AN INVESTIGATION INTO THE ROLE OF USF IN BETA-GLOBIN REGULATION

By

SHERMI YEN LIANG

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

2010

This work is dedicated to my grandfather Dr. I-Teh Wang, who passed away during my second year of graduate studies. As a child, I once promised him that I would become a doctor. I hope he would be proud of me, even if it’s “not that kind of doctor.”

ACKNOWLEDGMENTS

First and foremost, I would like to thank my mentor, Dr. Jörg Bungert, for giving me the opportunity to perform my doctoral studies in his laboratory. I owe him a great deal for all his help and guidance through these past few years. His dedication and endless optimism has been an inspiration for me.

Thanks go out to my committee members Drs. Thomas P. Yang, Brian Harfe, and Jianrong Lu for giving time out of their busy schedules to serve on my committee and to assist in overseeing my graduate studies. Not only have they provided insightful feedback to the thought processes behind the design and implementation of my experiments, but they also took an interest in my future career plans and have done much to help me on my way.

I am grateful to the many past and current members of the Bungert lab that have helped train me during my years there. Special thanks go out to Dr. Boris Thurish, Dr. Zhuo Zhou, Dr. Valerie Crusselle-Davis, Dr. Babak Moghimi, Michael Rosenberg, Joeva Barrow, and I-Ju Lin. I know that in the future, wherever I may go, I will miss Zhuo and Joeva very much. I highly enjoyed Michael's ability to perform comedic antics in the lab but focusing on experiments when necessary.

Neal Benson and Steve McClellan of the UF ICBR Flow Cytometry core helped me greatly by accommodating my flow cytometry experiments into their busy schedules. Dr. Dan Tuttle of the UF Animal Care Services provided his expertise and assistance in generating transient transgenic mice. Dr. Cortney Bouldin, formerly of the Harfe lab, helped me take pictures of mouse embryos.

I am and will remain eternally grateful to all my friends at the University of Florida that I leaned on for support and for whom I supported in return: Amanda W, Ada, Nicole

P, Aleixo, Arne, and Ben S. I honestly hope I haven't forgotten anyone; if so, I would like to express an apology for not including some names, as it was not my intention to deliberately neglect anyone. I am honored to have been able to make beautiful music—despite my meager violin skills—with all the members involved in the IDP Chamber Ensemble over the years and am especially grateful to Ellen for starting it.

I owe a great deal to my aunt and uncle, Jau and Richard Yoh, as well as my cousin Eric, for helping me get settled in when I first moved to Gainesville. I am glad to have been in this city to witness the transformation of my Uncle Richard's entire business from an auto shop to a successful restaurant, and I hope it will grow and continue to flourish for years to come. I wish them the best of luck in their endeavors.

I would like to thank my wonderful fiancé Justin Bickford for everything. No words could ever be enough to express how grateful I am to him for all he has done and all he continues to do. During these past few years, he has stayed by my side and supported me and witnessed all my ups and downs. He has seen me at my worst and at my best, and helped me through every step of the way. I look forward to starting our new life together.

Finally, I thank my parents, Dr. Alan Yuh-Lin and Yen Liang. Quite literally, none of this would ever have been possible without them. I would like to thank them for being the best role models in my life. They set the bar high, and I hope that like they have done, I will always manage to rise up to and overcome any challenge that comes my way.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µChIP	Small scale (“micro”) Chromatin Immunoprecipitation
3C	Chromatin Conformation Capture
ACH	Active Chromatin Hub
A-USF	Dominant-negative USF
bHLH-LZ	basic Helix-Loop-Helix Leucine Zipper
BRG1	Brahma-Related Gene 1
CBP	CREBBP; CREB Binding Protein
ChIP	Chromatin Immunoprecipitation
Co-IP	Co-Immunoprecipitation
CTD	Carboxy -Terminal Domain of RNA Polymerase II
DMSO	Dimethylsulfoxide
Dpc	Days *post coitum*
EKLF	Erythroid Krüppel-Like Factor
FACS	Fluorescence-Activated Cell Sorting
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
HLH	Helix-Loop-Helix
HS	Hypersensitive Site
HSC	Hematopoietic Stem Cell
IgG	Immunoglobulin G
K562 cells	Human Erythroleukemia cells
LCR	Locus Control Region
LZ	Leucine Zipper
MARE	Maf Recognition Element

MEL cells	Murine Erythroleukemia cells
p300	E1A binding protein p300; EP300
PIC	Transcription Pre-Initiation Complex
q-PCR	Quantitative (Real-Time) Polymerase Chain Reaction
qRT-PCR	Quantitative (Real-Time) Reverse Transcription Polymerase Chain Reaction
RBC	Red Blood Cell (Erythrocyte)
RNA Pol II	RNA Polymerase II
RT-PCR	Reverse Transcription Polymerase Chain Reaction
TG	Transgenic
TSS	Transcription Start Site
USF	Upstream Stimulatory Factor
WT	Wild-Type
YAC	Yeast Artificial Chromosome
YS	Yolk Sac
β-YAC	Human β-globin Gene Locus Yeast Artificial Chromosome

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

AN INVESTIGATION INTO THE ROLE OF USF IN BETA-GLOBIN GENE REGULATION

By

Shermi Yen Liang

July 2010

Chair: Jörg Bungert
Major: Medical Sciences--Genetics

The human β-globin gene locus contains five genes organized linearly in the order of their tissue and stage-specific expression (5'-ε-^{G}γ-^{A}γ-δ-β-3'). A region of DNA known as the β-globin locus control region (LCR) resides upstream of the genes, and is required for the high-level expression of these β-type globin genes in erythroid cells throughout development. The LCR is composed of five 200 to 400 bp DNase I hypersensitive sites (HS1 to HS5) which individually exhibit strong enhancer activity. Many transcription factors are known to bind at the LCR as well as the adult β-globin gene promoter, including the ubiquitously expressed transcription factor Upstream Stimulatory Factor (USF), which binds tightly to E-box elements and aids in the assembly of transcriptional complexes.

Previous studies showed that transfection of dominant-negative USF (A-USF) into murine erythroleukemia (MEL) cells diminished the adult β_{maj}-globin expression, while overexpression of USF1 increased β_{maj}-globin expression. I herein demonstrate that erythroid cell-specific expression of A-USF in transgenic mice reduces expression of all β-type globin genes, as well as the association of RNA polymerase II with HS2 and the β-globin gene promoter. Expression of several key erythroid cell-specific transcription

factors was also reduced. Furthermore, A-USF-expressing transgenic mice exhibited defective erythropoiesis.

USF has also been implicated in chromatin remodeling, and the phenotype of mice carrying a point mutation of the protein Brahma-Related Gene 1 (BRG1) is similar to that observed in A-USF mice. BRG1 is a subunit of the SWI/SNF chromatin remodeling complex, and is recruited to the locus to promote erythropoiesis through interaction with the erythroid-specific transcription factor Erythroid Krüppel-Like Factor (EKLF). We found that USF and BRG1 reside in a complex, although the exact properties of the interaction remain uncharacterized.

In summary, these data demonstrate that USF regulates globin expression indirectly by enhancing erythroid transcription factor expression, and directly by mediating the recruitment of transcriptional complexes to the globin gene locus. The ubiquitously expressed USF may have erythroid-specific effects when associated with erythroid-specific factors, enabling the optimal utilization of resources to promote erythropoiesis.

CHAPTER 1
INTRODUCTION

Hematopoiesis and Erythropoiesis

The process of hematopoiesis begins from hematopoietic stem cells (HSC), which give rise to all blood cell types of both myeloid and lymphoid lineages (1). Cell types in the myeloid lineage include granulocytes and erythrocytes (or red blood cells, RBC). Cell types in the lymphoid lineage include lymphocytes, a type of non-phagocytic white blood cell involved in the vertebrate immune system. Lymphocytes can be found in blood, lymph, and lymphatic tissues, while granulocytes and erythrocytes are restricted to blood. In a typical circulatory system, granulocytes are larger in size but less abundant than erythrocytes. They are also able to move about under their own power by amoeboid movement, unlike erythrocytes, which rely on the pumping of the heart and blood flow for mobility (2).

Because mature blood cells have undergone terminal differentiation, they have a finite life span and are generally incapable of mitosis. Thus, the body must rely on the constant self-renewal and pluripotency of HSCs in order to generate new mature blood cells. Environmental influences and various growth factors, such as erythropoietin, thrombopoietin, and interleukin, signal the differentiating cells to assume the desired lineage (1). HSCs first differentiate into either a multipotent lymphoid progenitor or a multipotent myeloid progenitor. Lymphoid progenitors then give rise to T-cell and B-cell progenitors, which are involved in the immune response against foreign substances (3). Myeloid progenitors give rise to myeloblasts and erythroblasts. Myeloblasts are the precursors to granulocytes, while erythroblasts eventually undergo erythropoiesis to become erythrocytes (4).

Under stimulation by specific growth factors, myeloid progenitors differentiate into erythroid burst forming units (BFU-E) and then erythroid colony forming units (CFU-E). CFU-Es then differentiate into a series of various erythroblasts, finally undergoing enucleation to yield reticulocytes and eventually mature erythrocytes (4). Erythrocytes contain no nucleus or organelles, and have a flat, biconcave disk shape, which helps maximize their surface area to facilitate gas exchange (3). These mature red blood cells are responsible for the delivery of oxygen from the lungs to tissues throughout the body and also for carrying carbon dioxide from those tissues back to the lungs.

In human adults, production of erythrocytes from HSCs in the bone marrow is carefully controlled by a negative-feedback system that senses the amount of oxygen reaching tissues from the blood (5). Under hypoxic conditions, in which tissues do not receive enough oxygen, the kidneys convert a membrane protein into the excreted hormone erythropoietin, which stimulates the bone marrow to produce more erythrocytes (3). When the tissues then receive more oxygen, the kidneys cease production of erythropoietin.

Hemoglobin Structure and Disease

In order to carry out their function of transporting oxygen and carbon dioxide throughout the body, erythrocytes utilize the hemoglobin molecule. The protein portion of hemoglobin is a tetramer comprised of a homodimer of two heterodimers of globin chains (6). In humans, embryonic hemoglobin consists of two ζ and two ε chains ($\zeta_2\varepsilon_2$), fetal hemoglobin (HbF) consists of two α and two γ chains ($\alpha_2\gamma_2$), and the majority of adult hemoglobin (HbA) consists of two α and two β chains ($\alpha_2\beta_2$) (7,8). Embryonic and fetal hemoglobin bind oxygen with higher affinity than the adult form, facilitating exchange from the mother's bloodstream to the developing embryo (9). Embedded in

each globin chain is a non-covalently bound ring-shaped heme group containing a central iron (Fe) atom, which is the site of reversible oxygen binding (10).

Mutations in the genes that encode the globin chains can cause defective and/or imbalanced globin chain synthesis, resulting in serious hemoglobinopathies, or blood diseases. However, many of these mutations can have protective effects against malaria, and thus these mutations are more prevalent in people of African, Indian, and Mediterranean descent—areas where there are also high incidences of malaria (11). Malaria is caused by parasitic protists of the genus *Plasmodium*, and humans can be infected by four different species: *P. falciparum* (the most common cause of severe malaria), *P. vivax*, *P. ovale*, and *P. malariae* (12). Infection by *Plasmodium* is facilitated by the mosquito, which injects the protists into human blood from its saliva. The parasites take up residence in the liver, and release more into the bloodstream where they then invade circulating erythrocytes and reticulocytes (12). The protists consume and metabolize hemoglobin, enlarging until they fill the cell completely, which subsequently lyses the infected erythroid cells (12). A set population does not lyse the cells, however, and instead produces gametocytes, which can be taken up by mosquitoes and consequently injected to other human hosts, restarting the infection process. Heterozygous individuals who carry one copy of a normal globin allele and one copy of a diseased globin allele often show increased resistance to the consequences of malaria due to the more rapid turnover of sickle-cell erythrocytes or other diseased erythrocytes. Thus, heterozygous individuals have higher relative fitness than individuals who are homozygous for either the normal globin allele or a diseased globin allele (13,14).

Overview of Hemoglobinopathies

Sickle-cell disease and thalassemias are both hemoglobinopathies that arise as a result of genetic defects, but differ in their classification and effect. Sickle-cell disease is caused by an aberrant globin chain structure, while thalassemias are caused by an abnormally low quantity of globin chains. The two disease phenotypes are not mutually exclusive, as a single individual may be affected by both. Currently, over 700 β-globin chain structural variants have been identified, and sickle-cell disease is specifically caused by the aberrant globin chain HbS (15). In HbS, a single base pair mutation in the adult β-globin gene causes a glutamic acid at amino acid position 6 to be mutated to a valine (1,16). Under low oxygen conditions, this causes the globin chains to form insoluble fibers, and erythrocytes containing these aberrant chains assume an abnormal, rigid, sickled shape characteristic of the disease. Sickle-cell erythrocytes can obstruct capillaries and restrict proper blood flow, causing extreme pain and possible internal organ damage (3).

Thalassemia is the most common monogenic disease worldwide. Approximately 300,000 to 500,000 infants a year are born with this disease, and it is estimated that 7% of the world's population are carriers (17). Thalassemias are classified first based on which globin chain is affected (α/β) and second based on the severity of the disease. Thus, β-thalassemias are caused by any number of mutations that either reduce or completely eliminate the expression of the adult β-globin gene. There are over 200 known mutations identified in β-thalassemia patients that cause this disease (18). The mildest form is known as β-thalassemia minor, a mild microcytic anemia that is typically asymptomatic, and is characterized by reduced levels of β-globin gene expression most often caused by mutations in transcription regulatory elements (19). The most severe

form is known as β-thalassemia major, which is associated with a severe microcytic hyperchromic anemia, and is characterized by the complete absence of β-chain synthesis. In between these two forms is β-thalassemia intermedia, which is characterized by a moderate to severe anemia (20). In β-thalassemia, the reduced production of β-globin chains can lead to imbalanced chain synthesis, causing aggregation and overproduction of functional α-globin chains. Erythrocyte precursors containing these aggregates are broken down in either the bone marrow or peripheral blood, which is what leads to the anemic phenotype (11).

