellules ne démontraient pas d'expression. Contrairement au modèle de variéation de l'expression, bien connu chez la drosophile et quelquefois rencontré chez les souris transgéniques, nous avons noté une variation dans les niveaux d'expression par cellule d'une cellule à l'autre et d'une lignée de souris transgénique à l'autre. Nous démontrons que le nombre de cellules qui colorent est dû à l'expression moyenne, elle même fixée par la position d'intégration. La présence de cellule qui ne colorent pas est un résultat de la distribution d'expression par cellule et d'une limite de détection de la coloration et non à un phénomène de variéation de l'expression.

Finalement, nous avons étudié le rôle de EKLF, un facteur transcriptionnel nécessaire pour la transition entre l'expression du gène  $\gamma$ -globine et celle du gène  $\beta$ -globine dans le foie foetal de la souris. Avec une construction où il n'y a pas de compétition génique, nous montrons que EKLF est requis pour la transcription basale du gène et non pour son interaction avec le LCR. Nous démontrons de plus que EKLF n'induit pas la transition puisqu'il est capable de stimuler le gène  $\beta$ -globine dans le stade foetal, avant la transition, en absence de compétition génique. Finalement, malgré la liaison de EKLF au motif CACC du promoteur de  $\beta$ -globine, nous démontrons que EKLF ne stimule pas la transcription de transgènes LCR- $\beta$ -*lacZ*, suggérant que EKLF pourrait stimuler l'expression de la  $\beta$ -globine par un autre élément que le motif CACC, probablement situé dans les séquences codantes.

## **Table des matières**

Sommaire	iii
Table des matières	v
Liste des tableaux	x
Liste des figures	xi
Liste des abréviations	xiii
Remerciements	xiv
<b>Chapitre 1 - Introduction</b>	<b>1</b>
1.1 La régulation de la transcription	2
1.2 La structure de la chromatine et l'expression des gènes	3
1.2.1 Le nucléosome	4
1.2.2 Les fibres de 10 nm et de 30 nm	5
1.2.3 Repléments supérieurs	5
1.2.4 Compaction de l'ADN et expression	6
1.2.5 Disposition en phase des nucléosomes	8
1.2.6 Domaines structuraux de chromatine	9
1.2.7 Transgénèse et effets de position	9
1.2.8 Variégation de l'expression	11
1.3 Éléments de régulation de la chromatine	13
1.3.1 Sites hypersensibles	13
1.3.2 Régions d'attachement à la matrice (MARs)	14
1.3.3 Méthylation de l'ADN	15
1.3.4 L'acétylation des histones	16
1.3.5 Systèmes de régulation de la structure de la chromatine	17
1.3.6 Isolateurs	19
1.3.7 Les régions de contrôle de locus (LCRs)	20
1.4 Les globines humaines du type $\beta$	22
1.4.1 Description des globines du type $\beta$	22
1.4.2 Les hémoglobinopathies	22
1.4.3 Systèmes utilisés pour étudier la régulation transcriptionnelle des globines	23
1.4.4 Structure de la chromatine du locus de $\beta$ -globine	25
1.4.5 Régulation transcriptionnelle locale des globines du type $\beta$	26
1.4.6 Activation de la transcription des gènes de globine par le LCR	27

1.4.7	L'activité du LCR est spécifiquement érythroïde	28
1.4.8	Les gènes de globine du type $\beta$ compétitionnent pour l'activation par le LCR	28
1.4.9	Le LCR de $\beta$ -globine et la structure de la chromatine du locus	31
1.4.10	Analyse des éléments composant le LCR	32
1.4.11	La structure du transgène influence la capacité du LCR à ouvrir la chromatine	34
1.4.12	Mécanismes suggérés pour l'activité du LCR	35
	Holo-complexe et durée d'activation	35
	Modèle de l'interaction mutuelle	36
	Modèle de la propagation de l'ouverture de la chromatine	37
	Le LCR pourrait prévenir la variéation	37
1.4.13	Les protéines régulant les gènes de globine	38
	GATA-1	38
	NF-E2	38
	EKLF	39
1.4.14	Séquences du gène de $\beta$ -globine impliquées dans l'activation	40
1.5	Problématique	41
 <b>Chapitre 2 - Article 1</b>		 51
 <b>The <math>\beta</math>-globin locus control region enhances transcription of but does not confer position-independent expression onto the <i>lacZ</i> gene in transgenic mice</b>		
	Abstract	53
	Introduction	54
	Results	57
	The $\beta$ -globin LCR does not confer position-independent expression onto a <i>lacZ</i> reporter transgene	57
	Addition of the whole $\beta$ -globin gene does not restore position-independent expression to the <i>lacZ</i> reporter	59
	$\beta$ -galactosidase activity levels in extracts corresponds to <i>lacZ</i> mRNA levels	60
	The LCR does confer position-independent expression onto the $\beta$ -globin gene itself	61

In drug-resistant-selected MEL cell clones, a LCR- <i>lacZ</i> transgene shows a much closer relationship between copy number and expression	61
Discussion	63
Culture cells are not a good model for position-independent studies	63
Why does the LCR not confer position-independent expression onto the <i>lacZ</i> reporter gene in transgenic mice?	64
Materials and methods	68
Acknowledgements	71
References	72
 <b>Chapitre 3 - Article 2</b>	 86
 <b>Position effects in mice carrying a <i>lacZ</i> transgene in <i>cis</i> with the <math>\beta</math>-globin LCR can be explained by a graded model</b>	
Abstract	88
Introduction	89
Materials and methods	93
Results	95
$\beta$ -galactosidase expression is graded in transgenic mice carrying the <i>lacZ</i> gene under the control of the $\beta$ -globin LCR	95
Gene silencing, as measured by methylation of the transgene DNA, was absent in samples that had detectable $\beta$ -galactosidase activity	99
Discussion	100
Acknowledgements	103
References	104
 <b>Chapitre 4 - Article 3</b>	 113
 <b>EKLF is essential for <math>\beta</math>-globin gene expression even in absence of gene competition, but is not sufficient to induce the switch from <math>\gamma</math>-globin to <math>\beta</math>-globin gene expression</b>	
Abstract	115
Introduction	116
Materials and methods	119



Results	120
The $\beta$ -globin gene requires EKLF for expression even when it is juxtaposed to the LCR and no gene competition is present	120
EKLF function is not developmental-specific	121
Discussion	122
EKLF is not the primary factor for globin switching	122
Does EKLF play a role in LCR- $\beta$ -globin gene interactions?	123
Acknowledgements	124
References	125
<b>Chapitre 5 - Discussion</b>	<b>134</b>
5.1 Validité de l'approche expérimentale	135
5.1.1 L'activité $\beta$ -galactosidase reflète les niveaux d'expression	135
5.1.2 Les souris fondatrices mosaïques ne nuisent pas à l'interprétation des résultats	136
5.2 Le LCR ne peut conférer une expression indépendante de la position lorsqu'un autre rapporteur que le gène intact de $\beta$ -globine est utilisé	137
5.2.1 L'activité enhancer du LCR est indépendante de l'ouverture de la chromatine	138
5.2.2 Il n'y a pas d'éléments dans le gène de $\beta$ -globine dont la seule présence rétablit l'activité du LCR	138
5.2.3 <i>LacZ</i> n'agit pas comme un isolateur qui bloquerait la propagation de l'ouverture de la chromatine	139
5.2.4 Une interaction mutuelle activant l'ouverture de la chromatine par le LCR pourrait être altérée par l'insertion de séquences dans le gène de $\beta$ -globine	140
5.2.5 Une interaction mutuelle entre le LCR et le gène $\beta$ -globine est compétitionnée par le gène $\beta$ - <i>lacZ</i>	141
5.3 Variégation de l'expression et modèle d'expression graduée	143
5.3.1 Il n'y a pas de variégation de l'expression des transgènes $\beta$ - <i>lacZ</i>	143
5.3.2 Le modèle d'expression graduée et mécanisme du LCR	144
5.4 Rôle du motif CACC du promoteur de $\beta$ -globine et de EKLF	145
5.4.1 Dans les cellules K562, le motif CACC permet la transcription stimulée par le LCR	146
5.4.2 Le motif CACC est important, mais non essentiel pour l'expression du rapporteur $\beta$ - <i>lacZ</i> chez les souris transgéniques	147

5.4.3	L'expression des transgènes $\beta$ - <i>lacZ</i> est indépendante de EKLF	148
5.4.4	EKLF pourrait activer le gène $\beta$ -globine par d'autres éléments que le motif CACC	149
5.4.5	Le LCR pourrait stimuler le gène le plus actif transcriptionnellement	150
5.4.6	Différents mécanismes gouvernent l'activité de stimulation de la transcription et l'ouverture de la chromatine	150
<b>Bibliographie des Chapitres 1 et 5 et Annexes I et II</b>		152
<b>Annexe I - Article 4</b>		171
<b>Une interaction mutuelle entre le LCR et le gène de <math>\beta</math>-globine peut prévenir les effets de position si la disposition des éléments du gène est correcte</b>		
	Résumé	172
	Résultats	173
<b>Annexe II - Article 5</b>		178
<b>Fonction du motif CACC du promoteur de <math>\beta</math>-globine dans les cellules K562 et les souris transgéniques</b>		
	Résumé	179
	Résultats	180
<b>Annexe III - Article 6</b>		187
<b>CAAT/enhancer-binding proteins are involved in <math>\beta</math>-globin gene expression and are differentially expressed in murine erythro-leukemia and K562 cells</b>		
	Abstract	188
	Introduction	188
	Materials and methods	189

Results	189
Discussion	193
Acknowledgements	194
References	194

## Liste des tableaux

### Chapitre 2

<b>Table I</b>	Analysis of $\beta$ A and $\beta$ B transgenic lines	76
<b>Table II</b>	Analysis of $\beta$ C, $\beta$ D and $\beta$ E transgenic fetuses	77
<b>Table III</b>	Analysis of $\beta$ F transgenic fetuses	78
<b>Table IV</b>	Analysis of MEL clones transfected with $\beta$ D construct	79

### Chapitre 3

<b>Table I</b>	Analysis of $\beta$ -galactosidase expression and X-gal staining in transgenic mice	107
----------------	-------------------------------------------------------------------------------------	-----

### Annexe I

<b>Tableau I</b>	Expression des transgènes $\beta$ I et $\beta$ J	173
<b>Tableau II</b>	Analysis of $\beta$ -TIMP and $\beta$ K transgenic fetuses	174

### Annexe II

<b>Tableau I</b>	Expression de la $\beta$ -galactosidase chez les souris transgeniques	180
<b>Tableau II</b>	Comparaison de l'expression $\beta$ -galactosidase chez les souris transgéniques de divers génotype EKLF à 14,5 jours	181

## Liste des Figures

### Chapitre 1

<b>Figure 1</b>	Expression des globines du type $\beta$ au cours du développement	43
<b>Figure 2</b>	Locus de $\beta$ -globine et les constructions mini- et micro-locus	44
<b>Figure 3</b>	Schéma simplifié de la différenciation hématopoïétique	45
<b>Figure 4</b>	Structure du gène de la $\beta$ -globine	46
<b>Figure 5</b>	Expression indépendante de la position au locus de $\beta$ -globine de poulet	47
<b>Figure 6</b>	Effets de position sur l'expression d'un gène $\gamma$ -globine humain	48
<b>Figure 7</b>	Modèle d'interaction mutuelle	49
<b>Figure 8</b>	Modèle de propagation de l'ouverture de la chromatine	50

### Chapitre 2

<b>Figure 1</b>	Structure of the $\beta$ -globin locus and microlocus constructs	80
<b>Figure 2</b>	<i>In situ</i> detection of $\beta$ -galactosidase activity in adult red blood cells and fetal tissues	81
<b>Figure 3</b>	Copy number determination	82
<b>Figure 4</b>	Hypersensitive site determination	83
<b>Figure 5</b>	RNase protection analysis of <i>lacZ</i> mRNA expression	84
<b>Figure 6</b>	S1 nuclease protection analysis of $\beta$ -globin gene expression in mice carrying the $\beta$ -globin gene without <i>lacZ</i> sequences	85

### Chapitre 3

<b>Figure 1</b>	Kinetics of appearance of stained cells <i>in situ</i>	108
<b>Figure 2</b>	The fraction of detectable cells and the intensity of blue coloration increases with time	109
<b>Figure 3</b>	Relationship between the average expression per cell and the total expression	110
<b>Figure 4</b>	Methylation of the <i>lacZ</i> transgene DNA in different transgenic lines	111
<b>Figure 5</b>	Theoretical curves of distribution of expression	112

### Chapitre 4

<b>Figure 1</b>	Structure of the human $\beta$ -like globin locus and the	130
-----------------	-----------------------------------------------------------	-----

	$\beta$ -globin gene minimal promoter	
<b>Figure 2</b>	Genotype determination	131
<b>Figure 3</b>	EKLF is required for $\beta$ -globin transgene expression in the adult stage	132
<b>Figure 4</b>	EKLF is required for $\beta$ -globin transgene expression in the embryonic stage	133

### Annexe I

<b>Figure 1</b>	Constructions	175
<b>Figure 2</b>	Analyse de l'expression des transgènes $\beta$ I and $\beta$ J chez les souris transgéniques	176
<b>Figure 3</b>	Analyse de l'expression du transgène $\beta$ K chez les souris transgéniques	177

### Annexe II

<b>Figure 1</b>	Constructions incluant des délétions du promoteur de $\beta$ -globine	182
<b>Figure 2</b>	Analyse de l'expression avec les différentes délétion du promoteur $\beta$ -globine dans les cellules K562	183
<b>Figure 3</b>	Constructions de diverses combinaisons de motifs CACC et CAAT	184
<b>Figure 4</b>	Analyse de l'expression avec les différentes combinaison des motifs CACC et CAAT	185
<b>Figure 5</b>	Constructions analysées chez les souris transgéniques	186

### Annexe III

<b>Figure 1</b>	Structure of the human $\beta$ -globin gene complex	189
<b>Figure 2</b>	DSF1 and DSRr are C/EBP family members	190
<b>Figure 3</b>	DSF1 and DSRr specifically bind to a C/EBP recognition sequence	191
<b>Figure 4</b>	A different form of C/EBP $\beta$ is expressed in K562 cells	192
<b>Figure 5</b>	K562 cells contain LIP	193
<b>Figure 6</b>	RNase protection analysis of transfected MEL cell populations	193
<b>Figure 7</b>	Both CP1 and C/EBP $\gamma$ are required for wild type levels of expression	194
<b>Figure 8</b>	GMS assays of promoter fragments	194

**Liste des abréviations**

- C/EBP : CAAT/enhancer-binding protein  
CPS : carbamoyl phosphate synthetase  
*dt* : *dystonia musculorum*  
EKLF : erythroid Krüppel-like factor  
GMS : retardement sur gel (Gel Mobility Shift assay)  
HSS : site hypersensible (HyperSensitive Site)  
LCR : région de contrôle de locus (Locus Control Region)  
MEL : érythroleucémie murine (Murine ErythroLeukemia)  
nt : non transgénique  
PBS : saline tamponnée au phosphate (Phosphate-Buffered Saline)  
SV40 : Simian virus 40  
tg : transgénique  
X-gal : 5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactoside

## Remerciements

Je tiens d'abord à remercier sincèrement mon directeur de doctorat le Dr. Lee Wall. Le soutien du Dr Wall, sa vision et ses idées ont grandement contribué à mon succès. J'ai particulièrement apprécié la liberté d'action qu'il m'a toujours donné de même que la grande confiance qu'il me portait. Merci également au Dr. Rashmi Kothary dont le savoir-faire et la collaboration ont été essentiels.

Je voudrais de plus remercier ma famille qui m'a toujours épaulé, enduré et encouragé. La patience, la compréhension et le support de ma compagne Marie-Claude Ouimet furent très importants; merci Marie. Je dois aussi remercier mes parents pour m'avoir donné un environnement propice, une soif de connaissances et d'excellence. Merci également pour tout le support durant mes études.

Finalement, j'aimerais souligner l'aide que j'ai reçu de plusieurs collaborateurs qui m'ont aidé dans le laboratoire ou dans l'animalerie, qui m'ont fourni des protocoles, des plasmides ou des souris ou enfin qui m'ont aidé pour la rédaction de divers rapport. Merci à André Toulouse, Andrew Perkins, Anne-Marie Mes-Masson, Arthur Brown, Benoit Houle, Dan Skup, David Donze, Daniel Houle, Denis Cournoyer, Gabriel Lepage, Gilbert Bernier, Graham Dellaire, Gratien Dalpé, Guy Charron, James Ellis, Jean-Pierre Julien, Karynn Veillette, Lee Wall, Luc Aubry, Marie-Claude Ouimet, Majid Belouchi, Mario Fillion, Martine Mathieu, Maryse Lamoureux, Maryse Mercier, Nathalie Delvoye, Nathalie Destroismaisons, Pascal Vachon, Qi Mei, Rashmi Kothary, Sandra Creigneur, Stéphan Denis, Stéphane Parent, Stuart Orkin, Ted Bradley, Tim Townes, Yves DeRepentigny et à tous ceux que j'oublie...

## **Chapitre 1**

### Introduction



## **INTRODUCTION:**

### **1.1 La régulation de la transcription**

Plusieurs processus biologiques, comme par exemple la différenciation cellulaire, le développement ou la transformation oncogénique, requièrent la modulation de molécules spécifiques. Pour ce faire, divers moyens ont été sélectionnés à presque tous les étapes de la production de protéines actives. Au début du processus, il y a le contrôle de la transcription qui permet par exemple l'expression des neurofilaments uniquement dans les neurones (1). La régulation de l'épissage permet d'obtenir des protéines légèrement différentes au besoin. Le gène *doublesex* de drosophile est épissé différemment chez le mâle et la femelle, résultant en une protéine mâle bloquant la différenciation femelle ou en une protéine femelle bloquant l'expression de gènes spécifiquement mâles (2). Parfois, on rencontre une régulation de la stabilité du messager qui permet par exemple de prolonger la durée de vie des ARNm de globine dans les érythrocytes (3). Il y a aussi des mécanismes qui contrôlent la traduction, comme le cas du transcrit *hunchback* chez la drosophile. La traduction de ce messager, est inhibée par le produit de *nanos* qui est présent selon un gradient postérieur-antérieur dans l'embryon de la mouche. Il y a donc traduction différentielle, selon un gradient antérieur-postérieur, de *hunchback* (2). Il peut aussi y avoir régulation de la stabilité des protéines; par exemple, c-fos, qui doit avoir une action très ponctuelle, est très rapidement dégradé (4). On rencontre aussi des modifications post-traductionnelles comme la phosphorylation de pRB qui permet la progression dans le cycle cellulaire (5). Enfin, des protéines peuvent être régulées par association comme NF- $\kappa$ B, inhibé par I- $\kappa$ B qui le retient dans le cytoplasme (6).

La **régulation transcriptionnelle** comporte de nombreux avantages sur les autres processus, dont voici quelques exemples. a) Cette approche requiert moins d'énergie. En effet, la régulation au début du processus évite des dépenses inutiles puisque la transcription, la traduction et les modifications post-traductionnelles sont coûteuses en énergie. b) La régulation transcriptionnelle permet de moduler de façon très fine et variée plusieurs gènes à la fois. Comme l'expression des gènes est contrôlée par

un grand nombre de facteurs protéiques plus ou moins spécialisés et que l'interrelation entre ces facteurs peut moduler leur activité, l'expression finale de chaque gène dépendra de la combinaison des facteurs présents. Les messages cellulaires (comme les voies de signalisation intracellulaire) affectant l'activité de ces facteurs pourront alors influencer plusieurs gènes à la fois et la réponse pour chacun de ces gènes sera spécifique. Ainsi, la cellule pourra, par de subtils changements, moduler l'expression de toute une série de gènes d'une façon précise, spécifique et ordonnée. c) Il y a amplification du message lorsqu'un signal se traduit par l'activation d'un gène. Prenons l'exemple de hsp68, une protéine anti-stress induite par un choc thermique. Lorsque la cellule subit un stress, il y a immédiatement un signal qui stimule la transcription du gène hsp68 (7). Dans les minutes qui suivent, plusieurs rondes de transcription se produisent, générant plusieurs exemplaires d'ARNm pour hsp68. Par la suite, chacun de ces messagers sera traduit plusieurs fois en protéine. Finalement, chacune de ces protéines pourra avoir une action de chaperonine et ce à plusieurs reprises. Ainsi, le signal unique au niveau de la transcription sera amplifié un très grand nombre de fois pour avoir une réponse cellulaire très importante. La régulation transcriptionnelle comporte aussi des désavantages, comme par exemple la lenteur du temps de réponse. En effet, la modulation transcriptionnelle d'un gène pourra prendre jusqu'à 20 minutes avant de modifier sensiblement la quantité du produit du gène.

Des dérèglements de la transcription peuvent avoir des conséquences dramatiques. D'ailleurs, plusieurs maladies en sont la conséquence. Par exemples, certains cancers résultent d'une surexpression d'un proto-oncogène, certains types d'hémophilie sont induits par une sous-expression de facteurs de coagulation et des thalassémies sont causées par une diminution de la transcription d'un gène de globine. L'étude des mécanismes impliqués dans la régulation transcriptionnelle, autant dans les situations "normales" que dans les maladies, permettra d'atteindre une meilleure compréhension de cet important processus cellulaire.

## 1.2 La structure de la chromatine et l'expression des gènes

La transcription à l'intérieur du noyau ne s'effectue pas aussi simplement que dans un essai *in vitro*. En effet, l'ADN y est assemblé en complexe avec des protéines, puis enroulé et compacté. La machinerie transcriptionnelle doit composer avec cette réalité et la structure de la chromatine aura une influence importante sur l'expression des gènes.

### 1.2.1 Le nucléosome

La chromatine de tous les eucaryotes apparaît comme une masse de petites particules compactes: les **nucléosomes**. Au coeur de ce complexe se trouve un bloc d'**histones** formant un octamère constitué de deux copies des histones H2A, H2B, H3 et H4 (8). L'ADN nucléaire se love sur ce complexe: 146 pb font environ deux tours autour de chaque bloc d'histones. Ensuite, une courte région d'ADN libre (~60 pb) sert de lien entre un nucléosome et le suivant; on retrouve donc un nucléosome toute les 200 pb environ. La longueur de ce lien peut varier d'un type cellulaire à l'autre. L'histone H1 peut venir se lier sur ce complexe pour stabiliser des structures supérieures. Le nucléosome ressemble alors à une bille d'environ 6 nm par 1 nm constitué à parts presque égales d'ADN et d'histones (250-300 kDa au total). L'ADN assemblé en nucléosome est environ 6 fois plus compacté que l'ADN nu et linéaire. L'ADN est toujours accessible aux facteurs transcriptionnels et à la machinerie transcriptionnelle puisqu'il est en périphérie du nucléosome. Toutefois, il est plus encombré et cela se reflète par exemple sur le patron de digestion aux nucléases: l'ADN lié à la particule centrale n'est pratiquement pas coupé par ces enzymes. De plus, le nucléosome pose un problème stérique à la transcription puisque l'ADN polymérase est de taille comparable à un nucléosome. Il y a donc possibilité de transcription malgré les nucléosomes mais dans des conditions plus difficiles (9,10).

Les histones de la particule centrale sont des protéines très fortement conservées chez toutes les espèces eucaryotes, ce qui suggère que leur fonction est également très similaire. L'histone H1 montre une plus grande variabilité, elle est même absente chez les

levures (2). On retrouve aussi d'autres protéines que les histones dans la chromatine mais elles sont beaucoup plus rares. Il est possible que ces protéines soient impliquées dans les repliements supérieurs de la chromatine (2).

### 1.2.2 Les fibres de 10 nm et de 30 nm

Dans certaines conditions, l'observation de la chromatine au microscope électronique démontre un fil perlé d'environ 10 nm de diamètre (11,12). Ce qui ressemble à un fil est en fait l'ADN faisant le lien entre les nucléosomes, eux-mêmes apparaissant comme les perles. L'histone H1 n'est pas requise pour obtenir cette **fibre de 10 nm**. En augmentant la force ionique et en présence de l'histone H1, on peut obtenir une fibre d'environ 30 nm de diamètre (13). Cette structure est formée à partir de la fibre de 10 nm qui s'enroule de façon à former une hélice solénoïde de nucléosomes. Environ 6 nucléosomes composent chaque tour de l'hélice. Le facteur de compaction de l'ADN passe alors à ~40 (si on déroulait l'ADN d'une **fibre de 30 nm** d'une longueur donnée, on obtiendrait une hélice d'ADN 40 fois plus longue).

### 1.2.3 Repliements supérieurs

Le degré de compaction de l'ADN dans le noyau est estimée à plusieurs centaines ou même quelques milliers de fois. Il faut donc que des **repliements supérieurs** viennent s'ajouter aux fibres de base. Plusieurs auteurs décrivent des structures ressemblant à des super-billes qui pourraient être formé par un enroulement de fibre de 30 nm (14,15).

Il faut toutefois faire attention et ne pas prendre ces modèles pour des représentations fidèles de la réalité. En effet, les fibres de 10 nm et de 30 nm sont obtenus *in vitro* et pourraient être des artefacts de préparation. D'autres études montrent des structures passant d'un agencement en zigzag ouvert à un zigzag fermé et finalement à une structure hélicoïdale à mesure que la force ionique augmente (16,17). Plusieurs modèles ont été proposés pour chercher une structure organisée mais une analyse poussée des images générées avec de la chromatine *in vivo* indique qu'il n'y a pas de preuves solides pour une structure organisée, hélicoïdale ou autre. Un des modèles les plus probables suggère que la taille du lien d'ADN influence la rotation du nucléosome par rapport au

précédent (18). *In vivo*, cette taille n'est pas stable, ce qui fait qu'aucune structure unifiée n'est obtenue. Au mieux obtient-on de courtes régions hélicoïdales ouvertes tournant à droite, d'autres tournant à gauche et des régions plus irrégulières. Le degré de compaction dans cette fibre irrégulière n'atteindra probablement pas 40 puisque le repliement n'est pas aussi parfait que dans la fibre de 30 nm. Toutefois, il devrait être bien supérieur à celui de la fibre de 10 nm. Une compaction axiale de cette fibre résulterait en une structure en accordéon irrégulier qui pourrait ressembler aux fameuses super-billes. La compaction de l'ADN pourrait alors monter à plusieurs centaines de fois. Peu importe le modèle, la structure supérieure réelle de la chromatine est loin d'être résolue. Toutefois, il faut bien comprendre que les repliements supérieurs induisent une forte compaction de l'ADN qui réduit son accessibilité.

Le degré de compaction n'est pas identique pour tout le génome. En observant la chromatine *in vivo*, deux états distincts peuvent être décrits (2): l'euchromatine et l'hétérochromatine. L'**euchromatine** est plus relâchée et est dispersée dans la plus grande partie du noyau. On estime que le degré de repliement de l'ADN y est près de 1000. Par contre, l'**hétérochromatine** est très dense, et présente des similarités avec l'ADN de chromosome en métaphase. Cette structure se retrouve en petits agrégats près de la membrane nucléaire et du nucléole. L'ADN y est près de 10 000 fois plus compacté que sous sa forme linéaire. Entre ces classifications grossières, la chromatine peut adopter une multitude de structures intermédiaires. En traitant des noyaux avec de l'uracile tritiée, l'incorporation du ribonucléotide radioactif, qui indique la transcription, se fait principalement dans l'euchromatine. L'euchromatine comprend donc les séquences transcrites alors que l'hétérochromatine est formée de séquences non-transcrites, de séquences répétées et, dans les cellules femelles, du chromosome X inactivé. L'euchromatine est donc plus active et elle est répliquée tôt dans le cycle cellulaire tandis que l'hétérochromatine est plus inerte et est répliquée tardivement.

#### 1.2.4 Compaction de l'ADN et expression

Dans certaines conditions, un plasmide assemblé sous forme de nucléosomes est moins bien transcrit *in vitro* qu'un plasmide libre (19-21). Cela peut s'expliquer par la disponibilité de l'ADN aux facteurs de la machinerie transcriptionnelle qui est réduite en présence des histones. L'ajout de certains facteurs de transcription avant les histones permet la transcription (22). Ces facteurs pourraient donc lier l'ADN et prévenir l'assemblage de nucléosomes sur cette région, permettant l'expression. D'autres facteurs ont la capacité de déplacer les nucléosomes déjà assemblés. Il est donc évident que la transcription peut avoir lieu malgré les histones, toutefois ces observations illustrent comment la compaction de l'ADN peut interférer avec la transcription. À l'intérieur de l'hétérochromatine, l'ADN est beaucoup plus compacté qu'un plasmide simplement recouvert de nucléosome. La structure de la chromatine a donc le potentiel de nuire fortement à la transcription. Les gènes se retrouvant sous une forme de chromatine plus compacte seront de toute évidence plus difficilement accessibles à la machinerie transcriptionnelle et ne pourront donc être transcrits aussi facilement.

Puisque la structure de la chromatine influence l'expression, il est utile de pouvoir la caractériser expérimentalement. On peut estimer le degré de compaction de la chromatine en testant la disponibilité de l'ADN nucléaire à une digestion enzymatique. Au cours de cette expérience, des noyaux ou des cellules sont directement traités à l'ADNase I ou d'autres nucléases. Les séquences fortement compactés dans la chromatine seront plus difficile d'accès et donc plus résistantes à la digestion enzymatique. Par contre, les séquences comprises dans la chromatine plus relâchée seront plus rapidement digérées. En estimant la quantité de nucléases ou le temps requis pour faire disparaître un fragment spécifique, nous pouvons donc évaluer la densité relative de la chromatine à ce locus. Cette technique mesure la **sensibilité générale** aux nucléases et elle permet de comparer le degré de compaction de régions de quelques kb (23). Il y a autant de nucléosomes dans les régions de chromatine à sensibilité générale que dans les régions plus compactées. La sensibilité générale n'est donc pas causée par la perte de nucléosomes, mais plutôt par le

déroulement des structures de repliements supérieurs ou alors par des contacts moins serrés entre l'ADN et les histones. Par exemple, des modifications sur les histones (voir 1.3.5) pourraient induire de légers changements de conformation du nucléosome pour que l'ADN soit plus exposé (24). Une autre possibilité est que l'histone H1 se dissocie de la chromatine à haute sensibilité générale, favorisant un déroulement des structure supérieures (25).

L'étude de plusieurs loci a démontré qu'une sensibilité générale élevée était toujours associée avec les régions transcrites. Par exemple, chez le saumon, l'administration de 17- $\beta$ -estradiol induit la production de vitellogénine (26). Le traitement augmente la sensibilité générale au locus de la vitellogénine uniquement dans le foie qui est le seul tissu exprimant ce gène. On peut aussi noter l'augmentation de la sensibilité générale du locus de  $\beta$ -globine qui accompagne le début de l'expression de ce gène lors de la différenciation des cellules MEL (27). De plus, au cours de la différenciation ou de l'induction de gènes, le relâchement de la chromatine précède l'expression de ces gènes (28). La sensibilité générale d'un locus, ou la faible densité de compaction de la chromatine, semble donc requise, quoique insuffisante, pour la transcription.

### 1.2.5 Disposition en phase des nucléosomes

Les octamères d'histones pourraient lier préférentiellement certaines séquences d'ADN. La disposition des nucléosomes à un locus donné n'est donc pas totalement aléatoire et certains nucléosomes peuvent se retrouver en phase dans toutes les cellules (23). Le phénomène pourrait s'expliquer par la propriété que possèdent certaines séquences d'ADN de s'enrouler (DNA bending) et donc de se retrouver préférentiellement au même endroit sur l'octamère d'histone: là où il adopte une structure thermodynamiquement favorable. Ce biais n'est toutefois pas très important et la disposition des nucléosomes est relativement aléatoire sur la plupart des loci. Toutefois, certains loci démontrent des nucléosomes parfaitement en phase *in vivo* avec une grande précision. Ce phénomène pourrait s'expliquer par la liaison de facteurs de transcription à des séquences spécifiques qui induisent une disposition précise du nucléosome à proximité. Ce nucléosome se

retrouvera donc à une position similaire dans toutes les cellules et les nucléosomes suivants seront alignés sur le premier. On trouvera donc, sur une région de quelques centaines ou milliers de paires de bases, une position des nucléosomes non aléatoire. L'absence de nucléosome sur une séquence donnée (site hypersensible, voir 1.3.1) pourrait aussi favoriser une localisation précise du nucléosome adjacent et induire une disposition en phase des nucléosomes suivants.

La liaison en phase des nucléosomes a été caractérisée avec le système *PHO* chez la levure (29a). *PHO4* est un facteur de transcription qui répond à un manque de phosphate en induisant la transcription du gène *PHO5*. En présence de phosphate, il y a un alignement précis de 4 nucléosomes sur le promoteur de *PHO5* et liaison de *PHO4* à un site entre deux nucléosomes. Puisque les nucléosomes sont en phase, la digestion de la chromatine par des nucléases donnera alors un patron de digestion très précis. Le retrait du phosphate du milieu de culture active la protéine *PHO4* qui induit alors la transcription du gène *PHO5*. Le patron de digestion est alors perdu, indiquant que l'organisation des nucléosomes est alors changée. Un autre exemple peut être donné au locus de l'albumine, où un enhancer distal est recouvert de nucléosomes en phase dans le foie, le tissu où l'enhancer est actif, mais non dans les autres tissus (29b). La régulation de la transcription est donc liée à l'organisation de la chromatine. Il existe des exemples de changements dans la structure de la chromatine en relation avec la transcription dans le développement, dans la spécificité tissulaire, dans la différenciation et dans l'induction génique.

#### 1.2.6 Domaines structuraux de chromatine

L'extraction des protéines solubles, incluant les histones, de chromosomes en métaphase révèle de grandes boucles d'ADN reliées au squelette protéique des chromosomes (l'échafaud) selon un intervalle d'environ 30 à 200 kb. (30) De la même façon, l'ADN en interphase est joint aux protéines de la matrice nucléaire - une structure protéique reliée à la membrane nucléaire interne en interphase - selon des boucles similaires. Chacune de ces boucles constitue un **domaine structurel**. Une théorie couramment rencontrée est que



chacune de ces boucles constitue une unité transcriptionnelle régulée indépendamment des autres. Cette théorie expliquerait comment les enhancers stimulent leur gène cible, parfois localisés plusieurs dizaines de kb de distance, et non les gènes des loci à proximité (voir MARs 1.3.2).

### 1.2.7 Transgénèse et effets de position

L'influence de la structure de la chromatine sur l'expression des gènes est révélée par la **transgénèse**. Cette technique consiste à introduire un fragment d'ADN, le transgène, dans un noyau de zygote (31a). Dans certains cas, l'ADN sera intégré dans le génome de la cellule hôte. L'insertion se fait au hasard dans le génome et un très grand nombre de sites d'intégration différents peuvent être obtenus. Lorsque le zygote est réimplanté dans une mère porteuse, le développement reprend normalement pour former un animal transgénique pouvant transmettre le transgène à ses descendants. Les drosophiles transgéniques sont régulièrement utilisées, entre autres avec le gène *white* dont l'expression est aisément détectable par la couleur des yeux de l'insecte (31b). Les souris transgéniques constituent un autre modèle courant. Dans la majorité des cas, plus d'une copie du transgène sera intégrée. La plupart du temps, les différentes copies sont intégrées au même locus et sont disposées les unes à la suite des autres, tête contre queue. Il peut arriver que l'intégration survienne après la première division cellulaire. Si tel est le cas, la souris fondatrice sera mosaïque, c'est-à-dire que seulement une fraction de ses cellules portent le transgène. Après accouplement de la souris fondatrice, il est possible d'obtenir des animaux normalement transgéniques si la gamète à l'origine de la F1 porte le transgène qui sera alors transmis à toutes les cellules de l'organisme. L'expression de la plupart des transgènes varie beaucoup d'une lignée de souris transgénique à l'autre. Ce phénomène, appelé **effets de position**, serait causé par l'influence de la structure de la chromatine au site d'intégration qui pourrait réduire et quelquefois stimuler la transcription du transgène (32). Puisqu'un très grand nombre de sites d'intégration différents sont possibles, la structure de la chromatine sera diversifiée d'une insertion à l'autre et son effet sur la transcription sera donc variable. L'étude de l'expression de transgènes constitue

donc un test fonctionnel pour déterminer l'effet de la structure de la chromatine sur l'activité transcriptionnelle.

### 1.2.8 Variégation de l'expression

Chez la drosophile et la levure, certaines translocations placent des gènes près d'une région d'hétérochromatine (comme le centromère ou les télomères), alors que ces gènes sont normalement dans des régions d'euchromatine. Dans certains cas, une partie des cellules vont perdre l'expression du gène en question alors que les autres cellules conserveront une expression normale (à des niveaux similaires à ceux du locus normal avant la translocation). Ce phénomène est dû à une propagation de compaction de l'ADN dans la région du transgène à une vitesse variable d'une cellule à l'autre (33,34). Dès que le gène se trouve sous forme d'hétérochromatine il devient silencieux. Puisque l'inactivation ne se fait pas au même moment dans toutes les cellules, il y aura expression hétérocellulaire ou **variégation** de l'expression. Plusieurs mutations peuvent stimuler ou empêcher la variégation (35). Certaines de ces mutations sont localisées directement dans des gènes d'histone (36) et d'autres mutations impliquent des protéines qui pourraient jouer un rôle dans le repliement de la chromatine ou l'acétylation des histones (37). Le lien entre la structure de la chromatine et la variégation semble donc être réel. La définition fonctionnelle stricte de la variégation est une expression hétérocellulaire bimodale avec des niveaux de transcription soit nuls dans les cellules silencieuses, soit équivalents (maximales) dans les cellules qui expriment. Par exemple, on peut faire un rapprochement avec un interrupteur qui peut permettre des niveaux selon une distribution bimodale: un niveau fixe ou alors pas d'expression du tout. Lorsque le gène *white* est soumis à de la variégation, on notera un oeil composé de cellules rouges et de cellules blanches à la fois mais non de couleurs intermédiaires (38a). Par opposition, en absence de variégation, les effets de position sur le gène *white* produiront une coloration uniforme de l'oeil selon un gradient entre le blanc (faible expression) et le rouge (forte expression) (38b).

Chez les mammifères, seulement quelques articles rapportent l'observation de phénomènes s'apparentant à la variégation de l'expression. Dans le cas du transgène de la

$\beta$ -lactoglobuline, dont l'expression dans les glandes mammaires révélée par hybridation *in situ* à l'ARNm démontrait quelques cellules n'exprimant pas le transgène (39). Il n'était toutefois pas clair si les niveaux d'expression dans les cellules positives étaient équivalents (distribution bimodale ou polymodale). D'autres études utilisent la cytométrie de flux, qui permet de séparer les cellules selon l'intensité d'un signal d'immunofluorescence. Des effets de position de transgènes CD2 peuvent par exemple révéler l'existence d'une population de cellules exprimant le transgène et d'une population silencieuse (40). Toutefois, cette technique se fait dans des conditions où le substrat, soit l'anticorps, est limitant. La mesure de la quantité d'antigène par cellule est donc biaisée, puisqu'au delà d'une certaine limite, il n'y aura pas suffisamment d'anticorps pour détecter la différence. De plus, ces appareils requièrent plusieurs réglages, comme par exemple une limite inférieure de détection. La présence d'un pic de cellules en apparence silencieuses pourrait donc être due à une faible expression, en deçà de la limite de détection. Enfin, un autre type d'expériences implique l'utilisation de transgènes *lacZ* dont l'expression peut être détectée dans des cellules individuelles par un test enzymatique sur une coupe histologique ou des cellules en suspension (41). Ces tests révèlent parfois une coloration dans seulement quelques cellules de la population totale, comme dans le cas d'un transgène  $\zeta$ -globine-*lacZ*. Il est donc possible qu'il y ait variéation pour plusieurs transgènes.

Walters et ses collaborateurs ont développé un modèle dérivé de la variéation de l'expression (42). En transfectant un gène avec ou sans enhancer, ils ont noté que l'enhancer n'augmentait pas les niveaux d'expression par cellule mais plutôt la quantité de cellules qui expriment. Leur interprétation de ces résultats était que les enhancers ne stimulent pas le taux de transcription, c'est-à-dire la densité d'ARN polymérase sur le gène, mais plutôt la probabilité qu'il y ait expression. Ce **modèle binaire** suggère que le niveau d'expression serait fixe, peu importe la position ou les éléments de régulation présents dans le transgène. Les enhancers ne pourraient pas influencer ce niveau mais pourraient par contre favoriser la formation d'une matrice de transcription active, influençant cet "interrupteur" pour favoriser la position "allumé". Toutefois, ces auteurs se basent sur des techniques qui biaisent leurs résultats. La sélection pour la résistance à

des drogues conférée l'intégration du plasmide rejette toutes les clones où l'expression serait faible en raison de la structure de la chromatine au site d'insertion. Donc, seuls les quelques clones où l'expression est élevée seront sélectionnés. Si l'expression est plus ou moins distribuée selon un gradient d'intensité, alors il y aura toujours quelques clones qui atteindront ces niveaux. Avec la présence d'enhancer, l'expression moyenne sera plus élevée et plus de clones seront sélectionnés. Le modèle binaire pourrait donc être erroné.

### **1.3 Éléments de régulation de la chromatine**

#### **1.3.1 Sites hypersensibles**

Parfois, de très courtes séquences (quelques centaines de bases) sont hypersensibles à la digestion de la chromatine par l'ADNase I, c'est-à-dire qu'elles seront coupées beaucoup plus rapidement que les régions de sensibilité générale. L'apparition de ces **sites hypersensibles** (43) est régulée au cours du développement, dans chaque type cellulaire et selon l'induction des gènes. Plusieurs sites hypersensibles sont associés à la transcription de gènes à proximité. Par exemple, le site hypersensibles situé dans l'enhancer à 3 kb en 5' du gène CD34, un marqueur des cellules précurseur hématopoïétique, n'est présent que dans ce type cellulaire et disparaît dès que la cellule s'engage dans une voie de différenciation (44). De façon similaire, l'induction de l'hormone glycoprotéine par l'AMPC provoque l'apparition d'un site hypersensible au locus codant pour cette protéine (45). Dans plusieurs cas, la liaison de facteurs de transcription à l'ADN induit les sites hypersensibles. Pour recréer *in vitro* un site hypersensible du locus de  $\beta$ -globine de poulet sur un plasmide, il faut ajouter un extrait protéique de globules rouges avant les histones (2). Si l'extrait est ajouté après les histone, aucun site hypersensible n'apparaît, suggérant que les protéines de l'extrait doivent lier l'ADN pour créer le site. Des expériences *in vivo* ont aussi été faite sur des souris transgéniques et des cellules en culture. Par exemple, la liaison des facteurs NF-E2 et GATA-1 au site hypersensible HSS4 du locus de  $\beta$ -globine humain est requise pour l'hypersensibilité (46,47). Le patron de digestion suggère qu'un nucléosome a été déplacé par la liaison de ces facteurs. Il est d'ailleurs généralement admis que la plupart des sites

hypersensibles sont dépourvus de nucléosomes. Par exemple, dans le minichromosome de SV40, il y a un site hypersensible et on peut directement voir en microscopie électronique que cette région est libre de nucléosome (48). L'étude des sites hypersensibles a démontrée que la plupart d'entre eux sont associés avec des régions régulatrices comme des promoteurs, des enhancers, des origines de réplication ou des régions de contrôle de locus (LCR) (43).

### 1.3.2 Régions d'attachement à la matrice (MARs)

Des régions d'attachements à la matrice (**MARs**) ou à l'échafaud (**SARs**) peuvent être caractérisées biochimiquement *in vitro* comme étant des séquences d'ADN liant les protéines de ces structures (voir 1.2.6) (30). Il est aussi possible d'isoler les séquences liées à la matrice *in vivo*. Les séquences pouvant lier la matrice selon ces deux techniques seront considérées comme MARs. Plusieurs MARs sont à la fois des SARs, il est donc possible qu'il s'agisse des mêmes éléments. Pourtant, la matrice et l'échafaud sont des structures très différentes. Elles ont toutefois quelques caractéristiques en commun, comme leur richesse en topoisomérase II. Il n'y a pas de séquence consensus pour les MARs, mais on y retrouve souvent des séquences riches en A/T, des palindromes, des éléments de déroulement de l'ADN, des origines de réplication, de l'ADN courbe et des sites hypersensibles. Plusieurs protéines lient les MARs comme SAT1B, CDP/Cux et hnRNP-U mais elles sont encore peu caractérisées. La fonction biologique réelle des MARs reste à élucider. Dans plusieurs cas ces éléments aident à l'expression des transgènes, suggérant qu'ils jouent un rôle dans la régulation de la transcription. Souvent, les MARs contribuent à la stimulation d'enhancers indépendants sans toutefois avoir d'activité de stimulation par elles-mêmes. C'est le cas du MAR et de l'enhancer du gène de la chaîne lourde de l'immunoglobine  $\mu$  qui doivent entrer en synergie pour stimuler le gène (49). Les MARs du gène de la kératine 18 peuvent prévenir les effets de position, probablement en empêchant la propagation d'hétérochromatine (50). Toutefois, il faut faire attention en interprétant ces résultats puisque d'autres éléments peuvent avoir cette activité. Par exemple, Stief *et al.* ont publié que les MARs du lysozyme de poulet pouvait protéger le gène des effets de position (51). Toutefois, leur construction incluait une

région de contrôle de locus (LCR) (52) (voir 1.3.7) qui a aussi cette activité. Plusieurs autres MAR n'ont pas vraiment d'activité. Par exemple, les MARs de KIT ne protègent pas contre les effets de position (53). Les MARs pourraient donc être des frontières structurelles dont l'effet réel sur la transcription est encore mal compris.

### 1.3.3 Méthylation de l'ADN

Dans les génomes de vertébrés, entre 2 et 7% des cytosines sont méthylées (54). L'étude de la **méthylation de l'ADN** démontre qu'elle se retrouve presque toujours dans des dinucléotides 5'-CG-3'. Ce biais ne semble pas une coïncidence puisque 50 à 90% des CG sont méthylés (55). Or, la fréquence du dinucléotide CG dans le génome est cinq fois plus petite que celle prévue par une distribution au hasard des nucléotides et ces CG sont largement localisés en 5' de gènes (56). Il est envisageable que la méthylation des îlots CG régule la transcription de ces gènes. Cette hypothèse a pu être vérifiée expérimentalement: 1) Le traitement de l'ADN nucléaire avec la nucléase micrococciale relâche premièrement de l'ADN hypométhylé (2). Conséquemment, les régions les plus accessibles (à haute sensibilité générale) et donc transcrites sont enrichies en ADN hypométhylé. 2) Des études de transfection ont démontré que la méthylation pouvait inhiber l'expression d'adénovirus 12 (57). Après transfection avec de l'ADN viral, seules quelques cellules exprimaient les gènes viraux. Il a pu être montré que l'ADN viral était hypométhylé dans les cellules qui exprimaient et hyperméthylé dans les autres. De plus, la transfection d'ADN méthylé *in vitro* résultait toujours en une absence d'expression. 3) L'induction de la déméthylation par diverses techniques (traitement au 5-azacytidine qui remplace les cytidine dans l'ADN mais ne peut pas être méthylé, des anti-sens qui inactivent l'ADN méthyl transférase (58) ou encore la transfection de v-Ha-Ras qui induit une cascade qui inactive le même enzyme (59)) provoque l'expression de plusieurs gènes qui étaient auparavant silencieux, comme par exemple MyoD (60) ou C21 (58). Ces changements peuvent provoquer plusieurs modifications cellulaires comme la différenciation ou la transformation.

Il est possible de mesurer la méthylation de l'ADN à un locus précis par le patron de digestion d'enzymes de restriction sensibles à la méthylation. Il a été noté que les gènes exprimés sont hypométhylés par rapport à des séquences silencieuses. Par exemple, l'ADN du gène ovalbumine est méthylé dans les tissus n'exprimant pas le gène et hypométhylé dans l'oviducte (61). D'ailleurs, l'hétérochromatine, incluant les télomères, les centromères, les séquences répétitives et le chromosome X inactivé, est plus riche en séquences méthylées (2). De façon générale, on note une hyperméthylation accompagnant l'inactivation d'un gène, mais il n'est pas clair si la méthylation induit l'inactivation ou si elle se produit par la suite.

La répression de la transcription par la méthylation de l'ADN pourrait se faire par au moins deux mécanismes. Premièrement, l'interaction entre plusieurs facteurs de transcription et leur site de liaison sur l'ADN est inhibée par la méthylation du site. C'est par exemple vrai pour CREB (62) et le facteur de transcription major late d'anéovirus (2). Ce n'est toutefois pas toujours le cas: Sp1 peut lier son site qu'il y ait méthylation ou non. Le deuxième mécanisme fait intervenir des protéines spécialisées: MeCP1 et MeCP2 (63). Ces protéines peuvent lier les CG méthylés et, de ce fait, inhibent la transcription.

La méthylation de l'ADN est aussi utilisée comme marqueur d'allèle lors de l'empreinte génétique parentale de certains gènes (64). Dans ces rares cas, la méthylation est parfois présente sur l'allèle exprimée, parfois sur l'allèle silencieuse et elle peut même varier d'une région du gène à l'autre. La régulation de la méthylation est donc une façon pour la cellule de réguler l'expression génique.

#### 1.3.4 L'acétylation des histones

A l'intérieur de la chromatine, les histones sont elles aussi affectées par des modifications covalentes. Pour l'histone H1, responsable de l'enroulement de la fibre de 10 nm en structure plus complexe, il y a phosphorylation juste avant la mitose et déphosphorylation juste après (65). Il est possible que cette modification induise en partie la condensation des chromosomes. Pour les autres histones, les modifications se produisent principalement sur des résidus lysine, normalement chargés positivement.

Après méthylation ou **acétylation des histones**, la charge négative de ces résidus est perdue (66). Ces modifications ont donc pour conséquence de réduire la basicité des histones ce qui provoque des changements d'interaction avec l'ADN, fortement acide. Ces changements pourraient diminuer la densité de la compaction en formant des contacts moins étroits entre les nucléosomes et l'ADN ou en changeant la capacité d'enroulement et de pliage (bending) de l'ADN (67). Il est aussi possible que les groupements acétyls nuisent à la liaison de l'histone H1 ou d'autres protéines non-histones, interférant ainsi avec le repliement supérieur de la chromatine. *In vitro*, l'histone H1 ne peut induire la condensation de la chromatine si les histones sont acétylés. L'acétylation des histones favorise aussi la liaison de certains facteurs de transcription comme par exemple Gal4 (68) et TF(III)A (69), ce qui pourrait directement stimuler la transcription.

De façon générale, on note que l'acétylation des histones est basse dans l'hétérochromatine (70). Il est aussi possible de tester l'acétylation des histones à des loci particuliers en immunoprécipitant spécifiquement les histones acétylées et en vérifiant si l'ADN du locus d'intérêt est enrichi dans le précipité par rapport au surnageant. Avec cette technique, on note que les histones recouvrant les séquences transcrites, comme celle du locus de  $\beta$ -globine dans les érythrocytes d'embryons de poulet, sont davantage acétylées (71). D'ailleurs, plusieurs protéines impliquées dans la transcription portent une activité d'acétylation des histones. Par exemple, TAF(II)230/250, des composantes de la machinerie basale de transcription et Gcn5p, p300/CBP et P/CAF, des co-activateurs interagissant avec des facteurs de transcription à domaine d'activation acide comme CREB ont toutes une activité acétyl transférase (72). L'acétylation des histones pourrait donc directement aider à la transcription en ouvrant la chromatine et en rendant l'ADN plus disponible.



### 1.3.5 Systèmes de régulation de la structure de la chromatine

#### 1.3.5.1 Répression de la transcription par inactivation de la chromatine

Plusieurs éléments d'ADN semblent pouvoir affecter en *cis* la chromatine environnante. Chez la levure, des transgènes intégrés près des télomères sont souvent silencieux et cet effet diminue avec la distance (73). Le phénomène est probablement dû à la formation d'hétérochromatine se propageant à partir des télomères. Une protéine, Rap1p, lie spécifiquement des séquences répétées des télomères. Par la suite, les protéines Sir3p et Sir4p interagissent avec Rap1p et aussi avec les histones H3 et H4 (74,75). Il s'en suit une polymérisation de protéines Sir3p et Sir4p sur les nucléosomes qui se propage le long de la chromatine. Des mutations dans la région des histones impliquées dans l'interaction avec les protéines SIR inhibent l'inactivation du locus. La modification des histones par acétylation pourrait empêcher l'interaction avec les protéines Sirp et la formation d'hétérochromatine. La chromatine ainsi prise au piège ne peut être transcrite et elle devient extrêmement résistante aux sondes moléculaires comme les nucléases. Le même phénomène se produit près du locus d'accouplement chez la levure qui doit être maintenu silencieux. En plus d'autres facteurs, Rap1p, Sir3p et Sir4p sont également impliqués dans l'inactivation de ce locus par polymérisation sur la chromatine. Un autre système est l'inactivation du chromosome X qui débute au locus "centre d'inactivation du chromosome X" (XIC) et se propage le long de la chromatine (76). Le processus débute lorsque le XIC est transcrit et que le messenger produit, Xist, vient se loger sur le locus. À partir de ces exemples, un modèle générale d'inactivation peut être proposé. Il faut un événement déclencheur comme la liaison d'une protéine ou d'un ARN à un site précis. Ensuite, à partir de cet élément, l'hétérochromatine pourra se former et elle se propagera le long de l'ADN. Son effet sera atténué à mesure que l'on s'éloigne du site déclencheur.

Chez la drosophile, un autre phénomène survient lors de l'inactivation de certains homéogènes par les protéines du groupe Polycomb (Pc-G) (77). Les loci concernés contiennent plusieurs séquences, encore mal caractérisées, appelées éléments de réponse aux Pc-G (PREs). Un modèle suggère que les protéines Pc-G lient ces PREs et forment

entre elles un complexe. Il en résulte des petites boucles d'ADN séquestré, formant une chromatine inactivée, quoique moins dense que le cas de la polymérisation aux télomères.

#### 1.3.5.2 Stimulation de la transcription par activation de la chromatine

Certains systèmes comportent des facteurs protéiques qui influencent la structure de la chromatine de façon à favoriser la transcription. Par exemple, le complexe SWI/SNF induit une réorganisation active des nucléosomes dans les cellules HeLa. Chez la levure, ce complexe est impliqué dans l'expression de plusieurs gènes et des mutations dans les gènes SWI1, 2 ou 3 ont des effets pléiotropiques semblables à ceux de mutations dans l'ARN polymérase II (78). SWI/SNF peut induire la liaison de facteurs de transcription à l'ADN (79). Par exemple, GAL4 requiert SWI/SNF pour lier son site dans le promoteur de *GALI* (80). Pour chacune de ces activités, l'hydrolyse d'ATP par une des sous-unités du complexe est requise. L'énergie libérée par cette réaction serait utilisée pour déplacer les nucléosomes et permettre la liaison de facteurs de transcription. À ce jour, comment SWI/SNF est dirigé spécifiquement vers les loci qu'il active n'est pas résolu.

#### 1.3.6 Isolateurs

D'autres séquences pourraient agir en **isolateurs**, n'affectant ni positivement ni négativement la transcription comme le font les enhanceurs mais bloquant la propagation de l'hétérochromatine. Chez la drosophile, les éléments transposables *gypsy* et les structures de chromatine spécialisées (scs) du locus de *hsp70* en sont des exemples (81). Placés de chaque côté d'un transgène, ces deux éléments sont capables d'annuler les effets de positions (82,83) et, insérés entre un enhanceur et un gène rapporteur, ils inhibent l'activation (82,84). Cette inhibition est polaire, c'est-à-dire qu'elle ne se produira pas si l'enhancer et le gène sont placés du même côté par rapport à l'isolateur. L'activité de *gypsy* provient de sites de liaisons pour le suppresseur d'hairywing (*su(Hw)*) (85) et elle requiert la présence de la protéine *su(Hw)* et *mod(mdg4)*. Le modèle proposé est que la liaison de la protéine *su(Hw)* sur l'ADN va ancrer cette séquence de telle sorte que la structure de la chromatine d'un côté de cette limite ne peut influencer l'ADN se trouvant de l'autre côté. Toutefois, si le gène codant pour *mod(mdg4)* est muté, alors l'isolateur n'est plus polaire.

Chez ce mutant, un site su(Hw) n'a pas besoin d'être entre l'enhancer et le gène et peut agir de l'extérieur. Cette particularité met en doute l'interprétation que *gypsy* est un isolateur et suggère qu'il pourrait être un inhibiteur d'enhancer. En effet, puisqu'il peut y avoir inhibition de l'extérieur, c'est que l'enhancer est inactivé par su(Hw) et non qu'il est dans un domaine de chromatine différent que le gène.

Les scs, quant à elles, sont localisées au locus de choc thermique 87A7 où sont disposés le gène *hsp70* proximal, scs, gène *hsp70* distal et scs' (86). On voit donc qu'un des deux gènes est entouré de séquences scs. Au cœur des scs, on retrouve une région protégée contre les sondes moléculaires telle l'ADNase I et, de chaque côté, des sites hypersensibles (82). Le patron de digestion de ces sites est modifié suivant un choc thermique, de même que la structure de la chromatine de tout le locus. Les scs peuvent isoler un gène *white* et le protéger des effets de position (82). Les scs sont liés par des protéines BEAF-32 que l'on retrouve localisées aux jonctions entre des bandes sur des chromosomes polytène (24). Ces chromosomes, retrouvés dans les glandes salivaires de drosophile, sont visibles en microscopie en raison de leur grande amplification par réplication, de telle sorte qu'il est possible d'avoir plusieurs centaines de brin d'ADN juxtaposés les uns aux autres (2). Des régions plus condensées forment des bandes foncées visibles en microscopie optique, séparées par des régions plus relâchées et plus pâles. Les protéines BEAF-32 pourraient donc être impliquées dans des frontières de domaines de chromosome et dans des domaines transcriptionnel.

Chez l'humain, un site hypersensible constitutif (HSS5) se retrouve devant le locus des globine du type  $\beta$ . Dans une expérience de blocage d'enhancer, ce site agit comme un isolateur (87). La séquence homologue chez le poulet, dupliquée plusieurs fois et disposé de chaque côté d'un gène rapporteur, est capable d'empêcher les effets de position (88). Des expériences plus détaillées permettront de déterminer si ce site constitue également une barrière de locus.

