IN VITRO AND *IN VIVO* ANALYSIS OF THE ESTABLISHMENT AND MAINTENANCE OF β-GLOBIN LOCUS CHROMATIN STRUCTURE

By

PADRAIC P. LEVINGS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

This work is dedicated to my father, Thomas J. Levings. He has been the most important force in my life.

ACKNOWLEDGMENTS

There are many people I wish to thank for contributing to my completion of my doctoral degree. Without them I would have never found fruition. The primary motivation and balance in this endeavor have been my parents. I would like to thank my mother for preaching patience and giving me comfort ion times when it was needed. My father has been the most important figure in my life. He is the reason I am the person I am today and when I could not find a reason for going on, his advice concerning my setbacks and accolades of my success was more than enough motivation.

I would also like to thank my mentor and friend Jörg Bungert. I think if I would have chosen another place to call home to my studies I would never have survived the past several years. He has been not only a mentor, but a role model as well. I appreciate his patience with my idiosyncrasies. He was able to direct my exuberance into something creative, and allowed me to go to the gym in the middle of the day. I am certain I will always look to him for guidance in the future.

My lab mates during the past seven years deserve a large degree of credit for my success, specifically, Kelly Leach, Karen Vieira, and Valerie Crusselle. Kelly Leach was my best friend for the first five years of graduate school until her graduation. She was a source of constructive criticism as well as praise. Karen Vieira was a stable presence in the lab and also a very close friend. She had a strong personality and a sharp, focused sense of purpose. She was a good friend not only in the lab, but also at home as we were housemates for more than a year. I would like to thank Valerie Crusselle for just being

Val. She has put up with my teasing for the last year and will have to do so for the next. It is always a pleasure to be in the lab when she is there; her outlook and voice are always a refreshing change from my own.

I am forever indebted to the members of my thesis committee, Drs. Laipis, Lewin, Muczycka, and Yang. They are the most intelligent men I have ever met. I never felt that a committee meeting was an onerous task; I always felt that it was a positive experience. Their approval of my research has been one of the most gratifying experiences of my life. I do not think I could find a more worthy group of peers.

I would also like to thank Thomas Conlon and Leah Cochran, the two people whom I spend the most time with outside of the lab. Their constructive and often unsolicited criticisms were often the kick in the right place I needed to get me off the couch and back to work. They have been inspirational.

Lastly I would like to thank Athena (my Siberian husky), Ogre, Thor, Achilles, Starburst, Lemonade, Sunfire, and Tigger (the Australian bearded dragons), and Turbo (the African spade foot tortoise), for the welcome distractions they have provided during the construction of this work.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iii

LIST OF FIGURES vii

ABSTRACT ix

CHAPTER

1 INTRODUCTION 1

Background and Significance 1
Chromatin Structure and Gene Regulation 3
Development and Hematopoiesis 6
In vitro Differentiation of Embryonic Stem Cells 9
Globin Gene Organization 11
Globin Gene Proximal Regulatory Elements and Transcription Factors Involved in Erythropoiesis 13
Locus Control Regions and Hemoglobin Gene Switching 17
Summation 22

2 MATERIALS AND METHODS 26

3 ESTABLISHMENT OF PATTERNS OF HISTONE MODIFICATION AND FACTOR RECRUITMENT WITHIN THE β–GLOBIN LOCUS DURING *IN-VITRO* DIFFERENTIATION OF MURINE EMBRYONIC STEM CELLs 36

Introduction 36
Results 39
Discussion 45

4 COMPOSITE β-GLOBIN LOCUS TRANSGENES EXHIBITING POSITION-INDEPENDENT EXPRESSION IN TRANSGENIC MICE 50

Introduction 50
Results 53
Discussion 59

5 SITE-SPECIFIC INTEGRATION OF MODIFIED YEAST ARTIFICIAL CHROMOSOMES IN MOUSE EMBRYONIC STEM CELLS BY RECOMBINASE MEDIATE CASSETTE EXCHANGE 64

Introduction 64
Results and Discussion 70

6 SUMMATION OF RESULTS 78

Functional Significance of the β-globin Locus Control Region 78
In Vitro Differentiation of Murine Embryonic Stem Cells 79
Site-Specific Integration and Position-Independent Expression of Human β-globin Transgenes 80

LIST OF REFERENCES 83

BIOGRAPHICAL SKETCH 93

LIST OF FIGURES

Figure page

1-1 The human β-globin locus....13

3-1 PCR analysis of DNase I treated and reverse-transcribed total RNA extracted from differentiating embryonic stem cells at the indicated time points. 40

3-2 Interaction of transcription factors and RNA polymerase II with the β-globin locus 41

3-3 Quantitative analysis of Pol II binding and histone modifications within the β-globin locus at "Day 0" in murine ES cells....43

3-4 Quantitative analysis of Pol II binding and histone modifications within the β-globin locus at "Day12" in differentaited murine ES cells 45

3-5 Fold change in factor binding and histone modifications at the murine β-globin locus during *in vitro* differentiation 47

3-6 Recruitment of transcription complexes and chromatin structure alterations at the β-globin locus in undifferentiated precursors and definitive erythrocytes....49

4-1 Structure and expression of the 432β4 transgene....54

4-2 Location of the human S1 integration site in the mouse genome 56

4-3 Expression analysis of a β-globin construct containing insulator sequences as well as LCR elements HS2 and 3 with their flanking DNA in transgenic mice. 57

4-4 Analysis of the integration site, β-globin gene expression, and DNase I HS sites in transgenic line 43f2β4B. 59

4-5 Relative expression levels of two β-globin transgenic constructs 61

5-1 Principle of RMCE....67

5-2 Structure of gene targeting cassettes for RMCE in ES cells....71

5-3 YAC constructs for RMCE. 72

5-4 *LoxP* sequences used in targeting and exchange cassettes. 73

5-5 Schematic representation of the structure of integrated plasmids used to generate YAC exchange cassettes. 74

5-6 Southern analysis of yeast transformants.. 75

5-7 Southern Analysis of human β-globin locus exchange constructs. 76

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

IN VITRO AND *IN VIVO* ANALYSIS OF THE ESTABLISHMENT AND MAINTENANCE OF β-GLOBIN LOCUS CHROMATIN STRUCTURE

By

Padraic P. Levings

May 2005

Chair: Jörg Bungert
Major Department: Biochemistry and Molecular Biology

Thalassemias are mutations in one or more of the globin genes causing defective hemoglobin synthesis and can result in severe anemia. They are the most common single gene disorders and affect thousands of people worldwide. Current therapies for these disorders require regular red cell transfusions coupled with iron chelation therapy to reduce the effects of iron overload that can cause organ failure. Bone marrow transplantation is the only currently available cure but is limited to those patients with matched siblings and carries an 80% mortality rate. Whereas, extensive knowledge concerning the transcriptional regulation of both the α- and β-globin gene loci has been accumulated, the cis-acting regulatory elements required to achieve therapeutic levels of gene expression in a therapeutic context have yet to be fully elucidated.

To this end we have investigated the association of various transcription factors and chromatin modifiers at the mouse β-globin locus during the process of *in vitro* embryonic stem (ES) cell differentiation. From this we hope to gain insight into the

elements/factors required for the establishment and maintenance of a transcriptionally active chromatin domain. Secondly, we have generated several lines of transgenic mice harboring various regulatory elements from the human β-globin locus to test their ability to imbue position-independent, copy number-dependent expression characteristics on a *cis*-linked β-globin gene.

Results from the *in vitro* analysis of the association of transcription factors and histone modification during differentiation of murine embryonic stem cells indicates that the locus control region acts as the primary site of recruitment of chromatin modifying activities and transcription complexes prior to gene activation. These activities are already present at this element in undifferentiated cells and do not appear at gene promoters later in development. Our *in vivo* studies show that the inclusion of sequences flanking the core regions of HS2 and 3 of the human locus effectively enhance transgene expression above that of constructs lacking these elements, even when integrated to generally repressive chromosomal regions. Furthermore, inclusion of boundary elements from the chicken locus may shield these constructs from the influence of chromatin at the site of integration. These results not only provide insight into the mechanisms involved in the activation of hematopoietic genes but may serve to provide more effective means of treatment of diseases associated with globin gene expression.

CHAPTER 1
INTRODUCTION

Background and Significance

Hemoglobin is a tetrameric protein composed of two α– and two β–like subunits ($\alpha_2\beta_2$) that functions to transport the oxygen in all vertebrate species from the lungs, gills, and skin to that of the capillaries for use in respiration. Hemoglobin was one of the first proteins whose molecular mass was accurately determined, whose existence was associated with a physiological function and the first in which a single point mutation was shown to result in an amino acid change (1). Whereas the hemoglobin protein has served as a model for cooperative interactions and allosteric mechanisms of regulation, it is the complex program of developmental stage- and tissue specific regulation of the genes that make up the hemoglobin tetramer that has received the most focus in the past two decades. Perturbations of this transcriptional program can result in a form of severe anemia occurring early in life and associated with splenomegaly and characteristic bone marrow changes termed thalassemia (2). Thalassemias are classified as a heterogeneous group of inherited disorders of hemoglobin synthesis classified by reduced or absent synthesis of one or more of the globin peptides of hemoglobin and are the most common single gene disorders in the world (3).

Homozygous β thalassemia is a condition in which there is a lack or severely reduced synthesis of β–like globin chains ultimately resulting in erythroid cell death. Current treatment of this disorder includes regular red cell transfusions that often lead to

iron overload and require chelation therapy. The observation that elevated levels of fetal hemoglobin synthesis beyond the perinatal period in some patients results in an amelioration of the severity of the anemia has led to the exploration of alternative treatments that utilize pharmacological intervention (3). These treatments involve the administration of various chemotherapeutic agents such as 5-azacytidine (4), hydroxyurea, and butyric acid analogues (5) aimed at increasing the amount of fetal hemoglobin production. However, concerns of carcinogenesis, myelotoxicity, and lack of effectiveness have made these treatments less than satisfactory for treatment of these conditions. Current experimental procedures involve the use of viral vectors to transduce corrected forms of the gene into stem cells and bone marrow engraftment (6). While gene therapy appears promising we have yet to fully understand the means for ensuring that the transferred gene is properly expressed. In order to do this we must fully elucidate the mechanisms and sequences required to properly regulate globin gene expression through detailed studies.

In humans the genes encoding the α– and β–globin loci are located on chromosomes 16 and 11, respectively. Both the α– and β–globin gene clusters are multi-gene loci that are spatially arranged in a linear fashion along the chromosome and the order of this arrangement reflects the order of their expression in erythroid cells during development. As the developing embryo progresses from the embryonic to fetal and from the fetal to adult stages of development there is a corresponding switch in the type of globin mRNA synthesized, resulting in changes of the peptide composition of the hemoglobin molecule. These sequential changes in gene expression that occur during development are defined as hemoglobin switching and have been shown to be regulated

primarily at the transcriptional level (3). Understanding how these transcriptional switches are regulated and the minimal genetic elements required to recapitulate them in a therapeutic context would allow for the development of more effective treatments for the thalassemic diseases.

Chromatin Structure and Gene Regulation

In order for eukaryotic organisms to fit the vast amounts of DNA encoding their genomes into a nucleus only a few microns in diameter it is required that the DNA be condensed and packaged into nucleoprotein structures composed of histone and non-histone proteins. These structures form the dynamic chromatin polymers that make up chromosomes. Nucleosomes are the basic units of chromatin and consist of ~147 bp of DNA wrapped around two subunits each of the highly conserved H2A, H2B, H3 and H4 core proteins. These nucleosomal units are compacted into increasingly more condensed structures ranging from the 10nm filament to that of chromosomes. Within the eukaryotic genomes the chromosomes themselves contain areas differing in their level of compaction and distinguished by varying degrees of sensitivity to nucleases. Nuclease sensitivity is thought to give an indication of the "openness" or degree of decondensation of chromatin. Areas that are nuclease sensitive are enriched for transcriptionally active sequences and called euchromatic, whereas areas that are comparatively resistant to nuclease digestion are usually transcriptionally silent and heterochromatic. Nucleosomes are stable structures under physiological conditions, are able to self-associate, and are extremely resistant to physical perturbations. Furthermore, in order to follow the left-handed spiral formed by the histone fold domains the DNA must be severely distorted resulting in topological constraints on the accessibility of nucleosome associated DNA sequences to various trans-acting factors. A fact made evident by the

repressive nature of histones to transcription both *in vitro* and *in vivo*. A central question arising from these observations is how such a transcriptionaly inhospitable structure could act as the physiological template of genetic information in eukaryotes. Are nucleosomes static structures that are functionally irrelevant to gene expression, simply functioning to package DNA into more manageable dimensions? It was not until Littau and colleagues noted an association between histone acetylation and transcription in eukaryotes that the idea of the passive nature of chromatin was challenged and its relevance to regulating gene expression explored (7).

In addition to the histone fold domains are external tail domains found at the N-termini of all four core proteins. These domains are not required for nucleosomal assembly or positioning although they are required for the formation of higher order chromatin structures and may mediate internucleosomal contacts (8). An extensive body of literature has been amassed studying the affects of various histone tail modifications on transcriptional regulation, replication, repair, recombination, cell cycle progression, and chromosome segregation. These modifications include acetylation, methylation, phosphorylation, ubiquitinylation, ADP-ribosylation, and sumoylation, the various combinations and permutations of which constitutes a language termed the "histone code" (9). This code of covalent modifications has been shown to be an integral and increasingly complex aspect of DNA metabolism particularly in terms of gene expression. It is predicated on the hypothesis that specific modifications of histone tail residues would result in the modulation of the affinities of chromatin-associated proteins for their targets. It also states that modifications may be interdependent, such that one modification can affect another and that interdependent modifications need not be on the

same tail. Lastly, it predicts that domains of higher order chromatin structure may be directly regulated by the local concentration of differentially modified nucleosomes (10, 11). These ideas have been thoroughly integrated into research studying changes in gene expression that occur without alterations in DNA sequence, or epigenetic changes.

The best characterized epigenetic modifications are that of lysine acetylation and methylation as well as that of serine phosphorylation as the systems catalyzing these modifications have been identified (12, 13). Lysine acetylation and deacetylation of core histones is catalyzed by histone acetyltransferases (HATs) and deacetylases (HDACs), respectively and has been shown to be causally linked to changes in transcriptional activity (12, 14). These modifications occur on specific lysine residues in the tails of both histones H3 and H4 and it is thought that acetylation of these lysine residues decreases the overall positive charge of the tail domain. This causes a decrease in its affinity for DNA and thus increases the accessibility of other factors to their cognate binding sequences. Additional research has also shown that various transcription factors and chromatin remodeling complexes recognize specific acetylation marks and bind to these regions using a bromodomain. Two such factors are PCAF and $TAF_{II}250$ both of which contain HAT activity and that act as activators of transcription (15, 16)

Histone methylation, similar to acetylation, occurs on specific lysine residues in the amino terminal tails of the core histones. Until recently it was thought that histone methylation was irreversible, as no histone demthylase had been characterized. A report by Shi *et al.* identifying the histone demthylase, LSD1 showed that there exist enzymes capable of removing this modification. The fact that it appears to specific for mono and dimethylated lysine four of histone H3 raises many question about its role in

transcriptional regulation (17). Histone methylation is catalyzed by histone methyl transferases (HMTs). It is thought that in some cases methylation of histones is associated with heterochromatin formation as a recent report showed that the *Drosophila* protein Su(var)3-9 localized to heterochromatin and that human and *S. pombe* homologues of this protein are histone H3 lysine 9 specific HMTs *in vitro* (18, 19). In addition, research shows that two heterochromatin associated proteins, HP1 and Swi6, contain a chromodomain that mediates preferential binding to methylated H3 lysine 9 *in vitro* (20, 21). In contrast to H3 lysine 9 methylation, methylation of H3 lysine 4 has been associated with transcriptional activation (20, 21). These data support the idea that it is not only the type of modification but also its location and the state of modification of neighboring residues of the same and adjacent core histones that dictate the transcriptional state of a genomic locus.

Development and Hematopoiesis

The use of pluripotent embryo- and adult tissue-derived cells for the therapeutic treatment of damaged or dysfunctional tissues has become the focus of a considerable amount of research. In order for stem cell therapy to become a realistic means of treatment and to accurately manipulate lineage choice and differentiation *in vitro* a detailed description of the molecular and cellular events involved must be developed.

Development is a complex and exquisitely regulated process that requires the proper spatio-temporal regulation of cell growth and signaling. Around the time of implantation the embryo is composed of three distinct cell types: the trophectoderm, the primitive endoderm, and the inner cell mass (ICM). The trophectoderm gives rise to the extraembryonic tissues of the placenta, while the primitive endoderm forms both visceral and parietal endoderm that lines the yolk sac (22). The ICM forms the epiblast that will

ultimately give rise to the embryo and other extra embryonic tissues (23). Gastrulation of the embryo results in the formation of the three primary germ layers of the embryo from cells of the epiblast. These cells are the progenitors of all the tissues of the fetus.

One of the earliest systems to be established during mammalian development is the hematopoietic system. In the mouse hematopoietic development initiates at approximately 7-7.5 days post coitus (d.p.c.) from cells of mesodermal origin that have migrated through the primitive streak (24). These cells are then allocated into discrete structures known as blood islands that are composed of a central focus of developing hematopoietic cells surrounded by primitive angioblasts. The close spatial and temporal development of these lineages has led to the proposal that they arise from a common progenitor, the "hemangioblast" (25). This idea is supported by the fact that these two cell lineages share a number of expressed genes including CD34, flk-1, flt-1, TIE2, scl/tal-1, GATA-2, and PECAM-1. In addition, several genes shown to be involved in the regulation of hematopoietic development (scl/tal-1, GATA-2, rbtn2) are expressed prior to the establishment of yolk sac hematopoiesis. This fact indicates that certain populations of early mesodermal precursor cells may have already initiated the genetic program that predisposes them to the hematopoietic lineage at a time shortly after gastrulation (26). At approximately 8.5 to 9 d.p.c. the vasculature forms and the heart begins to beat. At this time hematopoietic cells begin to circulate and seed the fetal liver, which becomes the predominant site of hematopoiesis until the bone marrow assumes this role around the time of birth.

In the mouse primitive hematopoeitic cells are first detected at about 7.5 days postcoitum (d.p.c.) in the blood islands. Primitive hematopoiesis is restricted to the

production of primitive erythrocytes and macrophage. Primitive erythrocytes differ from those of definitive origin in that they express different isoforms of globin, are larger, and remain nucleated. This population of cells is only generated in the yolk sac during this stage of development. The exact duration of yolk sac erythropoeisis is not known; production of primitive erythroid progenitors is detectable until day 9 of gestation indicating that they are a transient population existing for no more than 48 hours (27, 28). The primitive hematopoietic system appears to arise from a progenitor population distinct from that of the definitive system. This idea is supported by a series of experiments involving the identification of transcription factors required for all hematopoietic lineages except primitive erythrocytes (29, 30). The only other hematopoietic cells present in the yolk sac are macrophages (31). These cells appear distinct from those found at later stages due to the fact that they mature more rapidly and express certain genes at lower levels (25). By day 10-11 of gestation yolk sac hematopoiesis declines and the fetal liver becomes the dominant site of hematopoiesis although other intraembryonic sites have been identified. A region composed of splanchnopleural mesoderm, referred to as the AGM (aorta-gonad-mesenephros) was shown to have hemogenic activity via a number of engraftment studies as well as *in vitro* culture studies (32-35). These studies showed that only cells isolated from the AGM were capable of generating all adult hematopoeitic lineages supporting the idea that yolk sac and adult hematopoietic stem cells differ in nature. It is believed that yolk sac HSCs are devoted to a transient burst of primitive erythropoiesis and that long term repopulating stem cells (LTRSCs) required for the production of definitive lineages do not arise until hematopoiesis becomes active within the embryo proper. It is however, important to note that all of the above studies rely on *ex*

vivo culture on isolated embryonic explants and various culture conditions may affect the developmental potential of the engrafted cells. In fact Yoder *et al*. observed that in the mouse LTRSCs could be found in day 9 yolk sacs and that the AGM may simply represent a site of maturation for yolk sac derived LTRSCs as little evidence exists that cells of the AGM generates committed precursors (27, 36).

In vitro Differentiation of Embryonic Stem Cells

As stated earlier the hematopoietic system is one of the earliest systems to arise during mammalian development and it is now well established that the yolk sac represents the earliest site of both hematopoietic and endothelial cell development. Little is known about the mechanisms underlying the regulation of the events leading to the commitment and maturation of cells of mesodermal origin into cells of these lineages. The minute size of the embryo and number of cells available have made it untenable for the various molecular, cellular and biochemical assays required for full understanding of the processes underlying the developmental events following gastrulation. In order to circumvent these problems a number of groups have exploited both the availability and pluripotency of embryonic stem (ES) cells to develop *in vitro* systems to recapitulate the processes of *in vivo* development.

One system that can be used to study hematopoiesis exploits the ability of ES cells to form complex three-dimensional structures that contain developing precursor cells from multiple lineages called embryoid bodies (EBs). When cultured under appropriate conditions, EBs can be used to generate cells of both primitive and definitive hematopoiesis as well as that of the vascular system (37). The ability of this system to recapitulate the events of early embryonic development has been demonstrated by a number of studies (38). Analysis of the temporal development of various lineages has

shown that primitive erythroid and myeloid lineages are the first to develop followed by definitive lineages in a fashion similar to that observed *in utero*. Furthermore, lineage specific expression of marker genes occurs in a pattern consistent with that observed in the developing embryo (39).

A second system used to study hematopoiesis is that of the ES/OP9 system. This is a two-dimensional system that relies on a feeder layer of OP9 cells. OP9 cells are stromal cells isolated from the calvaria of newborn *op/op* mice. Stromal cells are derived from bone marrow mesenchymal cells and are capable of supporting the growth and differentiation of hematopoietic cells, however the rapid proliferation of macrophage often interferes with the examination of other hematopoietic lineages. The *op/op* mice have a mutation in the coding region of M-CSF gene that is essential for the differentiation of osteoclasts and formation of the bone marrow cavity (40). These mice suffer from osteoporosis due to defective osteoclast formation and the OP9 cell line was established from these mice to avoid macrophage proliferation When ES cells are co-cultured on stromal cell lines from normal mice the resulting cells are almost entirely macrophage. The OP9 cell line however, has been shown to be capable of supporting the differentiation of both myeloid and B lymphoid lineages *in vitro* (41).