Treatments

For many sickle-cell patients, the administration of hydroxyurea can alleviate symptoms. Hydroxyurea increases the amount of fetal hemoglobin in the blood, although the precise mechanism by which it acts is not completely known (16). Currently, no cure for severe forms of thalassemias exists, and the limited treatments available for patients include complications. Most conventional treatments are based primarily on blood transfusions. Unfortunately, with receiving frequent blood transfusions, iron chelation therapy also becomes necessary in order to prevent iron overload damage to internal organs (21,22). Since this process is tedious and painful, most patients choose not to endure it. In a 2000 study by Modell et al., it was found that 50% of thalassemia patients in the U.K. die before the age of 35, mostly as a result of noncompliance (23). Iron overload can also suppress the production of erythropoietin, which may further complicate a patient's symptoms of anemia (24). Alternatively, bone marrow transplants offer a more curative approach, but this requires finding a suitable donor and is dependent on the body's acceptance of foreign bone marrow.

The β-globin Gene Locus

Expression of the mammalian β-type globin genes, which encode for the ε, γ, and β chains of the hemoglobin molecule, occurs exclusively in erythroid cells. In humans, the β-globin gene locus is located on chromosome 11 while it is located on chromosome 7 in mice. The overall organization of the β-globin gene locus and its developmental regulation is conserved between the two species (Figure 1-1). The β-type globin genes are arranged linearly in the order of their developmental and tissue-specific expression, with the embryonic genes located at the 5' end of the β-globin gene locus, and the adult genes located at the 3' end (1,25). Both the human and murine β-type globin genes are also under the control of a powerful regulatory element known as the locus control region (LCR), which resides 6 to 22 kb upstream of the embryonic genes in both species (25,26).

In humans, ε-globin is expressed during the first 6 weeks of gestation in primitive erythroid cells originating from the embryonic yolk sack. During development, expression switches to erythroid cells generated during fetal liver hematopoiesis, where the two γ-globin, $^{G}\gamma$ and $^{A}\gamma$, genes are expressed. In a final switch completed shortly after birth, the γ-globin genes are silenced, and δ- and β-globin are expressed in erythroid cells that mature during bone marrow hematopoiesis. Although it also undergoes activation during this switch, δ-globin is expressed at 5% of β-globin expression levels due to a mutation in its promoter region (26). Unlike its human counterpart, the murine β-globin locus contains only four genes: εγ and β_{h1}, which are co-expressed in the embryonic yolk sac; and β_{maj} and β_{min}, which are expressed in the fetal liver and adult bone marrow. In mice, the switch from embryonic to fetal/adult expression occurs at 12 days *post coitum* (dpc).

Chromatin Structure and Function

In order to fit the entire genome into a single cell nucleus, eukaryotic DNA is packaged into nucleosomes and subsequent hierarchical organizational structures. Each nucleosome core particle is comprised of a histone octamer containing two copies each of histones H2A, H2B, H3, and H4, with 145-147 DNA base pairs wound around each octomer (27,28). This arrangement of nucleosomes is commonly referred to as "beads on a string." The linker histone H1 associates with DNA between single nucleosomes, establishing a higher level of organization coiled into a helical 30 nm fiber (29,30). Additional compaction occurs and leads to higher-order chromatin structure, but the exact architecture beyond the 30 nm fiber is unknown.

Histone Modifications

In addition to the globular carboxy-terminal domains that make up the core nucleosome, histones also have flexible amino-terminal tails that protrude outward from the nucleosome (27). These N-terminal tails are known to undergo post-translational modifications, which can affect the accessibility of the chromatin and underlying DNA (31). These modifications include, but are not limited to, acetylation, methylation, phosphorylation, glycosylation, ubiquitination, sumoylation, ADP-ribosylation, and carbonylation (32,33). The different combinations of these modifications at multiple and specific sites, and their effect on the surrounding chromatin, is termed the "histone code" (34,35). The functional outcome of these histone modifications, in terms of chromatin structure and gene expression, depends on the modification and the amino acid residue(s) modified.

Methylation of histones is carried out by transferring a methyl group from the donor *S*-adenosyl-*L*-methionine (SAM) to either the ε-NH_2 group of lysine (K) or the ω- or δ-

NH_2 of arginine (R) residues (36). Histone 3 (H3) K4, K9, K27, K36, and K79, and histone 4 (H4) K20 can be mono-, di-, or tri-methylated by the addition of one, two, or three methyl groups, respectively. Arginine residues are either mono-methylated, or symmetrically or asymmetrically di-methylated (36).

The transcriptional impact of histone methylation depends on the residue modified and the methylation status. Methylation of H3K9 and K27, as well as H4K20 residues has generally been found to be associated with heterochromatin and gene silencing, while methylation of H3K4, R17, R26, K36, and K79 has generally been found to be associated with euchromatin and active genes (37,38). Histone methylation was once considered irreversible, but in 2004 it was demonstrated by Shi et al. that LSD1/KDM1 is capable of demethylating mono- and di-methylated lysines, specifically H3K4 and H3K9 (39). Since then, several additional lysine demethylases have been identified, including Jumonji histone demethylases (JHDM), a family of demethylases which contain the Jumonji C (JmjC) domain (38,40).

Histone acetylation occurs at lysine residues and is carried out by a group of enzymes known as histone acetyl transferases (HATs). During acetylation, the acetyl moiety from the donor acetyl-Coenzyme A (acetyl-CoA) is transferred to the ε-NH_2 group of lysine. Histone deacetylases (HDACs) catalyze the reverse reaction, where CoA serves as an acceptor of the acetyl moiety instead (41). Histone 3 can be acetylated at H3K9, K14, K18, and K23, while histone 4 can be acetylated at H4K5, K8, K12, and K16. These acetylated histone tails can serve as signals for trans-acting factors containing bromodomains, which recognize acetylated lysine residues (42).

Chromatin Remodeling

Chromatin remodeling complexes are essential for regulating the binding of transcription factors as well as gene expression. Some complexes alter the chromatin structure to increase DNA accessibility for transcription, while others help generate chromatin structures that promote long-term gene silencing. Adenosine-5'-triphosphate (ATP)-dependent remodeling complexes use the energy from ATP hydrolysis to disrupt interactions between DNA and histones, which can cause changes in the structure and positioning of nucleosomes (43). At least five families of chromatin remodelers are known in eukaryotes: SWI/SNF, ISWI, NURD/Mi-2/CHD, INO80, and SWR1 (43-45). SWI/SNF is a 2 MDa multisubunit complex that is highly conserved in eukaryotes, and will be discussed later in this chapter in more detail. The SWI/SNF family of chromatin remodelers contain a bromodomain that targets these enzymes to histones with acetylated lysine residues on their N-terminal tails (46). They primarily act at gene promoters to promote transcription factor binding and transcriptional activation, but have also been implicated to have repressive activity (47). ISWI complexes appear to slide nucleosomes and play a role in the ordering and spacing of nucleosomes following DNA replication (48,49). They recognize histone tails as well as the linker DNA between nucleosomes, and act primarily to repress transcription (50,51). Similarly, the NURD complexes have been found to act as repressors, while INO80 complexes appear to evict nucleosomes from break sites to help facilitate the repair of double-stranded DNA breaks (52,53). SWR1 acts by replacing core histones with histone variants, such as exchanging H2A for histone variant H2A.Z (54,55). This histone variant has been implicated in both transcription activation and repression, and in some circumstances, is required for the recruitment of RNA Pol II and TBP (56,57). In mammals, H2A.Z is

essential, as H2A.Z$^{-/-}$ null mice are embryonic lethal (58). All of these complexes play a role in mediating gene regulation by altering the chromatin structure, which thereby affects the binding of factors.

Chromosome Territories and Nuclear Compartmentalization

The genome is arranged into discrete chromosome-specific territories, and chromatin can also be divided into two subcategories: heterochromatin and euchromatin (59,60). In general, heterochromatin is associated with silenced genes and condensed chromatin, while euchromatin is associated with active genes and a more open chromatin structure (59). Chromatin is highly dynamic, as the selective expression of genes residing in various chromatin structures is necessary to ensure proper gene regulation during development. Genes needed to be silent may remain condensed, while inducible or constitutively active genes must remain accessible. Chromatin itself can pose a barrier to transcription, as nucleosomes may hinder the association of the transcriptional machinery with the underlying DNA.

In addition to the physical packaging of DNA into nucleosomes and higher order structures, the eukaryotic cell nucleus itself is highly compartmentalized, both structurally and functionally (61). Studies have found that both centromeric and telomeric sequences, which are primarily heterochromatic, anchor chromatin to the nuclear periphery, an area of the cell that is typically associated with repression (62,63). During replication and mitosis, chromatin undergoes dramatic rearrangement in organization and structure. In addition to temporal regulation by the timing of the cell cycle, replication and transcription are also spatially regulated, and occur in separate and distinct areas of the nucleus known as factories (64). Presumably, a chromatin region that is engaged in active transcription would first need to dissociate from its

transcription factory and be relocated to a replication factory for replication, and after replication it would relocate back to transcription factories (65).

Transcription is not ubiquitous throughout the nucleus, but instead occurs in transcription factories, which are regions enriched with the active, elongating form of RNA polymerase II (66-68). Many active and highly expressed genes are transcribed in these transcription factories, and multiple genes undergoing transcription can also occupy the same factory or space (69). It is not known what nucleates transcription factories, and some data suggests that genes poised for transcription are repositioned to transcription factories, rather than directly undergoing the *de novo* assembly of a factory (69). The β-globin gene itself has been shown to associate with transcription factories in erythroid cells (61). During erythroid differentiation, the β-globin gene migrates away from the nuclear periphery and peri-centromeric heterochromatin (61). Deletion of the β-globin locus control region (LCR) not only reduces the expression of the β-type globin genes by 25- to 100-fold, but also affects the nuclear positioning of the locus during development (61,70).

Components of β-type Globin Gene Regulation

Transcription initiation is a multi-step process. It begins with nucleosome remodeling, then binding of transcription factors and co-activators to enhancers and promoters, which finally leads to the recruitment of basal transcription machinery to the core promoter (71). These events must all take place before transcription of the β-type globin genes can occur. In addition to general transcription factors, various *cis* and *trans* regulatory elements, both erythroid-specific and ubiquitous, are known to function in order to activate or enhance the expression of the β-type globin genes. These regulatory elements are described below.

General Transcription Factors

The process of generating RNA from DNA is termed transcription, and is accomplished by RNA polymerases. At least five different RNA polymerases are known to exist in eukaryotes, although two—RNA polymerase IV and V—are unique to plants. RNA polymerase I generates RNA sequences that are later processed into 18s and 28s rRNAs, while RNA polymerase II (RNA Pol II) generates mRNAs and a variety of functional small RNAs (e.g., U1 to U5 RNA involved in splicing), and RNA polymerase III generates cellular 5s rRNA and tRNAs (72-74). Because RNA Pol II generates mRNAs, which contain the information needed to produce the proteins necessary for the cell to function, it is the RNA polymerase most extensively studied. A wide range of transcription factors are required for RNA Pol II to bind to its promoters and initiate transcription.

The transcriptional machinery required for transcription by RNA Pol II include, but are not limited to, the transcription factors TFII-A, TFII-B, TFII-D, TFII-E, TFII-F, and TFII-H (71). Binding of TFII-D to the promoter is thought to be the first step in activating transcription. TFII-D is a multi-component transcription factor that contains a DNA binding subunit known as the TATA-binding protein (TBP) as well as approximately 13 TBP-associated factors (TAFs). This subunit recognizes a region of DNA located upstream of certain genes known as the TATA box, which contains the core sequence 5'-TATAAA-3'. In humans, the TATA box generally resides 25-30 nucleotides upstream of the transcription start site (TSS).

Following the binding of TFII-D, TFII-A binds next and serves to stabilize the interaction between TFII-D and the TATA-box (75,76). TFIIB associates with TFII-D at the promoter and directly recruits RNA Pol II to the TSS. TFII-F is required for the stable

association of RNA Pol II with the pre-initiation complex (PIC) (77). After formation of the TFII-D/TFII-B/RNA Pol II/TFII-F complex, TFII-E and TFII-H are recruited (78). TFII-E not only serves to both help recruit TFII-H to the PIC, but also to stimulate TFII-H activity. TFII-H has three crucial functions during the process of transcription initiation as an ATP-dependent ATPase, an ATP-dependent helicase, and a RNA Pol II carboxy-terminal domain (CTD) kinase (78).

Not all genes contain a TATA box, however, and these employ the use of an initiator element (INR), downstream promoter element (DPE), or downstream core element (DCE) in transcription initiation (71). The INR is a pyrimidine-rich region that surrounds the TSS, and is capable of directing accurate transcription initiation either alone or with TATA and other core elements (78). Interaction of TAFs with the INR, DPE, and DCE allow TFII-D to recognize TATA-less promoters, which TBP binds to with less affinity than the TATA sequence. The INR, DPE, and DCE can also act in TATA-containing genes by enhancing the strength of the promoter (71). Additionally, many eukaryotic gene promoters contain a TFII-B recognition element (BRE), which resides upstream and downstream of the TATA box and is bound by TFII-B. Binding of TFII-B to the BRE is thought to help orient the directionality of the transcription PIC (79).

After formation of the transcription PIC, TFII-H phosphorylates the CTD of RNA Pol II. The RNA Pol II CTD contains the tandem heptapeptide repeat Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS), which is repeated 52 times in humans (78,80). Both the serine 5 (Ser-5) and the serine 2 (Ser-2) residues of this heptapeptide repeat can be phosphorylated, which regulate different steps in the transcription cycle: initiation and elongation, respectively. Through its helicase activity, TFII-H melts the DNA. Both of

these events performed by TFII-H lead to transcription initiation and the binding of DRB Sensitivity Inducing Factor (DSIF) to RNA Pol II (81). DSIF then recruits Negative Elongation Factor (NELF), which momentarily arrests transcription (82). During this time, mRNA 5'-end capping enzymes are recruited through interactions with the Ser-5 phosphorylated CTD and DSIF (83). After 5'-end capping has taken place on the nascent mRNA, P-TEFb binds to RNA Pol II and phosphorylates both Ser-2 of the RNA Pol II CTD as well as DSIF. Phosphorylation of DSIF neutralizes the inhibitory behavior of NELF, which leads to transcription elongation and the eventual transcription-coupled recruitment of 3'-end processing factors (78,83).