### 1.3.7 Les régions de contrôle de locus (LCRs)

Certaines séquences, appelées régions de contrôle de locus (LCRs), peuvent prévenir les effets de positions, mais à la différence des isolateurs, une seule copie du LCR est requise dans le transgène. La définition fonctionnelle d'un LCR est une séquence qui permet l'expression d'un transgène lié en *cis* quelque soit le site d'intégration et ce de façon proportionnelle au nombre de copies intégrées (9). Lorsqu'un LCR est introduit dans un transgène, chaque copie intégrée est transcrite à son plein potentiel quelque soit le site d'intégration, l'expression devient donc indépendante de la position et proportionnelle au nombre de copies. On suppose que les LCRs peuvent induire une ouverture de la chromatine, permettant l'expression du transgène malgré les effets de position. D'ailleurs, plusieurs expériences démontrent que les LCR affectent la structure de la chromatine (voir 1.4.4). La plupart du temps, les LCRs sont constitués de sites hypersensibles.

Le premier LCR découvert et le mieux caractérisé à ce jour est le LCR du locus de  $\beta$ -globine (89-91). Ce LCR sera décrit en détail à la section 1.4.4. Plusieurs autres LCR ont été identifiés pour un grand nombre de gènes encodants des protéines variées: TCR- $\alpha$ , MHC classe I (B7) et classe II (Ea), adénosine déaminase, hormone de croissance, tyrosinase, myosine, S100 $\beta$ , CD2, phosphoénol-pyruvate carboxy-kinase, kératine 18, C/EBP $\beta$  etc. (50,92-101). Les lymphomes de Burkitt sont causés par une translocation qui place le proto-oncogène *c-myc* sous le contrôle du LCR de la chaîne lourde d'immunoglobuline (102).

CD2 est un marqueur des lymphocyte T. Le gène CD2 s'étend du 15 kb et un transgène contenant le gène avec 5 kb de séquences flanquantes en 5' et 9 kb en 3' est exprimé de façon indépendante de la position (103). Des sites hypersensibles ont été découverts en 5' et en 3' du gène et des analyses de délétion ont pu démontrer que les sites 1,5 kb en 3' du signal de polyadénylation constituent un LCR (104).

Le lysozyme de poulet est un enzyme antibactérien présent dans l'oviducte mature et dans les cellules de la lignée myéloïde (105). Le locus du gène encodant cette protéine est

constituée d'une région à haute sensibilité à l'ADNase I dans les cellules exprimant le gène (106). En amont du promoteur se trouvent au moins deux enhancers qui co-localisent avec des sites hypersensibles et, dans le cas du distal, un MAR. Lors de la différenciation des macrophages, la chromatine du locus est remodelée de telle sorte que les nucléosomes deviennent en phase dans la région de l'enhancer proximal et sont repositionnés sur l'enhancer distal alors qu'un site hypersensible apparaît au promoteur. Des souris transgéniques pour le locus de lysozyme entier expriment le transgène de façon indépendante de la position (107). Par contre, la délétion d'un seul des enhancers provoquera des effets de position, bien que l'expression soit toujours spécifique au bon type cellulaire et régulée dans le développement. Les effets de position sont accompagnés par une perte du réarrangement de la chromatine, générant une structure de la chromatine hétérogène d'une lignée de souris à l'autre. L'ensemble de ces enhancers constitue donc un LCR.

#### **1.4 Les globines humaines du type $\beta$**

##### 1.4.1 Description des globines du type $\beta$

Les globines sont des protéines spécifiques aux cellules érythroïdes qui lient un groupement hème et s'assemblent en tétramère pour former l'hémoglobine. Ce complexe est responsable des transports gazeux dans le sang. Chaque tétramère est constitué de deux sous-unités du type  $\alpha$  et deux sous-unités du type  $\beta$ . Chez l'humain, les globines de type  $\beta$  sont  $\epsilon$ -globine, produite dans le sac vitellin embryonnaire,  $A\gamma$ -globine et  $G\gamma$ -globine, toutes deux exprimées dans le foie foetal, et  $\delta$ -globine (très faiblement) et  $\beta$ -globine, présentes dans la moelle osseuse adulte (108) (voir Figure 1). Dans la famille des  $\alpha$ -globines on retrouve  $\zeta$ -globine chez l'embryon et  $\alpha$ -globine chez l'adulte. Au cours des dernières années, la régulation transcriptionnelle des gènes encodant ces protéines et les mécanismes régulant les transitions d'une globine à l'autre ont constitué un domaine de recherche intense (109).

Toutes les globines du type  $\beta$  se retrouvent au même locus sur le chromosome 11 (110,111). La disposition des gènes dans le locus est dans le même ordre que leur

expression dans le développement ( $5'-\epsilon-A\gamma-G\gamma-\delta-\beta-3'$ ; voir Figure 2). Tous les gènes sont relativement petits (environ 4 kb) et ils sont tous compris dans une région de moins de 60 kb. De façon, similaire, les gènes encodant les  $\alpha$ -globines se retrouvent au même locus avec  $\zeta$ -globine en 5' et deux gènes  $\alpha$ -globine en 3'.

#### 1.4.2 Les hémoglobinopathies

Plusieurs maladies génétiques sont associées avec des mutations dans le locus de  $\beta$ -globine, principalement les thalassémies, la persistance héréditaire de l'hémoglobine foetale (PHHF) et l'anémie falciforme.

Les thalassémies sont causées par une réduction dans l'expression d'un ou de plusieurs gènes de globine (112). On parle d' $\alpha$ -thalassémies et de  $\beta$ -thalassémies selon qu'il manque d' $\alpha$ -globine ou de  $\beta$ -globine. Le phénotype peut être asymptomatique ou très sévère en fonction de la gravité de la réduction d'expression. Une importante réduction d'un type de globine provoque un débalancement entre les chaînes  $\alpha$ - et  $\beta$ -globine. Les sous-unités surnuméraires qui ne sont pas associées en tétramères vont alors polymériser entre elles et la formation du précipité peut détruire la cellule, aggravant davantage la maladie. Une foule de mutations causant des thalassémies ont été caractérisées. Les plus simples constituent en une délétion d'un gène de globine ou d'une région du locus comprenant plusieurs gènes. Parfois, ce sont des régions régulatrices qui sont délétées plutôt que les gènes. Dans certains cas, des translocations déplacent les gènes de globine en dehors du locus. Enfin, des mutations ponctuelles peuvent causer des diminutions plus ou moins substantielles de globine fonctionnelles. Ces mutations peuvent affecter des régions régulatrices dans les promoteurs ou les enhancers, des sites d'épissages ou de polyadénylation ou encore causer des changements dans le cadre de lecture (113).

Certaines PHHFs sont des  $\beta$ -thalassémies dont le phénotype est légèrement amélioré par l'expression des  $\gamma$ -globines dans les stades adultes (114). Elles sont causées par la délétions de larges fragment en aval des gènes  $\gamma$ -globine qui font disparaître le gène  $\beta$ -globine. D'autres PHHF sont causées par des mutations dans les promoteurs d'un des

deux gènes de  $\gamma$ -globine. Dans ces cas, il y a toujours expression du gène  $\beta$ -globine, donc pas de thalassémie majeure. L'anémie falciforme, quant à elle, est causée par une mutation ponctuelle qui modifie un acide aminé de la chaîne peptidique de  $\beta$ -globine. Cette mutation provoque la polymérisation spontanée de la  $\beta$ -globine dans le globule rouge, et le polymère change la morphologie cellulaire qui prend une apparence de faucille.

### 1.4.3 Systèmes utilisés pour étudier la régulation transcriptionnelle des globines

#### 1.4.3.1 Les cellules MEL

Parmi les systèmes utilisés pour étudier l'expression des gènes de globine, il y a les lignées cellulaires MEL et K562 et les souris transgéniques. Les **cellules MEL** sont dérivées de leucémies érythroïdes murines induites par le virus Friend et elles expriment les globines de souris adultes  $\beta$ -minor et  $\beta$ -major (voir 1.4.3.3) (115). Un traitement chimique au diméthylsulfoxyde (DMSO) ou au hexa-méthylène bi-acétamide (HMBA) peut induire fortement l'expression de ces gènes par un mécanisme de différenciation cellulaire. Ces cellules sont considérées comme des précurseurs d'érythrocytes (CFU-Es) et l'induction chimique lance une différenciation vers les pro-érythroblastes (voir Figure 3).

#### 1.4.3.2 Les cellules K562

Les **cellules K562**, quant à elles, furent clonées à partir d'effusions pleurales d'un patient humain atteint de leucémie myéloïde chronique (116). On peut différencier partiellement ces cellules vers la lignée mégakaryocytaire avec un traitement au tétradécanoyl phorbol ester (117) ou vers la lignée érythroïde avec l'agent cytosine arabinoside (l'utilisation d'hémin peut aussi stimuler la production de globine sans toutefois vraiment différencier les cellules), ce qui indique que ces cellules représentent un précurseur précoce pluripotentiel (voir Figure 3). Les cellules K562 expriment les globines embryonnaires et foetales ( $\epsilon$ - et  $\gamma$ -globine).

#### 1.4.3.3 Les souris transgéniques

Le modèle idéal pour étudier la régulation développementale des gènes de globine est **la souris transgénique**. En effet, l'utilisation de souris permet de recréer, dans un même système, les différents stades de développement. Il est donc possible d'étudier la régulation des niveaux de transcription mais aussi la régulation temporelle et tissulaire. Chez la souris, les globines embryonnaires de type  $\beta$  sont  $\epsilon\gamma$ ,  $\beta^H0$  et  $\beta^H1$ . Leur expression, dans le sac vitellin, décline vers le 11e jour de développement intra utérin (118) (voir Figure 2). Vers la même période, les gènes adultes  $\beta$ -minor et  $\beta$ -major prennent la relève dans le foie foetal, puis dans la moelle osseuse après la naissance et dans la vie adulte. Deux stades d'érythropoïèse sont donc définis: **stade embryonnaire** avant le 11e jour de développement et **stade foetal-adulte** après ce point limite. Les érythrocytes du stade embryonnaires sont appelés primitifs et sont morphologiquement distinct des érythrocytes définitif du stade foetal-adulte (118).

Bien qu'il n'y ait pas de globine murine spécifiquement foetale, les gènes humains foetaux introduits chez la souris se comportent comme les gènes embryonnaires murins. Ainsi, le sac vitellin ou le sang sera prélevé vers 10.5 jour de développement embryonnaire seront source de tissus embryonnaires où les globines humaines  $\epsilon$ -globine et les  $\gamma$ -globine seront exprimées (119,120). Pour étudier le stade foetal-adulte, le foie et/ou le sang foetal entre le 13e et le 17e jour, ou encore le sang d'une souris adulte sont les tissus qui exprimeront la  $\beta$ -globine humaine (120-122).

Les souris transgéniques constituent un bon modèle pour étudier les effets de position. En effet, l'insertion du transgène se fait au stade de zygote, où toute la chromatine est condensée (31). Il ne devrait donc pas y avoir de sites d'insertion privilégiés comme cela pourrait se produire dans des cellules en culture où la chromatine est plus variée. De plus, le transgène sera soumis à tous les repliements et ouvertures qui se produisent au cours du développement alors qu'un plasmide transfecté est nu lors de son insertion et n'a pas été soumis à toutes les pressions qu'un transgène aura à subir. Dans le cas des érythrocytes, les effets de position devraient être beaucoup plus importants que dans d'autres types cellulaires. En effet, au cours de l'érythropoïèse, la chromatine se condense fortement avant d'être complètement inactivée lors de l'expulsion du noyau



(123). Donc, l'hétérochromatine devrait être plus importante dans le développement de ce type cellulaire. Enfin, l'obtention de souris transgéniques sans avoir recours à la sélection avec un gène de résistance évite des biais favorisant les sites où il y a expression du transgène comme cela pourrait se produire lors de transfection dans des cellules en culture.

#### 1.4.4 Structure de la chromatine du locus de $\beta$ -globine

Dans les cellules exprimant les globines, il y a une grande région où la sensibilité générale à l'ADNase I est élevée qui comprend tous les gènes de globines du type  $\beta$  et plusieurs dizaines de kb de chaque côté (124). Cette sensibilité est présente à tous les stades de développement. En amont de chaque promoteur, on peut retrouver un site hypersensible (125). La présence de ces sites est toutefois restreinte aux stades du développement où chacun des gènes est exprimé. Par exemple, au stade foetal les sites des gènes embryonnaires disparaissent pour laisser leur place aux sites des gènes foetaux alors que ceux des gènes adultes ne sont pas encore là. En amont du gène  $\epsilon$ -globine (à une distance entre 20 et 5 kb) se trouvent quatre sites hypersensibles spécifiques aux cellules érythroïdes (HSS1 à HSS4; voir Figure 2) (125,126). Ces sites sont stables dans le développement et constituent le LCR (voir 1.4.4). Il y a aussi deux autres sites hypersensibles qui ne sont quant à eux pas spécifiques aux érythrocytes. Le premier est localisé devant le LCR (HSS5) et le second 20 kb après le gène de  $\beta$ -globine (HSS-3') (127). Ensuite, les histones du locus sont fortement acétylées dans les cellules érythroïdes. L'immunoprécipitation des histones acétylées enrichit fortement le culot en ADN du locus (71). Le domaine d'acétylation correspond environ au domaine de sensibilité générale.

#### 1.4.5 Régulation transcriptionnelle locale des globines de type $\beta$

Le promoteur proximal du gène de  $\beta$ -globine contient trois motifs hautement conservés: CACC, CAAT et TATA (voir Figure 4) (128). Ces motifs se retrouvent dans tous les gènes de globine et aussi dans plusieurs autres gènes. Des mutations introduites dans le promoteur proximal, seules celles modifiant ces éléments réduisent significativement les niveaux d'expression en transfection cellulaire (129). Ces motifs

sont donc importants pour obtenir de hauts niveaux de transcription. L'activité de ces éléments n'est pas spécifique aux cellules érythroïdes car ils sont aussi fonctionnels dans plusieurs autres types cellulaires (HeLa, 3T6) (130,131). Il existe des mutations naturelles chez l'humain affectant le motif CACC et TATA et ces mutations provoquent des  $\beta$ -thalassémies légères chez les patients (113,132). Si ces mutations sont réintroduites dans des promoteurs dont l'activité est testée en transfection cellulaire, on note une baisse d'expression environ équivalente à celle observée chez les patients. Encore une fois, cette baisse se produit aussi dans des cellules non-érythroïdes. Bien que le promoteur puisse diriger une faible expression en transfection cellulaire, son activité ne semble pas suffisamment importante pour permettre la transcription chez les souris transgéniques. En effet, un gène rapporteur dirigé par le promoteur de  $\beta$ -globine était complètement silencieux chez toutes les souris analysées (121).

En aval du promoteur se situent deux enhanceurs de la transcription, l'un à la jonction du second intron et du troisième exon et l'autre environ 500 pb en aval du signal de polyadénylation (133). L'ajout de ces séquences stimule plusieurs fois les niveaux de transcription du gène de  $\beta$ -globine ou d'un rapporteur dans les cellules MEL. L'activité de ces enhanceurs est spécifique au tissu érythroïde et au stade adulte. Ils peuvent conférer à un gène non-globine une expression érythroïde et adulte (foetale-adulte chez les souris transgéniques) (134). L'étude de souris transgéniques pour le gène  $\beta$ -globine complet, incluant les enhanceurs, a révélé l'expression du transgène chez un peu plus de la moitié des souris transgéniques (120-122,135). Les niveaux d'expression étaient toutefois très bas, la majorité du temps en deçà de 10% de celle des globines murines endogènes et ce même si plusieurs copies du transgène étaient intégrées. Les enhanceurs sont donc fonctionnels chez les souris mais soumis à d'importants effets de position.

De façon similaire, les gènes  $\gamma$ -globine contiennent plusieurs éléments locaux de régulation. Environ 400 paires de bases en aval du signal de polyadénylation de  $A\gamma$  se trouve un enhancer de la transcription (136). Cet élément n'est toutefois pas très spécifique puisqu'il est actif dans divers types cellulaires. Par contre, le promoteur proximal comporte un enhancer spécifique au stade embryonnaire (137). Un gène de

$\gamma$ -globine complet introduit dans le génome de souris transgénique sera, comme le gène de  $\beta$ -globine, régulé dans le développement, donc exprimé spécifiquement dans les érythrocytes embryonnaires, mais à très faible niveau, selon les effets de position (119,120).

#### 1.4.6 Activation de la transcription des gènes de globine par le LCR

Les  $(\gamma\delta\beta)^0$ -thalassémies sont causées par la délétion des régions en 5' de  $\epsilon$ -globine (114). Le phénotype de ces maladies est très sévère: tous les gènes du locus sont complètement silencieux (138,139). Ces observations, couplées au fait que la région délétée comprend des sites hypersensibles (125,126), indiquaient que des éléments de régulation pouvait se trouver en amont du locus. Grosveld *et al.* ont liés une région de 20 kb comprenant les sites hypersensibles de la région délétée directement au gène de  $\beta$ -globine et des souris transgéniques ont été produites avec ce mini-locus (voir Figure 2) (140). Les niveaux d'expression du transgène dans les cellules érythroïdes de ces souris étaient beaucoup plus importants qu'avec le gène seul. Toutes les lignées produites exprimaient le transgène et ce, de façon proportionnelle au nombre de copies intégrées. Les souris portant un grand nombre de copies intégrées étaient même gravement anémiques en raison du déséquilibre entre les chaînes de  $\alpha$ -globine et de  $\beta$ -globine. Deux activités principales sont donc associées avec le LCR: a) une très forte activation de la transcription (activité enhancer) qui fait que le transgène est exprimé à des niveaux près de 100 fois plus élevés et b) la capacité de conférer une expression indépendante de la position (activité LCR) qui pourrait être associée avec une ouverture de la chromatine (voir 1.4.9).

#### 1.4.7 L'activité du LCR est spécifiquement érythroïde

Pour caractériser le LCR, sa capacité à stimuler la transcription a été testée dans plusieurs types cellulaires (139-143). Aucune stimulation n'a été détectée dans des cellules non-érythroïdes. Le LCR peut toutefois stimuler des gènes non-globines dans les érythrocytes. Dans les cellules MEL, la transcription du gène Thy-1 ou du gène de sélection tk-*neo* est stimulée par le LCR (141). De la façon similaire, des transgènes hsp68-*lacZ*, hprt-*lacZ* ou antigène grand T de SV40 liés au LCR et introduits dans des

souris transgéniques sont exprimés de façon élevée dans le compartiment érythroïde mais faible et variable dans les autres tissus (144-146). Le gène  $\alpha$ -globine peut aussi être stimulé par le LCR (143,147). En fait, l'insertion d'un gène dans le locus endogène de  $\beta$ -globine (entre  $\delta$ -globine et  $\beta$ -globine) permet son expression dans des cellules érythroïdes (148). Le LCR est donc actif seulement dans les cellules érythroïdes et il peut activer plusieurs gènes différents.

#### 1.4.8 Les gènes de globine de type $\beta$ compétitionnent pour l'activation par le LCR

Lorsqu'un seul gène de globine est lié directement au LCR, la spécificité développementale du gène est perdue. Une construction LCR- $\beta$ -globine (micro-locus, voir Figure 2) est exprimée dans les cellules humaine K562 (qui ont plutôt la capacité d'exprimer les globines embryonnaires) alors que le gène de  $\beta$ -globine seul y est silencieux (141). Des souris transgéniques pour une construction LCR- $\beta$ -globine expriment la  $\beta$ -globine aussi bien dans les cellules embryonnaires que dans les cellules adultes (149,150). De façon similaire, un gène  $\gamma$ -globine dirigé par le LCR chez des souris transgéniques est toujours exprimé dans le stade foetal et, selon les séquences introduites, fortement ou faiblement exprimé dans le stade adulte (149-151). Quel que soit le transgène, l'expression d'un gène  $\gamma$ -globine dans le stade foetal-adulte est toujours supérieure à celle rencontrée avec le locus complet, où le silencing des gènes de  $\gamma$ -globine est presque total (152-154). Par contre, si les gènes  $\gamma$ - et  $\beta$ -globine sont tous deux liés au LCR dans le même ordre que dans le locus, alors la régulation temporelle correcte est rétablie, c'est-à-dire que le gène  $\gamma$ - sera exprimé seulement dans le stade embryonnaire et le gène  $\beta$ - seulement dans le stade foetal-adulte murin (149,150). Ces observations révèlent d'autres caractéristiques du LCR: a) l'activation de la transcription est fonctionnelle au cours de tous les stades; b) elle peut supplanter la spécificité développementale des gènes (encodés dans les promoteurs et/ou enhanceurs locaux) et c) un seul gène à la fois peut être activé puisque le LCR ne stimule pas  $\gamma$ -globine et  $\beta$ -globine au même moment. Par surcroît, lorsqu'un gène de sélection *neo* est intégré à l'intérieur même du LCR dans des lignées cellulaires ou des souris transgéniques, non seulement devient-il fortement

stimulé, mais il compétitionne tous les gènes de globine qui deviennent complètement silencieux (155-157).

La régulation développementale des gènes de globine fait donc intervenir deux mécanismes: la régulation autonome qui se fait localement et la régulation par la stimulation du LCR à travers tout le locus. En présence du LCR, le gène  $\beta$ -globine est exprimé dans le stade embryonnaire à moins qu'un gène  $\gamma$ -globine ne puisse intercepter l'activation par le LCR. C'est donc dire qu'à l'intérieur du locus, les gènes sont en compétition pour l'activation par le LCR. Le gène  $\varepsilon$ -globine est régulé de façon autonome, c'est-à-dire qu'il est actif dans le stade embryonnaire et que des facteurs répresseurs viennent en inhiber l'expression dans les stades ultérieurs, même en présence du LCR (158,159). Les gènes  $\gamma$ -globine sont partiellement autonomes mais leur inactivation complète dans le stade adulte requière la présence de séquences en 3' et du gène  $\beta$ -globine (152). À cet égard, il est intéressant de noter que certaines PHHF sont causées par la perte de ces séquences en 3', incluant le gène  $\beta$ -globine (114). Donc, l'expression du gène  $\beta$ -globine pourrait être requise pour l'inactivation complète des gènes  $\gamma$ -globine par compétition pour le LCR. Toutefois, cette hypothèse est remise en question par le fait que les délétions causant des PHHF mettent les gènes  $\gamma$ -globine à proximité d'un enhancer (de l'autre côté de la jonction créée par la délétion) qui se trouve normalement à plusieurs dizaines de kb de distance. Cet enhancer pourrait expliquer l'expression adulte de  $\gamma$ -globine plutôt que l'absence de compétition (114). De plus, plusieurs  $\beta$ -thallasémies affectant le gène  $\beta$ -globine ne provoquent pas de PHHF.

Par contre, le gène  $\beta$ -globine est régulé principalement par compétition: le gène est actif tout au cours du développement lorsqu'il est lié seul au LCR. Pour conférer une spécificité adulte au gène  $\beta$ -globine, il faut ajouter à la construction un gène embryonnaire ou foetal qui pourra compétitionner pour le LCR et donc empêcher l'expression de  $\beta$ -globine dans les stades précoces du développement. Quelque soit la disposition des gènes (LCR- $\gamma$ - $\beta$  ou LCR- $\beta$ - $\gamma$ ), la régulation développementale des deux gènes est rétablie assez fidèlement si les deux sont présents à la fois. La compétition est légèrement moins efficace si le gène  $\gamma$ -globine est placé plus loin que le gène  $\beta$ -globine (il y a une faible

expression de  $\beta$ -globine dans les stades embryonnaire et foetal dans le cas du transgène LCR- $\beta$ - $\gamma$ ). Cette observation révèle que la position par rapport au LCR peut influencer légèrement la spécificité développementale mais que les éléments de régulation locaux des gènes sont les déterminants majeurs de cette spécificité. La répression autonome des gènes précoces permettrait donc l'activation du gène  $\beta$ -globine dans les stades adultes. Toutefois, nous ne savons pas s'il y aurait toujours une expression embryonnaire si le gène  $\beta$ -globine était placé à 60 kb de distance, comme dans le locus normal.

En plus de la spécificité développementale des gènes, la distance par rapport au LCR influence la compétition génique. Dans une construction LCR- $\beta$ -globin- $\gamma$ -globin, il y a expression des deux gènes dans le stade embryonnaire (160,161). C'est donc dire que la proximité entre le LCR et un gène de globine en fait un meilleur compétiteur. Cette observation aurait pu laisser croire que l'activation provenait du LCR et se propageait par le brin d'ADN jusqu'au gène. Ainsi, le plus près serait le mieux exprimé. La répression des gènes de  $\gamma$ -globine pourrait laisser passer l'activation jusqu'au gène de  $\beta$ -globine dans les stades foetal-adulte. Toutefois, de récentes expériences suggèrent que cette compétition polaire se fait plutôt par interaction LCR-gène impliquant une boucle d'ADN. En effet, la magnitude de la compétition d'un gène de  $\beta$ -globine marqué, inséré dans le locus complet, par rapport au gène de  $\beta$ -globine normal dépend de sa position dans le locus (162). Or, selon un modèle de propagation, peu importe la position du gène marqué, sa capacité de compétition aurait dû être la même du moment qu'il soit devant le gène normal. Donc, un modèle de compétition polaire, où le LCR stimule les gènes par contact direct, suggère que la proximité d'un gène augmente son affinité pour le LCR parce que la probabilité d'un contact est plus grande (parce que la distance est plus petite).

Pour tester la dynamique des interactions dans la chromatine *in vivo*, Wijgerde et al. ont étudié la synthèse des globines de type  $\beta$  au moment de la transition entre les globines embryonnaires et les globines foetales-adultes chez les souris transgéniques (au moment où les  $\gamma$ -globine sont encore présentes et  $\beta$ -globine est en hausse, voir Figure 1). Les auteurs ont mesuré la quantité d'ARN pré-messagers (avant épissage) dont la durée de vie est très courte avec des sondes dirigées contre les introns de globine en hybridation *in situ*.

Cette technique permet d'observer l'expression à un instant précis (163,164). Différentes cellules exprimaient soit seulement  $\gamma$ -globine, soit seulement  $\beta$ -globine ou alors les deux. Les cellules qui n'exprimaient que  $\gamma$ -globine avaient toutefois des messagers épissés de  $\beta$ -globine. Donc, le LCR peut activer les deux gènes consécutivement et faire des "flip-flops" de l'un à l'autre, mais à un instant donné, un seul gène est transcrit. Les cellules où un signal d'ARN pré-messager a été détectée pour  $\gamma$ -globine et  $\beta$ -globine venaient probablement tout juste de faire une transition entre les gènes et la dégradation du premier n'était pas terminée alors que la synthèse de l'autre était déjà commencée.

#### 1.4.9 Le LCR de $\beta$ -globine et la structure de la chromatine du locus

Il est possible d'étudier divers loci de  $\beta$ -globine en construisant des hybrides entre des cellules lymphoïdes de patients et des cellules MEL. On peut mesurer l'expression des globines provenant du chromosome humain de même que la sensibilité générale et les sites hypersensibles qui sont recréés dans l'hybride en culture. Par contre, lorsque l'hybride provient d'un patient  $(\gamma\delta\beta)^0$ -thalassémique (où le LCR est délété), le locus est résistant aux nucléase (165). Il n'y a pas de sites hypersensibles aux promoteurs ni de sensibilité générale dans le locus. La réplication, qui se fait habituellement tôt dans la phase S (avec un locus normal), devient alors tardive, tout comme la chromatine inactive (165). L'origine de réplication du locus est même changée en absence du LCR (166). Il semble donc que le LCR soit requis pour activer la chromatine du locus, tant au niveau des promoteurs que des séquences intergéniques.

Une autre indication que le LCR agit sur la chromatine est la différence de comportement dans des tests de transfections transiantes versus stables. En transiant, le LCR ne stimule un gène rapporteur que faiblement alors qu'en stable l'activation est plus importante (142). La différence entre les deux expériences est que l'expression transitoire provient de la transcription du plasmide libre. En stable, le plasmide doit être intégré dans le génome hôte et il se retrouve donc assemblé en chromatine. En étudiant les sites hypersensibles individuellement, seul HSS2 avait une activité enhancer en transiant (142), mais tous les sites stimulaient la transcription en stable. Donc HSS2 est un enhancer

classique alors que les autres sites ont une action différente, qui n'est observée qu'en présence de chromatine.

Il n'y a pas encore de preuves directes que le LCR ouvre la chromatine au site d'intégration. Sa capacité à conférer une expression indépendante de la position pourrait être explicable par d'autres mécanismes. Par exemple, il pourrait favoriser l'insertion dans des loci transcrits. Toutefois, en lumière de l'inactivation de la chromatine du locus lors de la délétion du LCR, l'hypothèse de l'ouverture de la chromatine semble plus réaliste. Le terme **ouverture de la chromatine** sera utilisé indistinctement de la capacité de conférer une expression indépendante de la position dans la suite de cette thèse.

#### 1.4.10 Analyse des éléments composant le LCR

Afin de déterminer quels étaient les éléments du LCR qui lui conféraient son activité, des dissections très précises du LCR ont été effectuées. Dans un premier temps, la localisation grossière des séquences importantes a été cherchée par des études de délétion. En liant côte à côte de courtes régions (1-2 kb) contenant chacun des 4 sites hypersensibles érythroïdes et le gène de  $\beta$ -globine dans une construction qui s'appelle le micro-locus (voir Figure 2) on peut obtenir les mêmes niveaux d'expression qu'avec le mini-locus chez les souris transgéniques et les cellules MEL (167,168). Les régions contenant les sites hypersensibles conservent donc l'activité du LCR.

L'étape suivante était de déterminer l'activité individuelle de chacune de ces régions. Un très grand nombre d'études avec des stratégies diverses ont été publiées. Certaines plaçaient une seule des régions de sites hypersensibles à la fois avec le gène de  $\beta$ -globine, ou encore avec le locus complet. D'autres approches étaient de déléter une à une ces régions du micro-locus ou encore du locus complet par knock out direct dans le LCR murin ou par recombinaison homologue sur un YAC contenant le locus de  $\beta$ -globine humain. Certains résultats sont contradictoires et la littérature est un peu confuse. On peut tout de même tirer certaines conclusions générales. Les sites hypersensibles 2, 3 et 4 peuvent empêcher les effets de positions dans des blocs de plusieurs copies du transgène



(169). Toutefois, seul HSS3 confère une expression indépendante de la chromatine lorsqu'une seule copie est intégrée (170,171). Avec une analyse très fine, de relativement faibles effets de positions ont été notés dans un micro-locus où un seul site a été délété (HSS1, 2 ou 3; voir 1.4.12.1) (172). L'activité d'ouverture de la chromatine ne provient donc pas d'une séquence discrète mais est plus ou moins diffuse sur plusieurs sites. Le site HSS1 est un très faible enhancer, il contribue de façon minoritaire à l'activité du LCR (173,174). HSS2 peut être considéré comme un vrai enhancer; il stimule tous les gènes de globine et sa perte réduit les niveaux d'expression environ trois fois à tous les stades du développement (157,170,175-180). HSS3 stimule lui aussi la transcription, mais peut-être surtout celle de gènes spécifiques. La littérature est très confuse à ce sujet et certains démontrent que HSS3 stimule davantage  $\epsilon$ -globine, d'autres favorisent les  $\gamma$ -globine et d'autres  $\beta$ -globine (173,180-184). HSS4 est quant à lui peu efficace en transfection dans les cellules MEL mais presque aussi puissant que HSS2 et HSS3 chez les souris transgéniques. Ce dernier site pourrait stimuler surtout les gènes adultes (173,185,186). Toutes ces données indiquent que le LCR, pour fonctionner pleinement, doit être composé de la presque totalité du micro-locus. La délétion de diverses régions résulte en une perte d'activité plus ou moins importante. L'activité globale ne peut pas être directement divisée en activités discrètes dispersées dans la séquence.

Finalement, l'activité de chacun des sites hypersensibles a été analysée finement pour identifier les séquences importantes et les protéines qui s'y lient. Dans tous les sites se trouvent une région de 200-300 pb qui contient le site hypersensible. Cet élément central (core element) retient presque toute l'activité de la région complète du site hypersensible (1-2 kb). Dans HSS2, cet élément central est constitué ainsi: 5' - motif G riche (semblable à un motif CACC et lié par Sp1 et BKLF/TEF-2) - dimère de sites NF-E2 - dimère de sites GATA-1 - sites USF (H-BP) - 3' (175,177) (une description de GATA-1 et NF-E2 sera donnée à la section 1.4.13). L'activité enhancer est principalement portée par les sites NF-E2 mais tous les autres y contribuent également. Aucune séquence précise n'est requise pour conférer une expression indépendante de la position. Toutefois, la mutation de deux sites parmi Sp1, GATA-1 et USF abolit cette activité

(175,176,178,179). NF-E2 semble donc surtout important pour stimuler les niveaux de transcription et les autres sites agissent ensemble pour créer l'activité LCR. Pour HSS3, on trouve 5' - dimère de sites GATA-1 - dimère de motifs G riche -dimère de sites GATA-1 - site GATA-1 - trimère de motifs G riche - site GATA-1 - motif G riche - 3' (181). L'activation de la transcription résulte d'une synergie de tous ces sites et aucun n'a une activité plus importante que les autres (182,183). L'expression indépendante de la position provient d'une combinaison de motifs GC entourés de sites GATA-1; il y a donc deux régions dans cet élément qui peuvent tous deux agir en LCR. L'analyse du site HSS4 révèle des sites de liaison pour NF-E2, GATA-1, jun/fos et des motifs G riche (185,186). Le rôle précis de chacun de ces sites reste à tester. En résumé, tous les sites hypersensibles sont composés de sites de liaison pour des facteurs de transcription érythroïdes (GATA-1 et NF-E2) et des facteurs ubiquitaires (Sp1 et TEF-2, USF et d'autres mineurs). Il semble donc que la juxtaposition de plusieurs sites crée un élément fonctionnel. Parmi ces séquences, les sites NF-E2 sont les enhanceurs les plus puissants mais toutes les autres séquences participent également à la stimulation. L'activité d'ouverture de la chromatine est plus dispersée dans tous les éléments mais impliquent principalement des sites pour GATA-1 et des sites G riches.

#### 1.4.11 La structure du transgène influence la capacité du LCR à ouvrir la chromatine

Avant le début de cette thèse, le LCR était considéré comme dominant en lui-même sur la chromatine environnante. Depuis, des articles qui remettent en question ce modèle sont parus. Reitman et ses collaborateurs ont fait un travail remarquable en étudiant le locus de  $\beta$ -globine de poulet chez les souris transgénique. Ils ont premièrement identifié un LCR situé entre les gènes  $\beta^A$ -globine adultes et  $\epsilon$ -globine embryonnaire (187). Ensuite, il a démontré que le LCR devenait inactif (l'hypersensibilité était dépendante de la position) chez des souris transgéniques où le gène  $\beta^A$ -globine ne pouvait pas être transcrit parce que le promoteur n'était pas inclus dans la construction (188). Donc, il semble que le gène transcrit doit activer le LCR pour que ce dernier puisse ouvrir la chromatine. Dans d'autres études, il a démontré que le locus entier de poulet était exprimé de façon indépendante de la position (189). L'inclusion de quatre sites hypersensibles en 5' du

locus (voir Figure 5) permet toujours une ouverture de la chromatine. Toutefois, si un seul de ces sites était ajouté, alors des effets de position sont observés (190). Donc, le LCR du locus de  $\beta$ -globine de poulet est sensible au contenu du transgène. Peut-être qu'un site hypersensible seul accapare de LCR et l'empêche de fonctionner alors que tous les sites hypersensibles ensemble forment un holo-complexe (voir 1.4.12.1) qui n'interfère pas avec le LCR.

D'autres études concernant le LCR de  $\beta$ -globine humain vont un peu dans le même sens. Si le LCR du micro-locus est lié au gène  $A\gamma$ -globine, l'expression sera indépendante de la position si un court promoteur est utilisé alors qu'elle sera dépendante si c'est un promoteur long (voir Figure 6) (191). Si on rajoute davantage de séquences en 3' du gène, ou encore le gène  $\beta$ -globine, au gène de  $\gamma$ -globine le plus long on retrouve une expression indépendante de la position (192,193). Ces résultats un peu étranges pourraient refléter le besoin du LCR d'être lui-même activé par un gène de globine. Cette activation pourrait nécessiter une unité de transcription bien particulière. Une structure précise des éléments du gènes pourrait être requise. À ce sujet, voir le modèle de l'interaction mutuelle en section 1.4.12.2 pour plus de détails.

#### 1.4.12 Mécanismes suggérés pour l'activité du LCR

##### 1.4.12.1 Le LCR pourrait former un holo-complexe agissant comme une seule entité et activant la transcription selon une durée et une fréquence déterminée

Ellis *et al.* ont suggéré que les protéines liant chacun des sites du LCR pourraient s'assembler pour former un large holo-complexe qui agirait comme une seule unité (170). Lors de l'activation d'un gène, le holo-complexe pourrait venir à proximité du promoteur en formant une boucle d'ADN et tenir la chromatine ouverte ou recruter des composantes de la machinerie transcriptionnelle comme cela a été proposé pour les enhancers. Ce modèle pourrait par exemple expliquer pourquoi le LCR ne stimule qu'un seul gène à la fois et qu'il a peu de spécificité pour un type de gène en particulier (voir 1.4.7). Il est de plus consistant avec l'observation que la structure de la chromatine du LCR est stable dans le développement.

Le modèle du holo-complexe est renforcé par le modèle de flip-flop qui indique qu'un seul gène à la fois est transcrit et qu'il peut y avoir alternance entre deux gènes (voir 1.4.8). Récemment, Milot *et al.* ont détecté des effets de position avec cette technique (172). En étudiant des souris transgéniques pour des constructions de micro-locus desquels les sites hypersensibles individuels avaient été délétés, ils ont pu démontrer que la transcription n'était pas toujours présente. En d'autres termes, les effets de position diminuaient la fraction du temps que le LCR passait à activer le gène. Ces observations s'expliqueraient très bien par un modèle de holo-complexe luttant contre la formation d'hétérochromatine.

#### 1.4.12.2 Modèle de l'interaction mutuelle

Comment le LCR peut-il prévenir les effets de positions? Plusieurs résultats suggèrent que le LCR interagit avec les gènes qu'il active en faisant une boucle d'ADN (149,150,162). D'autres expériences démontrent que le LCR doit lui-même être en contact avec une unité transcriptionnelle pour être activé (188). Il serait donc possible que le LCR entre en **interaction mutuelle** avec le gène qu'il active. Au cours de cette interaction, le gène pourrait maintenir le LCR fonctionnel et, en sens inverse, le LCR pourrait stimuler la transcription du gène (voir Figure 7). Le LCR devrait donc activer un seul gène à la fois et au cours du développement cette interaction pourrait favoriser un gène plutôt qu'un autre. Le fait que l'ouverture de la chromatine puisse être affectée par l'ajout de séquences dans le transgène pourrait alors s'expliquer par une interaction déficiente (190,191). Les séquences ajoutées pourraient s'insérer dans le complexe formé au cours de l'interaction et le déstabiliser. Alternativement, ces séquences pourraient séquestrer avec le gène ou le LCR et empêcher l'interaction mutuelle.

Le modèle d'interaction permet d'expliquer comment l'activation de la transcription pourrait se produire et elle peut également suggérer l'effet du LCR sur la chromatine. Comme une boucle d'ADN est formée entre le LCR et les gènes, il est possible que la chromatine comprise dans la boucle ne puisse tout simplement pas être compactée selon les repliements supérieurs. La boucle pourrait être incompatible avec une structure

compacte et donc prévenir la formation d'hétérochromatine. Alternativement, la propagation de l'hétérochromatine provenant de l'extérieur du transgène pourrait être bloquée par le complexe formée par l'interaction mutuelle.

#### 1.4.12.3 Modèle de propagation de l'ouverture de la chromatine par le LCR

Un autre mécanisme par lequel le LCR pourrai ouvrir la chromatine serait un signal de décompaction initié au LCR qui se transmet sur le brin d'ADN jusqu'au gène activé (décrit et critiqué par Dillon *et al.* (162)) (voir Figure 8). La chromatine étant alors ouverte, le gène pourra être transcrit à son plein potentiel. Ce modèle de **propagation** est toutefois plus difficile à concilier avec le fait qu'un seul gène à la fois soit activé, que le LCR agisse également en enhancer et que le LCR doive être activé par le gène qu'il stimule.

#### 1.4.12.4 Le LCR pourrait prévenir la variégation

Tel que mentionné à la section 1.2.10, il est possible que les effets de position soient en fait dû à la variégation de l'expression. Si tel était le cas, le LCR pourrait prévenir cette variégation de telle sorte que toutes les cellules expriment le transgène.

Ce modèle a été développé par Festenstein *et al.* qui ont montré que l'expression d'un transgène CD2 était parfois hétérocellulaire, mais jamais en présence du LCR fonctionnel (194). Lorsqu'il y avait variégation, l'insertion était près du centromère. Toutefois, l'expression par cellule a été déterminée par cytométrie de flux ou par immunohistochimie. Ces techniques ne permettent pas de déterminer clairement si l'expression est vraiment bimodale comme le modèle de variégation le suppose.

#### 1.4.13 Les protéines régulant les gènes de globines.

##### 1.4.13.1 GATA-1

Une foule de facteurs transcriptionnels sont impliqués dans la régulation des gènes de globine. Parmi les plus importants, il y a GATA-1, un facteur du type doigt de zinc spécifique aux tissus érythroïde et myéloïde (195). Des sites de liaisons pour GATA-1

sont retrouvés dans plusieurs promoteurs et enhancers de gènes de globine ( $\alpha$  et  $\beta$ ) (181,196-198) et dans plusieurs autres gènes érythroïdes (199-201). C'est un activateur de la transcription efficace (202), impliqué dans la régulation de virtuellement tous les gènes érythroïdes, et absolument requis pour la formation des érythroblastes. Le développement des cellules portant une allèle inactivée de GATA-1 est bloqué au stade de pro-érythroblaste et il n'y a plus de différenciation mais plutôt apoptose (203,204). Les sites GATA-1 sont importants pour les activités d'activation de la transcription et d'indépendance de la position du LCR (176,182). Souvent, des sites consensus pour GATA-1 stimulent la transcription de façon synergique avec des motifs CACC ou GC riches (201,205). GATA-1 peut former des homodimères et/ou des hétérodimères avec d'autres facteurs comme Sp1 et EKLF (205,206). Il est donc possible que GATA-1 soit impliqué dans des liaisons protéine-protéine lors de l'interaction mutuelle entre le LCR et les gènes de globine.

#### 1.4.13.2 NF-E2

NF-E2 est une protéine de type leucine zipper liant des sites de type AP-1 du LCR dont l'expression se limite à quelques types de cellules hématopoïétiques (175,185,207). La fonction majeure de NF-E2 semble être l'activation de la transcription. L'inactivation génétique de NF-E2 par insertion virale ou l'utilisation de mutants dominants négatifs qui empêche le fonctionnement de la protéine réduit fortement l'expression des globines dans les cellules MEL (208,209). Toutefois, l'ablation génétique de NF-E2 chez la souris n'affecte pas les globines mais plutôt les plaquettes (210,211). Ceci pourrait être expliqué par l'utilisation de protéines très homologues comme Nrf-1, Nfr-2 ou LCR-F1 (212-214). Ces protéines pourraient par exemple être présentes chez les souris dépourvues de NF-E2 mais pas chez les cellules en culture. Un domaine de NF-E2, requis pour l'activation, interagit avec des composantes de la machinerie transcriptionnelle de base (TAF-130), même à très grande distance (215); NF-E2 pourrait donc lui aussi être impliqué dans l'interaction LCR-gènes de globine.

### 1.4.13.3 EKLf

Le facteur érythroïde de type Krüppel (EKLf) est une protéine dont les trois domaines doigt de zinc sont homologues à ceux de la protéine Krüppel de drosophile et Sp1 chez les mammifères (216). Son expression est restreinte aux cellules érythroïdes, autant embryonnaires qu'adultes (217). Le meilleur site de liaison connu pour EKLf est le motif CACC du promoteur du gène de  $\beta$ -globine. Bien que plusieurs autres gènes de globine et des sites hypersensibles du LCR contiennent un tel motif, l'affinité de liaison pour EKLf y est plus faible (218). EKLf stimule la transcription d'un rapporteur dirigé par le motif CACC dans des tests de cotransfection (216,219). Les mutations du motif CACC causant chez l'humain une réduction d'environ trois fois dans les niveaux de  $\beta$ -globine abolissent la liaison de EKLf et la transactivation du rapporteur (220). L'ablation génétique de EKLf provoque une très forte diminution d'expression du gène  $\beta$ -major dégénéralant en anémie létale durant la vie foetale, bien que les autres gènes de globine étudiés, tels  $\alpha$ -,  $\zeta$ -,  $\epsilon\gamma$ - et  $\beta$ h1-globine, étaient exprimés à des niveaux normaux (221,222). Contrairement à ce qui se produit lors de l'inactivation de GATA-1, les érythrocytes des souris EKLf  $-/-$  se développent jusqu'à la différenciation terminale, leur seule anomalie est un manque de  $\beta$ -globine adulte (223). Ces résultats suggèrent que la fonction principale de EKLf est d'activer spécifiquement la transcription du gène de  $\beta$ -globine. Des souris transgéniques pour le locus entier de  $\beta$ -globine humaine portant les allèles inactivées de EKLf ont permis de montrer que l'expression du gène de  $\beta$ -globine humaine était elle aussi fortement réduite (164,224). De plus, l'expression des gènes de  $\gamma$ -globine chez ces souris était prolongée après le 11<sup>e</sup> jour, date normale de la transition des gènes de  $\gamma$ -globine vers celui de  $\beta$ -globine. Il a même été démontré par des tests d'hybridation *in situ* de transcrits primaires que les gènes de  $\gamma$ -globine étaient plus souvent transcrits en absence de EKLf. Cette absence de "silencing" des gènes de  $\gamma$ -globine a amenée l'hypothèse que EKLf jouait un rôle majeur dans la transition d'expression de  $\gamma$ -vers  $\beta$ -globine en augmentant l'affinité d'interaction entre le LCR et le promoteur de  $\beta$ -globine (219). Alternativement, EKLf pourrait activer le gène de  $\beta$ -globine au début du

stade foetal-adulte et ainsi induire le transfert d'activation par le LCR des gènes  $\gamma$ - vers  $\beta$ -globine.

Cette interaction pourrait avoir lieu entre EKLF lié au promoteur de  $\beta$ -globine et des facteurs liés au LCR. Par exemple, le LCR comporte plusieurs sites de liaison pour GATA-1 qui interagit directement avec les facteurs de la famille Krüppel: EKLF et Sp1 (201,205). D'autre part, Sp1 peut former des homo-complexes et de l'ADN portant plusieurs sites de liaison pour Sp1 forme une boucle pour permettre la formation de ces complexes (225,226). EKLF, très proche parent de Sp1, pourrait faire la même chose avec le motif CACC du gène de  $\beta$ -globine et des protéines liant le LCR.

#### 1.4.14 Séquences du gène $\beta$ -globine impliquées dans l'activation par le LCR

Le gène de  $\beta$ -globine répond à la présence du LCR de plus d'une façon: les niveaux d'expression sont stimulés et l'expression devient indépendante de la position. Délimiter les séquences du gène qui sont capables de répondre au LCR pourrait clarifier le mécanisme d'activation du LCR. Par exemple, si une interaction directe se produit entre le LCR et le gène, nous pourrions trouver les régions impliquées par des études de délétion. En transfection dans les cellules MEL, il a été démontré que seulement le promoteur minimal, contenant les motifs CACC proximal, CAAT et TATA, était nécessaire pour la stimulation par le LCR (227). Le motif TATA à lui seul peut soutenir l'activation et l'ajout des motifs CAAT et CACC augmentent encore plus les niveaux de transcription. Par contre, les enhancers situés en 3' ne stimulent pas davantage l'expression en présence du LCR (167). Les motifs conservés du promoteur peuvent donc répondre à l'activation par le LCR alors que les enhancers deviennent redondants. Ces études ont aussi révélé que les niveaux d'expression sont beaucoup plus bas si le LCR est lié à un ADNc de  $\beta$ -globine qu'avec un gène complet (167). En rajoutant les introns devant le transgène, donc dans une position où il n'y aura pas d'épissage, on ne rétablit pas les niveaux normaux. Ce n'est donc pas la simple absence de ces séquences qui cause problème mais l'absence d'épissage comme tel.

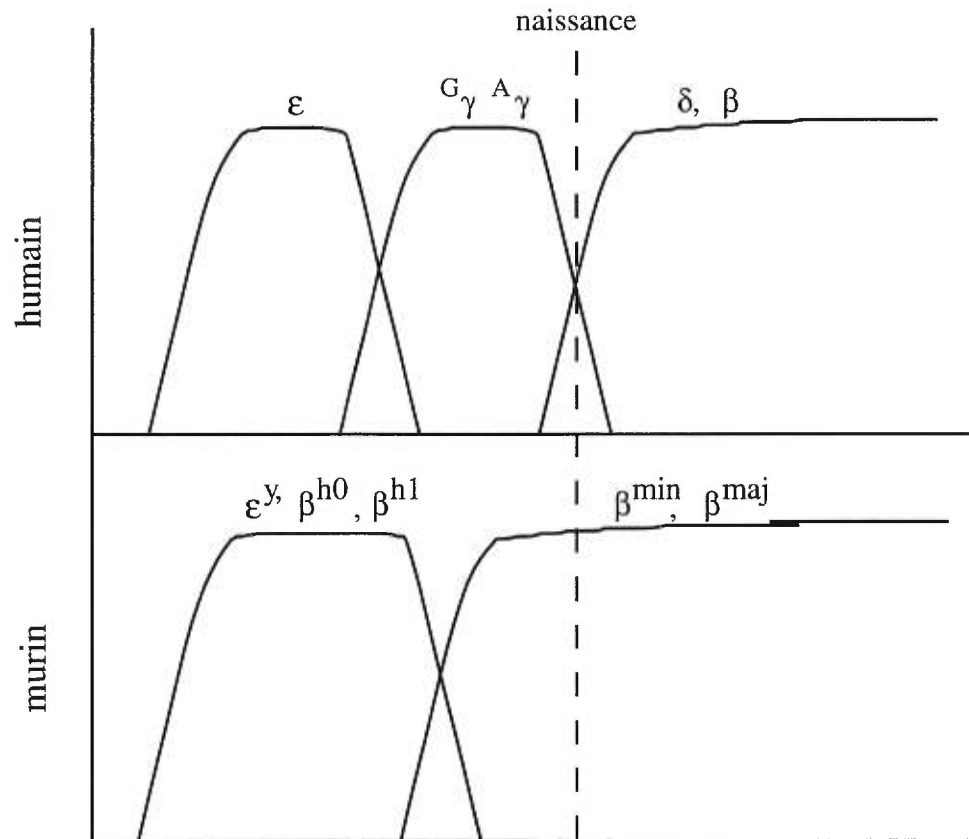


Ces études de cellules MEL transfectées avec des constructions du LCR permettent d'étudier l'activation de la transcription mais non les effets de position. En effet, la mesure de l'expression se fait dans population de clones différents et nous donne l'expression moyenne de tous ces clones, on pourra donc seulement noter une augmentation ou une diminution générale. Il faudrait tester l'expression dans des clones cellulaires individuels pour savoir si on peut étudier l'indépendance de la position en tranfection cellulaire.

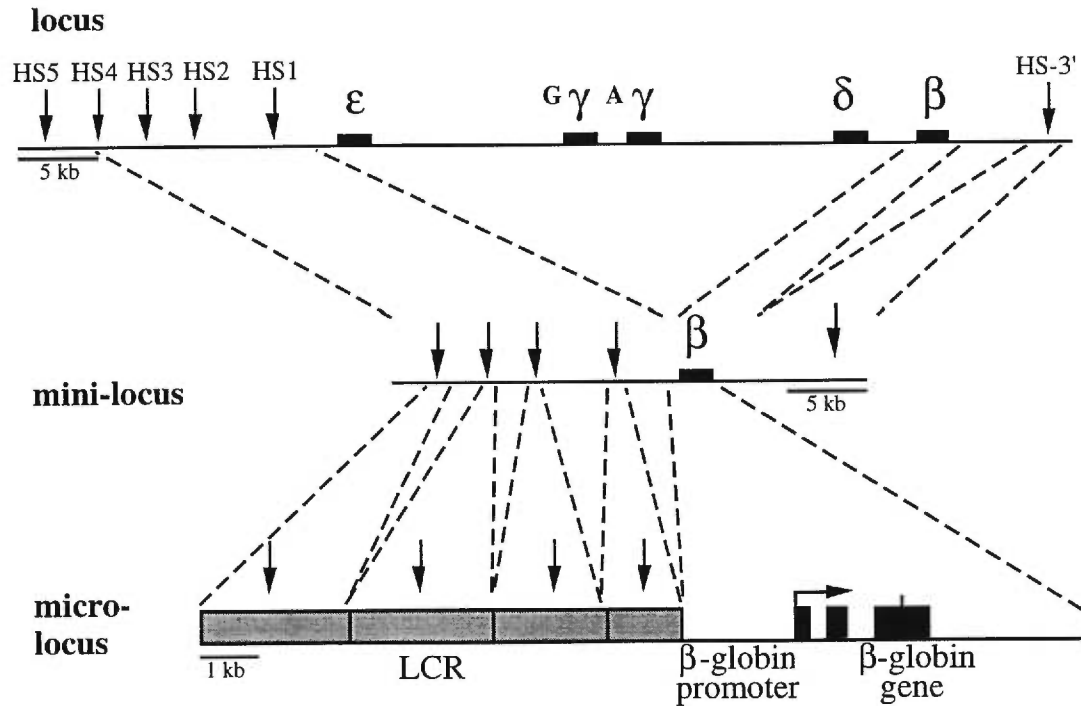
### **1.5 Problématique**

À l'origine, l'objectif général de cette thèse était de caractériser les éléments du gène de  $\beta$ -globine impliqué dans l'activation par le LCR au cours du développement. Dans le laboratoire du Dr Wall, Delvoye *et al.* avaient démontré que le motif CAAT du promoteur de  $\beta$ -globine permettait la stimulation par le LCR dans les cellules MEL murine au phénotype érythroïde adulte mais non dans les cellules K562 humaine au phénotype embryonnaire (228). Une des hypothèses qui pouvait expliquer cette différence est que l'activation du motif CAAT par le LCR est régulée dans le développement et est acquise dans le stade adulte. À la lumière de ce résultat, un des objectifs spécifiques de cette thèse était de tester la spécificité développementale de l'activation du motif CAAT par le LCR chez les souris transgéniques. Nous avons choisi de remplacer les séquences codantes de  $\beta$ -globine par le gène rapporteur *lacZ* encodant la  $\beta$ -galactosidase dont l'activité enzymatique facilite la quantification de l'expression. Ces transgènes permettait de plus de tester l'activité du LCR sur un gène non-globine, ce qui n'avait jamais été fait. L'utilisation de *lacZ* a aussi permis de vérifier s'il y avait variéation de l'expression.

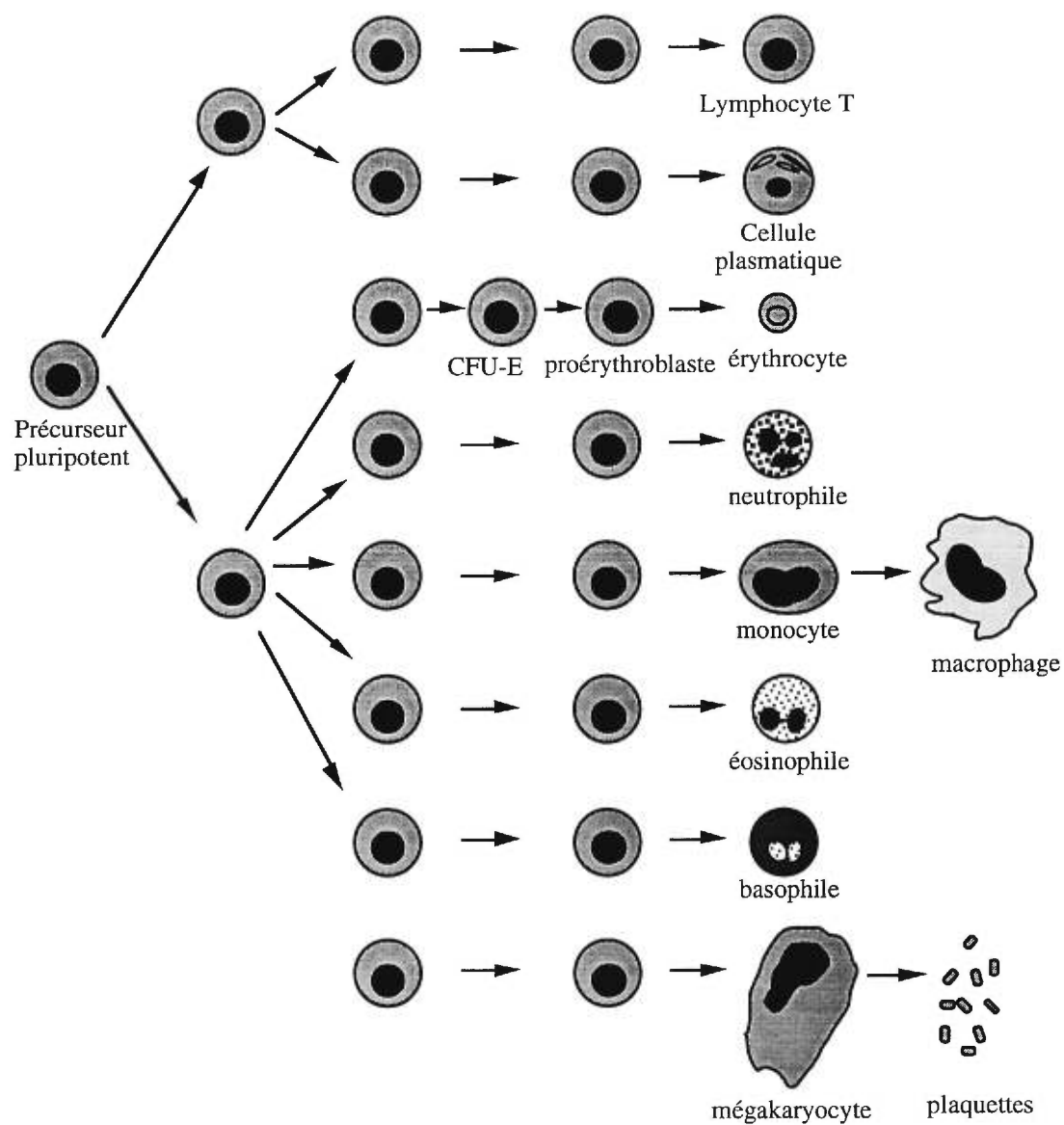
En présence du LCR, l'expression d'un gène de  $\beta$ -globine transfecté dans les cellules K562 est permise avec un promoteur complet (le LCR supplante la spécificité adulte conférée par les enhancers en 3' du gène) mais non avec un promoteur minimal ne contenant que les motifs CAAT et TATA. Donc, le LCR est capable de stimuler le gène  $\beta$ -globine par un élément situé entre le promoteur complet (-800) et le motif CAAT. Un objectif secondaire de cette thèse était d'identifier cet élément par des études de délétion.