Although both the EB and ES/OP9 systems can effectively recapitulate the events of *in vivo* hematopoiesis there are several differences between the two: The EB system is structurally more complex as it generates cells from all three germ layers simultaneously in a three-dimensional sphere. The OP9 system is two-dimensional and is more amenable to observation as one cannot see what is occurring within the embryoid body. EB formation results in the production of hematopoietic progenitors, whereas; the OP9

system produces both progenitors and mature blood cells. The OP9 method allows for the production of both megakaryocytes and B cell lineages. The EB method is not capable of generating these cells. Lastly, the EB method is restricted to the ES cell lines that efficiently produce embryoid bodies. No such restrictions exist for the OP9 method as all cells tested to date have efficiently been used in this system (42).

Though the two systems have their differences, both are useful for the observation and analysis of development of the hematopoietic system. In addition, these two systems have been used in conjunction to generate transplantable hematopoietic stem cells (43).

Globin Gene Organization

All vertebrate species utilize hemoglobin as a means of transporting the oxygen critical for respiration. The proteins that compose the hemoglobin tetramer are members of a small family of proteins encoded by the α–and β–like globin gene loci. The two loci are located on separate chromosomes in mammals and avians and are composed of several genes that are regulated in a tissue- and developmental stage-specific fashion, such that the peptide composition of the hemoglobin tetramer varies dependent upon the developmental state of the organism. The β-globin gene clusters reside in a linear array along the chromosome, the order of which reflects their order of expression during red blood cell ontogeny; the more 5' or embryonic genes are expressed first followed by their simultaneous silencing and activation of fetal and adult specific genes as development progresses. These developmental stage-specific changes in activation and suppression of gene expression has been termed hemoglobin switching and has made the globin loci a paradigm system for the study of how multi-gene loci are regulated. The central focus of globin gene research has been focused upon how the exquisitely regulated pattern of gene

expression is achieved, with a particular emphasis upon the β–like globins as understanding the molecular events involved in their regulation may provide new insights into treatment of β–globin-associated hemoglobinopathies.

The human β-globin locus resides on chromosome 11 and is composed of five genes. The 5' most embryonic gene (ε) is expressed first when the primary site of hematopoiesis is within the blood islands of the embryonic yolk sac. Shortly thereafter the embryonic gene is silenced and the fetal γ–globin genes (Aγ and Gγ) are activated as the primary site of hematopoeisis switches from that of the yolk sac to the fetal liver. A second switch occurs shortly before birth and the two fetal genes are silenced and the adult δ– and β–globin genes are activated and remain the predominant forms of globin expressed in the adult organism.

Approximately six to twenty kilobases upstream of the embryonic globin gene is a region of DNA containing a series of elements exhibiting a high degree of sensitivity to the nuclease DNase I. These elements were termed hypersensitive sites (HSs) and the region was dubbed the Locus Control Region (LCR) due to the fact that a number of naturally occurring deletions of this region can result in the disruption of globin gene expression. There exists a wealth of data indicating that cooperation between gene proximal promoter and enhancer elements along with that of those found within the LCR is required to achieve proper spatial and temporal expression of β–globin genes (44-48).

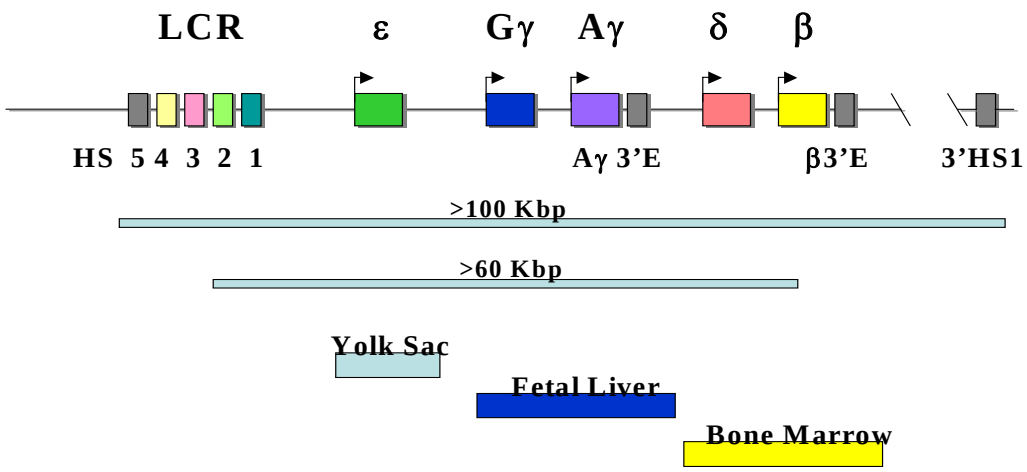


Figure 1-1. The human β-globin locus. The five globin genes are depicted as colored boxes with labels above. The arrows indicate the direction of transcription. LCR hypersensitive sites and 3' enhancer elements are labeled below. Horizontal bars indicate the stage at which the gene is expressed.

Globin Gene Proximal Regulatory Elements and Transcription Factors Involved in Erythropoiesis

The promoter regions and gene proximal regulatory elements of the β-globin locus contain binding sites for both ubiquitous and tissue specific factors. These include TATA boxes, Initiator sequences, GATA, CCAAT, and CACCC motifs (49).

GATA binding sites are recognized by members of the GATA family of transcription factors denoted by a novel zinc finger DNA binding domain, and in the case of the globin loci are most likely bound by the hematopoietic lineage specific factors GATA-1 and GATA-2 (50, 51). GATA binding sites are found in numerous erythroid gene promoters and in all LCR elements (51). The importance of these factors in hematopoiesis and erythroid differentiation has been shown by studies in which the genes were disrupted in ES cells. GATA-1 null ES cells are unable to contribute to erythropoiesis and GATA-2 mutant embryos die around 9.5 d.p.c. of severe anemia (52, 53). Because GATA-1 null ES cells show growth arrest at the proerythroblast stage, it has

been proposed that GATA-1 has a role in regulating the genes required for erythroid cell maturation and commitment (54). Indeed, *in vitro* hematopoietic differentiation in GATA-1 ablated ES cells revealed a total loss of primitive erythropoiesis although a number of putative GATA-1 target genes, including globins, SCL, erythropoietin receptor, and EKLF, are still expressed (55). This may be due to the ability of one of the other GATA proteins to compensate for loss of GATA-1. GATA-2 expression increases as much as 50-fold in the absence of GATA-1, suggesting that it is negatively regulated by GATA-1. GATA-2 appears to be necessary for progenitor cell expansion and survival. When GATA-2 null ES cells were introduced into normal mouse blastocysts, the resulting chimeric animals revealed that the GATA-2 mutant cells failed to contribute to the erythroid lineage, particularly that of the definitive stage (53). In addition, forced expression of GATA-2 in avian erythrocytes leads to an arrest in differentiation and promotes the proliferation of primitive erythroid progenitors, indicting that its suppression my be required for progression through later stages of development (56). It has also been shown that globin gene activation in multipotential hematopoietic progenitors may be dependent upon protein complexes nucleated by GATA-1 and GATA-2 (57). GATA binding sites are found in close proximity to canonical Sp1 sites (51) and Merika and colleagues were able to show that GATA-1 was able to physically interact with Sp1 and EKLF via their zinc finger domains (58). Furthermore, GATA-1 has been shown to physically interact with the histone acetyltransferase CBP linking GATA function with that of chromatin structure alterations that are intrinsic to gene expression (59).

NF-E2 is a heterodimeric protein composed of 45 and 18 kd subunits and recognizes an AP-1 like sequence (TGA[C/G]TCA) found in many erythroid promoter and enhancer elements, including that of the α– and β–globin loci. NF-E2 motifs often exist in close juxtaposition to that of GATA motifs and the two have been shown to act synergistically to activate elements of the β–globin LCR (60). p45 is an erythroid specific protein containing a basic region leucine zipper motif (b-zip) and a DNA binding domain related to the *Drosophila* protein cap n' collar (cnc). The smaller p18 subunit also contains a b-zip domain but is ubiquitously distributed and related to the v-*maf* oncogene. Although it lacks a transactivation domain the p18 subunit can still dimerize and bind DNA independent of p45. Using a p45 NF-E2 null immortalized cell line (CB3) and a dominant negative form of p18 expressed in MEL cells; Kotow and Orkin were able to show that the p45/p18 heterodimer is the active NF-E2 complex (61). Johnson and colleagues proceeded to show that p45NF-E2 was required for histone hyperacetylation and RNA polymerase II recruitment at the adult β–globin gene promoter (62). Furthermore, homodimers of p45 are incapable of binding NF-E2 sites although homodimers of small *maf* proteins retain this ability, indicating that changes in dimer composition may play a role in regulating sets of erythroid genes (63, 64). Research involving p45 NF-E2 null mice indicate that the p45 peptide may be dispensable for erythropoiesis because only subtle perturbations in red cell development such as microcytosis and decreased hemoglobin content have been observed (65). Thus, although p45 NF-E2 may be required for optimal globin gene expression, other AP-1 like proteins may be able to substitute for the transcriptional activities provided by p45. Interestingly, p45 NF-E2 null mice do suffer from a complete absence of platelets as well as

hemorrhaging resulting in extreme morbidity during the neonatal period, indicating that p45 is indispensable for certain hematopoietic functions.

Erythroid Krüppel-like factor (EKLF) is an erythroid and mast cell specific transcription factor containing three TFIII A-like zinc fingers homologous to other Krüppel-like factors. EKLF is expressed in both primitive and definitive erythroid lineages and its promoter contains a GATA motif shown to be required for its activation. EKLF binds to elements containing a CCACACCT sequence such as that found in the globin gene promoters and LCR cores. Naturally occurring mutations of this element in the adult β–globin gene promoter lead to β–thalassemia in humans (66). Although other zinc finger proteins such as Sp1 are also capable of binding CACC sequences, the *in vivo* role of EKLF has been more accurately defined using gene-targeting experiments. EKLF null embryos die of anemia during the fetal stage due to severe β–thalassemia (67). EKLF activity appears to be restricted to adult lineages as both embryonic and fetal hematopoiesis is unaffected, however, it may play some role in the fetal to adult switch in globin gene expression as EKLF null mice carrying a human β–globin locus transgene display a delayed fetal to adult switch (68). Over-expression of EKLF in fetal erythroid cells enhances β–globin gene expression more than a 1000-fold and mutation of the β–globin gene CACCC element abrogates this affect. Homozygous knock out of EKLF in mice results in persistent expression of human γ-globin transgenes in adult erythroid cells (68, 69). It appears that the γ–globin gene is principally regulated in an autonomous fashion and that the adult β–globin gene is active at basal levels during the yolk sac stage, suggesting that the role of EKLF in fetal to adult globin gene switching is not primary.

An additional factor that has been shown to play a critical role in hematopoietic development is stem cell leukemia factor (SCL). It is a basic helix-loop-helix (b-HLH) protein expressed in several hematopoietic lineages. Abnormal expression of SCL is associated with acute T-cell leukemia (70). Many b-HLH proteins can act as transcriptional activators and bind DNA specifically at E-box elements (CANNTG) found in a variety of erythroid promoter and enhancer elements (71). Gene targeting experiments in mice have shown that SCL is critical for red blood cell development and SCL anti-sense RNA blocks differentiation of MEL cells (72). SCL mutant mice die around 8.5-9.5 d.p.c., exhibiting a total absence of nucleated red blood cells and mimicking the defects caused by loss of GATA-1. The SCL gene itself may be regulated by one or more GATA factors as one of its two promoters contains a GATA site.

Locus Control Regions and Hemoglobin Gene Switching

The β–globin locus control region LCR is a distal regulatory element located approximately 6-22 kb(73) upstream of the ε–globin gene. The LCR is characterized by a series of elements that are extremely accessible to the nuclease DNase I in erythroid cells. The core regions of these hypersensitive sites are roughly 200-400 bp and are separated by one to two kilobases of flanking sequences. The cores contain a high density of binding sites for both ubiquitous and tissue-specific transcription factors. The overall structure of these domains is conserved in several species, suggesting a functional significance (74). The motifs that are most conserved are maf recognition elements (MAREs) and GATA sequences in HS2, 3, and 4, KLF sites in HS2 and 3, as well as an E-box motif in HS2. The GATA sites are most likely bound by GATA-1 or GATA-2, as these are the only known GATA factors expressed in erythroid cells. The MARE sequences are most likely bound by heterodimers of small maf proteins and other b-zip

family members such as p45 NF-E2 and Bach 1. The CACCC elements in HS2 and 3 are most likely bound by EKLF *in vivo* as transgenic mice lacking EKLF show a specific decrease in hypersensitivity at HS3 as shown by Lee and colleagues (73). E-box motifs are bound by helix-loop-helix proteins such as USF and SCL and *in vitro* studies indicate that the E-box in HS2 is bound by both these proteins, although the functional significance of these interactions have yet to be elucidated (73). Protein-protein interactions most likely play a critical role in LCR function as most of the proteins associated with HS elements have the potential to interact with one or more partners, perhaps to mediate the formation of higher order structures that are involved in changes in gene expression and chromatin structure observed throughout the locus (46).

Locus control regions are defined as elements capable of conferring position-independent and copy number-dependent expression on *cis*-linked genes irrespective of chromosomal position. They are capable of directing tissue-specific expression at physiological levels and may also be involved in the replication timing of gene loci. The founding member of this class of regulatory elements is that of the mammalian β–globin locus and accordingly an extensive amount of research has been aimed at elucidating its role in globin gene regulation. However, the exact *in vivo* activities of the LCR are still a matter of much debate, the reasons for which are manifold. It is generally agreed now that the LCR is required for high-level expression of all the genes at all developmental stages. Whether or not the LCR simply acts as a classical, although somewhat more complex, enhancer is one of the fundamental questions yet to be answered. The importance of the LCR was first realized through the study of a series of naturally occurring deletions of the region, which result in the clinical manifestation of

β–thalassemia. The smallest of these deletions (Hispanic Thalassemia) removes just 35 kb of sequence upstream of the ε–globin gene but results in the silencing of all the genes and a loss of general DNase I sensitivity throughout the remainder of the locus. This observation indicated that the LCR must contain some positive *cis*-acting elements required for the activation of the locus and expression of all the β–globin genes (47, 75). Transgenic studies found that globin transgenes were expressed at variable levels and only in a small portion of animals, indicating the constructs were subject to position effects. However when linked in *cis* to an LCR or combinations of various HS sites, the transgenes were expressed at physiological levels in a position-independent and copy number-dependent manner (76). These results indicated that not only could the β–globin LCR act as a strong enhancer but also harbored a dominant chromatin opening activity. Although all of the above data suggest a prominent role for the LCR in globin locus activation and resistance to the nature of chromosomal position effects other research argues against the same function at endogenous loci.

The individual HS sites themselves have been shown to contain both unique as well as redundant activities. In one group of experiments deletion of a 375 bp core element from HS2 in the context of an otherwise complete yeast artificial chromosome containing the human β–globin locus resulted in catastrophic reductions in globin gene expression as well impaired HS site formation(77). Replacement of HS2 by HS3 resulted in the restoration of HS formation and thus chromatin opening, but was unable to fully compensate for the transcriptional enhancement activities of HS2 (78). In a similar series of studies by the same group it was found that deletion of HS sites 3 and 4 resulted in a drastic decrease in expression of all the globin genes at all stages . When HS3 replaced

the deleted HS4, proper developmental stage expression at physiologically relevant levels was restored. However, when HS4 was used to replace HS3 gene expression levels was attenuated at all stages (76-78). In a parallel group of experiments Peterson *et al*., created larger deletions of HS2 and 3 encompassing not only the cores but the flanking sequences as well(79). These deletions resulted in a decrease in ε–globin expression in the case of HS3 and a minor decrease in ε–, γ–, and β–globin gene expression when HS2 was deleted, contrasting the results of deletion of the core HS elements alone (79). These results provide some insight into the discrete nature of LCR elements and suggest that a certain functional synergism is required for full activity. They also suggest that certain sequences act as structural components for the formation of a higher ordered structure such as that postulated in the holocomplex model of LCR activity (68, 77). The hypothesis is that the LCR HS sites physically interact to form a higher ordered chromatin structure and that this interaction is mediated by protein-protein and protein-DNA interactions (see below).

Studies of LCR function at the endogenous human and mouse loci provided somewhat contradictory results. In one such study Reik and colleagues used homologous recombination and the DT40 shuttle system to generate human β–globin LCR mutations and study their effects in MEL cells. The authors found that deletion of the LCR in this system resulted in the loss of globin gene expression but that the chromatin structure of the remainder of the locus was essentially unchanged. They concluded that for the human locus the LCR was necessary for high-level gene expression, but not for the maintenance of chromatin structure (80). A similar deletion of the mouse LCR resulted in a drastic decrease in the expression of all the genes although temporal regulation was maintained

and HS site formation at the promoters was unperturbed (81, 82). These results suggest two possibilities: that the LCR simply acts as an enhancer ensuring that the genes are expressed at high levels, or that some redundant chromatin opening activities 5' to the currently defined LCR exist. In both mouse and humans the β–globin locus is embedded in inactive odorant receptor loci and HS sites outside the current boundaries of the defined LCR have been shown to have some impact on globin gene expression and formation of higher order chromatin structures associated with developmental switching. Interestingly a recent report by Farrell and colleagues observed that an 11 kb deletion of HS sites 5' to the LCR sequences had little or no effect on HS site formation and gene expression that may shift the focus of current research to sequences even farther upstream (83).

The fact that the function of the LCR, when integrated at ectopic sites, appears to differ from that of the locus in the context of its endogenous location within the genome has confounded attempts to fully integrate the entire body of research into a comprehensive model for globin gene switching. Consequently several models have been formulated describing this phenomena, the two most prominent are the "linking" and "looping" models of gene activation. The linking model states that stage specific activities bound at the LCR are transmitted via chromatin associated facilitator proteins that interact with factors in the transcriptionally active domain (84). The looping model posits that the LCR holocomplex interacts physically with specific gene promoter and proximal regulatory elements by looping of the intervening DNA (89). This interaction allows for the transfer of activities first recruited to the LCR to be deposited at the genes at the proper time. Two recent reports indicate that not only do sequences within the LCR

and active globin genes reside in close proximity in the nucleus but that the conformation of such interactions is dynamic and dependent upon the developmental state (44, 48). Whereas both models are consistent with current data there is no irrefutable evidence supporting either, although the looping model has gained support in recent years.

Summation

The complex pattern of gene expression and dynamic changes in chromatin structure that are associated with expression of the β-like globin genes have made hemoglobin gene switching a paradigm for the study of how coordinated regulation of multi-gene loci can be achieved. Although a complete understanding of the events leading to the activation and silencing of particular genes has yet to be obtained they can conceptually be divided into several steps: generation of a highly accessible holocomplex, recruitment of transcription factors and chromatin modifiers to the LCR, activation of stage specific globin sub-domains, and transfer of transcription complexes to the appropriate gene promoters.

The initial event of globin gene expression must be an opening of the domain rendering is accessible to various *trans*-acting factors; an event that may occur prior to commitment to the erythroid lineage and supported by the fact LCR HS sites are detectable in uncommitted progenitors (85). Hematopoieitic-specific transcription factors expressed in uncommitted progenitors could diffuse into inactive or non-permissive chromosomal domains and bind to target sequences within the locus. These initial binding events could result in a general increase in accessibility of the locus to other factors such as HATs and ATP-dependent chromatin remodeling complexes as well as sequestration of the locus in an active region of the nucleus. GATA factors may be

involved at this step as data indicates that they are expressed in multi-potential progenitor cells and GATA binding sites are found throughout the locus (86). In addition, GATA-1 is known to associate with the HAT CBP, an interaction that stimulates GATA-1 activity in transient transfection assays (87). The general opening of the locus does not appear to require activities present within the LCR as even when it is deleted the remaining elements retain nuclease sensitivity and increased levels of histone H4 acetylation (82, 88). The increased permeability of the locus to proteins present within the nucleus and the high concentration of binding sites within the core LCR HS elements most likely results in the accumulation of these factors at the LCR, perturbing the DNA. It is this perturbation that most likely results in hypersensitivity and holocomplex formation as many of the proteins involved not only contain DNA binding domains but protein-protein interaction domains as well.

Once formed the LCR holocomplex could potentially act as a sink for the macro-molecular complexes that directly mediate transcription, first recruiting them to the locus and than distributing them to globin gene promoters in a developmental stage-specific fashion. *In vitro* studies indicate that RNA Pol II is present at LCR HS2 in undifferentiated ES cells prior to its appearance at promoters (89). It is possible that recruitment of transcription complexes to the LCR could further serve to modify the sub-domains. Transcription of the LCR itself as well as that of intergenic regions has been observed and movement of polymerase complexes through these regions may facilitate the formation of transcriptionally active sub-domains (90). These intergenic transcripts have been used to delineate developmental stage-specific chromatin domains within the locus, however their exact role is not currently known. They may simply be a bi-product

of active transcription, although arguing against this the fact that deletion of the adult-specific transcription initiation site results in a decrease in general DNase I sensitivity of the domain as well as a reduction in expression of the β–globin gene (90). Despite this lack of information concerning the role of intergenic transcription, the idea of stage specific sub-domains is supported by observed changes in the degree of DNase I sensitivity and patterns of histone modification that occur during development (88, 91). These changes are most likely due to the dynamic association of factors involved in the regulation of chromatin structure including, but not limited, to histone acetyl transfereases and deacetylases as well as nucleosome remodeling complexes.