Cis-acting Regulatory Elements

Cis-acting DNA regulatory elements contain binding sites which recruit transcription factors and/or members of the PIC, and thereby regulate transcription. These regulatory elements include promoters, enhancers, and locus control regions, which are classified based on their varying functions and locations relative to the gene which they act upon. Promoters are proximal regulators, typically located upstream (5') and adjacent to the genes they regulate, while enhancers and locus control regions (LCRs) are distal regulators. Enhancers can function with heterologous promoters in a position and orientation independent manner, while LCRs are able to activate gene expression in a tissue specific, copy number-dependent, and position-independent manner (84). The β-type globin genes are subject to regulation by the β-globin LCR, which resides upstream of the genes. LCRs were also first identified in the β-globin locus, and are thought to have chromatin opening activity. This was observed in transgenic mice carrying the human β-globin locus, where the locus remained in a relative open conformation despite its location of integration, indicating resistance to

position effects (85). A position effect is described as the effect on expression of a gene depending on the gene's location in the genome. The genes surrounding a transgene's location of integration can affect its conformation, and may consequently reduce or enhance its expression levels. Additionally, in β-thalassemia patients carrying the Hispanic deletion, in which an approximately 35 kb region of the LCR is deleted but the rest of the β-globin locus remains intact, the entire locus resides in a closed chromatin conformation (86).

In addition to the LCR, the main regulatory elements that mediate tissue-specific regulation of the adult β-globin gene are located proximal to the gene, including the 5' promoter, downstream promoter elements, an intronic enhancer, and a 3' enhancer (87,88). All of these elements likely work together in order to mediate the tissue and the developmental stage specific expression of the β-type globin genes.

The β-globin locus control region (LCR)

The β-globin LCR is an important *cis*-acting regulatory domain that is crucial for high-level expression of the β-type globin genes during all stages of erythroid development (89). It is comprised of several erythroid specific DNase I hypersensitive sites (HS), each separated from another by 2-4 kb of flanking DNA. These HSs contain several putative and confirmed binding sites for various regulators and co-regulators. Thus, the LCR can serve as the primary recruitment site for protein complexes that mediate both chromosome remodeling and transcription of the β-type globin genes (90).

In humans, the LCR contains five HSs (HS1 to HS5), which have varying functions but work together to activate the β-type globin genes. Each HS contains a core region of 200-400 bp. HS5 is present in multiple cell lineages and behaves as an insulator, while HS1 to 4 are only present in erythroid cells (91-93). HS3 is required for chromatin

opening, while HS2 behaves as a classical enhancer and also has the most powerful enhancing activity (94,95). HS2 contains binding sites for regulators and co-regulators that also bind to the adult β-globin promoter (Figure 1-2).

Some transcription factors known to be involved in LCR-mediated activation of adult β-globin gene transcription include the hematopoietic specific proteins GATA-1, EKLF, NF-E2, and TAL1, as well as the ubiquitously expressed transcription factor USF. These *trans*-acting factors involved in the transcription pre-initiation complex formation colocalize to the LCR and are thought to be subsequently transferred to the appropriate gene for transcription initiation (26). Several mechanisms for this interaction have been postulated, such as tracking and looping models.

The tracking model of LCR function. A tracking model of LCR function was proposed based on the observation that LCR HS2 initiates the formation of long non-coding transcripts (96). Most of this non-coding transcription proceeds uni-directionally toward the globin genes. According to the tracking model, the LCR recruits transcription complexes that track along the DNA until they reach the globin gene promoters. The model explains the chromatin opening function of the LCR because elongating transcription complexes in eukaryotic cells harbor chromatin modifying activities that either remodel nucleosomes or modify histone tails (97). It also explains the enhancer function, because the mechanism of tracking would deliver the polymerase to the genes.

Several recent findings support an RNA Pol II-tracking based mechanism for globin activation. Insulators placed between the LCR and the globin genes diminish expression of the β-globin genes (98,99). Furthermore, it has recently not only been

demonstrated that an insulator placed between the LCR and the embryonic ε-globin gene represses expression of the downstream gene but also, and perhaps more importantly, that RNA Pol II transcription complexes accumulate at the insulator (99). The role of intergenic transcription in modulation of chromatin structure is somewhat controversial. Inhibiting transcription elongation by DRB (5,6-dichloro-1-β-*D*-ribofuranosylbenzamidazole) in erythroid cells has no effect on chromatin modifications downstream of the LCR, suggesting that transcription elongation is not required for chromatin opening (100). DRB mediated inhibition also fails to prevent conformational changes that bring LCR and globin genes in close proximity (101). It should be noted however that recent studies demonstrated that long non-coding transcription is relatively insensitive to DRB (102). The fact that deletion of the murine endogenous LCR does not affect DNase I sensitivity or increase in histone acetylation in the remainder of the globin gene locus also argues against the hypothesis that LCR initiated intergenic transcription is required to open chromatin structure (87). Furthermore, a recent study has shown that intergenic transcription does not correlate with open chromatin domains in the globin gene locus, and inhibition of dicer causes an increase in the abundance of intergenic transcripts (103). This suggests that intergenic transcription could, in fact, have a negative role in mediating silencing of globin locus domains that are inactive at specific developmental stages.

The looping model of LCR function. The looping model proposes that LCR HSs interact with the globin genes to activate transcription (104,105). This interaction may be mediated by transcription factors and co-factors that interact with the HSs and the globin gene promoters. In this view, the looping model is a contact model, suggesting

direct contacts between the regulatory elements and the genes. However, it is also possible that there are no direct interactions between the LCR and the genes but that these elements instead are brought together in close physical proximity to allow transfer of activities from the LCR to the genes or to allow modifications of promoter bound activities by those recruited to the LCR. The recently applied chromatin conformation capture (3C) technology, which assays proximities between genes and regulatory elements, does demonstrate that the LCR and actively transcribed globin genes are in close proximity in the context of an ACH, or active chromatin hub (106). However, how close these elements come together is not known. A more open and perhaps dynamic configuration appears to be consistent with previous observations showing that γ-globin and β-globin genes rapidly switch their interactions with the LCR and also with previously proposed competition models (104,107).

While the transcription tracking mechanism may explain gene activation, it is not known how looping can lead to the enhanced expression of the globin genes. LCR-mediated gene activation either results in enhanced recruitment of transcription complexes to the globin genes and/or in the conversion of transcription initiation complexes to elongation active complexes (26,108). This could be achieved by providing activities that are first recruited to the LCR and subsequently transferred to globin genes. For example, transcription complexes could first be recruited to the LCR and looping would mediate the transfer to the globin gene promoters. The ACH would provide a high local concentration of transcription factors that could efficiently capture transcription complexes which would then be recruited to and positioned at the basal promoters of the globin genes to engage in productive transcription. Alternatively, or

additionally, elongation incompetent transcription complexes may be recruited to the genes while the LCR provides activities necessary for activation, for example, kinases that phosphorylate Ser-2 at the RNA Pol II CTD or co-regulator complexes that modify chromatin structure at the globin gene promoters either to enable recruitment of transcription complexes and/or to allow elongation.

Promoter regions

The human adult β-globin promoter consists of basal promoter elements and regulatory sequences (88). The basal promoter contains a TATA-like sequence (5'-CATAAA-3') and an initiator, both of which interact with components of the TFII-D complex (109,110). The upstream promoter region contains important elements for tissue-and stage specific regulation, including multiple binding sites for GATA-1 and EKLF (or Sp1). Both EKLF and Sp1 are thought to interact with the promoter at the -90 CACCC motif (111,112).

The promoter region also includes two E-box motifs that are conserved between humans and mice (Figure 1-3). One E-box overlaps with the initiator while the other one resides downstream from the transcription start site at +60. The downstream E-box is required for high-level *in vitro* transcription of the β-globin gene (113). It is believed that the upstream E-box is bound by the general transcription factor TFII-I and ubiquitous transcription factor USF1 while the downstream E-box is bound by a USF1/USF2 heterodimer. There is also a partial Ap1/MARE-like (maf recognition element) binding site at +24, which has been shown to bind NF-E2 with low affinity (113).

Proximal enhancer elements

In the human β-globin locus, both the fetal γ-globin gene and the adult β-globin gene contain enhancer elements at their 3' ends. The γ-globin 3' enhancer resides 410

bp downstream of the polyadenylation site of the $^{A}\gamma$-globin gene and is less than 750 bp in length (114). The role of this 3' enhancer during development and in the expression of the β-type genes is widely disputed. It has been implicated as a stage-specific silencer as well as reported to have no function at all (115). However, in a study where the region was deleted in transgenic mice carrying a yeast artificial chromosome containing the entire human β-globin gene locus (β-YAC), no effect on the high level, stage-specific, or position-independent expression of any of the human globin genes was observed (116). The γ-globin 3' enhancer may have a functionally redundant role in conjunction with other *cis*-regulators in the locus.

The adult β-globin 3' enhancer is located about 2.2 kb downstream of the promoter region, 500 to 850 bp downstream of the polyadenylation site (117). This enhancer contains four binding sites for GATA-1, and confers high level expression (118). Whereas deletion of the $^{A}\gamma$-globin 3' enhancer in β-YAC transgenic mice had no effect, deletion of the β-globin 3' enhancer in these mice caused a reduction in β-globin gene expression (119). The adult β-globin 3' enhancer is known to regulate transcription in both a tissue and developmental stage-specific manner (120-123). When the 3' end of the human β-globin gene was linked to the 5' end of the human $^{A}\gamma$-globin gene in transgenic mice, the hybrid 5'-$^{A}\gamma$-β-3' globin gene displayed an expression pattern similar to that of the adult gene (124).

Gene competition

It has been suggested that the β-globin genes compete for interaction with the LCR, and that only one gene is transcribed at a time (104,105). This theory postulates that only one active gene can come into contact with the LCR at any given time (125). Gene competition in the β-globin gene locus was first observed when transgenic mice

were generated using various constructs containing the LCR linked to specific globin genes (126,127). When the LCR was linked directly to only the γ-globin gene, γ-globin was expressed in the embryo, and to a lesser degree, in the adult mouse. When the LCR was linked directly to only the β-globin gene, however, β-globin was expressed at high levels at all stages of development. When both the γ-globin and β-globin genes were linked to the LCR, normal developmental expression was restored. This indicates that there may be a preferential interaction between the LCR and the γ-globin gene during fetal development, which changes to a preference for the β-globin gene during the adult stage of expression. Additionally, the relative position of the genes with respect to the LCR is also important for correct developmental expression. If the gene order relative to the LCR is inverted in transgenic mice, an aberrant expression pattern is observed, in which β-globin is expressed in embryonic cells, and ε-globin gene expression remains silent throughout development (128,129).

Trans-acting Regulatory Elements

There are many proteins involved in the regulation of β-globin gene expression. These proteins function as DNA binding transcription factors, as co-regulators modulating chromatin structure and/or recruiting transcription complexes, and as architectural proteins that change the conformation and perhaps nuclear localization of the globin locus during differentiation and development. These factors include the ubiquitously expressed transcription factor USF, and the hematopoietic specific proteins GATA-1 & GATA-2, NF-E2 (p45), TAL1, and EKLF.

Upstream stimulatory factor (USF)

USF is a ubiquitously expressed transcription factor that binds to DNA E-box motifs and has been associated with the transcription of many cellular and viral genes

(130). It belongs to a family of transcription factors characterized by their basic helix-loop-helix leucine zipper (bHLH-LZ) DNA binding domains (131). There are currently two known members of this family: USF1 (44 kDa) and USF2 (43 kDa). These two USF proteins bear close similarities to one another except at the N-terminus, and they also share a small but extremely conserved nuclear localization domain termed the USF-specific region (USR), located just upstream of the basic region (132). The predominant form of USF is a USF1/USF2 heterodimer, although homodimers are known to exist in various degrees across cell types (133).

The bHLH-LZ DNA binding domain is not unique to the USF family, and is also present in the Myc family of regulatory proteins. Like c-Myc, the USF proteins have similarly been implicated in controlling cell growth and proliferation, and they play a role in G_1/S and G_2/M cell cycle progression by regulating the expression of cyclin and Cdk genes (134-136). USF1 has been found to be an active regulator of cyclin B1 and cdc2 gene expression, which are involved in the G_2/M transition (134,135). The CDK4 gene, which is involved in the G_1/S transition, is activated by USF2 (136).

In mouse studies, USF1 and USF2 knockout mice display distinct phenotypes, further illustrating the regulatory differences between the two USF proteins. $USF1^{-/-}$ null mice are viable and fertile, but display elevated amounts of USF2, a possible compensatory reaction to the lack of USF1. $USF2^{-/-}$ null mice, however, have reduced levels of USF1 and display a growth defect. Double knockouts of USF1 and USF2 are embryonic lethal (137).

Though the USF proteins are ubiquitously expressed, they appear to primarily regulate genes that are expressed in a differentiation and tissue-specific manner (130).

Most genes activated by USF are expressed at high levels in differentiated cells, including the β-globin gene (130,138). USF binds with high affinity to DNA fragments containing a central 5'-CANNTG-3' consensus E-box motif and is known to be involved in the transcriptional regulation of genes containing this E-box (139,140). Additionally, a genome-wide mapping study of USF interaction sites in hepatocytes revealed that it preferentially binds DNA in close proximity to transcription start sites, and that this interaction correlates with increased levels of H3 acetylation (141). This suggests that USF may be involved in the recruitment of transcription complexes.

USF in β-globin gene regulation. Previous studies have shown that USF interacts with conserved E-box elements located in Hypersensitive Site II (HS2) of the LCR, as well as in the adult β-globin downstream promoter region (Figure 1-3) (138,142-144). USF interacts with co-activators and histone modifiers in erythroid cells, suggesting that it functions through chromatin remodeling and RNA Pol II recruitment (145,146). The general transcription factor TFII-I has also been shown to interact with USF at INR and E-box elements in order to coordinate gene activation or repression, and can recruit USF to these sites (147,148). TFII-I and USF have been found to co-localize at the adult β-globin INR, although they appear to have antagonistic effects on the adult β-globin gene expression, where TFII-I is a repressor and USF is an activator (113,138).

USF has also been shown to interact with the HAT CBP/p300, as well as with large co-regulator complexes such as the histone methyltransferases PRMT1 and SET1 (146,149). This suggests that USF, at least in part, regulates chromatin accessibility which may facilitate the assembly of transcription complexes at the LCR and at the adult

β-globin gene promoter. Additionally, USF also is known to function at chromatin barriers, such as the chicken β-globin 5'HS4 insulator element, and may serve to help maintain an environment of active chromatin (149,150).