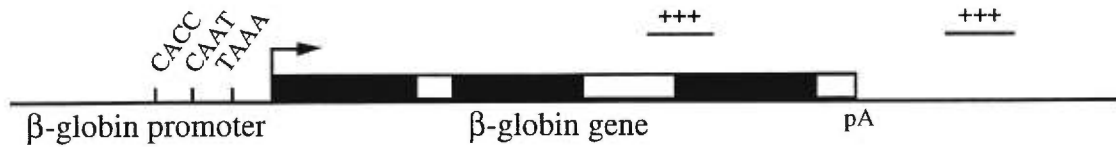
**Figure 1:** Expression des globines du type  $\beta$  au cours du développement. Chez l'humain (partie supérieure), deux transitions se produisent, l'une durant la vie foetale et la suivante à la naissance. Chez la souris, une seule transition a lieu au 11<sup>e</sup> jour de développement embryonnaire.



**Figure 2:** Locus de  $\beta$ -globine et les constructions mini- et micro-locus. Tous les gènes de globine du type  $\beta$  sont regroupés au même locus sur le chromosome 1. Dans les parties supérieures de la figure, les gènes sont représentés par des rectangles noirs et les sites hypersensibles par des flèches. Dans la partie inférieure, les sites hypersensibles HSS1 à HSS4 ont été liés directement au gène  $\beta$ -globine dont les exons sont représentés par des rectangles noirs.

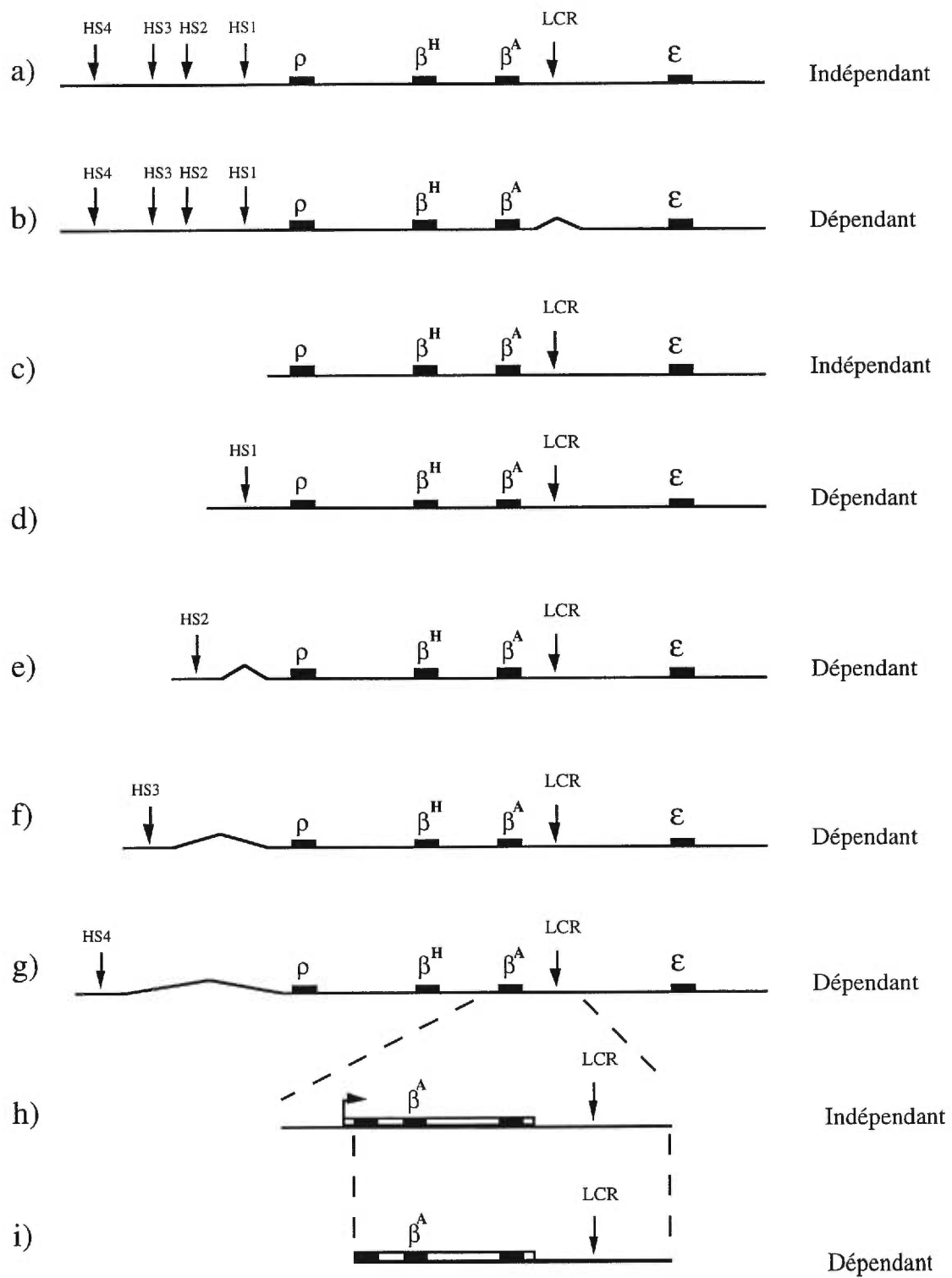


**Figure 3:** Schéma simplifié de la différenciation cellulaire au cours de l'hématopoïèse.

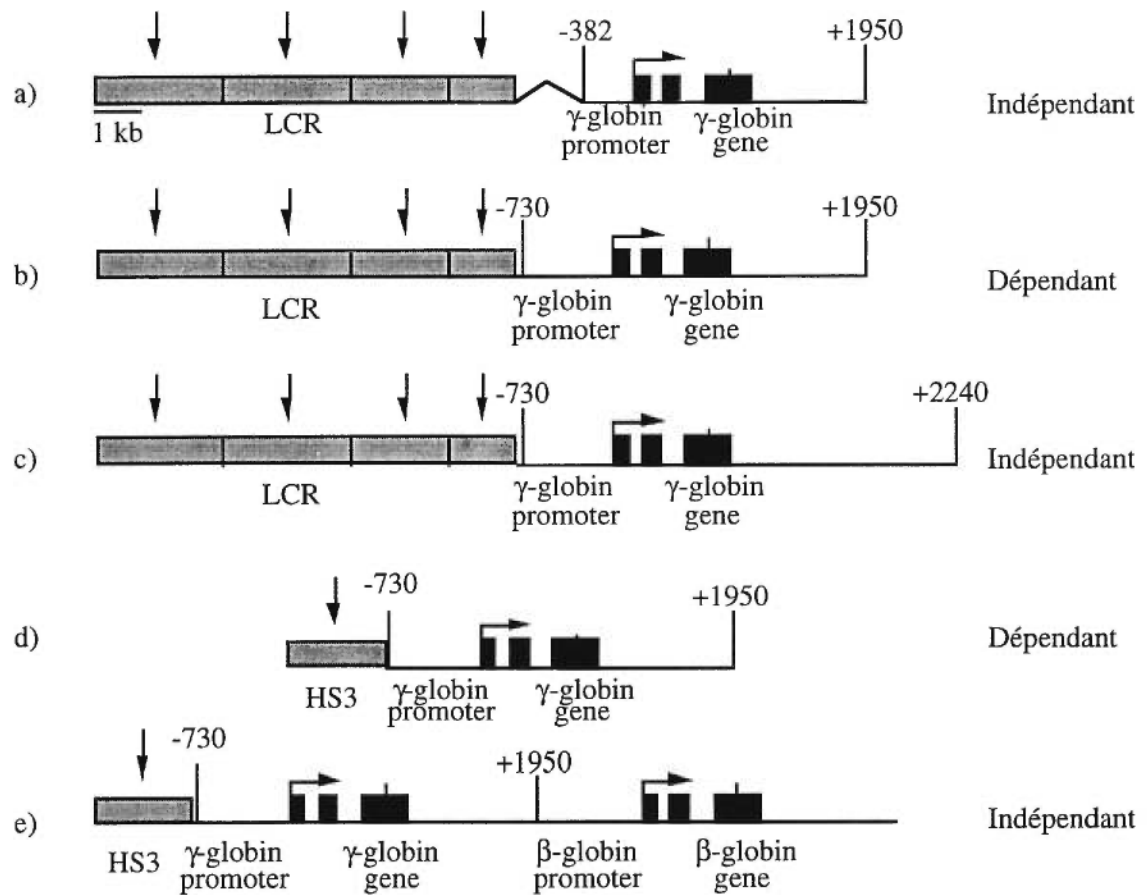


**Figure 4:** Structure du gène de  $\beta$ -globine. Les exons sont représentés en noir. Des enhancers de la transcription sont indiqués par les symboles (+++).

**Figure 5:** (Page suivante) Expression indépendante de la position au locus de  $\beta$ -globine de poulet. Il y a quatre sites hypersensibles en amont du locus, à la même position que le LCR humain. Toutefois, ces sites sont des enhancers et non un LCR. L'activité LCR se localise entre les gènes  $\beta^A$  et  $\epsilon$ . Les globines  $\rho$  et  $\epsilon$  sont embryonnaires alors que les globines  $\beta^H$  et  $\beta^A$  sont adultes. Chaque construction a été testé pour sa capacité de conférer une expression indépendante de la position.

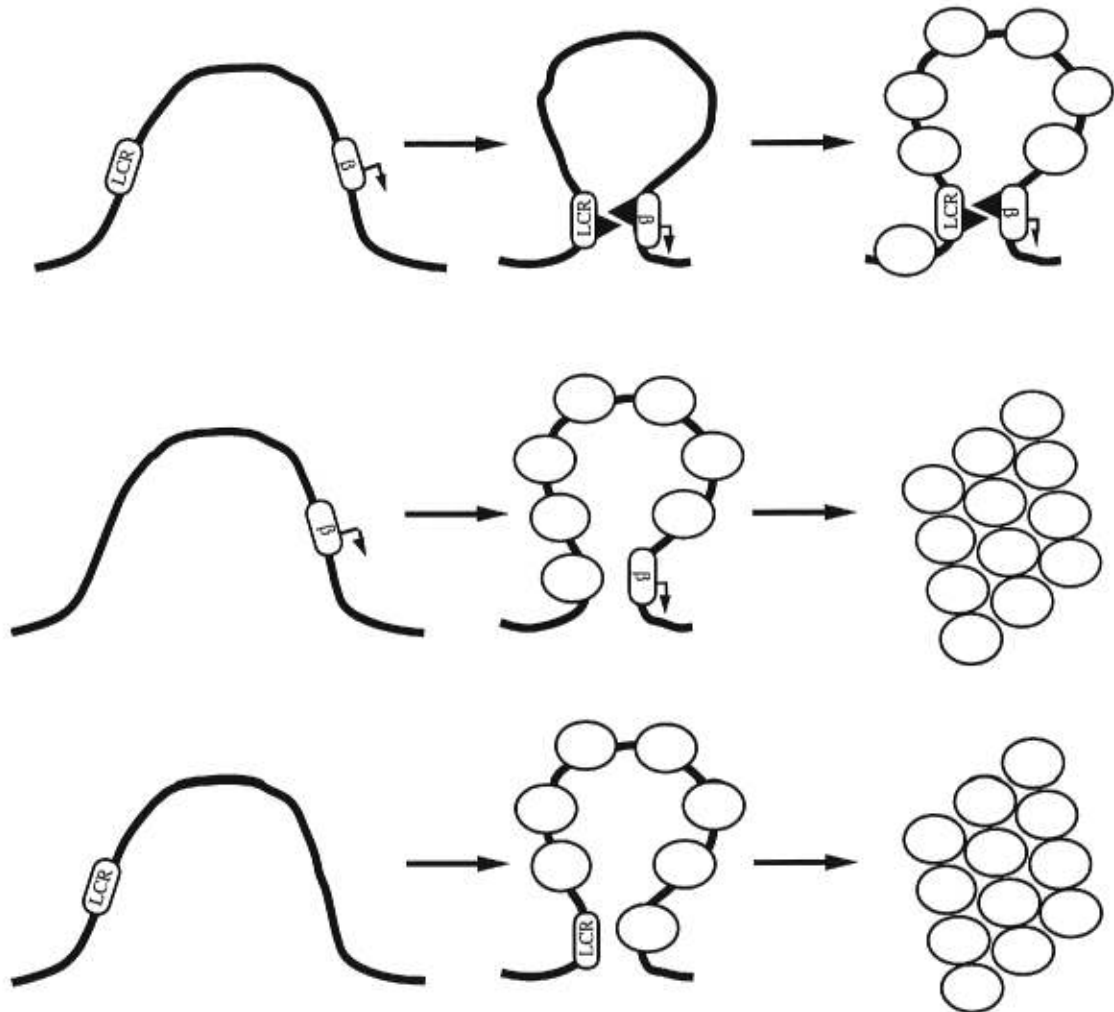


**Figure 5:** (Légende page précédente)



**Figure 6:** Effets de position sur l'expression d'un gène  $\gamma$ -globine humain. Selon les éléments du gène  $\gamma$ -globine liés au LCR du micro-locus, il y aura ou non des effets de position.

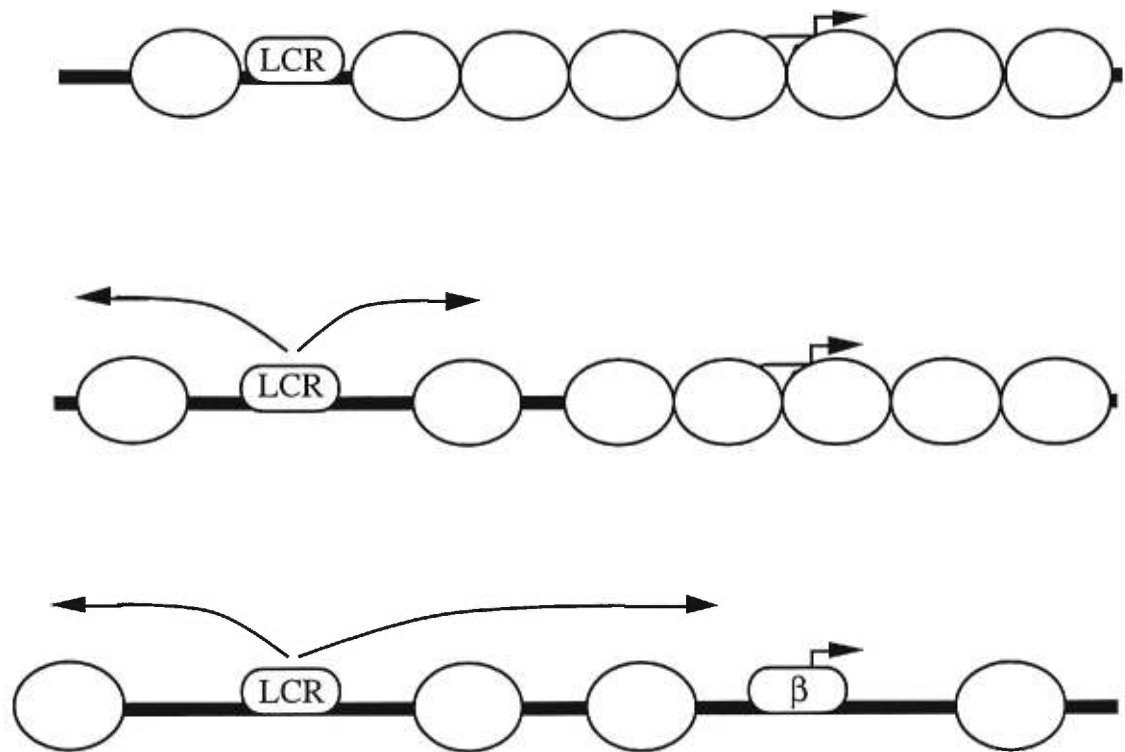
## Interaction mutuelle



**Figure 7:** Modèle d'interaction mutuelle. Ce modèle suggère que le LCR active la transcription du gène par contact en formant une boucle d'ADN. Par le fait même, le gène active le LCR qui requiert cette interaction pour ouvrir la chromatine. La figure n'est qu'une représentation schématique pour aider à la compréhension. La structure exacte de la chromatine est sans doute beaucoup plus complexe.



## Propagation



**Figure 8:** Modèles de propagation d'ouverture de la chromatine. Selon cette hypothèse, le LCR induit localement une décondensation de la chromatine qui se propage le long de la chromatine jusqu'à libérer le gène distal. La figure n'est qu'une représentation schématique pour aider à la compréhension. La structure exacte de la chromatine est sans doute beaucoup plus complexe.

## CHAPITRE 2

### Article 1

The  $\beta$ -globin locus control region enhances transcription of  
but does not confer position-independent expression  
onto the *lacZ* gene in transgenic mice.

Louis-Georges Guy, Rashmi Kothary, Yves DeRepentigny,  
Nathalie Delvoye, James Ellis et Lee Wall. (1996)  
*The Embo Journal*, **15**: 3713-3727.

**The  $\beta$ -globin locus control region enhances transcription of but does not confer position-independent expression onto the *lacZ* gene in transgenic mice.**

L.-G. Guy<sup>1</sup>, R. Kothary<sup>1,2</sup>, Y. DeRepentigny<sup>1</sup>, N. Delvoye<sup>1</sup>, J. Ellis<sup>3</sup> and L. Wall<sup>1,2\*</sup>

<sup>1</sup>Institut du cancer de Montréal, Centre de Recherche L.-C. Simard and <sup>2</sup>Department of Medicine, Université de Montréal, 1560 Sherbrooke Street East, Montreal, Quebec, Canada, H2L 4M1 and <sup>3</sup>Department of Genetics, Hospital for Sick Children, 555 University Avenue, Toronto, Canada M5G 1X8

Running title: *LacZ* interferes with LCR position independence

Key words:  $\beta$ -globin LCR/*lacZ*/position-independence/transgenic mice

\*To whom correspondence should be addressed:

tel.: 514-876-5497

fax.: 514-876-5476

E mail: Walll@ERE.UMontreal.CA

**Abstract:**

The  $\beta$ -globin locus control region (LCR) confers high levels of position-independent, copy number-dependent expression onto globin transgenes. Here >40 independent transgenic mouse lines and founders that carried the LCR in *cis* with the  $\beta$ -globin gene promoter driving a *lacZ* reporter gene were studied. Expression of the *lacZ* transgene was assayed by measuring  $\beta$ -galactosidase enzyme activity in fetal liver extracts, the levels of which correlated with the quantity of *lacZ* mRNA determined using RNase protection assays. Unexpectedly, expression of the *lacZ* transgene was found to show strong position effects, varying as much as 700 fold per transgene copy. These position effects occurred even if the whole  $\beta$ -globin gene was incorporated as part of the *lacZ* reporter gene. Moreover, DNase I hypersensitive sites appeared in the transgene LCR in high expressing but not in low expressing lines, suggesting that the LCR itself was position-dependent. In contrast, MEL cell clones, in which transcriptionally active integration sites were selected for, gave <13 fold variation in expression per copy of an LCR-*lacZ* construct. These results show that the *lacZ* reporter affects the ability of the LCR to activate chromatin in mice and that culture cells are not an adequate model for position-independent gene expression studies.

**Introduction:**

The human  $\beta$ -globin gene locus contains five active genes that are arranged in their developmental order of expression (Figure 1a).  $\epsilon$ -Globin is expressed only during erythropoiesis in the yolk sac of the early embryo, the two  $\gamma$ -globins are expressed at high levels in the red blood cells derived from the fetal liver, while  $\delta$ -globin, a minor contributor, and  $\beta$ -globin are expressed during erythropoiesis in the bone marrow late in fetal life and throughout adult life (reviewed in Stamatoyannopoulos and Nienhuis, 1994). To express at high levels, these genes rely on the activity of a locus control region (LCR). The human  $\beta$ -globin LCR, which is located 6-18 kb upstream from the  $\epsilon$ -globin gene (Figure 1a), is characterized by four strong DNase I hypersensitive site (HSS) regions that appear specifically in nuclei of erythroid cells (Tuan *et al.*, 1985; Forrester *et al.*, 1986).

The  $\beta$ -globin LCR has two important properties associated with it: a very strong erythroid-specific enhancer activity and the ability to confer position-independent, copy number-dependent expression onto globin genes under its control (reviewed in Townes and Behringer, 1990; Epner *et al.*, 1992; Dillon and Grosveld, 1993). Although it is still unclear how this latter property functions, it is sometimes referred to as chromatin opening activity. Globin transgenes that include the LCR are expressed in all independent transgenic mouse lines and erythroid cell clones (Grosveld *et al.*, 1987; Talbot *et al.*, 1989). The levels of expression are comparable with that of the endogenous globin genes and increase proportionally to the number of copies of the transgene present in each line, independently of where the transgene integrates into the genome. In contrast, globin transgenes without the LCR show strong position effects, which are characterized by very low and highly variable levels of expression with no relationship to copy number. In fact, 30%-70% of mice harbouring globin transgenes without the LCR give no detectable expression (Chada *et al.*, 1985; Costantini *et al.*, 1985; Townes *et al.*, 1985). LCRs have also been described for a number of other gene loci (see Greaves *et al.*, 1989; Palmiter *et al.*, 1993; Bonifer *et al.*, 1994; Madisen *et al.*, 1994; Talbot *et al.*, 1994; May and Enver,

1995 for examples ) and are expected to be essential for the correct expression of many genes.

Although only the HSS2 region of the  $\beta$ -globin LCR shows enhancer activity in transient assays, three of the four HSS regions (HSS2, HSS3, and HSS4) demonstrate enhancer activity in stably transfected erythroid culture cells and/or transgenic mice (Philipsen *et al.*, 1990; Talbot *et al.*, 1990; Pruzina *et al.*, 1991). The full enhancer activity of the LCR requires a combination of at least two and possibly all three of these HSS regions. DNA fragments containing each of the four HSS regions have been combined into a plasmid referred to as the microlocus cassette (Talbot *et al.*, 1989; Collis *et al.*, 1990). This microlocus retains full LCR enhancer activity and the ability to confer position-independent expression onto globin transgenes in mice.

The presence of DNase I hypersensitive sites in LCRs may reflect whether or not the LCR is situated in transcriptionally active chromatin (Reitman *et al.*, 1993; Huber *et al.*, 1994) and/or whether it is bound by specific transcription factors (Stamatoyannopoulos *et al.*, 1995). With the chicken  $\beta$ -globin LCR in *cis* with a globin gene void of a promoter, DNase I hypersensitivity in the LCR was observed in only 60% of transgenic mouse lines (Reitman *et al.*, 1993). This was in contrast to 100% of mice showing DNase I hypersensitive sites when the gene promoter was present. Thus, certain gene components may be necessary for the LCR itself to be active. In murine erythroleukemia (MEL) cells, with the  $\beta$ -globin promoter driving a  $\beta$ -globin cDNA in the microlocus vector, only very low levels of expression were obtained (Collis *et al.*, 1990). Addition of more  $\beta$ -globin gene sequences, including the second intron, was necessary to restore high expression. This was attributed to increased transcription rates as well as increased mRNA stability. Therefore, it also appears that particular gene sequences and/or structures may be necessary for the LCR to enhance transcription. However, in studies with MEL cells it was not determined whether expression was also position-independent. Hence, whether low expression with just a  $\beta$ -globin cDNA could also be due to an inability of the LCR to act in a position-independent manner is not known. Moreover, stably transfected populations and clones of MEL cells are selected with G418 for

expression of the *neo* gene marker in the microlocus plasmid (Talbot *et al.*, 1989; Collis *et al.*, 1990). Since the promoter of the *neo* gene must be transcriptionally active, one only selects cells in which this gene has been activated. Thus, cell culture systems may not be ideal for studies regarding position-independent gene expression.

Here it is shown that the  $\beta$ -globin LCR is incapable of conferring position-independent expression onto a *lacZ* reporter transgene in mice, even when the whole  $\beta$ -globin gene is included in the transcription unit. Yet, LCR enhancer activity in the *lacZ* gene containing constructs is at least partially maintained in mice and the same LCR-*lacZ* constructs express at near position-independent, copy number-dependent levels in MEL cells. These results show unequivocally that MEL cells are not an adequate model for position-independent studies. The data are discussed in relation to previous results regarding the mechanism of LCR activation.

## **Results:**

The  $\beta$ -globin LCR does not confer position-independent expression onto a *lacZ* reporter transgene.

In the transgenic mice studies described below, the  $\beta$ -globin gene promoter from -800 to +32 [relative to the transcription start (Cap) site] or a  $\beta$ -globin promoter deleted to -87, containing only the TATA and CAAT box motifs of the minimal promoter, was used. The promoter fragments were linked to a *lacZ* reporter gene (which codes for the  $\beta$ -galactosidase enzyme) and this was followed by SV40 polyadenylation signals. These reporter constructs were cloned into the microlocus cassette described by Collis *et al.*, (1990). The final constructs, denoted  $\beta$ A and  $\beta$ B, are shown in Figure 1. Throughout the results when we refer to LCR we mean as it exists in the microlocus cassette.

Five and seven stable transgenic lines for the  $\beta$ A and  $\beta$ B constructs, respectively, were established. Adult blood cells isolated from transgenic animals were incubated in the presence of X-gal to detect  $\beta$ -galactosidase activity *in situ*. Three of the five  $\beta$ A lines ( $\beta$ A2,  $\beta$ A4 and  $\beta$ A5) and all seven of the  $\beta$ B lines demonstrated blue colouration of some of the red blood cells in this assay. An example of this is shown in the top panel of Figure 2. The percentage of cells that coloured blue varied considerably between the different lines, although it varied very little between animals of the same transgenic origin. No red blood cells stained in two of the  $\beta$ A lines ( $\beta$ A1 and  $\beta$ A3) (not shown). When whole 15.5 day transgenic fetuses from the  $\beta$ A and  $\beta$ B lines were similarly stained *in situ*, the circulatory system and the fetal liver, the site of erythropoiesis at this stage of development, were coloured blue (Figure 2, middle and bottom panels). The intensity of staining of the circulatory system and fetal liver varied among the different transgenic lines, and the two  $\beta$ A lines ( $\beta$ A1 and  $\beta$ A3) that gave no detectable staining of adult red blood cells also gave no colouration at the fetal stages (not shown). No blue colouration was observed outside of the circulatory system and fetal liver for any of the  $\beta$ A and  $\beta$ B lines. Thus, expression of the *lacZ* reporter gene was erythroid-specific, as would be



expected for a  $\beta$ -globin LCR driven transgene (Grosveld *et al.*, 1987; Blom van Assendelft *et al.*, 1989).

To determine the expression level per transgene copy, the copy number was quantified on Southern blots of genomic DNA (Figure 3) and  $\beta$ -galactosidase activity was measured in extracts prepared from fetal liver. The results are summarized in Table I. The two  $\beta$ A lines that showed no staining ( $\beta$ A1 and  $\beta$ A3) gave no measurable  $\beta$ -galactosidase activity (Table I). In contrast, the other three  $\beta$ A lines gave relatively high  $\beta$ -galactosidase activity ( $\beta$ A2,  $\beta$ A4, and  $\beta$ A5 lines in Table I).

Four of the  $\beta$ B lines ( $\beta$ B3,  $\beta$ B4,  $\beta$ B5, and  $\beta$ B6) gave an expression per transgene copy that was comparable to or greater than the activity seen with the expressing  $\beta$ A lines. On the other hand, three of the  $\beta$ B lines ( $\beta$ B1,  $\beta$ B2, and  $\beta$ B7) gave relatively low expression per transgene copy. Thus, for both the  $\beta$ A and  $\beta$ B lines  $\beta$ -galactosidase activity levels were highly variable and did not correlate to transgene copy number.

To ensure that none of the lines had sustained deletions and or rearrangements of the transgenes, a detailed DNA analysis was done for all of the  $\beta$ A and  $\beta$ B lines (data not shown). Genomic DNA was digested separately with *Sac*I and *Xba*I, which cut the transgene constructs into nine and three fragments, respectively. Southern blots of the digested DNA were hybridized in succession with a number of probes that cover the entire construct (not shown). All fragments of the correct size were detected in the lines included in the analyses. Taken together, these results suggest that the  $\beta$ -globin LCR does not confer position-independent expression onto the *lacZ* gene in mice, at least when only a  $\beta$ -globin promoter is present.

We also determined whether DNase I hypersensitive sites were present in the transgene LCR of different transgenic lines. Hypersensitive sites in the transgene human  $\beta$ -globin LCR were detected in fetal liver nuclei isolated from two high expressing lines ( $\beta$ B5 and  $\beta$ B6; Figure 4, panels A and B) but were not detectable in nuclei isolated from two low expressing lines ( $\beta$ B1 and  $\beta$ B7; panels C and D). As a positive control, DNase I hypersensitive sites in the endogenous mouse  $\beta$ -globin LCR were detected in these latter

two cases using the same DNA (panels E and F). Therefore, in terms of hypersensitive sites, the LCR itself was demonstrating position effects in these LCR-*lacZ* transgenic mice.

Addition of the whole  $\beta$ -globin gene does not restore position-independent expression to the *lacZ* reporter.

As it was found previously that a  $\beta$ -globin cDNA linked to the  $\beta$ -globin promoter in the microlocus cassette was not expressed at high levels in MEL cells unless additional  $\beta$ -globin gene sequences were included (Collis *et al.*, 1990), we tested the effect of adding further  $\beta$ -globin gene sequences to the *lacZ* reporter gene. Three different constructs were tested. In  $\beta$ C (Figure 1), a DNA fragment from the *Bam*HI site in the second exon of the  $\beta$ -globin gene to the *Eco*RV site located 1.7 kb past the polyadenylation signals of the gene was cloned after the SV40 polyadenylation signals in the  $\beta$ B construct. This  $\beta$ -globin fragment includes the second intron, third exon, polyadenylation signals and the downstream enhancer that was mapped previously (Behringer *et al.*, 1987; Kollias *et al.*, 1987; Trudel and Costantini, 1987). However, since the SV40 polyadenylation signals will be used in this construct, the downstream  $\beta$ -globin gene will not be part of the unprocessed mRNA. In  $\beta$ D (Figure 1) the same  $\beta$ -globin gene fragment was cloned 3' of *lacZ* in the  $\beta$ B construct, but the SV40 polyadenylation signals were removed in this case so that the  $\beta$ -globin sequences will become part of the transcription unit. The second intron will be spliced out and the  $\beta$ -globin gene polyadenylation signals will be used. Although a hybrid mRNA will be produced, the  $\beta$ -galactosidase protein synthesized will be the same because the *lacZ* translational stop site is present. In the third construct ( $\beta$ E), the *lacZ* gene, without a promoter or SV40 polyadenylation signals, was cloned directly into the *Nco*I site situated at the protein initiation codon (ATG) in exon 1 of the  $\beta$ -globin gene. Thus, the whole  $\beta$ -globin gene is present in this case. Introns 1 and 2 will be spliced from the precursor message and the  $\beta$ -globin polyadenylation signals will be used. For these three constructs, fetal livers of founder embryos were studied directly in most cases, although a few lines were also bred. When founder fetal livers were tested directly it is possible that some animals were mosaic. However, this should not have affected the

results as the transgene copy number was determined using fetal liver DNA in these cases; the same tissue used to measure  $\beta$ -galactosidase activity. If the fetal liver was mosaic for the transgene, expression and copy number would be affected in the same manner. All transgenes were also tested for deletions as described above. None were detected in transgenic animals that are included in the analyses.

Three lines carrying the  $\beta$ C construct were established and between them there was a 15 fold variation in expression per transgene copy (compare  $\beta$ C2 and  $\beta$ C3 in Table II). Many more lines and/or founders carrying the  $\beta$ D or  $\beta$ E transgenes were generated and the level of  $\beta$ -galactosidase expression per copy was seen to vary enormously for both of these constructs (Table II). For  $\beta$ D, a >700-fold variation was observed (compare  $\beta$ D9 and  $\beta$ D13), while for  $\beta$ E a 300 fold variation was obtained (compare  $\beta$ E6 and  $\beta$ E9). Moreover, several of the  $\beta$ D and  $\beta$ E transgenic founders ( $\beta$ D6,  $\beta$ D12,  $\beta$ E5, and  $\beta$ E9 in Table II) gave total levels of activity that were near or only within a few fold of the background, which was determined to be 1 mU/mg. Thus, although all the transgenic derivatives with the  $\beta$ D and  $\beta$ E constructs demonstrated detectable levels of expression, in addition to showing a high variation in the level of expression per transgene copy, the absolute level of expression in some cases was extremely low.

$\beta$ -galactosidase activity levels in extracts corresponds to *lacZ* mRNA levels.

To test whether the levels of  $\beta$ -galactosidase activity in extracts reflected the levels of transgene message, fetal liver RNA from several  $\beta$ A lines and  $\beta$ D founders was assayed by RNase protection for *lacZ* mRNA expression and for mouse  $\beta$ -major globin expression as an internal control (Figure 5). The  $\beta$ -major signal varied only slightly among the different samples. In contrast, the *lacZ* signal varied tremendously and this variation correlated with the amount of  $\beta$ -galactosidase activity measured in extracts, which is given at the bottom of the lanes in Figure 5 for comparison. Thus, the  $\beta$ -galactosidase activity accurately reflected the amount of *lacZ* mRNA produced in the different transgenic animals.

The LCR does confer position-independent expression onto the  $\beta$ -globin gene itself.

The microlocus LCR does confer position-independent, copy number-dependent expression onto the  $\beta$ -globin gene itself in transgenic mice (Talbot *et al.*, 1989). However, the version of the microlocus used in that study was slightly different than that designed by Collis *et al.*, (1990). To ensure the microlocus LCR used in the present report could confer position-independent expression onto the  $\beta$ -globin gene alone and to determine what variation in expression per copy would be obtained, the human  $\beta$ -globin gene alone in the same microlocus was studied ( $\beta$ F construct in Figure 1). Founder transgenic fetal livers were tested for expression of the human  $\beta$ -globin transgene mRNA relative to the endogenous mouse  $\beta$ -major mRNA by S1 protection analysis (Figure 6). Six out of six independent founders showed high level expression of the transgene and there was <4-fold variation in expression per transgene copy (summarized in Table III). Thus, this version of the LCR does confer position-independent expression onto the  $\beta$ -globin gene.

In drug-resistant-selected MEL cell clones, a LCR-*lacZ* transgene shows a much closer relationship between copy number and expression levels.

For comparative purposes, the  $\beta$ D construct, for which we had obtained the most independent transgenic samples (Table II), was also transfected into MEL cells. In this case, the vector contained the G418 drug selectable gene *neo* that was removed in preparing DNA for the generation of transgenic mice. Individual, stably transfected clones were isolated and DNA was analyzed by Southern blot to determine transgene copy number and to test for deletions. Clones that showed no deletions were induced to differentiate with dimethylsulfoxide and extracts from the cells were measured for  $\beta$ -galactosidase activity (Table IV). All of the clones gave relatively high levels of enzymatic activity. Moreover, the levels of expression per transgene copy correlated to a much better degree than the same construct studied in transgenic mice. There was <13-fold variation in expression on a per transgene copy basis in MEL clones versus a

>700-fold variation for the same construct studied in transgenic mice (compare results for  $\beta$ D construct in Table IV to those in Table II).

## **Discussion:**

### Cultured cells are not a good model for position-independent studies.

Although the microlocus LCR confers position-independent, copy number-dependent expression onto the  $\beta$ -globin gene in transgenic mice, with less than a four fold variation in expression per transgene copy (Talbot *et al.*, 1989, and see Table III), we demonstrated here that LCR constructs running a *lacZ* reporter gene in mice give up to a 700 fold variation in expression of  $\beta$ -galactosidase activity per transgene copy (Tables I and II). This occurs even when the entire  $\beta$ -globin gene is incorporated as part of the *lacZ* reporter gene construct. Recently, it has been suggested that a *lacZ* reporter gene in *cis* with the major regulatory element of the human  $\alpha$ -globin gene locus gives cell variegated expression in transgenic mice (Robertson *et al.*, 1995). This results in variable levels of expression between different transgenic lines. However, the  $\beta$ -globin LCR-*lacZ* constructs studied in the present report do not show cell-variegated expression (manuscript submitted). Our results show that the  $\beta$ -globin LCR is unable to confer position-independent, copy number-dependent expression onto the *lacZ* reporter gene in transgenic mice.

In contrast to the results in transgenic mice, a LCR-*lacZ* construct gave only 13-fold variation in expression per copy in stably transfected clones of MEL cells (Table IV). There are at least two reasons why such a difference between the mouse and cell culture models may exist: (i) Stably transfected cell clones are drug selected. For a clone to be able to grow in the presence of drug, the promoter of the selectable gene must be at least mildly active. If it happened that the LCR were unable to open chromatin itself, one would then only select clones in which the transgene had integrated into sites of chromatin that are normally transcriptionally active. In such cases, the LCR enhancers may be able to function at high levels independently from the LCR chromatin opening activity. On the other hand, both transcriptionally inactive and active sites of chromatin integration are obtained in transgenic mice, as no selection for the transgene is used. Thus, transgene expression would be expected to show position effects in mice if the LCR were unable to

open chromatin properly. (ii) Alternatively, the LCR in our constructs may be more able to properly open chromatin and allow the LCR enhancers to function in a position-independent manner in MEL cells than in transgenic mice. During development, chromatin remodelling occurs at different stages and acts to activate and repress different regions of chromatin in a tissue-specific manner (Boulikas, 1991; Zlatanova and Holde, 1992). For the LCR to function fully in erythroid cells and in any chromatin site in the developing organism it must be able to override the mechanisms that are attempting to repress gene expression at each developmental stage. On the other hand, MEL cells represent a single and late (adult) stage of mouse development. The requirements of the LCR to open and maintain chromatin in an active configuration may therefore be less stringent in MEL cells than in transgenic mice. Whatever the reason, our results unequivocally show that a LCR containing construct that does not give position-independent expression in transgenic mice can give a false indication of position-independence in cell culture clones. Thus, drug-selected cell culture clones are not a good model for studies regarding position-independent expression. Others have also demonstrated that culture cells may sometimes be a poor model for gene expression studies. For example, Zimmerman *et al.* (1990) found that expression of N-*myc* transgenes was tissue restricted in mice but not in culture cells.

Why does the LCR not confer position-independent expression onto the *lacZ* reporter gene in transgenic mice?

Previously, when globin genes were studied without the LCR it was found that the  $\beta$ -globin promoter alone, as the only defined regulatory sequence, was not sufficient for expression in transgenic mice (Behringer *et al.*, 1987; Trudel and Costantini, 1987). However, with the  $\beta$ -globin promoter extending to -800 linked to the *lacZ* gene ( $\beta$ B construct), or with the promoter extending to just -87 ( $\beta$ A construct), with no additional regulatory elements other than the LCR, seven of seven and three of five transgenic lines, respectively, gave detectable  $\beta$ -galactosidase activity in mice (Table I). Thus, the enhancers of the LCR can work at least partly in transgenic mice with constructs containing the *lacZ* gene. Moreover, the *lacZ* reporter gene was expressed at high levels

in every MEL cell clone and in a near copy number-dependent fashion (Table IV). Thus, in the context of active chromatin, regardless of whether or not it was the LCR that provided the active chromatin in MEL cells, the LCR enhancers are capable of invoking high levels of transcription onto the *lacZ* reporter gene. However, *lacZ* expression in transgenic mice was very influenced by position, suggesting that in the presence of the *lacZ* reporter gene the chromatin opening activity of the LCR is definitely impaired.

There may be many reasons why *lacZ* is able to interfere with the chromatin opening activity of the LCR in mice but still allow the enhancers to partly function. Two possibilities, that are dealt with individually below, are suggested by previous findings. The first possibility is that the LCR chromatin opening activity and its enhancer properties are partially or entirely separable functions that might involve different mechanisms. If this were the case, the *lacZ* sequences might be able to interfere with the chromatin opening function without affecting enhancer activity directly. That the two LCR properties are separable entities is suggested by the finding that the element of the chicken globin locus that can confer position-independent expression has only weak enhancer activity (Reitman *et al.*, 1990), and by the recent demonstration by one of us (Ellis *et al.*, 1996) that the chromatin-opening activity of the human  $\beta$ -globin LCR, as measured by the ability to give position-independent expression in mice containing single copy transgenes, resides solely in the HSS3 region. Thus, the chromatin opening activity in the human  $\beta$ -globin LCR is both physically and functionally separate from the enhancers in HSS2 and HSS4. It remains to be determined if it is also separable from the enhancer activity of HSS3.

Do the chromatin opening activity and enhancer properties of the LCR involve different mechanisms? To explain why the globin genes compete for activation by the LCR during development, it has been suggested that the LCR functions through a looping mechanism, in which it directly interacts with the gene it is activating (Hanscombe *et al.*, 1991). This model has been strongly supported by further experimentation (Kim *et al.*, 1992; Fiering *et al.*, 1995). However, in most investigations, only the relative level of gene expression is considered. As the levels of transcription only reflect enhancer activity



*per se*, the studies have not directly addressed whether the chromatin opening activity may function through a similar mechanism to the enhancers. In fact, the demonstration that a  $\gamma$ - and  $\beta$ -globin gene together in *cis* with the LCR are both expressed in a position-independent manner at all stages of development, although the genes do show a developmental-specific competition for levels of expression (Li and Stamatoyannopoulos, 1994), suggest the genes compete for the enhancer activity but not for the chromatin opening activity of the LCR. This would be explained if LCR chromatin opening activity involves a mechanism different from that of the enhancers. As an example of an alternative to the looping model, a scanning mechanism has been suggested for other systems (Migeon, 1994; Hecht *et al.*, 1995). In a scanning scenario a change in chromatin structure may originate and propagate from the LCR. There may exist sequences in the *lacZ* gene that block the necessary alteration in chromatin from passing through them, but do not directly prevent the LCR enhancers from interacting with the  $\beta$ -globin promoter. This would result in the position-dependent expression in transgenic mice we observe with LCR-*lacZ* constructs.

The second plausible explanation as to why the  $\beta$ -globin LCR is unable to activate the *lacZ* reporter gene in a position-independent manner is that the LCR chromatin opening activity is spatially distributed and requires a specific arrangement of defined regulatory sequences. For example, although the chicken  $\beta$ -globin gene in *cis* with its downstream enhancer (LCR) is expressed independently from position in mice, addition of any one of four upstream HSS regions, but not all four together, causes the gene to be transcribed in a position-dependent manner (Reitman *et al.*, 1995). In addition, whereas a human  $\gamma$ -globin gene with its promoter to -382, in *cis* with the LCR, is expressed independently of position, extension of the promoter to -730 prevents it, while also including the  $\beta$ -globin gene in 3' permits the recovery of position-independent expression (Li and Stamatoyannopoulos, 1994). Thus, LCR chromatin opening activity may need specific regulatory sequences at both the LCR and gene level, as well as a specific spatial arrangement of these sequences. Although the  $\beta$ -globin gene would be expected to provide the necessary gene elements, as it is activated in a position-independent fashion by

the LCR, inserting the *lacZ* gene into the  $\beta$ -globin gene might disturb the spatial arrangement required and/or prevent interactions between the different elements.

The two possibilities discussed above are not mutually exclusive. The LCR chromatin opening activity may involve a different mechanism from that of the enhancer activity and at the same time need a specific arrangement of regulatory sequences to function properly. In any case, our results have shown that the LCR does not confer position-independent expression onto just any gene in transgenic mice. Future work needs to be focused on determining what is required at the gene level for the LCR to be able to confer position-independent expression and/or whether chromatin opening functions as a separate entity from enhancement. The *lacZ* gene may be a useful tool in this regard.

## **Materials and methods:**

### Transgenic mouse production.

Microlocus constructs were digested with *Sac*II to remove plasmid and *neo* gene sequences, the DNA fragments generated were separated by agarose gel electrophoresis, and the fragment of interest was purified using GeneClean II (Bio/Can Scientific). Transgenic mice were produced as described (Bérard *et al.*, 1994). Transgenic founders were either allowed to go to term and bred to produce stable lines or sacrificed 14 or 15 days following microinjection.

### DNA analysis.

For bred lines, tail DNA was used to determine copy number and transgene integrity. Fetal liver DNA was used for non-bred founders. The  $\beta$ A5 line was deemed to contain a single transgene copy by end fragment analysis (not shown), and was used as a standard to determine copy number for all other founders and lines. Fifteen  $\mu$ g of genomic DNA was digested with *Sac*I and analyzed by Southern blot, hybridizing successively to a 2 kb *Xba*I fragment from the 5' end of the LCR in the microlocus vector and a 1.3 kb *Sac*I fragment from the *dystonia musculorum* (*dt*) locus (Brown *et al.*, 1995) as a loading control. The signals were quantified by scanning autoradiograms with an LKB Ultra Scan laser densitometer linked to a LKB 2200 integrator.

### $\beta$ -galactosidase assays.

Transgenic fetuses were isolated 15 days after observation of the spermatic plug for bred fetuses or 15 days after microinjection for founder fetuses. For *in situ* detection of  $\beta$ -galactosidase activity, fetuses were fixed 1 hour in the dark at 4°C in 1 % formaldehyde, 0.2 % glutaraldehyde, 0.02 % NP-40, and 1 mM MgCl<sub>2</sub> in phosphate buffered saline (PBS). The fetuses were rinsed twice in PBS and then incubated for the indicated time in the dark at 37°C in PBS containing 0.01 M potassium ferrocyanide, 0.01 M potassium ferricyanide, 0.1 M 5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactoside (X-gal) and 1 mM MgCl<sub>2</sub>. For adult blood cells, a few drops of tail blood were collected, the cells were

fixed in 0.2 % glutaraldehyde in PBS for 10 minutes, rinsed in PBS, and incubated 18 hours in X-gal solution as above. MEL cells were assayed *in situ* as described for adult blood.

To measure  $\beta$ -galactosidase activity in extracts, a quarter of a fetal liver was pipetted through a plastic pipette tip to disperse the cells into 100  $\mu$ l of 0.25 M Tris-HCl (pH 8). Similarly,  $1 \times 10^5$  induced MEL cells were diluted in 100  $\mu$ l of the same buffer. The suspensions were frozen on dry ice and thawed at 37° three times and the insoluble matter was removed by centrifugation for 5 minutes at 4°C. Protein concentration was estimated using the Bio-Rad Protein Assay system (Bio-Rad Laboratories) with bovine serum albumin as a standard. Twenty  $\mu$ g of extract protein was assayed for  $\beta$ -galactosidase activity as described (Sambrook *et al.*, 1989) and activity was measured relative to a commercial  $\beta$ -galactosidase standard (Boehringer Mannheim). All assays were performed under linear conditions and several concentrations of the standard were assayed for different times in each case.

#### Hypersensitive site detection.

DNase I hypersensitive sites were measured as described (Forrester *et al.*, 1990) with modifications. Three to five fetal livers were homogenized in 5 ml of 10 mM Tris-HCl (pH 7.5) containing 10 mM NaCl and 0.1 % NP-40 on ice and 20 ml of the same buffer without NP-40 was added. Nuclei were pelleted and resuspended into 1.6 ml of 10 mM Tris-HCl (pH 7.5) containing 10 mM NaCl and 1 mM CaCl<sub>2</sub>. Aliquots were incubated for 5 minutes at 37°C with 0, 0.1, 0.4, 1, 4 or 16  $\mu$ g of DNase I (Boehringer Mannheim). Genomic DNA was isolated and digested with *Cla*I and *Xho*I, and Southern blots of the DNA were hybridized with a 1 kb *Not*I-*Cla*I fragment (HSS1) from the human LCR in the microlocus vector. Alternatively, the DNA was digested with *Pst*I and Southern blots were hybridized with a 1.1 kb *Eco*RI-*Pst*I fragment from the mouse LCR (Jimenez *et al.*, 1992).

### RNA analysis.

The  $\beta$ F founder fetal livers (Table III) were isolated at 14 days following microinjection. Cells from half of a liver were dispersed in 500  $\mu$ l of Trizol (Gibco-BRL) and RNA was isolated as described by the manufacturer. DNA from the cell debris in the organic phase was precipitated with 1 ml of ethanol and purified by usual procedures for tissues. S1 analysis was done as described (Ellis *et al.*, 1993). The  $\beta$ A and  $\beta$ D fetal livers (Figure 5) were isolated at 15.5 days and RNA purified as above. RNase protection assays were performed as described previously (Delvoye *et al.*, 1993). The antisense RNA probes used represented a *lacZ* gene fragment from *Cla*I to *Eco*RV (+1129 to +1417 relative to the translation initiation site) in the *lacZ* gene that protects a 288 base RNA fragment and a  $\beta$ -major fragment from *Sau*3A to *Hga*I (-15 to +60 relative to the Cap site) of the mouse  $\beta$ -major gene that protects a 60 base RNA fragment.

### Cell culture.

MEL (C88) cells were grown and transfected, as previously described (Delvoye *et al.*, 1993). Single clones were selected by serially diluting cell suspensions to contain on average 1 cell per 200  $\mu$ l in 20% conditioned medium, plating 200  $\mu$ l in individual wells of 96 well plates, and microscopically testing for individual cells and colonies for several days. Selected clones were induced with 2 % dimethylsulfoxide for 5 days and protein extracts for  $\beta$ -galactosidase measurements were prepared.

**Acknowledgments:**

This work was supported by a grant from the National Cancer Institute of Canada to L. W. and grants from the Medical Research Council of Canada to R. K. and J. E.. L. W. and R. K. are recipients of a scholarship from the Fonds de Recherche en Santé du Québec.

**References:**

Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. and Townes, T.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7056-7060.

Bérard, J., Gaboury, L., Landers, M., De Repentigny, Y., Houle, B., Kothary, R. and Bradley, W.E. (1994) *EMBO J.*, **13**, 5570-5580.

Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989) *Cell*, **56**, 969-977.

Bonifer, C., Yannoutsos, N., Kruger, G., Grosveld, F. and Sippel, A.E. (1994) *Nucleic Acids Res.*, **22**, 4202-4210.

Boulikas, T. (1991) *Anticancer Research*, **11**, 489-527.

Brown, A., Bernier, G., Mathieu, M., Rossant, J. and Kothary, R. (1995) *Nature Genetics*, **10**, 301-306.

Chada, K., Magram, J. and Costantini, F. (1985) *Nature*, **319**, 685-689.

Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.*, **9**, 233-240.

Costantini, F., Radice, G., Magram, J., Stamatoyannopoulos, G., Papayannopoulou, T. and Chada, K. (1985) *Cold Spring Harbor Symposia on Quantitative Biology*, **50**, 361-370.

Delvoe, N.L., Destroismaisons, N.M. and Wall, L.A. (1993) *Mol. Cell. Biol.*, **13**, 6969-6983.

Dillon, N. and Grosveld, F. (1993) *Trends Gen.*, **9**, 134-137.

Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993) *EMBO J.*, **12**, 127-134.

Ellis, J., Talbot, D., Tan-Un, K.C., Harper, A., Michalocich, D., Yannoutsos, N., Philipsen, S. and Grosveld, F. (1996) *EMBO J.*, In press.

Epner, E., Kim, C.G. and Groudine, M. (1992) *Curr. Biol.*, **2**, 262-264.

Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I.K., Enver, T., Ley, T.T. and Groudine, M. (1995) *Genes Dev.*, **9**, 2203-2213.

- Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) *Genes Dev.*, **4**, 1637-1649.
- Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359-1363.
- Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989) *Cell*, **56**, 979-986.
- Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) *Cell*, **51**, 975-985.
- Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) *Genes Dev.*, **5**, 1387-1394.
- Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M. and Grunstein, M. (1995) *Cell*, **80**, 583-592.
- Huber, M.C., Bosch, F.X., Sippel, A.E. and Bonifer, C. (1994) *Nucleic Acids Res.*, **22**, 4195-4210.
- Jimenez, G., Griffiths, S.D., Ford, A.M., Greaves, M.F. and Enver, T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10618-10622.
- Kim, C.G., Epner, E.M., Forrester, W.C. and Groudine, M. (1992) *Genes Dev.*, **6**, 928-938.
- Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) *Nucleic Acids Res.*, **15**, 5739-5747.
- Li, Q. and Stamatoyannopoulos, J.A. (1994) *Mol. Cell. Biol.*, **14**, 6087-6096.
- Madisen, L., Groudine, M., Bonifer, C., Yannoutsos, B., Kruger, G., Grosveld, F., Sippel, A.E., Talbot, D., Descombes, P., Schibler, U., Palmiter, R.D., Sandgren, E.P., Koeller, D.M., Brinster, R.L. and Boulikas, T. (1994) *Genes Dev.*, **8**, 2212-2226.
- May, G. and Enver, T. (1995) *EMBO J.*, **14**, 564-574.
- Migeon, B.R. (1994) *Trends Gen.*, **10**, 230-235.
- Palmiter, R.D., Sandgren, E.P., Koeller, D.M. and Brinster, R.L. (1993) *Mol. Cell. Biol.*, **13**, 5266-5275.



- Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2159-2167.
- Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. and Philipsen, S. (1991) *Nucleic Acids Res.*, **19**, 1413-149.
- Reitman, M., Lee, E. and Westphal, H. (1995) *Nucleic Acids Res.*, **23**, 1790-1794
- Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G. (1990) *Nature*, **348**, 749-752.
- Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G. (1993) *Mol. Cell. Biol.*, **13**, 3990-3998.
- Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5371-5375.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Stamatoyannopoulos, G. and Nienhuis, A.W. (1994) In Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P.W. and Varmus, H. (eds.), *Hemoglobin switching*. In *The molecular Basis of Blood Diseases.*, W.B Saunders Co., Philadelphia, pp. 107-136.
- Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T. and Lowrey, C.H. (1995) *EMBO J.*, **14**, 106-116.
- Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D.R. (1989) *Nature*, **338**, 352-355.
- Talbot, D., Descombes, P. and Schibler, U. (1994) *Nucleic Acids Res.*, **22**, 556-566.
- Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2169-177.
- Townes, T.M. and Behringer, R.R. (1990) *Trends Gen.*, **6**, 219-223.
- Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1985) *EMBO J.*, **4**, 1715-1723.
- Trudel, M. and Costantini, F. (1987) *Genes Dev.*, **1**, 954-961.
- Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384-6388.

Zimmerman, K., Legouy, E., Stewart, V., Depinho, R. and Alt, F.W. (1990) *Mol. Cell. Biol.*, **10**, 2096-2103.

Zlatanova, F.S. and Holde, K.E. (1992) *Critical Reviews in Eucaryotic Gene Expression*, **2**, 211-224.

**Table I:** Analysis of  $\beta$ A and  $\beta$ B transgenic lines.

Line	$\beta$ -gal. <sup>1</sup> (mU/mg)	copy # <sup>2</sup>	Expression/copy (mU/mg)/copy
$\beta$ A1	0	2	0
$\beta$ A2	220	3	73
$\beta$ A3	0	~200	0
$\beta$ A4	270	5	54
$\beta$ A5	60	1	60
$\beta$ B1	25	3	8.3
$\beta$ B2	25	19	1.3
$\beta$ B3	590	14	42
$\beta$ B4	720	3	240
$\beta$ B5	610	4	152
$\beta$ B6	900	3	300
$\beta$ B7	8	7	1.1

<sup>1</sup> Extract from fetal liver was assayed for  $\beta$ -galactosidase activity.

<sup>2</sup> Copy number was determined using F1 tail DNA.

**Table II:** Analysis of  $\beta$ C,  $\beta$ D and  $\beta$ E transgenic fetuses.

Line <sup>1</sup>	$\beta$ -gal. (mU/mg)	copy # <sup>2</sup>	Expression/copy (mU/mg)/copy
$\beta$ C1 (I)	1650	20	82
$\beta$ C2 (I)	1020	10	100
$\beta$ C3 (I)	105	16	6.6
$\beta$ D1	2750	50	55
$\beta$ D2	125	3	42
$\beta$ D3	15	5	3.0
$\beta$ D4	30	12	2.5
$\beta$ D5	750	5	150
$\beta$ D6	3	1	3.0
$\beta$ D7	130	3	43
$\beta$ D8	1060	12	88
$\beta$ D9	3900	25	156
$\beta$ D10	300	15	20
$\beta$ D11	300	2	150
$\beta$ D12	2	1	2.0
$\beta$ D13	5	30	0.2
$\beta$ D14	50	20	2.5
$\beta$ D15 (I)	12	20	0.6
$\beta$ D16 (I)	23	30	0.8
$\beta$ D17 (I)	1020	14	73
$\beta$ E1	10	5	2.0
$\beta$ E2	6	4	1.5
$\beta$ E3	9	0.6	15
$\beta$ E4	6	4	1.5
$\beta$ E5	1	2	0.5
$\beta$ E6	120	2	60
$\beta$ E7	8	1	8.0
$\beta$ E8	50	1	50
$\beta$ E9	1	5	0.2
$\beta$ E10	25	0.7	36

<sup>1</sup>(I) depicts bred lines which were analyzed as in Table I. The other mice were analyzed as founder fetuses 15 days following microinjection.

<sup>2</sup>The copy number was determined in fetal liver DNA.

---

**Table III:** Analysis of  $\beta$ F transgenic fetuses<sup>1</sup>.

---

Line	% hu $\beta$ /mo $\beta$ <sup>2</sup>	copy # <sup>3</sup>	Expression/copy
$\beta$ F1	11	0.2	55
$\beta$ F2	37	1	37
$\beta$ F3	50	3	17
$\beta$ F4	75	5	15
$\beta$ F5	180	8	23
$\beta$ F6	328	6	55

---

<sup>1</sup> Fetal livers were isolated 14 days after microinjection

<sup>2</sup> The ratio of human  $\beta$ -globin expression to mouse  $\beta$ -major expression was determined by S1 protection (see Figure 5).