The ultimate outcome of stage-specific reconfiguration of the β–globin locus would be the direct interaction of the LCR with the appropriate gene promoter and the inhibition of promiscuous interactions with regulatory elements outside the domain being transcribed. This would result in the transfer of activities required for high-level expression of the globin genes from the LCR to these promoters. Evidence supporting direct LCR-promoter communication comes from studies showing a localized hyperacetylation of histone H3 at the LCR and active gene promoter that is not observed in the absence of the LCR, an event consistent with the "looping" model of LCR function (88). In addition, HS2 and HS3 have been shown capable of recruiting RNA PolII *in vitro* and *in vivo* (62, 92). The *in vivo* data indicate that the recruitment of PolII to the LCR and its transfer to the β–globin promoter requires NF-E2 as in MEL cells lacking this protein PolII can be found at the promoter but is absent from the gene.

Although this is a simplified view of the events occurring at the globin locus during development it provides the framework from which to base further studies into the

molecular nature of developmentally regulated changes in gene expression. A fundamental understanding of these processes will be critical in identifying genes required for proper development and ameliorating conditions brought about by their disregulation.

CHAPTER 2
MATERIALS AND METHODS

ES cell differentiation. Mouse ES cells were differentiated to generate cells of the hematopoietic lineage using the ES/OP9 method established and described by Kitajima *et al*.(42). Briefly, ESD3 cells (ATCC, CRL-1934) were seeded onto a confluent monolayer of mouse embryonic fibroblasts (MEFs) at a density of 10^5 cells/25 cm^2 in ES media (DMEM, 4.5g/l glucose, 1.5g/l sodium bicarbonate, 15% FBS, 0.1 mM 2-mercaptoethanol and 10^6U/ml LIF), grown for two days, and then passaged (1:6) and grown for another day. An aliquot of the cells (3-4x10^7) was taken at this time (Day 0) and subjected to RT-PCR and ChIP analysis. The remaining day 0 cells were then seeded onto confluent OP9 stromal cells in OP9 media (α-MEM with ribonucleosides and deoxyribonucleosides; 20% FBS) in the absence of LIF at a density of 10^4 cells/well in 6 well tissue culture dishes. At day 3 Epo or Epo and SCF was added (2U/ml and 50ng/ml, respectively) for the remainder of the course of induction. On day five of induction cells were passaged and reseeded onto fresh OP9 cultures at a density of 10^5 cells/well. On day 12 cells were collected and subjected to RT-PCR and ChIP analysis.

RT-PCR . RNA was isolated for RT-PCR using the Arum Total RNA Mini Kit (Bio Rad) according to the manufacturer's protocol. Reverse Transcription was performed using 200 to 250 ng RNA and the iScript cDNA synthesis Kit (Bio-Rad) as described by the manufacturers protocol. PCR amplification was performed using the Eppendorf PCR Mastermix (Eppendorf). Primer sequences specific for Flk-1, Epo-R, GATA-2, β-actin, Rex-1 were obtained from Elefanty *et al.* (93). Additional primers used

are as follows: mouse HS2 US: 5'GGGTCTCTCTAGGAGGAAGTCCACAGG 3' and DS: 5'CAGATCTAATGACCCTAACTCTAAC 3'; mouse β^{maj}-globin: US, 5'CACCTTTGCCAGCCTCAGTG3', DS, 5'GGTTTAGTGGTACTTGTGAGCC3'; mouse Ey, US, 5'AACCCTCATCAATGGCCTGTGG, DS, 5'TCAGTGGTACTTGTGGGACAGC 3'.

Chromatin immunoprecipitation (ChIP). ChIP was performed as described by Leach *et al.*(92). For acylamide gels the following primers were used: Mouse β_{maj}-globin: US 5' TAATTTGTCAGTAGTTTAAGGTTGC 3' and DS 5' CAT TGTTCACAGGCAAGAGCAGG 3'; Mouse Ey-globin: US 5' CAAAGAGAG TTTTTGTTGAAGGAGGAG 3' and DS 5' AAAGTTCACCATGATGGCAAGTCTGG 3'; Mouse HS2: US 5' TTCCTACACATTAACGAGCCTCTGC 3' and DS 5'AACATC TGGCCACACACCCTAAGC 3'; mouse HS2 5'flank, US 5' CTATTTGCTAACAGTCTGACAATAGAGTAG3' and DS 5'GTTACATATGCAGCTAAAGCCACAAATC 3'.

Real-time PCR analysis was carried out using the DyAmo HS SYBR green qPCR kit (MJ Reaserch) and the following primers: Mouse β^{maj}-globin: US 5' CAGGGAGAAATATGCTTGTCATCA 3' and DS 5'GTGAGCAGATTGGCCCTTACC 3'; Mouse Ey-globin: US 5' CAAAGAGAG TTTTTGTTGAAGGAGGAG 3' and DS 5' AAAGTTCACCATGATGGCAAGTCTGG 3'; ßH1: US 5' AGGTCCAGGGTGAAGAATAAAAGG 3' and US 5'ATCTCAAGTGTGCAAAAGCCAGA 3'; Mouse HS2core: US 5' AGTCAATTCTCTACTCCCCACCCT 3' and DS 5'ACTGCTGTGCTCAAGCCTGAT 3'; 3/2flank, US 5' TTAAAGCCTCATTATCTCCAAACCA3' and DS

5’GTGTGCACTGGGTGGGTAGA 3’; IVR3: US 5’ TGTGCTAGCCTCAAGCTCACA 3’ and DS 5’ TCCCAGCACTCAGAAGAAGGA 3’; Mouse Rex-1: US 5’AACTGCATCCTCTGCTTGTG 3’ and DS 5’ TGCGCTCTATTTCCTCCTTG3 ’. Antibodies: TFIIB sc-225, Pol II (N-20) sc-899, NF-E2 (C-19) sc-291, (all purchased from Santa Cruz Biotechnology), Pol II 05-623, histone H3 di-methylated at lysine 4, and acetylated histone H4 (all purchasd fromUpstate Biotech.). All antibodies were tested in Western blotting experiments using MEL or K562 nuclear extracts as described by Leach *et al.* (94).

Transgene Construction. Components of the chicken and human globin locus were combined to generate the plasmid 432β4. Briefly, the core region from human HS3 was isolated as an XbaI-XhoI fragment from the plasmid HS434 (77). The HS2 core was generated by PCR from a YAC containing the entire human β−globin locus as a template (YACA201F4.3; (95)) using the following primers: HS2US 5’ACCTCGAGCCCTCTATCCCTTCCAGCATCC 3’; and DS, 5’ACGATTCGAATATCACATTCTGTCTCA 3’; XhoI and EcoRI sites were included 5’ and 3’, respectively, to facilitate cloning. The fragments were cloned between the XbaI and EcoRI sites of the plasmid pGEM7 to create pGEM7HS32cores. HS4 from the chicken globin locus was PCR amplified from chicken genomic DNA using primers that introduced AatII and SphI sites 5’and 3’ to the 250 bp core element (96): 5’ HS4US, 5’ ACGACGTCGAGCTCAGGGGACAGCCCCCCC 3’; DS, 5’GTGGACCCCCTATGCCCCTTTTGCATGCAC 3’. The resulting fragment was cloned into pGEM7HS32cores using AatII and SphI sites. From this plasmid a SacI-KpnI fragment containing all three core regions (chicken-HS4, Human-HS3 and -2 cores) was

isolated and cloned into pUC19 to create pUC432cores. Into this plasmid the human β–globin gene/3'enhancer was cloned as a 4.6kb KpnI-XbaI fragment isolated from the plasmid βA/X (92)using KpnI and XbaI sites. Finally the 3' chicken HS4 was PCR generated to contain 5' and 3' SalI sites: 3'HS4US, 5'ATATGTCGACCTCACGGGGACAGCC 3'; 3' and DS, 5' CCCGGTCGACCCCCGTATCCCCCCA 3'. The resulting fragment was cloned into the SalI site of pUC432cores to create the plasmid 432β4pUC. The plasmid containing the β-globin integration construct (p43f2β4) was constructed using the pNEB193 plasmid (New England Biolabs) as the backbone. A 500 bp EcoRI-PstI fragment containing the AAV2 5'ITR and p5 promoter was blunt ended on the 3' end and cloned into the EcoRI and SmaI sites of pNEB193. A linker containing PacI, SpeI, NsiI, ApaI and XbaI restriction enzyme sites were then cloned into the PacI and XbaI sites. The 246 bp 3' cHS4 fragment was PCR amplified using the 3'HS4 primers described above and ligated into the SalI site of the vector. Next a 3.8 kb NsiI-XbaI fragment containing the human β-globin gene including the promoter and 3' enhancer elements was cloned into the respective sites in the vector. A 6.9 kb NsiI fragment containing the human β-globin LCR from HS3 through HS2 was ligated into the vector and clones with the correct orientation were identified by restriction enzyme analysis. The 5'HS4 fragment was PCR amplified using the following primers:
HS4 US, 5'CCTTAATTAACTCACGGGGACAGCC-3', HS4DS 5'CTAGTCTAGACCCCGTATCCCCCA-3' introducing PacI and XbaI sites at the 5' and 3' ends, respectively. This fragment was ligated into the PacI and SpeI sites of the vector. This cloning step removed 742 bp of the 5' end of the HS3 flank HS2 LCR

fragment previously ligated into this vector. The complete p43f2β4 plasmid was purified using Qiagen's maxiprep kit. Both plasmids were sequenced to verify the integrity of the regulatory elements. The 3.6 kb EcoRI-KpnI fragment containing the human AAVS1 site was ligated into pBS246 (Invitrogen).

Transgenic Mouse Production. We used FvB/B6 mice (Jackson laboratories) to generate all transgenic lines. Plasmid p432β4 was linearized with AatII; pAAVS1 was linearized with EcoRI. The linearized plasmids were purified from agarose gels and resuspended in injection buffer at a concentration of 2 ng/μl. Transgenic mice were generated as described previously (77). Transgenic founders were first identified by PCR on DNA isolated from tail clips. Copy number and integrity was analyzed by Southern blotting. AAVS1 transgenic mice were mated to generate mice homozygous for the transgene. In experiments using the β-globin integration construct (p43f2β4) 1-5 ng of the supercoiled plasmid DNA was complexed on ice with a 1:5, 1:10, or 1:15 molar ratio of DNA to purified AAV2 rep68 protein and injected into fertilized oocytes homozygous for the AAVS1 transgene.

DNA Isolation, PCR Screening, Inverse PCR, and Southern Blot Analysis. DNA was isolated from mouse tail and the presence of human AAVS1 or β-globin locus sequences was first determined by PCR using primers against the human AAVS1 site and the flanking region between human β-globin HS3 and HS2: AAVS1 US 5'ATCTGCCCGGCATTTCTGAC 3', AAVS1 DS 5'CGCAAAATGTCGCAAAACAC 3'. The primers amplifying a region between HS2 and HS3 were published by Leach *et al.* (94). DNA from tails that contained AAVS1 and/or human β-globin sequences were then subjected to Southern blot analysis. Approximately 10 μg of tail DNA was digested

with restriction enzymes, size fractionated on 1.2% agarose gels, and transferred onto Nylon membranes as described (78). The membranes were then probed with DNA fragments corresponding to regions of the β-globin transgene or of the human AAVS1 site. The β mid probe is a 917 bp BamHI-EcoRI fragment derived pβA/X (92) and corresponding to the coding region of the human β-globin gene. The 3' β-globin probe is a PstI fragment encompassing the β-globin 3'enhancer and derived from pβA/X. The AAVS1 probe was derived by PCR using the primers described above (AAVS1 US and DS). Copy number of transgenes was determined by hybridizing the nylon membranes with a radioactive probe corresponding to the murine snrp N gene. This probe is a 300bp EcoRI/SacI fragment derived from the snrp N locus which hybridizes to a 4kb EcoRI fragment in Southern blotting experiments. This probe was made available to us by Dr. Camilynn Brannan (UF).

Inverse PCR was carried out as described by Hartl and Ochman (97). Briefly, genomic DNA from S1 transgenes was digested with SacI or MspI, ligated and subjected to PCR using the following primers, S1tg DSI: 5'CACAGCCCCAGGTGGAGAAACT3', S1tg DSII: 5'CCCGGGTTGGAGGAAGAAGACT3', S1tg US: 5'TTCTCCAGGCAGGTCCCCAA3'. PCR products were subcloned into the TopoII vector (Invitrogen) for sequencing.

Metaphase Preparation and FISH Analysis. F1 animals containing the transgene were sacrificed and the spleens were isolated in 2-3 ml sterile PBS for metaphase chromosome preparations (98). Cells were isolated from the spleen and pelleted in a total volume of 10 ml PBS at 500g for 10 min. The cells were then immediately resuspended

in 10 ml 0.075M KCl (prewarmed to 37°C) and incubated at 37°C for 20 min. 2 ml fixative (3 parts methanol: 1 part glacial acetic acid) was immediately added to the cells and the cells were pelleted at 500 g for 10 min. The metaphase cells were washed 3 times in 10 ml fixative and stored at 4°C in fixative.

Metaphase cells were placed on microscope slides and allowed to air dry. The slides were aged in an 80°C incubator for 1 hr and then immediately used for FISH. In the dark, 10-15 μl fluorescently labeled probe was placed on each slide, covered with a glass slide, and then placed in a HyBrite (Vysis) apparatus overnight where the slides were denatured at 75°C for 15 min and hybridized at 37°C for16 hrs. Cover glasses were then removed in the dark, and the slides were washed for 2 min at 75°C in 0.4 X SSC, 0.3% NP-40, followed immediately by 1 wash in 2 X SSC, 1% NP-40 for 1 min at room temperature. The slides were allowed to air dry in the dark and counterstained with 10 μl DAPI II solution (Vysis). The slides were visualized by fluorescence microscopy or stored in the dark at 4°C.

The entire β-globin transgene (p43f2β4) or the AAVS1 plasmid was fluorescently labeled using the BioPrime kit (Invitrogen) by substituting rhodamine-tagged dUTP (Tetramethylrhodamine-5-2'-deoxy-uridine-5'-triphosphate, Roche) for the biotin-tagged dUTP that comes with the kit. The chromosome paint probes specific for mouse chromosomes 7 and 15, were obtained from ID Labs Inc. and used according to the manufacturer's instruction.

RNA Isolation and Semi-quantitative RT-PCR . F1 animals containing the β-globin transgene were made anemic by injecting phenylhydrazine as described previously (77). RNA was extracted from the spleen and cDNA was prepared as described in

Bungert *et al*.(77). 10% of the RT reaction was used for semi-quantitative PCR analysis using primers against the human β-globin and mouse α-globin genes using primers published previously(77). We used a new mouse α-globin downstream primer to span an intron in these experiments: ma DS, 5'TCCACACGCAGCTTGTGGGCATGCAG 3'. PCR samples were removed at 14, 16, and 18 cycles and size fractionated on a 10% polyacrylamide gel. The gels were stained with SyBr-green and quantitated by phosphorimager analysis using a storm scanner. Human β-globin transgene expression level was calculated relative to the expression level of the endogenous mouse α-globin gene.

DNase I Hypersensitivity Analysis. Cells taken from a spleen of mice made anemic by phenylhydrazine injection (see RNA isolation section) were washed with PBS, pelleted, and subjected to DNase I digestion as described by Kang *et al.* (99). 10 μg of DNase I treated DNA was digested with EcoRI, size fractionated on a 0.8% agarose gel, and subjected to Southern blot analysis using the β-mid probe as described above.

YAC Modification. Generation of mutant human β-globin locus YACs by homologous recombination in yeast was carried out as described previously (77). Flanking regions for recombination vectors were constructed by PCR amplification. The vectors were linearized within the 5' or 3' flanking regions and used to transform yeast cells containing the human β-globin locus YAC (A201F4.3). Transformed yeast cells were plated onto agarose plates lacking uracil. DNA was prepared from cells growing on uracil-deficient medium and analyzed by Southern blotting for correct integration of the loxP containing mutation. Clones that had the plasmid integrated into the homologous site of the β-globin locus YAC were grown in uracil-containing medium and plated onto

agarose plates containing 5'-fluoroorotic acid (FOA). Growth on FOA-containing plates indicates removal of the *URA3* gene by homologous excision within the YAC, generating cells with either wild type or the desired mutant β-globin locus structure. DNA was prepared from these cells and analyzed by Southern blotting for of the inserted loxP site our selectable marker. Yeast clones bearing the mutant β-globin YACs were embedded in agarose plugs for pulsed-field gel electrophoretic analyses or DNA isolation.

Four inserstion vectors were generated for introducing wild-type (L1) or mutated (L2) loxP sites, as well as a puromycin selectable marker into the human β-globin YAC. Sequences for mutated and wild-type sites differed by a single G-A basepair mutation in the spacer region of the site (100). For insertion of the L2 site 5' to the LCR a vector, 5'LCRL2, was created by PCR using sequences approximately 3 kb to the 5' EcoR1 at the 5' of the locus as defined by the HUMHBB sequence. The 5' homology of this vector was approximately 500 bp and was delimited 5' by a HindIII site and 3' by a BamH1 site. The 3' homology was 700 bp and contained SpeI and SfiI restriction sites at its 5' and 3' ends, respectively. The fragments were cloned into the pBS246 loxP vector (Gibco). An L2 site was introduced as a synthetic oligonucleotide ligated into unique BamH1 and SpeI sites . In addition to the loxP site these oligos also contained a unique RsrII site that would later be ued for removal of the YAC vector arms and circularization. The cassette was cloned into the pRS306 plasmid as a NotI fragment, linearized using a unique BtrI site that cuts in the 5' homology and introduced via electroporation into the A201F4.3 strain of *S. cerevisiea*. Similar strategies were used for producing the 3'LCR L2 and 3' β-globin genes L1 (3'BGL1) constructs. The 3'LCRL2 construct contained regions of homology corresponding to bp 13000-13900 (5' homology) and 13900-15100

(3’ homology) of the HUMHBB locus and contained unique, PCR generated EcoRV and BamHI sites (5’homology) and SpeI and SfiI sites (3’homology). BamH1 and SpeI sites were included at the 5’ and 3’ ends to facilitate cloning of the L2 site into the vector. The same L2 oligo was used to introduce the L2 site between the 5’and 3’ homologous segments with the following exceptions: a SmaI site was included at the 5’ end and a BamHI site was present at the 3’ end. Again, these fragments were sub-cloned into the pBS246 vector followed by cloning into the pRS vector as a Not I cassette. The vector was linearized in the 3’ homology using XhoI. The 3’BGL1 plasmid utilized sequences ranging from 66000 to 68100 for its 5’ and 3’ homologies. The wild-type loxP site in this vector was derived from the pBS246 plasmid and the final vector was linearized in the 5’homology using SphI. All finished vectors were sequenced to confirm the presence of L1 and L2 sites. The puromycin gene insertion vector was created using homologous sequences spanning bp’s 65423 to 66900 of the HUMHBB locus. PCR generated homologies were ligated into pRS306 and the puromycin gene was subcloned as AscI fragment from the vector pKOSelectPuro (Stratagene). The vector was linearized with the enzyme BbvCI prior to electroporation.

L2 Oligos: 5’LCRL2, Forward: 5’ GATCCGGTCCGATAACTTCGTATAATGTATACTATACGAAGTTATA 3’; Reverse:5’CTAGTATAACTTCGTATAGTATACATTATACGAAGTTATCGGACCG G 3’; 3’LCRL2, Forward: 5’GGGCGGACCGATAACTTCGTATAATGTATACTATACGAAGTTATAG3’ Reverse:5’GATCCATAACTTCGTATAGTATACATTATACGAAGTTATCGGTCCG CCC3’

CHAPTER 3
ESTABLISHMENT OF PATTERNS OF HISTONE MODIFICATION AND FACTOR RECRUITMENT WITHIN THE β–GLOBIN LOCUS DURING *IN-VITRO* DIFFERENTIATION OF MURINE EMBRYONIC STEM CELLS

Introduction

Multicellular organisms are composed of a variety of cell types, all derived from a common precursor and characterized by different patterns of gene expression. It is the transcriptional profile of a specific cell type that determines its morphology and function. The establishment of the expression patterns of terminally differentiated cells is mediated by various ubiquitously expressed and tissue-specific transcription factors and repressors, as well as nucleosome modifying and remodeling factors whose activity results in the proper spatial and temporal expression of specific subsets of genes. The sequential silencing of genes involved in maintenance of pluripotent and multipotent states and the activation of those involved in differentiation is believed to be a dominant factor in the progression from multi-lineage precursors to that of specific cell typeS. The maintenance of this transcriptional state following cell division depends upon not only the direct action of *trans*-acting factors, but also the heritable epigenetic status they impart. Data accumulated in recent years indicates that combinations of covalent histone modifications may constitute a "histone code" that regulates the use of genetic information (10). The manner in which the acquisition of various epigenetic states is regulated during development is only partially understood.