GATA-1 and GATA-2

The GATA transcription factors are a family of six small zinc finger proteins that bind to DNA regions containing a 5'-(A/T)GATA(A/G)-3' sequence motif. GATA factors are known to interact with various co-regulators, including FOG-1 (friend of GATA-1), CBP (a co-activator with HAT activity), and mediator (a large co-activator complex associated with RNA Pol II) (151). During development, GATA factors play a prominent role in differentiation, proliferation, and organ morphogenesis. GATA-1, -2, and -3 are important regulators of hematopoietic stem cells and their derivatives, while GATA-4, -5, and -6 are expressed in various mesoderm and endoderm-derived tissues.

GATA-1 has been shown to both increase histone acetylation in the LCR and in globin gene promoters, as well as to participate in the recruitment of RNA Pol II to the LCR and to the promoters (152). Additionally, GATA-1 represses *c-Kit* signaling at multiple levels to coordinate decreased cell proliferation (153). However, GATA-1 does not indiscriminately bind to all GATA sites in accessible chromatin, and interaction with specific sites is facilitated by the zinc finger protein FOG-1 (151). Both GATA-1 and FOG-1 are essential for erythroid and megakaryocyte development. FOG-1 interacts with the amino (N) finger of GATA-1 and cooperates to promote differentiation by mediating the proximity between the LCR and the adult β-globin gene (154). GATA-1 has also been implicated in transcription repression and associates with the NURD co-repressor complex (155).

As GATA-1 levels increase upon cell maturation, GATA-2 expression is silenced; thus, erythroid differentiation is accompanied by an exchange of GATA factors. It is believed that GATA-2 promotes the survival and expansion of hematopoietic cells and acts to block differentiation. GATA-2 is expressed in hematopoietic progenitors, including early erythroid cells, mast cells, and megakaryocytes, and also in non-hematopoietic embryonic stem cells.

NF-E2

NF-E2, also known as Nuclear Factor Erythroid 2, is a heterodimer comprised of a small ubiquitously expressed subunit (p18) and a large hematopoietic subunit (p45). Both subunits of NF-E2 both contain leucine zipper and DNA binding domains, and interact with MAREs (151). Mice deficient for p45 do not exhibit a significant erythroid phenotype or reduction in β-globin gene expression. Studies have shown that a decrease in p45 activity is accompanied by an increase in occupancy of β-globin locus associated NF-E2 sites by the NF-E2 related protein, NRF2 (156). The same group demonstrated that NF-E2 is not required for conformational changes that reduce the distance between the LCR and β-globin gene promoter. NF-E2 has been shown to interact with other proteins, including components of the TFII-D complex and protein complexes with chromatin modifying activities (151,157). It appears that in erythroid progenitor cells, MARE sites in the LCR are occupied by Bach1, an NF-E2 related protein that functions as a repressor (158). Heme-induced differentiation leads to the dissociation of Bach1 and the subsequent association of the activator NF-E2.

TAL1/SCL

TAL1 stands for T-cell acute lymphocytic leukemia protein 1, and is also known as SCL. TAL1 is a hematopoietic specific helix-loop-helix protein that heterodimerizes with

ubiquitously expressed E12/E47 proteins and interacts with E-box sequences (151). It has also been shown to interact with TFII-H, a member of the basal transcription complex (159). TAL1 has an essential role during early stages of erythroid specification, but the function of TAL1 in more differentiated and adult erythroid cells is far less understood. TAL1 and its heterodimeric partner interact with a variety of proteins, and can recruit both co-activators and/or co-repressors to target regulatory elements. It also forms a complex with NL1/Ldb1 and GATA-1, which interacts with a modular *cis*-element composed of GATA and E-box sequences, although it is not clear whether individual sites are able to recruit the entire complex *in vivo* (151). TAL1 has been shown to interact *in vitro* with an E-box located in the LCR element HS2 (143). More recently, it has been shown that TAL1 can be crosslinked to the LCR as well as to the β-globin gene promoter, and that NL1/Ldb1 is required for conformational changes that bring the LCR and the adult β-globin gene into close proximity (160).

Erythroid Krüppel-like factor (EKLF)

EKLF, also known as Krüppel-Like Factor 1 (KLF1), is an erythroid-specific zinc-finger transcription factor that binds to DNA sequences containing a CACCC motif (112). It belongs to a family of transcription factors called the Krüppel-like factors, which derives its name from the family's high degree of similarity to the *Drosophila* transcription factor, Krüppel. In humans, EKLF is expressed in bone marrow and erythropoietic cells, but not in myeloid or lymphoid cell lines.

EKLF is known to interact with LCR elements HS3 and HS2 as well as the adult β-globin promoter in both mice and humans (112). In humans, EKLF regulates the adult β-globin promoter specifically through the -90 CACCC sequence. EKLF is also required for mediating proximity between the LCR and the adult β-globin gene promoter (161).

EKLF$^{-/-}$ knockout mice die at ~14.5 dpc due to severe anemia resulting from ineffective erythropoiesis (162). The fetal livers of these embryos are pale, and they fail to express high levels of adult β_{maj}, although no change is observed in the embryonic εγ and β_{h1} expression. Additionally, these EKLF-deficient mice reveal a reduction in the formation of LCR elements HS2 and HS3.

Although the main function of EKLF appears to be in activating adult β-globin gene expression, it is present in primitive erythroid cells and surprisingly was found to bind the β-globin gene in these cells (163). Although it is expressed at all stages, earlier works suggested that EKLF is not required for yolk sac erythropoiesis, erythroid commitment, or expression of other potential target genes (164). Its stage-specific and β-globin gene-specific requirement suggests that EKLF may facilitate completion of the fetal-to-adult switch. EKLF expression increases during development, and it is speculated that increased expression leads to greater association with co-factors that would then be recruited to the β-globin gene in adult erythroid cells. The zinc-finger DNA binding domain of EKLF has also been shown to interact with the BRG1 subunit of the SWI/SNF complex, indicating that it also plays a role in recruiting chromatin remodeling enzymes to the β-globin locus (165,166).

Co-Regulators

In addition to *cis* and *trans* regulatory elements, several co-regulators also contribute to the regulation of the β-globin locus such as the SWI/SNF family of chromatin remodeling complexes, and CREB binding protein (CBP). SWI/SNF plays a role in re-positioning nucleosomes to allow access for transcriptional machinery to the underlying DNA and has been shown to be involved in erythropoiesis. CBP is a co-

activator that is known to bind at enhancer regions. These two are described in more detail below.

SWI/SNF

SWI/SNF is a chromatin remodeling complex that uses energy derived from ATP hydrolysis to disrupt the chromatin architecture of target promoters (167). The SWI/SNF complex was first identified in yeast through screening for genes that regulated mating-type switching (SWI) and sucrose non-fermenting phenotypes (SNF). In general, genes targeted by the SWI/SNF complex are often regulated by highly inducible promoters, respond to growth conditions, or are important for cellular development and differentiation (167). The SWI/SNF complex is also known to play a role in DNA repair.

Since its discovery, SWI/SNF homologues have also been found in other species, including *Drosophila*, mice, and humans. Mammalian SWI/SNF consists of 8~15 subunits of various BRG1-associated factors (BAFs), and contains either Brahma (BRM) or Brahma-Related Gene 1 (BRG1; also known as SMARCA4, which stands for SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4) as the central catalytic ATPase subunit (166). Although BRM and BRG1 are similar in sequence and have comparable *in vitro* activity, knockout studies have revealed distinct differences. $BRM^{-/-}$ null mice are viable and express increased levels of BRG1, indicating a possible compensatory response (168). Conversely, $BRG^{-/-}$ null mice fail to develop beyond the peri-implantation stage of early embryogenesis (169). In order to circumvent this issue for the purposes of studying the role of BRG1 in development, mutagenesis studies have produced mice carrying a partial loss-of-function mutation of BRG1 ($BRG1^{ENU1}$). In these mice, BRG1 has been implicated in β-globin regulation and erythropoiesis (170).

Additionally, BRG1 contains a unique N-terminal domain not present in BRM, with which it can bind to zinc-finger proteins such as EKLF, GATA-1, and Sp1 (166,171). During cellular differentiation, BRM protein concentrations increase, while BRG1 protein levels remain relatively constant in all cells. The unique properties of each catalytic ATPase subunit may allow them to direct different cellular processes through chromatin remodeling by making use of various specific protein-protein interactions.

CBP

CBP stands for CREB (cAMP response element binding protein) binding protein. As evidenced by its name, the CBP protein is associated with and co-activates the transcription factor CREB (172). CREB binds to certain DNA sequences called cAMP response elements (CRE) in order to increase or decrease the transcription of genes. CBP contains a bromodomain, a protein domain that recognizes acetylated lysine residues, such as those in modified histone N-terminal tails (173). These domains are often found in co-activators involved in signal-dependent, but not basal, transcription.

CBP is associated with the E1A binding protein p300 (EP300 or p300), and the two share high homology. Both proteins play critical roles in embryonic development and growth; knockout studies on CBP and p300 in mice have shown that both are required for proper development, as loss of either is embryonic lethal (174,175). Both of these proteins have histone acetyltransferase activity, and are capable of acetylating all the core histones in a nucleosome (176,177). Recently, genome-wide studies have established that many, if not most, enhancers are bound by CBP/p300 and H3K4me1, and that these enhancers are often located very distal from known transcription start sites (178-180).

In studies involving transient transfection of murine non-hematopoietic cells, it was found that overexpression of CBP stimulated GATA-1 expression. Co-IP studies on MEL nuclear extract found an association between the two endogenous proteins, which could be pinpointed to the zinc finger region of GATA-1 and to the E1A-binding region of CBP (181). Additionally, CBP is known to acetylate both GATA-1 and EKLF, which may assist their role in activating the adult β-globin gene (182,183).

Summary

The β-globin gene locus is one of the most extensively studied eukaryotic loci due to several unique aspects of its regulation, as well as for its role in sickle-cell disease and β-thalassemia. The regulation and maintenance of β-type globin gene expression during development and throughout life is essential to survival, and gaining a better understanding of how the adult β-globin gene is regulated is the first step toward finding better treatments for patients suffering from these hemoglobinopathies. Identifying key players in its regulation during development would greatly contribute to this understanding.

It had previously been shown that the ubiquitously expressed transcription factor USF as well as the BRG1 subunit of the SWI/SNF complex both have a role in activating the adult β-globin gene. The goal of the work presented here is to further characterize that knowledge. Although ubiquitously expressed transcription factors have been implicated in erythropoiesis, it is likely that they act in an erythroid-specific manner by association with erythroid-specific factors; this would allow the optimal utilization of available resources in the context of an erythroid system in order to promote erythropoiesis.

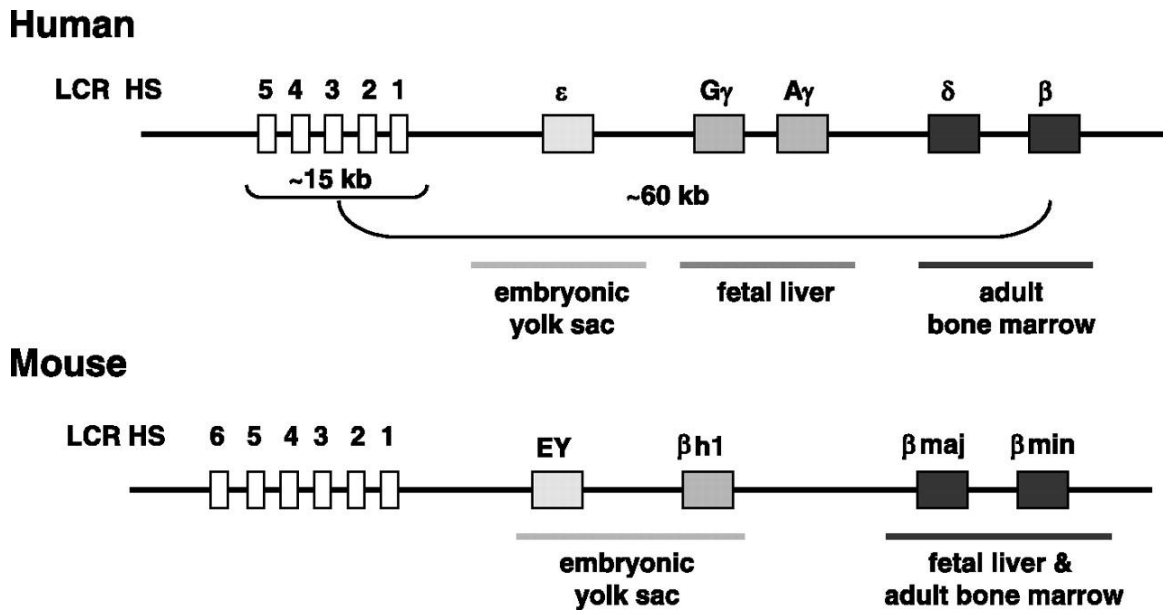

Figure 1-1. Schematic representation of the structural organization of the human (top) and mouse (bottom) β-globin gene loci (138). Tissue- and stage-specific expression is indicated below the genes.