The copy number was determined as in Table II.

<sup>3</sup> Copy number was determined from the fetal liver DNA.

**Table IV:** Analysis of MEL clones transfected with  $\beta$ D construct.

Transfected clone	$\beta$ -gal. (mU/mg)	relative <sup>1</sup> copy #	Expression/copy
$\beta$ D-C1	570	2	290
$\beta$ D-C2	1200	2	600
$\beta$ D-C3	1200	2	600
$\beta$ D-C4	1300	2	650
$\beta$ D-C5	2300	4	575
$\beta$ D-C6	1600	6	270
$\beta$ D-C7	1300	3	430
$\beta$ D-C8	3300	4	830
$\beta$ D-C9	800	2	400
$\beta$ D-C10	1100	2	550
$\beta$ D-C11	510	3	170
$\beta$ D-C12	1000	1	1000
$\beta$ D-C13	970	2	490
$\beta$ D-C14	1300	1	1300
$\beta$ D-C15	780	2	390
$\beta$ D-C16	710	2	355
$\beta$ D-C17	620	2	310
$\beta$ D-C18	630	6	105
$\beta$ D-C19	510	2	260
$\beta$ D-C20	410	1	410
$\beta$ D-C21	270	1	270
$\beta$ D-C22	430	1	430
$\beta$ D-C23	360	2	180

<sup>1</sup> A copy number of 1 was assigned to the clone with the lowest LCR probe intensity relative to the *dt* control probe intensity on Southern blot analysis and the others were assigned by comparison to it.

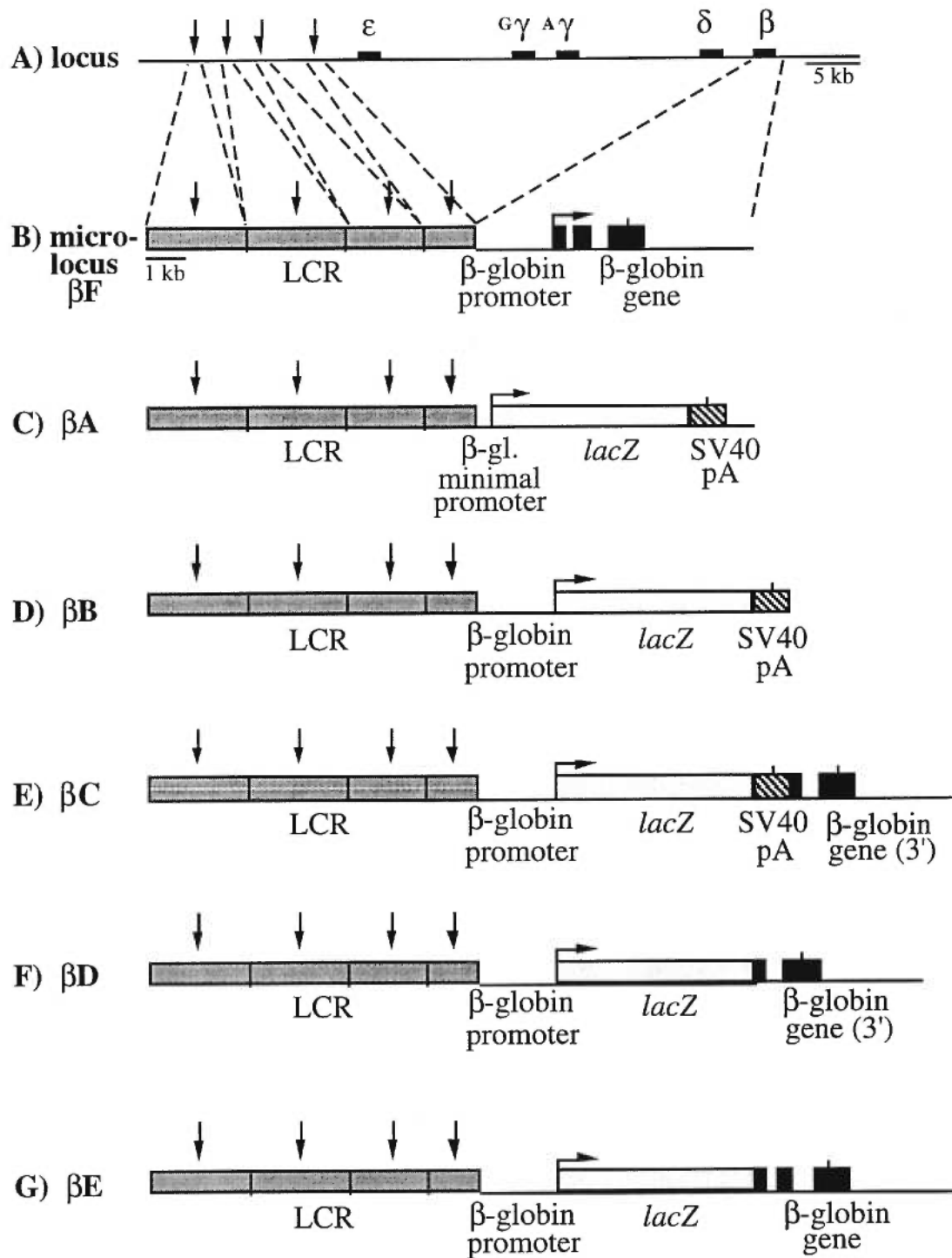
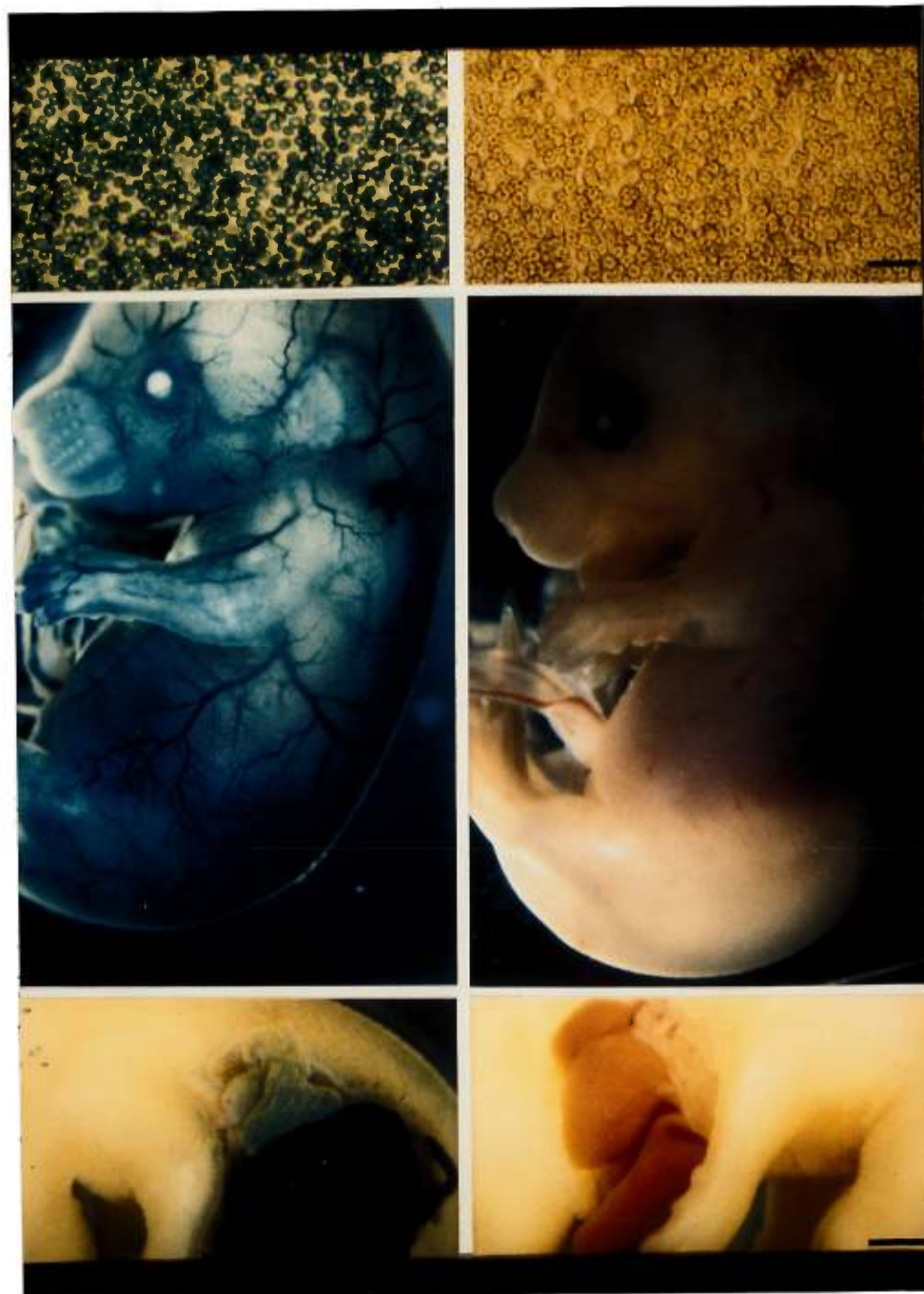
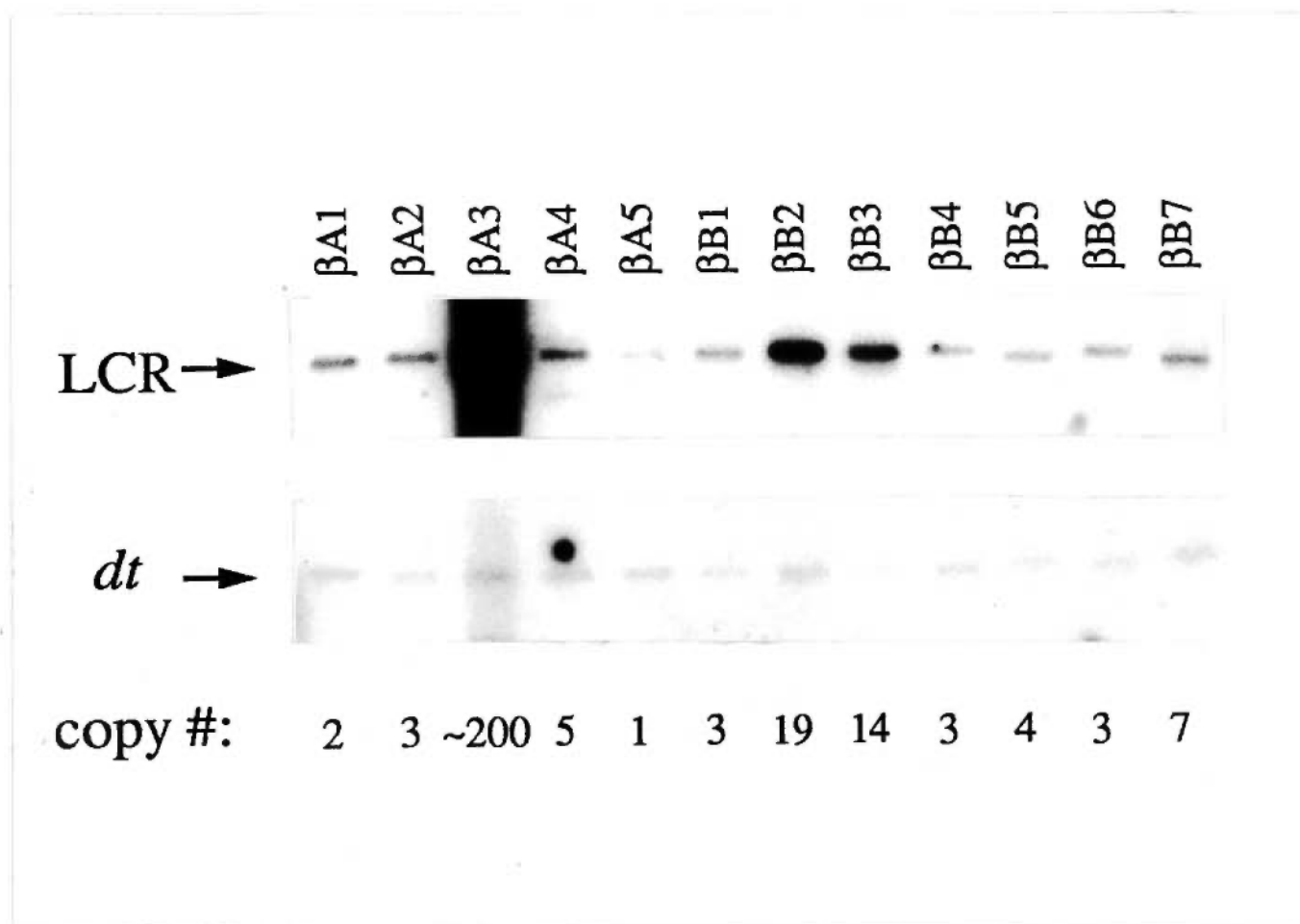


Figure 1: Structure of the  $\beta$ -globin locus and microlocus constructs. (A) The structure of the  $\beta$ -globin locus with the position of HSS regions in the LCR indicated by arrows. (B) the structure of the  $\beta$ -globin microlocus designed by Collis *et al.* (1990) is shown. Black boxes indicate the 3 exons of the  $\beta$ -globin gene. (C-G) The structure of the various  $lacZ$  reporter constructs.

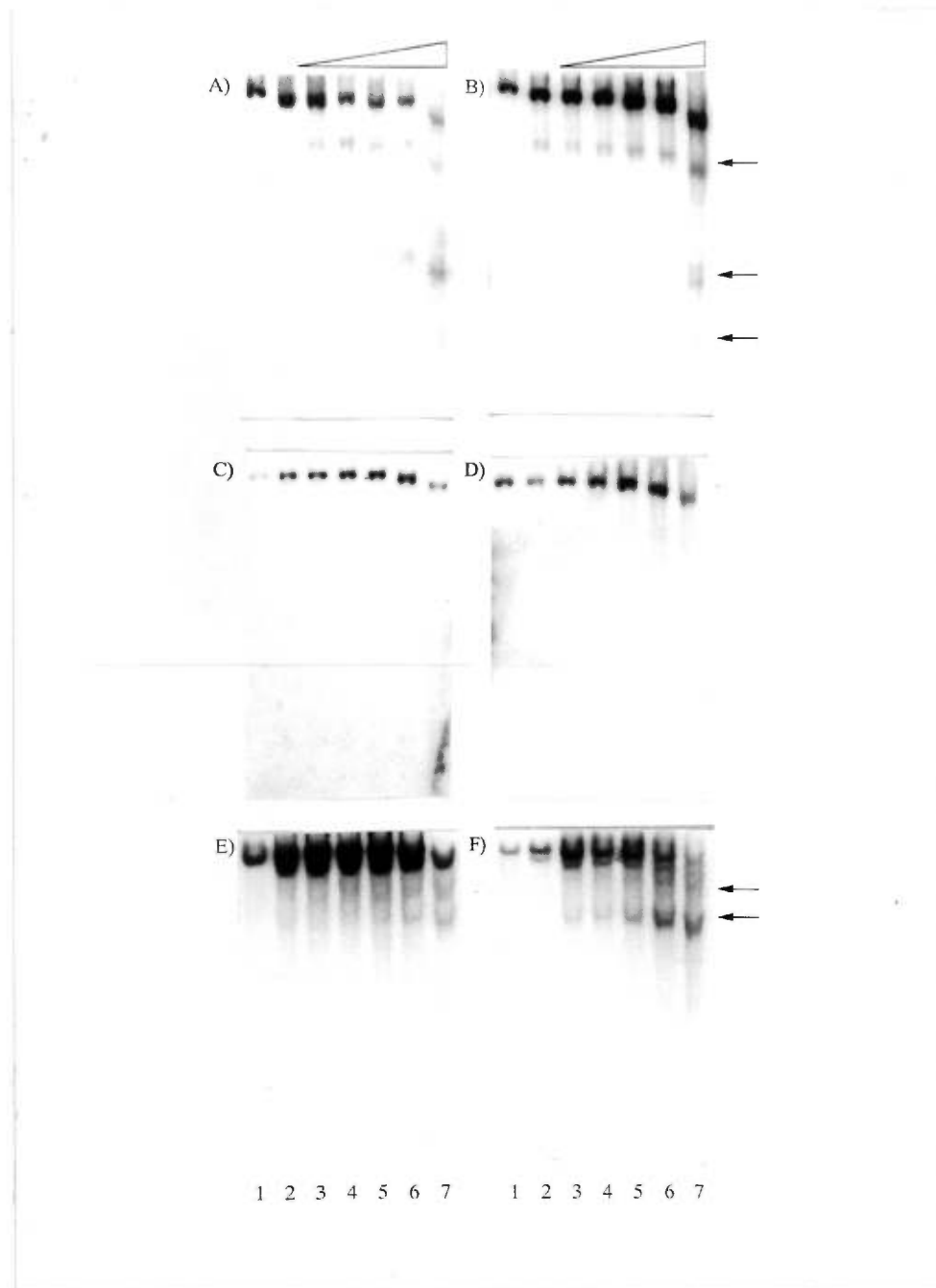


**Figure 2:** *In situ* detection of  $\beta$ -galactosidase activity in adult red blood cells and fetal tissues. The top panel shows the tail blood from adult mice (transgenic line  $\beta$ B6), the middle panel shows 15.5 day fetuses (transgenic line  $\beta$ B1) and the bottom panel shows 15.5 day fetuses (transgenic line  $\beta$ A4) with the liver dissected out. The right of each panel is a non-transgenic litter mate. The scale bars in the top and bottom panel represent 20 and 1000  $\mu$ m, respectively. The fetuses in the middle panel were incubated in X-gal solution overnight to stain the circulatory system, whereas the fetuses in the bottom panel were only incubated for 30 minutes.

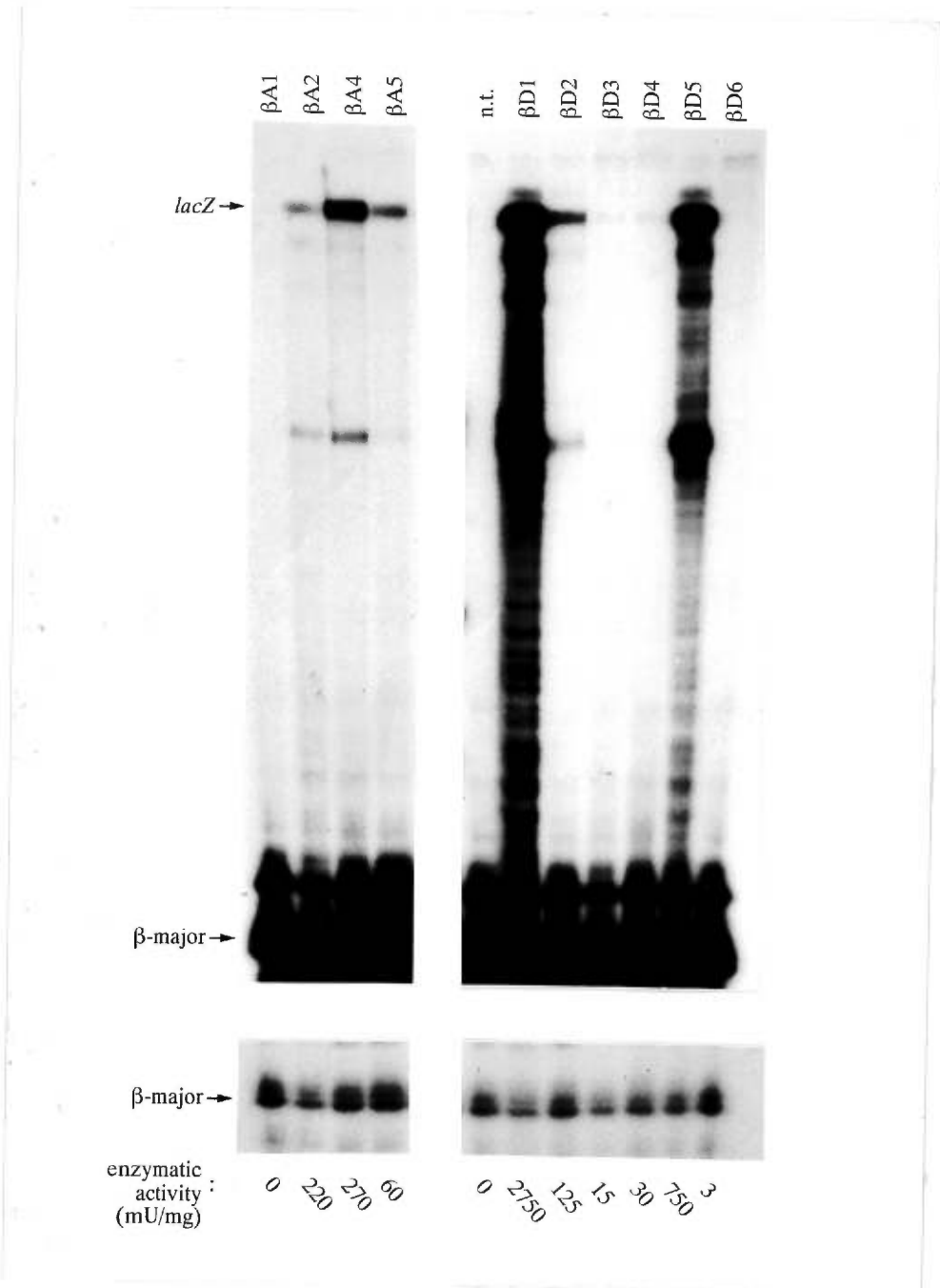




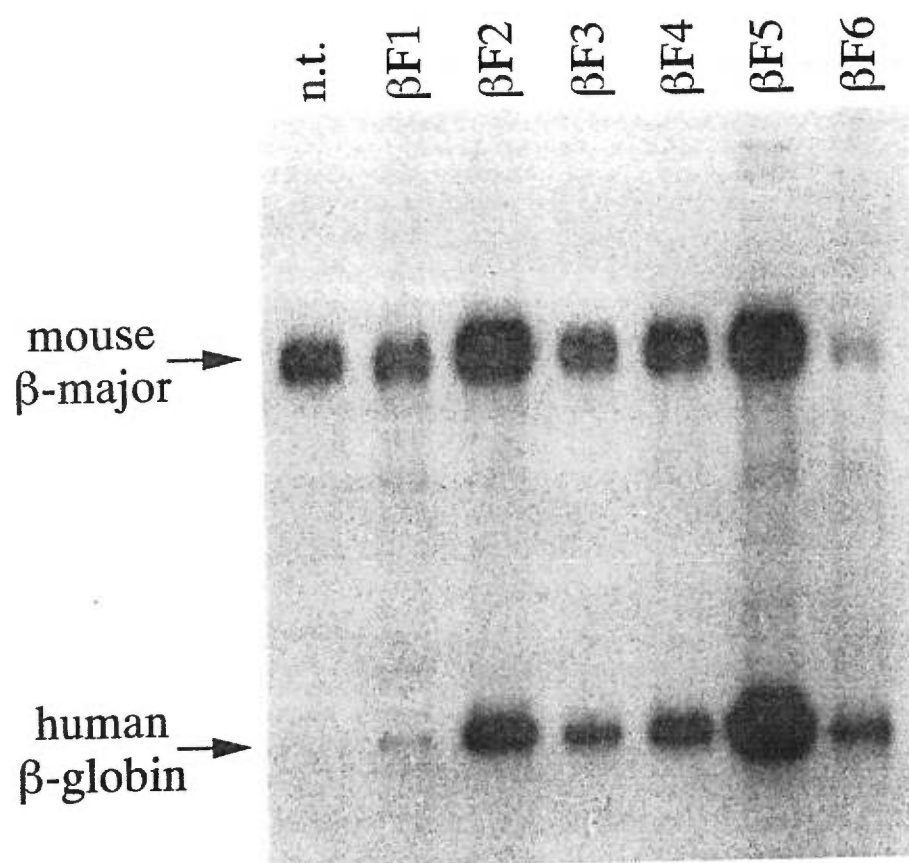
**Figure 3:** Copy number determination. DNA from the transgenic mouse line shown at the top of each lane was digested with *SacI* and then analyzed by Southern blot hybridization. The membrane was hybridized with a DNA fragment from the LCR in the microlocus and with a control probe from the mouse *dt* locus. Copy number was calculated as the ratio of the LCR band intensity over the *dt* band intensity.



**Figure 4:** Hypersensitive site determination. Nuclei isolated from 15.5 day fetal livers were assayed for DNase I hypersensitivity of the transgene LCR or the endogenous mouse  $\beta$ -globin LCR. (A-D) Transgenic lines  $\beta B5$ ,  $\beta B6$ ,  $\beta B1$  and  $\beta B7$ , respectively, were analyzed with a human HSS1 probe from the transgene LCR.  $\beta B1$  and  $\beta B7$  were also analyzed with a mouse LCR probe in (E) and (F), respectively. In lanes 1, DNA was isolated from untreated fetal liver nuclei, while in lanes 2-7 the DNA was isolated from nuclei that were incubated with increasing amounts of DNase I (0, 0.1, 0.4, 1.0, 4.0 and 16  $\mu$ g, respectively). The arrows to the right of the figure indicate the position of DNA fragments that appear on the Southern blot due to the presence of DNase I hypersensitive sites in the respective LCR.



**Figure 5:** RNase protection analysis of *lacZ* mRNA expression. Total RNA from the fetal liver of the  $\beta A$  line or  $\beta D$  founder indicated at the top of each lane (n.t., non-transgenic) was assayed by RNase protection using a *lacZ* and  $\beta$ -major antisense probes. The position of the signals for *lacZ* and  $\beta$ -major are indicated by the arrows to the left of the figure. The top panel shows a 1 week autoradiographic exposure of the complete gel, while in the bottom panel a 5 hour exposure of the  $\beta$ -major signal is also shown for clarity. For comparison, the level of  $\beta$ -galactosidase activity measured in fetal liver extracts of each transgenic line or founder is indicated at the bottom of each lane.



**Figure 6:** S1 nuclease protection analysis of  $\beta$ -globin gene expression in mice carrying the  $\beta$ -globin gene without *lacZ* sequences. Total RNA, isolated from the fetal liver of the founder transgenic mouse indicated at the top of each lane (n.t., non-transgenic) carrying the  $\beta$ F construct, was assayed by S1 protection for human  $\beta$ -globin mRNA and for the endogenous mouse  $\beta$ -major transcript.

## Chapitre 3

### Article 2

Position effects in mice carrying a *lacZ* transgene  
in *cis* with the  $\beta$ -globin LCR can be explained by a graded model

Louis-Georges Guy, Rashmi Kothary et Lee Wall. (1997)

*Nucleic Acids Research*, **21**: 4400-4407.

**Position effects in mice carrying a *lacZ* transgene in *cis* with the  $\beta$ -globin LCR can be explained by a graded model**

Louis-Georges Guy, Rashmi Kothary<sup>1</sup> and Lee Wall<sup>1\*</sup>

From the Institut du cancer de Montréal and the Centre de recherche du Centre hospitalier de l'Université de Montréal, 1560 Sherbrooke Street East, Montreal, Quebec, Canada H2L 4M1 and the <sup>1</sup>Department of Medicine, Université de Montréal.

Running title: Position effects and graded transgene expression

\*Correspondent footnote: Send correspondence to Lee Wall at the Institut du cancer de Montréal, Room Y-4605, 1560 Sherbrooke Street East, Montreal, Quebec, Canada H2L 4M1

Tel: 514-281-6000, ext. 5497

FAX: 514-896-4689

E-mail: Walll@ere.UMontreal.CA

**Abstract:**

We studied transgenic mice carrying the *lacZ* reporter gene linked to the erythroid-specific  $\beta$ -globin promoter and  $\beta$ -globin locus control region (LCR). Previously, we had demonstrated that the total level of expression of  $\beta$ -galactosidase enzyme, which is the product of the *lacZ* gene, varies widely between different transgenic mice due to position effects at the sites of transgene integration. Here, using the X-gal based *in situ* assay for  $\beta$ -galactosidase activity, we found that the percent erythroid cells that expressed the transgene also varied widely between the mice. Moreover, a kinetic analysis showed that the average  $\beta$ -galactosidase content per expressing cell varied both between samples of different transgenic decent and between erythroid cells within each sample; demonstrating that the variable expression of this *lacZ* transgene was being controlled in a graded manner. These results suggest that the  $\beta$ -globin LCR enhancers function through a graded model, which is described, rather than the binary mechanism that has been proposed previously for other enhancers.

## **Introduction:**

When transgenes integrate into the genome of mice or other organisms, they are generally prone to a phenomenon called position effects or position-dependent expression (reviewed in 1,2). The levels of expression show a wide variation between different sites of integration in the genome with no relationship to transgene copy number. These position effects are thought to be the result of the influence of neighboring chromatin structure on the functioning of enhancers controlling expression of the transgene. Thus, by studying position-effects we can gain insights into the mechanism of up-regulation of gene transcription by enhancers.

Originally, it was proposed that enhancers increase the transcription initiation rate of a gene by increasing the density of RNA polymerase molecules transcribing it (3,4). However, several studies have disputed this rate model. For example, using an assay for *lacZ* transgene activity that detects expression in individual cells, Walters *et al.* (5) found that enhancers increase the proportion of expressing cells, but not the level of expression in individual cells. To explain this result, they proposed the binary model, which suggests that the level of expression is always the same in each transcribing cell, but only some cells are able to form an active transcription complex. By affecting this binary (on/off) switch, enhancers increase the percentage of cells that form an active transcription complex. A similar hypothesis was originally put forth by Weintraub (6) and the possibility was also discussed earlier by Moreau *et al.* (7). More recently, Walters *et al.* showed that enhancers prevented transgene repression that occurred upon removal of drug selection from stably transfected cell clones (8). They therefore suggested that enhancers increase the number of cells able to undergo active transcription by preventing gene repression in individual cells.

It is well known that integration of a transgene near telomeric or centromeric regions in yeast or *Drosophila* (reviewed in 9,10), as well as in mice (reviewed in 11), can result in heterocellular expression of a transgene such that it is expressed in some cells but not in others. This phenomenon is called position effect variegation. In looking at expression in



individual red blood cells of transgenic mice carrying a *lacZ* reporter gene under the control of  $\alpha$ -globin regulatory sequences, Robertson *et al.* (12) found that position effects resulted in a variable number of expressing cells between different transgenic mouse lines, but did not appear to greatly affect the level of expression per cell. Moreover, the addition of an enhancer was found to increase the number of expressing cells with no apparent effect on the level of transcription in individual cells (13). These observations support the notion that position effects and enhancer function can be controlled by a binary model in some cases. However, the binary model alone may not fit all situations. For example, position effects on expression of the *white* gene in *Drosophila* sometimes result in variegation that can be seen as a mosaic (or variegated) eye color pattern. This is caused by the fact that some cells do not express the transgene at all while all other cells express at wild type levels, as the binary model would predict. In contrast, many position effects on the *white* gene result in a variety of eye colors, ranging from light yellow, to orange, brown and dark red, without apparent cell to cell variegation (14). Similarly, even when a *ade6* transgene is integrated near a centromeric region in yeast, cells of a variable color between white and red arise, suggesting that there can be a gradient of transgene expression between individual cells (15). Thus, an alternative mechanism to the binary model may function in some cases.

Locus control regions (LCRs) are DNA elements that can overcome position effects (reviewed in 16). LCRs have been isolated from several gene loci and they are often associated with enhancers. However, the ability to enhance gene expression and the ability to confer position-independent expression onto transgenes appear to be independent and separable functions of LCRs.

To date, the best characterized LCR is that of the human  $\beta$ -globin gene locus (reviewed in 17,18). The  $\beta$ -globin locus contains five active genes ( $\epsilon$ ,  $G\gamma$ ,  $A\gamma$ ,  $\delta$ , and  $\beta$ ) that are arranged in their developmental order of expression and all five genes rely on an LCR for high levels of expression. This  $\beta$ -globin LCR is located ~15 and 60 kb upstream of the  $\epsilon$ - and  $\beta$ -globin genes, respectively, and it is signified by five regions of DNase I hypersensitive sites (HS). Three of these HS regions have enhancer activities that can act

independently of one another to give partial levels of expression (19 and references therein). How the full  $\beta$ -globin LCR protects against position effects is unresolved. It may require at least three of the HS regions (20), and may also involve specific sequences flanking the globin genes (21-23).

If the enhancers of the LCR act via the binary mechanism as proposed (5), one might predict the LCRs position-independence function is to prevent repression in all cells so that they all express at the same level. However, it has been shown that the genes of the  $\beta$ -globin locus compete for activation by the LCR during development (24-26), such that when a  $\gamma$ - and  $\beta$ -globin gene are together with the LCR in transgenic mice, the  $\gamma$ -globin gene is only expressed during embryonic and early fetal life, while the  $\beta$ -globin gene is expressed only in late fetal and adult life. This is so, even though the embryonic cells are perfectly competent to express the  $\beta$ -globin gene, since if the  $\beta$ -globin gene is alone with the LCR it is expressed at all stages of development in mice (24-26). Moreover, Wijgerde *et al.* showed that during the switching process the cells flip-flop between transcribing the  $\gamma$ - and  $\beta$ -globin gene from the same chromosome (27). In addition, some transgenic mice carrying a position-dependent  $\gamma$ -globin-LCR construct gave a heterogeneous cell to cell intensity of staining with  $\gamma$ -globin antibodies (28,29), and erythroid colonies isolated from patients with heterocellular hereditary persistence of fetal hemoglobin, a human disorder in which there is an increase in  $\gamma$ -globin producing cells in adults, show a heterogeneity in the amount of  $\gamma$ -globin production (30). Taken together, these results show that activation of the LCR enhancers does not necessarily mean that all cells will express a given gene at the same level. Therefore, although the LCR action may involve a binary mechanism, this model alone fails to explain all observations.

To understand how the  $\beta$ -globin LCR brings about position-independent expression, we need to know how the LCR enhancers will act on a globin gene promoter under a situation where the ability of the LCR to protect against position effects has been crippled. Recently, we showed that a *lacZ* transgene linked to the  $\beta$ -globin gene promoter in *cis* with the LCR is highly susceptible to position effects in mice (21). That this same transgene gave full levels of expression in individual drug-selected, transfected cell culture

clones suggested that the LCR enhancers were functioning, but the LCRs ability to prevent position effects does not work with such a transgene in mice. Here we look at *lacZ* transgene expression in individual blood cells of >40 different transgenic mice. We show that expression on a cell to cell basis is graded both within a single transgenic sample and between different samples. Thus, the  $\beta$ -globin LCR enhancers function through a graded mechanism that is described.

## **Materials and Methods:**

### **Transgenic mice:**

The generation of transgenic mice used in this study has been described (21). Briefly, the  $\beta$ A transgene contains the *lacZ* coding sequences, followed by Simian Virus 40 (SV40) polyadenylation signals, linked to a minimal  $\beta$ -globin promoter (-87 to +32 relative to the Cap site) in *cis* with the LCR from the  $\beta$ -globin  $\mu$ -locus plasmid (31). In the  $\beta$ B,  $\beta$ C,  $\beta$ D and  $\beta$ E transgenes the  $\beta$ -globin promoter extends from -800 to +32.  $\beta$ B is otherwise identical to  $\beta$ A. In  $\beta$ C and  $\beta$ D, sequences starting from the *Bam*HI site in the second intron to 1.7 kb downstream of the polyadenylation signal of the  $\beta$ -globin gene were cloned 3' to the *lacZ* sequences in  $\beta$ B. For  $\beta$ C the SV40 polyadenylation signals were left intact so that the further  $\beta$ -globin gene sequences are not transcribed, while the SV40 polyadenylation signals were removed in the case of the  $\beta$ D construct. The  $\beta$ E transgene contains the whole  $\beta$ -globin gene, from -800 to 1.7 kb past the polyadenylation signal, with the *lacZ* sequences (without the SV40 polyadenylation signals) cloned directly into the *Nco*I site situated at the  $\beta$ -globin protein initiation codon. As indicated in Table 1, transgenic animals were either studied as founders isolated 15 days after microinjection or stable lines were produced and 15.5 day old fetuses of third and further generations were studied.

### **$\beta$ -galactosidase assays:**

For *in situ* assays with the X-gal (5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactopyranoside) substrate, one quarter of a fetal liver was dispersed into 100  $\mu$ l of 0.25 M Tris.HCl (pH 8.0). Ten  $\mu$ l of the cell suspension was deposited into a well of a 96 well tissue culture plate, and then the cells were fixed by incubating in 100  $\mu$ l of phosphate buffered saline (PBS), containing 0.2% glutaraldehyde, for 10 min. The cells were rinsed with PBS and incubated in the dark at 37°C in PBS containing 0.1 M X-gal, 0.01 M potassium ferrocyanide, 0.01 M potassium ferricyanide and 1 mM MgCl<sub>2</sub>, for either 24 hours, after which time no further staining was observed, or for the times

indicated in the figures. Measurements of the percent blue stained cells and photographs were done under phase contrast optical microscopy. The method to measure total  $\beta$ -galactosidase activity was as described previously (21).

#### Densitometry.

The density of blue coloration of individual cells developed with the X-gal *in situ* assay (see Fig. 2b) was determined using an IS 1000  $\alpha$ -imager (Canberra Packard Inc.). For each photograph, which represented a different time point (see Fig. 1), the density of 90-100 randomly chosen cells was determined and from this the background determined from areas of the photograph without cells was subtracted. The most intensely colored cells from all photographs were arbitrarily given a value of 10, while the least intense (unstained) cells were given a value of 0. The intensity of other cells was scored between 0 and 10, where 0 means the cell had a relative intensity of <1, 1 means the cell had a relative intensity of between 1 and 2, 2 means the cell had a relative intensity between 2 and 3, and so on.

#### Methylation assays.

DNA was assayed for methylation as described previously (32). Details are given in the legend to Figure 4. Note the different lanes in Fig. 4 were all from the same Southern blot which was exposed to film for different times to even out the intensity of the individual lanes.

**Results:**

$\beta$ -Galactosidase expression is graded in transgenic mice carrying the *lacZ* gene under the control of the  $\beta$ -globin LCR.

Although the  $\beta$ -globin LCR directs position-independent expression onto globin transgenes in mice (33), transgenes containing the  $\beta$ -globin LCR and the bacterial *lacZ* gene linked to the  $\beta$ -globin promoter are prone to strong position effects in mice (21). Among forty two transgenic samples studied, the level of *lacZ* transgene expression, as measured both at the mRNA and  $\beta$ -galactosidase activity level, varied >700 fold on a per copy basis between those mice that had detectable expression, and two samples gave no expression at all. Fetal liver cells from 41 of these transgenic animals, which included fifteen bred lines, were dispersed into solution and were stained for  $\beta$ -galactosidase activity using the X-gal reagent (34). The percent cells that colored blue after a 24 hour incubation, after which time no further staining was observed, varied widely among the different transgenic animals (Table 1). A low of ~1%, to a high of >85% was obtained, excluding the two samples that had no detectable expression at all. Greater than 80% apparently  $\beta$ -galactosidase positive cells likely represented all of the red blood cells, since fetal liver also contains hepatic cells which would not be expected to express the transgene as it is erythroid-specific (35).

The variation in the fraction of staining cells in the different transgenic animals was not due to a lack of precision in the assay since fetal livers derived from animals of the same transgenic line gave very little difference in the percent staining cells (not shown). It is possible that some of the variation between samples from founders was due to the presence of mosaic animals. If a founder was mosaic, some of the red blood cells may not have contained the transgene. However, a >10-fold variation in the fraction of staining cells was also observed between the different bred lines (Table 1). Since the transgene must have passed through the germ line for it to be transferred to offspring, all red blood cells in these lines must have contained the transgene. Therefore, a kinetic

analysis of the *in situ*  $\beta$ -galactosidase assay was done to gain an appreciation of the mechanism of this heterocellular staining-position effect phenomenon.

Transgenic fetal liver samples derived from bred lines that showed either a relatively low ( $\beta$ D15, 11%), medium ( $\beta$ A5, 28%), or high ( $\beta$ B5, 82%) proportion of staining cells were viewed at different times of incubation with X-gal (Fig. 1). Note that all the other lines and founders were studied similarly and the results presented and discussed below were typical.

The binary model predicts that when comparing cells within a single transgenic sample the expressing cells should have approximately the same content of  $\beta$ -galactosidase enzyme. Thus, one would expect there would be little variation in the staining intensity of individual cells in a given sample. Indeed, this was seen to be the case for a *lacZ* transgene under the control of  $\alpha$ -globin regulatory sequences in transgenic mice (13). However, in all of our samples the number of cells that could be discerned to contain blue coloration increased with the time of incubation with X-gal (Figs. 1 and 2a), suggesting that cells within a single sample did not contain the same amounts of  $\beta$ -galactosidase protein. This conclusion was supported by the observation that there was also a large variation in the intensity of staining of individual cells at any given time. This could be seen visually (see arrows in Fig. 1 for example, but note that the variation in cell intensity can be easily seen in all samples in Fig. 1 and was seen with every transgenic sample) and by determining the staining intensity of individual cells using densitometry (Fig. 2b). Note also that cells could be seen to increase their intensity of staining with the time of incubation (see arrow heads in Fig. 1), as would be expected if the intensity of blue coloration versus time reflected the level of  $\beta$ -galactosidase enzyme in individual cells. Thus, cells within each transgenic sample were not expressing at a set level as the binary model would suggest and as has been seen in other studies (13). Rather, there was a gradient in  $\beta$ -galactosidase activity.

To ensure the cell to cell variation in staining intensity within a single transgenic sample was not due to an incubation phenomena, such as diffusion of  $\beta$ -galactosidase

enzyme or reaction product from a positive cell to a negative cell, we mixed, in varying proportions, fetal liver cells from a transgenic animal that showed a high percent of positive cells with fetal liver cells from a non-transgenic animal. In all cases, the percent cells that could be detected as  $\beta$ -galactosidase positive was predictable based on the proportion of transgenic sample in the mix (results not shown). It also seemed possible that the cell to cell variation could at least partially reflect the changing differentiation status of developing erythroid cells in fetal liver. However, when fetal blood cells, all of which should be mature cells, were stained and looked at at different times they showed a similar gradient in staining intensity (results not shown). Hence, the differentiation status was playing a minor role if any. We have also previously shown that the level of  $\beta$ -galactosidase activity correlates with the level of *lacZ* mRNA (21). Thus, the results taken together suggested that the variation in staining intensity reflected different levels of transgene expression in individual red blood cells.

The binary model also suggests that the level of expression per positive cell should be the same between different transgenic samples as, again, was observed to be the case for the *lacZ* gene in *cis* with the  $\alpha$ -globin regulatory sequences (13). However, this was not the case for our transgenic samples. For example, in a sample that gave a low percent of detectable cells (top row in Fig. 1), stained cells could only barely be detected after 2 h of incubation. In fact, even after 24 h of incubation, none of the cells had attained an intense blue color. Rather, they were just detectable. In contrast, a transgenic sample with a high fraction of detectable cells (bottom row in Fig. 1) gave intense staining cells even after only 2 h of incubation, and many stained cells were already easily discernible after only 45 min of incubation. A transgenic sample that gave a medium proportion of apparent  $\beta$ -galactosidase positive cells (middle row in Fig. 1a; compare to top and bottom rows) gave a profile of staining, in terms of the time blue cells first appeared and the average intensity of staining at any given time, that was between that for the low and high expressing samples. These results suggest that the average amount of  $\beta$ -galactosidase per expressing cell was varying considerably between the different lines. In fact, it appeared that samples that showed a low fraction of positive cells had less  $\beta$ -galactosidase enzyme



per cell, on average, than samples that had a higher proportion of positive cells. To test these possibilities more rigorously, we calculated the amount of  $\beta$ -galactosidase activity per expressing cell for each sample and plotted the values against the total expression measured in extracts. With the binary model, since expression per positive cell should be constant between the different samples, one would predict the curve shown in Figure 3a. However, as can be seen in Figure 3b, the average level of expression per positive cell we observed was not constant. Rather it increased proportionally to the total expression ( $R= 0.981$ ). Thus, not only was there a grading of expression between cells within a given sample, there was also a grading of expression when comparing cells between different samples. It is also interesting to note that the maximum level of  $\beta$ -galactosidase enzyme that cells could possibly produce was not reached since the curve in Figure 3b did not plateau at the higher values of expression per cell. Another manner of looking at the results is shown in Figure 3c. As the average level of expression per positive cell increases so does the number of cells that can be detected as positive. These two values are approximately proportional at the lower part of the curve and then plateau at >80% positive cells, which likely represents the total red cell population. This agrees with the finding that samples with a low fraction of positive cells give rise to less intense stained cells than samples with a higher fraction of positive cells (Fig. 1). The data in Figure 3c clearly indicate that it is the average level of expression per cell in the population of cells that determines what fraction of cells can be detected as positive.

In presenting the binary model the effect of a variation in transgene copy number was not discussed (5). In any case, we have shown previously that expression of the *lacZ* gene in the transgenic samples studied here shows absolutely no relationship to copy number in terms of total  $\beta$ -galactosidase activity or mRNA levels in fetal liver (21). Moreover, there is also no relationship to transgene copy number when expression, in terms of percent cells stained or expression per cell, is looked at on an individual cell basis (Table 1). For example, a single copy transgenic can give more intensely stained cells (i.e. higher expression per cell) than a multi-copy transgenic (compare the 1 copy  $\beta$ A5 line

to the 20 copy  $\beta$ D15 line in Fig.1). Thus, transgene copy number does not have any bearing on the interpretation of the data presented here.

Gene silencing, as measured by methylation of the transgene DNA, was absent in samples that had detectable  $\beta$ -galactosidase activity.

In vertebrates, DNA methylation at CpG dinucleotides is associated with gene silencing, although whether this is a direct cause of gene inactivity is not resolved (reviewed in 36,37). Since in many of our transgenic samples a large proportion of cells appeared to contain no detectable  $\beta$ -galactosidase activity, we wondered if this might have been due to direct silencing of the transgene in a proportion of cells. Therefore, we looked at the methylation status of the *lacZ* gene in fetal liver DNA by differential digestion with methylation insensitive (*MspI*) and methylation sensitive (*HpaII*) restriction enzymes. As can be seen in Figure 4, in transgenic lines with a relatively low ( $\beta$ A5, 28% and  $\beta$ B1, 38%) or a relatively high ( $\beta$ B5, 82% and  $\beta$ B6, 84%) fraction of staining cells, the *lacZ* transgene was not at all methylated. It was cut equally well and to near completion by both enzymes. On the other hand, in the two samples ( $\beta$ A1 and  $\beta$ A3) that gave no detectable  $\beta$ -galactosidase expression at all (no detectable activity in extracts and no detectable positive cells), the *lacZ* transgene was completely methylated, as it was only digested by the methylation insensitive enzyme (Fig. 4, compare lane 2 to lane 3 and lane 5 to lane 6). Thus, in transgenic samples in which absolutely no *lacZ* transgene activity could be detected, it appeared that this involved silencing of the transgene in the total cell population. On the other hand, when the *lacZ* transgene was expressed in only a fraction of the cells, this was not a result of the *lacZ* gene being preferentially silenced in the cells that did not express the transgene. It is also worthy to note that Garrick *et al.* (38) also found that a *lacZ* transgene remained unmethylated in mice whether it was expressed at a relatively low or high level. However, these authors did not study lines in which the transgene was totally silenced, which we show here is associated with methylation.

## **Discussion:**

In the results presented here, it was found that a *lacZ* transgene in *cis* with the  $\beta$ -globin LCR is expressed in a graded fashion on a cell to cell basis, both between cells within a single transgenic sample (Figs. 1 and 2) and between different transgenic samples (Fig. 3b). As far as we are aware, this is the first demonstration of definite cell to cell gradation of transgene expression in mice, although it has been documented in yeast studies previously (15). The graded cell to cell variation of expression within a single transgenic sample would not appear to be a position effect phenomenon *per se*, as it was observed in all 39 expressing transgenic samples studied. Therefore, this gradation occurs independent of the integration site. Nor does this gradient in expression appear to be something that is inherent to the *lacZ* transgene, since it did not occur in mice carrying the *lacZ* transgene controlled by  $\alpha$ -globin regulatory sequences (13). Thus, we must conclude that this gradation in expression between cells is something that is inherent to the regulatory regions directing expression of the *lacZ* transgene in our mice; namely the LCR enhancers in conjunction with the  $\beta$ -globin promoter.

The position effects in our mice provoked a variable level of total expression in the population of red blood cells from one transgenic line to the next, with no relationship to copy number (21), and here we observed that in many transgenic samples, including bred lines, only a fraction of the red blood cells contained  $\beta$ -galactosidase activity. If this was primarily due to a binary mechanism controlling the position effects, we would have expected that the level of expression per expressing cell would be constant and that the total level of expression in the population would simply be determined by the fraction of expressing cells. However, the average level of expression per positive cell was not constant between the different transgenic samples, but increased proportionally to the total level of expression in the population (Fig. 3b). Thus, the position effect, in this case, determined the average level of expression per cell in the population. We refer to this as a graded model for position effects, as it describes both the gradation in the average level of

expression per cell between different transgene integration sites and the cell to cell gradation in individual transgenic samples.

Since position effects are presumed to reflect the level of functioning of enhancers controlling the transgene, our results argue that the LCR enhancers do not work by a simple binary mechanism, for example, acting to prevent silencing of the gene in individual cells. Rather the LCR enhancers act to increase the average level of expression in individual cells. This conclusion contrasts with those of others (5,13) whose results have indicated that enhancers work via a binary mechanism. However, regulatory sequences different from those being studied here were investigated by these researchers. Thus, we suggest that the  $\beta$ -globin LCR enhancers work by the graded mechanism as we have described, whereas some enhancers may work via the binary model.

If the LCR enhancers act to increase expression in individual cells, why then do we observe heterocellular staining in some transgenic samples? There are at least two possible explanations. One would be that in a certain fraction of cells the transgene is silenced as the binary model would predict (5). However, we do not prefer this explanation, as we found that there is no methylation of the *lacZ* transgene DNA in samples that had detectable expression in only a proportion of the cells (Fig. 4). Moreover, in studies of yeast it was found that isolated non-expressing cells can revert back to expressing cells after several generations (15), suggesting that the transgene in non-expressing cells is not irreversibly silenced in such mixed populations. Still, we cannot rule out the possibility of gene silencing in a sub-population of the non-expressing cells in our case, and it should be realized that methylation of DNA may represent only the final stage of gene silencing. On the other hand, the heterocellular staining we observe in some samples could be related to the graded model. When a cell expresses the product of a gene, it must initiate transcription, elongate the transcript, process the mRNA precursor, transport the mRNA from the nucleus to the cytoplasm and translate the mRNA into protein. Moreover, enough protein must accumulate within the cell to form enough active enzyme in the case where gene expression is being measured at the enzyme activity level. The steady state level of mRNA and protein within a cell is a net result of synthesis and degradation at each

of the steps. The graded model for position effects in combination with the random distribution of expression within individual cells in the population, would predict that a transgene strongly repressed by its position in the genome can fall below the level of actually making active protein or even initiating transcription in many cells (we use the term repression to mean a more dynamic state of gene inactivity, while the term silencing is meant as a more irreversible state of gene inactivity). The more strongly the transgene is repressed by its genomic position the higher the likelihood that a large fraction of cells will fall below the detection limit of the assay being employed to measure expression and this is because there is a cell to cell variation in expression levels (refer to Fig. 5). On the other hand, when the transgene is expressed at higher levels because of a more favorable genomic position, more, and possibly all cells will be above the detection limits of the assay (Fig. 5). Therefore, the graded model predicts that as the average level of expression increases in the population of cells so does the fraction of cells that express the transgene. This will plateau when all the cells can be detected as positive, as was seen to be the case here (Fig. 3c). Hence, the graded model explains why a fraction of cells do not express the transgene in samples with a low overall activity without invoking gene silencing.

At this point we can only speculate as to how the graded mechanism functions at the molecular level and how the  $\beta$ -globin LCR in conjunction with globin genes acts to prevent position effects. It has been suggested that the total level a globin gene is expressed in a cell is determined by the total time that gene is transcribed which, in turn, is dictated by the time it spends interacting with the LCR. If we can relate position effects to the amount of time a gene is transcribed and find a way to make it so that the  $\beta$ -globin LCR can confer position-independent expression onto the *lacZ* gene such that we can look at expression in individual cells, we should be able to address the mechanisms involved more directly in the future.

**Acknowledgments:**

This work was supported by grants from the Medical Research Council of Canada to L.W. and R.K. and from the Fondation de l'Hôpital Notre-Dame to L.W.. Both L.W. and R.K. are Scholars of the Fonds de la Recherche en Santé du Québec.

**References:**

1. Palmiter, R. D. and Brinster, R. L. (1986) *Annu. Rev. Genet.*, **20**, 465-499.
2. Dillon, N. and Grosveld, F. (1994) *Curr. Biol.*, **4**, 260-264.
3. Weber, F. and Schaffner, W. (1985) *Nature*, **315**, 75-77.
4. Treisman, R. and Maniatis, T. (1985) *Nature*, **315**, 73-75.
5. Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. and Martin, D. I. K. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7125-7129.
6. Weintraub, H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5819-5823.
7. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P. and Chambon, P. (1981) *Nucleic Acids Res.*, **9**, 6047-6068.
8. Walters, M. C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. and Martin, D. I. K. (1996) *Genes Dev.*, **10**, 185-193.
9. Tartof, K. D. (1994) *BioEssays*, **16**, 713-714.
10. Henikoff, S. (1992) *Curr. Opin. Genet. Dev.*, **2**, 907-912.
11. Dobie, K., Mehtali, M., McClenaghan, M. and Lathe, R. (1997) *Trends Genet.*, **13**, 127-130.
12. Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5371-5375.
13. Sutherland, H. G. E., Martin, D. I. K. and Whitelaw, E. (1997) *Mol. Cell. Biol.*, **17**, 1607-1614.
14. Levis, R., Hazelrigg, T. and Rubin, G. M. (1985) *EMBO J.*, **4**, 3489-3499.
15. Allshire, R. C., Javerzat, J.-P., Redhead, N. J. and Cranston, G. (1994) *Cell*, **76**, 157-169.
16. Felsenfeld, G. (1992) *Nature*, **355**, 219-224.

17. Epner, E., Kim, C. G. and Groudine, M. (1992) *Curr. Biol.*, **2**, 262-264.
18. Townes, T. M. and Behringer, R. R. (1990) *Trends Genet.*, **6**, 219-223.
19. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) *Genes Dev.*, **7**, 106-113.
20. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P. (1996) *Cell*, **87**, 105-114.
21. Guy, L.-G., Kothary, R., DeRepentigny, Y., Delvoye, N., Ellis, J. and Wall, L. (1996) *EMBO J.*, **15**, 3713-3721.
22. Li, Q. and Stamatoyannopoulos, J. A. (1994) *Mol. Cell. Biol.*, **14**, 6087-6096.
23. Stamatoyannopoulos, G., Josephson, B., Zhang, J. W. and Li, Q. (1993) *Mol. Cell. Biol.*, **13**, 7636-7644.
24. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. and Townes, T. M. (1990) *Genes Dev.*, **4**, 380-389.
25. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) *Genes Dev.*, **5**, 1387-1394.
26. Peterson, K. R. and Stamatoyannopoulos, G. (1993) *Mol. Cell. Biol.*, **13**, 4836-4843.
27. Wijgerde, M., Grosveld, F. and Fraser, P. (1995) *Nature*, **377**, 209-213.
28. Morley, B. J., Abbott, C. A., Sharpe, J. A., Lida, J., Chan-Thomas, P. S. and Wood, W. G. (1992) *Mol. Cell. Biol.*, **12**, 2057-2066.
29. Enver, T., Ebens, A. J., Forrester, W. C. and Stamatoyannopoulos, G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7033-7037.
30. Dubart, A., Testa, U., Mulumeci, S., Vainchenker, W., Beuzard, Y., Henri, A., Schirilo, G., Romeo, M. A., Russo, G., Rochant, H. and Rosa, J. (1980) *Scandinavian J. Haematol.*, **25**, 339-346.
31. Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.*, **9**, 233-240.



32. Waalwijk, C. and Flavell, R. A. (1978) *Nucleic Acids Res.*, **5**, 4631-4644.
33. Grosveld, F., Blom van Assendelft, M., Greaves, D. and Kollias, G. (1987) *Cell*, **51**, 21-31.
34. Vernet, M., Bonnerot, C., Briand, P. and Nicolas, J.-F. (1993) *Methods Enzymol.*, **225**, 434-451.
35. Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D. R. and Grosveld, F. (1989) *Genes Dev.*, **3**, 1572-1581.
36. Chomet, P. S. (1991) *Curr. Opin. Cell Biol.*, **34**, 438-443.
37. Meehan, R., Lewis, J., Gross, S., Nan, X., Jeppesen, P. and Bird, A. (1992) *J. Cell Sci.*, **16**, 9-14.
38. Garrick, D., Sutherland, H., Robertson, G. and Whitelaw, E. (1996) *Nucleic Acids Res.*, **24**, 4902-4909.