The vertebrate globin gene family has provided a model system to study the molecular basis of developmentally regulated differential gene expression. It contains a

number of tissue-specific genes that are coordinately regulated and whose expression changes during development of the hematopoietic system, a process termed “hemoglobin gene switching”. Epigenetic modifications have been shown to play an important role in the expression of the β–like globin genes. The chicken β–globin locus has been shown to reside in a domain of uniform histone hyperacetylation with the active genes being acetylated on lysine 4 of histone H3 and inactive genes exhibiting H3 lysine 9 methylation (101, 102). Differential acetylation has also been observed in the murine β–globin locus. Forsberg and colleagues observed dynamic changes in histone acetylation of the globin genes during development with the LCR and active genes marked by increased H3 and H4 acetylation (91). These observations suggest epigenetic modifications may be an important factor in the maintenance of an active locus, however, how and when these patterns are established is not entirely known. Bottardi and others investigated the epigenetic state of the human β–globin locus in hematopoietic progenitor cells (HPCs) and transgenic mice (103). They found that histone H3 at the β promoter was hyperacetylated and dimetylated at lysine 4 in HPCs but deacetylated in mature erythroid cells. In contrast, the human γ promoters lacked these modifications in HPCs and transgenic fetal liver cells. These results indicate acetylation plays a critical role in the transcriptional potentiation and developmental regulation of these genes in progenitor cells or cells that have yet to express the genes at physiologically relevant levels. Chromatin structure modifications in uncommitted progenitor cells have also been observed for the murine β–globin locus (85, 104). A recent study by Johnson and colleagues showed that RNA Pol II is recruited in a strictly localized fashion within the LCR and was only detected at the core regions. Localization of Pol II to the LCR was

independent of active transcriptional elongation; the addition of DRB did not affect recruitment (105). The observation that a gene can exhibit a chromatin structure similar to that of active loci in precusor cells has been made at other loci, such as the lysozyme locus (106), c-fms (107), and the myeloperoxidase gene (108). Understanding how epigenetic states are acquired during development and how they impact globally on gene expression is a critical step in the treatment of a number of diseases, ranging from birth defects to cancer (109). A logical first step in this process would be to determine the mechanisms involved in this process at the level of individual genes. Previously we had proposed a model describing the processes involved in the activation of the β–globin locus during development (46). A key aspect of this model was the initial transition in chromatin structure from a closed, inaccessible conformation to that of a more open one. We hypothesized that this first step could be mediated by transcription factors such as GATA-1, which would bind first at the LCR and perturb the structure of nucleosomes in this region. This would be followed by the recruitment of various chromatin remodeling complexes and transcriptional co-activators as well as complexes containing RNA polymerase II (Pol II). This would result in the nucleation of a transcriptionally competent chromatin conformation that could be perpetuated to gene promoter proximal regions to activate transcription. Thus, we proposed that the LCR acts as a center of attraction at early stages of hematopoietic development to ensure the proper activation of the genes in erythroid cells.

In this study we wished to investigate the hypothesis that chromatin structure modifications and factor recruitment occurs first at the LCR and sunsequently to the genes. We analyzed the events of Pol II recuitment and chromatin structure alterations at

the murine β–globin locus in un-induced embryonic stem cells (day 0), mesodermal cells (day 5), as well as that of primitive and definitive erythroid cells (day 12). Using chromatin immunoprecipitation (ChIP) we demonstrate that core elements of the LCR adopt a structure characteristic of transcriptionally active chromatin and recruit RNA polymerase II prior to erythroid differentiation in murine ES cells. Real time PCR analysis indicates that the locus is first activated at the LCR and that this state is perpetuated to more distal regions as the process of differentiation proceeds. Histone modifications and factor recruitment corresponding to a transcriptionally permissive state appear to be acquired prior to gene expression

Results

We examined the association of Pol II with the murine β-globin gene locus during *in vitro* differentiation of murine embryonic stem (ES) cells. In these experiments we utilized the ES/OP9 cell *in vitro* differentiation system described by Kitayima *et al*. (42). The ability of these cells to generate mice was not examined so their pluripotency was not directly confirmed, however they did express markers of early development such as Rex-1 and did not express any of the globin genes (Fig.3-1). Furthermore, we were able to generate cells of both the hematopoietic and nervous systems (data not shown). Total RNA was isolated from ES/STO and ES/OP9 cultures at the indicated points following the start of induction and treated with DNase-1 to remove genomic DNA. Reverse-transcription polymerase chain reaction (RT-PCR) was used to examine the developmental progression of cell samples and primer sets span introns in all cases except for the Rex-1 gene. The appearance of transcripts was visualized by polyacrylamide gel electrophoresis and staining with SYBR green. Day 0 cells are

composed of ES and STO cells grown in ES media containing LIF. These cells express the Rex-1 and β-actin genes but do not express any of the genes of the β-globin locus.

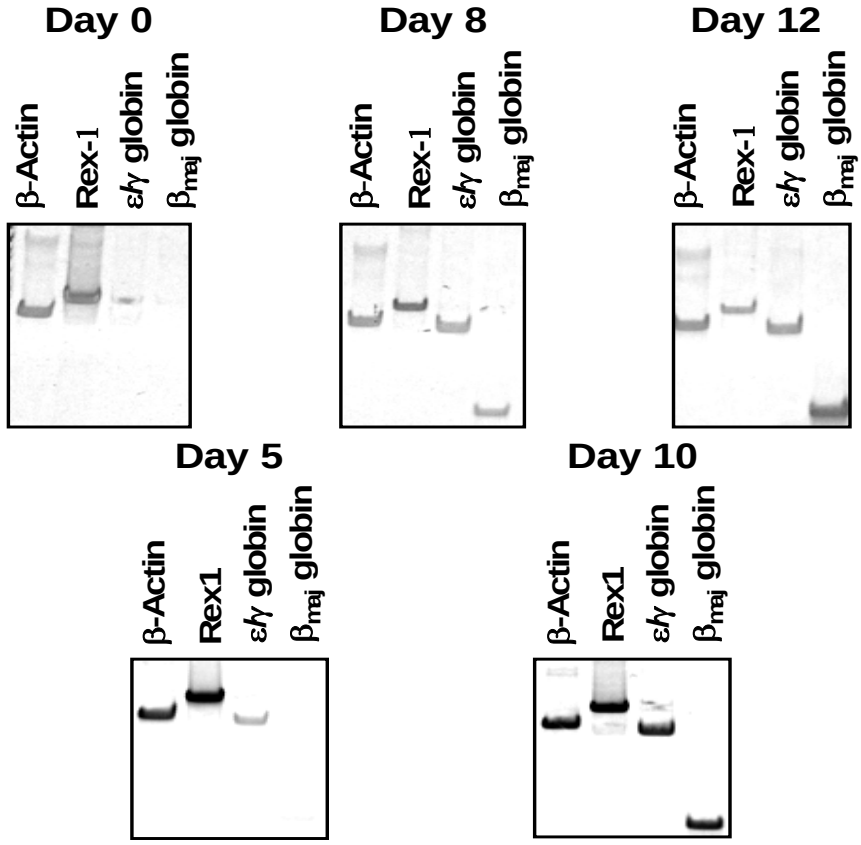


Figure 3-1. PCR analysis of DNase I treated and reverse-transcribed total RNA extracted from differentiating embryonic stem cells at the indicated time points. All primer sets span introns with the exception of Rex-1 and size of each RT-PCR product is as follows; Rex-1, ~600bp; β-actin, 480 bp; ε/γ̃globin, 400 bp; βmaj, 220 bp. None of the samples showed genomic DNA amplification (not shown).

Upon differentiation the embryonic- and adult-specific β-globin genes are activated sequentially. The εγ gene is activated first with transcripts appearing as early as day 5 of the time course (Fig.3-1). Expression of the adult specific gene is first observed at low levels at day 8 and is then up regulated upon the initiation of definitive erythropoiesis (Days 10-12). Expression of Rex-1 is reduced, although still detectable;

this is most likely due to residual undifferentiated cells present in culture rather than genomic DNA contamination.

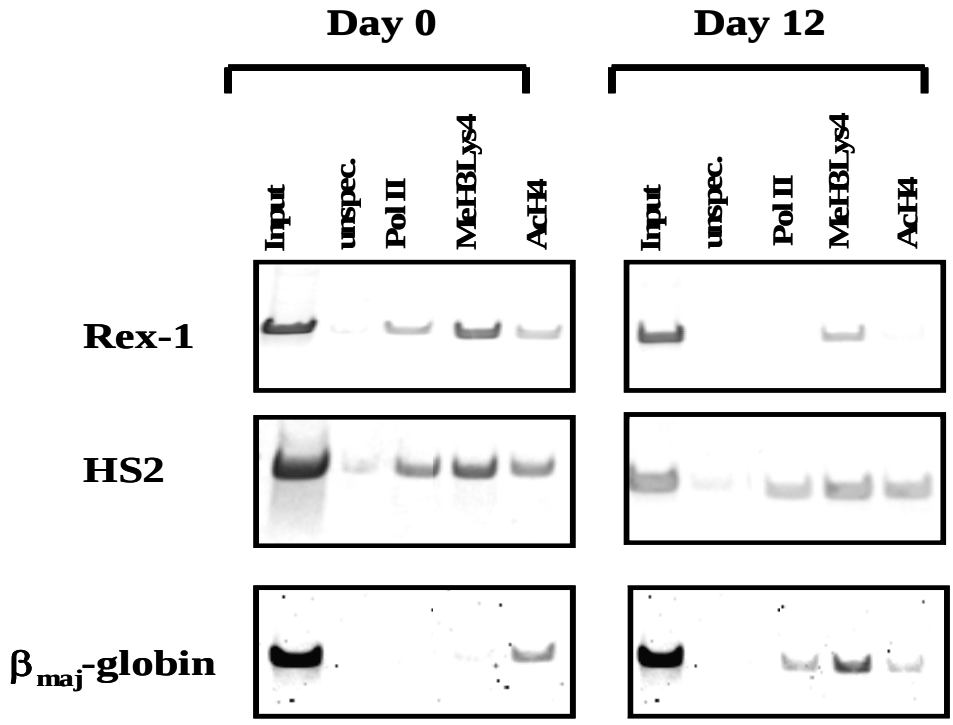


Figure 3-2. Interaction of transcription factors and RNA polymerase II with the β-globin locus. Differentiating ES cells were incubated in formaldehyde and the crosslinked chromatin was fragmented, isolated, and precipitated with antibodies specific for with chicken anti-IgG (unspec.), RNA polymerase II (Pol II), di-methylated histone H3 lysine 4 (MeH3Lys4), and acetylated histone H4 (AcH4). DNA purified from the precipitate was analyzed by PCR with primers corresponding to regions in the murine β-globin locus as indicated.

This is supported by the lack of genomic bands for the other primer sets. We analyzed the interaction of Pol II and the appearance of modified histones within the globin locus during the course of differentiation using the ChIP assay (Fig.3-2). We used antibodies specific for Pol II, acetylated histone H4 (AcH4) and for di-methylated lysine 4 of histone H3 (Me_2K4H3). Di-methylation of H3 at lysine 4 is associated with regions

permissive for transcription (110). The Pol II-specific antibody detected both phosphorylated and un-phosphorylated forms of the protein. Each antibody was used in three independent experiments.

The results show that Pol II and Me_2K4H3 are present at the LCR but not at the β-globin gene in undifferentiated ES cells (day 0) suggesting that Pol II recruitment to the LCR occurs before activation of any of the globin genes (Fig.3-2 and 3-3). The presence of H3 di-methylated at K4 indicates that these elements are permissible to active transcription; an observation supported by the detection of transcription through HS2 and HS3 cores (data not shown). This lysine methylation is specific to the core regions of the HS sites and this mark is not detected in a region between the HS2 and 3 cores (3/2Flank) and no transcription within this region is detected (not shown). The β-globin gene is associated with acetylated histone H4 but not H3 dimethyl-K4 suggesting that the chromatin structure has been modified but is not transcriptionally active. The Rex-1 gene is associated with a chromatin structure characteristic of an open, transcriptionally active domain. We examined both the promoter and the transcribed region of the Rex-1 gene (not shown).

In differentiated erythroid cell samples containing both mature primitive and definitive precursors (Day 12) Pol II is associated with the LCR and the β-globin gene (Fig.3-2 and 3-4). The enrichment of H3 di-methylated at lysine 4 in these regions is compatible with an open, transcriptionally active chromatin structure. In contrast, the Rex-1 gene is repressed at this stage and this is accompanied by a reduction in K4 di-methylated H3 and acetylated H4 as well as a decrease in Pol II recruitment at the promoter. These results demonstrate that prior to erythroid differentiation, the LCR lies

within a transcriptionally competent chromatin structure and is associated with components of the basal transcription machinery (Pol II).

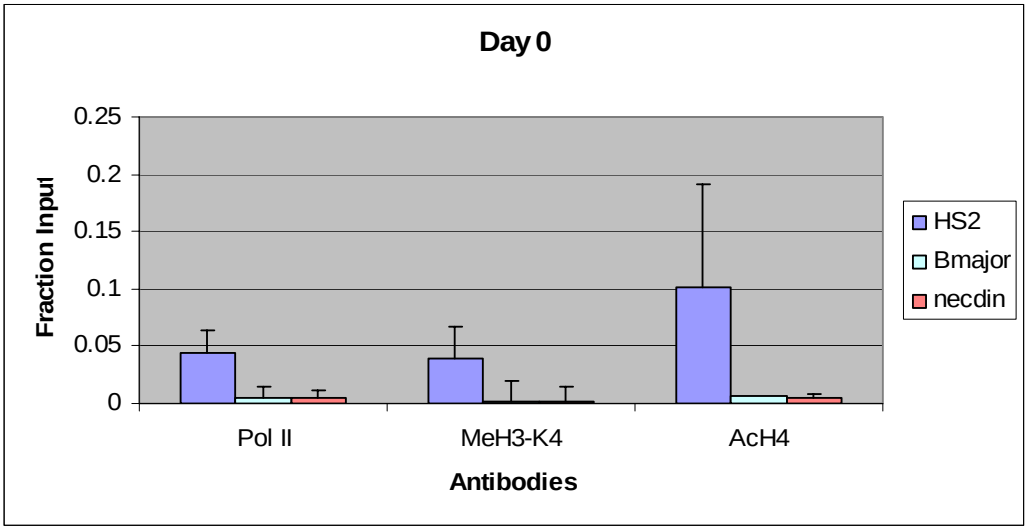


Figure 3-3. Quantitative analysis of Pol II binding and histone modifications within the β-globin locus at "Day 0" in murine ES cells. Values were calculated using standard curve titration of input samples.

We next employed real-time PCR analysis to obtain a quantitative measure of factor binding and histone modifications during differentiation (Figs.3-3 through 3-5). We hoped to show that the locus underwent changes in factor recruitment and chromatin structure as the genes became active. Quantitative analysis of these samples and comparison of the values obtained throughout the time course to those obtained for the neuronal specific necdin gene show that Pol II recruitment increased as much as five-fold at the globin gene promoters and approximately 20-fold at the core regions of the LCR (Fig.3-5). The concentration of histone modifications associated with transcriptional activation at the promoter regions increased from two to five-fold. Distinct differences between the changes at the embryonic and adult stage-specific promoters were observed. The significance of this observation is not known but may reflect different mechanisms involved in the regulation of these genes. At the embryonic βH1 gene promoter there was

a detectable increase in methylation of histone H3 lysine 4 but not in acetylated H4. The exact opposite was observed at the βmajor gene promoter. At HS2 both di-methylation of histone H3 lysine 4 and acetylation of H4 increased during differentiation, although the increase in histone H4 acetylation was more dramatic.

Although not shown we also examined the association of these factors at time points that should contain pre-hematopoietic cells (day 5) and definitive hematopoietic precursors (day 10). We also chose to analyze only the adherent cells in these cultures in contrast to the previous experiments where both adherent and floating fractions were obtained. The major motivation for this was the fact that a day 5 very few cells were actually floating freely in the media. We wished to use equal numbers of cells in these assays. Using the same ES/OP9 system Suwabe *et al*. found that cells isolated from the floating fraction contained a greater abundance of mRNAs for the εγ- and βmajor–globin genes compared to that of the adherent fraction (111). The adherent fraction of these cultures contained approximately 10 times more erythroid- and granulocyte-monocyte colony-forming units than the floating fraction (111). We hypothesized that the adherent fraction may contain cells whose chromatin structure was indicative of a less differentiated cell type than that of those floating freely, a conclusion supported by the presence of more precursor cells in these samples. In day 5 cells we observed low levels of transcription of the embryonic εγ-globin gene and little or no detectable transcription of the adult specific βmajor-globin gene. By day 10 both genes were active. Consistent with previous experiments RNA Pol II was present at the LCR in significant quantities, however very little was observed at the gene promoters and levels did not appear to increase as before. Similar results were obtained for MeH3K4 and AcH4 antibody

fractions. Most importantly we were unable to observe the formation of developmental stage –specific subdomains. Our results agreed with the analysis performed by Suwabe and colleagues using a similar ES/OP9 cells culture system. They showed that the adherent cells present in ES/OP9 cultures were enriched for precursor cells and that the structure of the β-globin locus was similar to that found in precursor cells isolated from other systems (105). Further analysis of factor recruitment and histone modifications at these points is ongoing.

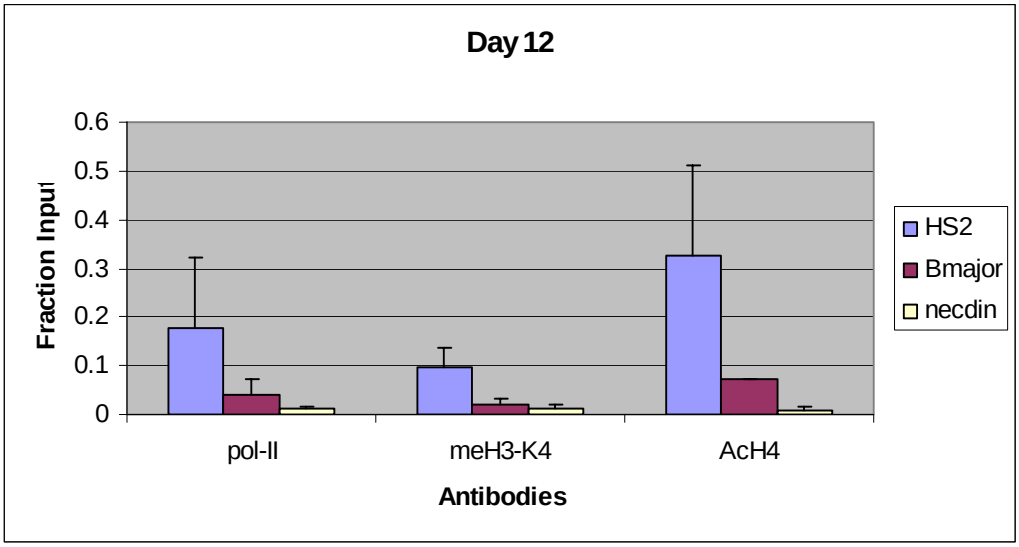


Figure 3-4. Quantitative analysis of Pol II binding and histone modifications within the β-globin locus at "Day12" in differentaited murine ES cells. Values were calculated using standard curve titration of input samples

Discussion

The commitment of pluripotent stem cells to successively less plastic progenitors and finally, differentiated cells exhibiting stable expression patterns is thought to involve the reorganization of the chromatin environment of many lineage-specific genes. The timing of these changes, in many cases, has been shown to precede gene transcription (103, 105, 107). In the present study we have assessed the temporal nature and extent of

covalent histone modifications and association of transcrition complexes at the murine β-globin locus during the *in vitro* differentiation of murine embryonic stem cells. We have observed that elements of the β-globin LCR are capable of recruiting RNA polymerase II and histone modifications compatible with transcription prior to lineage specification. We also observe transcription of HS elements of the LCR in undifferentiated ES cells. These results suggest that the β-globin locus may already exist, in part, in a transcriptionally active state very early during development. It appears that in the context of this system it remains so in a number of pre-hematopoietic precursor cell populations and undergoes a number of alterations in chromatin structure and factor recruitment as these cells progress towards hematopoietic commitment. Quantitative analysis shows that recruitment of transcription complexes and histone modifications are present in greater abundance at the LCR compared to the gene promoters. This is consistent with the idea that the LCR may be activated in a number of hematopoietic and pre-hematopoietic cell types, whereas the activation of the genes is restricted to that of the erythroid lineage. Whether or not this is a requirement for the proper stage-specific activation of the genes is not known. It may be that these observations are a bi-product of the generally non-repressive chromatin environment of ES cells and that the locus lacks these marks in a number of cells that are still capable of expressing the globin genes. It would be of interest to study the conformation of other gene loci that contain LCRs to examine the possibility that a high degree of accessibility in multipotential precursors and ES cells is an intrinsic quality of these dominant regulatory elements.

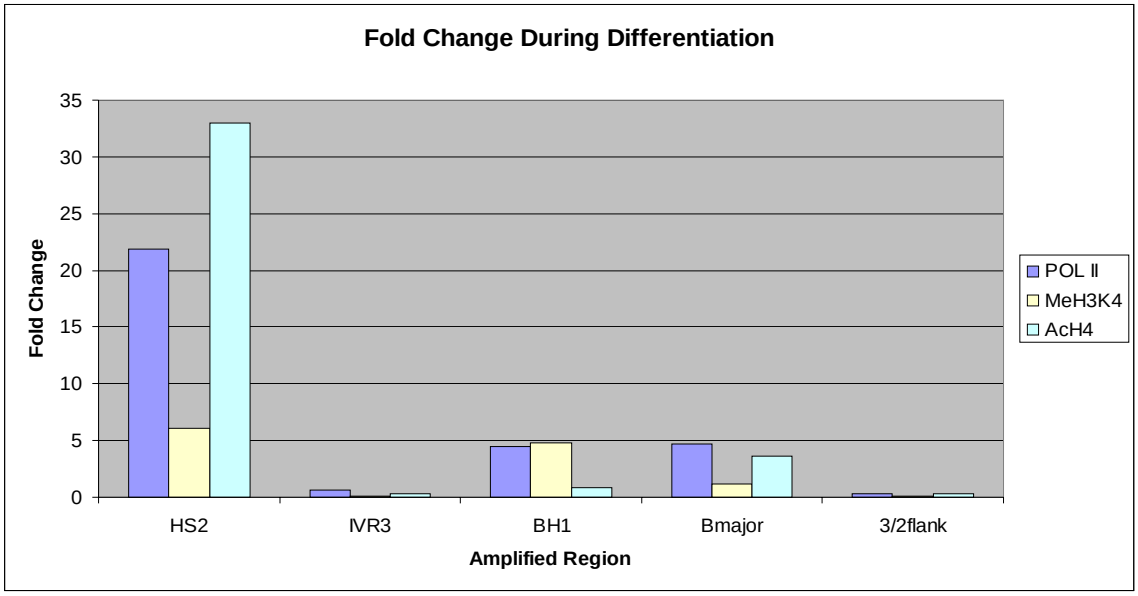


Figure 3-5. Fold change in factor binding and histone modifications at the murine β-globin locus during *in vitro* differentiation. Values are shown on the Y-axis as the fold difference compared to the values obtained for the necdin promoter.