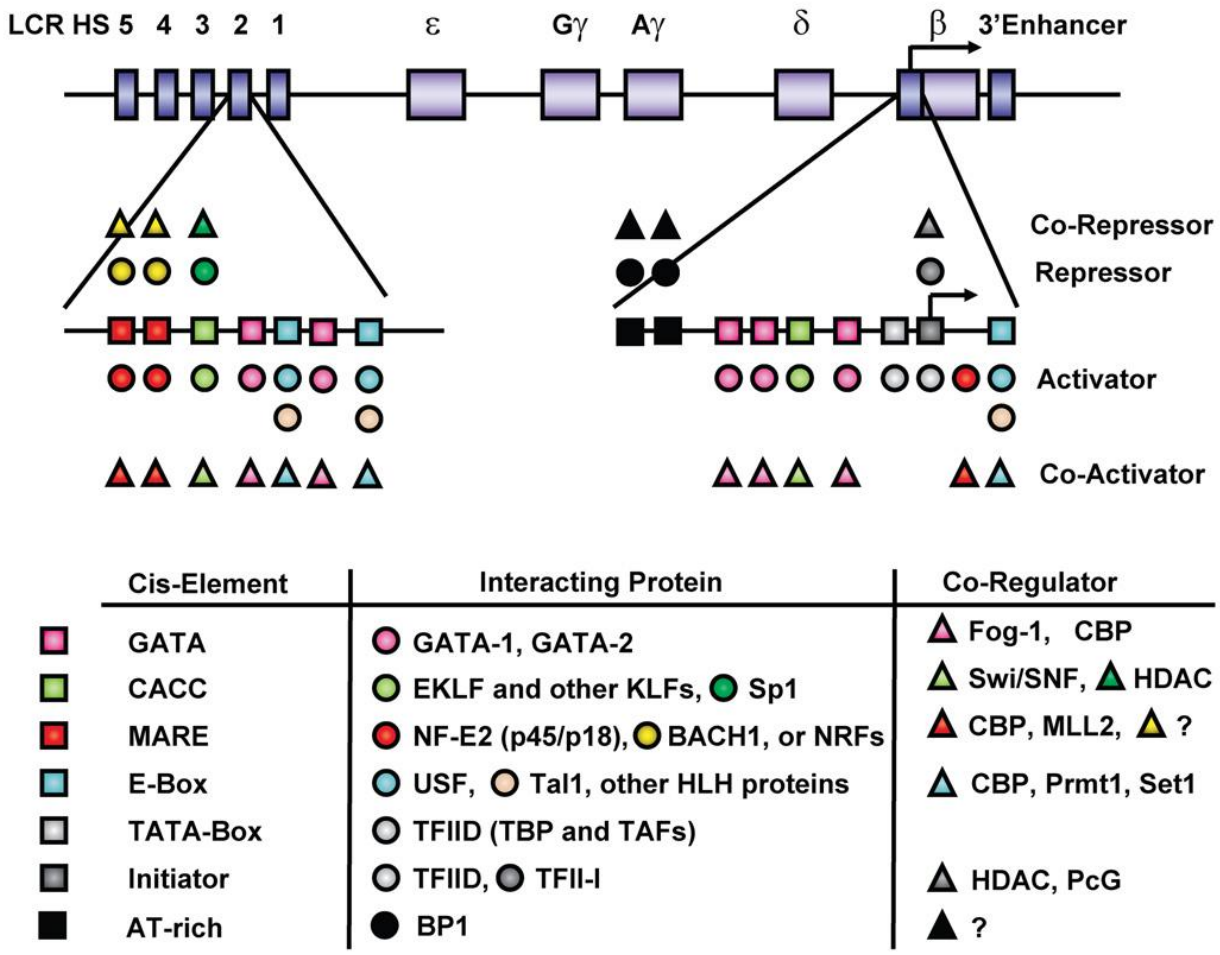

Figure 1-2. Summary of proteins and co-regulators interacting with LCR element HS2 and with the adult β-globin gene promoter (184). Shown on top is the overall organization of the human β-globin gene locus depicting the LCR and the embryonic (ε), fetal (Gγ and Aγ), and adult (δ and β) globin genes. The diagram below illustrates the overall organization of regulatory elements associated with adult β-globin gene regulation. Transcription factor binding sites (boxes) in LCR element HS2 (on the left) and the β-globin promoter (on the right) are bound by transcription factors (circles) that either activate (shown below the binding sites) or repress (shown above the binding sites) transcription through the recruitment of co-regulators (triangles). The binding sequences, interacting proteins, and co-regulators are listed in the table. DNA binding motifs as well as interacting proteins and their known co-regulators are color matched. No co-regulators have been described for Bach1 and Bp1 (indicated by question marks).

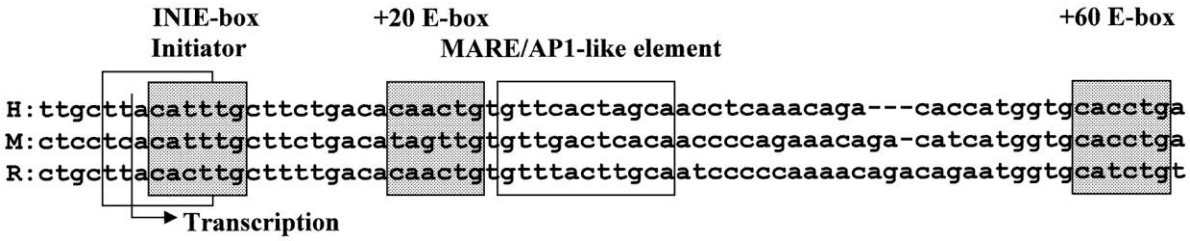

Figure 1-3. Sequence alignment of human (H), mouse (M), and rabbit (R) downstream promoter region of the adult β-globin gene (113). Shaded rectangles indicate regions containing consensus E-boxes motifs (CANNTG) where USF may bind.

CHAPTER 2
MATERIALS AND METHODS

Cell Culture

Human erythroleukemia (K562) cells were grown in RPMI 1640 containing L-glutamine and supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (P/S). Cells were grown in 5% CO_2 at 37°C and maintained at a density of around 1.5×10^5 to 2.5×10^5 cells/mL.

Murine erythroleukemia (MEL) cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing L-glutamine, 4.5 g/L glucose, and sodium pyruvate, and supplemented with 10% FBS and 1% P/S. Cells were grown in 5% CO_2 at 37°C and maintained at a density between 1×10^5 and 2×10^6 cells/mL. Generally, induction of erythroid differentiation of MEL cells was achieved by incubating cells in medium containing 1.5% dimethylsulfoxide (DMSO) for 72 h. In certain situations, induction was achieved by incubation with 2% DMSO for 24 h, and this deviation will be noted in the appropriate figure legends.

Cell Cycle Synchronization and Mitotic Arrest

Cell synchronization of K562 cells at the border of G_1- (gap 1) and S-phase (synthesis) was performed as described previously in Vieira, et al. (90). In order to arrest cells in M-phase (mitosis), cells were incubated in 60 ng/mL Nocodazole (Sigma) for an additional 10-12 h after incubation in fresh medium for 5 h following synchronization. Verification of synchronization and arrested cells was achieved by propidium iodide staining followed by flow cytometry as described by Vieira, et al. (90). Cells were taken for flow cytometry, chromatin immunoprecipitation (ChIP), and

chromosome conformation capture (3C) analyses at specific time points after release from synchronization or block.

A-USF Transgene Construction

The pITRp543f2AUSF4 plasmid used to generate transgenic mice was created previously by Dr. Valerie Crusselle-Davis (138). In pITRp543f2AUSF4, the dominant-negative Upstream Stimulatory Factor (A-USF) gene was placed under the control of the human adult β-globin promoter and 3' enhancer in order to confer erythroid-specific expression (Figure 2-1). Additionally, the human β-globin locus control region (LCR) elements HS3, HS3/HS2 linker, and HS2 were included for high-level expression. Finally, the transgene was flanked by two chicken HS4 sequences which have been shown to have insulator activity, thus protecting the transgene from position effects attributable to its location of integration (185).

Generation and Identification of Transgenic Mice

The pITRp543f2AUSF4 plasmid DNA was first linearized by digestion with Asc I, and then resuspended in injection buffer (5 mM Tris-HCl [pH 7.4], 0.1 mM EDTA) at a concentration of 2 ng/µL. The linearized plasmid DNA was injected into fertilized murine oocytes as described by Bungert et al. and implanted into the uterus of pseudopregnant surrogate females (186). For the analysis of transient transgenic mice, the implanted embryos were isolated at 11.5 dpc rather than allowing the pseudopregnant females to give birth. DNA was prepared from offspring using tail clips from adult mice or the head of embryos by overnight digestion in DNA lysis buffer containing proteinase K and was purified by a series of one phenol, one phenol-chloroform-isoamyl alcohol (25:24:1), and one chloroform-isoamyl alcohol (49:1) extraction. The presence of the A-USF expression construct in transgenic mice was determined by PCR. Primers used for

amplification of the A-USF transgene were 5’-TGACGAAGAAGAACTCGAGGA-3’ (upstream [US]) and 5’-ACGACCTCTAATCCGTGGTG-3’ (downstream [DS]).

Phenylhydrazine Treatment

To induce hemolytic anemia in adult mice, a total of three intraperitoneal injections of phenylhydrazine hydrochloride (Sigma) were administered at an estimated amount of 40 μg/g of bodyweight. Mice were weighed, and the appropriate amount of phenylhydrazine was dissolved in a 0.9% saline solution so that the total volume of each injection was approximately 0.1 mL. The first injection was administered in the morning with an additional injection 8 hours later, while the third injection was administered in the morning of the following day (0, 8, and 24 h). Spleens were isolated from anemic animals 6 days after the first injection and subjected to RNA or ChIP analyses.

RNA Isolation and Analysis

RNA was isolated using the guanidinium-thiocyanate method as described previously and reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad) (186,187). RNA was subjected to analysis by quantitative real-time reverse transcription PCR (qRT-PCR) using the MyiQ (Bio-Rad) system, and reactions were carried out using the iQ SYBR green super mix (Bio-Rad). qRT-PCR conditions were the following: 95°C for 5 min, followed by 40 cycles of 94°C for 30 s, 59°C for 30 s, and 72°C for 1 min (readings were taken after every cycle). A melting curve was performed from 60 to 95°C (readings every 0.5°C). Standard curves were generated using 10-fold serial dilutions of wild-type complementary DNA (cDNA) from the appropriate source. Final quantification analysis was performed using the relative standard curve method. RNA subjected to analysis by reverse transcription PCR (RT-PCR) was electrophoresed on either agarose

gels and visualized with ethidium bromide, or on 5% Tris-borate-EDTA (TBE) polyacrylamide gels, stained with SYBR Green I (Invitrogen), and scanned with a Storm scanner.

Gene expression analysis was performed as described previously, with results reported as expression relative to wild-type levels after the normalization of the transcript data to those of a control gene, GAPDH (138,146). Primers for amplifying murine β_{maj}-globin, murine β-actin, and murine GAPDH have been described previously (138). RNA from wild-type and transgenic mice was isolated from embryonic yolk sac, fetal liver, or the spleen of phenylhydrazine-treated anemic mice as described in Bungert et al. (186). Primers for the murine β_{maj}-globin gene and the murine LCR element HS2 used to amplify cDNA in qRT-PCR analyses were as described in Crusselle-Davis et al. (138). Additional primers used to amplify cDNA can be found in Table 2-1. Primers for mouse GATA-1 and EKLF analysis were described by Tanabe et al., primers for mouse β_{h1} were described by Basu et al., and primers for mouse Band3 were described by Nilson et al. (188-190).

Chromatin Immunoprecipitation (ChIP) and MicroChIP Analyses

Conventional chromatin immunoprecipitation (ChIP) assays were performed as described previously (113,138). Fetal liver cells were homogenized in phosphate-buffered saline (PBS) containing protease inhibitors and passed through a 70 μm cell strainer. The following antibodies were used for ChIP in this study: USF1 and USF2 (H-86 and N-18, respectively; Santa Cruz Biotechnology), RNA Pol II (CTD45H8; Upstate Biotechnology), serine 5 phosphorylated RNA Pol II (H14; Covance), serine 2 phosphorylated RNA Pol II (YSPTSPS phospho S2; Abcam), CBP (A-22; Santa Cruz Biotechnology), TFII-B (C-18; Santa Cruz Biotechnology), and rabbit polyclonal BRG1

(H-88; Santa Cruz Biotechnology). For ChIP assays using IgM antibodies, Dynabeads rat anti-mouse IgM (Invitrogen) was used instead of Protein A-Sepharose beads, with normal mouse IgM as a negative control. Dynabeads were prepared as per the manufacturer's protocol.

MicroChIP (μChIP) was used according to a previous protocol with minor modifications (191,192). Antibody-bead complexes were prepared using Dynabeads Protein A beads (Invitrogen). Embryonic yolk sacs were cross-linked with 1% formaldehyde in 500 μL PBS and quenched with 125 mM glycine. Prior to sonication, yolk sacs were homogenized using a glass tissue grinder (Radnoti) and washed with PBS. Based on trial and error, sonication conditions were optimized to yield fragments of ~500 bp, and sonication products were diluted 10-fold in μChIP RIPA buffer (10 mM Tris-HCl [pH 7.4], 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-desoxycholate) containing 1mM PMSF and additional protease inhibitors (Roche). Sonicated chromatin was incubated with various antibody-bead complexes, and after a series of washes, DNA was purified using a QIAprep Spin Miniprep kit (Qiagen).

Quantitative (real-time) PCR (q-PCR) conditions were the following: 95°C for 5 min, followed by 40 cycles of 94°C for 30 s, 59°C for 20 s, and 72°C for 30 s. A melting curve was performed from 60 to 95°C (reading every 0.5°C). Negative control experiments were performed using immunoglobulin G (IgG) antibodies and primers amplifying a region between LCR elements HS2 and HS3 that does not interact with USF and RNA Pol II (90,146, and data not shown). Standard curves were generated

using 10-fold serial dilutions of the input DNA. Final quantification analysis was performed using the relative standard curve method.

ChIP primers used for q-PCR analyses can be found in Table 2-2. Additional primers used for amplifying specific regions of the human and murine β-globin locus have been described previously (90,138). Primers for the mouse EKLF gene promoter were described by Vakoc, et al. (154).

Chromatin Conformation Capture (3C) and ChIP-3C (ChIP-loop)

3C was performed on K562 cells as described by Dekker, et al., with minor modifications (193,194). The restriction enzyme Hind III was used for the initial digestion, and 3C primers for the human β-globin gene locus were described by Chan, et al. (195). Additionally, ChIP-3C (a.k.a. ChIP-loop) studies were performed on MEL cells as described by Song, et al., with minor modifications (160). This procedure combines ChIP and 3C, and allows for the identification of chromatin interactions that are bound or facilitated by specific proteins (196). Cells were lysed in 3C lysis buffer (10 mM Tris-HCl [pH 8.0], 10 mM NaCl, 0.2% NP-40) containing protease inhibitors and homogenized in a glass tissue grinder (Radnoti). The initial digestion was performed with either Hind III or Bgl II, and antibodies used to pull down protein-DNA complexes are the same as described in ChIP, with the addition of di-methylated H3K4 (Upstate).

Analysis of 3C or ChIP-3C ligation products were analyzed by conventional PCR and electrophoresis on 1.2% agarose gels or on 5% TBE gels (Bio-Rad). Conventional PCR conditions for 3C and ChIP-3C were the following: 95°C for 5 min, followed by 34 cycles of 95°C for 1 min, 60°C for 45 s, 72°C for 2 min, with an additional 95°C for 1 min, 65°C for 45 sec, and 72°C for 8 min afterward. ChIP-3C primers for amplifying Hind

III digests of the murine β-globin locus were provided by Dr. Ann Dean and can be found in Table 2-3 (160).