Table 1: Analysis of  $\beta$ -galactosidase expression and X-gal staining in transgenic mice

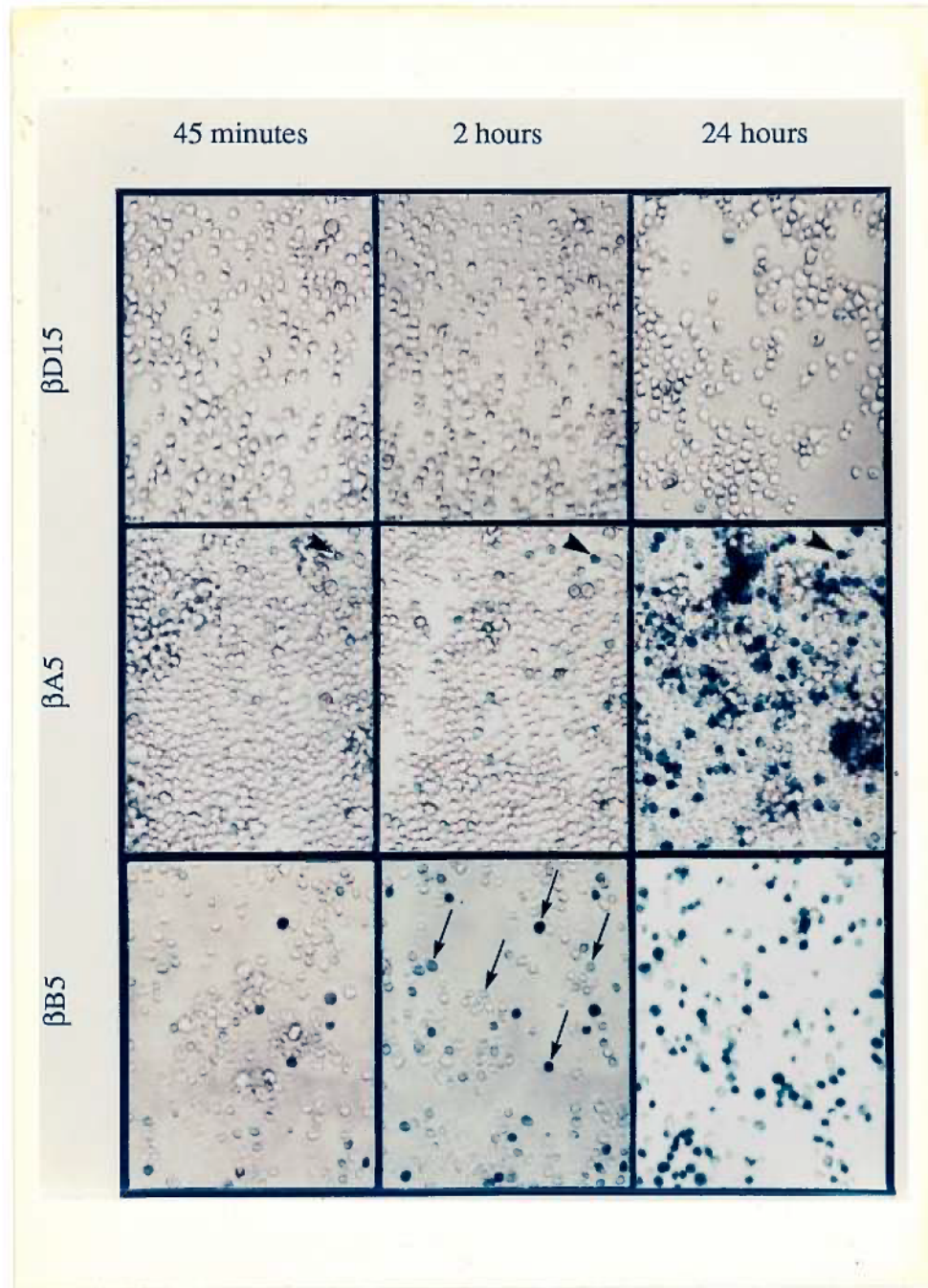
Line	<sup>1</sup> copy number <sup>2</sup>	% staining cells <sup>3</sup>	Expression <sup>4</sup> (mU/mg)	Exp./staining cell (mU/mg)/% cell
$\beta$ A1	(l) 2	0	0	-
$\beta$ A2	(l) 3	46	220	4.8
$\beta$ A3	(l) ~200	0	0	-
$\beta$ A4	(l) 5	58	270	4.7
$\beta$ A5	(l) 1	28	60	2.1
$\beta$ B1	(l) 3	38	25	0.66
$\beta$ B3	(l) 14	80	590	7.4
$\beta$ B4	(l) 3	87	720	8.3
$\beta$ B5	(l) 4	82	610	7.4
$\beta$ B6	(l) 3	84	900	11
$\beta$ B7	(l) 7	11	8	0.73
$\beta$ C1	(l) 20	89	1650	18
$\beta$ C2	(l) 10	75	1020	14
$\beta$ C3	(l) 16	35	105	3.0
$\beta$ D1	12	48	1060	22
$\beta$ D2	25	81	3900	48
$\beta$ D3	50	70	2750	39
$\beta$ D4	3	49	125	2.6
$\beta$ D5	3	78	130	1.7
$\beta$ D6	5	11	15	1.4
$\beta$ D7	12	22	30	1.4
$\beta$ D8	5	49	750	15
$\beta$ D9	1	3	3	1.0
$\beta$ D10	15	72	300	4.2
$\beta$ D11	2	31	300	9.7
$\beta$ D12	1	1	2	2.0
$\beta$ D13	30	4	5	1.3
$\beta$ D14	20	64	50	0.78
$\beta$ D15	(l) 20	11	12	1.1
$\beta$ D16	(l) 30	9	23	2.6
$\beta$ D17	(l) 14	81	1020	13
$\beta$ E1	5	22	10	0.45
$\beta$ E2	4	5	6	1.2
$\beta$ E3	0.6	25	9	0.36
$\beta$ E4	4	14	6	0.43
$\beta$ E5	2	5	1	0.20
$\beta$ E6	2	82	120	1.5
$\beta$ E7	1	15	8	0.53
$\beta$ E8	1	43	50	1.2
$\beta$ E9	5	1	1	1.0
$\beta$ E10	0.7	29	25	0.86

<sup>1</sup> (l) depicts bred lines. The other mice were analysed as founder fetuses 15 days following microinjection.

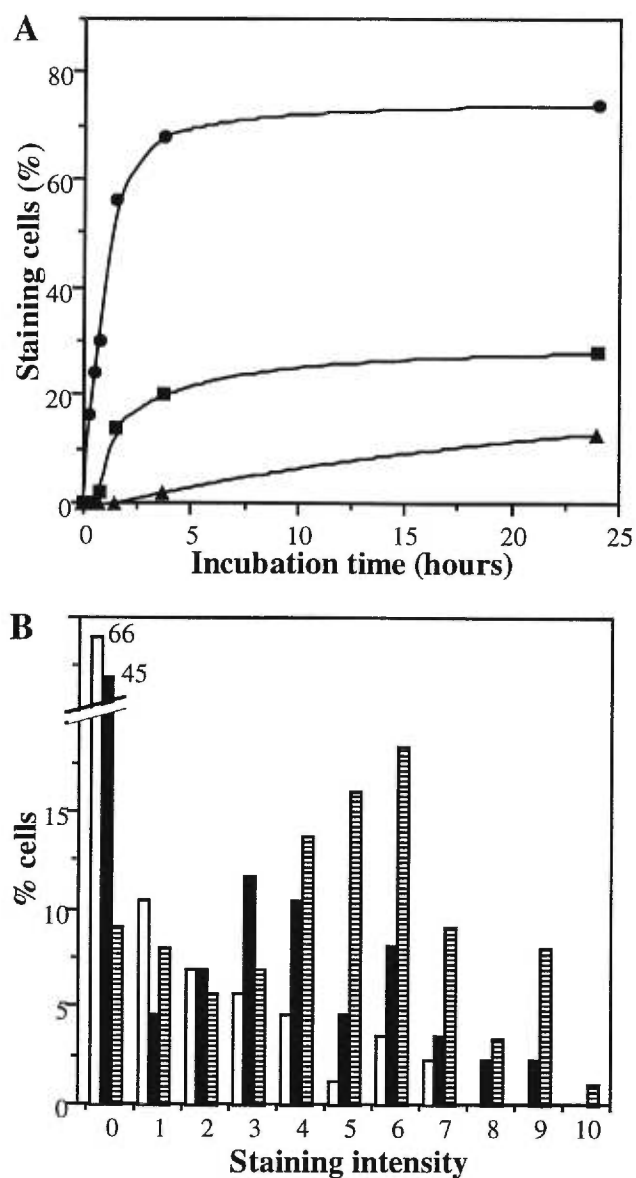
<sup>2</sup> Copy number was measured from F1 tail DNA for bred lines and from fetal liver DNA for founder fetuses.

<sup>3</sup> Cells from the fetal liver were stained with X-Gal for  $\beta$ -galactosidase and blue cells were counted after 24 hours.

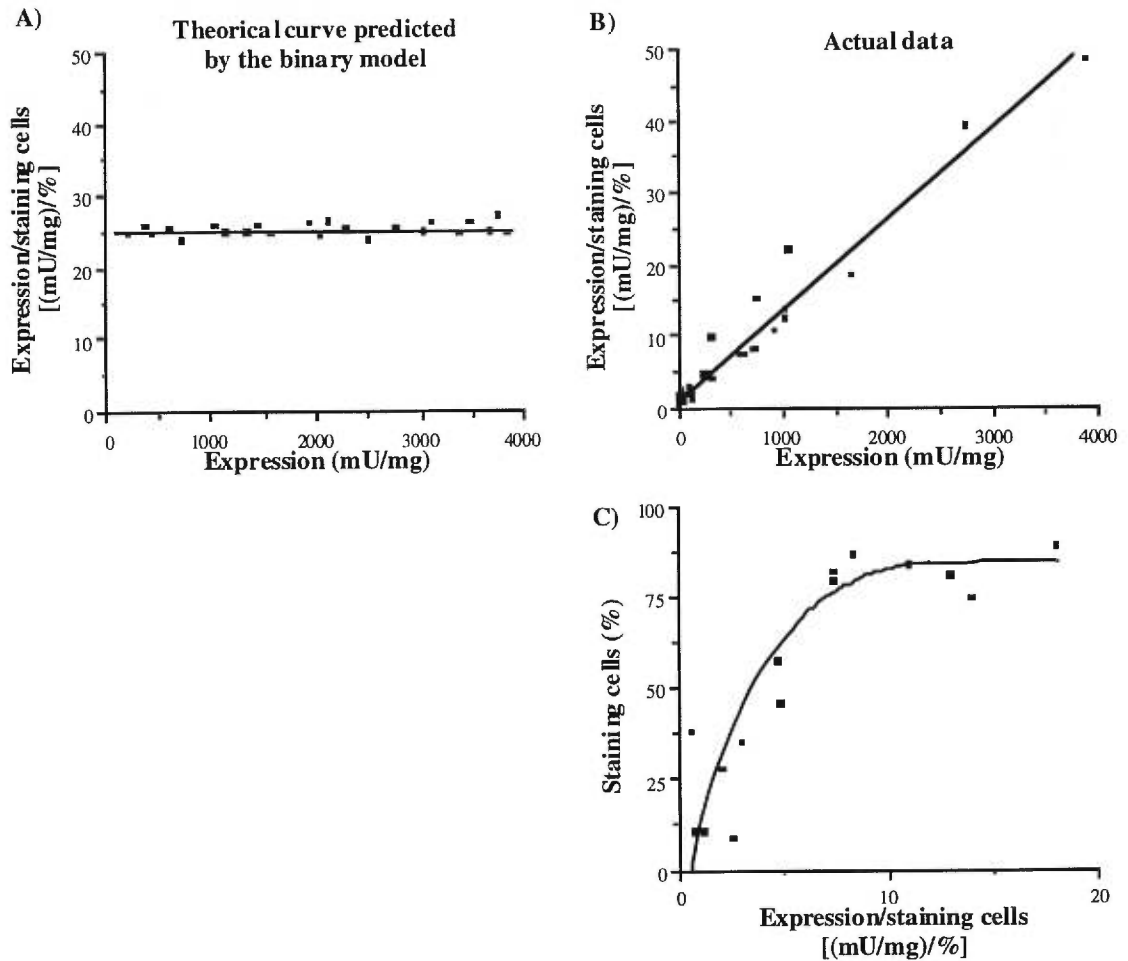
<sup>4</sup>  $\beta$ -galactosidase activity measured in protein extracts from fetal liver.



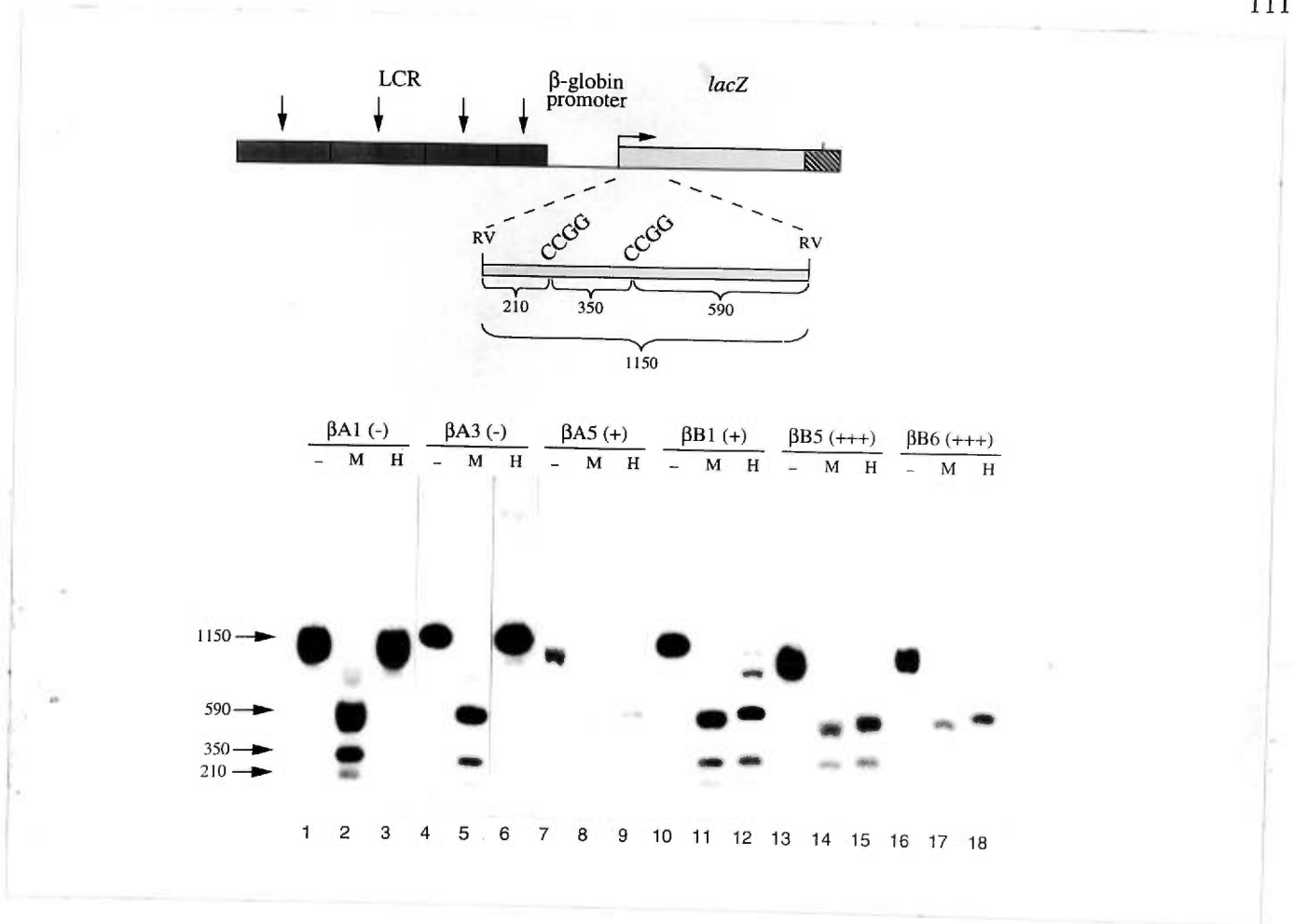
**Figure 1:** Kinetics of appearance of stained cells *in situ*. Fetal liver cells from transgenic lines that gave a low ( $\beta$ D15), medium ( $\beta$ A5) or high ( $\beta$ B5) fraction of staining cells were incubated with X-gal and were photographed at different times. The transgenic lines studied are indicated to the left of the rows, while the time of incubation is indicated at the top of the columns.



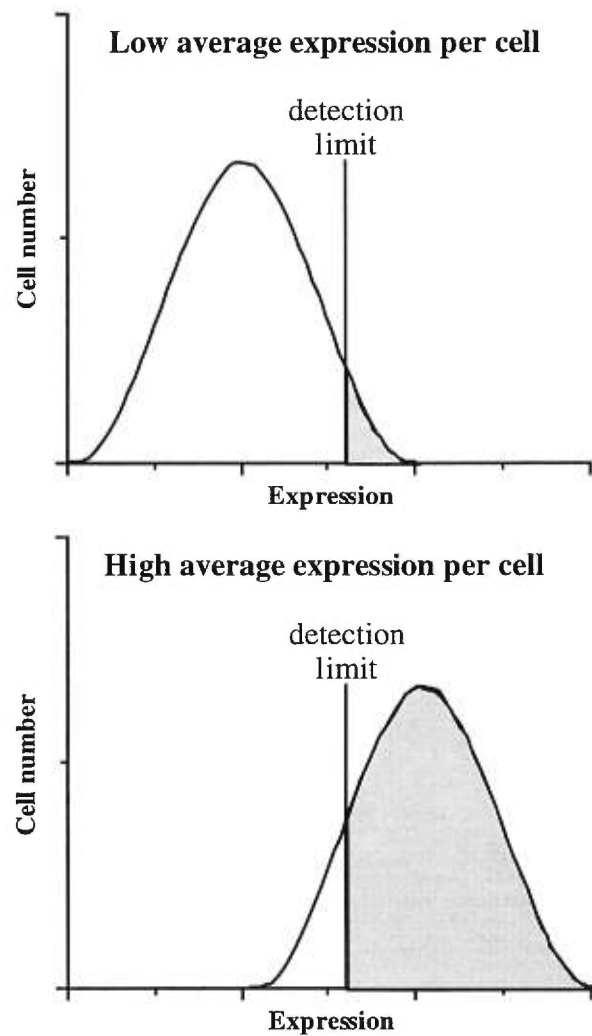
**Figure 2:** The fraction of detectable cells and the intensity of blue coloration increases with time. (A) Fetal liver cells from different transgenic lines ( $\beta$ B5, circles;  $\beta$ A5, squares; and  $\beta$ D15, triangles) were treated with X-gal and the percent cells that could be discerned to contain blue coloration was determined at different times. (B) Fetal liver cells of an animal from the  $\beta$ B5 line were treated with X-gal and the intensity of blue coloration of individual cells at 45 min (empty boxes), 2 h (filled boxes), and 24 h (dashed boxes) was determined using densitometry as indicated under Materials and Methods.



**Figure 3:** The average expression per cell determines both the fraction of cells that can be detected as  $\beta$ -galactosidase positive and the total level of  $\beta$ -galactosidase activity in the cell population. (A) The relationship between expression per staining cell and total expression expected for a binary model is shown. A straight line with zero slope is expected since the model predicts that all integration sites will give the same expression per cell and only the number of cells expressing should vary. (B) The actual data obtained are plotted. It can be seen that the expression per staining cell actually increases proportionally to the total expression. All transgenic samples, including lines and founders, but excluding the two transgenic lines that had no detectable expression, were used for this plot. (C) The fraction of cells that were  $\beta$ -galactosidase positive for each of the lines that had detectable expression is plotted against the average expression per positive cell. It can be seen that the number of cells that are detectable as  $\beta$ -galactosidase positive is determined by the average level of expression per cell. Note that founder transgenic samples were not used for the plot in (C), since the possibility that a founder was mosaic could have resulted in a fraction of the cells not containing the transgene, which would specifically affect the percent cells staining parameter.



**Figure 4:** Methylation of the *lacZ* transgene DNA in different transgenic lines. Genomic DNA isolated from fetal liver was digested with *EcoRV*, which generates a 1150 bp fragment from the 5' end of the *lacZ* reporter gene (see top of Figure). The DNA was either not treated further (lanes marked -) or was digested with *MspI* (lanes marked M) or *HpaII* (lanes marked H). Both of these enzymes can restrict the sequence CCGG and will generate fragments of 590, 350, and 210 bp from the 1150 bp *EcoRV* fragment (see top and the positions of the fragments on the Southern blot are indicated by the arrows to the left). However, *HpaII* is sensitive to methylation of the internal C residue and will not restrict the DNA if it is methylated, whereas *MspI* is not sensitive to methylation here. The DNA was separated on agarose gels, transferred to nylon membrane and probed with  $^{32}\text{P}$ -labeled *EcoRV lacZ* fragment. The transgenic line studied is indicated on the top of each triplet of lanes and the fraction of cells for each line that were  $\beta$ -galactosidase positive are indicated in brackets as -, no positive cells; +, a low fraction of positive cells; +++, a high fraction of positive cells. Note that the different lanes were spliced from autorads in which the same Southern blot was exposed to film for different periods of time to equalize the signal intensities.



**Figure 5:** Theoretical curves showing that when there is a random cell to cell distribution of expression in a population of cells, a population with a low average level of expression per cell will have a much higher fraction of cells falling below the detection limit of the assay being employed [and many cells may not have any expression at all (see text)] than a population with a high average expression per cell.

## Chapitre 4

### Article 3

EKLF is essential for  $\beta$ -globin gene expression even in absence of gene competition, but is not sufficient to induce the switch from  $\gamma$ -globin to  $\beta$ -globin gene expression

Louis-Georges Guy, Qi Mei, Andrew C. Perkins, Stuart H. Orkin  
et Lee Wall. (Sous revision)



**EKLF is essential for  $\beta$ -globin gene expression even in absence of gene competition, but is not sufficient to induce the switch from  $\gamma$ -globin to  $\beta$ -globin gene expression.#**

Louis-Georges Guy<sup>1</sup>, Qi Mei<sup>1</sup>, Andrew C. Perkins<sup>2,5</sup>, Stuart H. Orkin<sup>2,3</sup> and Lee Wall<sup>1,4\*</sup>

<sup>1</sup>Centre de recherche du Centre hospitalier de l'Université de Montréal and the Institut du cancer de Montréal, Montreal, Quebec, Canada

<sup>2</sup>Division of Hematology/Oncology, Children's Hospital and the Dana Farber Cancer Institute, Boston, Massachusetts, U.S.A.

<sup>3</sup>Howard Hughes Medical Institute, Boston, Massachusetts, U.S.A.

<sup>4</sup>Department of Medicine, Université de Montréal, Montreal, Quebec, Canada

<sup>5</sup>Current address: Department of Physiology, Monash University, Australia

Running title: **Role of EKLF in  $\beta$ -globin gene switching.**

\*Send Correspondence to Lee Wall at Centre de recherche du Centre hospitalier de l'Université de Montréal, Room 4605, Pavillon Notre-Dame, 1560 Sherbrooke East, Montreal, Quebec, Canada H2L 4M1      Tel: 514-281-6000, ext. 5497

Fax: 514-896-4689

E-mail: WALLL@ERE.UMONTREAL.CA

#This work was supported by a grant from the Medical Research Council of Canada to L.W., through a donation from the Fondation de l'Hôpital Notre-Dame and by salary support from Defi Coporatif Canderel to Q.M.. L.W. is a Scholar of the Fonds de la Recherche en Santé du Québec.

**Abstract:**

Different genes in the  $\beta$ -like globin locus are expressed at specific times during development. This is controlled, in part, by competition between the genes for activation by the locus control region. In mice, gene inactivation of the EKLF transcription factor results in a lethal anemia due to a specific and substantial decrease in expression of the fetal/adult stage-specific  $\beta$ -globin gene. In transgenic mice carrying the complete human  $\beta$ -globin locus, EKLF ablation not only impairs human  $\beta$ -globin gene expression, but also results in increased expression of the human  $\gamma$ -globin genes during the fetal/adult stages. Hence, it may appear that EKLF is a determining factor for the developmental switch from  $\gamma$ -globin to  $\beta$ -globin transcription. However, we show here that the function of EKLF for  $\beta$ -globin gene expression is necessary even in absence of gene competition. Moreover, EKLF is not developmental specific and is present and functional before the switch from  $\gamma$ -globin to  $\beta$ -globin gene expression occurs. Thus, EKLF is not the primary factor that controls the switch. Repression of  $\gamma$ -globin transcription that occurs during late fetal development is likely to be the initiating event that induces the switch.

### **Introduction:**

In mammals, expression of globin genes is erythroid specific and is developmentally regulated<sup>reviewed in 1</sup>. The human  $\beta$ -globin locus contains five active genes:  $\epsilon$ , which is transcribed in the embryonic yolk sac;  $G\gamma$  and  $A\gamma$ , which are expressed in the fetal liver, and  $\delta$  (a minor contributor) and  $\beta$  which are transcribed in the bone marrow throughout adult life. These five genes are arranged in their temporal order of expression (5'- $\epsilon$ - $G\gamma$ - $A\gamma$ - $\delta$ - $\beta$ -3'); see Fig. 1). There is only one well defined switch in globin gene expression from the mouse  $\beta$ -like globin locus. Mice express  $\epsilon^y$ ,  $\beta^h0$  and  $\beta^h1$  globins during embryonic life and in early fetal development. Thereafter, expression switches to the  $\beta$ -minor and  $\beta$ -major globins genes, and these two genes continue to be transcribed into adult life.

For high level expression, the  $\beta$ -like globin genes are dependent on a locus control region (LCR). The LCR in the  $\beta$ -globin locus is signified by four DNase I hypersensitive site regions situated upstream of the genes (Fig. 1). This DNA element has very strong erythroid-specific enhancer activity and is characterized as an LCR by its ability to confer copy number-dependent, integration site-independent, endogenous levels of expression onto globin transgenes<sup>reviewed in 2-4</sup>.

In mice, transcription of a human  $\gamma$ -globin transgene present alone in *cis* with the LCR is down regulated, but only partially, during fetal development<sup>5-9</sup>. The human  $\beta$ -globin gene is expressed at equivalent levels at all developmental stages in transgenic mice when it is present alone with the LCR<sup>6,7</sup>. On the other hand, if  $\gamma$ -globin and  $\beta$ -globin genes are linked together with the LCR (LCR- $\gamma$ - $\beta$ ), both genes are appropriately regulated during development: the  $\beta$ -globin gene is not expressed in the embryo and then is switched on near 11.5 days in fetal life, while the  $\gamma$ -globin gene is expressed in the embryo and then is markedly down-regulated during fetal life when the  $\beta$ -globin gene is up-regulated<sup>6,7,10</sup>. These observations, in conjunction with several other experiments<sup>11-15</sup>, have suggested a competitive model to describe the developmental regulation of

globin genes. It has been postulated that a globin gene must interact directly with the LCR through a looping mechanism for it to be activated and the LCR can only interact with one gene at a time. Thus, the globin genes compete with each other to contact the LCR<sup>6,7,10</sup>. At each developmental stage, the array of proteins that interact with the LCR and/or the globin genes favors interaction with a particular gene. As switching takes place, modifications in the transcription factor environment alter the relative affinity between the LCR and the genes such that the interaction switches from one gene to the next. Therefore, competition between the globin genes partly dictates their developmental pattern of expression<sup>6,7,9</sup>. Identifying the transcription factors involved in LCR-gene interactions and/or the switching process would be a major step towards elucidating the mechanism regulating globin gene developmental specificity and the LCR mode of action.

The erythroid krüppel-like factor (EKLF) plays an important role in transcription of the mouse fetal/adult and human adult stage-specific  $\beta$ -globin genes (See Fig. 1). Upon binding to the CACC box element in the promoters of these globin genes, EKLF can act as a transcriptional activator<sup>16-20</sup>. Mice in which the EKLF gene is knocked out suffer from a severe anemia and die at approximately sixteen days *post coitus*<sup>21,22</sup>. The anemia is caused by a major ( $\approx 20$  fold) down-regulation in expression of the mouse  $\beta$ -major gene. Other erythroid-specific genes, including other globin genes, are expressed at normal levels and erythropoiesis is apparently otherwise unaffected. Therefore, the requirement for EKLF function appears to be very specific to the fetal/adult-specific  $\beta$ -major globin gene in mice. Similarly, in transgenic mice carrying the complete human  $\beta$ -like locus, expression of the human  $\beta$ -globin gene is severely hampered in an EKLF  $-/-$  background<sup>15,23</sup>. However, in the absence of EKLF the human  $\gamma$ -globin transgenes are actually expressed at higher than normal levels during fetal life, up to the time of embryonic lethality. This result indicates that the switch in LCR interaction from the  $\gamma$ -globin to the  $\beta$ -globin gene that is expected to occur in early fetal development in mice may not take place in the absence of EKLF.

A model to account for the above results would suggest that EKLF is directly implicated in the interaction of the  $\beta$ -globin gene with the LCR and that EKLF activity is developmental specific. EKLF does interact with the erythroid-specific protein GATA-1<sup>24,25</sup>, for which there exist many DNA-binding sites in the LCR<sup>26-29</sup>. Thus, there is a potential for EKLF to be involved in LCR-gene interactions. Whether or not the function of EKLF is developmentally regulated has not been tested directly, although it is known that the mRNA is present at early stages<sup>20,30</sup>. On the other hand, the fact that the  $\beta$ -globin gene is expressed at all stages of development when it is alone with the LCR (see above) questions the requirement for a developmental specific modifier acting on the  $\beta$ -globin gene and involved in the switching process. EKLF function might, therefore, only be necessary for competition against other globin genes. To address these questions, we have looked at EKLF function in the absence of gene competition and during development.

## **Materials and Methods:**

### Transgenic mice

EKLF +/- mice and  $\mu$ D14 mice, which have been described previously<sup>22,31</sup>, were maintained on a C57B16 x 129 and C57B16 x CBA background, respectively. Fetuses were dissected out at the indicated times following the observation of the spermatic plug.

### DNA analyses

Fetal DNA was digested with *Hind*III, separated on agarose gels and blotted onto nylon membranes. The membranes were hybridized to a *Nco*I-*Sma*I 550 bp fragment from the 5' region of EKLF cDNA that distinguishes between the wild type and the targeted EKLF allele. The membranes were re-hybridized with a 750 bp *Hind*III fragment from HSS2 of the LCR that detects this same 750 bp fragment from the  $\mu$ D transgene.

### RNA analyses

RNA was isolated using Trizol reagent (Gibco-BRL) as described by the manufacture. RNase protection assays were conducted as described<sup>32</sup> with antisense probes that span the 5' region of the mRNA.

## **Results:**

The  $\beta$ -globin gene requires EKLF for expression even when it is juxtaposed to the LCR and no gene competition is present.

In the complete  $\beta$ -like loci, the LCR is about 60 kb away from the  $\beta$ -globin gene (Fig. 1), there are several genes that compete with the  $\beta$ -globin gene for interaction with the LCR and this competition plays a role in the developmental regulation of these genes<sup>6,7,10</sup>. Here, we aimed to test the function of EKLF in  $\beta$ -globin gene expression in the absence of gene competition.

We studied a transgenic mouse line carrying the human micro-locus construct<sup>33,34</sup> in which the 4 hypersensitive site regions of the LCR are directly linked in *cis* to the complete human  $\beta$ -globin gene. In this construct, the LCR is within 4 kb of the TATA box of the  $\beta$ -globin gene promoter, as measured from the middle of DNase hypersensitive sites 2 and 3 of the LCR. The expression of this transgene is erythroid-specific and position-independent<sup>33</sup>. We used the  $\mu$ D14 transgenic line, which carries a single copy of the transgene and expresses human  $\beta$ -globin at 43% the level of the endogenous mouse  $\beta$ -major gene<sup>31</sup>.  $\mu$ D14 transgenic mice were bred to mice heterozygous for the EKLF inactivated allele. F1 animals that were transgenic and EKLF +/- were bred again to EKLF heterozygous mice to obtain  $\mu$ D14 fetuses null for the EKLF gene. The genotype of the animals was determined by Southern blot using a probe that distinguishes the EKLF wild type allele from the knockout allele and a probe that is specific for the human LCR region of the transgene (Fig. 2). Human  $\beta$ -globin transgene expression was measured relative to mouse  $\beta$ -major mRNA via RNase protection (Fig. 3). Comparison of human  $\beta$ -globin mRNA levels in fetal livers from EKLF knock-out, heterozygous and wild type animals demonstrated that expression of the transgene is as dependent on EKLF as the endogenous mouse  $\beta$ -major gene. Both genes gave 15-20 fold lower expression in the absence of EKLF. In comparison, GATA-1 expression was not affected by EKLF inactivation (Fig. 3). Thus, although the  $\beta$ -globin promoter is juxtaposed to the LCR and

no gene competition exists in the  $\mu$ D transgene,  $\beta$ -globin expression is still highly dependent on EKLF.

EKLF function is not developmental-specific.

When the human  $\beta$ -globin gene is alone in *cis* with the LCR, it is expressed at all developmental stages<sup>7,10,35</sup>. Thus, we could determine whether the EKLF protein is active at early embryonic stages by measuring the expression of the  $\mu$ D transgene in the presence and absence of EKLF. RNA from yolk sacs (the site of primitive erythropoiesis) was isolated from EKLF wild type, heterozygous and null 10.5 day embryos. Human  $\beta$ -globin transgene expression was again measured by RNase protection, but the embryonic stage-specific mouse  $\epsilon^y$ -globin mRNA served as the control in this case. Expression of the transgene was strongly dependent on EKLF. Very little human  $\beta$ -globin expression could be detected in EKLF *-/-* embryos, but the transgene was expressed at approximately 30% the level of the mouse  $\epsilon^y$  gene in wild type and heterozygous embryos (Fig. 4, compare lane 3 to lanes 1 and 2). Results identical to those shown in Fig. 4 were obtained for a second set of embryos in which two additional EKLF *-/-* animals carried the transgene. Thus, the dependence of  $\beta$ -globin gene expression on EKLF, and hence EKLF function, is not developmental specific.



**Discussion:****EKLF is not the primary factor for globin switching.**

It had been suggested previously that EKLF function may be developmental specific, partly because it is absent from K562 cells which express human embryonic and fetal globin genes, but not the adult  $\beta$ -globin gene<sup>20</sup>. If this were the case, EKLF might serve as the predominant switch protein in the regulation of globin genes. Previously, it had been shown that EKLF mRNA and DNA-binding activity are indeed present in embryonic erythroid tissues<sup>20,22,30</sup>. However, it had not been excluded that EKLF activity is controlled in developmental specific manner. For example, EKLF function might dependent on protein modifications and/or protein partners that are developmentally regulated. However, as reported here, we have found that the  $\beta$ -globin transgene in the  $\mu$ D construct is as dependent on EKLF for expression in embryonic erythroid tissues as it is in fetal liver (Fig. 4). Thus, the function of EKLF is not developmental specific. We conclude that EKLF cannot be responsible for the initiation of globin gene switching, since it is present and fully active before this process occurs. Similarly, EKLF alone cannot be the primary determinant that dictates the  $\beta$ -globin gene as a better substrate for interacting with the LCR in later development. In this respect, the intrinsic down-regulation of the  $\gamma$ -globin genes that occurs during development<sup>6-9</sup> may be the primary factor that allows for the LCR to preferentially interact with the  $\beta$ -globin gene in later developmental stages. We suggest that the  $\beta$ -globin gene may be fully primed to interact with the LCR before it is actually expressed, but the interaction between the  $\beta$ -globin gene and LCR does not occur until the  $\gamma$ -globin genes are down regulated and their interaction with the LCR is weakened. We cannot rule out that a protein other than EKLF is involved in strengthening LCR- $\beta$ -globin gene interactions in a developmental regulated fashion. However, expression of the  $\beta$ -globin gene at all developmental stages when present alone with the LCR<sup>6,7</sup> suggests this is not the case.

### Does EKLF play a role in LCR- $\beta$ -globin gene interactions?

In the absence of the LCR, a human  $\beta$ -globin transgene is expressed at more than 100 fold lower levels than when it is in *cis* with the LCR<sup>6</sup>. Yet, in EKLF knockout mice,  $\beta$ -globin expression is only down 15 to 20 fold<sup>36</sup> and our own results<sup>37</sup>. Thus, the interaction between the LCR and  $\beta$ -globin gene can occur, at least to some extent, without EKLF.

The probability of interaction between a gene and the LCR decreases with increasing distance between the two elements. The probability is predicted to be roughly inversely proportional to the cube of the distance between the gene and LCR<sup>10</sup>. In the endogenous mouse and human  $\beta$ -like loci, the  $\beta$ -globin genes are approximately 60 kb away from the LCR, while in the  $\mu$ D construct, there is only about 4 kb of DNA between the LCR and  $\beta$ -globin gene. Thus, if EKLF is directly involved in LCR-gene interactions, one might have expected the  $\mu$ D construct to be much less dependent on EKLF. However, the  $\mu$ D construct was found to be as dependent on EKLF for expression as is the mouse  $\beta$ -major gene. On an EKLF<sup>-/-</sup> background, both gave about 15 fold lower expression in fetal liver (Fig. 3). Thus, EKLF may not play a predominant role in LCR-gene interactions, but might participate along with other more critical proteins. On the other hand, the decrease in  $\beta$ -globin gene expression that is caused by natural occurring point mutations of the  $\beta$ -globin promoter CACC box, which result in a loss of EKLF binding, can be recapitulated *in vitro* using reporter constructs that do not include the LCR<sup>37,38</sup>. In addition, deletion or mutation of the  $\beta$ -globin CACC box results in a substantial decrease in promoter activity, whether or not reporter constructs contain the LCR<sup>39-41</sup>. Hence, EKLF function might be more directly concerned with the promoter activity of the  $\beta$ -globin gene than with LCR-gene interactions. If this is the case, we need to determine how LCR-globin gene interactions are controlled to understand why  $\gamma$ -globin is expressed at higher levels in the absence of EKLF.

**Acknowledgments:**

We are grateful to R. Kothary for critical reading of the manuscript. We thank Dr. Tim Townes and Dr. Jim Bieker for providing an EKLF clone and Dr. James Ellis for providing the  $\mu$ D14 transgenic mice.

**References:**

1. Stamatoyannopoulos G, Nienhuis AW: Hemoglobin switching, in Stamatoyannopoulos G, Nienhuis AW, Majerus PW and Varmus H (eds): *The Molecular Basis of Blood Diseases*. Philadelphia, PA, Saunders Co., 1994, p 107
2. Townes TM, Behringer RR: Human globin locus activation region (LAR): role in temporal control. *Trends Genet* 6:219, 1990
3. Epner E, Kim CG, Groudine M: What does the locus control region control? *Curr Biol* 2:262, 1992
4. Dillon N, Grosveld F: Transcriptional regulation of multigene loci: multilevel control. *Trends Genet* 9:134, 1993
5. Enver T, Ebens AJ, Forrester WC, Stamatoyannopoulos G: The human  $\beta$ -globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice. *Proc Natl Acad Sci USA* 86:7033, 1989
6. Behringer RR, Ryan TM, Palmiter RD, Brinster RL, Townes TM: Human  $\gamma$ - to  $\beta$ -globin gene switching in transgenic mice. *Genes Dev* 4:380, 1990
7. Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G: Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* 344:309, 1990
8. Dillon N, Grosveld F: Human  $\gamma$ -globin genes silenced independently of other genes in the  $\beta$ -globin locus. *Nature* 350:252, 1991
9. Stamatoyannopoulos G, Josephson B, Zhang JW, Li Q: Developmental regulation of human  $\gamma$ -globin genes in transgenic mice. *Mol Cell Biol* 13:7636, 1993
10. Hanscombe O, Whyatt D, Fraser P, Yannoutsos N, Greaves D, Dillon N, Grosveld F: Importance of globin gene order for correct developmental expression. *Genes Dev* 5:1387, 1991

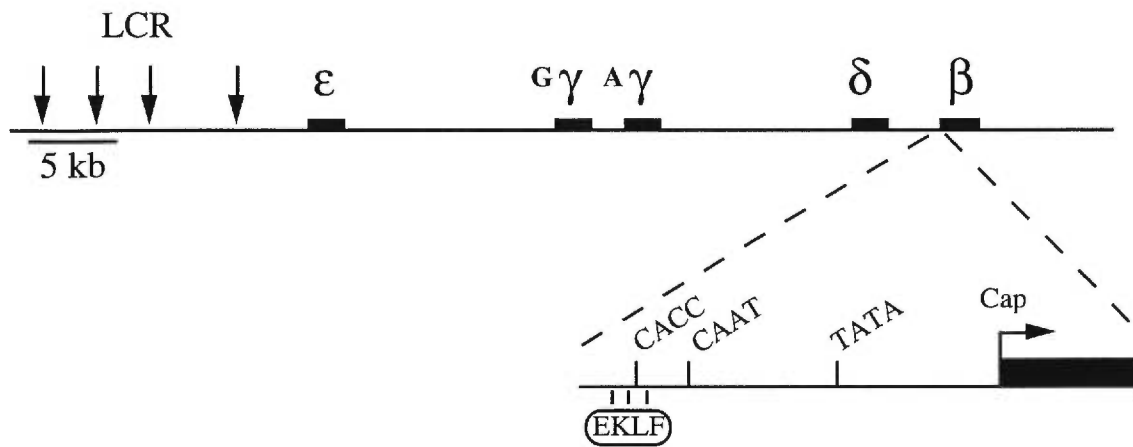
11. Kim CG, Epner EM, Forrester WC, Groudine M: Inactivation of the human  $\beta$ -globin gene by targeted insertion into the  $\beta$ -globin locus control region. *Genes Dev* 6:928, 1992
12. Dillon N, Strouboulis J, Grosveld F: The regulation of human  $\beta$ -globin gene expression: Polarity of transcriptional competition in the human  $\beta$ -globin locus, in Stamatoyannopoulos G (ed): *Molecular biology of hemoglobin switching, Proceedings of the 9th conference on hemoglobin switching*. Andover, Hants, UK, Intercept Limited, 1995, p 23
13. Fiering S, Epner E, Robinson K, Zhuang Y, Telling A, Hu M, Martin DIK, Enver T, Ley TT, Groudine M: Targeted deletion of 5'HS2 of the murine  $\beta$ -globin LCR reveals that it is not essential for proper regulation of the  $\beta$ -globin locus. *Genes Dev* 9:2203, 1995
14. Wijgerde M, Grosveld F, Fraser P: Transcription complex stability and chromatin dynamics in vivo. *Nature* 377:209, 1995
15. Wijgerde M, Gribnau J, Trimborn T, Nuez B, Philipsen S, Grosveld F, Fraser P: The role of EKLF in human  $\beta$ -globin gene competition. *Genes Dev* 10:2894, 1996
16. Miller IJ, Bieker JJ: A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol Cell Biol* 13:2776, 1993
17. Feng WC, Southwood CM, Bieker JJ: Analyses of  $\beta$ -thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* 269:1493, 1994
18. Bieker JJ: Role of erythroid Krüppel-like factor (EKLF) in erythroid-specific transcription, in Stamatoyannopoulos G (ed): *Molecular biology of hemoglobin switching, Proceedings of the 9th conference on hemoglobin switching*. Andover, Hants, UK, Intercept Limited, 1995, p 231
19. Bieker JJ, Southwood CM: The erythroid Krüppel-like factor transactivation domain is a critical component for cell-specific inducibility of a  $\beta$ -globin promoter. *Mol Cell Biol* 15:852, 1995

20. Donze D, Townes TM, Bieker JJ: Role of erythroid Krüppel-like factor in human  $\gamma$ -to  $\beta$ -globin gene switching. *J Biol Chem* 270:1955, 1995
21. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F: Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375:316, 1995
22. Perkins AC, Sharpe AH, Orkin SH: Lethal  $\beta$ -thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* 375:318, 1995
23. Perkins AC, Gaensler KM, Orkin SH: Silencing of human fetal globin expression is impaired in the absence of the adult  $\beta$ -globin gene activator protein EKLF. *Proc Natl Acad Sci USA* 93:12267, 1996
24. Merika M, Orkin SH: Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. *Mol Cell Biol* 15:2437, 1995
25. Gregory RC, Taxman DJ, Seshasayee D, Kensinger MH, Bieker JJ, Wojchowski DM: Functional interaction of GATA-1 with erythroid Krüppel-like factor and Sp1 at defined erythroid promoters. *Blood* 87:1793, 1996
26. Philipsen S, Talbot D, Fraser P, Grosveld F: The beta-globin dominant control region: hypersensitive site 2. *EMBO J* 9:2159, 1990
27. Talbot D, Philipsen S, Fraser P, Grosveld F: Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO J* 9:2169, 1990
28. Pruzina S, Hanscombe O, Whyatt D, Grosveld F, Philipsen S: Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Research* 19:1413, 1991
29. Talbot D, Grosveld F: The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J* 10:1391, 1991

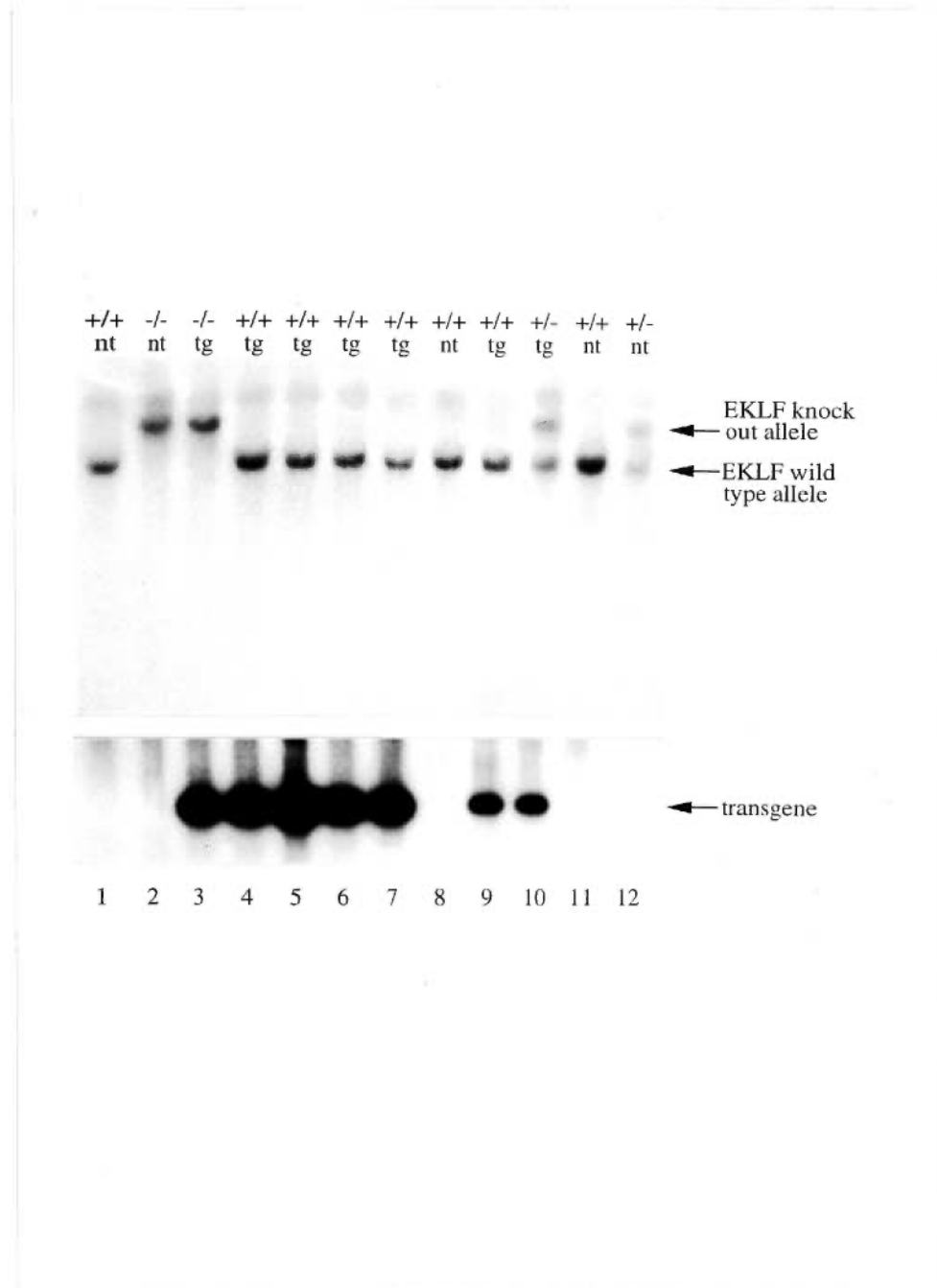
30. Southwood CM, Downs KM, Bieker JJ: Erythroid Krüppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dynamics* 206:248, 1996
31. Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipsen S, Grosveld F: A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human  $\beta$ -globin locus control region. *EMBO J* 15:562, 1996
32. Delvoeye NL, Destroismaisons NM, Wall LA: Activation of the  $\beta$ -globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol Cell Biol* 13:6969, 1993
33. Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, Greaves DR: A dominant control region from the human  $\beta$ -globin locus conferring integration site-independent gene expression. *Nature* 338:352, 1989
34. Collis P, Antoniou M, Grosveld F: Definition of the minimal requirements within the human  $\beta$ -globin gene and the dominant control region for high level expression. *EMBO J* 9:233, 1990
35. Blom van Assendelft G, Hanscombe O, Grosveld F, Greaves DR: The  $\beta$ -globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* 56:969, 1989
36. Lim S-K, Bieker JJ, Lin C-S, Costantini F: A shortened life span of EKLF  $-/-$  adult erythrocytes, due to a deficiency of  $\beta$ -globin chains, is ameliorated by human  $\gamma$ -globin chains. *Blood* 90:1291, 1997
37. Orkin SH, Antonarakis SE, Kazazian HH, Jr.: Base substitution at position -88 in a  $\beta$ -thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J Biol Chem* 259:8679, 1984
38. Kulozik AE, Bellan-Koch A, Bail S, Kohne E, Kleihauer E: Thalassemia intermedia: moderate reduction of  $\beta$  globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood* 77:2054, 1991

39. Dierks P, van Ooyen A, Cochran MD, Dobkin C, Reiser J, Weissmann C: Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit  $\beta$ -globin gene in mouse 3T6 cells. *Cell* 32:695, 1983
40. Myers RM, Tilly K, Maniatis T: Fine structure genetic analysis of a  $\beta$ -globin promoter. *Science* 232:613, 1986
41. Antoniou M, Grosveld F:  $\beta$ -globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev* 4:1007, 1990

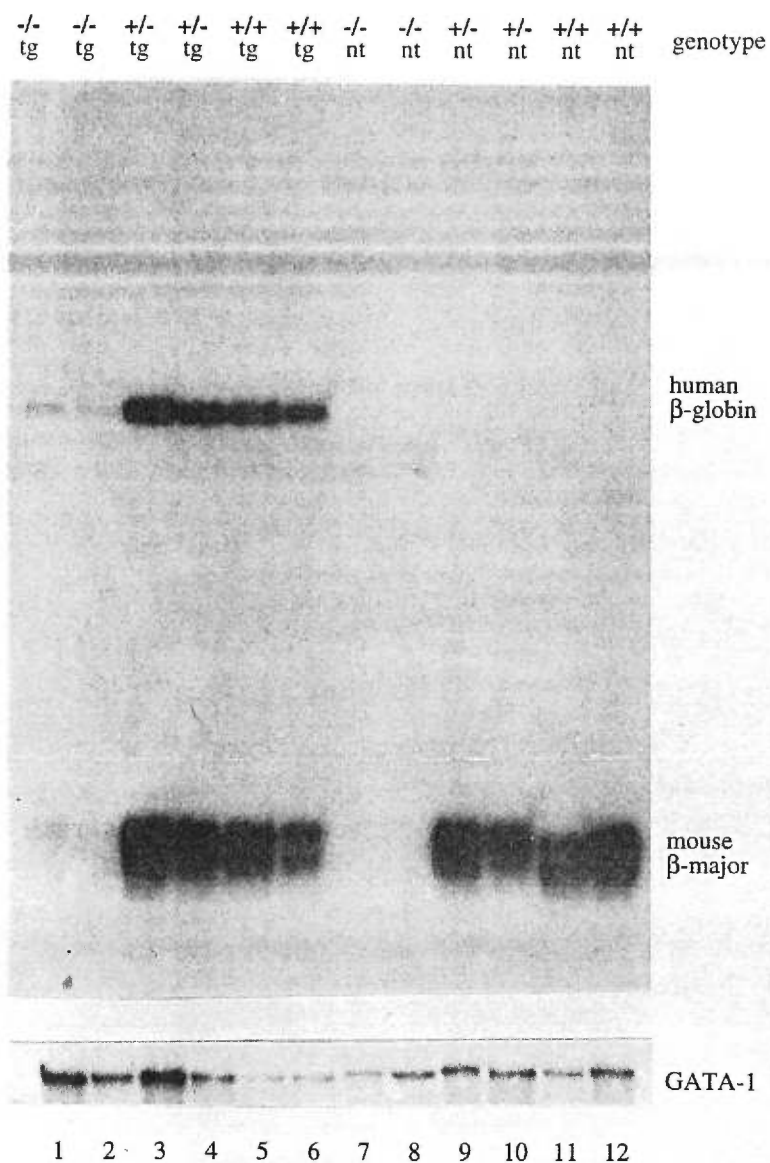




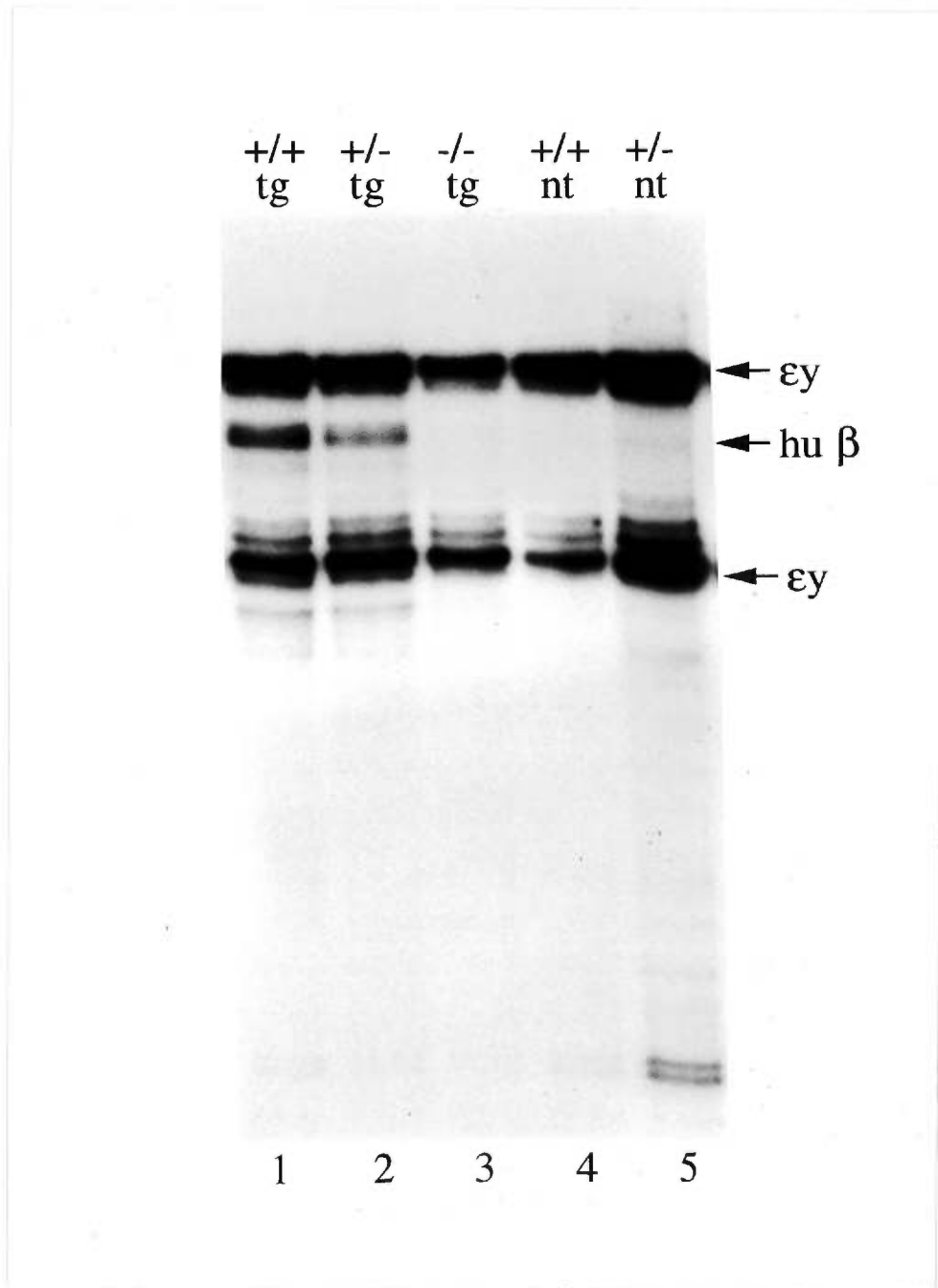
**Figure 1:** Structure of the human  $\beta$ -like globin locus and the  $\beta$ -globin gene minimal promoter. The figure also shows EKLf that interacts with the CACC motif element of the promoter.



**Figure 2:** Genotype determination: Fetus DNA digested with *Hind*III was analyzed by Southern blot with an EKLf probe that distinguishes between the knockout allele (10 kb fragment detected) and the wild type allele (6 kb fragment detected). The same blot was re-probed with LCR sequences that hybridize to a 0,75 kb *Hind*III fragment from the transgene. The EKLf genotype and the presence (tg) or absence (nt) of the transgene is indicated at the bottom of each lane.



**Figure 3:** EKLf is required for  $\beta$ -globin transgene expression in the fetal-adult stage: Fourteen and a half day fetal liver RNA was isolated from  $\mu$ D14 transgenic (tg) animals or non-transgenic (nt) litter mates with the EKLf genotype indicated at the top of each lane. The RNA was analyzed by RNase protection using a human  $\beta$ -globin and a mouse  $\beta$ -major probe. The RNA was also simultaneously analyzed with a GATA-1 probe as a loading control. The protected fragments are indicated to the right of the figure.



**Figure 4:** EKLf is also required for  $\beta$ -globin transgene expression at the embryonic stage: RNA isolated from 10.5 day yolk sac of embryos with the various EKLf genotypes and transgenic for the  $\mu$ D14 construct indicated at the top of the lanes were analyzed by RNase protection with a human  $\beta$ -globin transgene probe and a mouse  $\epsilon^y$ -globin probe. The position of the protected fragments detected by each probe is indicated by the arrows to the right of the figure.

## **Chapitre 5**

### Discussion

## **DISCUSSION:**

### **5.1 Validité de l'approche expérimentale**

#### 5.1.1 L'activité $\beta$ -galactosidase reflète les niveaux d'expression

Une importante partie des résultats analysés dans cette thèse se base sur la variation de niveaux d'activité  $\beta$ -galactosidase pour détecter les effets de position. Habituellement, ce sont plutôt les niveaux d'ARNm qui sont mesurés et, pour valider notre étude, nous nous sommes assurés que notre approche expérimentale nous permettait d'évaluer efficacement les effets de position.