However, from this data we can construct a very basic picture describing the two observed states of the locus in our studies (Fig. 3-6). In "Day 0" precursor cells the LCR already exhibits characteristics of a transcritpionally active region. These include histone modifications and recruitment of transcription complexes containing RNA Pol II. These activities may be recruited to the LCR so that they may be deposited at the gene promoter regions to activate transcription at the proper developmental stage. Transcription through the LCR core regions may facilitate these interactions. In differentiated cells that express the globin genes these marks are present at both the LCR and the gene promoters. How these activities are transferred is not known. It may be through direct interaction of the LCR with gene proximal elements (looping) or by the movement of transcription complexes and associated chromatin modifiers along the DNA (linking). Whatever the mechanism it is clear that the LCR plays an important role in the initial reorganization of the locus prior to high-level expression of the globin genes in erythroid cells. Recently a

study by Szutorisz and others produced similar observations for the B-cell specific *VpreB1* and λ5 genes (112). They characterize a *cis*-acting element in this locus marked by H3 acetylation, H3 lysine 4 di-methylation, and Pol II recruitment in ES cells and show that these marks occur independently of the recruitment of any lineage-specific transcription factors such as PU.1. Furthermore, they observe the presence of components of the TFIID complex (TAF 10 and TBP) to this element in ES cells. They label these marks collectively as the early transcription competence mark (ETCM) and substantiate its importance by making light of the fact that subsequent, similar modifications appear to spread outward in both directions to the genes it controls. This is identical to the observed appearance of these marks at the LCR of the globin locus in ES cells followed by the genes in our ES/OP9 cultures. These results call into question the notion of differences between the pluripotent and multipotent states. These results suggest that the progression from a progenitor cell to that of a terminally differentiated and functional cell cannot be explained by a simple cascade of gene repression and activation, since many tissue-specific genes appear to be poised for activation prior to commitment.

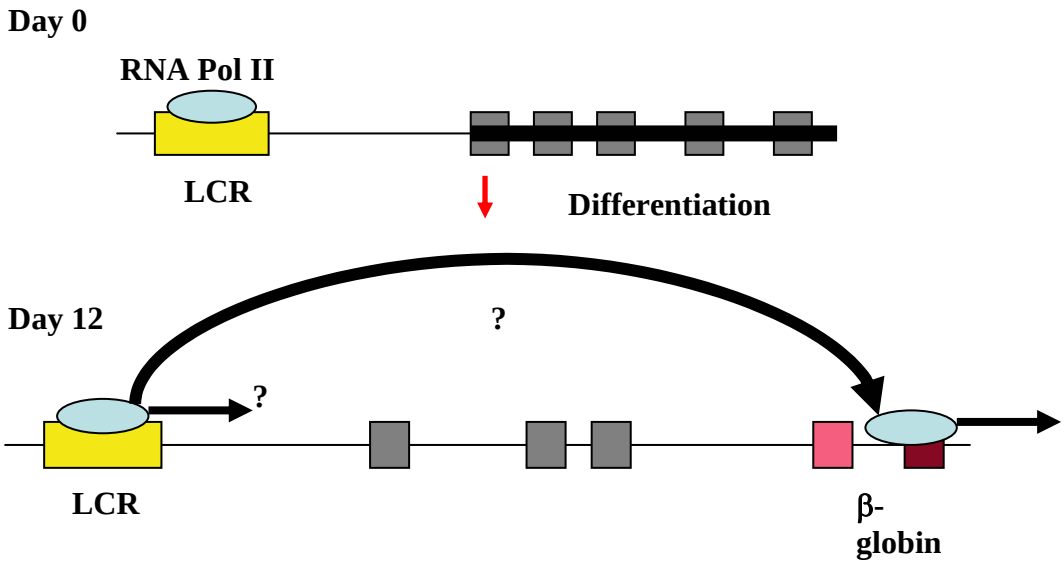


Figure 3-6. Recruitment of transcription complexes and chromatin structure alterations at the β-globin locus in undifferentiated precursors and definitive erythrocytes. Transcription complexes recruited to the LCR in uncommitted progenitors could be delivered to gene promoters directlt through looping of the intervening DNA or via a tracking mechanism.

CHAPTER 4
COMPOSITE β-GLOBIN LOCUS TRANSGENES EXHIBITING POSITION-INDEPENDENT EXPRESSION IN TRANSGENIC MICE

Introduction

Transgenic mice have proven to be an excellent system for the *in vivo* analysis of how the regulatory elements involved in gene expression function. From such studies we have identified elements critical for the expression of a large number of genes including HOX genes as well as those of the growth hormone and globin loci. As informative as such studies are the results gleaned from analyzing the behavior of genes at ectopic locations within the genome is often confounded by the influence of the surrounding chromatin environment, a phenomena known as position effect variegation (PEV) (113). Position effects can often cause disparate levels of gene expression between independent lines harboring identical transgenes, an important fact when considered in the context of gene therapy. In order to circumvent the difficulties associated with chromatin environment many researchers have sought to identify dominant regulatory elements able to protect genes from position effects in transgenic assays. One class of elements capable of such a feat is locus control regions (LCRs). LCRs are often composite elements containing multiple core regions that exhibit heightened sensitivity to nucleases in specific cell types (114-116). These HS sites can be clustered or spread throughout a gene locus (115, 117). Exactly how LCRs function is not entirely known, although many studies have shown that the LCR HS sites of the human β-globin gene locus synergize to confer high-level and position-independent expression (77, 118-121). Results from

genetic as well as conformational studies suggest that the HS sites of the β-globin LCR interact with each other and with the genes they activate at a particular developmental stage (45, 77, 78, 122). These interactions could establish a configuration that protects the genes from negative effects exerted by neighboring chromatin (48). Another group of regulatory DNA elements that protect genes from position of integration effects are the so-called boundary elements. These elements were first discovered as sequences that protect transgenes from position effect variegation in *drosophila* (113). Similar elements were discovered in higher eukaryotic cells. For example matrix attachment regions (MARs) flanking the chicken lysozyme gene locus protect the gene from position effects in transgenic mice (123). Likewise, elements flanking the chicken β-globin gene locus, cHS4, are able to protect reporter genes in transgenic assays (96, 124). Chicken HS4 harbors two distinguishable activities; it blocks the function of enhancers on activating promoters, and it establishes a boundary between open and closed chromatin (124). While enhancer blocking activity of cHS4 is mediated by CTCF (125), proteins mediating the boundary function remain to be determined, although USF proteins have recently been shown to be involved through binding the element and recruitment of histone modifying activities (126). Results also suggest that some insulators may be tethered to nuclear compartments, e.g. nuclear pore complex or nucleolus, via interactions with proteins known to reside in these structures (127, 128).

A promising strategy for avoiding position of integration effects is to direct the transgene into a specific site in the genome. Several strategies have been applied to directing transgenes into specific genomic sites (100, 129, 130). In this study we utilized components of the adeno associated virus (AAV) integration machinery with the goal to

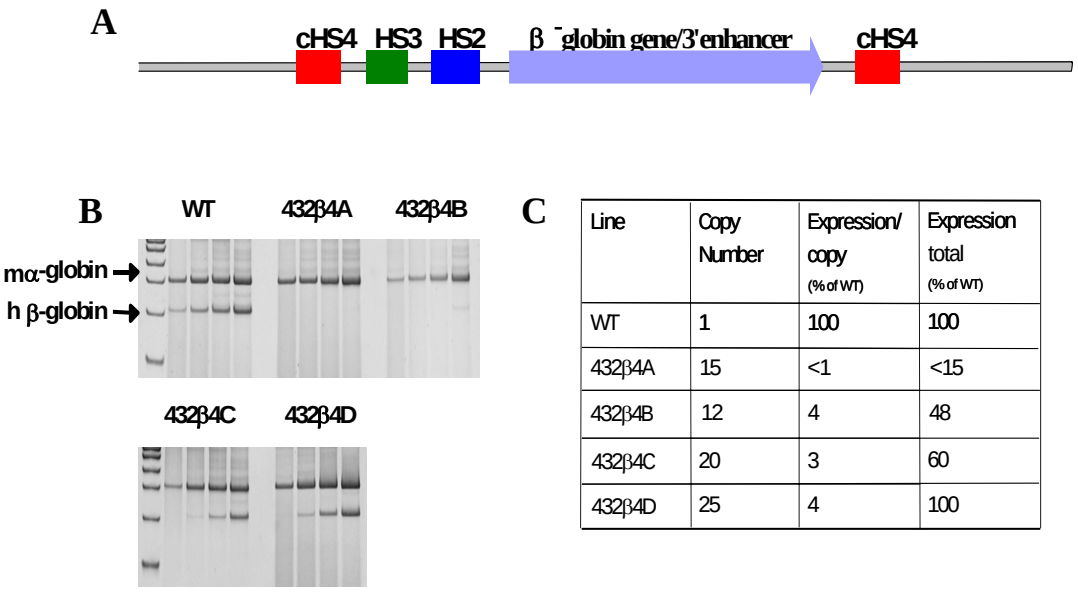
direct a human β-globin expression construct into a specific site in the mouse genome. AAV is a small human virus that is able to establish latent infection by integrating its DNA into a specific site on human chromosome 19, called the AAVS1 site (131, 132). Integration is mediated by DNA sequences present in the viral genome as well as by the AAV encoded rep protein, which contains ATPase, helicase, and DNA-nicking activities (133-135). We generated transgenic mice containing the human AAVS1 integration site. The β-globin expression construct contained sequences from AAV that have previously been shown to be critical or to enhance integration into the AAVS1 site (136). This construct was incubated with recombinant AAV rep protein and the mixture injected into the pronuclei of fertilized murine oocytes transgenic for the AAVS1 site. Despite using different conditions and rep/DNA molar ratios none of the transgenic mice had the transgene integrated into the human AAVS1 site.

During the course of this study we have analyzed integration sites and expression levels of two different β-globin gene constructs. The first construct contained LCR core elements HS2 and HS3, the β-globin gene, and the β-globin gene 3'enhancer, flanked by insulator elements from the chicken β-globin gene locus (cHS4). The second construct differed from the first one in that we included the HS2/HS3 flanking sequences. Both of these constructs expressed the β-globin gene from different positions in the murine genome. However, the construct containing the HS2/HS3 flanking sequence consistently revealed higher β-globin expression levels, even when integrated in or close to a centromere. This suggests that the HS2/3 flanking region facilitates the activation by LCR core elements.

Results

We began our studies by generating and analyzing transgenic mice harboring an expression construct of the human β-globin gene that is small enough to be packaged into recombinant AAV (rAAV). This construct contained the core sequences of LCR HS sites 2 and 3, the human β-globin gene, and the β-globin 3'enhancer. These elements were flanked on either site with a single copy of an insulator sequence derived from the chicken β-globin gene locus (cHS4). The rationale for the inclusion of the regulatory elements was the following: HS2 has been shown to have strong enhancer activity when linked to globin or other reporter gene constructs (137) and HS3 has been shown to harbor both enhancer and chromatin opening activity (138). We did not include human LCR element HS4 because results from previous studies suggest that it does not contribute unique activities for β-globin gene activation (77, 119, 139). Chicken HS4 (cHS4) exhibits both enhancer blocking as well as boundary activities (124). Finally the β-globin 3'enhancer has been shown to be important for high level β-globin gene expression in the context of the complete locus in β-globin yeast artificial chromosome (β-globin YAC) transgenic mice (140). We hypothesized that HS2 and HS3 would open the chromatin regardless of the transgene integration site and that the presence of cHS4 would protect the β-globin gene expression construct from any negative effect exerted by surrounding chromatin at the site of integration. We have generated and analyzed four transgenic lines with this construct. The copy number of the transgene was determined by southern blotting experiments in which a single copy β-globin YAC transgene was used as a standard and the murine snrp N gene as an internal control (data not shown). The single copy line contained the entire human β-globin locus in the context of a YAC (141). β-globin gene expression in the transgenic lines was analyzed by semi-

quantitative RT-PCR using pairs of primers specific for human β- and murine α-globin cDNAs (77). Murine α-globin expression levels were used as an internal control and β-globin gene expression (β/α-globin) was calculated as percent expression of that of the single copy β-globin YAC transgenic line (set at 100%).



Line	Copy Number	Expression/ copy (% of WT)	Expression total (% of WT)
WT	1	100	100
432β4A	15	<1	<15
432β4B	12	4	48
432β4C	20	3	60
432β4D	25	4	100

Figure 4-1. Structure and expression of the 432β4 transgene. A) Structure of the transgene showing cHS4 insulators, hypersensitive site cores 3 and 2 (HS3, 2) from the human β-globin locus and a β-globin gene/3'enhancer. B) PCR analysis of cDNA reverse transcribed from RNA isolated from anemic spleens of transgenic mice. Human β-globin and endogenous mouse α-globin are shown (shown are the signals for 14, 16, 18, and 20 PCR cycles). C) Summary of expression levels and copy numbers in the individual transgenic lines shown in Panel B. RT-PCR signals were quantitated by phosphorimaging. Expression levels were calculated based on expression of the mouse α-globin gene and presented as % with expression levels in the human β-globin YAC transgenic line set as 100%.

β-globin gene expression is shown as expression per copy or total. The data show that expression per copy is low in these transgenic mice, demonstrating that the combination of regulatory elements present in the 432β4 expression construct is not sufficient to confer high-level β-globin gene expression (Fig.4-1). The fact that all of the lines do

express the β-globin gene indicates that the construct is likely protected from position effects due to the presence of the cHS4 insulator elements.

Because the inclusion of additional regulatory elements would render the transgene too large for packaging into AAV, we decided to pursue an alternative strategy that, if successful, would allow us to integrate the transgene into a defined site into the mouse genome and which would not be limited by the size of the DNA construct. Wild-type AAV integrates its DNA into a specific site on human chromosome 19, called AAVS1 (131). Although the mechanism of integration is not entirely clear, the process requires *cis*-acting elements present in the viral genome as well as the function of the AAV encoded rep protein. The AAVS1 sequence was ligated as a 3.6 kb fragment into the EcoRI/KpnI restriction sites of the vector pBS246 leaving a single *loxp* site at the 3' end of the AAVS1 site. The presence of a single *loxP* site would allow us to eventually reduce the copy-number by Cre mediated recombination (142). The construct was linearized with EcoRI and used to generate transgenic mice. We generated two transgenic lines, however, only one of these lines continued to transmit the AAVS1 sequence. Southern blotting experiments showed that this line contained the AAVS1 plasmid integrated in 5 tandem copies (data not shown). We mapped the position of the transgene by inverse PCR and DNA FISH analysis. The AAVS1 transgene integrated within the intron of a gene predicted to encode a metalloprotease and located near the telomere of chromosome 15 (Fig. 4-2).

We next generated a larger β-globin expression construct that is similar to the one described in Fig.1 but contains in addition the flanking region of the HS2 and HS3 core enhancers (43f2β4, Fig. 3A). Previous work has shown that inclusion of DNA flanking

the core regions allows the LCR HS sites to synergistically activate globin gene expression (120). The final construct is about 11 kb in size and contains DNA sequence elements derived from the AAV genome, the 5'inverted terminal repeat (ITR) and the p5 promoter region. Both of these sequences have been implicated in the integration of wild-type AAV into the S1 site (136). These elements were placed outside of the 5'cHS4 sequence. The supercoiled plasmid DNA was incubated on ice with recombinant rep68 and the mixture injected into the pronuclei of fertilized oocytes homozygous for the AAVS1 integration site. The offspring was first analyzed by PCR for the presence of the β-globin transgene. We generated four transgenic lines with the second β-globin gene

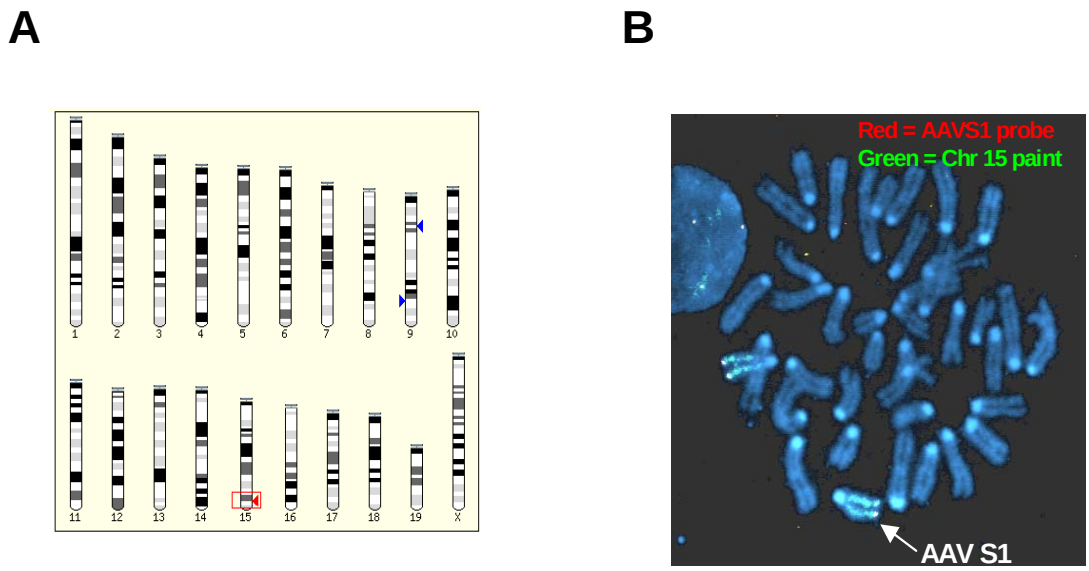
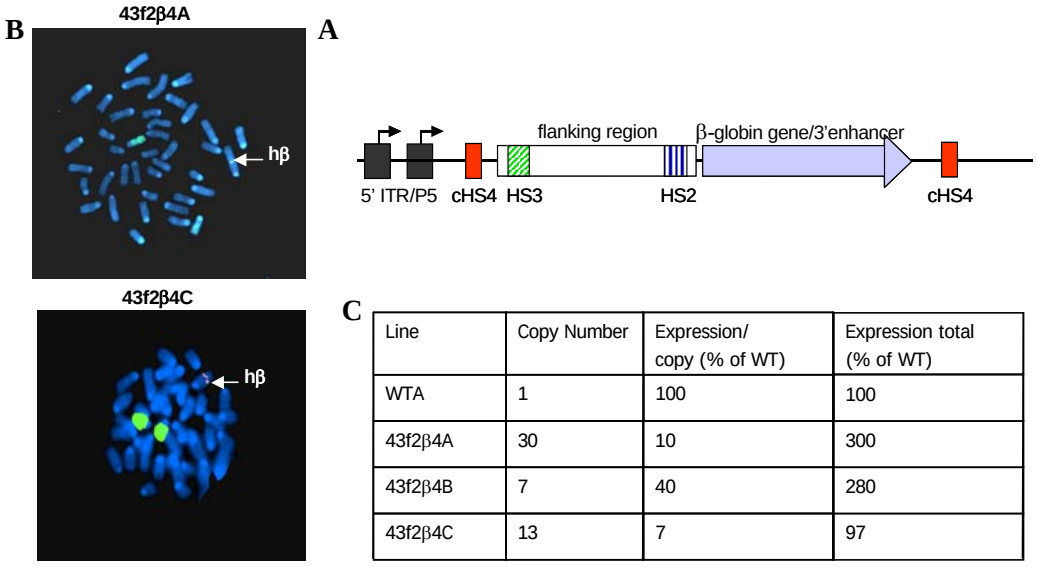


Figure 4-2. Location of the human S1 integration site in the mouse genome. A)Inverse PCR was used to obtain sequence information from the site of transgene integration. This sequence was blasted against the mouse genomic database (www.ensembl.org). A perfect match to the sequence is located close to the telomeric end of chromosome 15. B) Confirmation of the integration site by DNA FISH. Metaphase spreads of spleen cells from transgenic mice were hybridized using a fluorescent probe specific for the human AAVS1 site (red) and a chromosome paint specific for chromosome 15 (green).

construct (43f2β4); the first line did not transmit the transgene. The three remaining lines were bred with AAVS1 transgenic mice to generate mice heterozygous for the β-globin expression construct. Determination of copy number and integrity as well as expression analysis was performed as described for the first construct. We also analyzed the integration patterns of these mice using DNA FISH. The expression analysis in these mice is summarized in Fig. 4-3C. All three lines expressed the β-globin gene at higher levels than those harboring the smaller expression construct, supporting previous findings that the flanking sequences allow the HS core sites to synergistically activate β-globin gene expression (120).



C

Line	Copy Number	Expression/ copy (% of WT)	Expression total (% of WT)
WTA	1	100	100
43f2β4A	30	10	300
43f2β4B	7	40	280
43f2β4C	13	7	97

Figure 4-3. Expression analysis of a β-globin construct containing insulator sequences as well as LCR elements HS2 and 3 with their flanking DNA in transgenic mice. A) Structure of the 43f2β4 plasmid. B) Integration pattern of 43f2β4 transgenic lines (43f2β4 A and C) analyzed by DNA FISH. C) Summary of human globin gene expression and copy number in transgenic mice harboring the 43f2β4 construct (43f2β4 A to C). Analysis performed as described in Fig.4-1.

We next used DNA FISH to determine the site of integration of these constructs within the mouse genome. The results demonstrate that the human β-globin gene

construct did not integrate into the transgenic human AAVS1 site on chromosome 15 (Fig.4-3 and 4). Recently, a mouse ortholog of the human AAV S1 site has been identified. This sequence is located in the peri-centromeric region of chromosome 7. Because the 43f2β4B transgene integrated into or close to a murine centromere we examined whether the murine ortholog of the human AAVS1 site has been targeted. However, southern blotting experiments and DNA FISH with a chromosome 7 paint revealed that it did not integrate into the mouse S1 site (data not shown).