Fluorescence-Activated Cell Sorting (FACS)

Cells obtained from the yolk sacs of transgenic and wild-type murine male embryos at 10.5 days *post coitum* (dpc), as determined by Y chromosome-specific PCR, were subjected to fluorescence-activated cell sorting (FACS) analysis using antibodies against fluorescein isothiocyanate CD71 (BD Biosciences) and phycoeryrthrin-Cy7 Ter-119 (eBioscience). Cells were homogenized in PBS containing 2% FBS using a glass tissue grinder (Radnoti), passed through a 70 µm cell strainer, and incubated on ice with antibodies for 30 min. After a series of washes to remove unbound antibodies, cells were subjected to FACS using a BD LSRII system. Ter-119$^+$ yolk sac cells collected for RNA extraction were homogenized with collagenase in PBS containing 20% FBS. Y chromosome primers used for identifying male embryos were described by Kunieda et al. (197). CD71$^+$ cells represent early erythroid progenitor cells, while Ter-119$^+$ cells represent more mature erythroid cells (198).

Nuclear Extraction and Co-Immunoprecipitation (Co-IP)

Nuclear extraction and co-immunoprecipitation (Co-IP) was performed on MEL cells as described by Leach, et al. (144). Protein concentration from both uninduced and induced MEL nuclear extracts was determined by Bradford assay, and RIPA buffer (50 mM Tris-HCl [pH 7.4], 100 mM NaCl, 10 mM EDTA, 0.25% Na-Desoxycholate, 1% NP-40, 0.1% SDS) was added to lysates to yield a final protein concentration of 1 mg/mL. Precipitation was carried out by incubating diluted extracts with 2.5 µg of antibody for 2.5 h. The antibodies used in Co-IP are the same as those used in the ChIP assays, with the exception of BRG1. For Co-IP studies, mouse monoclonal antibody against

BRG1 was used, which was provided by Dr. David Reisman (UF). Complexes were then captured by adding Protein-A Sepharose beads (GE Healthcare) and incubating for an additional 2 h. Samples were washed 3 times with RIPA buffer and eluted with Laemmli buffer (Bio-Rad) at 95°C for 10 min. Eluted samples were loaded onto 7.5% Tris-HCl Ready Gels (Bio-Rad) and analyzed by immunoblot.

Protein Isolation and Immunoblotting

Protein extract for immunoblot analysis was obtained by resuspending cells in RIPA buffer (50 mM Tris-HCl [pH 7.4], 100 mM NaCl, 10 mM EDTA, 0.25% Na-Desoxycholate, 1% NP-40, 0.1% SDS) containing protease inhibitors (Roche) and incubation on a rotating wheel at 4°C for 30 minutes. Following incubation, lysates were centrifuged and the pellet was discarded. The protein concentration was determined using the Bradford assay. A total of 20 µg was incubated with Laemmli buffer (Bio-Rad) at 95°C for 10 min before being loaded onto 7.5% or 10% Tris-HCl Ready Gels (Bio-Rad). After transfer to a nitrocellulose membrane (Bio-Rad), proteins were detected using Immobilon Western Detection Reagents (Millipore) according to the manufacturer's protocol. Primary antibodies used are the same as those used in the ChIP assays. Secondary antibodies used include goat anti-rabbit (IgG-HRP, sc-2004) and goat anti-mouse (IgG-HRP, sc-2005), both obtained from Santa Cruz Biotechnology.

Table 2-1. Partial list of names and sequences of primer pairs used for qRT-PCR

Primers	Sequence
Mouse Hba α1	US 5'-CCTGGGGGAAGATTGGTG-3'
	DS 5'-GCCGTGGCTTACATCAAAGT-3'
Mouse β_{min}	US 5'-TGAGCTCCACTGTGACAAGC-3'
	DS 5'-TACTTGTGAGCCAGGGCAGT-3'
Mouse HoxB4	US 5'-TGGATGCGCAAAGTTCACG-3'
	DS 5'-GGTCTTTTTTCCACTTCATGCG-3'
Mouse USF1	US 5'-GATGAGAAACGGAGGGCTCAACATA-3'
	DS 5'-TTAGTTGCTGTCATTCTTGATGACG-3'
Mouse NF-E2 (p45)	US 5'-TCAGCAGAACAGGAACAGGT-3'
	DS 5'-GCTTTGACACTGGTATAGCT-3'
Mouse TAL1	US 5'-TAGCCTTAGCCAGCCGCTCG-3'
	DS 5'-GCGGAGGATCTCATTCTTGC-3'

Table 2-2. Partial list of names and sequences of primer pairs used for ChIP

Primers	Sequence
Human 3' $^{A}\gamma$-globin gene	US 5'-CCATGATGCAGAGCTTTCAA-3'
	DS 5'-TTTGCTCATCAAAACCCACA-3'
Human Necdin promoter	US 5'-GTGTTATGTGCGTGCAAACC-3'
	DS 5'-CTCTTCCCGGGTTTCTTCTC-3'
Mouse GATA-1 promoter	US 5'-AGCCTCTGCTTGAAATGCTC-3'
	DS 5'-CCTTTGGCTTCTGTGGAGTC-3'
Mouse TAL1 promoter	US 5'-CAGATCCGTTAGAGGGTTCG-3'
	DS 5'-CTGGGAATTACCTCGTGTGC-3'
Mouse NF-E2 (p45) promoter	US 5'-GCAGACACAGTGAGCACTCC-3'
	DS 5'-GAGGGTCCTTAGGTGGGAGA-3'
Mouse Necdin promoter	US 5'-TTTACATAAGCCTAGTGGTACCCTCC-3'
	DS 5'-ATCGCTGTCCTGCATCTCACAGTCG-3'

Table 2-3. List of names and sequences of primer pairs used for ChIP-3C, Hind III digest

Primers	Sequence
Mouse HS2	5'-GTTAATGAAATGCTATTTGGAATGG-3'
Mouse εγ	5'-CTAAAGAAACGCCAGACTGATTTAC-3'
Mouse β_{h1}	5'-AAGGGTGAGAGTTTAGCCTTCTCTA-3'
Mouse β_{maj}	5'-TACTCCCTCTGAATAATGTTTGTCC-3'
Mouse 3' HS1	5'-TTAAATTCATCTGGAAAGGCAAATA-3'

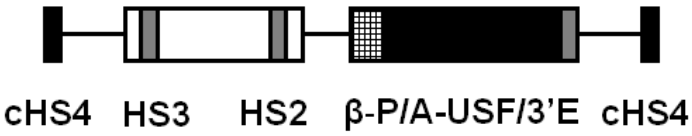

Figure 2-1. DNA construct pITRp543f2AUSF4 used to generate transgenic mice expressing dominant-negative USF (A-USF). The A-USF coding region is under the control of the human β-globin gene promoter (β-P) and 3' enhancer (3'E) as well as human LCR elements HS2 and HS3 plus flanking DNA. The DNA construct is flanked on either site by insulator elements derived from the chicken β-globin gene locus (cHS4).

CHAPTER 3
THE CONSEQUENCE OF EXPRESSING DOMINANT-NEGATIVE USF IN MICE

Introduction

Previous research demonstrated that expression of dominant-negative Upstream Stimulatory Factor (A-USF) in MEL cells leads to the inhibition of adult β_{maj}-globin gene expression, as well as a reduction in the recruitment of RNA polymerase II (RNA Pol II) to HS2 and to the β_{maj}-globin gene promoter (138). Additionally, overexpression of USF1 led to an increase in β_{maj}-globin gene expression. Overall, these data suggest that USF acts as an activator of the adult β-globin gene. The purpose of the following study was to examine the *in vivo* role of USF in erythropoiesis by generating transgenic mouse lines that express A-USF.

A-USF contains the USF heterodimerization domain but lacks the USF-specific region (USR), which is required for transcriptional activation. Additionally, the basic DNA binding region has been mutated to contain an acidic extension (132,199). Thus, any USF heterodimers that contain A-USF will be unable to bind or activate DNA. A-USF has been shown to successfully inhibit normal activity of both USF1 and USF2 (199-201).

The A-USF construct used to generate A-USF transgenic mice was designed to confer high level, erythroid specific expression of A-USF (Figure 2-1). The intention was to express A-USF exclusively or preferentially in erythroid cells to interfere with the function of both USF1 and USF2 without affecting the vital functions of these proteins in other tissues and organs.

Results

Three transgenic founders were generated with this construct (founders I, II, and III), although two of these founders did not transmit the transgene to their progeny (founders I and III). Reverse transcription PCR (RT-PCR) and immunoblot analyses showed that A-USF was expressed in all transgenic lines, but not in the liver of founder mouse I (Figure 3-1, panel B). Additionally, RT-PCR analysis of A-USF in the spleen of three F_1 females from founder II revealed that the expression of A-USF varied between littermates (Figure 3-1, panel A).

To analyze the effect of expressing A-USF on β_{maj}-globin gene expression, transgenic (founder I and three F_1 females from founder II) and four control wild-type (WT) littermates were treated with phenylhydrazine hydrochloride, which induces hemolytic anemia. Under anemic conditions, the mouse spleen becomes a major site of erythrocyte production, which increases the amount of nucleated red blood cells in the organ. It was found that β_{maj}-globin gene expression was reduced by 50% in the non-transmitting transgenic mouse (founder I) compared to expression in a wild-type control mouse (Figure 3-1, panel C). The expression of A-USF in the spleen of phenylhydrazine-treated F_1 animals from the transmitting line varied (Figure 3-1, panel B), and it resulted in a two- to fivefold decrease in β_{maj}-globin gene expression in three transgenic littermates (Figure 3-1, panel C) compared to that of three wild-type littermates. Five transgenic female mice but none of the wild-type mice died as a result of the phenylhydrazine treatment, indicating a possible defect in erythropoiesis. Consistent with this observation is the fact that 4-week-old transgenic female mice weighed 2 to 3 g less than wild-type littermates (23 ± 1 and 26 ± 1 g, respectively).

Next, the recruitment of RNA Pol II at the murine β_{maj}-globin gene in mice expressing A-USF was analyzed and compared to wild-type littermates, which do not express A-USF (Figure 3-1, panel D). It was found that the expression of A-USF in erythroid cells led to a reduction in USF2 and RNA Pol II binding to the β_{maj}-globin promoter. The recruitment of RNA Pol II to the control β-actin gene was not affected in A-USF-expressing mice (data not shown).

Because the founder of line II was male, but failed to yield transgenic male offspring, it was hypothesized that the A-USF transgene had integrated into the X chromosome, and that the expression of A-USF in all erythroid cells was not compatible with survival. Females from line II showed a somewhat variegated phenotype, likely because of differences in the silencing of the transgene on the X chromosome due to X chromosome inactivation. Since adult transgenic males were unable to be examined, male transgenic embryos were examined at different developmental stages instead. In two different 14.5 days *post coitum* (dpc) litters, which were obtained by mating a transgenic female (line II) with a wild-type male, several re-absorbed and pale embryos were detected. Genotyping with primers specific for both the A-USF transgenic construct and for the Y chromosome revealed that the re-absorbed embryos were transgenic males. Embryos at earlier stages—10.5, 11.5, and 12.5 dpc—were examined next. At 10.5 and 11.5 dpc, all embryos appeared to be alive and normally developed; however, several of the embryos were pale, and these were identified by PCR as transgenic males (Figure 3-2). At 12.5 dpc, the male transgenic embryos ceased to develop further, demonstrating that the male transgenic embryos did not survive beyond 11.5 dpc.

The expression of globin genes in 10.5 and 11.5 dpc embryos was examined next, and it was found that the expression of A-USF caused a reduction in the expression of all globin genes compared to that of wild-type littermates (Figure 3-3). In 10.5 dpc yolk sac samples, the expression of the embryonic εγ and β_{h1} genes as well as that of the Hba α1 and β_{min}-globin genes was reduced 5 to 10-fold, and in 11.5 dpc fetal liver samples there was a 5 to 10-fold reduction in the expression of the adult β_{maj}-globin gene. The effect of A-USF on the expression of other erythroid cell-specific genes was also examined, including those that encode for transcription factors that regulate erythropoiesis, such as GATA-1, Erythroid Krüppel-Like Factor (EKLF), TAL1, NF-E2 (p45), and HoxB4 (Figure 3-4). HoxB4, a homeobox transcription factor expressed in primitive hematopoietic stem cells, has previously been shown to be regulated by USF in K562 cells (202). We found that the expression of HoxB4 and GATA-1 is reduced by only about twofold in the yolk sac of transgenic males. In contrast, the expression of transcription factors EKLF, TAL1, and NF-E2 (p45) was reduced by 5 to 10-fold, suggesting that USF is required for the expression of these genes during primitive erythropoiesis (Figure 3-4). The expression of another well-characterized erythroid cell-specific gene, Band3, also was reduced by more than 5-fold in the transgenic yolk sac samples. Band3 is a major glycoprotein of the erythrocyte membrane, and plays a key role in the uptake of carbon dioxide by red blood cells by mediating the exchange of chloride and bicarbonate across the phospholipid bilayer (203). USF1 transcription was unaffected by the expression of A-USF (Figures 3-3 and 3-4).

To exclude the possibility that the phenotype observed in the male transgenic embryos was due to the integration of the transgene and subsequent disruption of a

specific cellular function, transient transgenic embryos were generated and analyzed at 11.5 dpc. In total, 4 transient transgenic embryos were generated, although only two of them expressed A-USF in the yolk sac, as determined by RT-PCR (Figure 3-4). Both embryos appeared pale and revealed reductions in the expression of Hba α1 and β-globin genes as well as that of EKLF and Band3 (Figure 3-4). The expression of USF1 was not affected in these mice. These data demonstrate that the expression of A-USF in erythroid cells of transgenic mice leads to consistent defects in erythropoiesis in multiple independent transgenic embryos. Therefore, the erythroid phenotype observed in the transmitting line (line II) is unlikely to be due to the disruption of gene expression patterns at the site of transgene integration.

To verify that the expression of A-USF affects the binding of USF in transgenic embryos, we examined the binding of USF1 to LCR element HS2 in yolk sac samples taken from 10.5 dpc transgenic embryos (line II) and wild-type litter mates using the μChIP assay, which allows the detection of protein-chromatin interactions with a small number of cells. The binding of USF1 to the LCR was reduced in transgenic embryos compared to that of wild-type littermates (Figure 3-5, panel A). The interaction of RNA Pol II with LCR element HS2 also was reduced in 10.5 dpc yolk sac samples from transgenic mice compared to that of littermates (Figure 3-5, panel B), whereas there was no change in the association of RNA Pol II with the GAPDH gene between wild-type and transgenic embryos (Figure 3-5, panel B).