Puisqu'une expression indépendante de la position permet la transcription de chaque copie du transgène à son plein potentiel, les niveaux d'expression deviennent proportionnels au nombre de copies. Dans ces conditions, l'expression par copie devrait être environ constante entre les différentes lignées de souris transgéniques (qui reflètent différents sites d'intégration). Pour tester la capacité des transgènes à conférer une expression indépendante de la position, nous avons mesuré les niveaux d'activité  $\beta$ -galactosidase par copie. Toutefois, il n'était pas impossible que l'activité enzymatique varie pour des raisons autres (affectant le messager ou la protéine) que des variations de transcription du transgène. Il fallait donc prouver que les niveaux d'activité  $\beta$ -galactosidase correspondait aux niveaux de transcription du transgène.

Nous avons mesuré la quantité d'ARNm *lacZ* dans plusieurs lignées différentes. L'intensité du signal *lacZ* variait fortement d'une lignée à l'autre et elle était proportionnelle à l'activité enzymatique (Chapitre 2, Figure 5). Donc, les variations d'activité enzymatique sont dues à des variations dans la quantité de messagers produite.

Par surcroît, la position du site d'intégration semble influencer directement la structure de la chromatine du transgène. Nous avons noté des variations dans l'hyperméthylation du LCR du transgène qui étaient reliées aux niveaux d'expression. De la même façon, la méthylation du transgène a été observée dans les lignées silencieuses. Donc, les niveaux d'expression mesurés sont directement liés avec l'état du transgène.

En se basant sur ces résultats, nous concluons que la variation dans l'activité enzymatique est réellement liée à des effets de position sur l'expression du transgène.

#### 5.1.2 Les souris fondatrices mosaïques ne nuisent pas à l'interprétation des résultats

Au début de l'étude, nous avons analysé des animaux de deuxième génération (ou davantage), donc provenant de lignées stables. L'interprétation des résultats est plus simple avec ce type d'animaux puisque le transgène, provenant d'une cellule germinale transgénique, se retrouvera dans le génome de tout l'organisme. Pour la suite de l'étude, nous avons privilégié l'analyse de souris transgéniques fondatrices dont certaines pourraient être mosaïques pour le transgène, c'est-à-dire que le fragment d'ADN s'est inséré dans le génome du zygote après la première division cellulaire, générant un animal formé d'un mélange de cellules normales et de cellules transgéniques. Les niveaux d'expression par copie pourraient donc varier malgré une expression indépendante de la position si, par exemple, il y avait peu de cellules transgéniques pour générer l'extrait protéique (pour mesurer l'activité  $\beta$ -galactosidase) mais beaucoup pour isoler l'ADN (pour mesurer le nombre de copies). Pour que les résultats soient toujours valables, même chez les animaux mosaïques, nous avons mis au point un protocole où des cellules de foie foetal sont utilisées comme source à la fois de protéines et d'ADN. Ainsi, la même proportion de cellules transgéniques se retrouvera dans les deux échantillons.

Par contre, les mesures de niveaux d'expression du transgène de  $\beta$ -globine humaine ont été faites par protections aux ribonucléases. Dans ces cas, l'ARN et l'ADN n'ont pas été isolés à partir des mêmes cellules. Nous avons donc opté pour une approche où le foie est divisé en deux parties qui seront utilisées pour isoler l'ARN pour les protections et l'ADN pour les buvardages. Il y a donc risque que la dilution des cellules transgéniques ne soit pas la même dans les deux échantillons. Toutefois, ce risque nous apparaît mineur pour plusieurs raisons. Premièrement, chaque tissu est probablement formé à partir d'un nombre relativement petit de cellules précurseurs; donc la mosaïcité devrait être plutôt constante à l'intérieur d'un même tissu. Ensuite, seul un petit nombre de souris fondatrices semblent être mosaïques. En effet, il est parfois possible de détecter les souris

mosaïques par des différences de nombre de copies moyen d'un tissu à l'autre. Ce type d'étude ne nous a permis de détecter qu'une petite fraction de souris mosaïques (environ 10%; données non-montrées). De plus, lors de la reproduction de souris fondatrices adultes, seules quelques unes n'ont pas donné naissance à des rejetons transgéniques dans les proportions mendéliennes, indiquant que la proportion de mosaïques était faible (données non-montrées).

Finalement, si la mosaïcité faisait varier l'expression par copie, seuls quelques animaux (les mosaïques) auraient une expression par copie divergente par rapport aux autres où la valeur serait à peu près stable. Or, de vastes écarts sont notés pour les constructions  $\beta I$ ,  $\beta J$  et  $\beta K$  (position dépendants, Annexe I) qui ne pourraient donc être dus à la mosaïcité de quelques animaux. De plus, la construction du micro-locus  $\beta F$  était exprimée de façon indépendante de la position (Chapitre 2) malgré ce risque de mosaïcité. Donc, même si certains animaux étaient mosaïques, cela ne devrait pas avoir d'effet important sur l'interprétation des résultats.

## **5.2 Le LCR ne peut conférer une expression indépendante de la position lorsqu'un autre rapporteur que le gène intact de $\beta$ -globine est utilisé**

Avant cette étude, le LCR de  $\beta$ -globine était considéré comme étant un régulateur dominant sur la chromatine environnante, suffisant en lui-même pour contrer les effets de position (89-91). Une des découvertes principales de cette thèse est le fait que le LCR ne peut réguler la chromatine du locus environnant si le gène qu'il active n'est pas un gène de globine. Le LCR contient donc des éléments essentiels pour conférer une expression indépendante de la position, toutefois ces éléments ne sont pas suffisants et doivent fonctionner en conjonction avec un gène adéquat. Ce phénomène nous renseigne sur le mécanisme du LCR. Il nous indique qu'il doit y avoir une quelconque "communication" entre le LCR et le gène, c'est-à-dire que le gène doit en quelque sorte induire l'activité du LCR. Une bonne partie de la thèse cherchait à découvrir comment cette communication pouvait avoir lieu.



### 5.2.1 L'activité enhancer du LCR est indépendante de l'ouverture de la chromatine

Malgré la perte de la capacité d'ouvrir la chromatine, nous avons montré que le LCR conserve toujours son activité enhancer. Avec des dissections moléculaires fines, plusieurs études tentaient de localiser ces deux activités sur des séquences discrètes du LCR (169-186). Au mieux ont-elles trouvé des sites principalement enhancers et des régions plus ou moins capables de conférer une expression indépendante de la position. Par exemple, HSS2 est le seul enhancer classique, c'est-à-dire fonctionnant en transfection transiente, à l'intérieur du LCR (174). HSS3 est quant à lui le seul site à pouvoir ouvrir la chromatine lors d'insertion d'un seul transgène (171). Toutefois, dans les faits, chacun de ces deux sites est important pour les deux activités à la fois. La délétion d'un des sites réduit les niveaux d'expression (activité enhancer altérée) et induit de légers effets de position (indépendance de la position altérée) (172,173). Donc, les deux activités semblent distribuées sur l'ensemble du LCR. Ces observations renforcent la théorie selon laquelle le LCR forme un holo-complexe qui agit comme une seule entité (171). Toutefois, nous démontrons qu'il est possible d'altérer l'ouverture de la chromatine sans empêcher l'activité enhancer, suggérant que ces activités pourraient fonctionner selon des mécanismes différents. Il est tout à fait possible que le holo-complexe ne soit impliqué que dans l'une des deux activités. Alternativement, il pourrait agir de deux façons différentes pour protéger contre les effets de position et pour stimuler la transcription.

### 5.2.2 Il n'y a pas d'éléments dans le gène de $\beta$ -globine dont la seule présence rétablit l'activité du LCR

La première hypothèse de travail était que des éléments du gène de  $\beta$ -globine autres que le promoteur et donc absents du transgène LCR- $\beta$ -*lacZ* ( $\beta$ B) étaient requis pour l'activité d'expression indépendante de la position. Les séquences situées en aval du promoteur de  $\beta$ -globine ont été rajoutées à ce transgène ( $\beta$ E) mais cela n'a pas suffi à prévenir les effets de position, indiquant qu'il n'y aurait pas de séquence discrète nécessaire pour l'activité du LCR dans le gène  $\beta$ -globine. Pourtant, le LCR lié exactement aux mêmes séquences, mais sans *lacZ* (micro-locus ou  $\beta$ F), permet une expression

indépendante de la position. Nous concluons donc que c'est directement l'insertion de *lacZ* dans le transgène qui altère la capacité du LCR d'ouvrir la chromatine.

### 5.2.3 *LacZ* n'agit pas comme un isolateur qui bloquerait la propagation de l'ouverture de la chromatine

Plusieurs modèles peuvent être mis de l'avant pour expliquer comment *lacZ* peut inhiber l'activité du LCR. Premièrement, le LCR pourrait empêcher la propagation de l'ouverture de la chromatine (Chapitre 1, Figure 8) d'une de ces façons: a) Par exemple, les séquences *lacZ*, qui proviennent d'un génome procaryote, pourraient adopter une conformation non compatible avec l'activité du LCR. b) Elles pourraient aussi agir en isolateur naturel qui empêcherait la propagation du signal de décompaction de la chromatine. c) Les séquences de *lacZ* sont peut-être liées des protéines toxiques pour l'activité du LCR. Si cette hypothèse est la bonne, d'autres séquences que *lacZ*, mise dans la même situation (remplaçant *lacZ* dans le transgène  $\beta E$ , par exemple) n'auraient pas le même effet.

Si *lacZ* agissait comme un isolateur, alors un gène de  $\beta$ -globine (complet et intègre) juxtaposé au LCR pourrait être exprimé indépendamment de la position. Au contraire, un gène de  $\beta$ -globine qui serait toujours séparé du LCR par une séquence *lacZ* serait automatiquement soumis à des effets de position. Le transgène  $\beta J$  (LCR- $\beta$ -*lacZ*- $\beta$ -globine; Annexe I) offre la possibilité d'une ouverture de la chromatine au gène de  $\beta$ -globine par propagation directe en provenance du LCR dans le cas d'une insertion de plusieurs copies tête contre queue. Par contre,  $\beta I$ , est un transgène similaire mais où un autre exemplaire des séquences *lacZ* a été ajoutée en 3' du gène de  $\beta$ -globine pour bloquer cette possibilité de juxtaposition directe. En testant l'expression de ces transgènes, nous avons noté de forts effets de position du gène de  $\beta$ -globine dans les deux cas. Puisqu'il n'y a pas de différences entre le patron d'expression des deux transgènes, nous concluons que *lacZ* n'agit pas en isolateur.

Une autre façon de prouver cette affirmation est de démontrer que d'autres séquences que *lacZ* ont le même effet sur l'activité du LCR. Nous avons utilisé l'ADNc

eucaryote TIMP dans deux constructions.  $\beta$ L place TIMP à la même position qu'avait *lacZ* dans  $\beta$ D (Annexe I, Figure 1). De la même façon,  $\beta$ K est l'équivalente de  $\beta$ J mais les séquences *lacZ* ont été substituées par TIMP. Ces deux transgènes étaient eux aussi incapable de conférer une expression indépendante de la position. Ces résultats prouvent que d'autres séquences que *lacZ* empêchent le LCR de manifester l'ouverture de la chromatine. Puisque *lacZ* n'est pas spécifiquement un élément qui bloque le LCR, nous concluons que c'est simplement son insertion qui cause le problème. Son action serait donc tout simplement de déplacer les éléments du gène de  $\beta$ -globine les uns par rapports aux autres (par exemple, le promoteur se trouve plus loin des enhancers) et ainsi empêcher la formation d'une "structure" capable d'activer le LCR.

#### 5.2.4 Une interaction mutuelle activant l'ouverture de la chromatine par le LCR pourrait être altérée par l'insertion de séquences dans le gène de $\beta$ -globine

La seconde hypothèse décrivant comment *lacZ* pourrait nuire à l'ouverture de la chromatine par le LCR serait qu'une interaction mutuelle entre le LCR et le gène de  $\beta$ -globine (Chapitre 1, Figure 7) (188) ne puisse avoir lieu si des séquences sont introduites dans le gène. Bien que toutes les séquences de  $\beta$ -globine soient présentes dans le transgène  $\beta$ E, puisque les divers éléments ne se situent pas tous à leur position normale les uns par rapport aux autres en raison de l'insertion de *lacZ*, il est possible que l'interaction n'ait pas lieu ou ne soit pas fonctionnelle. Plusieurs résultats pourraient être compatibles avec cette hypothèse. Les études de Reitman sur le locus de  $\beta$ -globine de poulet démontrent le besoin de combinaisons des divers éléments de régulation du locus très bien définies pour éviter les effets de position (Chapitre 1, figure 5) (190). De plus, l'expression du gène  $\gamma$ -globine sera indépendante ou non selon les séquences à proximité du gène incluses dans le transgène et la présence ou non du gène  $\beta$ -globine (191,193). L'ajout ou le retrait de nouveaux éléments de régulation pourrait donc nuire ou aider à la formation d'une interaction mutuelle productive. Dans les constructions où il y a des effets de position, des interaction secondaires (par exemple entre un site hypersensible seul et le LCR) pourraient nuire à l'interaction mutuelle entre le gène et le LCR et donc

permettre les effets de position. Selon cette interprétation, en ajoutant ou déplaçant des éléments, la capacité de former une interaction mutuelle fonctionnelle pourrait être réduite.

Afin de tester cette hypothèse, nous avons ajouté un gène de  $\beta$ -globine, complet et intègre (comme il apparaît dans le mini- et le micro-locus), au transgène LCR- $\beta$ -*lacZ* ( $\beta$ B) pour savoir s'il pouvait former une interaction mutuelle fonctionnelle avec le LCR et protéger contre les effets de position. Plusieurs souris fondatrices portant les transgènes  $\beta$ I et  $\beta$ J ont été analysées. Moins de la moitié des lignées exprimaient le transgène et, lorsqu'il y avait expression, celle-ci était relativement faible en comparaison aux globines endogènes (Annexe I, figure 2). Un fait à noter est que l'expression de *lacZ* et de  $\beta$ -globine était relativement proportionnelle dans la même souris. Une grande variation dans l'expression corrigée par le nombre de copie a été observée pour chacun des transgènes. Ces résultats nous indiquent qu'il n'est pas possible pour le LCR d'interagir avec le gène de  $\beta$ -globine afin de prévenir les effets de position lorsque  $\beta$ -*lacZ* se retrouve entre les deux.

#### 5.2.5 Une interaction mutuelle entre le LCR et le gène $\beta$ -globine est compétitionnée par le gène $\beta$ -*lacZ*

Il est donc possible que le modèle d'interaction mutuelle ne soit pas approprié pour expliquer le mécanisme d'action du LCR, toutefois, nous privilégions une autre explication. Il est par exemple possible que le LCR ne puisse interagir correctement avec le gène de  $\beta$ -globine parce qu'il y a compétition du gène  $\beta$ -*lacZ*, une telle compétition pourrait favoriser *lacZ* puisqu'il se trouve plus près du LCR. En effet, la compétition polaire démontre que le LCR active davantage les gènes proximaux (160-162). Si tel était le cas, il pourrait y avoir une interaction entre le LCR et le gène  $\beta$ -*lacZ*, mais non productive, c'est-à-dire incapable de créer les conditions requises pour protéger le transgène contre les effets de position. Cette interaction empêcherait l'entrée en contact du LCR et du gène  $\beta$ -globine.

Le modèle du flip-flop qui suggère que le LCR peut activer un gène de globine, changer de gène cible, puis revenir au premier gène (163), pourrait être impliqué dans le

processus. En effet, il est possible que le LCR tente d'interagir ou d'activer tour à tour les gènes  $\beta$ -*lacZ* et  $\beta$ -globine. Le flip-flop de l'un à l'autre pourrait faire en sorte que chaque gène soit moins activé que s'il avait été seul avec le LCR. Ceci pourrait réduire les niveaux de stimulation mais aussi accentuer les effets de position. En effet, si la formation d'hétérochromatine sur le transgène est bloquée par une interaction entre le LCR et le gène, elle pourrait se produire durant le temps que le LCR passe à interagir avec l'autre gène. En bout de ligne, l'expression pourrait être fortement réduite et même cesser complètement. Ceci pourrait expliquer que les effets de positions sont beaucoup plus graves pour les transgènes où deux gènes rapporteurs sont activés par le LCR par rapport aux transgènes où il n'y en a qu'un seul. Par exemple, les transgènes où un promoteur complet de  $\beta$ -globine dirige *lacZ* sont toujours exprimés ( $\beta$ B à  $\beta$ BE, 37/37) alors que l'ajout d'un autre gène de  $\beta$ -globine rend la moitié des intégrations silencieuses ( $\beta$ I et  $\beta$ J, 5/10).

Une expérience selon laquelle un gène de  $\beta$ -globine complet, intègre et sans compétition pour le LCR pourrait prouver cette hypothèse s'il n'y a pas d'effets de position. En fournissant au LCR un gène avec lequel il peut former une interaction productive sans compétition, nous pourrions peut-être créer un transgène dont l'expression serait indépendante de la position malgré la présence de *lacZ*. Un tel résultat serait une autre preuve que les séquences *lacZ* ne contiennent pas d'éléments qui, intrinsèquement, interfèrent avec l'activité du LCR. Plusieurs constructions peuvent être envisagées. En premier lieu, nous pourrions placer les séquences *lacZ* sans promoteur dans le micro-locus (avant ou après le gène de  $\beta$ -globine). L'absence de promoteur empêcherait la compétition et selon notre modèle, l'expression de  $\beta$ -globine devrait être indépendante de la position. Nous pourrions aussi placer un gène  $\beta$ -*lacZ* après le gène de  $\beta$ -globine dans le micro-locus. Dans ce cas, il y aurait compétition mais *lacZ* serait défavorisé par sa position. Donc si des effets de position sont observés, ils devraient être moins importants que dans  $\beta$ J, indiquant que le modèle de compétition pour l'interaction est valable.

### **5.3 Variégation de l'expression et modèle d'expression graduée**

#### **5.3.1 Il n'y a pas de variégation de l'expression des transgènes $\beta$ -lacZ**

Nous avons observé qu'un certain nombre de cellules ne coloraient pas au X-gal, ce qui pouvait être compatible avec une variégation de l'expression ou, plus particulièrement, un modèle binaire (voir Chapitres 1 et 3). Toutefois, nos données suggèrent que la gradation dans les niveaux d'expression par cellules dans une même lignée provoquent une coloration dans seulement une partie des cellules. Puisque l'expression moyenne par cellule diffère d'une lignée à l'autre, le nombre de cellules détectées variera en fonction de l'expression. Le modèle de l'expression graduée présente une similitude avec la variégation: il y a un phénomène qui induit une baisse d'expression dans certaines cellules. L'expression graduée prévoit que les niveaux d'expression par cellule sont très variés alors que la variégation suggère une distribution bimodale (allumé/éteint). Toutefois, la différence majeure entre les deux modèles provient surtout du fait que l'expression graduée explique comment l'expression détermine le nombre de cellules positives alors que dans la variégation, c'est le nombre de cellules positives qui fixe l'expression. Cette différence fondamentale est très importante car elle a des conséquences sur notre vision des effets de position, à savoir est-ce que les effets de position sont dus à la variégation ou alors est-ce seulement une conséquence des effets de position?

D'autres transgènes sont-ils exprimés de façon graduée? Plusieurs études semblent démontrer la variégation de l'expression de transgènes chez la souris. Dans plusieurs cas, la cytométrie de flux est utilisée (140). Tel que mentionné en introduction, plusieurs autres explications que la variégation peuvent expliquer les résultats. Bien qu'il soit probable que d'autres transgènes génèrent une réelle variégation de l'expression, il est aussi possible que certains auteurs obtiennent une expression graduée mais la confondent avec la variégation de l'expression.

### 5.3.2 Modèle d'expression graduée et mécanisme du LCR

Des auteurs ont suggéré que le LCR du gène CD2 permettait de supprimer la variabilité de l'expression, elle-même générée par des effets de position (194). Dans notre cas, nous avons pu démontrer que les transgènes LCR- $\beta$ -*lacZ*, aussi soumis à des effets de position, avaient plutôt un patron d'expression graduée. Toutefois, il est difficile d'établir un lien clair avec les mécanismes des enhanceurs et du LCR. En effet, aucune construction sans LCR n'a été analysée, on ne peut donc établir son rôle direct dans le processus. De plus, puisque l'activité du LCR qui confère une expression indépendante de la position n'est fonctionnelle dans aucune construction, nous ne pouvons pas non plus comparer le patron d'expression avec et sans effets de position. Il est possible que les effets de position induisent une expression graduée et que le rôle d'un LCR fonctionnel est de permettre une expression à un niveau identique dans toutes les cellules. Toutefois, puisque toutes les lignées démontraient une expression graduée, peu importe la position, cette hypothèse semble improbable. Nous croyons plutôt que le LCR ne doit que fixer l'expression moyenne et qu'il y aura expression graduée, même sans effets de position. Des expériences intéressantes seraient d'étudier des lignées sans LCR et des lignées avec LCR mais qui exprimeraient la  $\beta$ -galactosidase de façon indépendante de la position.

Le modèle proposé par Milot *et al.* selon lequel certains effets de positions sont dus à une réduction de la fréquence et de la durée de stimulation par le LCR (172) apporte ici un éclairage différent. Si ce modèle est le bon, alors l'expression graduée pourrait être due à une différence dans la proportion du temps durant laquelle le LCR stimule le gène dans chaque cellule.

Si ce ne sont pas les effets de position qui induisent l'expression graduée, est-ce que les éléments de régulation du locus de globine sont impliqués ou est-ce un phénomène lié directement au rapporteur? En utilisant les éléments de régulation du locus des gènes de type  $\alpha$ -globine, Sutherland *et al.* (229) ont démontré une expression équivalente de  $\beta$ -galactosidase dans chaque cellule. Il semble donc que le phénomène ne soit pas spécifique à *lacZ*. De plus, d'autres gènes ont une expression graduée. Par exemple, les

colorations par immunofluorescence de globules rouges de souris transgéniques avec des anticorps contre la globine humaine encodée par le transgène (HSS2 murin- $\beta$ -globine, position indépendant et LCR- $\gamma$ -globine, position dépendant) ont démontré une coloration variable d'une cellule à l'autre (191,230). Le phénomène est donc bien réel et pas seulement spécifique à *lacZ*. D'ailleurs, les patrons obtenus en cytométrie de flux pour une foule de gènes ou de transgènes sont très souvent distribués sur une grande région du graphique. Rarement voit-on un pic d'expression très bien défini. Dans des conditions normales, probablement tous les globules rouges adultes expriment la  $\beta$ -globine, il est toutefois possible qu'il y ait une légère variation dans les niveaux par cellule. Donc le rapporteur n'est pas intrinsèquement exprimé de façon graduée et le phénomène semble induit par les séquences de régulation de  $\beta$ -globine.

#### **5.4 Rôle du motif CACC du promoteur de $\beta$ -globine et de EKLF**

Plusieurs études suggèrent que les motifs conservés CACC, CAAT et TATA du promoteur minimal de  $\beta$ -globine jouent un rôle très important dans la régulation transcriptionnelle du gène (129-131). Par exemple, ils sont essentiels pour obtenir des bons niveaux d'expression basale, l'activité promotrice est donc essentiellement contenue dans ces séquences. Ensuite, leur présence est requise pour permettre la stimulation par le LCR (227). L'étude de la transcription dirigée par ces motif et stimulée par le LCR dans divers modèles cellulaires et à divers stades de développement chez les souris transgéniques et des facteurs transcriptionnels qui s'y lient aide à cerner leur rôle dans la régulation du gène de  $\beta$ -globine.

EKLF est un facteur dont la fonction est très intéressante. Il peut interagir avec le motif CACC et stimuler la transcription en transfection cellulaire (216,219,220). L'ablation génétique du gène EKLF chez la souris résulte en une perte d'expression du gène de  $\beta$ -globine et une augmentation de celle des gènes de  $\gamma$ -globine (164,221,222, 224). Il ne semble donc pas y avoir de transition vers les globines adultes. Conséquemment, l'hypothèse que EKLF soit un régulateur majeur de la régulation développementale des gènes de globine a été posée (219).



#### 5.4.1 Dans les cellules K562, le motif CACC permet la transcription stimulée par le LCR

Afin de définir les interactions entre le LCR et le promoteur de  $\beta$ -globine, nous avons caractérisé les éléments requis pour une bonne expression dans les cellules humaines K562 et les souris transgéniques. En transfections stables dans les cellules K562, nous avons analysé l'expression de transgènes où plusieurs délétions du promoteur ont été clonés entre le LCR et le gène rapporteur *H-2K*. Nous avons pu démontrer que le motif CACC, à lui seul, pouvait soutenir la stimulation du LCR dans les cellules K562 (Annexe II).

Quel rôle le motif CACC joue-t-il dans l'expression du gène  $\beta$ -globine dans les cellules K562? Le motif CACC stimule environ 3 fois les niveaux d'expression en transfection dans des cellules non-érythroïdes (HeLa ou 3T6) (129-131). Toutefois, il n'est pas totalement essentiel: il y a toujours expression, bien que plus faible, sans ce motif. Au contraire, dans les cellules K562, le motif CACC est absolument requis pour l'expression (Annexe II). Sa fonction pourrait être différente que celle de simplement stimuler la transcription basale. De plus, les cellules K562 n'expriment pas normalement le gène de  $\beta$ -globine en raison de la régulation développementale. L'expression détectée est donc causée par la stimulation du LCR, créée plus ou moins artificiellement par l'absence de compétition génique (141). Il serait raisonnable de croire que l'activation de la transcription observée correspond à une stimulation provenant du LCR à travers le motif CACC et non pas un simple effet sur la transcription basale. Il y a donc des facteurs liant le motif CACC dans ces cellules qui peuvent soutenir l'activation par le LCR.

Les facteurs liant le motif CACC ont pour la plupart été bien caractérisés (231). Il y a entre autres Sp1, un facteur plus ou moins ubiquitaire liant les motifs G/C, dont le motif CACC fait partie; EKLF, qui est spécifique aux érythrocytes et BKLF, qui est principalement présent dans les cellules hématopoïétiques mais aussi retrouvé dans d'autres types cellulaires; et CAC-A, encore non-identifié. Chez les cellules K562, la caractérisation est moins poussée. Nous savons toutefois que EKLF n'est pas exprimé dans ce type cellulaire (219,232). Ces résultats indiquent que d'autres facteurs que EKLF

peuvent activer la transcription et permettre la stimulation par le LCR à travers le motif CACC. L'identification de ces facteurs serait donc très intéressante.

#### 5.4.2 Le motif CACC est important, mais non essentiel pour l'expression du rapporteur $\beta$ -lacZ chez les souris transgéniques

Nous avons procédé à des études similaires chez les souris transgéniques. Des délétions du promoteur de  $\beta$ -globine ont été analysées dans le contexte de constructions LCR- $\beta$ -lacZ. Les différents promoteurs qui ont été comparés sont le promoteur complet et des promoteurs minimaux incluant les motifs CACC, CAAT et TATA ou des combinaisons de motifs: CACC et TATA ou CAAT et TATA (Annexe II, Figure 5). L'expression de la  $\beta$ -galactosidase chez des souris transgéniques portant ces constructions est soumise à des effets de position, toutefois, les niveaux d'expression moyens nous donnent une idée de la capacité transcriptionnelle de chacun des promoteurs. Nous avons noté que les motifs CAAT et CACC peuvent soutenir des niveaux d'expression environ équivalents et qu'ils semblent avoir un effet additif lorsque présents ensemble. L'ajout des séquences plus distales augmente encore un peu les niveaux obtenus.

Ces résultats nous indiquent comment le LCR peut stimuler les gènes de globine. Des études ont montré que l'expression d'un transgène  $\beta$ -globine (sans LCR) chez les souris était très faible (120-122,135). Si, en plus de l'absence du LCR, les enhancers en 3' du gène sont délétés, alors il n'y a plus d'expression du tout (121). Le promoteur du gène de  $\beta$ -globine seul est donc inactif chez les souris. Les niveaux d'expression du rapporteur lacZ, dirigé par ce même promoteur, sont donc nécessairement dus à la stimulation par le LCR. Nous pouvons donc affirmer que les motifs CAAT, CACC et/ou TATA sont en quelque sorte des vecteurs par lesquels le LCR peut stimuler la transcription. Il y a donc une possibilité que des interactions - probablement par les protéines liées à ces séquences - entre le LCR et les motifs du promoteur minimal de  $\beta$ -globine se produisent dans la chromatine des souris transgéniques.

### 5.4.3 L'expression des transgènes $\beta$ -lacZ est indépendante de EKLF

Chez les souris transgéniques, l'expression de la  $\beta$ -galactosidase dépendait du motif CACC dans la mesure où les niveaux étaient supérieurs lorsqu'il était inclus dans le promoteur. Toutefois, sa présence n'était pas absolument requise: la délétion des séquences CACC ne résultait qu'en une perte d'expression d'environ 50% (transgène  $\beta$ A par rapport au transgène  $\beta$ G). Par contre, l'ablation génétique de EKLF provoque une perte d'expression du gène  $\beta$ -globine endogène de près de 20 fois (221,222). Il semble paradoxal que la délétion du site de liaison d'un facteur aussi important pour l'expression n'ait pas d'effet plus drastique. Nous avons donc supposé qu'EKLF pouvait avoir un effet ailleurs qu'au motif CACC. Il aurait par exemple été possible que EKLF lie le LCR et parvienne ainsi à activer les transgènes LCR- $\beta$ -lacZ. Afin de tester cette hypothèse, nous avons croisé certaines des lignées transgéniques avec des souris knock out pour EKLF. Contre toute attente, l'activité  $\beta$ -galactosidase n'a pas été réduite chez les embryons dépourvus de EKLF. Au contraire, une augmentation plus ou moins substantielle (allant jusqu'à 7 fois) a été observée (Annexe II, Tableau II). L'intensité de la stimulation ne semble pas liée à la présence du motif CACC et varie dans les lignées où le motif CACC est présent et dans les lignées où il est absent. Nous concluons donc que l'expression des transgènes LCR- $\beta$ -lacZ ne requiert en aucune façon EKLF. Ce facteur ne joue donc aucun rôle, du moins détectable, dans la stimulation de la transcription en liant le LCR. De la même façon, il ne stimule pas la transcription des transgènes en se liant au motif CACC du promoteur de  $\beta$ -globine.

Comment expliquer que l'expression de ces transgènes soit supérieure en absence de EKLF? Ce facteur de transcription pourrait carrément être un répresseur. Toutefois, EKLF pourrait aussi être un activateur mais tout simplement moins efficace que d'autres protéines liant le même motif (par exemple Sp1 ou BKLF). EKLF serait alors un compétiteur agissant en agoniste et sa perte permettrait une meilleure stimulation par les autres facteurs. Cependant, cette hypothèse n'expliquerait pas pourquoi les lignées  $\beta$ A, dont le promoteur ne contient pas le motif CACC, sont elles aussi mieux exprimées dans

les souris knock out pour EKLF. Il semble donc plus probable qu'un effet indirect soit en cause. Par exemple, le transgène pourrait être en compétition avec le locus de  $\beta$ -globine murine endogène pour des facteurs de transcription limitants. Dans les knock out, puisque ce locus n'exprime pas la globine  $\beta$ -major murine, ces facteurs deviendraient disponibles pour le transgène. Une autre possibilité est que ces animaux, fortement anémiques, tentent de stimuler l'érythropoïèse pour remédier à la situation. De tels signaux existent réellement lors de perte de sang importante par exemple. Ces signaux pourraient influencer l'expression des transgènes.

#### 5.4.4 EKLF pourrait activer le gène $\beta$ -globine par d'autres éléments que le motif CACC

Puisque nous ne détectons pas de stimulation par EKLF par le LCR ou le promoteur de  $\beta$ -globine dans nos souris transgéniques, EKLF pourrait stimuler l'expression de la  $\beta$ -globine par d'autres éléments. Il est donc possible que le rôle supposé de EKLF, soit d'activer la transcription du gène de  $\beta$ -globine en se liant au motif CACC, ne corresponde pas à la réalité. Or, la seule différence entre ce transgène et  $\beta$ B (LCR- $\beta$ -*lacZ*) est la séquence codante ( $\beta$ -globine versus *lacZ*). EKLF pourrait par exemple activer la  $\beta$ -globine par les enhancers 3' présents dans le micro-locus mais non dans  $\beta$ B. D'ailleurs, ces enhancers comportent des motifs de type CACC qui sont liés par des protéines. La liaison de EKLF dans cette région n'a, à notre connaissance, pas encore été testée. Il est aussi possible que EKLF active la transcription par le motif CACC mais seulement si un autre élément, celui-là localisé dans les séquences codantes, est présent.

Une autre possibilité est que nous observons ici un artefact créé par l'expression aberrante de la  $\beta$ -galactosidase. En effet, les transgènes LCR- $\beta$ -*lacZ* sont soumis à d'importants effets de position, qui pourraient annuler la stimulation de EKLF. Dans de telles conditions, la présence ou l'absence du facteur n'aurait pas d'effet important sur l'expression du transgène. L'étude de la conséquence du knock out de EKLF sur l'expression de transgènes utilisant les séquences codantes pour la  $\beta$ -globine comme rapporteur, et non pas *lacZ*, pourrait confirmer l'une ou l'autre de ces hypothèses.

#### 5.4.5 LCR pourrait stimuler le gène le plus actif transcriptionnellement

Au Chapitre 4, nous démontrons que EKLF, contrairement à ce que l'on aurait pu supposer à partir des expériences antérieures, n'induit pas la transition entre les gènes de  $\gamma$ -globine et le gène de  $\beta$ -globine. La transition pourrait plutôt avoir lieu lors de la répression partielle des gènes de  $\gamma$ -globine (152), sans toutefois se produire si le gène de  $\beta$ -globine n'est pas au préalable activé par EKLF. Malheureusement, ces expériences ne permettent pas de conclure sur le type d'interaction entre le LCR et les gènes. Par exemple, ces résultats seraient compatibles avec un mécanisme où le LCR choisirait d'interagir avec le gène le plus actif à proximité, sans égard à l'affinité d'interaction avec des protéines spécifiques liées au LCR et aux gènes. D'ailleurs, le fait qu'un gène *neo*, dirigé par aucun élément de régulation érythroïde, peut efficacement compétitionner tous les gènes de globine lorsqu'inséré à proximité du LCR (155-157) nous indique que l'activation ne semble pas faire intervenir des protéines très spécifiques. Le LCR pourrait par exemple directement stimuler directement l'ARN polymérase II.

#### 5.4.6 Différents mécanismes gouvernent l'activité de stimulation de la transcription et l'ouverture de la chromatine

L'hypothèse que le LCR peut stimuler la transcription d'un gène sans interaction spécifique n'est pas incompatible avec celle d'une interaction mutuelle protégeant contre les effets de position. Il est possible que ces deux interactions soient distinctes et indépendantes. a) Une interaction mutuelle entre le LCR et le gène qu'il active pourrait avoir lieu à un moment précis du développement et la conséquence de cette interaction serait de prévenir les effets de position. La structure de la chromatine serait alors plus ou moins fixée pour le reste du développement (avec peut-être une répression qui apparaîtrait graduellement). Cette interaction devrait être très spécifique puisque l'insertion de certains éléments de régulation dans le transgène pourrait empêcher sa formation. Ce modèle est compatible avec les effets de position causés par l'insertion de *lacZ* dans un transgène de  $\beta$ -globine (Chapitre 1 et annexe I) et avec les dispositions d'éléments de régulation bien précises requises pour l'indépendance de la position des divers transgènes du locus de  $\beta$ -globine de poulet ou de  $\gamma$ -globine humaine (Chapitre 1, figures 7 et 8). Il explique aussi

pourquoi deux gènes ( $\beta$ -*lacZ* et  $\beta$ -globine) inclus dans le même transgène sont stimulés de façon plus ou moins proportionnelle. Enfin, il pourrait expliquer comment un certain silencing se produit alors que la souris vieillit ( et données non montrées). b) Par la suite, il pourrait y avoir une activation de la transcription du gène le plus actif de la région environnante. Puisque les enhancers stimulent à peu près n'importe quel gène, il ne doit pas y avoir d'interaction très spécifique. Une hypothèse souvent rencontrée est que ces éléments recrutent des facteurs de la machinerie transcriptionnelle générale, qui est à peu près commune à tous les gènes transcrits par l'ADN polymérase II. Ce modèle global pourrait réconcilier tous les résultats présentés dans cette thèse, et plusieurs autres déjà publiés.

### Références (Chapitres 1 et 5 et Annexes I et II)

1. Charron, G., Guy, L.-G., Bazinet, M. et Julien, J.-P. (1995) Multiple neuron-specific enhancers in the gene coding for the human neurofilament light chain. *J. Biol. Chem.*, **270**: 30604-10.
2. Lewin, B. (1994) *Genes V*, Oxford University Press, New York.
3. Volloch, V. et Housman, D. (1981) Stability of globin mRNA in terminally differentiating murine erythroleukemia cells. *Cell*, **23**: 509-14.
4. Papavassiliou, A. G., Treier, M., Chavrier, C. et Bohmann, D. (1992) Targeted degradation of c-Fos, but not v-Fos, by a phosphorylation-dependent signal on c-Jun. *Science*, **258**: 1941-4.
5. Lin, B. T. et Wang, J. Y. (1992) Cell cycle regulation of retinoblastoma protein phosphorylation. *Ciba Found. Symp.*, **170**: 227-41.
6. Verma, I. M., Stevenson, J. K., Schwarz, E. M., Van Antwerp, D. et Miyamoto, S. (1995) Rel/NF- $\kappa$ B/I  $\kappa$  B family: intimate tales of association and dissociation. *Genes Dev.*, **9**: 2723-35.
7. Molto, M. D., Pascual, L., Martinez-Sebastian, M. J. et de Frutos, R. (1992) Genetic analysis of heat shock response in three *Drosophila* species of the obscura group. *Genome*, **35**: 870-80.
8. Kornberg, R. D. (1977) The nucleosome. *Ann. Rev. Biochem.*, **46**: 931-954.
9. Felsenfeld, G. (1992) Chromatin as an essential part of the transcriptional mechanism. *Nature*, **355**: 219-224.
10. Felsenfeld, G., Boyes, J., Chung, J., Clark, D. et Studitsky, V. (1996) Chromatin structure and gene expression. *Proc. Natl Acad. Sci. USA*, **93**: 9384-8.
11. Miller, O. L., Jr. et Bakken, A. H. (1972) Morphological studies of transcription. *Acta Endocrinol. Suppl.*, **168**: 155-77.
12. Olins, A. L. et Olins, D. E. (1974) Spheroid chromatin units. *Science*, **183**: 330-2.
13. Finch, J. T., Noll, M. et Kornberg, R. D. (1975) Electron microscopy of defined lengths of chromatin. *Proc. Natl Acad. Sci. USA*, **72**: 3320-2.
14. Tsanev, R., Russev, G., Pashev, I. et Zlatanova, J. (1992) *Replication and transcription of chromatin*, CRC Press, Inc., Boca Raton, FL.
15. van Holde, K. E. (1988) *Chromatin*, Springer Verlag, New York.

16. Thoma, F., Koller, T. et Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.*, **83**: 403-27.
17. Rattner, J. B., Saunders, C., Davie, J. R. et Hamkalo, B. A. (1982) Ultrastructural organization of yeast chromatin. *J. Cell Biol.*, **93**: 217-22.
18. van Holde, K. et Zlatanova, J. (1995) Chromatin higher order structure: chasing a mirage? *J. Biol. Chem.*, **270**: 8373-6.
19. Lorch, Y., LaPointe, J. W. et Kornberg, R. D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, **49**: 203-10.
20. Losa, R. et Brown, D. D. (1987) A bacteriophage RNA polymerase transcribes in vitro through a nucleosome core without displacing it. *Cell*, **50**: 801-8.
21. Knezetic, J. A. et Luse, D. S. (1986) The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, **45**: 95-104.
22. Emerson, B. M. et Felsenfeld, G. (1984) Specific factor conferring nuclease hypersensitivity at the 5' end of the chicken adult  $\beta$ -globin gene. *Proc. Natl Acad. Sci. USA*, **81**: 95-9.
23. Kornberg, R. D. (1992) Chromatin structure and transcription. *Ann. Rev. Cell Biol.*, **8**: 563-587.
24. Lewin, B. (1994) Chromatin and gene expression: constant questions, but changing answers. *Cell*, **79**: 397-406.
25. Shen, X., Yu, L., Weir, J. W. et Gorovsky, M. A. (1995) Linker histones are not essential and affect chromatin condensation in vivo. *Cell*, **82**: 47-56.
26. Waters, S. et von der Decken, A. (1992) Tissue-specific sensitivity of chromatin and the vitellogenin gene to micrococcal nuclease after continuous exposure of salmon (*Salmo salar*) to 17  $\beta$ -estradiol. *Gen. Comp. Endocrinol.*, **87**: 105-10.
27. Sheffery, M., Rifkind, R. A. et Marks, P. A. (1982) Murine erythroleukemia cell differentiation: DNase I hypersensitivity and DNA methylation near the globin genes. *Proc. Natl Acad. Sci. USA*, **79**: 1180-4.
28. Jimenez, G., Griffiths, S. D., Ford, A. M., Greaves, M. F. et Enver, T. (1992) Activation of the  $\beta$ -globin locus control region precedes commitment to the erythroid lineage. *Proc. Natl Acad. Sci. USA*, **89**: 10618-22.
- 29a. Schmid, A., Fascher, K. D. et Horz, W. (1992) Nucleosome disruption at the yeast PHO5 promoter upon PHO5 induction occurs in the absence of DNA replication. *Cell*, **71**: 853-64.



- 29b. McPherson, C. F., Horowitz, R., Woodcock, C. L. Jian, C. et Zaret, K. S. (1996) Nucleosome positioning properties of the albumin transcriptional enhancer. *Nucl. Acids Res.*, **24**: 397-404.
30. Laemmli, U. K., Kas, E., Poljak, L. et Adachi, Y. (1992) Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. *Curr. Opin. Genet. Dev.*, **2**: 275-85.
- 31a. Hogan, B., Costantini, F. et Lacy, E. (1986) Manipulating the mouse embryo, Cold Spring Harbor Laboratory, New York.
- 31b. Hazelrigg, T., Lewis, R. et Robin, G. M. (1984) Transformation of white locus DNA in drosophila: dosage compensation, zeste interaction and position effects. *Cell*, **36**: 469-81.
32. Palmiter, R. D. et Brinster, R. L. (1986) Germ-line transformation of mice. *Annu. Rev. Genet.*, **20**: 465-499.
33. Tartof, K. D. (1994) Position effect variegation in yeast. *BioEssays*, **16**: 713-714.
34. Henikoff, S. (1992) Position effect and related phenomena. *Curr. Opin. Genet. Dev.*, **2**: 907-912.
35. Locke, J., Kotarski, M. A. et Tartof, K. D. (1988) Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics*, **120**: 181-98.
36. Moore, G. D., Sinclair, D. A. et Grigliatti, R. A. (1983) Histone gene multiplicity and position effect variegation in *drosophila melanogaster*. *Genetics*, **105**: 327.
37. Dorn, R., Heymann, S., Lindigkeit, R. et Reuter, G. (1986) Suppressor mutation of position-effect variegation in *drosophila melanogaster* affecting chromatin properties. *Chromosoma*, **93**: 398-403.
- 38a. Muller, H. J. (1930) Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.*, **22**: 299-334.
- 38b. Levis, R., Hazelrigg, T. et Rubin, G. M. (1985) Separable cis-acting elements for expression of the white gene of *drosophila*. *EMBO J.*, **4**: 3489-3499.
39. Dobie, K. W., Lee, M., Fantes, A. A., Graham, E., Clark, J. A., Springbett, A., Lathe, R. et McClenaghan, M. (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc. Natl Acad. Sci. USA*, **93**: 6659-6664.
40. Elliott, J. I., Festenstein, R., Tolaini, M. et Kioussis, D. (1995) Random activation of a transgene under the control of a hybrid hCD2 locus control region/Ig enhancer regulatory element. *EMBO J.*, **14**: 575-584.

41. Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. et Whitelaw, E. (1995) Position-dependent variegation of globin transgene expression in mice. *Proc. Natl Acad. Sci. USA*, **92**: 5371-5375.
42. Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. et Martin, D. I. K. (1995) Enhancers increase the probability but not the level of gene expression. *Proc. Natl Acad. Sci. USA*, **92**: 7125-7129.
43. Gross, D. S. et Garrad, W. T. (1988) Nuclease-hypersensitive sites in chromatin. *Ann. Rev. Biochem.*, **57**: 159-197.
44. May, G. et Enver, T. (1995) Targeting gene expression to haemopoietic stem cells: a chromatin-dependent upstream element mediates cell type-specific expression of the stem cell antigen CD34. *EMBO J.*, **14**: 564-74.
45. Campaign, J. A. et Cox, G. S. (1992) Deoxyribonuclease-hypersensitive sites in the glycoprotein hormone  $\alpha$ -subunit gene from trophoblastic and nontrophoblastic human tumor cell lines: correlation with expression and effect of chemical inducers. *Mol. Endocrinol.*, **6**: 677-93.
46. Stamatoyannopoulos, J. A., Goodwin, A., Joyce, T. et Lowrey, C. H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human  $\beta$ -globin locus control region. *EMBO J.*, **14**: 106-16.
47. Lowrey, C. H., Bodine, D. M. et Nienhuis, A. W. (1992) Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc. Natl Acad. Sci. USA*, **89**: 1143-7.
48. Choder, M., Bratosin, S. et Aloni, Y. (1984) A direct analysis of transcribed minichromosomes: all transcribed SV40 minichromosomes have a nuclease-hypersensitive region within a nucleosome-free domain. *EMBO J.*, **3**: 2929-36.
49. Forrester, W. C., van Genderen, C., Jenuwein, T. et Grosschedl, R. (1994) Dependence of enhancer-mediated transcription of the immunoglobulin m gene on nuclear matrix attachment regions. *Science*, **265**: 1221-5.
50. Neznanov, N., Thorey, I. S., Cecena, G. et Oshima, R. G. (1993) Transcriptional insulation of the human keratin 18 gene in transgenic mice. *Mol. Cell. Biol.*, **13**: 2214-23.
51. Stief, A., Winter, D. M., Stratling, W. H. et Sippel, A. E. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature*, **341**: 343-5.
52. Bonifer, C., Yannoutsos, N., Kruger, G., Grosveld, F. et Sippel, A. E. (1994) Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucl. Acids Res.*, **22**: 4202-10.

53. De Sepulveda, P., Salaun, P., Maas, N., Andre, C. et Panthier, J. J. (1995) SARs do not impair position-dependent expression of a *kit/lacZ* transgene. *Bioch. Bioph. Res. Comm.*, **211**: 735-41.
54. Vanyushin, B. F., Tkacheva, S. G. et Belozersky, A. N. (1970) Rare bases in animal DNA. *Nature*, **225**: 948-9.
55. Doscocil, J. et Sorm, F. (1962) Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim. Biophys. Acta*, **55**: 953.
56. Meehan, R., Lewis, J., Cross, S., Nan, X., Jeppesen, P. et Bird, A. (1992) Transcriptional repression by methylation of CpG. *J. Cell Sci. Supp.*, **16**: 9-14.
57. Sutter, D., Westphal, M. et Doerfler, W. (1978) Patterns of integration of viral DNA sequences in the genomes of adenovirus type 12-transformed hamster cells. *Cell*, **14**: 569-85.
58. Ramchandani, S., MacLeod, A. R., Pinard, M., von Hofe, E. et Szyf, M. (1997) Inhibition of tumorigenesis by a cytosine-DNA, methyltransferase, antisense oligodeoxynucleotide. *Proc. Natl Acad. Sci. USA*, **94**: 684-9.
59. Szyf, M., Theberge, J. et Bozovic, V. (1995) Ras induces a general DNA demethylation activity in mouse embryonal P19 cells. *J. Biol. Chem.*, **270**: 12690-6.
60. Jones, P. A., Wolkowicz, M. J., Rideout, W. M. d., Gonzales, F. A., Marziasz, C. M., Coetzee, G. A. et Tapscott, S. J. (1990) De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. *Proc. Natl Acad. Sci. USA*, **87**: 6117-21.
61. Mandel, J. L. et Chambon, P. (1979) DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucl. Acids Res.*, **7**: 2081-103.
62. Inamdar, N. M., Ehrlich, K. C. et Ehrlich, M. (1991) CpG methylation inhibits binding of several sequence-specific DNA-binding proteins from pea, wheat, soybean and cauliflower. *Plant Mol. Biolo.*, **17**: 111-23.
63. Bird, A. (1992) The essentials of DNA methylation. *Cell*, **70**: 5-8.
64. Turker, M. S. et Bestor, T. H. (1997) Formation of methylation patterns in the mammalian genome. *Mutation Res.*, **386**: 119-30.
65. Roth, S. Y. et Allis, C. D. (1992) Chromatin condensation: does histone H1 dephosphorylation play a role? *Trends Biochem. Sci.*, **17**: 93-8.
66. Turner, B. M. (1991) Histone acetylation and control of gene expression. *J. Cell Sci.*, **99**: 13-20.

67. Vidali, G., Boffa, L. C., Bradbury, E. M. et Allfrey, V. G. (1978) Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. *Proc. Natl Acad. Sci. USA*, **75**: 2239-43.
68. Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane- Robinson, C., Allis, C. D. et Workman, J. L. (1996) Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J.*, **15**: 2508-18.
69. Rhodes, D. (1985) Structural analysis of a triple complex between the histone octamer, a *Xenopus* gene for 5S RNA and transcription factor IIIA. *EMBO J.*, **4**: 3473-82.
70. Muller, S., Erard, M., Burggraf, E., Couppez, M., Sautiere, P., Champagne, M. et Van Regenmortel, M. H. (1982) Immunochemical detection of changes in chromatin subunits induced by histone H4 acetylation. *EMBO J.*, **1**: 939-44.
71. Hebbes, T. R., Thorne, A. W. et Crane-Robinson, C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.*, **7**: 1395-402.
72. Wade, P. A., Pruss, D. et Wolffe, A. P. (1997) Histone acetylation: chromatin in action. *Trends Bioch. Sci.*, **22**: 128-32.
73. Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K. et Gottschling, D. E. (1993) Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev.*, **7**: 1133-45.
74. Moretti, P., Freeman, K., Coodly, L. et Shore, D. (1994) Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.*, **8**: 2257-69.
75. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. et Grunstein, M. (1995) Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell*, **80**: 583-92.
76. Latham, K. E. (1996) X chromosome imprinting and inactivation in the early mammalian embryo. *Trends Genet.*, **12**: 134-138.
77. Pirota, V. (1995) Chromatin complexes regulating gene expression in *Drosophila*. *Curr. Opin. Genet. Dev.*, **5**: 466-472.
78. Peterson, C. L., Kruger, W. et Herskowitz, I. (1991) A functional interaction between the C-terminal domain of RNA polymerase II and the negative regulator SIN1. *Cell*, **64**: 1135-43.

79. Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. et Green, M. R. (1994) Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature*, **370**: 477-81.
80. Côté, J., Quinn, J., Workman, J. L. et Peterson, C. L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, **265**: 53-60.
81. Eissenberg, J. C. et Elgin, S. C. (1991) Boundary functions in the control of gene expression. *Trends Genet.*, **7**: 335-40.
82. Kellum, R. et Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol.*, **12**: 2424-31.
83. Roseman, R. R., Pirrotta, V. et Geyer, P. K. (1993) The su(Hw) protein insulates expression of the *Drosophila melanogaster* white gene from chromosomal position-effects. *EMBO J.*, **12**: 435-42.
84. Cai, H. et Levine, M. (1995) Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature*, **376**: 533-6.
85. Spana, C., Harrison, D. A. et Corces, V. G. (1988) The *Drosophila melanogaster* suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes Dev.*, **2**: 1414-23.
86. Udvardy, A., Maine, E. et Schedl, P. (1985) 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.*, **185**: 341-58.
87. Li, Q. et Stamatoyannopoulos, G. (1994) Hypersensitive site 5 of the human  $\beta$  locus control region functions as a chromatin insulator. *Blood*, **84**: 1399-401.
88. Chung, J. H., Whiteley, M. et Felsenfeld, G. (1993) 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*, **74**: 505-14.
89. Crossley, M. et Orkin, S. H. (1993) Regulation of the  $\beta$ -globin locus. *Curr. Opin. Genet. Dev.*, **3**: 232-7.
90. Epner, E., Kim, C. G. et Groudine, M. (1992) What does the locus control region control? *Curr. Biol.*, **2**: 262-4.
91. Townes, T. M. et Behringer, R. R. (1990) Human globin locus activation region (LAR): role in temporal control. *Trends Genet.*, **6**: 219-23.
92. Diaz, P., Cado, D. et Winoto, A. (1994) A locus control region in the T cell receptor  $\alpha/\delta$  locus. *Immunity*, **1**: 207-17.

93. Chamberlain, J. W., Vasavada, H. A., Ganguly, S. et Weissman, S. M. (1991) Identification of cis sequences controlling efficient position-independent tissue-specific expression of human major histocompatibility complex class I genes in transgenic mice. *Mol. Cell. Biol.*, **11**: 3564-72.
94. Carson, S. et Wiles, M. V. (1993) Far upstream regions of class II MHC Ea are necessary for position-independent, copy-dependent expression of Ea transgene. *Nucl. Acids Res.*, **21**: 2065-72.
95. Aronow, B. J., Silbiger, R. N., Dusing, M. R., Stock, J. L., Yager, K. L., Potter, S. S., Hutton, J. J. et Wiginton, D. A. (1992) Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position-independent transgene expression. *Mol. Cell. Biol.*, **12**: 4170-85.
96. Jones, B. K., Monks, B. R., Liebhaber, S. A. et Cooke, N. E. (1995) The human growth hormone gene is regulated by a multicomponent locus control region. *Mol. Cell. Biol.*, **15**: 7010-21.
97. Ganss, R., Montoliu, L., Monaghan, A. P. et Schutz, G. (1994) A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. *EMBO J.*, **13**: 3083-93.
98. Knotts, S., Rindt, H. et Robbins, J. (1995) Position independent expression and developmental regulation is directed by the  $\beta$  myosin heavy chain gene's 5' upstream region in transgenic mice. *Nucl. Acids Res.*, **23**: 3301-9.
99. Friend, W. C., Clapoff, S., Landry, C., Becker, L. E., O'Hanlon, D., Allore, R. J., Brown, I. R., Marks, A., Roder, J. et Dunn, R. J. (1992) Cell-specific expression of high levels of human S100  $\beta$  in transgenic mouse brain is dependent on gene dosage. *J. Neurosci.*, **12**: 4337-46.
100. Cheyette, T. E., Ip, T., Faber, S., Matsui, Y. et Chalkley, R. (1992) Characterization of the factors binding to a PEPCK gene upstream hypersensitive site with LCR activity. *Nucl. Acids Res.*, **20**: 3427-33.
101. Talbot, D., Descombes, P. et Schibler, U. (1994) The 5' flanking region of the rat LAP (C/EBP  $\beta$ ) gene can direct high-level, position-independent, copy number-dependent expression in multiple tissues in transgenic mice. *Nucl. Acids Res.*, **22**: 556-66.
102. Madisen, L., Groudine, M. (1994) Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-myc expression in plasmacytoma and Burkitt's lymphoma cells. *Genes Dev.*, **8**: 2212-26.

103. Greaves, D. R., Wilson, F. D., Lang, G. et Kioussis, D. (1989) Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell*, **56**: 979-86.
104. Lang, G., Mamalaki, C., Greenberg, D., Yannoutsos, N. et Kioussis, D. (1991) Deletion analysis of the human CD2 gene locus control region in transgenic mice. *Nucl. Acids Res.*, **19**: 5851-6.
105. Schutz, G., Nguyen-Huu, M. C., Giesecke, K., Hynes, N. E., Groner, B., Wurtz, T. et Sippel, A. E. (1978) Hormonal control of egg white protein messenger RNA synthesis in the chicken oviduct. *Cold Spring Harbor Symposia Quanti.Biol.*, **42**: 617-24.
106. Bonifer, C., Jäggle, U. et Hubert, M. C. (1997) The chicken lysozyme locus control as a paradigm for the complex developmental regulation of eukaryotic gene loci. *J. Biol. Chem.*, **272**: 26075-26078.
107. Bonifer, C., Vidal, M., Grosveld, F., Sippel, A. E. (1990) Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.*, **9**: 2843-8.
108. Maniatis, T., Fritsch, E. F., Lauer, J. et Lawn, R. M. (1980) The molecular genetics of human hemoglobins. *Annu. Rev. Genet.*, **14**: 145-78.
109. Orkin, S. H. (1990) Globin gene regulation and switching: circa 1990. *Cell*, **63**: 665-72.
110. Stamatoyannopoulos, G. (1991) Human hemoglobin switching. *Science*, **252**.
111. Stamatoyannopoulos, G. et Nienhuis, A. W. (1994) In Stamatoyannopoulos G, N. A., Majerus PW and Varmus H (ed.), *The Molecular Basis of Blood Diseases*. Saunders Co., Philadelphia, PA, pp. 107-136.
112. Huisman, T. H., Kutlar, F. et Gu, L. H. (1991)  $\gamma$  chain abnormalities and  $\gamma$ -globin gene rearrangements in newborn babies of various populations. *Hemoglobin*, **15**: 349-79.
113. Orkin, S. H., Kazazian, H. H., Jr., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G. et Giardina, P. J. (1982) Linkage of  $\beta$ -thalassaemia mutations and  $\beta$ -globin gene polymorphisms with DNA polymorphisms in human  $\beta$ -globin gene cluster. *Nature*, **296**: 627-31.
114. Bollekens, J. A. et Forget, B. G. (1991) Delta  $\beta$  thalassemia and hereditary persistence of fetal hemoglobin. *Hematol.Oncol. Clinics North America*, **5**: 399-422.