Analysis showed that average expression of the β-globin gene in the context of 43f2β4 construct was roughly four fold higher per copy than that of the 432β4 construct, and total expression was five fold higher in 43f2β4 transgenic lines (Fig.4-5). The line 43f2β4B expressed the human β-globin gene at 40% of levels found in a single copy β-globin YAC transgenic mouse. This was the highest level of expression per copy in all of the transgenic lines analyzed in this study. The FISH analysis revealed that the transgene integrated into a region located at the centromeric end of a mouse chromosome as indicated by the intense DAPI staining (Fig.4-4). The expression analysis as well as the mapping of DNase I hypersensitive sites demonstrates that the transgene is open and active in this location.

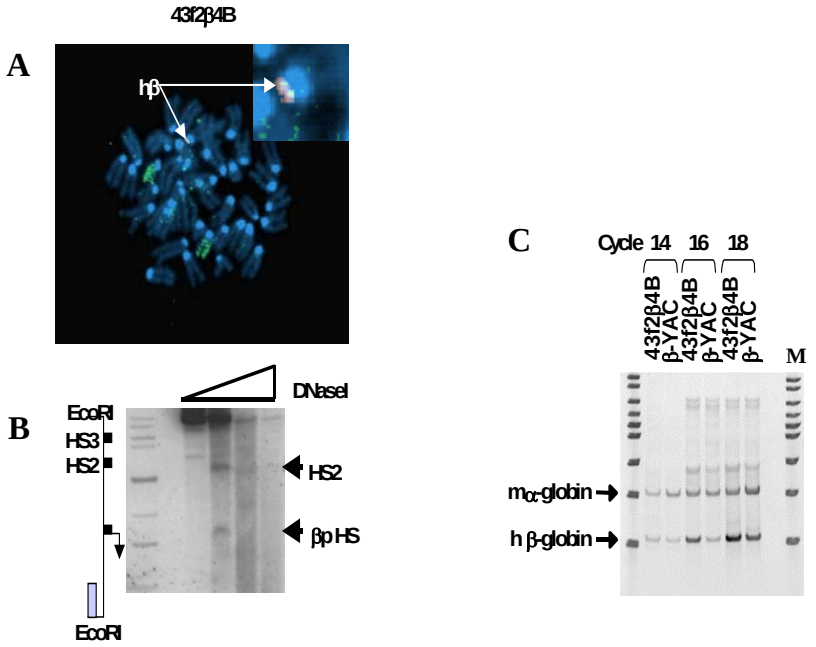


Figure 4-4. Analysis of the integration site, β-globin gene expression, and DNase I HS sites in transgenic line 43f2β4B. A) Analysis of the 43f2β4B integration site by DNA FISH. Metaphase spreads of spleen cells from transgenic mice were hybridized to fluorescent probes specific for the plasmid 43f2β4 (red) and chromosome 15 (green), stained with DAPI and visualized by fluorescence microscopy. The insert is a magnification of the region containing the 43f2β4 integration site. B) Analysis of DNase I HS sites in the 43f2β4 transgene. Spleen nuclei of anemic transgenic mice were digested with increasing concentrations of DNase I. The genomic DNA, isolated from these samples, was digested with EcoRI, size-fractionated by gel-electrophoresis, and transferred to a nylon membrane. The DNA was hybridized to a radioactive probe corresponding to a region just 5' of the downstream EcoRI site as indicated. The arrows indicate the position of HS sites associated with LCR element HS2 and the β-globin promoter. Lane M represents a 1 Kb ladder. C) Analysis of β-globin gene expression by RT-PCR in 43f2β4B and b-globin YAC (WT) transgenic mice. Expression of the human β- and mouse α- globin genes was analyzed as described in Figure 4-1 Panel B. Shown are the signals for 14, 16, and 18 PCR cycles.

Discussion

Hemoglobinopathies are among the most common inherited diseases in the human population (143). Due to problems associated with current treatments, alternative therapies are highly sought after. Ideally one would like to use gene therapy to deliver corrected copies of the mutated gene into cells of the hematopoietic system. For example,

investigators have successfully used lentiviral vectors to deliver therapeutic β-globin expressing constructs into hematopoietic cells of mice carrying mutations in the globin locus (121, 144). However, the use of lentiviral vectors may not be without problems because the viral DNA is integrated more or less randomly into the genome (145). Adeno-associated virus is considered to be safe because it does not cause strong immune reactions and DNA of recombinant viruses used in gene therapy experiments often remain episomal (146), although some data suggest that AAV serotype 2 integrates preferentially into transcriptionally active regions in the nucleus (147). The disadvantage of AAV is its low packaging capacity. This is particularly problematic in terms of the β-globin locus as it appears to require a complex set of regulatory elements to achieve physiologically relevant levels of gene expression. We wished to avoid this drawback by exploiting elements of the AAV system involved in site-specific integration without actually packaging the DNA into a virus.

Despite numerous attempts, using a number of different conditions, this strategy proved unsuccessful. The components used in our studies have been shown sufficient for AAV integration into the S1 site (136). However since the normal means of delivery is through viral infection critical steps may be bypassed in our injection procedure. The chromatin environment surrounding the integration site of our S1 transgene may also play a role as it appears to be integrated within a telomeric region of chromosome 15. This however, is unlikely as transgenes are capable of integrating randomly throughout the genome with no apparent bias for chromatin structure. Further investigation into the conditions required for rep-mediated integration into our transgenic AAV S1 site is ongoing.

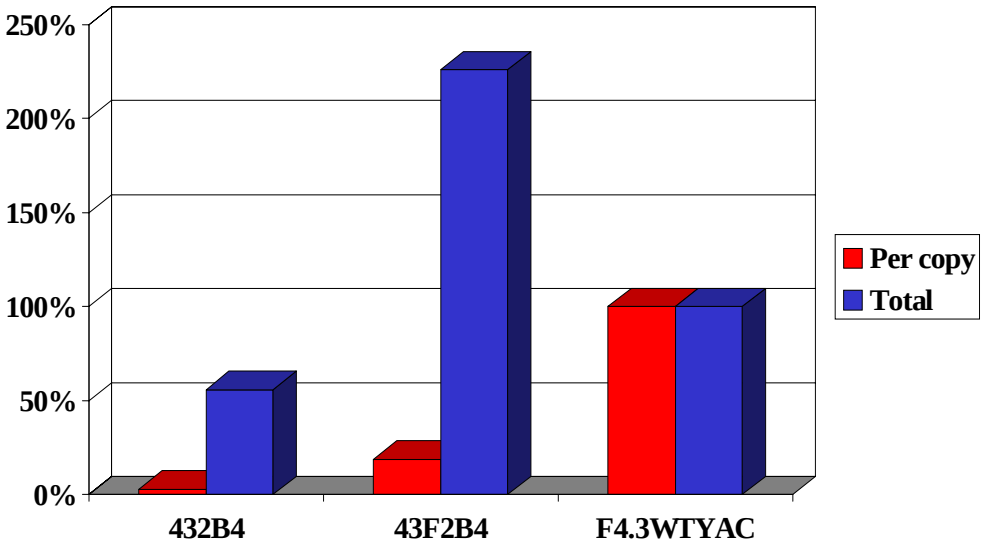


Figure 4-5. Relative expression levels of two β-globin transgenic constructs. Average expression levels of 432β4 and 432fβ4 transgenic constructs as compared to that of a single copy F4.3 YAC containing and intact human β-globin locus. The total expression and expression per copy is shown.

Our initial experiments were aimed at investigating the ability of a construct that was capable of being packaged into the AAV virus to exhibit position-independent expression at physiologically relevant levels. We included what we considered to be the minimal genetic elements from the β–globin locus in combination with a known boundary element in these constructs. The data show that inclusion of these elements is not sufficient to confer high-level β-globin gene expression. The levels of β-globin gene transcription observed could be due to the fact that the regulatory elements fail to independently establish an active chromatin domain permissible for transcription. The second construct containing the HS2 and HS3 flanking DNA led to higher expression levels in all the transgenes analyzed supporting previous conclusions that the cores function better in the presence of flanking regions (120). This is in line with the LCR holocomplex model according to which the HS sites interact with each other to activate globin gene expression (68, 77). It may be that the flanking sequences serve an

architectural role allowing the HS cores to interact with one another. Recent conformational studies suggest that the LCR HS sites are in close proximity in erythroid cells and that the genes that are expressed at a specific developmental stage contact the LCR holocomplex (48). Formation of the LCR holocomplex and its association with other HS sites in the globin locus, a complex termed the "chromatin hub", could establish an architecture that is resistant to negative influences exerted by neighboring chromatin at the site of transgene integration. In this respect, it is possible that the larger construct is able to establish such an architecture or micro-domain, whereas the smaller one, due to the absence of the HS core flanking DNA, is unable to do so. Alternatively, the HS 2/3 core flanking region may harbor additional regulatory elements that cooperate with the core HS sites to enhance globin gene transcription a notion supported by the fact that evolutionarily conserved sequence elements exist within in these regions (74, 120). There is also the possibility that the AAV ITR/p5 promoter may influence transgene expression although this is most likely not the case as it lies outside the cHS4 elements that have been shown to be highly efficient in enhancer blocking (96).

The highest level of β-globin gene expression was observed in a transgenic line in which the construct integrated into a centromeric region. Centromeres are usually incompatible with transcription, although it has been reported that genes can be expressed within functional centromeres (148). We do not yet know whether the human β-globin construct integrated into heterochromatic centromeric repeats, but the DNA FISH result shows that the transgene is located at the very tip of the chromosome in a region of intense DAPI staining. The data thus demonstrate that HS2 and HS3 in combination with insulator sequences are able to confer high-level expression to the β-globin gene in a

centromeric location. It should also be noted that this mouse line carries the lowest number of transgenic copies among all the transgenes analyzed here. Copy-number itself can influence the expression levels of transgenes. Previous work has shown that a high number of copies can repress transgene expression, possibly due to the fact that highly repetitive DNA tends to fold into a heterochromatic structure (142).

In summary, our data demonstrate that the activity of LCR HS sites is enhanced in the presence of their flanking DNA in transgenic mice. The combination of specific LCR HS sites and boundary elements can provide high-level expression to a β-globin transgene even when integrated in a centromeric location.

CHAPTER 5
SITE-SPECIFIC INTEGRATION OF MODIFIED YEAST ARTIFICIAL CHROMOSOMES IN MOUSE EMBRYONIC STEM CELLS BY RECOMBINASE MEDIATE CASSETTE EXCHANGE

Introduction

The best way of examining the function of genetic regulatory elements is in the context of an intact organism. For the study of mammalian gene regulation the transgenic mouse has proven to be an extremely useful animal model due to its rapid proliferation and development as well as extensive genomic sequence information. However, the generation of transgenic mice or any transgenic organism is not without its pitfalls. Position of integration effects (PEV) first observed in *drosophila* often complicate transgenic studies concerning the roles of regulatory elements by altering the context in which they are observed. Position effects describe the influence of the chromatin environment around the position that a transgene integrates; a transcriptionally permissive environment can have drastically different effects on transgene expression than one that is by nature repressive, such as peri-centromeric regions. Furthermore, the orientation of integration may impact upon transgene expression. Feng and colleagues demonstrated that position of integration effects can, in some instances be highly dependent upon the orientation of the construct within the integrated locus, with one orientation being repressive and the other permissive to expression (149). The random manner in which transgenes integrate into the genome when introduced into fertilized oocytes often precludes comparative assessment of transgene behavior. Moreover, in many cases one would like to study the behavior of gene and corresponding regulatory

elements in a specific context, say that of constitutive heterochromatin, or juxtaposed to an inactive gene to test for promiscuous interactions. Homologous gene targeting strategies (HR) are commonly applied in order to introduce mutations or genes of interest into defined locations, however HR is highly inefficient often occurring at frequencies as low as 1:1000 when compared to that of random insertion (150). Random insertion can result in the introduction of multiple copies of a transgene, an event with a number of known consequences on gene expression, such as that of repeat induced silencing and copy-number dependent expression patterns (142). In order to circumvent these difficulties many investigators have employed the use of site-specific recombinases or integrases to introduce transgenes into defined genomic locations (151, 152). Enzymatically catalyzed integration by site-specific recombinases is more efficient than random integration and allows for the introduction of a single copy of the transgene. There are several, well defined systems used for this purpose the most widely studied being that of the Cre (causes recombination) isolated from the bacteriophage P1, and the *S. cerevisiae* Flp (flippase) recombinase. Both act in a sequence-specific fashion recognizing short target sequences and share a common mechanism of DNA recombination that involves strand cleavage, exchange and ligation (153). Both the Cre and Flp systems use two 13 bp inverted repeat sequences recognized by the enzyme that are separated by an 8 bp spacer sequence, *lox*P sites for Cre, frt sites for Flp. In the presence of two target sites recombinase monomers bound to the inverted repeats promote DNA synaptic complex formation and recombination between the two sites (152). Because it is the length of the spacer and not its sequence that is critical for recombination, substitution mutations within this region will not affect the ability of the

site to act as a substrate for the enzyme (154). Sequence mutations will affect the ability of the site to efficiently recombine with a wild type site, such that sites with identical spacer sequences will recombine much more readily with one another than a wild type site (155). It is this characteristic of these systems that is the basis for recombinase mediated cassette exchange or RMCE. Use of these enzymes in mammalian cells initially involved the integration of a single *lox*P or FRT site followed by trapping of rare integration events. This approach was hindered by the fact that they are inefficient, the persistence of the enzyme often results in the re-excision of the "floxed" or "flrted" allele, which is always more efficient in this context than the integration. Secondly, the entire plasmid integrates and often leaves a positive selection marker behind. Sequences of prokaryotic origin or co-expressed genes can severely perturb the expression of neighboring genes as well as the gene of interest (156). By employing variant target sites that recombine inefficiently or not at all it is possible to introduce selected sequences site-specifically and stably into the genome. The first step in the process is to integrate a selectable marker flanked by heterotypic target sites into the genome. These recognition site-containing cassettes can be site-specifically targeted via homologous recombination to any location in the genome, allowing for the examination of chromatin environment prior to the integration of the gene of interest. Once characterized these target cassettes can act as substrates for the exchange reaction involving a circularized exchange cassette containing the gene of interest flanked by the same sites present in the genomically integrated targeting cassette. The exchange plasmid can then integrate via a two-step mechanism into the site, exchanging the original selectable marker for the desired transgenic sequences (Fig.5-1). Since only homotypic target sites and not

heterotypic ones are capable of efficient recombination, the integration event is stable. Furthermore, due to the nature of the integration subsequent targetings to the same location can occur, allowing for a comparison of the behavior of a number of transgenes from the same location.

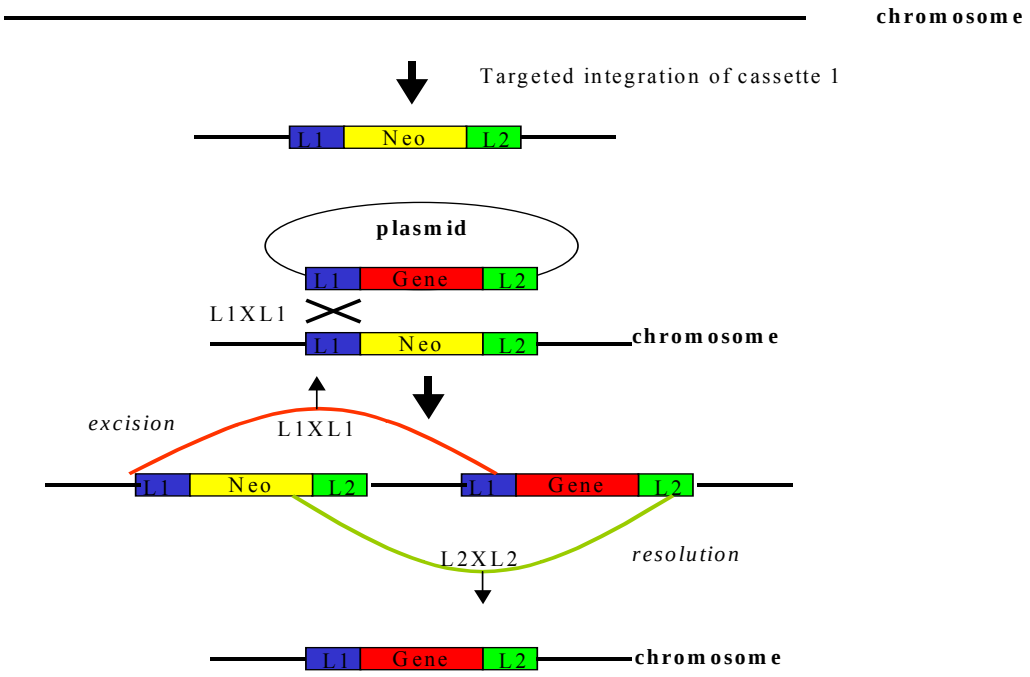


Figure 5-1. Principle of RMCE. The system is based on the single copy integration of cassette 1 containing a positive selection marker flanked by heterotypic *lox*P sites. Cotransfection of a plasmid containing cassette 2 with a Cre expression plasmid can result in two outcomes, integration followed by excision or integration followed by resolution resulting in the stable integration of cassette 2 (adapted from ref. 149).

We sought to exploit the Cre based system of RMCE to introduce human β-globin locus transgenes into the inactive X chromosome in murine embryonic stem cells through a combination of homologous recombination and RMCE, and to subsequently generate transgenic mice from these clones. The central aims of this study are to examine the role

of the LCR in resisting the repressive events that occur during the process of X-inactivation and secondly, to establish a system for reproducibly introducing large chromosomal domains, those in excess of 50 Kb, into defined locations within the mouse genome. The first step in this process is to introduce the target cassette containing a neomycin or other selectable marker via homologous recombination into the X-linked HPRT locus in murine ES cells, creating a target site for exchange cassettes. We would then use yeast artificial chromosomes (YACs) bearing an intact human β-globin locus that we have modified to contain heterotypic loxP sites as well as a puromycin selectable marker to act as exchange cassettes for RMCE. Co-transfection of our circularized YACs with a Cre expression plasmid would allow for the stable integration of the globin transgenes into the targeted locus. By employing an elegant breeding strategy involving the use of Searle's mice in which X-inactivation is not random we could analyzed the chromatin structure and expression characteristic of littermates bearing the locus on either inactive or active X chromosome (Searl's mice harbor a T(X;16)16H translocation and it is the wild type X chromosome that is preferentially in-activated) (157).

X chromosome inactivation is an extensively studied phenomenon whereby in female mammals a "random choice" mechanism decides which of the two X chromosomes will be inactivated. This inactivation results in the transcriptional silencing of most of the genes and the condensation of the chromosome into the heterochromatic Barr body. The purpose of this process is to ensure equal levels of expression of X-linked genes in XX females and XY males. In eutherian mammals this is accomplished by inactivating one of the female X chromosomes early in embryogenesis (158). Studies have revealed that a specific region of the X-chromosome, the X inactivation center (Xic)

is required in *cis* for an X chromosome to be inactivated (159). The Xic is a complex element believed to be involved not only in counting how many, but choosing which X chromosome will be inactivated (160). Furthermore, a cell must contain at least two Xics for inactivation to occur. Found within the Xic is the Xist transcript, a non-coding RNA that coats the chromosome in *cis* and triggers its silencing. A second critical element to this process is the Xce or X controlling element found within the Xic. It is believed that the choice of which chromosome to be inactivated occurs at this element. This assumption is based on the discovery that differences in Xce alleles can result in skewed patterns of X-inactivation (161). Once the choice is made the to-be inactivated chromosome undergoes a number of changes to make it distinct from that of the active one. These include the up-regulation of the *Xist* transcript located in the Xic, replication at later point in S-phase (162, 163), differential methylation of selected CpG islands (164), and characteristic changes in histone modification patterns, including hypoacetylation of histones H3 and H4, the appearance of histone modifications associated with gene silencing including di-methylation of H3 lys-9 and tri-methylation of H3 lys-27 (165). It is this unique and well defined process of gene silencing and heterochromatization that led us to choose this environment to test the ability of the human β-globin locus control region (LCR) to act as a dominant regulatory element capable of resisting the repressive nature of the inactive X chromosome. As defined previously the LCR is capable of conferring position-independent expression on *cis* linked genes. However, whether or not this is the case for all locations is still a matter of contention. Similar studies examining the ability of another LCR containing locus, the

lysozyme locus to escape X-inactivation associated silencing revealed that the elements included in the domain were unable to protect from gene silencing (166).

We believe intrinsic differences in β-globin and lysozyme locus structure and regulation may result in a different outcome for globin locus transgenes subjected to the same conditions. We wish to examine the chromatin structure and expression characteristics of the globin genes in the presence or absence of the LCR by creating two constructs in which either the entire locus ("floxed locus") or the genes alone ("floxed genes") are contained within the boundaries of heterotypic loxP sites. From this we hope to discern the activities provided by the LCR versus those found intrinsically within the gene proximal regulatory elements. If the intact locus does indeed remain active when integrated into the Xi we would also like to study active domain formation and spreading during the developmental process using an *in vitro* differentiation system. Lastly, if successful these studies will be the first example of a system capable of site specific-integration of large chromosomal domains (>100Kb) into defined positions within the mouse genome.