Because EKLF, TAL1, and NF-E2 (p45) failed to be expressed at high levels in the hematopoietic tissue of transgenic mice, we examined the possibility that USF directly regulates these genes. We performed ChIP to examine the interaction of USF with the

gene loci encoding these transcription factors during the differentiation of MEL cells. One of the multiple DNA regulatory elements in the EKLF gene locus contains an E-box, which previously has been shown to interact with TAL1 (204,205). Both subunits of USF associated with the E-box-containing regulatory region of the EKLF gene in MEL cells (Figure 3-6). We also observed interactions of USF with the TAL1 gene locus, which also contains an E-box motif in a regulatory element. Interestingly, the interaction of USF with the TAL1 gene decreased during dimethylsulfoxide (DMSO)-induced MEL cell differentiation. USF binding also was detectable at the GATA-1 gene locus (Figure 3-6). There are no previous data concerning E-box elements regulating the GATA-1 gene. We failed to detect significant interactions of USF1 with the NF-E2 (p45) gene; however, the recovery of USF2-precipitated p45 gene fragments was higher than that of the IgG control. The data suggest that the TAL1 and EKLF genes are direct targets of both USF1 and USF2 in differentiating erythroid cells. There was no significant binding of USF to the control Necdin gene, which is not expressed in erythroid cells (Figure 3-6). We confirmed the interactions of USF2 with the erythroid cell-specific gene loci in primary erythroid cells taken from 16.5 dpc mouse fetal liver samples (Figure 3-7). The ChIP results demonstrated that USF2 interacts with the EKLF, GATA-1, and TAL1 gene loci but not with the Necdin gene locus. We observed a reproducible interaction of USF2 with the NF-E2 (p45) gene locus in fetal liver cells.

We next examined the possibility that the expression of A-USF in erythroid cells impairs their differentiation potential. We began these studies by examining 10.5 dpc yolk sac cells from transgenic embryos and wild-type littermates for the expression of the transferrin receptor CD71, which is expressed at high levels in developing erythroid

cells and serves as a marker for erythroid progenitors (198). The CD71-mediated sorting of yolk sac cells revealed that the number of $CD71^+$ cells was about threefold lower in the transgenic embryos than that of the wild-type embryos (Figure 3-8). Furthermore, we observed a decrease in the number of cells that express high levels of Ter-119 (Figure 3-8), which is a marker for more differentiated erythroid cells (206). The number of benzidine-positive cells also was reduced by three to fourfold in the yolk sac cells from transgenic embryos compared to those taken from wild-type littermates (data not shown). Taken together, these results demonstrate that USF is an important contributor to erythroid cell differentiation and mediates the high-level expression of erythroid transcription factors and the expression of the globin genes.

To examine whether USF not only regulates the differentiation of erythroid cells but also functions within the context of differentiating cells, we analyzed the expression of the β_{h1}-globin gene in Ter-119-sorted cells obtained from transgenic or wild-type embryos (Figure 3-8). The expression of the β_{h1}-globin gene was reduced in Ter-119^+ cells isolated from two transgenic embryos compared to that of their wild-type littermates.

Discussion

It currently is unknown how ubiquitously expressed and tissue-specific transcription factors coordinate the activation of highly expressed genes during differentiation. Perhaps tissue-specific factors mediate the accessibility of regulatory sites, whereas ubiquitously expressed proteins perform basic functions involved in the local remodeling of nucleosomes and the recruitment of transcription complexes.

USF was one of the first transcription factors shown to activate transcription mediated by RNA Pol II, and it plays a role in the high-level expression of many genes

in differentiated cells (130,139). Accumulating evidence points to the possibility that highly expressed genes are transcribed in specialized nuclear domains enriched for splicing factors and RNA Pol II, often referred to as transcription factories or transcription domains (207-209). It is possible that the LCR nucleates such a transcription domain in erythroid cells (184). USF is a likely candidate protein that could mediate the association of genes or regulatory elements with transcription domains in the nucleus. USF mediates the high-level expression of genes during cellular differentiation, and the global analysis of the interaction of USF with chromatin revealed that USF mostly binds to regions close to transcription start sites (141).

The data presented here suggest that USF regulates many genes involved in erythropoiesis, including genes encoding key erythroid transcription factors. Inactivating USF thus causes a defect in the differentiation of erythroid cells. This is supported by our observation that the expression of A-USF in transgenic mice causes reductions in the number of $CD71^+$ and $Ter\text{-}119^+$ cells in the yolk sac (Figure 3-8). In addition to regulating erythropoiesis, several lines of evidence suggest that USF also directly regulates the recruitment of transcription complexes to the β-globin gene locus. First, both USF1 and USF2 interact with LCR element HS2 and with the adult β-globin gene promoter in vitro and in the context of intact erythroid cells (113,138,142,143). Electrophoretic mobility shift assays (EMSA) using protein extracts from erythroid cells demonstrated that a single complex is formed on the E-box derived from the adult β-globin downstream promoter region (113). This complex is super-shifted with USF antibodies, which suggests that no other helix-loop-helix (HLH) protein present in the protein extract is capable of interacting with this site in vitro. The expression of a

dominant-negative mutant of USF in MEL cells or transgenic mice reduces the recruitment of USF and RNA Pol II to the β-globin gene locus (138,210). Furthermore, it has also been demonstrated that A-USF reduces the recruitment of RNA Pol II to immobilized LCR templates in vitro (211). Finally, a reduction in globin gene expression in Ter-119^{+} cells isolated from A-USF-expressing transgenic embryos was observed (Figure 3-8). All of these data are consistent with the hypothesis that USF plays a direct role in the regulation of erythroid-specific genes, including the globin genes. However, USF does not appear to act globally on transcription, as no changes in the expression of housekeeping genes in the A-USF-transgenic mice were observed.

The reduction of embryonic globin gene expression in A-USF-expressing transgenic mice could be due to reduced interactions of USF and transcription complexes with the LCR, which may impair its activity. A reduction in the expression levels of EKLF, TAL1, and NF-E2 (p45) was observed, whereas the expression of GATA-1 and HoxB4 was only mildly reduced. It is likely that the reduced expression of these key erythroid transcription factors also contributes to the decreased globin gene expression observed in the yolk sac of A-USF transgenic mice. USF appears to interact with E-box-containing regulatory elements in the EKLF, GATA-1, and TAL1 gene loci. USF may function within the context of erythroid-specific transcription domains in the nucleus, and that genes expressed during erythropoiesis associate with these domains, which is consistent with data from Osborne et al. (69).

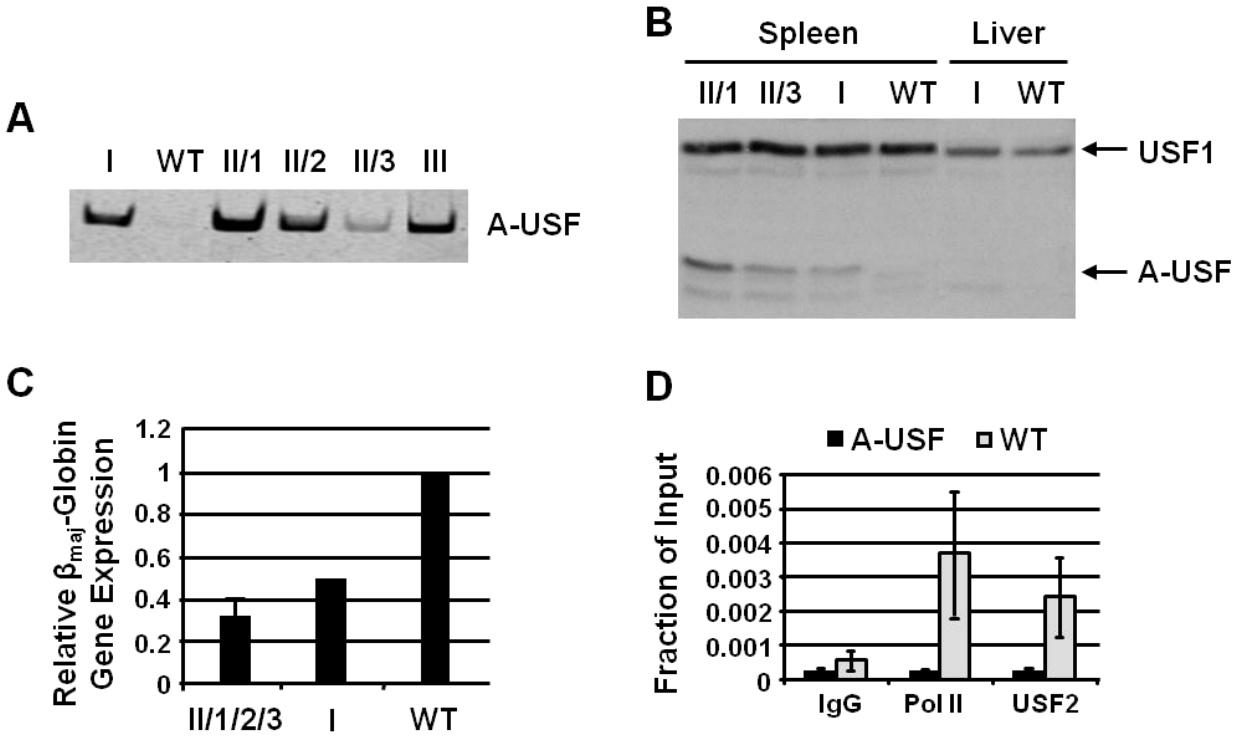

Figure 3-1. Analysis of mice expressing A-USF. A) SYBR green stain of the RT-PCR analysis of A-USF expression in transgenic (founders I and III and line II F_1 littermates 1 to 3 [II/1 to II/3]) and wild-type (WT) mice. RNA was isolated from the spleens of phenylhydrazine-treated mice, reverse transcribed, subjected to PCR analysis with primers specific to the A-USF coding region, and electrophoresed in 5% TBE gels. B) Immunoblot analysis of A-USF expression in transgenic or wild-type mice. Protein was isolated from the spleen or liver of phenylhydrazine-treated mice and subjected to immunoblot analysis using an antibody against USF1, which also detects A-USF. C) qRT-PCR analysis of β_{maj}-globin gene expression in spleens of A-USF transgenic line II F_1 littermates, A-USF founder mouse I, and wild-type mice. Data from the three line II F_1 littermates were combined and are designated II/1/2/3. GAPDH was used as a loading control, and results from samples were normalized to those of the wild type. D) q-PCR ChIP analysis of RNA Pol II and USF2 interactions with the β_{maj}-globin gene promoter control mice (WT) and transgenic mice (A-USF). Spleens taken from two phenylhydrazine-treated F_1 females (derived from line II) or wild-type mice were homogenized and subjected to ChIP analysis using antibodies against IgG, RNA Pol II, or USF2. Error bars reflect standard deviations from two independent experiments. These data were generated by Babak Moghimi.

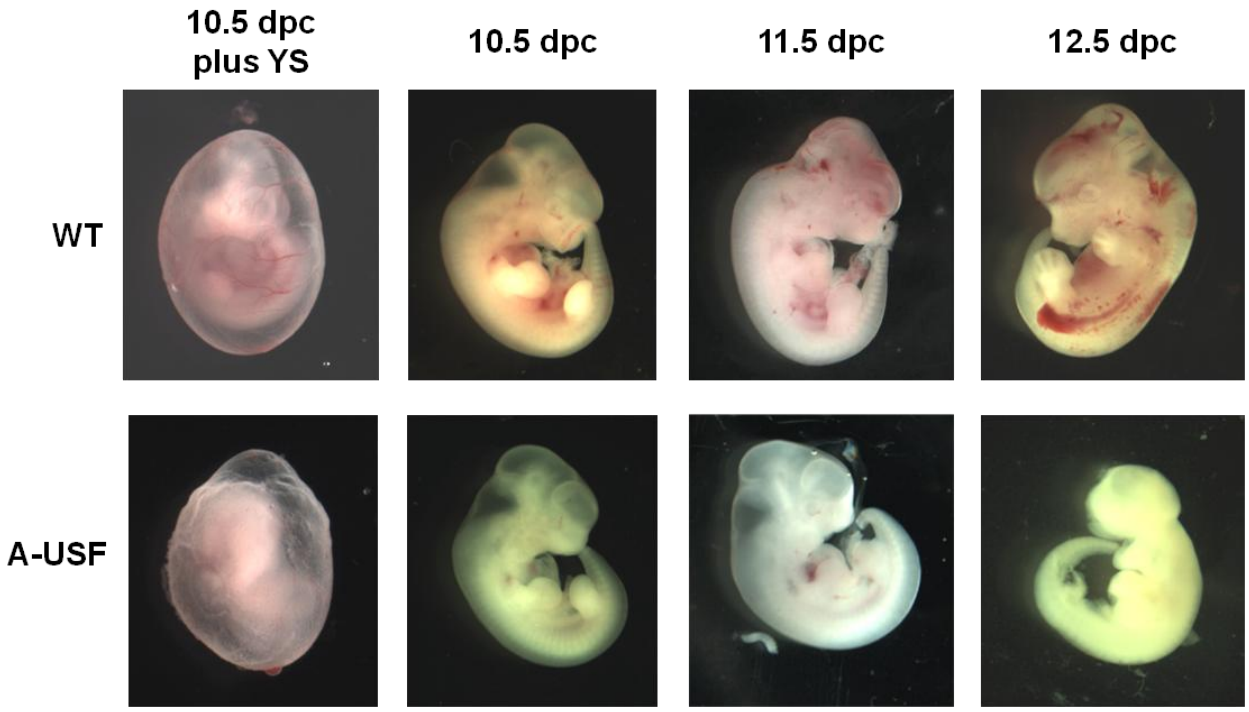

Figure 3-2. Analysis of transgenic mouse embryos at different stages of development. Male embryos were isolated at the indicated time points of development from A-USF transgenic females (F_1 females from line II) mated with wild-type (WT) males. Embryos were placed in a culture dish with PBS either in the presence or absence of the yolk sac (YS) and photographed using a Leica MZ16F4 instrument and the Qcapture program. Embryos were genotyped for sex and determined to be wild-type or transgenic (A-USF).