115. Marks, P. A., Sheffery, M. et Rifkind, R. A. (1987) Induction of transformed cells to terminal differentiation and the modulation of gene expression. *Cancer Res.*, **47**: 659-66.
116. Lozzio, C. B. et Lozzio, B. B. (1975) Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*, **45**: 321-34.
117. Tetteroo, P. A., Massaro, F., Mulder, A., Schreuder-van Gelder, R. et von dem Borne, A. E. (1984) Megakaryoblastic differentiation of proerythroblastic K562 cell-line cells. *Leukemia Res.*, **8**: 197-206.
118. Dzierzak, E. et Medvinsky, A. (1995) Mouse embryonic hematopoiesis. *Trends Genet.*, **11**: 359-366.
119. Chada, K., Magram, J. et Costantini, F. (1985) An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature*, **319**: 685-9.
120. Kollias, G., Wrighton, N., Hurst, J. et Grosveld, F. (1986) Regulated expression of human A  $\gamma$ -,  $\beta$ -, and hybrid  $\gamma\beta$ -globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell*, **46**: 89-94.
121. Behringer, R. R., Hammer, R. E., Brinster, R. L., Palmiter, R. D. et Townes, T. M. (1987) Two 3' sequences direct adult erythroid-specific expression of human  $\beta$ -globin genes in transgenic mice. *Proc. Natl Acad. Sci. USA*, **84**: 7056-60.
122. Trudel, M. et Costantini, F. (1987) A 3' enhancer contributes to the stage-specific expression of the human  $\beta$ -globin gene. *Genes Dev.*, **1**: 954-61.
123. Martin, D. I. et Whitelaw, E. (1996) The vagaries of variegating transgenes. *Bioessays*, **18**: 919-23.
124. Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G. et Papayannopoulou, T. (1983) Human fetal to adult hemoglobin switching: changes in chromatin structure of the  $\beta$ -globin gene locus. *Proc. Natl Acad. Sci. USA*, **80**: 7551-7555.
125. Forrester, W. C., Thompson, C., Elder, J. T. et Groudine, M. (1986) A developmentally stable chromatin structure in the human  $\beta$ -globin gene cluster. *Proc. Natl Acad. Sci. USA*, **83**: 1359-63.
126. Tuan, D., Solomon, W., Li, Q. et London, I. M. (1985) The " $\beta$ -like-globin" gene domain in human erythroid cells. *Proc. Natl Acad. Sci. USA*, **82**: 6384-8.
127. Dhar, V., Nandi, A., Schildkraut, C. L., Skoultschi and A.I. (1990) Erythroid-specific nuclease-hypersensitive sites flanking the human  $\beta$ -globin domain. *Mol. Cell.Biol.*, **10**: 4324.



128. Efstratiadis, A. (1980) The structure and evolution of the human  $\beta$ -globin gene family. *Cell*, **21**: 653-668.
129. Charnay, P., Mellon, P. et Maniatis, T. (1985) Linker scanning mutagenesis of the 5'-flanking region of the mouse  $\beta$ -major-globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. *Mol. Cell. Biol.*, **5**: 1498-511.
130. Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. et Weissmann, C. (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit  $\beta$ -globin gene in mouse 3T6 cells. *Cell*, **32**: 695-706.
131. Myers, R. M., Tilly, K. et Maniatis, T. (1986) Fine structure genetic analysis of a  $\beta$ -globin promoter. *Science*, **232**: 613-8.
132. Kulozik, A. E., Bellan-Koch, A., Bail, S., Kohne, E. et Kleihauer, E. (1991) Thalassaemia intermedia: moderate reduction of  $\beta$  globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood*, **77**: 2054-8.
133. Antoniou, M., deBoer, E., Habets, G. et Grosveld, F. (1988) The human  $\beta$ -globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *EMBO J.*, **7**: 377-84.
134. Magram, J., Niederreither, K. et Costantini, F. (1989)  $\beta$ -globin enhancers target expression of a heterologous gene to erythroid tissues of transgenic mice. *Mol. Cell. Biol.*, **9**: 4581-4.
135. Kollias, G., Hurst, J., deBoer, E. et Grosveld, F. (1987) The human  $\beta$ -globin gene contains a downstream developmental specific enhancer. *Nucl. Acids Res.*, **15**: 5739-47.
136. Bodine, D. M. et Ley, T. J. (1987) An enhancer element lies 3' to the human A  $\gamma$  globin gene. *EMBO J.*, **6**: 2997-3004.
137. McDonagh, K. T., Lin, H. J., Lowrey, C. H., Bodine, D. M. et Nienhuis, A. W. (1991) The upstream region of the human  $\gamma$ -globin gene promoter. Identification and functional analysis of nuclear protein binding sites. *J. Biol. Chem.*, **266**: 11965-74.
138. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. et Grosveld, F. G. (1983)  $\beta$ -globin gene inactivation by DNA translocation in  $\gamma$   $\beta$ -thalassaemia. *Nature*, **306**: 662-6.
139. Forrester, W. C., Novak, U., Gelinis, R. et Groudine, M. (1989) Molecular analysis of the human  $\beta$ -globin locus activation region. *Proc. Natl Acad. Sci. USA*, **86**: 5439-43.

140. Grosveld, F., van Assendelft, G. B., Greaves, D. R. et Kollias, G. (1987) Position-independent, high-level expression of the human  $\beta$ -globin gene in transgenic mice. *Cell*, **51**: 975-85.
141. Blom van Assendelft, G., Hanscombe, O., Grosveld, F. et Greaves, D. R. (1989) The  $\beta$ -globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, **56**: 969-77.
142. Tuan, D. Y., Solomon, W. B., London, I. M. et Lee, D. P. (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human " $\beta$ -like globin" genes. *Proc. Natl Acad. Sci. USA*, **86**: 2554-2558.
143. Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D. R. et Grosveld, F. (1989) High-level, erythroid-specific expression of the human  $\alpha$ -globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes. *Genes Dev.*, **3**: 1572-81.
144. Bonnerot, C. et Nicola, J.-F. (1993) Interactions d'un transgène dépendant du site d'intégration et d'un élément de contrôle de la région du locus de la  $\beta$ -globine. *C. R. Acad. Sci. Paris, Sciences de la vie*, **316**: 352-357.
145. Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. et Philipson, S. (1996) The human  $\beta$ -globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development*, **122**: 3991-9.
146. Teitz, T., Chang, J. C., Kitamura, M., Yen, T. S. et Kan, Y. W. (1993) Rhabdomyosarcoma arising in transgenic mice harboring the  $\beta$ -globin locus control region fused with simian virus 40 large T antigen gene. *Proc. Natl Acad. Sci. USA*, **90**: 2910-4.
147. Ryan, T. M., Behringer, R. R., Townes, T. M., Palmiter, R. D. et Brinster, R. L. (1989) High-level erythroid expression of human  $\alpha$ -globin genes in transgenic mice. *Proc. Natl Acad. Sci. USA*, **86**: 37-41.
148. Andi, A. K., Roginski, R. S., Gregg, R. G., Smithies, O. et Skoultchi, A. I. (1988) Regulated expression of genes inserted at the human chromosomal  $\beta$ -globin locus by homologous recombination. *Proc. Natl Acad. Sci. USA*, **85**: 3845-3849.
149. Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Costantini, F. et Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature*, **344**: 309-13.
150. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. et Townes, T. M. (1990) Human  $\gamma$ - to  $\beta$ -globin gene switching in transgenic mice. *Genes Dev.*, **4**: 380-389.

151. Enver, T., Ebens, A. J., Forrester, W. C. et Stamatoyannopoulos, G. (1989) The human  $\beta$ -globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice. *Proc. Natl Acad. Sci. USA*, **86**: 7033-7.
152. Dillon, N. et Grosveld, F. (1991) Human  $\gamma$ -globin genes silenced independently of other genes in the  $\beta$ -globin locus. *Nature*, **350**: 252-4.
153. Strouboulis, J., Dillon, N. et Grosveld, F. (1992) Developmental regulation of a complete 70-kb human  $\beta$ -globin locus in transgenic mice. *Genes Dev.*, **6**: 1857-64.
154. Peterson, K. R., Li, Q. L., Clegg, C. H., Furukawa, T., Navas, P. A., Norton, E. J., Kimbrough, T. G. et Stamatoyannopoulos, G. (1996) Use of yeast artificial chromosome (YACs) in studies of  $\beta$ -globin locus YAC mice carrying human globin developmental mutants. *Proc. Natl Acad. Sci. USA*, **92**: 5655-5659.
155. Pham, C. T., MacIvor, D. M., Hug, B. A., Heusel, J. W. et Ley, T. J. (1996) Long-range disruption of gene expression by a selectable marker cassette. *Proc. Natl Acad. Sci. USA*, **93**: 13090-5.
156. Kim, C. G., Epner, E. M., Forrester, W. C. et Groudine, M. (1992) Inactivation of the human  $\beta$ -globin gene by targeted insertion into the  $\beta$ -globin locus control region. *Genes Dev.*, **6**: 928-38.
157. Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D. I. K., Enver, T., Ley, T. T. et Groudine, M. (1995) Targeted deletion of 5'HS2 of the murine  $\beta$ -globin LCR reveals that it is not essential for proper regulation of the  $\beta$ -globin locus. *Genes Dev.*, **9**: 2203-13.
158. Shih, D. M., Wall, R. J. et Shapiro, S. G. (1993) A 5' control region of the human epsilon-globin gene is sufficient for embryonic specificity in transgenic mice. *J. Biol. Chem.*, **268**: 3066-71.
159. Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T. et Stamatoyannopoulos, G. (1990) Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science*, **250**: 1147-9.
160. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. et Grosveld, F. (1991) Importance of globin gene order for correct developmental expression. *Genes Dev.*, **5**: 1387-94.
161. Peterson, K. R. et Stamatoyannopoulos, G. (1993) Role of gene order in developmental control of human  $\gamma$ - and  $\beta$ -globin gene expression. *Mol. Cell. Biol.*, **13**: 4836-43.
162. Dillon, N., Strouboulis, J. et Grosveld, F. (1995) In Stamatoyannopoulos, G. (ed.), *Molecular biology of hemoglobin switching*, Proceedings of the 9th conference on hemoglobin switching. Intercept Limited, Andover, Hants, UK, pp. 23-28.

163. Wijgerde, M., Grosveld, F. et Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, **377**: 209-13.
164. Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipson, S., Grosveld, F. et Fraser, P. (1996) The role of EKLF in human  $\beta$ -globin gene competition. *Genes Dev.*, **10**: 2894-902.
165. Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T. et Groudine, M. (1990) A deletion of the human  $\beta$ -globin locus activation region causes a major alteration in chromatin structure and replication across the entire  $\beta$ -globin locus. *Genes Dev.*, **4**: 1637-49.
166. Aladjem, M. I., Groudine, M., Brody, L. L., Dieken, E. S., Fournier, R. E., Wahl, G. M. et Epner, E. M. (1995) Participation of the human  $\beta$ -globin locus control region in initiation of DNA replication. *Science*, **270**: 815-9.
167. Collis, P., Antoniou, M. et Grosveld, F. (1990) Definition of the minimal requirements within the human  $\beta$ -globin gene and the dominant control region for high level expression. *EMBO J.*, **9**: 233-40.
168. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. et Greaves, D. R. (1989) A dominant control region from the human  $\beta$ -globin locus conferring integration site-independent gene expression. *Nature*, **338**: 352-5.
169. Fraser, P., Hurst, J., Collis, P. et Grosveld, F. (1990) DNaseI hypersensitive sites 1: 2 and 3 of the human  $\beta$ -globin dominant control region direct position-independent expression. *Nucl. Acids Res.*, **18**: 3503-8.
170. Ellis, J., Talbot, D., Dillon, N. et Grosveld, F. (1993) Synthetic human  $\beta$ -globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *EMBO J.*, **12**: 127-34.
171. Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipson, S. et Grosveld, F. (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human  $\beta$ -globin locus control region. *EMBO J.*, **15**: 562-568.
172. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., et Fraser, P. (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**: 105-114.
173. Fraser, P., Pruzina, S., Antoniou, M. et Grosveld, F. (1993) Each hypersensitive site of the human  $\beta$ -globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.*, **7**: 106-113.

174. Tuan, D., Abeliovich, A., Lee-Oldham, M. et Lee, D. (1987) Identification of regulatory elements of human  $\beta$ -like globin genes. *Progress Clinical Biol.*, **251**: 211-220.
175. Talbot, D., Philipsen, S., Fraser, P. et Grosveld, F. (1990) Detailed analysis of the site 3 region of the human  $\beta$ -globin dominant control region. *EMBO J.*, **9**: 2169-2177.
176. Talbot, D. et Grosveld, F. (1991) The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J.*, **10**: 1391-1398.
177. Caterina, J. J., Ryan, T. M., Pawlik, K. M., Palmiter, R. D., Brinster, R. L., Behringer, R. R. et Townes, T. M. (1991) Human  $\beta$ -globin locus control region: Analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. *Proc. Natl Acad. Sci. USA*, **88**: 1626-1630.
178. Liu, D., Chang, J. C., Moi, P., Liu, W., Kan, Y. W. et Curtin, P. T. (1992) Dissection of the enhancer activity of  $\beta$ -globin 5' DNase I-hypersensitive site 2 in transgenic mice. *Proc. Natl Acad. Sci. USA*, **89**: 3899-903.
179. Caterina, J. J., Ciavatta, D. J., Donze, D., Behringer, R. R., and Townes, T. M. (1994) Multiple elements in human  $\beta$ -globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucl. Acids Res.*, **22**: 1006-11.
180. Peterson, K. R., Clegg, C. H., Navas, P. A., Norton, E. J., Kimbrough, T. G. et Stamatoyannopoulos, G. (1996) Effect of deletion of 5'HS3 or 5'HS2 of the human  $\beta$ -globin locus control region on the developmental regulation of globin gene expression in  $\beta$ -globin locus yeast artificial chromosome transgenic mice. *Proc. Natl Acad. Sci. USA*, **93**: 6605-9.
181. Philipsen, S., Talbot, D., Fraser, P. et Grosveld, F. (1990) The  $\beta$ -globin dominant control region: hypersensitive site 2. *EMBO J.*, **9**: 2159-2167.
182. Philipsen, S., Pruzina, S. et Grosveld, F. (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the  $\beta$  globin locus control region. *EMBO J.*, **12**: 1077-85.
183. Pruzina, S., Antoniou, M., Hurst, J., Grosveld, F. et Philipsen, S. (1994) Transcriptional activation by hypersensitive site three of the human  $\beta$ -globin locus control region in murine erythroleukemia cells. *Biochi. Bioph. Acta*, **1219**: 351-60.
184. Hug, B. A., Wesselschmidt, R. L., Fiering, S., Bender, M. A., Epner, E., Groudine, M. et Ley, T. J. (1996) Analysis of mice containing a targeted deletion of  $\beta$ -globin locus control region 5' hypersensitive site 3. *Mol. Cell. Biol.*, **16**: 2906-12.

185. Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. et Philipson, S. (1991) Hypersensitive site 4 of the human  $\beta$  globin locus control region. *Nucl. Acids Res.*, **19**: 1413-1419.
186. Walters, M. C., Kim, C. et Gelinias, R. (1991) Characterization of a DNA binding activity in DNase I hypersensitive site 4 of the human locus control region. *Nucl. Acids Res.*, **19**: 5385-5393.
187. Reitman, M., Lee, E., Westphal, H. et Felsenfeld, G. (1990) Site-independent expression of the chicken  $\beta$ A-globin gene in transgenic mice. *Nature*, **348**: 749-52.
188. Reitman, M., Lee, E., Westphal, H. et Felsenfeld, G. (1993) An enhancer/locus control region is not sufficient to open chromatin. *Mol. Cell. Biol.*, **13**: 3990-8.
189. Mason, M. M., Lee, E., Westphal, H. et Reitman, M. (1995) Expression of the chicken  $\beta$ -globin gene cluster in mice: correct developmental expression and distributed control. *Mol. Cell. Biol.*, **15**: 407-414.
190. Reitman, M., Lee, E. et Westphal, H. (1995) Function of the upstream hypersensitive sites of the chicken  $\beta$ -globin gene cluster in mice. *Nucl. Acids Res.*, **23**: 1790-4.
191. Stamatoyannopoulos, G., Josephson, B., Zhang, J. W. et Li, Q. (1993) Developmental regulation of human  $\gamma$ -globin genes in transgenic mice. *Mol. Cell. Biol.*, **13**: 7636-7644.
192. Stamatoyannopoulos, J. A., Clegg, C. H. et Li, Q. (1997) Sheltering of  $\gamma$ -globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the A  $\gamma$ -globin gene. *Mol. Cell. Biol.*, **17**: 240-7.
193. Li, Q. et Stamatoyannopoulos, J. A. (1994) Position independence and proper developmental control of  $\gamma$ -globin gene expression require both a 5' locus control region and a downstream sequence element. *Mol. Cell. Biol.*, **14**: 6087-6096.
194. Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. et Kioussis, D. (1996) Locus control region function and heterochromatin-induced position effect variegation. *Science*, **271**: 1123-1125.
195. Evans, T. et Felsenfeld, G. (1989) The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*, **58**: 877-85.
196. Wall, L., deBoer, E. et Grosveld, F. (1988) The human  $\beta$ -globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev.*, **2**: 1089-1100.
197. Gong, Q. et Dean, A. (1993) Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NF-E2. *Mol. Cell. Biol.*, **13**: 911-7.

198. Jarman, A. P., Wood, W. G., Sharpe, J. A., Gourdon, G., Ayyub, H. et Higgs, D. R. (1991) Characterization of the major regulatory element upstream of the human  $\alpha$ -globin gene cluster. *Mol. Cell. Biol.*, **11**: 4679-89.
199. Rahuel, C., Vinit, M. A., Lemarchandel, V., Cartron, J. P. et Romeo, P. H. (1992) Erythroid-specific activity of the glycophorin B promoter requires GATA-1 mediated displacement of a repressor. *EMBO J.*, **11**: 4095-102.
200. Crossley, M., Tsang, A. P., Bieker, J. J. et Orkin, S. H. (1994) Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J. Biol. Chem.*, **269**: 15440-4.
201. Frampton, J., Walker, M., Plumb, M. et Harrison, P. R. (1990) Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. *Mol. Cell. Biol.*, **10**: 3838-42.
202. Martin, D. I. et Orkin, S. H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev.*, **4**: 1886-98.
203. Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H. et Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**: 257-60.
204. Weiss, M. J. et Orkin, S. H. (1995) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc. Natl Acad. Sci. USA*, **92**: 9623-7.
205. Merika, M. et Orkin, S. H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.*, **15**: 2437-47.
206. Gregory, R. C., Taxman, D. J., Seshasayee, D., Kensinger, M. H., Bieker, J. J. et Wojchowski, D. M. (1996) Functional interaction of GATA-1 with erythroid Krüppel-like factor and Sp1 at defined erythroid promoters. *Blood*, **87**: 1793-801.
207. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. et Orkin, S. H. (1993) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature*, **362**: 722-8.
208. Lu, S. J., Rowan, S., Bani, M. R. et Ben-David, Y. (1994) Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in Friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc. Natl Acad. Sci. USA*, **91**: 8398-402.

209. Kotkow, K. J. et Orkin, S. H. (1995) Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol. Cell. Biol.*, **15**: 4640-7.
210. Shivdasani, R. A., Rosenblatt, M. F., Zucker-Franklin, D., Jackson, C. W., Hunt, P., Saris, C. J. et Orkin, S. H. (1995) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, **81**: 695-704.
211. Shivdasani, R. A. et Orkin, S. H. (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc. Natl Acad. Sci. USA*, **92**: 8690-8694.
212. Chan, J. Y., Han, X. L. et Kan, Y. W. (1993) Cloning of Nrf1: an NF-E2-related transcription factor, by genetic selection in yeast. *Proc. Natl Acad. Sci. USA*, **90**: 11371-5.
213. Moi, P., Chan, K., Asunis, I., Cao, A. et Kan, Y. (1994) Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the  $\beta$ -globin locus control region. *Proc. Natl Acad. Sci. USA*, **91**: 9926-9930.
214. Caterina, J. J., Donze, D., Sun, C. W., Ciavatta, D. J. et Townes, T. M. (1994) Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression. *Nucl. Acids Res.*, **22**: 2383-91.
215. Amrolia, P. J., Cunningham, J. M., Ney, P., Nienhuis, A. W. et Jane, S. M. (1995) Identification of two novel regulatory elements within the 5'-untranslated region of the human A  $\gamma$ -globin gene. *J. Biol. Chem.*, **270**: 12892-8.
216. Miller, I. J. et Bieker, J. J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol. Cell. Biol.*, **13**: 2776-86.
217. Southwood, C. M., Downs, K. M. et Bieker, J. J. (1996) Erythroid Krüppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev. Dynamics*, **206**: 248-59.
218. Bieker, J. J. (1995) In Stamatoyannopoulos, G. (ed.), Molecular biology of hemoglobin switching, Proceedings of the 9th conference on hemoglobin switching. Intercept Limited, Andover, Hants, UK, pp. 231-241.
219. Donze, D., Townes, T. M. et Bieker, J. J. (1995) Role of erythroid Krüppel-like factor in human  $\gamma$ - to  $\beta$ -globin gene switching. *J. Biol. Chem.*, **270**: 1955-9.



220. Feng, W. C., Southwood, C. M. et Bieker, J. J. (1994) Analyses of  $\beta$ -thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J. Biol. Chem.*, **269**: 1493-500.
221. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. et Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**: 316-8.
222. Perkins, A. C., Sharpe, A. H. et Orkin, S. H. (1995) Lethal  $\beta$ -thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**: 318-22.
223. Lim, S.-K., Bieker, J. J., Lin, C.-S. et Costantini, F. (1996) A shortened life span of EKLF  $-/-$  adult erythrocytes, due to a deficiency of  $\beta$ -globin chains, is ameliorated by human  $\gamma$ -globin chains. *Blood*, **90**: 1291-1299.
224. Perkins, A. C., Gaensler, K. M. et Orkin, S. H. (1996) Silencing of human fetal globin expression is impaired in the absence of the adult  $\beta$ -globin gene activator protein EKLF. *Proc. Natl Acad. Sci. USA*, **93**: 12267-71.
225. Su, W., Jackson, S., Tjian, R. et Echols, H. (1991) DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev.*, **5**: 820-6.
226. Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P. et Hough, P. V. (1991) DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc. Natl Acad. Sci. USA*, **88**: 5670-4.
227. Antoniou, M. et Grosveld, F. (1990)  $\beta$ -globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev.*, **4**: 1007-13.
228. Delvoe, N. L., Destroismaisons, N. M. et Wall, L. A. (1993) Activation of the  $\beta$ -globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol. Cell. Biol.*, **13**: 6969-83.
229. Sutherland, H.G.E., Martin, D.I.K. et Whitelaw, E. (1997) A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene. *Mol. Cell. Biol.*, **17**: 1607-14.
230. Enver, T., Li, Q., Gale, K.B., May, G.E., Karlinsey, J.E., Jiminez, G. Papayannoupoulou, T. et Costantini, F. (1994) Analysis of the developmental and transcriptional potentiation functions of 5'HS2 of the murine  $\beta$ -globin locus control region in transgenic mice. *Dev. Biol.* **165**: 574-84.
231. Hartzog, G.A., Myers, R.M. (1993) Discrimination among potential activators of the  $\beta$ -globin CACCC element by correlation of binding and transcriptional properties. *Mol. Cell. Biol.*, **13**: 44-56.

232. Bieder, J.J. Isolation, genomic structure, and expression of human erythroid krüppel-like factor (EKLF). (1996) *DNA Cell Biol.* **5**: 347-52.
233. Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. et Whitelaw, E. (1996) Age-dependent silencing of globin transgenes in the mouse. *Nucl. Acids Res.* **24**: 1465-1471.

**Annexe I**

## Article 4

Une interaction mutuelle entre le LCR et le gène  $\beta$ -globine  
peut prévenir les effets de position  
si la disposition des éléments du gène est correcte

Louis-Georges Guy, Qi Mei, Rashmi Kothary, Yves DeRepentigny  
et Lee Wall.

## **Résumé**

Nous avons déjà démontré que l'insertion du gène rapporteur *lacZ* dans le microlocus de  $\beta$ -globine empêchait une expression indépendante conférée par le LCR de la position chez les souris transgéniques (Chapitre 1). Afin de comprendre le mécanisme de cette interférence, nous avons analysé de nouvelles constructions où les séquences *lacZ* ont été remplacées par l'ADNc eucaryote TIMP. Ces transgènes sont aussi soumis à des effets de position, démontrant que *lacZ* ne porte pas de façon intrinsèque d'éléments inhibant l'activité du LCR. Nous proposons un modèle où une interaction mutuelle doit avoir lieu entre le LCR et un complexe constitué d'éléments du gène de  $\beta$ -globine pour prévenir les effets de position. Le déplacement de ces éléments par l'insertion de *lacZ* empêcherait une interaction productive de se produire. De plus, nous démontrons qu'un gène de  $\beta$ -globine complet et intact ne peut interagir avec le LCR de façon à prévenir les effets de position si un gène  $\beta$ -*lacZ* est inclus dans le transgène mais plus près du LCR (donc meilleur compétiteur). Selon notre modèle, ce gène  $\beta$ -*lacZ* pourrait compétitionner avec le gène  $\beta$ -globine pour l'interaction avec le LCR mais cette interaction ne serait pas productive, c'est-à-dire incapable de prévenir les effets de position.

Tableau I: Expression des transgènes  $\beta$ I et  $\beta$ J

Lignée	nombre <sup>1</sup> de copies	$\beta$ -gal. <sup>2</sup> (mU/mg)	$\beta$ -gal./copie	% $\beta$ hu/ $\beta$ mu <sup>2</sup>	Exp./copie
$\beta$ I1	4	220	55	1.3	0.33
$\beta$ I2	23	0	0	0	0
$\beta$ I3	80	190	2.3	1.6	0.02
$\beta$ I4	28	0	0	0	0
$\beta$ I5	55	0	0	0	0
$\beta$ I6	22	14	0.63	<0.1	<0.01
$\beta$ J1	5	1100	220	11.8	2.36
$\beta$ J2	1	0	0	0	0
$\beta$ J3	3	0	0	0	0
$\beta$ J4	2	130	70	5.1	1.05

1. Le nombre de copies a été déterminé avec de l'ADN de foie foetal.

2. Les analyses ont été effectuées sur des échantillons prélevés sur des souris transgénique fondatrices de 14,5 jours.

Tableau 2: analyse des souris transgéniques  $\beta$ K et  $\beta$ L

Lignée	nombre <sup>1</sup> de copies	% $\beta$ hu/ $\beta$ mu <sup>2,3</sup>	Exp./copie
$\beta$ K1	3	0.14	4.7
$\beta$ K2	1	0.008	0.77
$\beta$ K3	16	0.026	0.16
$\beta$ K4	25	0.001	<0.01
$\beta$ K5	3	0.003	0.11
$\beta$ K6	2	0.028	1.4
$\beta$ K7	5	0.020	0.40
$\beta$ K8	9	0.004	0.05

Lignée	nombre de copies	Timp <sup>2,4</sup>	Exp./copie
$\beta$ L2	31	0.090	0.0029
$\beta$ L3	1	0.018	0.018
$\beta$ L5	7	0.0046	0.00066
$\beta$ L18	0.9	0.0033	0.037

1. Le nombre de copies a été déterminé avec de l'ADN de foie foetal.
2. Les analyses ont été effectuées sur des échantillons prélevés sur des souris transgéniques fondatrices de 14,5 jours.
3. L'expression de TIMP était si faible qu'elle n'a pu être détectée par protection aux ribonucléases.
4. L'expression de TIMP a été calculée de façon relative au gène  $\beta$ -major.

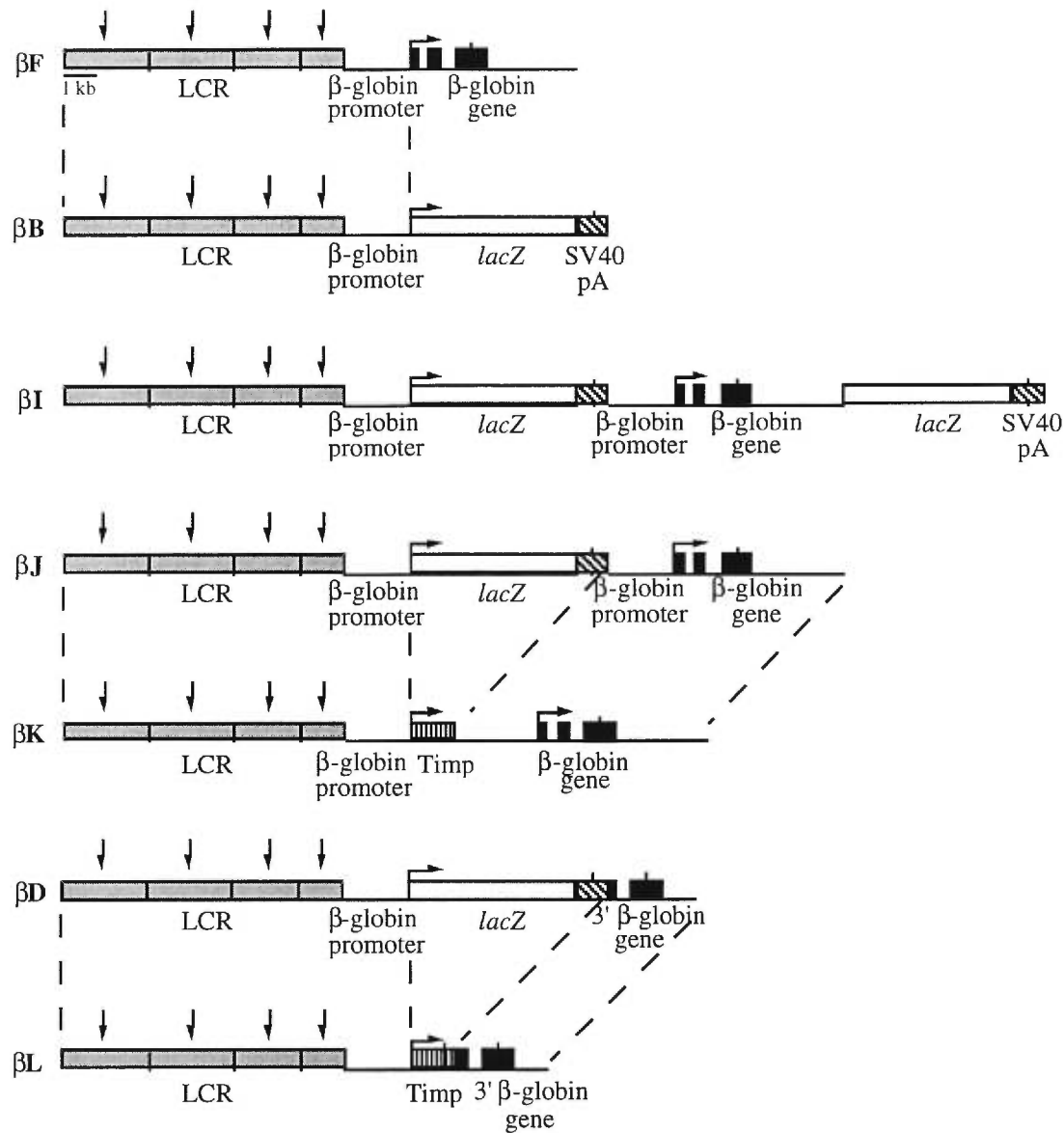


Figure 1: Constructions.  $\beta F$  est le micro-locus normal où les sites hypersensibles du LCR (indiqués par des flèches) sont liés au gène de  $\beta$ -globine. Dans  $\beta B$ , *lacZ*, encodant la  $\beta$ -galactosidase, remplace les séquences codantes de  $\beta$ -globine, le reste du transgène est identique à  $\beta F$ . Dans  $\beta I$  et  $\beta J$ , un gène complet et intact de  $\beta$ -globine a été ajouté en 3' de  $\beta B$  afin de tester la capacité de ce gène de rétablir une expression indépendante de la position malgré la présence de *lacZ*. Dans  $\beta I$ , un second exemplaire des séquences *lacZ* a été ajouté en 3' afin d'empêcher une juxtaposition directe entre le gène de  $\beta$ -globine et le LCR dans le cas de copies multiples. Cette juxtaposition sera cependant possible dans le cas de  $\beta J$ .  $\beta K$  est identique à  $\beta J$  à l'exception des séquences *lacZ* qui ont été remplacées par l'ADNc eucaryote TIMP. De façon similaire,  $\beta$ -globine est l'équivalent de  $\beta D$  ( $\beta B$  avec la partie 3' du gène de  $\beta$ -globine, incluant le dernier intron, les enhancers et le signal de poly-adénylation) mais où *lacZ* a été substitué par TIMP.

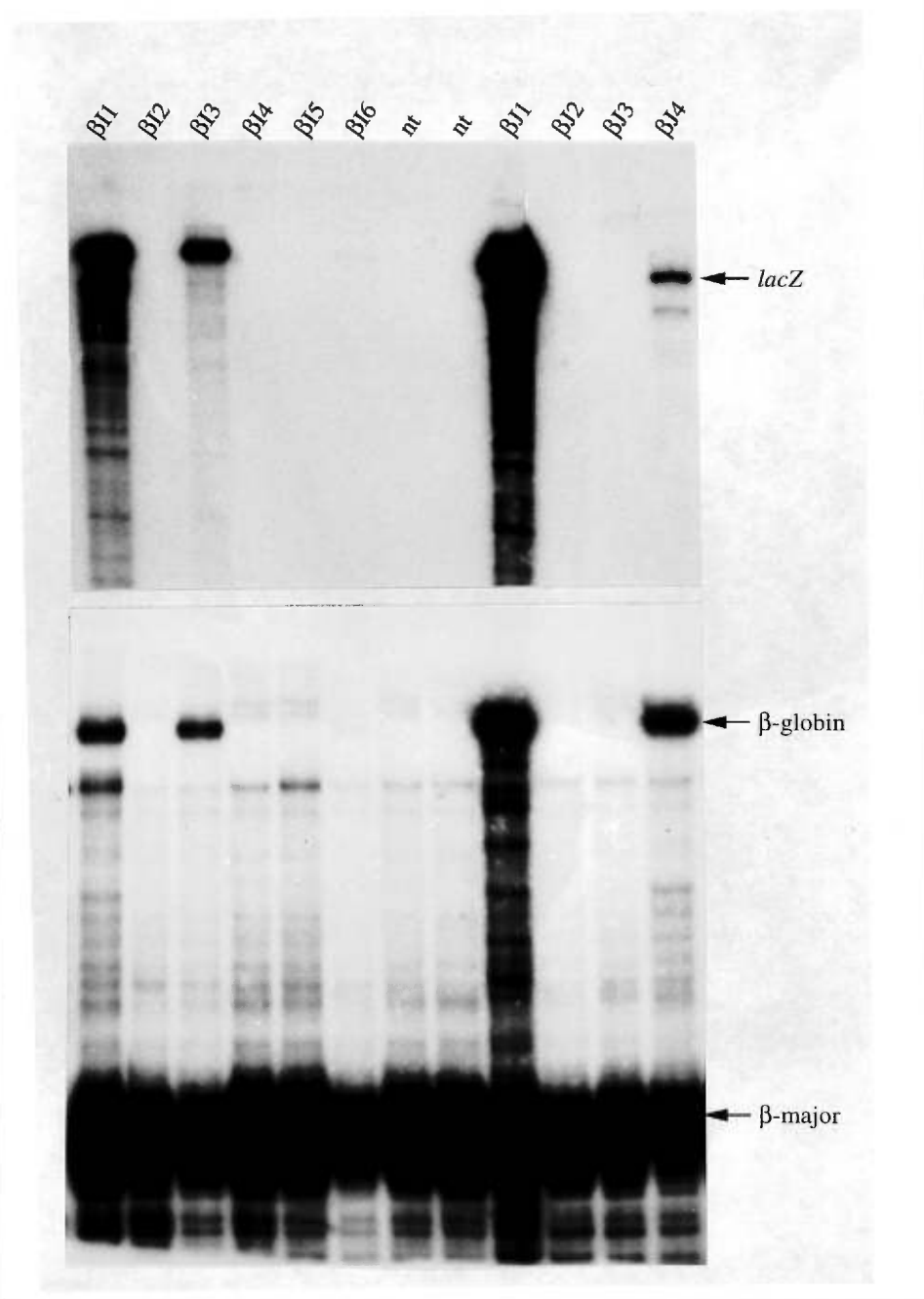


Figure 2: Analyse de l'expression des transgènes  $\beta I$  et  $\beta J$  chez les souris transgéniques. Les ARN totaux isolés du foie foetal des souris transgéniques  $\beta I$  et  $\beta J$  ont été analysés par protection aux ribonucléases pour détecter l'expression de *lacZ* et de  $\beta$ -globine humaine du transgène et de  $\beta$ -major endogène comme contrôle. Des échantillons d'ARN de souris non transgéniques (nt) ont été ajoutés comme contrôle négatif.



$\beta$ K1  $\beta$ K2  $\beta$ K3  $\beta$ K4  $\beta$ K5  $\beta$ K6  $\beta$ K7  $\beta$ K8 nt  $\mu$ D14

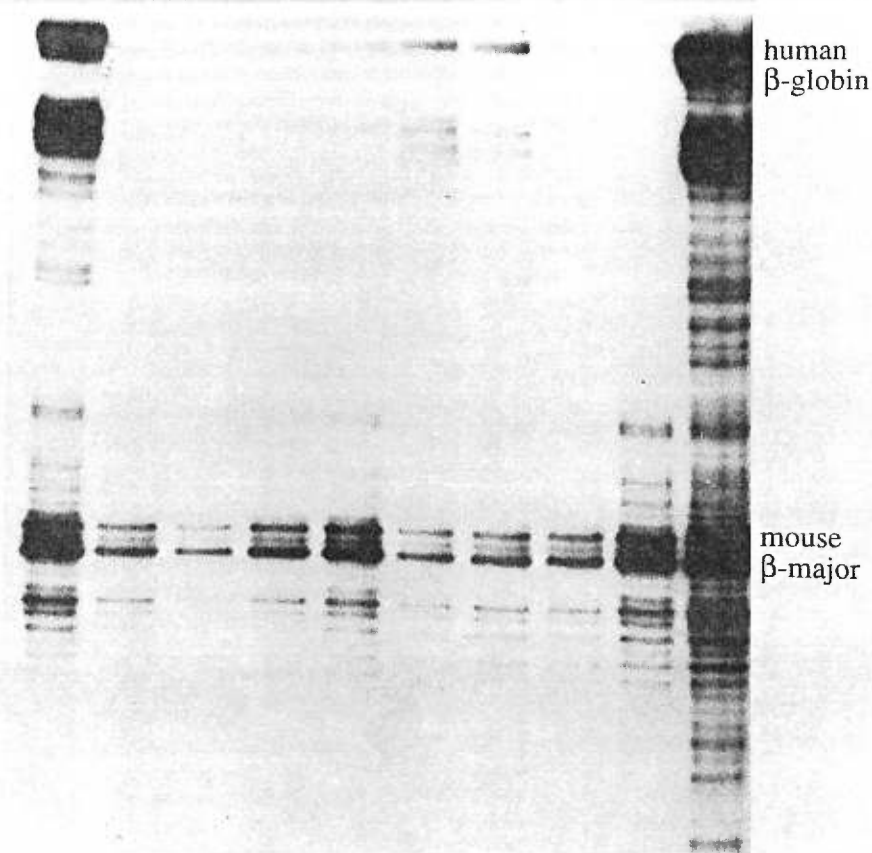


Figure 3: Analyse de l'expression du transgène  $\beta$ K chez les souris transgéniques. L'expression de la  $\beta$ -globine humaine du transgène et de  $\beta$ -major endogène a été mesurée dans l'ARN de foie foetal par protection aux ribonucléases. L'expression de TIMP était si faible qu'elle n'a pu être détectée (donnée non-montrée). Comme contrôle, des échantillons d'ARN provenant d'une souris non-transgénique (nt) et d'une souris portant le transgène micro-locus de  $\beta$ -globine humaine ( $\mu$ -D14) ont aussi été analysés.

Annexe II

Article 5

Fonction du motif CACC du promoteur de  $\beta$ -globine  
dans les cellules K562 et les souris transgéniques

Louis-Georges Guy, Yves DeRepentigny,

Rashmi Kothary et Lee Wall.

## **Résumé**

Les cellules K562 expriment les globines embryonnaires et foetales mais pas le gène de  $\beta$ -globine adulte. Toutefois, si le transgène est lié en *cis* au LCR, un puissant activateur du locus de  $\beta$ -globine, alors le gène est transcrit. La cause de cette perte de spécificité développementale est que le LCR est actif dans les stades embryonnaire et foetal et peut activer le gène  $\beta$ -globine malgré sa spécificité adulte. Toutefois, le laboratoire du Dr. Lee Wall a démontré qu'un promoteur minimal, ne contenant que les motifs CAAT et TATA ne répondait plus à la stimulation du LCR dans les cellules K562 (228). Dans la présente étude, nous montrons que le motif CACC du promoteur de  $\beta$ -globine, un élément hautement conservé et important pour l'activité transcriptionnelle du promoteur, était suffisant pour permettre l'activation de la transcription par le LCR dans les cellules K562. Afin de caractériser davantage la stimulation du promoteur de  $\beta$ -globine par le LCR, nous avons analysé l'expression de plusieurs constructions chez les souris transgéniques. Ces résultats démontrent que le motif CACC, en conjonction avec le LCR, est important mais non essentiel pour l'expression d'un gène rapporteur *lacZ*. Pourtant, l'ablation génétique de EKLF, un facteur de transcription liant le motif CACC, résulte en une diminution de l'expression de la  $\beta$ -globine d'environ 20 fois. En croisant nos souris transgéniques avec les souris knock out pour EKLF, nous démontrons que l'expression du rapporteur *lacZ* est totalement indépendante de EKLF. Il est donc possible que l'action de EKLF sur l'expression du gène de  $\beta$ -globine passe par d'autres séquences que le motif CACC, par exemple les enhancers de  $\beta$ -globine absents des constructions LCR- $\beta$ -*lacZ*.

Tableau 1: Expression de la  $\beta$ -galactosidase chez les souris transgéniques

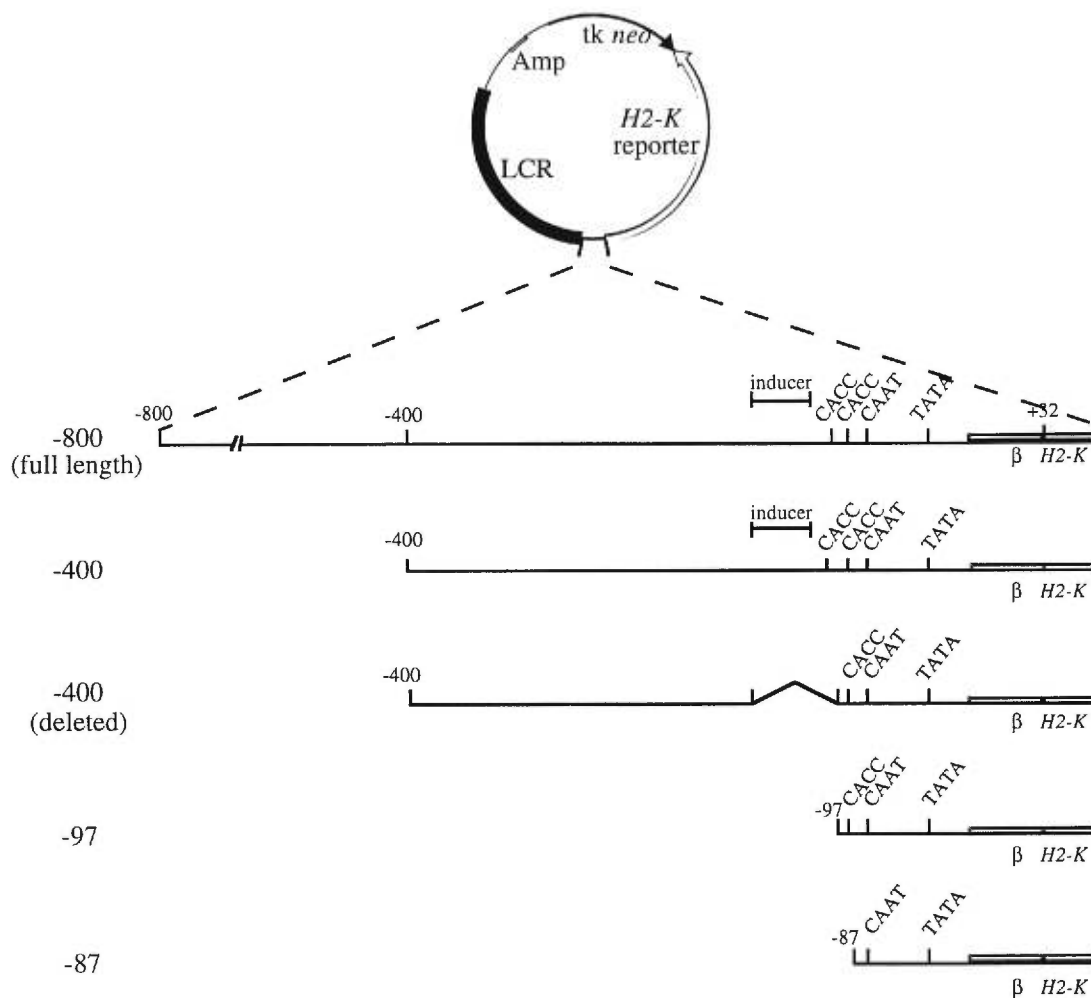
Lignée	$\beta$ -gal. <sup>1</sup> (mU/mg)	nombre <sup>2</sup> de copies	Exp./copy
$\beta$ B1	25	3	8.3
$\beta$ B2	25	19	1.3
$\beta$ B3	590	14	42
$\beta$ B4	720	3	240
$\beta$ B5	610	4	150
$\beta$ B6	900	3	300
$\beta$ B7	8	7	1.1
Moyenne	410		100
$\beta$ G1	1300	3	430
$\beta$ G2	340	30	11
$\beta$ G3	2	80	0.03
$\beta$ G4	320	5	64
$\beta$ G5	0	3	0
$\beta$ G6	550	13	42
$\beta$ G7	1	5	0.2
$\beta$ G8	50	14	3.6
Moyenne	320		70
$\beta$ H1	0	60	0
$\beta$ H2	2	90	0.2
$\beta$ H3	50	10	5
$\beta$ H4	140	7	20
$\beta$ H5	440	3	150
Moyenne	130		35
$\beta$ A1	0	2	0
$\beta$ A2	220	3	73
$\beta$ A3	0	~200	0
$\beta$ A4	270	5	54
$\beta$ A5	60	1	60
Moyenne	110		37

1. Les foies foetaux ont été prélevés 15 jours suivant l'observation du culot spermatique.

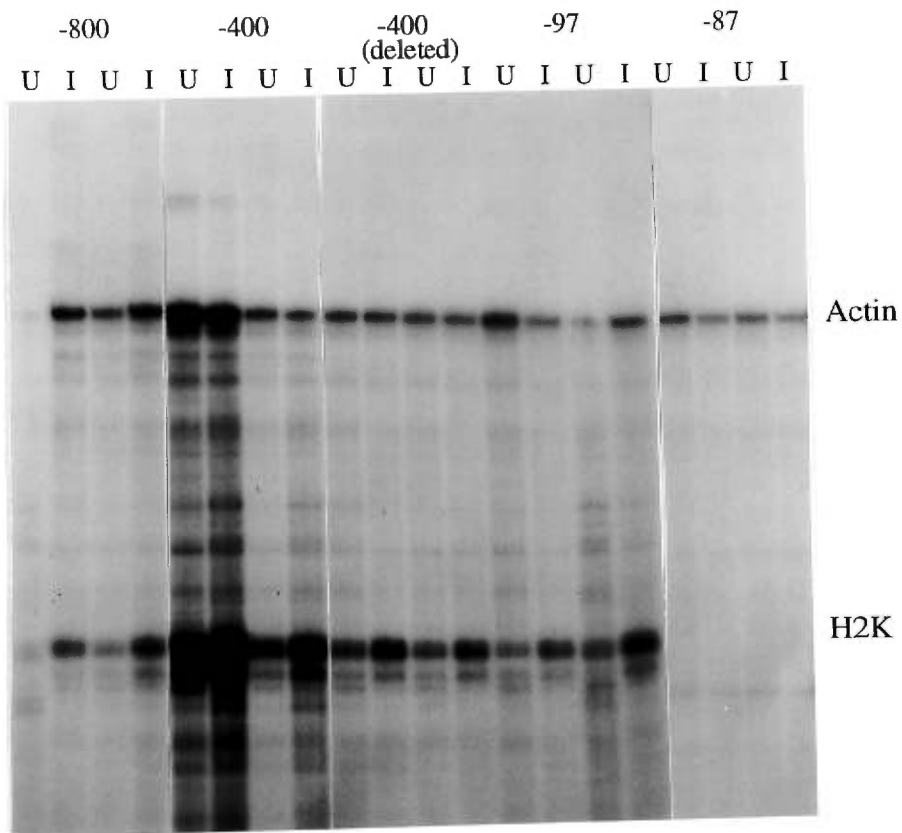
2. Le nombre de copies déterminé avec de l'ADN de F1.

Tableau 2: Comparaison de l'expression  $\beta$ -galactosidase chez les souris transgénique des divers génotype EKLF à 14.5 jours.

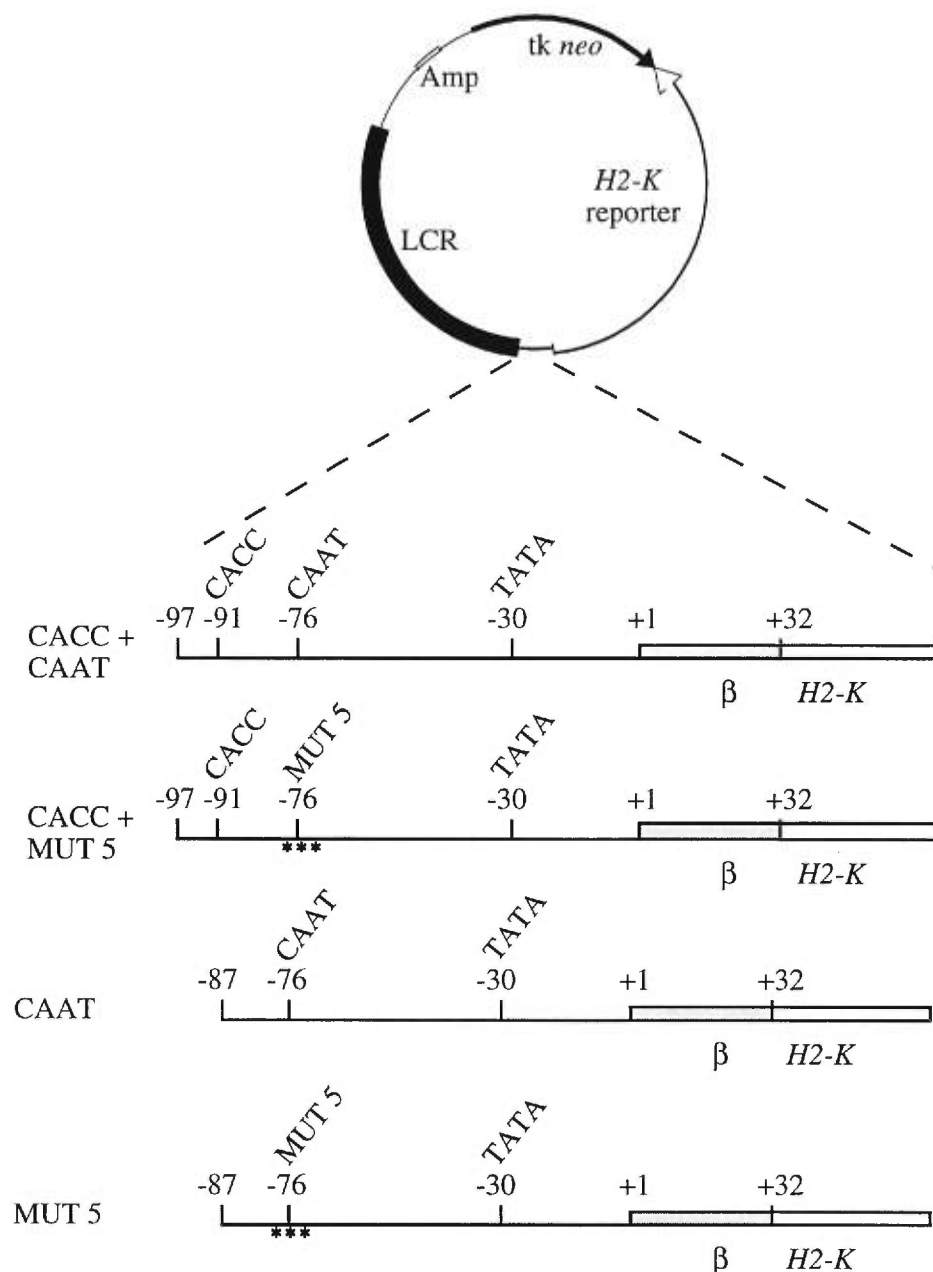
Lignée	EKLF -/-	EKLF +/-	EKLF +/+
$\beta$ B4 (-800)	1900	950	780
$\beta$ G6 (CACC+ CAAT)	260	300	220
$\beta$ G8 (CACC+ CAAT)	180	26	28
$\beta$ H3 (CACC)	140	28	24
$\beta$ H4 (CACC)	1000	100	140
$\beta$ A2 (CAAT)	260	90	120
$\beta$ A4 (CAAT)	890	550	370



**Figure 1:** Constructions incluant des délétions du promoteur de  $\beta$ -globine. Un plasmide contenant le gène de résistance *neo*, le LCR et un gène rapporteur *H-2K* (haut de la figure) a été utilisé pour insérer différentes versions du promoteur de  $\beta$ -globine (bas de la figure). Le promoteur -800 est le promoteur complet tel qu'utilisé dans le micro-locus. Le promoteur -400 est constitué d'une délétion des séquences -800 à -400 du promoteur complet. La version -400 deleted est identique à -400 mais une région interne comprenant les séquences entre -163 et -97, qui contiennent un élément de réponse à l'induction des cellules MEL, a été déléetée. Le promoteur -97 est un promoteur minimal ne comprenant que les motifs CACC proximal, CAAT et TATA. Le promoteur -87 est similaire mais les nucléotides du motif CACC ont été déléetés.

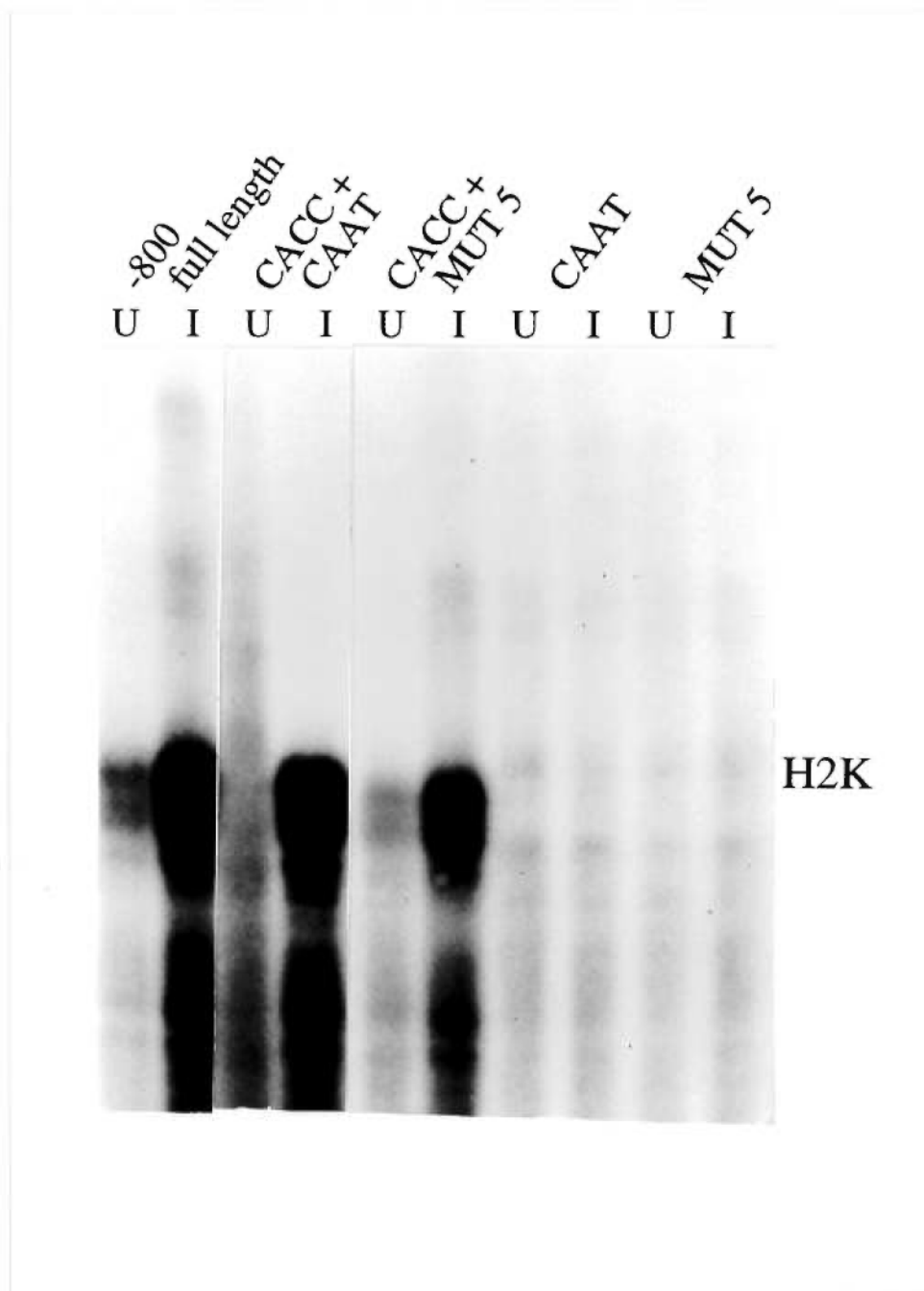


**Figure 2:** Analyse de l'expression avec les différentes délétions du promoteur  $\beta$ -globine dans les cellules K562. L'expression du gène rapporteur *H2-K* a été mesurée dans des populations de cellules K562 sélectionnées au G-418 pour l'insertion stable des plasmides indiqués au haut de la figure (voir Figure 1) a été mesurée par protection aux ribonucléases. Comme contrôle, l'expression de la  $\beta$ -actine endogène a aussi été mesurée. Pour chaque construction, deux populations différentes ont été analysées et l'expression a été testée dans des cellules non induites (U) ou induites à l'hémine (I). La figure révèle que toutes les constructions incluant le motif CACC proximal (-800 à -97) expriment le gène rapporteur alors que la délétion du motif (-87) empêche l'expression.