Results and Discussion

As stated, the most important aspect of site-specific RMCE is to introduce a target cassette into the desired location within the genome. For this purpose we have generated two DNA constructs for targeting the murine Hprt gene locus (Fig.5-2A and B). In the first construct we have embedded a neomycin resistance gene and two heterospecific *lox*P sites within Hprt homologous sequences derived from a vector Del 19.2 acquired from the laboratory of Allan Bradly and shown to be capable of targeting the locus. These sequences were cloned into the pKO916 gene targeting vector (Stratagene). The second

construct is somewhat more complex and contains; in addition to the Hprt homologous sequences extracted from Del 19.2, neomycin resistance gene, and the two heterotypic loxP sites, an intact β-globin LCR. The larger construct was generated by ligation of Hprt sequences flanking LCR homologous sequence elements derived from

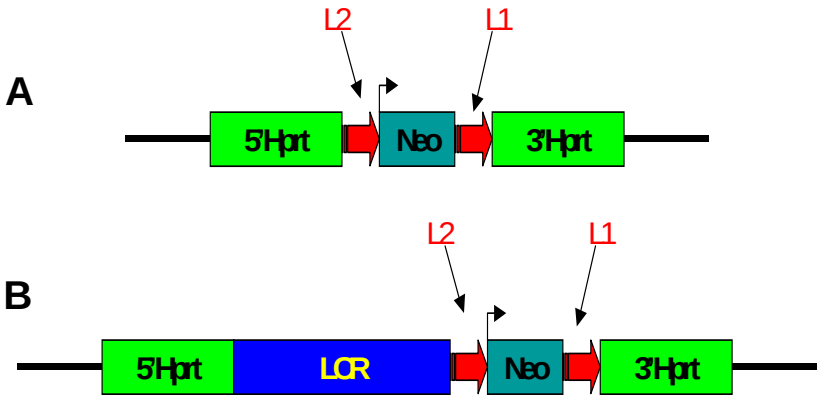


Figure 5-2. Structure of gene targeting cassettes for RMCE in ES cells. A) Small targeting construct containing only "floxed" neomycin gene and Hprt homologous sequences. B) Larger construct containing the human β-globin LCR in addition to Hprt homologous sequences and "floxed" selectable marker.

the 5' and 3' end of the LCR. The neomycin resistance gene and the two *lox*P sites were ligated 3' to the LCR homologous sequences. The yeast shuttle vector pRS316 was used as the backbone of this construct as it is capable of being replicated in both bacteria and as an episome in yeast. It also contains the ura3 gene. The strain of *S. cerevisiae* used for these experiments is a heterotrophic mutant for components of lysine, tryptophan, and uracil metabolism. The lysine and tryptophan mutations are complemented by the presence of these genes in the vector arms of the A2014.3 YAC that contains the human β-globin locus. The YAC is maintained via growth in medium lacking lysine and tryptophan. Upon linearization of the pRS316 plasmid between selected regions of globin locus homology, the plasmid is introduced via electroporation into the A201F4.3 strain.

The transformed cells are then grown in medium lacking uracil as well lysine and tryptophan (lys-,trp-,ura- media). The surviving colonies are those that have undergone the desired recombination event. This process will be discussed further in the next section. Plasmids that have incorporated the entire LCR were shuttled into bacteria, amplified, purified, and used to transform ES cells.

Human β-globin locus cassette exchange constructs were made using homologous recombination in yeast to retrofit the YAC with heterotypic loxP sites and a puromycin selectable marker. Four recombination vectors were used to create two human β-globin locus exchange cassettes (Fig.5-3).

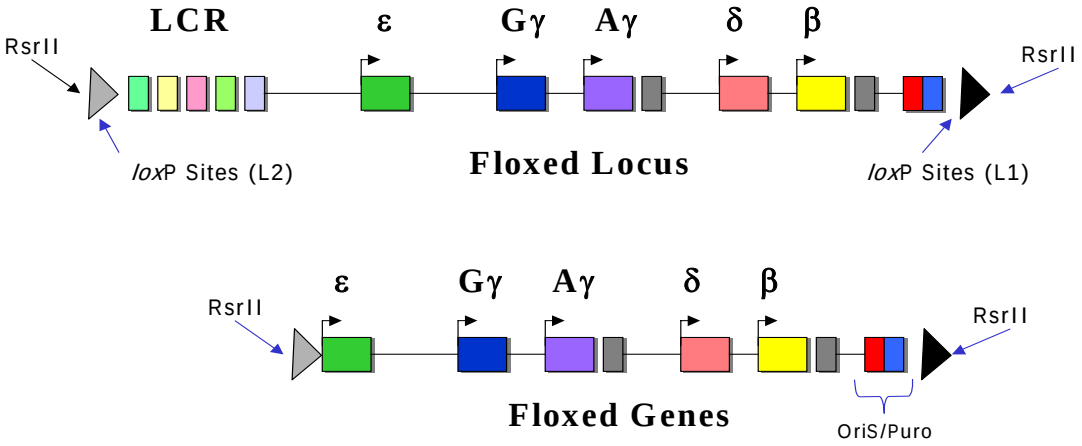


Figure 5-3. YAC constructs for RMCE. F4.3 human β-globin locus yeast artificial chromosomes were modified via homologous recombination to create exchange cassettes for the integration of domains spaning the entire locus or the genes alone. OriS/Puro, a bacterial origin of replication from the *F* factor and a puromycin selectable marker.One construct spans the entire locus

("floxed locus") and the second contains only the five β-globin genes, lacking an LCR ("floxed genes"). The vectors were created by PCR generation of homologous sequences to the regions flanking the site to which the loxP mutation was directed. LoxP sites were generated from synthetic oligos (L2 sites) or derived from the pBS246 plasmid (L1 sites). Based on earlier studies by Bouhassira et al. we generated cassettes in which the loxP

sites differed in a single G-A mutation within the spacer region and were termed L1 and L2 (100). The sequences are shown in Fig.5-4.

L1:ATAACTTCGTATAATGTAT**G**CTATACGAAGTTAT

L2: ATAACTTCGTATAATGTAT**A**CTATACGAAGTTAT

Figure 5-4 *LoxP* sequences used in targeting and exchange cassettes. G-A mutation is underlined

These sites recombine at low efficiencies, as little as 1-5% in the absence of selection (167), and under selective pressure the desired integrants can be found at levels of 100% (100). The four target vectors included 5'LCRL2, 5'GenesL2, 3'GenesL1, and 3'Genes Puro. These vectors would allow us to construct to modified YACs capable of integrating via RMCE into defined loci within the genome and exchange a neomycin selection marker for that of a puromycin marker, allowing us to enrich for populations that had undergone legitimate site-specific integration. Transformation is carried out by linearizing the vectors in the 5' or 3' homology and electroporation into yeast that are in mid log-phase growth. Recombination between the 5' homologies results in the integration of the vector as shown in Fig5-5. Similar to the prs316 shuttle vector the backbone of these constructs (pRS306) contains a ura3 gene allowing transformants to grow in lys-,trp-,ura- media. The surviving colonies are analyzed by PCR and at least two different southern blots (Fig.5-6). Following confirmation of correct targeting selected clones are grown for 2 days in non-selective media then on solid media containing 5-fluor-orotic acid (5-FOA). This step is referred to as the "pop-out" step, and results in either the excision of entire construct or only the pRS vector sequences due to the duplication of the homologous sequences upon integration. Excised sequences are lost due to the lack of an autonomous replication sequence (ARS). 5-FOA is

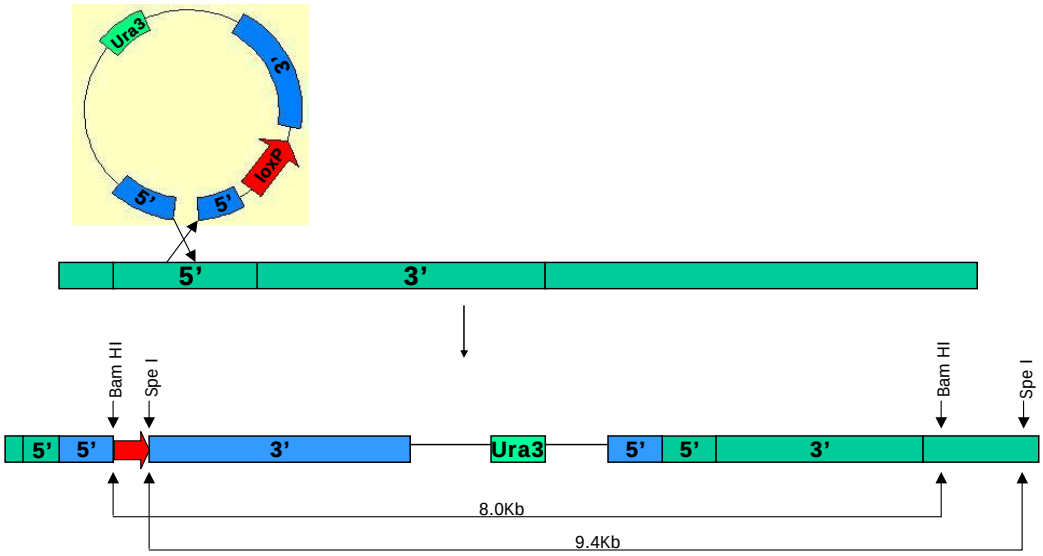


Figure 5-5. Schematic representation of the structure of integrated plasmids used to generate YAC exchange cassettes. Restriction sites and fragments generated in Southern analysis of transformants are shown. Homologies are shown as horizontal bars labeled 5' and 3'. Ura3, uracil biosynthesis gene contained in plasmid backbone and used to select integrants.

metabolized by the ura3 gene product into a toxic metabolite and selects against those cells that have retained the vector after growth in non-selective media. This selects for those clones that have lost the plasmid sequences and allows for multiple mutations to be introduced. The unique RsrII site within the vectors will allow for removal of all "un-floxed" DNA as well as circularization of the YAC for replication in the *DH10B* strain of E. coli. To date we have inserted all of the required *lox*P sites, puromycin markers and shown the RsrII sites to be functional (Fig.5-7). However, we have yet to successfully shuttle them into bacteria. We believe that this is due to our attempts to use the bacterial origin of replication found in the pRS plasmids. The large (>50 kb) constructs are most likely too large for this origin. We are currently constructing a vector to incorperate an *Ori*S from the fertility factor into the cassette. This origin allows for the replication of much larger plasmids (300 Kb).

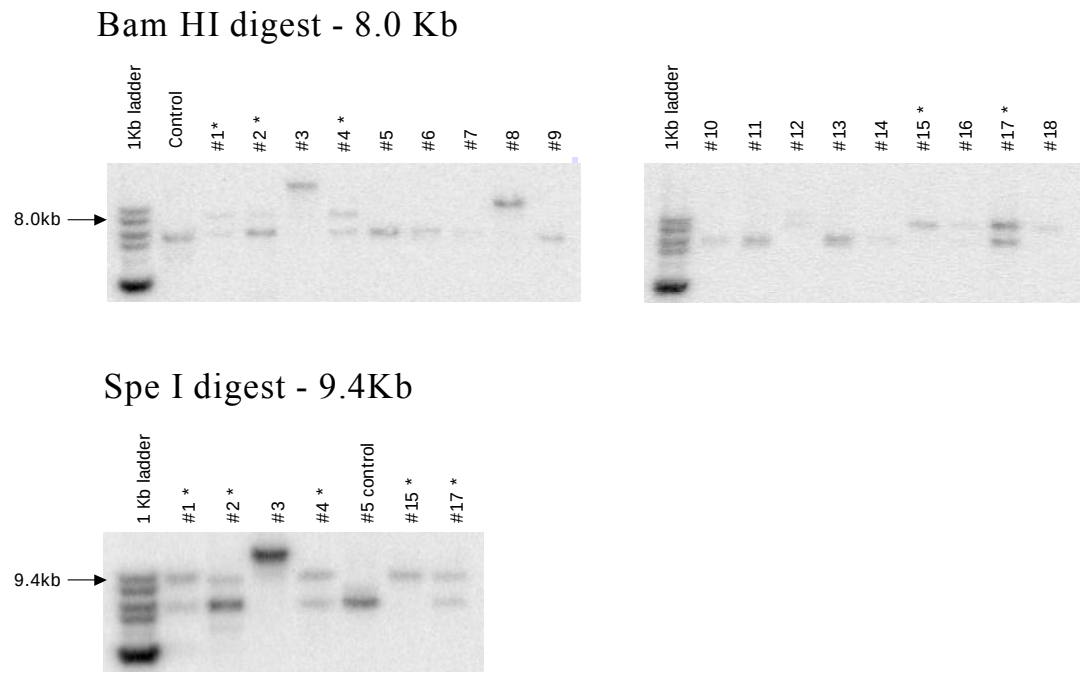


Figure 5-6. Southern analysis of yeast transformants. Clones were isolated following growth on selective media and DNA was extracted, purified, and digested. The restriction enzyme used and the expected fragment size are shown above each blot. Positive clones are shown with an asterisk.

ES cells were transformed via electroporation using 25-30μg of target construct. Hprt targeted ES cells should be neomycin resistant and insensitive to 6-thioguanine (due to disruption of the Hprt gene). We have isolated several such clones from transformations with the two constructs. Although the clones were positive for the transfected DNA, so far we were not successful in targeting the Hprt gene as indicated by southern and PCR analysis. However, as revealed by fluorescence *in situ* hybridization (FISH) several clones appear to have integrated into telomeric or centromeric regions of the genome (data not shown). We will use these clones to first establish the second step of targeting, which is the Cre-mediated site-specific recombination of globin locus constructs into the heterospecific loxp sites. At the same time we will modify our targeting constructs so as to substitute the current segments of Hprt homology, which were derived from Balbc mice with those isolated from a 129 strain line library. The

reason being that our lack of success in targeting the desired locus may be due to the fact that the homology incorporated into our target vector is non-isogenic as our ES cells are of 129 strain origin. We intend to redesign the targeting vector to include regions of homology to Hprt derived from a 129 library.

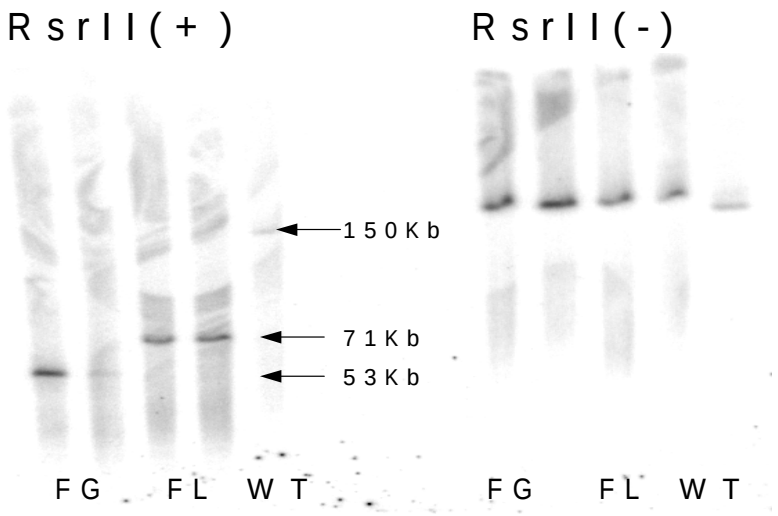


Figure 5-7. Southern Analysis of human β-globin locus exchange constructs. Isolated YAC DNA was digested with RsrII, run of a pulse field gel and analyzed by southern protocol using a probe specific for the β-globin gene. The unmodified A201F4.3 human β-globin locus YAC is approximately 150 Kb. The "floxed locus" and "floxed genes" cassettes should be approximately 71 and 52 Kb. following digestion. Digested and undigested samples are shown.

Despite preliminary difficulties in our targeting and exchange strategy we believe the subtle modifications suggested will prove successful. The use of RMCE for the site-specific integration of large chromosomal domains has several merits. As stated, understanding the chromatin environment at the site of integration is critical in understanding the behavior of any transgene. Being able to choose the context in which you can examine the impact of various regulatory elements on transcription has obvious advantages. The application of this technique to large YAC or BAC transgenes will allow for the inclusion of more distal elements involved in gene regulation without assembling them in a completely artificial construct. We propose integrating the human β-globin

locus into the inactive X-chromosome, but once the target cassette has been properly recombined into the Hprt locus any properly modified genetic construct could be studied in this context. A recent publication by Adams and colleagues has shown that cassette exchange is possible in ES cells, thus eliminating one possible variable and supporting the continuation of the project (168).

CHAPTER6
SUMMATION OF RESULTS

Functional Significance of the β-globin Locus Control Region

Since its discovery several decades ago much research and speculation concerning the role of the LCR in β-globin locus gene regulation has been put forth. Even today a complete picture of its function has not fully been described. There is however a large body of work describing LCR function in both natural and artificial contexts (143). At its endogenous position it appears that the LCR may only be required for high-level expression of the globin genes whereas; transgenic studies have shown that globin genes lacking an intact LCR or critical combinations of LCR elements were expressed at variable levels or not at all (76, 80). These observations and others like them led authors to conclude that perhaps the LCR contained two, dissociable activities; a tissue-specific enhancing activity and the ability to dominantly open chromatin structure and ectopic sites within the genome. This over-simplified description of LCR function belies the vastly complex nature of events occurring at this element during erythroid ontogeny. During the establishment of the hematopoietic system a number of complex and icompletely defined series of protein:DNA and protein:protein interaction occur and result in the exqusitely regulated series of transcriptional switches collectively known as hemoglobin gene switching. A second facet of the LCR that has gained much attention is its ability to ensure expression of *cis*-linked genes in a transgenic context. This ability has been attributed to a dominant chromatin opening activity present with elements of the LCR.

Although neither the LCR's role in gene switching and the means by which it is able to remodel the structure of chromatin at ectopic sites has been fully elucidated we wished to dissect certain aspects of these phenomena in this work. Using an *in vitro* system that allowed us to observe some of the earliest events in globin locus activation we examined gene expression and chromatin structure changes in uncommitted progenitors and definitive erythrocytes. Secondly, using transgenic mice we attempted to define the minimal genetic elements of β-globin locus LCRs from various vertebrate species required to establish independently regulated chromosomal domains and hence, position-independent patterns of gene expression. Lastly, we attempted to use components of the Adeno Associated Virus integration machinery to deliver human β-globin transgenes in a site-specific fashion.

In Vitro Differentiation of Murine Embryonic Stem Cells

The ES/OP9 *in vitro* differentiation system provided an excellent means of studying the initial events occurring at the globin locus prior to the onset of gene expression. Our goal was to test the hypothesis that during the developmental process the progressive activation of the locus would first be imitiated by chromatin structure changes and factor recruitment at the LCR. Our results have shown that histone modifications such as acetylation of histone H4 and di-methylation of lysine of histone H3 were present at the LCR preceding detectable expression of the globin genes. We also observed recruitment of transcription complexes containing RNA polymerase II to the LCR but not to the promoters of the globin genes in these cells. At later time points these marks and factors were detectable at actively transcribed genes as well as that the core regions of the LCR HS sites.

Analysis of intermediate time points such as that corresponding to the onset of embryonic hematopoiesis were inconclusive as we were unable to detect the establishment of an embryonic-specific chromatin domain as described in our model (46). Despite this, we were able to establish that the LCR appears to be the initial site of activation and recruitment of transcription-specific activities to the locus. The means by which these activities are transferred to the gene promoters is not known and a number of models have been proposed (169).

These results clearly challenge the role of the LCR as acting as a simple enhancer. We wish to expand upon data obtained in these studies in several ways. We first wished to examine other time points during the differentiation experiments. Of particular interest is day 5 because this is when the embryonic but not adult genes appear to be active. We would like to determine if the region spanning the embryonic genes exhibits a chromatin structure different from that the adult genes as suggested by Gribnau *et al*.(90). The inclusion of additional antibodies, such as TAF 10 would also be if interest. The possibility that transcription complexes recruited to the LCR prior to gene expression posses different subunit compositions could suggest a role for Pol II in remodeling of the LCR and holocomplex formation. Lastly, a FACS analysis to determine the fraction of cells in our assays that are actually hematopoietic and/or erythroid could provide evidence that the LCR is active in a number of cell types.

Site-Specific Integration and Position-Independent Expression of Human β-globin Transgenes

Many attempts have been made to ascertain the exact role of the various regulatory elements found within both gene proximal elements and that of the more distal locus control region of the β-globin locus. Many of these studies have been performed using

transgenic models, most notably that of the mouse. Confounding the results from these studies has been the influence of chromatin structure at the site of integration of many of these transgenes. These "position effects" can cause variable levels of gene expression between independent lines bearing the same transgenic construct. One method of avoiding this problem has been to include dominant regulatory sequences or boundary elements that can resist the influences of neighboring chromatin. A second strategy used to avoid position effects has been to direct transgenes to specific locations within the genome and a number of systems have been used for this purpose (100, 129, 130). We sought to create a human β-globin containing transgene containing the minimal genetic elements from vertebrate globin gene loci that would ensure position-independent expression at therapeutically relevant levels as well as direct our gene, using the AAV virus, to a specific location within the mouse genome using a mouse transgenic human AAVS1 site. This construct proved to exhibit a position independent expression pattern as human β-globin gene transcripts were detectable in four, independent lines. However in all cases expression levels were low indicating that additional regulatory elements were required. We therefore included the sequences flanking the core regions of the HS2 and 3 normally found in the context if an intact locus. Inclusion of flanking sequence had previously been shown to enhance expression of linked genes (121, 144). This rendered the construct too large for packaging of our construct into AAV. We then modified our strategy and included elements from the AAV genome critical for integration (136). The larger construct was incubated with various molar concentrations of the Rep 68 protein and injected into fertilized S1 transgenic oocytes. Whereas, expression levels were higher we failed to site-specifically integrate the construct into our target site. We can only

speculate that either our DNA:protein ratios need adjustment our that some aspect of the viral cycle required for integration of the genome was bypassed by non-viral delivery of the DNA. Despite this, one of the lines generated was of particular interest. There were two reasons for this. The first being that is expressed at levels 40% of that of a single-copy, intact wild type β-globin YAC. Secondly, upon FISH analysis this transgene appeared to have integrated within or near the centromere of the chromosome. It is known that certain centromeres are transcriptionally permissible, but in general these regions are highly heterochromatic and transcriptionally repressive (148). Hypersensitive sites were also detectable in the β promoter and in HS2. Currently we are focused and identifying the exact site of integration of this transgene and will characterize the chromatin structure within the transgene as well as the regions flanking the site of integration. We also intend to continue to optimize the conditions required for Rep-mediated, site-specific integration of globin constructs.

The results of these studies support the hypothesis that LCR acts as the primary site of recruitment for transcription and co-activator complexes during development. Furthermore, we have shown that synergistic interactions between HS sites of the LCR are critical to high-level expression of the globin genes.

LIST OF REFERENCES

1. Pauling, L., Itano, H. A. (1949) *Science* **110,** 543-8.

2. Weatherall, D. J. (2004) *Nature Genetics* **5,** 625-31.

3. Stamatoyannopoulos, G. N., A.W. Mjerus, P.W. Varmus, H. (1994) *The Molecular Basis of Blood Diseases* (W. B. Saunders, Philadelphia).