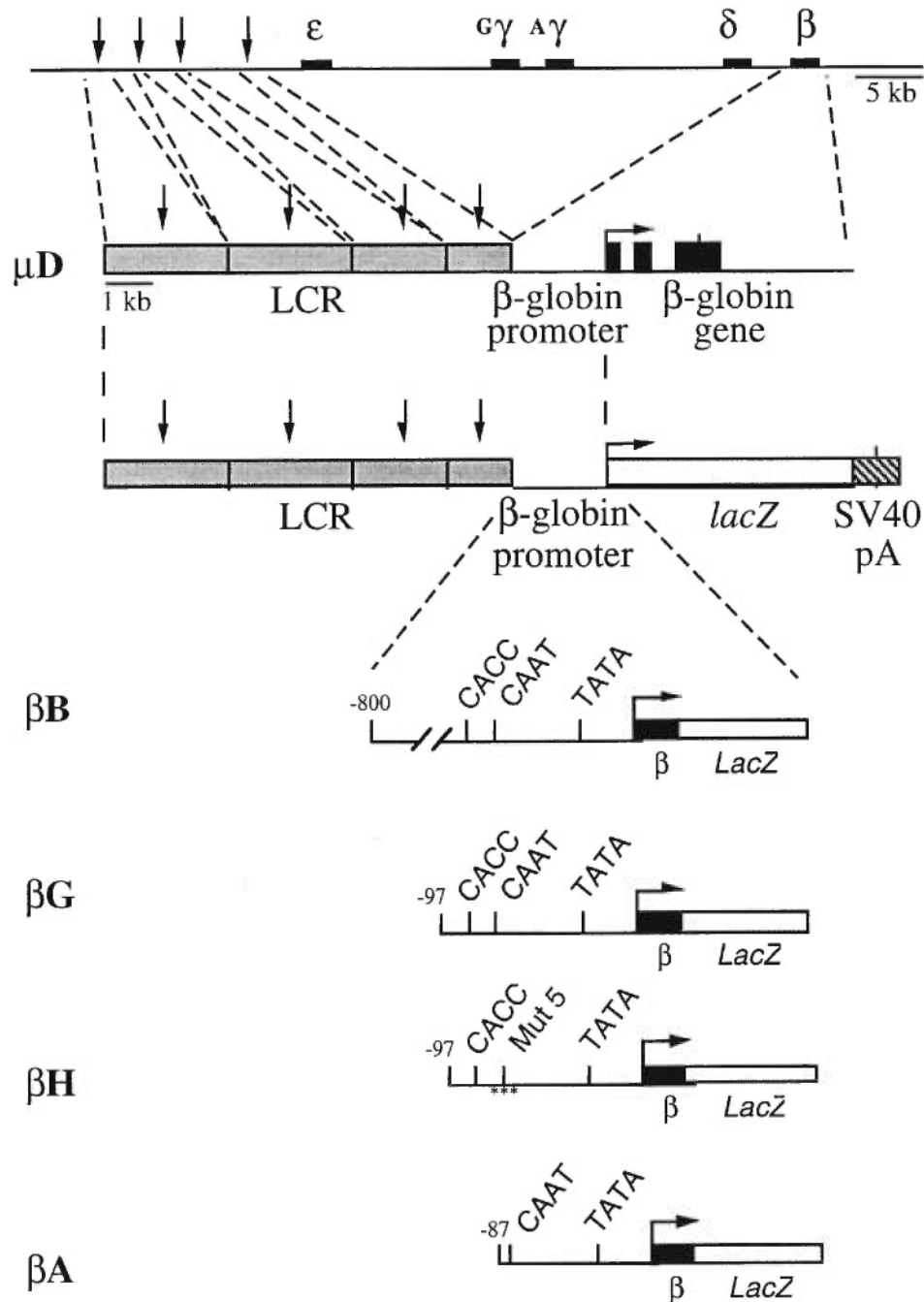


**Figure 3:** Constructions de diverses combinaison des motifs CACC et CAAT. Afin de vérifier si les motifs CACC et CAAT fonctionnent en conjonction ou si le motif CACC peut, à lui seul, stimuler l'expression du rapporteur dans les cellules K562, diverses combinaison des motifs CACC ont été insérées dans le plasmide décrit à la figure 1. Le promoteur CACC + CAAT est le même que le promoteur -97 de la Figure 1. Le promoteur CACC + MUT5 est un promoteur -97 où ont été insérés des mutations ponctuelles inactivant le motif CAAT (). Le promoteur CAAT était appelé -87 à la Figure 1 et le promoteur MUT5 introduit des mutations inactivant le motif CAAT dans le promoteur -87.





**Figure 4:** Analyse de l'expression avec les différentes combinaisons des motifs CACC et CAAT dans les cellules K562. L'expression du gène rapporteur *H-2K* a été mesuré dans des populations de cellules K562 sélectionnées au G-418 pour l'insertion stable des plasmides indiqués a été mesuré par protection aux ribonucléases. Quoique les niveaux ne soient pas directement comparable en raison de l'absence de contrôle, il est évident que le motif CACC peut soutenir une expression élevé sans la présence du motif CAAT (voir CACC + MUT5, I). Noter que l'ARN des pistes (U) était de moins bonne qualité que dans les pistes (I), expliquant la différence de niveaux.



**Figure 5:** Constructions analysées dans les souris transgéniques. Différents promoteurs de  $\beta$ -globine ont été placés entre le LCR et le gène rapporteur *lacZ*. Dans  $\beta B$ , le promoteur complet (-800 à +32), tel qu'utilisé dans le micro-locus ( $\mu D$ ) a été utilisé. Dans  $\beta G$ , un promoteur minimal ne comprenant que les motifs CACC proximal, CAAT et TATA remplace le promoteur complet. Dans  $\beta H$ , des mutations ponctuelles inactivant le motif CAAT (MUT5) ont été introduites dans le promoteur -97, ne laissant que les motifs CACC et TATA fonctionnels.  $\beta A$  est un promoteur -87 où ne subsistent que les motifs CAAT et TATA.

## Annexe III

## Article 6

CAAT/enhancer-binding proteins are involved in  
 $\beta$ -globin gene expression and are differentially expressed  
in murine erythroleukemia and K562 cells

Lee Wall, Nathalie Destroismaisons, Nathalie Delvoye et  
Louis-Georges Guy. (1996) *The Journal of Biological Chemistry*,  
**271**: 16477-16484.

## CAAT/Enhancer-binding Proteins Are Involved in $\beta$ -Globin Gene Expression and Are Differentially Expressed in Murine Erythroleukemia and K562 Cells\*

(Received for publication, November 6, 1995, and in revised form, April 12, 1996)

Lee Wall†§¶, Nathalie Destroismaisons‡, Nathalie Delvoye‡, and Louis-Georges Guy‡

From the ‡Institut du cancer de Montréal of the Centre de Recherche L.-C. Simard and the §Department of Medicine, Université de Montréal, 1560 Sherbrooke Street East, Montreal, Quebec, H2L 4M1 Canada

Acting in *cis* with the  $\beta$ -globin locus control region, the CAAT box of the  $\beta$ -globin gene promoter stimulates transcription 10-fold in murine erythroleukemia (MEL) cells but is without effect in K562 cells. Our previous studies suggested that of four proteins from MEL cells that bind to this CAAT box region (CP1, GATA-1, and two factors that were denoted DSFr and DSF1) DSFr is involved in the up-regulation of transcription. In the present report, the DSFr protein in MEL cells was identified as C/EBP $\gamma$  through expression cloning and antibody studies. C/EBP $\gamma$  DNA binding activity could not be detected in K562 cells. However, K562 cells, but not MEL cells, were found to express LIP, which is a truncated form of C/EBP $\beta$  and is an inhibitor of transcription. Thus, the differential expression of C/EBP members could account for the ability of the  $\beta$ -globin CAAT box to stimulate transcription in MEL cells, but not function in K562 cells. Juxtaposing a specific C/EBP binding sequence next to the  $\beta$ -globin promoter, in constructs in which the CAAT box had been rendered inactive by mutation or deletion, restored full promoter activity in MEL cells only if CP1 still bound to the promoter. In conjunction with previous mutation analyses, these results suggest that C/EBP $\gamma$  may collaborate with CP1 to enhance transcription through the  $\beta$ -globin CAAT box.

The human  $\beta$ -globin locus (Fig. 1) contains five active genes that are expressed specifically in red blood cells at different times during development (reviewed in Ref. 1).  $\epsilon$ -Globin is only expressed early in embryogenesis, the two  $\gamma$ -globin genes are transcribed at high levels during fetal life, while the  $\delta$ - (a minor contributor) and  $\beta$ -globin genes are expressed in late fetal life and throughout adult life. To express at high levels, all of these genes rely on the activity of the LCR<sup>1</sup> that is situated upstream of the  $\epsilon$ -globin gene (see Fig. 1). The LCR is a powerful, erythroid-specific enhancer that has the additional property of being able to confer position-independent, copy number-dependent expression onto globin transgenes under its control (reviewed in Ref. 2).

Transgenic mouse studies of the  $\beta$ - and  $\gamma$ -globin genes, alone or together in *cis* with the LCR (3-6), as well as studies of the

whole human  $\beta$ -globin locus (7, 8), have shown that the globin genes compete for activation by the LCR during development. Investigations have also indicated that the LCR can only activate one gene at a time and that the LCR interacts directly with the gene it is enhancing through a looping mechanism (3, 6, 9). Thus, it is believed that the developmental switching of globin gene expression is regulated, at least in part, through a competition between the genes for interactions with the LCR. In this model, regulatory sequences surrounding each gene would interact with the LCR and, therefore, would play a direct role in determining the affinity each gene has for the LCR during the different stages of development.

Using MEL cells as a model system, Antoniou and Grosveld (10) demonstrated that only a minimal  $\beta$ -globin gene promoter, containing the TATA box, the CAAT box, and the proximal CACC box sequences (Fig. 1), is necessary for the LCR to enhance high levels of transcription onto a *H-2K* reporter gene. In conjunction with the TATA box, the CAAT, and the proximal CACC motifs can each stimulate activity 10-fold higher than the TATA box alone in MEL cells. Thus, both the CAAT and proximal CACC elements and the cognate transcription factors that bind to these sequences play an important role in regulating the  $\beta$ -globin promoter in the presence of the LCR and, thus, may be involved in LCR interactions.

Recent work by other groups has implicated an erythroid-specific kruppel-like factor in the involvement of the  $\beta$ -globin proximal CACC box in LCR activated transcription (11, 12). Mice in which the erythroid-specific kruppel-like factor gene has been crippled through homologous recombination die at the fetal stage of development due to a severe anemia. This may be caused by the specific absence of expression of the mouse  $\beta$ -globin gene (13, 14).

Our studies on the  $\beta$ -globin CAAT box have shown that four distinct factors in MEL cells can bind to this region; the erythroid-specific protein GATA-1, the ubiquitous CAAT-binding protein CP1 (also called NF-Y), and two DNA binding activities that were denoted DSF1 and DSFr (15). DSF1 DNA binding activity is only detectable in MEL cells that are induced to terminally differentiate, while DSFr activity is present in both uninduced and induced cells. A mutation analysis that was done suggested that DSFr is involved in regulating the activity of the  $\beta$ -globin CAAT box in MEL cells. However, in these same studies it was also found that the CAAT box, in *cis* with the LCR, was completely unable to increase transcription from the  $\beta$ -globin promoter in K562 cells (15). Yet, it appeared that K562 cells did contain DSFr DNA binding activity. These results suggested that there might be a difference in the way the LCR functions in MEL versus K562 cells and/or that there are differences in the CAAT box-binding factors that were not detected in the previous studies. Therefore, we sought to identify the DSFr and DSF1 proteins to be able to determine if there

\* This work was supported in part by a grant from the National Cancer Institute of Canada (to L. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Recipient of a scholarship from the Fonds de Recherche en Santé du Québec. To whom correspondence should be addressed. Tel.: 514-876-5497; Fax: 514-876-5476; E-mail: Walll@ere.UMONTREAL.CA.

<sup>1</sup> The abbreviations used are: LCR, locus control region; GMS, gel mobility shift; MEL, murine erythroleukemia; CPS, carbamyl phosphate synthetase; C/EBP, CAAT/enhancer-binding protein.

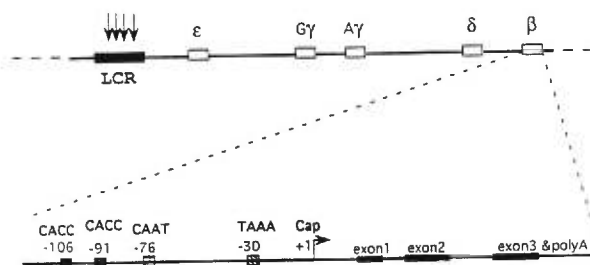


FIG. 1. Structure of the human  $\beta$ -globin gene complex. The promoter and the rest of the  $\beta$ -globin gene are not drawn to scale relative to each other.

exist important differences between the proteins in MEL cells versus K562 cells.

The C/EBP family of transcription factors belong to the basic-ZIP class of DNA-binding proteins (Refs. 16 and 17, and references therein). C/EBP members contain a domain in their C-terminal half that is rich in basic amino acids. This basic domain is followed by a leucine zipper motif (4 to 5 repeats of leucine at every seventh amino acid) through which the C/EBP members form homodimers or heterodimerize with one another. Dimerization is required for the proteins to interact with DNA through the basic domains.

At least six distinct members of the C/EBP family in mammals have been identified: C/EBP $\alpha$ , C/EBP $\beta$  (also known as NFIL6 and LAP), NF-IL6 $\beta$ , C/EBP $\gamma$  (also called IgEBP), C/EBP $\delta$ , and CHOP (16–24). With the exception of CHOP, which does not appear to bind DNA (22), all C/EBP members have a nearly identical DNA-binding specificity. Although the DNA sequence consensus recognized by these proteins is highly redundant, a CAAT-like sequence is frequently observed.

In the present report we have identified DSF1 and DSFr as C/EBP $\beta$  and C/EBP $\gamma$ , respectively, and we have found that these DNA binding activities are differentially expressed in MEL and K562 cells. Moreover, our studies suggest that C/EBP $\gamma$  may act in collaboration with CP1 to up-regulate LCR enhanced  $\beta$ -globin gene expression through the  $\beta$ -globin CAAT box in MEL cells.

#### MATERIALS AND METHODS

**Library Screening**—Custom cDNA libraries, representing dimethyl sulfoxide-induced MEL cells, in  $\lambda$ gt11 (random primed) and  $\lambda$ gt10 (mixed random and oligo(dT) primed) vectors were purchased from Clontech. For screening  $\lambda$ gt11 expression libraries, approximately 50,000 plaque-forming units were grown on each 10-inch square plate and these were screened by the Southwestern method (25). Filters were probed with the synthetic oligonucleotides GAGCCCTAGGGTTGGC-CAATCTACTCCAG and GCTCCTGGGAGTAGATTGGCCAACCCCT-AGG, which were end labeled with [ $\gamma$ - $^{32}$ P]ATP, annealed to form GAGC and GCTC 5' overhangs, and then ligated to form concatamers.

**Gel Mobility Shift Assays**—GMS assays and nuclear extract preparation were done as described previously (15). Extracts were used at a protein concentration of 0.8 mg/ml in the final reaction mixtures unless indicated otherwise. Antisera were diluted 3-fold in phosphate-buffered saline and then extracts were mixed with equal amounts of diluted antisera and incubated on ice for 30 min. Antibodies to C/EBP $\gamma$  were kindly provided by Kathryn Calame (26). Antibodies to C/EBP $\beta$  and C/EBP $\delta$  were purchased from Santa Cruz Biotechnology Inc. Antibodies to GATA-1 were as described (27). The  $\beta$ -CAAT oligonucleotide used and mutants 1, 3, 5, and 7 of this sequence have been described previously (15). The CPS oligonucleotide was (both strands are given 5' to 3') ATTTGTTATGTTGCAACTTGTATGACA/ATGTCATACAAGTTGCCA-CATAACAAA.

**Synthesis of DNA Constructs**—Plasmid constructs containing the  $\beta$ -globin promoter from position -87, which is just upstream of the CAAT box (Fig. 1), to +32 linked to a *H-2K* reporter gene and cloned into the  $\beta$ -globin microlocus vector, which contains the  $\beta$ -globin LCR, have been described previously (15). A promoter containing a deleted

CAAT box was synthesized by the polymerase chain reaction using a 5' primer with the sequence 5'-CGCGGGGATCCTACTCCAGGAG-CAGGG-3'. The underlined sequence represents a *Bam*HI site that was added for cloning purposes and the T following this represents the start of the sequence of the  $\beta$ -globin promoter beginning two nucleotides 3' to the CAAT sequence. The DNA product was linked to the *H-2K* reporter and cloned into the microlocus vector as described (15). To clone the CPS sequences 5' to the various  $\beta$ -globin promoters, the CPS oligonucleotides (given above) were re-synthesized to contain a 5' CG overhang on the first strand and a 5' GATC overhang on the second strand. The oligonucleotides were annealed and then cloned juxtaposed 5' to the promoters using *Bam*HI and *Cla*I restriction sites that are unique when the promoter fragments are cloned into BlueScript vector before they are transferred to the microlocus vector (15). The final products were sequenced to ensure only one copy had inserted and that it was in the correct orientation. The rat LIP expression vector was kindly provided by Ueli Shihler (21).

**Western Blots**—Fifty  $\mu$ g of proteins from nuclear extracts were separated on 13% SDS-polyacrylamide gels, the proteins were transferred electrophoretically to nitrocellulose membranes and the membranes were probed with a 1:2000 dilution of C/EBP $\beta$  antibodies using the ECL system of Amersham Life Sciences.

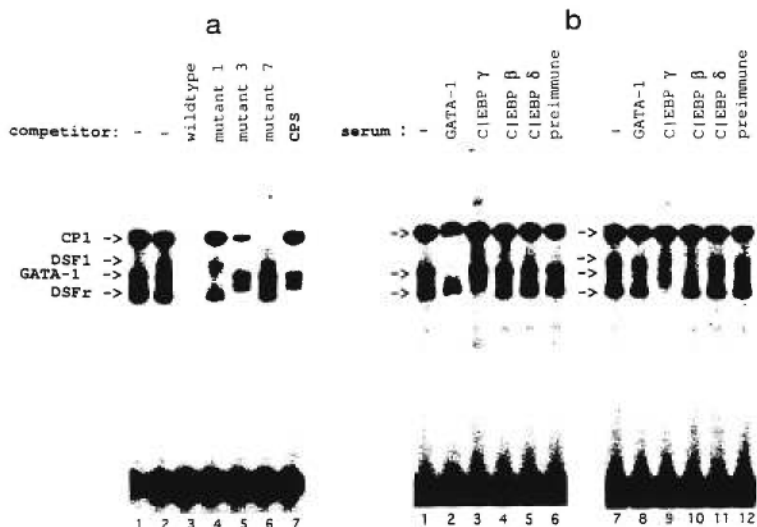
#### RESULTS

**DSF1 and DSFr in MEL Cells Are C/EBP $\beta$  and C/EBP $\gamma$ , respectively**—To clone the corresponding cDNAs for the DSF proteins, a  $\lambda$ gt11 expression library was screened with a DNA probe representing the CAAT box of the  $\beta$ -globin promoter. One positive clone was obtained from  $0.8 \times 10^5$  recombinants screened. This clone contained part of the open reading frame and 3' sequences of the C/EBP $\gamma$  cDNA that was previously cloned from mouse and rat cells (18–20). A  $\lambda$ gt10 cDNA library was screened to isolate the complete coding sequences for C/EBP $\gamma$ . Interestingly, the 5' end of the longest cDNA isolated extended to the G 39 base pairs upstream from the C that was suggested to be the site of transcription initiation for the C/EBP $\gamma$  gene (Ref. 18, data not shown). This result agreed with studies on rat C/EBP $\gamma$ , in which the 5' end of the longest cDNA isolated is highly homologous to the mouse gene and extends 5' to just 2 nucleotides before our longest clone (20). Thus, either the transcription start site for the mouse C/EBP $\gamma$  gene was incorrectly mapped or multiple start sites are used.

To determine if DSF1 and/or DSFr represent C/EBP members, MEL nuclear extracts were incubated with antibodies against various C/EBP proteins before performing GMS assays with the  $\beta$ -globin CAAT box oligonucleotide as the labeled DNA probe. To identify the CP1, DSF1, GATA-1, and DSFr bands on the autoradiograms, competition assays were performed (Fig. 2a) with various mutations of the CAAT box that were previously described (15). These can be summarized as follows: CP1, GATA-1, DSF1 and DSFr bind to the wild type  $\beta$ -CAAT sequence; only GATA-1 recognizes mutant 1; mutant 3 is recognized by DSFr and DSF1, and, to a greatly reduced degree, by CP1; while only CP1 binds to the mutant 7 sequence. Each oligonucleotide will only compete binding of those proteins that bind to it.

With uninduced MEL extracts (Fig. 2b, lanes 1–6) the C/EBP $\gamma$  antisera inhibited formation of the DSFr protein-DNA complex and caused a supershift to form, but it did not affect either GATA-1 or CP1 (lane 3). None of the bands were affected by antisera against C/EBP $\beta$  or C/EBP $\delta$ , or by preimmune serum (lanes 4, 5, and 6, respectively), while GATA-1 antibodies specifically inhibited formation of the GATA-1 complex (lane 2). With induced MEL extracts (Fig. 2b, lanes 7–12) formation of the DSFr complex was again only inhibited by the antibodies against C/EBP $\gamma$ . The C/EBP $\gamma$  antiserum also partially inhibited formation of the DSF1 complex (lane 9). In addition, the DSF1 band was completely inhibited by the C/EBP $\beta$  antiserum (lane 10). However, the DSF1 signal was not affected by GATA-1 antiserum, C/EBP $\delta$  antiserum, or preimmune serum

**FIG. 2. DSF1 and DSFr are C/EBP family members.** In *a*, GMS assays were performed with uninduced (*lane 1*) or induced (all other lanes) MEL extracts with the  $\beta$ -globin CAAT sequence as the DNA probe, in the absence of competitor (-) or in the presence of a 100-fold molar excess of the competitor oligonucleotide shown at the top of each lane. Mutant 1 binds DSF1 and DSFr (CP1 very weakly), mutant 3 binds DSF1 and DSFr (CP1 only), mutant 7 binds CP1 only (15). The CPS sequence binds to DSF1 and DSFr specifically (see text). In *b*, uninduced (*lanes 1-6*) or induced (*lanes 7-12*) MEL extracts were preincubated with the antiserum shown at the top of each lane before performing the GMS assays. -, no additions.



(lanes 8, 11, and 12, respectively), which demonstrated that the effects were specific.

To extend these results, we determined if the DSF1 and/or DSFr factors could recognize DNA sequences known to interact with C/EBP members. A GMS analysis using a sequence from the promoter of the CPS gene, which interacts strongly with C/EBP proteins (28), is shown in Fig. 3. Two bands were detected in induced MEL extracts (Fig. 3*a*, lane 2) that showed exactly the same competition profile as DSFr and DSF1 detected with the  $\beta$ -CAAT probe (compare lanes 2 to 7 in Figs. 2*a* and 3*a*). The faster migrating band was virtually completely inhibited by the anti-C/EBP $\gamma$  serum, but this band was not affected by any of the other sera tested, while the upper band was partially affected by the anti-C/EBP $\gamma$  serum and was completely inhibited by the anti-C/EBP $\beta$  serum (Fig. 3*b*, lanes 6-10). The apparent slight reactivity of the upper band to anti-C/EBP $\delta$  and the slight decrease in the lower band in response to anti-C/EBP $\beta$  and anti-C/EBP $\delta$  was not reproducible. Thus, the two factors in induced MEL extracts that bound to the CPS oligonucleotide gave the same DNA-binding specificity and the same antibody reactivity as DSFr and DSF1 that bind to the  $\beta$ -globin CAAT sequence.

Surprisingly, the CPS oligonucleotide also formed two DNA-protein complexes in GMS assays with uninduced MEL extracts (Fig. 3*a*, lane 1). However, the slower migrating band seen with uninduced extracts migrated at a slightly higher position than the slower migrating band seen with induced extracts (compare lanes 1 and 2 in Fig. 3*a*), and antibody studies indicated that the upper band obtained with uninduced extracts was likely due to two molecules of DSFr binding to a single molecule of the CPS DNA probe. For example, both of the DNA-protein complexes detected with uninduced extracts were almost completely blocked by the anti-C/EBP $\gamma$  serum, whereas neither was affected by the other sera tested, including the C/EBP $\beta$  antiserum (Fig. 3*b*, lanes 1-5 and compare lanes 3 and 4 to lanes 8 and 9). This was in contrast to the upper band obtained with induced extracts, which was completely reactive to C/EBP $\beta$  antiserum.

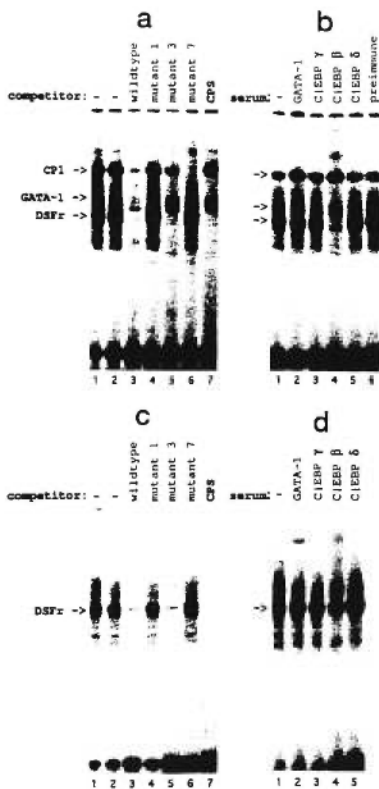
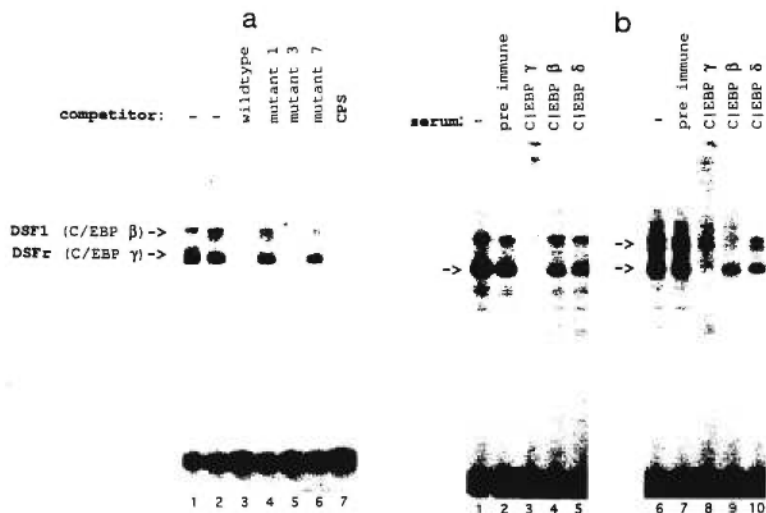
In summary, both DSFr and DSF1 DNA binding activities can be detected with C/EBP-binding sites, as well as with the  $\beta$ -globin CAAT sequence, in GMS assays. DSFr is specifically reactive to C/EBP $\gamma$  antibodies. DSF1 is partly reactive to

MEL cells, while DSF1 is likely C/EBP $\beta$ . If DSF1 is C/EBP $\beta$ , it must be present in part as a homodimer and in part as a heterodimer with the C/EBP $\gamma$  (DSFr) protein. This would explain why it is partly reactive to the C/EBP $\gamma$  antiserum. This interpretation is exactly the same conclusion drawn by Cooper *et al.* (26) for mouse B-cells. Both early and mature B-cells contain active C/EBP $\gamma$ , but only more mature cells contain detectable C/EBP $\beta$  DNA binding activity and a large proportion of the C/EBP $\beta$  protein is present as a heterodimer with C/EBP $\gamma$ , as appears to be the case here for MEL cells.

**K562 Cells Contain LIP, a Truncated Form of C/EBP $\beta$  That Is an Inhibitor of Transcription**—K562 cells were also studied using antisera and GMS assays as outlined above for MEL extracts. In panel *a* of Fig. 4, the identity of the various radio-labeled bands obtained in GMS assays of K562 extracts using the  $\beta$ -CAAT oligonucleotide as the probe was established by competition assays, while the effect of preincubation with the various antibodies is shown in panel *b* of this figure. Previously (15), we found that GATA-1, CP1, and a band migrating near the position of and showing the same competition pattern as DSFr (C/EBP $\gamma$ ) can be detected with K562 extracts. However, the band migrating at the position of DSFr was not at all reactive to the C/EBP $\gamma$  antiserum and treatment with this serum did not cause any super shift to form (lane 3 in Fig. 4*b*). This indicated that none of the complexes seen with K562 extracts is related to C/EBP $\gamma$  (DSFr). On the other hand, the protein-DNA complex seen at the DSFr position was strongly inhibited by the C/EBP $\beta$  antiserum and supershifted bands were observed upon treatment with this antiserum (Fig. 4*b*, lane 4). With the CPS oligonucleotide as probe, only one major labeled protein-DNA complex was obtained in GMS assays of K562 extracts (Fig. 4*c*; minor bands observed were not reproducible). This complex gave the competition profile expected for a C/EBP protein (Fig. 4, compare results in panel *c* to those in panel *a*) and again, this DNA-protein complex was reactive to the C/EBP $\beta$  antibodies specifically (Fig. 4*d*). Thus, the DSF related complex observed with K562 extracts is related to C/EBP $\beta$  (DSF1), not to C/EBP $\gamma$  (DSFr).

The C/EBP $\beta$  mRNA contains three in-frame methionines that can be used as protein initiation codons *in vivo* (21). The first two ATGs are near the 5' end of the message and code for proteins in the 35-40-kDa range. The third ATG is near the 5' end of the open reading frame and codes for a protein of only

**FIG. 3. DSF1 and DSFr specifically bind to a C/EBP recognition sequence.** The figure is exactly as described for Fig. 2, except that the CPS oligonucleotide was used as the labeled DNA probe. In *b*, lanes 1–5 are uninduced MEL extracts and lanes 6–10 are induced MEL extracts.



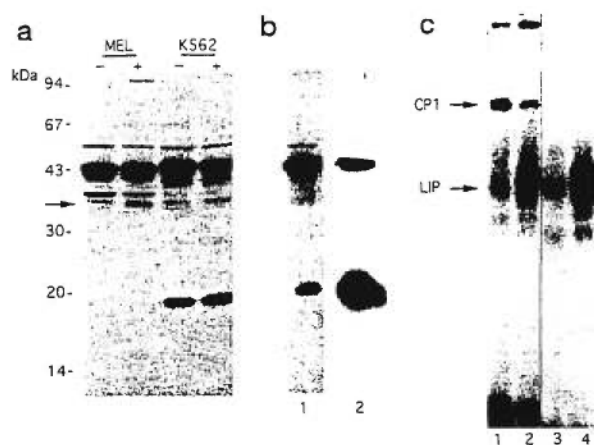
**FIG. 4. A different form of C/EBP $\beta$  is expressed in K562 cells.** EMSA assays were performed with extracts from K562 cells, as described in Figs. 2 and 3, with the  $\beta$ -CAAT oligonucleotide (panels *a* and *b*) or the CPS oligonucleotide (panels *c* and *d*) as the labeled DNA probe. The competitor or antibody conditions are indicated at the top of each lane. Uninduced K562 nuclear extracts were used for lanes 1 in panels *a* and *c*, and extracts from K562 cells treated with 80  $\mu$ M hemin for 48 h were used for all other lanes; uninduced extracts having given identical results.

amide gel electrophoresis near 20 kDa. Since in the EMSA assays the C/EBP-related DNA binding activity in K562 extracts co-migrates with C/EBP $\gamma$  (DSFr) from MEL cells, and C/EBP $\gamma$  itself is a 16.3-kDa protein (18), it was suspected that the K562 DNA binding activity might represent LIP. In West-

ern blots probed with C/EBP $\beta$  antibodies (Fig. 5*a*) a prominent band migrating near 20 kDa was seen with K562 extracts, whereas no protein at this position was detected with MEL extracts. Moreover, in a K562 clone that was stably transfected with an expression vector for the LIP protein of rat, the intensity of this 20-kDa protein detected in Western blots with the C/EBP $\beta$  antibodies was increased substantially (Fig. 5*b*). In EMSA assays of extracts prepared from this clone, an intense DNA-protein complex, migrating near the position of the DSF related band seen with normal K562 extracts, was obtained with both the  $\beta$ -CAAT and CPS sequences as probes (Fig. 5*c*). Note that the major band seen with the LIP expressing clone migrated slightly above the position of the DSF related activity seen with normal K562 cells. This likely represents a difference in migration between the transfected rat LIP protein and the endogenous human K562 protein. Taken together, these data conclusively show that the DSF related DNA binding activity in K562 extracts represents the LIP protein. LIP is present in K562 cells but not in MEL cells.

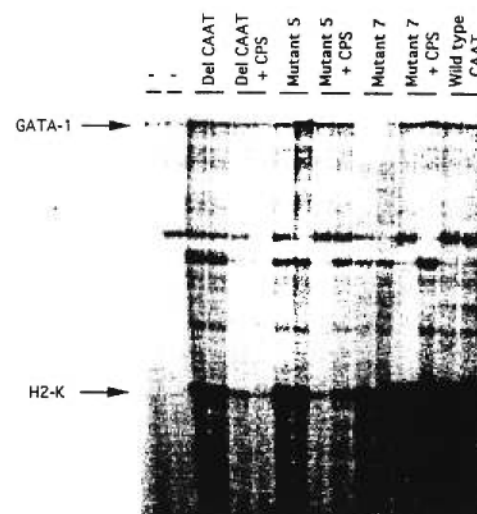
In the Western blot analyses, the C/EBP $\beta$  antiserum also detected major protein bands in the 35–40-kDa range with both K562 and MEL extracts and these were detected equally well with uninduced and induced extracts from either cell type (Fig. 5*a*). It is expected that these represent the larger C/EBP $\beta$  proteins (see above). Yet, C/EBP $\beta$  (DSF1) DNA binding activity can only be detected with induced extracts of MEL cells. It cannot be detected with uninduced MEL extracts, nor with any K562 extracts (see Figs. 2 and 4 for example). Thus, it was thought that only induced MEL cell extracts would contain full-length C/EBP $\beta$  proteins. However, the Western blot analyses suggest that both K562 and MEL cells, whether they are induced or not, contain the full-length proteins. One possibility is that the C/EBP proteins are not active for DNA binding. For example, it has been shown that the DNA binding activity of C/EBP proteins can be modified by phosphorylation (29). In this respect, a minor band migrating near 35 kDa is detected with induced MEL extracts by the C/EBP $\beta$  antibodies in the Western blot analysis, but this band is not detected with uninduced MEL or any K562 extracts (arrow in Fig. 5*a*). Thus, it may be that it is a specific form of the C/EBP $\beta$  protein that is active for DNA binding in induced MEL cells. Further studies will be required to address this possibility.

*An Oligonucleotide That Binds Specifically to the C/EBP Factors Can Reconstitute the  $\beta$ -Globin CAAT Box, but Only in Conjunction with CP1*—Our previous results (15) had sug-



**FIG. 5. K562 cells contain LIP.** In *a*, equivalent amounts of protein extract from K562 or MEL cells, were separated by SDS-polyacrylamide gel electrophoresis and then analyzed by Western blot using antibodies against C/EBP $\beta$ . The position of protein molecular weight standards run in parallel are indicated on the left-hand side, and the source of the extracts are indicated at the top of the lanes. Note that LIP is the 20-kDa protein band that is only seen in the K562 lanes, while the major bands near 40 kDa seen in all lanes are expected to be the larger forms of C/EBP $\beta$ . The arrow to the left indicates the position of a band that is only seen with induced MEL extracts. -, uninduced extracts; +, induced extracts. In *b*, extracts from normal K562 cells (*lane 1*), or from a K562 cell clone transfected with a rat LIP expression vector (*lane 2*) were analyzed by Western blot as in *a*. In *c*, GMS assays were performed using extracts from normal K562 cells (*lanes 1* and *3*) or the LIP expressing clone of K562 cells (*lanes 2* and *4*) with either the  $\beta$ -CAAT oligonucleotide (*lanes 1* and *2*) or the CPS oligonucleotide as the DNA probe. Extracts were used at a final protein concentration of 0.2 mg/ml to emphasize the difference in DNA binding activity.

gested that C/EBP $\gamma$  is involved in LCR enhanced activation of transcription through the  $\beta$ -globin CAAT box in MEL cells. We, therefore predicted that a DNA sequence that binds specifically to C/EBP proteins in MEL cells, such as the CPS oligonucleotide (Fig. 3), may be able to reconstitute transcriptional activity to a  $\beta$ -globin promoter in which the CAAT box had been inactivated. Previously, we had derived several multiple point mutations of the  $\beta$ -globin CAAT sequences (15). These were investigated using constructs that contained the LCR *in cis* to the  $\beta$ -globin promoter (from -87 to +32, which includes just the TATA box and CAAT box motifs (Fig. 1)) linked to a *H-2K* reporter gene. Two of the CAAT box mutations studied, mutant 5, which no longer binds to any proteins, and mutant 7, which only retains CP1 binding, had a 10-fold reduced activity in LCR enhanced transcription in MEL cells, suggesting that in these two cases the  $\beta$ -globin CAAT box is not at all functional (Ref. 15; a summary of these two mutations is shown in Fig. 7). In addition, here we derived an additional construct in which the CAAT box and further upstream sequences were deleted from the  $\beta$ -globin promoter (-71 to +32). This deleted promoter, denoted deleted CAAT, also had a 10-fold reduced activity in MEL cells (see below). We tested whether juxtaposing the CPS oligonucleotide 5' of the mutant 5, mutant 7, or deleted CAAT promoters could restore transcription in MEL cells to the levels seen with the wild type promoter extending to -87. The promoter fragments (with or without the CPS sequence juxtaposed 5' to them) were linked to a *H-2K* reporter gene in the LCR containing microlocus plasmid. Stably transfected populations of MEL cells were selected and then dimethyl sulfoxide-induced populations were assayed by RNase protection (Fig. 6)

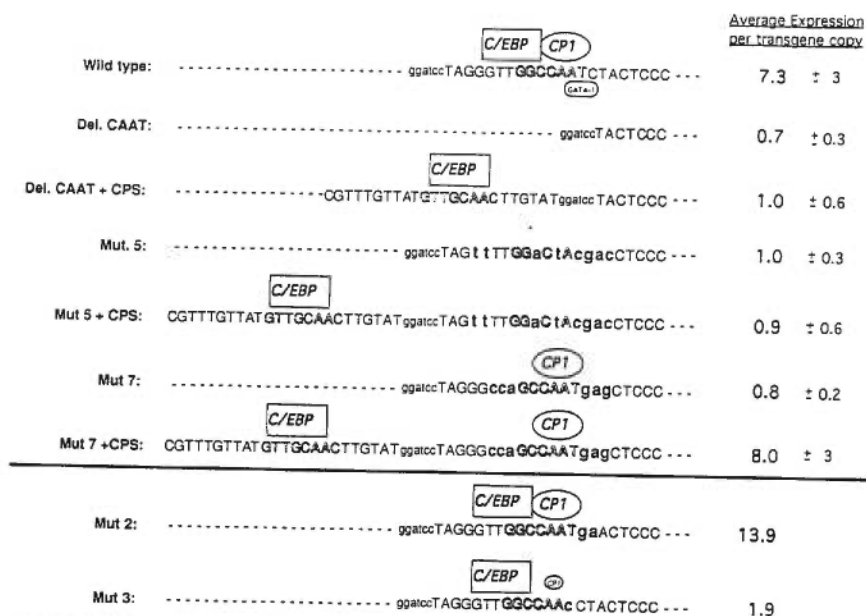


**FIG. 6. RNase protection analysis of transfected MEL cell populations.** MEL cells were transfected with constructs containing the  $\beta$ -globin promoter from -87 to +32 cloned in front of a *H-2K* reporter gene in the microlocus LCR containing vector as described previously (15). The CAAT box of the promoter was either deleted (del CAAT) or mutated to bind no factors (mutant 5) or to bind only CP1 (mutant 7) and the promoters were studied with and without the CPS oligonucleotide cloned 5' to them as indicated at the top of each pair of lanes. Two individually transfected populations of MEL cells for each construct were induced for 3 days with 1.8% dimethyl sulfoxide, then total RNA was isolated and was assayed by RNase protection simultaneously for expression of GATA-1 as a control and for expression of the *H-2K* reporter gene. The position of the signals on the autoradiogram are indicated by the arrows. In the first 2 lanes marked -, untransfected MEL cells were assayed as a negative control for the *H-2K* signal.

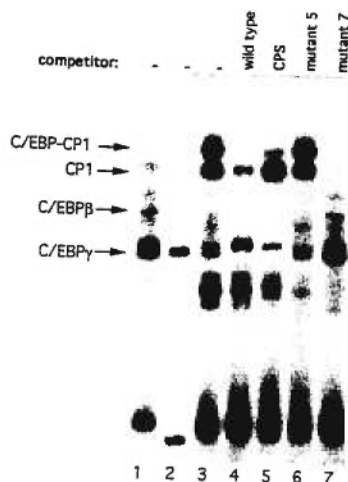
number (which was determined by Southern blot analysis) in each population. A summary of the promoter constructs tested, what factors each binds to, as well as the expression results, is given in the top part of Fig. 7. The CPS sequences did not stimulate activity of either the deleted CAAT or the mutant 5 promoter. In contrast, the CPS oligonucleotide did stimulate the mutant 7 promoter approximately 8-fold, such that the promoter was now as active as the wild type promoter. None of the constructs gave any appreciable activity in K562 cells (results not shown).

To ensure that the ability of the CPS sequence to stimulate transcription of the mutant 7 promoter in MEL cells was not because a new protein binding site had been created, polymerase chain reaction was used to synthesize each promoter fragment containing the CPS sequence (CPS-del CAAT, CPS-mutant 5, and CPS-mutant 7) and these were studied by GMS assays with extracts from MEL cells (Fig. 8). The results showed that in addition to CP1, C/EBP $\beta$ , and C/EBP $\gamma$ , the CPS-mutant 7 DNA fragment formed one additional specific DNA-protein complex. This complex migrated above the CP1 position on the gels (marked C/EBP-CP1 in the figure). Other bands observed were not reproducible and/or were nonspecific. For example, the band migrating below the C/EBP $\gamma$  position was not observed in all experiments and it was nonspecific, as it could be competed to some degree by all competitors tested, especially mutant 5 and mutant 7. Competition assays showed that the band migrating above CP1 represented CP1 and C/EBP $\gamma$  or  $\beta$  binding together on the CPS-mutant 7 DNA fragment. For example, it was competed by the CPS, wild type  $\beta$ -CAAT, and mutant 7 oligonucleotides, but it was not inhib-



CAAT/Enhancer-binding Proteins in  $\beta$ -Globin Gene Expression

**FIG. 7. Both CP1 and C/EBP $\gamma$  are required for wild type levels of expression.** In the top part of the figure, the  $\beta$ -globin promoter constructs studied in this report are summarized along with the factors that bind to them. The CAAT box region is outlined, mutations from the wild type  $\beta$ -globin sequence are shown in bold lower case letters, and a BamHI restriction enzyme site (ggaacc) that was used to clone in the CPS sequence is shown in smaller lower case letters. The consensus C/EBP-binding site in the CPS sequence is also outlined. To the right of the figure, the average level of expression per transgene copy ( $\pm$ , average variation) of each construct, as measured by RNase protection and Southern blot analysis, is given. Each value is an average from the two different transfected populations shown in Fig. 6, except for the mutant 5 + CPS, mutant 7, and mutant 7 + CPS constructs, for which three, three, and four different populations were studied, respectively. (Note that the expression values given here may not appear to correspond to the expression levels in Fig. 6. However, the calculated values have been corrected for copy number since the LCR confers copy number-dependent expression onto transgenes in MEL cells (36). For example, deleted CAAT + CPS would appear to have a much lower level of expression than mutant 5 in Fig. 6. However, the average copy number for the mutant 5 populations was approximately 5-fold higher than for the deleted CAAT + CPS populations. Upon correction, they turn out to have approximately the same level of expression.) In the bottom of the figure, the expression results for mutant 2 and mutant 3, as described previously (15) are summarized for comparative purposes. These expression values have been corrected by a factor of 1.28 to reflect the difference in wild type values obtained here (7.3) versus the values obtained previously (9.4).



**FIG. 8. GMS assays of promoter fragments.** The  $\beta$ -globin promoter fragments containing the CPS sequences (mut 5 + CPS, lane 1; del. CAAT + CPS, lane 2; and mut 7 + CPS, lanes 3-7) were synthesized using polymerase chain reaction and then used as the labeled DNA probe in GMS assays with MEL extracts. The assays were done in the presence of approximately a 100-fold excess of the oligonucleotide shown at the top of each lane. -, no competitor added.

mutant 7 (competes CP1 specifically) the bands at the positions of C/EBP $\gamma$  and C/EBP $\beta$  increased (lane 7). Moreover, when the CPS-mutant 7 fragment was competed with the CPS oligonucleotide, the GMS pattern observed exactly coincided with the

pattern obtained for CPS-mutant 5 and CPS-deleted CAAT (compare lane 7 to lanes 1 and 2 in Fig. 8), showing that the only extra protein that binds to CPS-mutant 7, relative to these two other promoter fragments, is CP1.

## DISCUSSION

In previous studies (15) we showed that whereas the wild type CAAT box of the  $\beta$ -globin promoter can stimulate LCR enhanced transcription 10-fold in MEL cells, the CAAT box is completely inactive in K562 cells. Mutation and stability analyses indicated that a factor that was then denoted DSFr is involved in the CAAT box function in MEL cells. In the present report, we showed that the C/EBP $\gamma$  transcription factor represents the DSFr DNA binding activity detected in MEL cells. However, no C/EBP $\gamma$  DNA binding activity could be detected with extracts from K562 cells, showing that K562 cells do not contain active C/EBP $\gamma$ . We cannot say whether or not K562 cells contain inactive C/EBP $\gamma$  protein, as the antibodies to this protein did not work well in Western blot analyses. However, K562 cells were found to contain substantially less C/EBP $\gamma$  mRNA than MEL cells (not shown), suggesting that the regulation might be at the level of transcription. Further studies will be necessary to clarify this point.

The absence of active C/EBP $\gamma$  in K562 cells might be expected to account for the inability of the  $\beta$ -globin CAAT box to function in cis with the LCR in these cells. However, with clones of K562 cells that were stably co-transfected with a  $\beta$ -globin promoter-LCR reporter plasmid (the promoter from -87 to +32 with an unmutated CAAT box) and a C/EBP $\gamma$  expression plasmid, we failed to detect any expression of the

linked *H-2K* reporter gene (not shown). This suggests that the absence of active C/EBP $\gamma$  in K562 cells, alone, may not account for the inactivity of the CAAT box in these cells. However, we also found that K562 cells, but not MEL cells, express active LIP. LIP has been shown to be an inhibitor of transcription (21). LIP does not contain transcriptional activation domains that are located in the N-terminal half of the larger C/EBP $\beta$  proteins (30). In binding to its cognate sequence, LIP will prevent potential transcription activators, including other C/EBP members, from binding there. Therefore, it is likely the combination of not containing active C/EBP $\gamma$  protein and having LIP that keeps the CAAT box tightly silent in K562 cells. We also cannot rule out the possibility that CP1 is defective in K562 cells (see below).

*C/EBP $\gamma$  May Cooperate with CP1*—Our results showed that a CPS oligonucleotide, which binds only to C/EBP proteins in MEL cells, was unable to restore activity to the mutant 5 and the deleted CAAT  $\beta$ -globin promoters, but was able to restore wild type levels of activity to the mutant 7 promoter. In terms of factors that normally bind to the wild type CAAT sequence, the deleted CAAT and mutant 5 sequences do not bind to any factors, whereas mutant 7 retains full binding of CP1, but does not bind to any of the other factors in MEL cells (Ref. 15; summarized in Fig. 7). Thus, these expression results suggest the possibility that C/EBP $\gamma$  cooperates with CP1 to enhance transcription through the  $\beta$ -globin CAAT box. In fact, this interpretation would concur with our previous findings (15). A CAAT box mutation (mutant 2) that retained binding of C/EBP $\gamma$  (DSFr), C/EBP $\beta$  (DSF1), and CP1 stimulated transcription above wild type levels, whereas a mutation (mutant 3) that also retained binding of C/EBP $\gamma$  and C/EBP $\beta$ , but had a greatly reduced affinity for CP1, only stimulated transcription slightly above background (these results are summarized in the bottom part of Fig. 7 for comparison). This again could be explained by the difference in CP1 binding to the two mutants. Note that CP1 binding alone has no stimulatory activity above background (Fig. 7). Thus, it appears that C/EBP $\gamma$  and CP1 may indeed collaborate. However, this possibility does cause a dilemma. The binding sites for C/EBP $\gamma$  and CP1 in the  $\beta$ -globin CAAT box overlap each other (15). Hence, it would be expected that the binding of the two proteins is mutually exclusive. Indeed, we have been unable to detect both proteins binding together onto the wild type CAAT box. However, the situation is similar to what has been described for C/EBP factors activating the albumin promoter (31). The C/EBP DNA-binding site is also tightly juxtaposed to a binding site for CP1 in the albumin promoter. Although the binding of both factors is not totally mutually exclusive in this case, C/EBP binding interferes with CP1 binding. Yet, binding of both proteins is again necessary for transcriptional activation (31). Therefore, in some cases C/EBP proteins act in synergism with CP1, even when the factors might interfere with each others binding. If C/EBP $\gamma$  and CP1 do collaborate to enhance transcription through the  $\beta$ -globin CAAT box, the possibility that both proteins can bind to the CAAT box together under *in vivo* conditions cannot be excluded. It would also seem possible that C/EBP $\gamma$  and CP1 are necessary for different steps in the formation of the preinitiation transcription complex and that each protein's function is required at a different time, *i.e.* the binding of both proteins may be necessary, but may not have to occur simultaneously. Further experiments will be necessary to determine conclusively whether C/EBP $\gamma$  and CP1 do collaborate at the  $\beta$ -globin CAAT box, and, if so, what the function of each protein is in enhancing  $\beta$ -globin gene expression.

*Does the Differential Expression of C/EBP Proteins in MEL and K562 Cells Reflect a Developmental Role?*—As K562 cells

only express the endogenous  $\epsilon$ - and  $\gamma$ -globin genes and not the adult stage-specific  $\beta$ -globin gene, they have been considered a model for the embryonic-fetal stages of erythroid development (see Ref. 11, for example). On the other hand, MEL cells are considered a fetal-adult model. Thus, it seems possible that the differential expression of C/EBP factors in K562 cells (active LIP and no active C/EBP $\gamma$ ) versus MEL cells (active C/EBP $\gamma$  and also active full-length C/EBP $\beta$  in induced cells) and the functioning of the  $\beta$ -globin CAAT box only in MEL cells might be related to a developmental specificity. However, it has previously been shown that both C/EBP $\gamma$  and C/EBP $\beta$  are expressed in rat yolk sac and fetal liver, at least at the RNA level (20). This sheds some doubt on the possibility that the differential expression of C/EBP factors in MEL and K562 cells is related to developmental switching of  $\beta$ -globin gene expression.

On the other hand, the differential expression of C/EBP factors has been correlated with cellular differentiation and cell maturation in a variety of systems. For example, the differential expression of three C/EBP members,  $\alpha$ ,  $\beta$ , and  $\delta$ , has been directly related to adipocyte differentiation (16, 32). The differential expression of LIP and full-length C/EBP $\beta$  has been correlated with liver regeneration. Whereas more LIP is present in actively dividing liver cells, more full-length C/EBP protein is present in quiescent cells and this is directly related to cell growth (33). In this sense, it is important to note that MEL cells and K562 cells also appear to represent different precursors in the erythroid pathway. For example, MEL cells are completely committed to becoming red cells, but this is not the case for K562 cells. If K562 cells are treated with phorbol esters, they lose their erythroid characteristics and induce megakaryocyte markers (34, 35). Moreover, we have obtained initial results that suggest that de-regulated expression of C/EBP factors in K562 cells can alter both erythroid and megakaryocytic differentiation of these cells.<sup>2</sup> Thus, we suspect that the differential expression of C/EBP factors in K562 cells versus MEL cells, as well as the inactivity of the  $\beta$ -globin CAAT box in K562 cells, may be related to the different precursor status of these cells.

*Acknowledgments*—We thank Kathryn Calame (Columbia University) for her kind gift of antiserum against C/EBP $\gamma$  and we thank Ueli Schibler (Switzerland) for the LIP expression vector. We also thank Gordon Shore (Montreal) for providing the initial CPS oligonucleotides and Rashmi Kothary (this institute) for critical reading of the manuscript.

#### REFERENCES

1. Stamatoyannopoulos, G., and Nienhuis, A. W. (1987) *Developmental Control of Globin Gene Expression*. Alan R. Liss, Inc., New York
2. Dillon, N., and Grosfeld, F. (1993) *Trends Genet.* **9**, 134–137
3. Behringer, R. R., Ryan, T. M., Palmiter, R., Brinster, R. L., and Townes, T. M. (1990) *Genes & Dev.* **4**, 380–389
4. Enver, T., Raich, N., Ebens, A. J., Papyannopoulou, T., Costantini, F., and Stamatoyannopoulos, G. (1990) *Nature* **344**, 309–313
5. Peterson, K. R., and Stamatoyannopoulos, G. (1993) *Mol. Cell. Biol.* **13**, 4836–4843
6. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., and Grosfeld, F. (1991) *Genes & Dev.* **5**, 1387–1394
7. Kim, C. G., Epner, E. M., Forrester, W. C., and Groudine, M. (1992) *Genes & Dev.* **6**, 928–938
8. Bungert, J., Davé, U., Lim, K.-C., Lieuw, K. H., Shavit, J. A., Liu, Q., and Engel, J. D. (1995) *Genes & Dev.* **9**, 3083–3096
9. Dillon, N., Strouboulis, J., and Grosfeld, F. (1995) in *Molecular Biology of Hemoglobin* (Stamatoyannopoulos, G., ed) pp. 23–28, Intercept Ltd., Andover, United Kingdom
10. Antoniou, M., and Grosfeld, F. (1990) *Genes & Dev.* **4**, 1007–1013
11. Donze, D., Townes, T. M., and Bieker, J. J. (1995) *J. Biol. Chem.* **270**, 1955–1959
12. Feng, W. C., Southwood, C. M., and Bieker, J. J. (1994) *J. Biol. Chem.* **269**, 1493–1500
13. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosfeld, F. (1995) *Nature* **375**, 316–318
14. Perkins, A. C., Sharpe, A. H., and Orkin, S. H. (1995) *Nature* **375**, 318–322
15. Delvoe, N. L., Destroismaisons, N. M., and Wall, L. A. (1993) *Mol. Cell. Biol.* **13**, 6969–6983

<sup>2</sup> L. Wall and N. Delvoe, unpublished observations.

16484

CAAT/Enhancer-binding Proteins in  $\beta$ -Globin Gene Expression

16. Cao, Z., Umek, R. M., and McKnight, S. L. (1991) *Genes & Dev.* **5**, 1538-1552
17. Williams, S. C., Cantwell, C. A., and Johnson, P. F. (1991) *Genes & Dev.* **5**, 1553-1567
18. Nishizawa, M., Wakabayashi-Ito, N., and Nagata, S. (1991) *FEBS Lett.* **282**, 95-97
19. Roman, C., Platero, J. S., Shuman, J., and Calame, K. (1990) *Genes & Dev.* **4**, 1404-1415
20. Thomassin, H., Hamel, D., Bernier, D., Guertin, M., and Belanger, L. (1992) *Nucleic Acids Res.* **20**, 3091-3098
21. Descombes, P., and Schibler, U. (1991) *Cell* **67**, 569-579
22. Ron, D., and Habner, J. F. (1992) *Genes & Dev.* **6**, 439-453
23. Akira, S., Ieshiki, H., Sugita, T., Tanbe, O., Kinoshita, Y., Nakajima, T., Hirano, T., and Kishimoto, T. (1990) *EMBO J.* **9**, 1897-1906
24. Kinoshita, S., Akira, S., and Kishimoto, T. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1473-1476
25. Vinson, C. R., LaMarco, K. L., Johnson, W. H., Landschultz, W. H., and McKnight, S. L. (1988) *Genes & Dev.* **2**, 801-806
26. Cooper, C., Johnson, D., Roman, C., Avitahl, N., Tucker, P., and Calame, K. (1992) *J. Immunol.* **149**, 3225-3231
27. Aumont, F., Trudel, P., and Wall, L. (1993) *Differentiation* **52**, 169-176
28. Lagace M., Goping, I. S., Mueller, C. R., Lazzaro, M., and Shore, G. C. (1992) *Gene (Amst.)* **118**, 231-238
29. Mahoney, C. W., Shuman, J., McKnight, S. L., Chen, H.-C., and Huang, K.-P. (1992) *J. Biol. Chem.* **267**, 19396-19403
30. Kowenz-Leutz, E., Twamley, G., Anisau, S., and Leutz, A. (1994) *Genes & Dev.* **8**, 2781-2791
31. Milos, P. M., and Zaret, K. S. (1992) *Genes & Dev.* **6**, 991-1004
32. Umek, R., Friedman, A. D., and McKnight, S. L. (1991) *Science* **251**, 288-292
33. Buck, M., Turler, H., and Chojkier, M. (1994) *EMBO J.* **13**, 851-860
34. Lumelsky, N., and Forget, B. G. (1991) *Mol. Cell. Biol.* **11**, 3528-3536
35. Tetteroo, P. A. T., Massaro, F., Mulder, A., Schreuder-van Gelder, R., and von dem Borne, A. E. G. Kr. (1984) *Leuk. Res.* **8**, 197-206
36. Blom van Assendelft, Hanscombe, O., Grosveld, F., and Greaves, D. R. (1989) *Cell* **56**, 969-977