4. Olivieri, N. F., Brittenham, G. M., McLaren, C. E., Templeton, D. M., Cameron, R. G., McClelland, R. A., Burt, A. D. & Fleming, K. A. (1998) *N Engl J Med* **339,** 417-23.

5. Blau, C. A., Constantoulakis, P., al-Khatti, A., Spadaccino, E., Goldwasser, E., Papayannopoulou, T. & Stamatoyannopoulos, G. (1993) *Blood* **81,** 227-33.

6. Ellis, J. & Pannell, D. (2001) *Clin. Genet.* **59,** 17-24.

7. Littau, V. C., Burdick, C. J., Allfrey, V. G. & Mirsky, S. A. (1965) *Proc. Natl. Acad. Sci. U S A* **54,** 1204-12.

8. Hayes, J. J., Tullius, T. D. & Wolffe, A. P. (1990) *Proc. Natl. Acad. Sci. U S A* **87,** 7405-9.

9. Strahl, B. D. & Allis, C. D. (2000) *Nature* **403,** 41-5.

10. Jenuwein, T. & Allis, C. D. (2001) *Science* **293,** 1074-80.

11. Turner, B. M. (2000) *Bioessays* **22,** 836-45.

12. Roth, S. Y., Denu, J. M. & Allis, C. D. (2001) *Annu. Rev. Biochem.* **70,** 81-120.

13. Hsu, J. Y., Sun, Z. W., Li, X., Reuben, M., Tatchell, K., Bishop, D. K., Grushcow, J. M., Brame, C. J., Caldwell, J. A., Hunt, D. F., Lin, R., Smith, M. M. & Allis, C. D. (2000) *Cell* **102,** 279-91.

14. Sterner, D. E. & Berger, S. L. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 435-59.

15. Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K. & Zhou, M. M. (1999) *Nature* **399,** 491-6.

16. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. (2000) *Science* **288,** 1422-5.

17. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A. & Casero, R. A. (2004) *Cell* **119,** 941-53.

18. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. & Jenuwein, T. (2000) *Nature* **406,** 593-9.

19. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. (1998) *Cell. Mol. Life Sci.* **54,** 80-93.

20. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. (2001) *Nature* **410,** 116-20.

21. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. & Kouzarides, T. (2001) *Nature* **410,** 120-4.

22. Roach, M. L. & McNeish, J. D. (2002) *Methods Mol. Biol.* **185,** 1-16.

23. Loebel, D. A., Watson, C. M., De Young, R. A. & Tam, P. P. (2003) *Dev. Biol.* **264,** 1-14.

24. Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A. & Tam, P. P. (1999) *Development* **126,** 4691-701.

25. Keller, G., Lacaud, G. & Robertson, S. (1999) *Exp. Hematol.* **27,** 777-87.

26. Silver, L. & Palis, J. (1997) *Blood* **89,** 1154-64.

27. Wong, P. M., Chung, S. W., Chui, D. H. & Eaves, C. J. (1986) *Proc. Natl. Acad. Sci. U S A* **83,** 3851-4.

28. Palis, J. & Segel, G. B. (1998) *Blood Rev.* **12,** 106-14.

29. Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. (1996) *Cell* **84,** 321-30.

30. Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H. & Speck, N. A. (1996) *Proc. Natl. Acad. Sci. U S A* **93,** 3444-9.

31. Cline, M. J. & Moore, M. A. (1972) *Blood* **39,** 842-9.

32. Medvinsky, A. & Dzierzak, E. (1996) *Cell* **86,** 897-906.

33. Muller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F. & Dzierzak, E. (1994) *Immunity* **1,** 291-301.

34. Godin, I. & Cumano, A. (2002) *Nat. Rev. Immunol.* **2,** 593-604.

35. Cumano, A., Dieterlen-Lievre, F. & Godin, I. (1996) *Cell* **86,** 907-16.

36. Yoder, M. C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D. M. & Orlic, D. (1997) *Immunity* **7,** 335-44.

37. Keller, G. M. (1995) *Curr. Opin. Cell Biol.* **7,** 862-9.

38. Robertson, S., Kennedy, M. & Keller, G. (1999) *Ann. N. Y. Acad. Sci.* **872,** 9-15; discussion 15-6.

39. Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M. V. (1993) *Mol. Cell. Biol.* **13,** 473-86.

40. Yoshida, H., Hayashi, S., Kunisada, T., Ogawa, M., Nishikawa, S., Okamura, H., Sudo, T. & Shultz, L. D. (1990) *Nature* **345,** 442-4.

41. Kodama, H., Nose, M., Niida, S. & Nishikawa, S. (1994) *Exp. Hematol.* **22,** 979-84.

42. Kitajima, K., Tanaka, M., Zheng, J., Sakai-Ogawa, E. & Nakano, T. (2003) *Methods Enzymol.* **365,** 72-83.

43. Kyba, M., Perlingeiro, R. C. & Daley, G. Q. (2002) *Cell* **109,** 29-37.

44. Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. & de Laat, W. (2003) *Nat. Genet.* **35,** 190-4.

45. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. (2002) *Mol. Cell* **10,** 1453-65.

46. Levings, P. P. & Bungert, J. (2002) *Eur. J. Biochem.* **269,** 1589-99.

47. Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T. & Groudine, M. (1990) *Genes. Dev.* **4,** 1637-49.

48. de Laat, W. & Grosveld, F. (2003) *Chromosome Res.* **11,** 447-59.

49. Hardison, R., Chao, K. M., Schwartz, S., Stojanovic, N., Ganetsky, M. & Miller, W. (1994) *Genomics* **21,** 344-53.

50. Bungert, J. & Engel, J. D. (1996) *Ann. Med.* **28,** 47-55.

51. Orkin, S. H. (1995) *J. Biol. Chem.* **270,** 4955-8.

52. Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H. & Costantini, F. (1991) *Nature* **349,** 257-60.

53. Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W. & Orkin, S. H. (1994) *Nature* **371,** 221-6.

54. Zon, L. I., Mather, C., Burgess, S., Bolce, M. E., Harland, R. M. & Orkin, S. H. (1991) *Proc. Natl. Acad. Sci. U S A* **88,** 10642-6.

55. Weiss, M. J., Keller, G. & Orkin, S. H. (1994) *Genes. Dev.* **8,** 1184-97.

56. Briegel, K., Lim, K. C., Plank, C., Beug, H., Engel, J. D. & Zenke, M. (1993) *Genes. Dev.* **7,** 1097-109.

57. Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G. & Higgs, D. R. (2004) *Embo. J.* **23,** 2841-52.

58. Merika, M. & Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 2437-47.

59. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M. & Orkin, S. H. (1998) *Proc. Natl. Acad. Sci. U S A* **95,** 2061-6.

60. Stamatoyannopoulos, J. A., Goodwin, A., Joyce, T. & Lowrey, C. H. (1995) *Embo. J.* **14,** 106-16.

61. Kotkow, K. J. & Orkin, S. H. (1996) *Proc. Natl. Acad. Sci. U S A* **93,** 3514-8.

62. Johnson, K. D., Christensen, H. M., Zhao, B. & Bresnick, E. H. (2001) *Mol. Cell.* **8,** 465-71.

63. Engel, J. D. (1994) *Nature* **367,** 516-7.

64. Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. & Yamamoto, M. (1994) *Nature* **367,** 568-72.

65. Shivdasani, R. A., Rosenblatt, M. F., Zucker-Franklin, D., Jackson, C. W., Hunt, P., Saris, C. J. & Orkin, S. H. (1995) *Cell* **81,** 695-704.

66. Crossley, M., Tsang, A. P., Bieker, J. J. & Orkin, S. H. (1994) *J. Biol. Chem.* **269,** 15440-4.

67. Perkins, A. C., Sharpe, A. H. & Orkin, S. H. (1995) *Nature* **375,** 318-22.

68. Wijgerde, M., Grosveld, F. & Fraser, P. (1995) *Nature* **377,** 209-13.

69. Donze, D., Townes, T. M. & Bieker, J. J. (1995) *J. Biol. Chem.* **270,** 1955-9.

70. Mouthon, M. A., Bernard, O., Mitjavila, M. T., Romeo, P. H., Vainchenker, W. & Mathieu-Mahul, D. (1993) *Blood* **81,** 647-55.

71. Bresnick, E. H. & Felsenfeld, G. (1993) *J. Biol. Chem.* **268,** 18824-34.

72. Aplan, P. D., Nakahara, K., Orkin, S. H. & Kirsch, I. R. (1992) *Embo. J.* **11,** 4073-81.

73. Lee, J. S., Lee, C. H. & Chung, J. H. (1999) *Proc. Natl. Acad. Sci. U S A* **96,** 10051-5.

74. Hardison, R. (1998) *J. Exp. Biol.* **201,** 1099-117.

75. Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. (1987) *Nucleic Acids Res.* **15,** 10159-77.

76. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. (1987) *Cell* **51,** 975-85.

77. Bungert, J., Dave, U., Lim, K. C., Lieuw, K. H., Shavit, J. A., Liu, Q. & Engel, J. D. (1995) *Genes Dev.* **9,** 3083-96.

78. Bungert, J., Tanimoto, K., Patel, S., Liu, Q., Fear, M. & Engel, J. D. (1999) *Mol. Cell Biol.* **19,** 3062-72.

79. Peterson, K. R., Clegg, C. H., Navas, P. A., Norton, E. J., Kimbrough, T. G. & Stamatoyannopoulos, G. (1996) *Proc. Natl. Acad. Sci. U S A* **93,** 6605-9.

80. Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. & Groudine, M. (1998) *Mol. Cell. Biol.* **18,** 5992-6000.

81. Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M. A., Fiering, S., Enver, T., Martin, D. I., Kennedy, M., Keller, G. & Groudine, M. (1998) *Mol. Cell.* **2,** 447-55.

82. Bender, M. A., Bulger, M., Close, J. & Groudine, M. (2000) *Mol. Cell.* **5,** 387-93.

83. Farrell, C. M., Grinberg, A., Huang, S. P., Chen, D., Pichel, J. G., Westphal, H. & Felsenfeld, G. (2000) *Proc. Natl. Acad. Sci. U S A* **97,** 14554-9.

84. Bulger, M. & Groudine, M. (1999) *Genes Dev.* **13,** 2465-77.

85. Jimenez, G., Griffiths, S. D., Ford, A. M., Greaves, M. F. & Enver, T. (1992) *Proc. Natl. Acad. Sci. U S A* **89,** 10618-22.

86. Orkin, S. H. (1995) *Eur. J. Biochem.* **231,** 271-81.

87. Blobel, G. A. (2000) *Blood* **95,** 745-55.

88. Schubeler, D., Francastel, C., Cimbora, D. M., Reik, A., Martin, D. I. & Groudine, M. (2000) *Genes Dev.* **14,** 940-50.

89. Vieira, K. F., Levings, P. P., Hill, M. A., Crusselle, V. J., Kang, S. H., Engel, J. D. & Bungert, J. (2004) *J. Biol. Chem.* **279,** 50350-7.

90. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. (2000) *Mol. Cell.* **5,** 377-86.

91. Forsberg, E. C., Downs, K. M., Christensen, H. M., Im, H., Nuzzi, P. A. & Bresnick, E. H. (2000) *Proc. Natl. Acad. Sci. U S A* **97,** 14494-9.

92. Leach, K. M., Nightingale, K., Igarashi, K., Levings, P. P., Engel, J. D., Becker, P. B. & Bungert, J. (2001) *Mol. Cell Biol.* **21,** 2629-40.

93. Elefanty, A. G., Robb, L., Birner, R. & Begley, C. G. (1997) *Blood* **90,** 1435-47.

94. Leach, K. M., Vieira, K. F., Kang, S. H., Aslanian, A., Teichmann, M., Roeder, R. G. & Bungert, J. (2003) *Nucleic Acids Res.* **31,** 1292-301.

95. Gaensler, K. M., Burmeister, M., Brownstein, B. H., Taillon-Miller, P. & Myers, R. M. (1991) *Genomics* **10,** 976-84.

96. Chung, J. H., Bell, A. C. & Felsenfeld, G. (1997) *Proc. Natl. Acad. Sci. U S A* **94,** 575-80.

97. Hartl, D. L. & Ochman, H. (1996) *Methods Mol. Biol.* **58,** 293-301.

98. Trask, B. J. (1991) *Curr. Opin. Genet. Dev.* **1,** 82-7.

99. Kang, S. H., Vieira, K. & Bungert, J. (2002) *Nucleic Acids. Res.* **30,** e44.

100. Bouhassira, E. E., Westerman, K. & Leboulch, P. (1997) *Blood* **90,** 3332-44.

101. Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. & Felsenfeld, G. (2001) *Science* **293,** 2453-5.

102. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. (1994) *Embo. J.* **13,** 1823-30.

103. Bottardi, S., Aumont, A., Grosveld, F. & Milot, E. (2003) *Blood* **102,** 3989-97.

104. Delassus, S., Titley, I. & Enver, T. (1999) *Blood* **94,** 1495-503.

105. Johnson, K. D., Grass, J. A., Park, C., Im, H., Choi, K. & Bresnick, E. H. (2003) *Mol. Cell Biol.* **23,** 6484-93.

106. Kontaraki, J., Chen, H. H., Riggs, A. & Bonifer, C. (2000) *Genes Dev.* **14,** 2106-22.

107. Tagoh, H., Himes, R., Clarke, D., Leenen, P. J., Riggs, A. D., Hume, D. & Bonifer, C. (2002) *Genes. Dev.* **16,** 1721-37.

108. Ford, A. M., Bennett, C. A., Healy, L. E., Towatari, M., Greaves, M. F. & Enver, T. (1996) *Proc. Natl. Acad. Sci. U S A* **93,** 10838-43.

109. Lund, A. H. & van Lohuizen, M. (2004) *Genes Dev.* **18,** 2315-35.

110. Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. (1999) *Proc. Natl. Acad. Sci. U S A* **96,** 14967-72.

111. Suwabe, N., Takahashi, S., Nakano, T. & Yamamoto, M. (1998) *Blood* **92,** 4108-18.

112. Szutorisz, H., Canzonetta, C., Georgiou, A., Chow, C. M., Tora, L. & Dillon, N. (2005) *Mol. Cell. Biol.* **25,** 1804-20.

113. Kellum, R. & Schedl, P. (1991) *Cell* **64,** 941-50.

114. Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. (1989) *Proc. Natl. Acad. Sci. U S A* **86,** 5439-43.

115. Bonifer, C. (2000) *Trends Genet.* **16,** 310-5.

116. Tuan, D., Solomon, W., Li, Q. & London, I. M. (1985) *Proc. Natl. Acad. Sci. U S A* **82,** 6384-8.

117. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. (2002) *Blood* **100,** 3077-86.

118. Li, G., Lim, K. C., Engel, J. D. & Bungert, J. (1998) *Genes Cells* **3,** 415-29.

119. Milot, E., Fraser, P. & Grosveld, F. (1996) *Trends Genet.* **12,** 123-6.

120. Molete, J. M., Petrykowska, H., Bouhassira, E. E., Feng, Y. Q., Miller, W. & Hardison, R. C. (2001) *Mol. Cell. Biol.* **21,** 2969-80.

121. May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K. M., Luzzatto, L. & Sadelain, M. (2000) *Nature* **406,** 82-6.

122. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. & Fraser, P. (2002) *Nat. Genet.* **32,** 623-6.

123. Stief, A., Winter, D. M., Stratling, W. H. & Sippel, A. E. (1989) *Nature* **341,** 343-5.

124. West, A. G., Gaszner, M. & Felsenfeld, G. (2002) *Genes. Dev.* **16,** 271-88.

125. Bell, A. C., West, A. G. & Felsenfeld, G. (1999) *Cell* **98,** 387-96.

126. West, A. G., Huang, S., Gaszner, M., Litt, M. D. & Felsenfeld, G. (2004) *Mol. Cell* **16,** 453-63.

127. Ishii, K., Arib, G., Lin, C., Van Houwe, G. & Laemmli, U. K. (2002) *Cell* **109,** 551-62.

128. Yusufzai, T. M. & Felsenfeld, G. (2004) *Proc. Natl. Acad. Sci. U S A* **101,** 8620-4.

129. Porteus, M. H., Cathomen, T., Weitzman, M. D. & Baltimore, D. (2003) *Mol. Cell. Biol.* **23,** 3558-65.

130. Belteki, G., Gertsenstein, M., Ow, D. W. & Nagy, A. (2003) *Nat. Biotechnol.* **21,** 321-4.

131. Kotin, R. M., Linden, R. M. & Berns, K. I. (1992) *Embo. J.* **11,** 5071-8.

132. Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. (1996) *Proc. Natl. Acad. Sci. U S A* **93,** 11288-94.

133. Ryan, J. H., Zolotukhin, S. & Muzyczka, N. (1996) *J. Virol.* **70,** 1542-53.

134. Weitzman, M. D., Kyostio, S. R., Kotin, R. M. & Owens, R. A. (1994) *Proc. Natl. Acad. Sci. U S A* **91,** 5808-12.

135. Zhou, X., Zolotukhin, I., Im, D. S. & Muzyczka, N. (1999) *J. Viro.l* **73,** 1580-90.

136. Philpott, N. J., Gomos, J., Berns, K. I. & Falck-Pedersen, E. (2002) *Proc. Natl. Acad. Sci. U S A* **99,** 12381-5.

137. Ellis, J., Talbot, D., Dillon, N. & Grosveld, F. (1993) *Embo. J*. **12,** 127-34.

138. Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S. & Grosveld, F. (1996) *Embo. J.* **15,** 562-8.

139. Navas, P. A., Peterson, K. R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S. E., Clegg, C. H., Asano, H. & Stamatoyannopoulos, G. (1998) *Mol. Cell. Biol.* **18,** 4188-96.

140. Liu, Q., Bungert, J. & Engel, J. D. (1997) *Proc. Natl. Acad. Sci. U S A* **94,** 169-74.

141. Tanimoto, K., Liu, Q., Bungert, J. & Engel, J. D. (1999) *Nature* **398,** 344-8.

142. Garrick, D., Fiering, S., Martin, D. I. & Whitelaw, E. (1998) *Nat. Genet.* **18,** 56-9.

143. Higgs, D. R., Thein, S.L. & Wood, W.G. (2001) *The Thalassaemia Syndromes,* 133-191. (Blackwell Science, Oxford)

144. Pawliuk, R., Westerman, K. A., Fabry, M. E., Payen, E., Tighe, R., Bouhassira, E. E., Acharya, S. A., Ellis, J., London, I. M., Eaves, C. J., Humphries, R. K., Beuzard, Y., Nagel, R. L. & Leboulch, P. (2001) *Science* **294,** 2368-71.

145. Marshall, E. (2002) *Science* **298,** 510-1.

146. Flotte, T. R. (2004) *Gene. Ther.* **11,** 805-10.

147. Nakai, H., Montini, E., Fuess, S., Storm, T. A., Grompe, M. & Kay, M. A. (2003) *Nat. Genet.* **34,** 297-302.

148. Saffery, R., Sumer, H., Hassan, S., Wong, L. H., Craig, J. M., Todokoro, K., Anderson, M., Stafford, A. & Choo, K. H. (2003) *Mol. Cell.* **12,** 509-16.

149. Feng, Y. Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. (2001) *Mol. Cell. Biol.* **21,** 298-309.

150. Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E. & Nikolaev, L. (2000) *Biol. Chem.* **381,** 801-13.

151. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. (2000) *Genesis* **28,** 106-10.

152. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992) *Trends Genet.* **8,** 432-9.

153. Sadowski, P. D. (1995) *Prog. Nucleic. Acid. Res. Mol. Biol.* **51,** 53-91.

154. Hoess, R. H., Wierzbicki, A. & Abremski, K. (1986) *Nucleic Acids Res.* **14,** 2287-300.

155. Baer, A. & Bode, J. (2001) *Curr. Opin. Biotechnol.* **12,** 473-80.

156. Branda, C. S. & Dymecki, S. M. (2004) *Dev. Cell* **6,** 7-28.

157. Lyon, M. F., Searle, A. G., Ford, C. E. & Ohno, S. (1964) *Cytogenetics* **15,** 306-23.

158. Lyon, M. F. (1961) *Nature* **190,** 372-3.

159. Rastan, S. & Brown, S. D. (1990) *Genet. Res.* **56,** 99-106.

160. Heard, E. (2004) *Curr. Opin. Cell Biol.* **16,** 247-55.

161. Cattanach, B. M. & Papworth, D. (1981) *Genet. Res.* **38,** 57-70.

162. Tada, T. & Takagi, N. (1992) *Mol. Reprod. Dev.* **31,** 20-7.

163. Keohane, A. M., O'Neill L, P., Belyaev, N. D., Lavender, J. S. & Turner, B. M. (1996) *Dev. Biol.* **180,** 618-30.

164. Miller, O. J., Schnedl, W., Allen, J. & Erlanger, B. F. (1974) *Nature* **251,** 636-7.

165. Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P. & Brockdorff, N. (2003) *Dev. Cell* **4,** 481-95.

166. Chong, S., Kontaraki, J., Bonifer, C. & Riggs, A. D. (2002) *Mol. Cell. Biol.* **22,** 4667-76.

167. Feng, Y. Q., Seibler, J., Alami, R., Eisen, A., Westerman, K. A., Leboulch, P., Fiering, S. & Bouhassira, E. E. (1999) *J. Mol. Biol.* **292,** 779-85.

168. Adams, L. D., Choi, L., Xian, H. Q., Yang, A., Sauer, B., Wei, L. & Gottlieb, D. I. (2003) *Brain. Res. Mol. Brain Res.* **110,** 220-33.

169. Engel, J. D. & Tanimoto, K. (2000) *Cell* **100,** 499-502.

BIOGRAPHICAL SKETCH

Padraic Levings is one of nine children and was born on Long Island to Thomas and Miriam Levings. He attended the University at Stony Brook where he played three seasons of division IA lacrosse. He was also a member of the ΣB and Gold Key honor societies. He graduated cum laude in 1998 with a bachelor's degree in biochemistry.