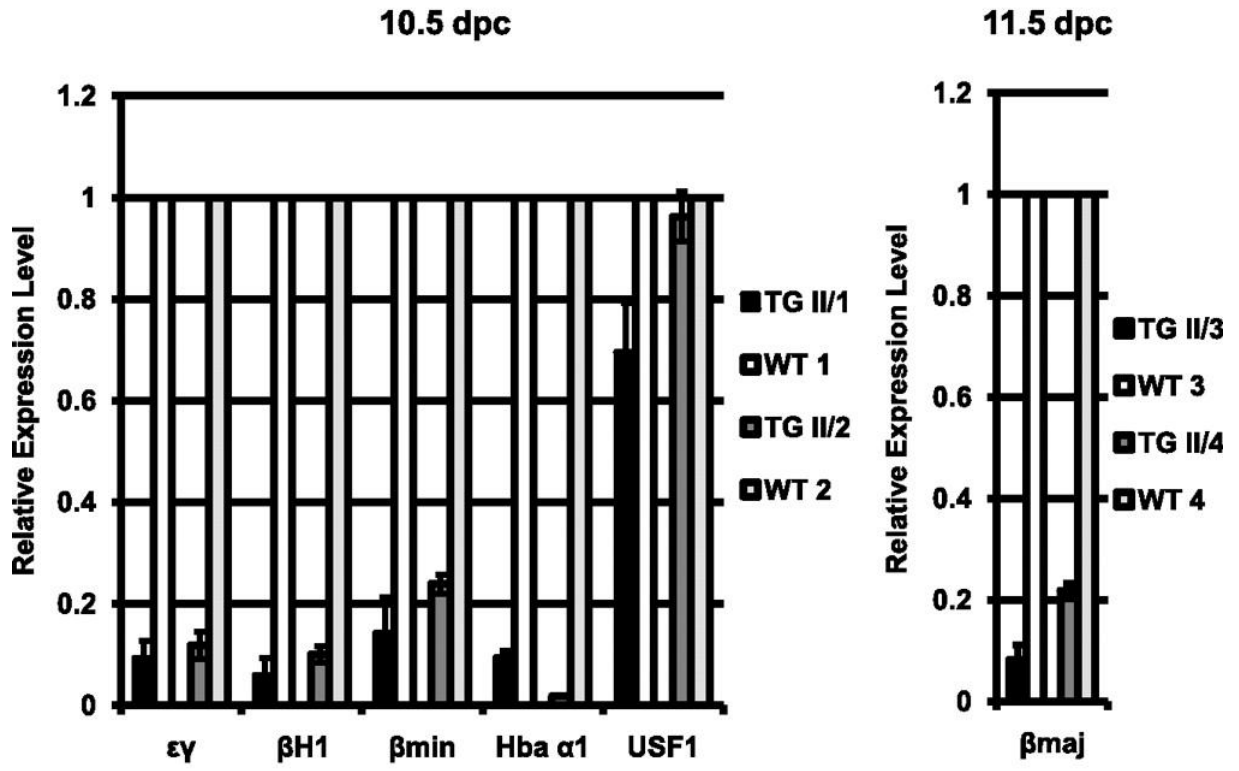

Figure 3-3. Effects of A-USF expression on the expression of erythroid genes in embryonic yolk sac cells. RNA was extracted from 10.5- or 11.5-dpc embryos, reverse transcribed, and subjected to qRT-PCR performed in triplicate. qRT-PCR analysis of εγ-globin, β_{h1}-globin, β_{min}-globin, Hba α1, USF1 (left; 10.5 dpc), and β_{maj}-globin (right; 11.5 dpc) gene expression in A-USF transgenic (TG II/1 to II/4) and wild-type (WT 1 to 4) mouse embryos. Two sets of four embryos, each containing two TG and two WT animals, were examined. GAPDH was used as an internal control, and sample data were normalized to those for a respective WT littermate. Data are represented as means ± standard errors of the means of at least three PCRs on each sample.

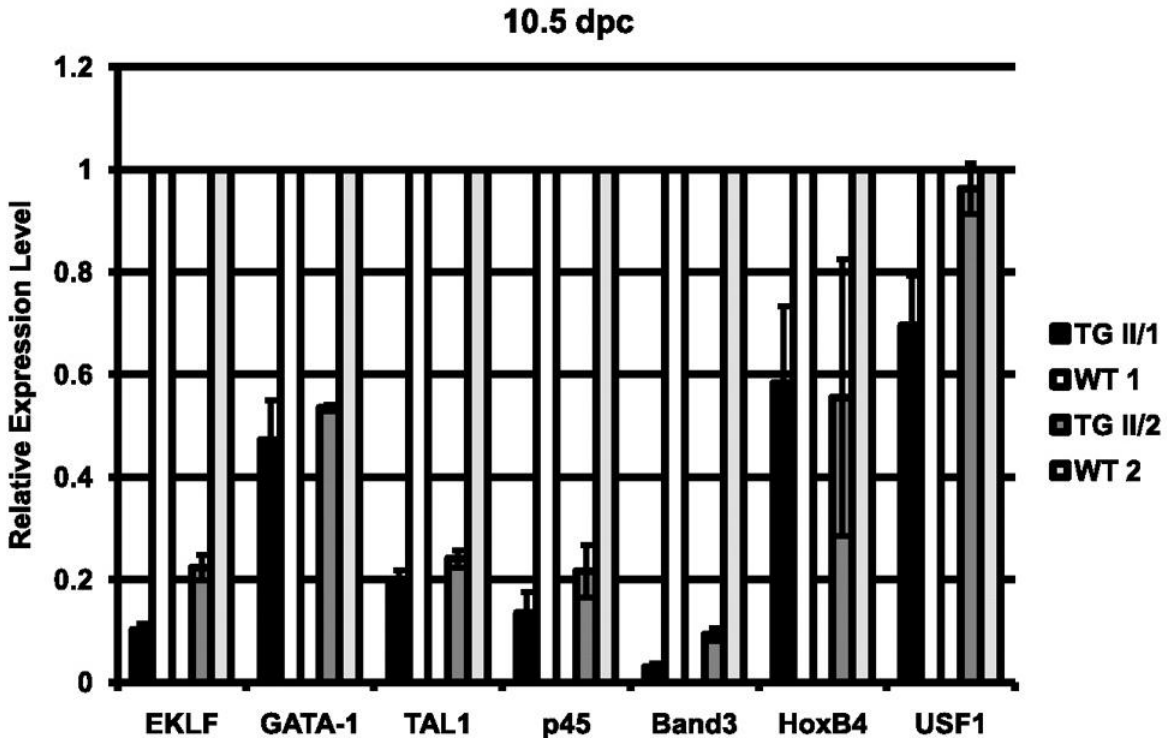

Figure 3-4. Effects of A-USF expression on the expression of erythroid cell-specific transcription factors in 10.5 dpc embryonic yolk sac cells. qRT-PCR analysis of EKLF, GATA-1, Tal-1, p45, Band3, HoxB4, and USF1 gene expression in the yolk sac of the transgenic (TG II/1 and II/2) and wild-type (WT 1 and 2) 10.5-dpc embryos examined in Figure 3-3. Data are presented as described for Figure 3-3.

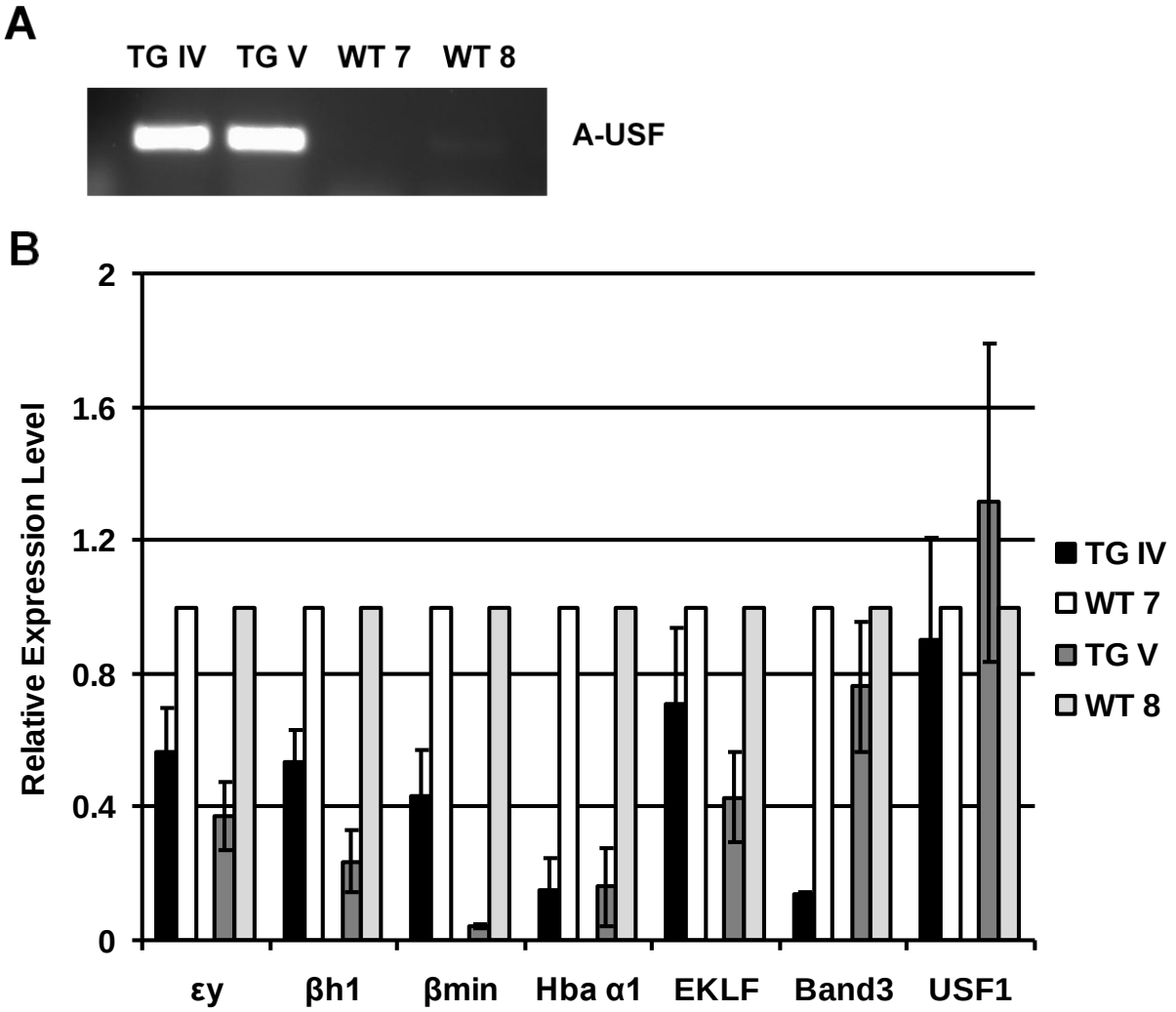

Figure 3-4. Generation and analysis of 11.5 dpc transient transgenic mouse embryos expressing A-USF. Fertilized oocytes were injected with the A-USF expression construct and implanted into the uterus of a pseudopregnant foster mother. Embryos (11.5 dpc) were isolated and subjected to DNA (embryo) and RNA (yolk sac) extraction. A) cDNA from the embryos was analyzed by RT-PCR using primers specific for the A-USF transgene to verify A-USF expression. All four embryos, two transgenic (TG IV and TG V) and two wild-type (WT 7 and WT 8) embryos, were taken from the same litter. B) RNA was subjected to qRT-PCR performed in triplicate for the analysis of εγ, β_{h1}, β_{min}, Hba α1, EKLF, Band3, and USF1 gene expression. Data were analyzed and are represented as described in the legend of Figure 3-3.

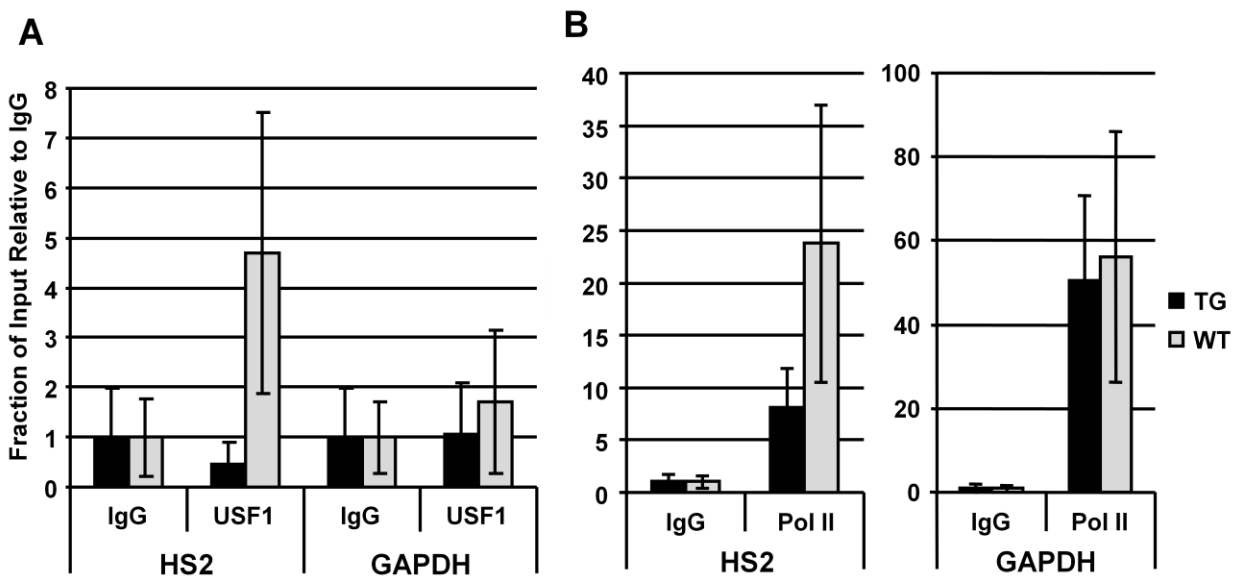

Figure 3-5. µChIP analysis of RNA Pol II and USF1 association with LCR element HS2 and the GAPDH gene in the yolk sac of wild-type and A-USF transgenic embryos. Embryos (10.5 dpc) were taken from an A-USF transgenic female mated to a wild-type male. Yolk sacs were isolated and subjected to µChIP analysis. A) µChIP was performed with antibodies against negative control IgG and USF1. DNA was analyzed by qPCR using primers specific for LCR element HS2 as well as for the control GAPDH gene, as indicated. B) µChIP was performed with antibodies against the negative control IgG and RNA Pol II. The DNA was analyzed by qPCR using primers specific for LCR element HS2 as well as for the control GAPDH gene, as indicated. Data were normalized to those of IgG and are represented as means ± standard errors of the means of three independent µChIP experiments with qPCRs performed in triplicate.

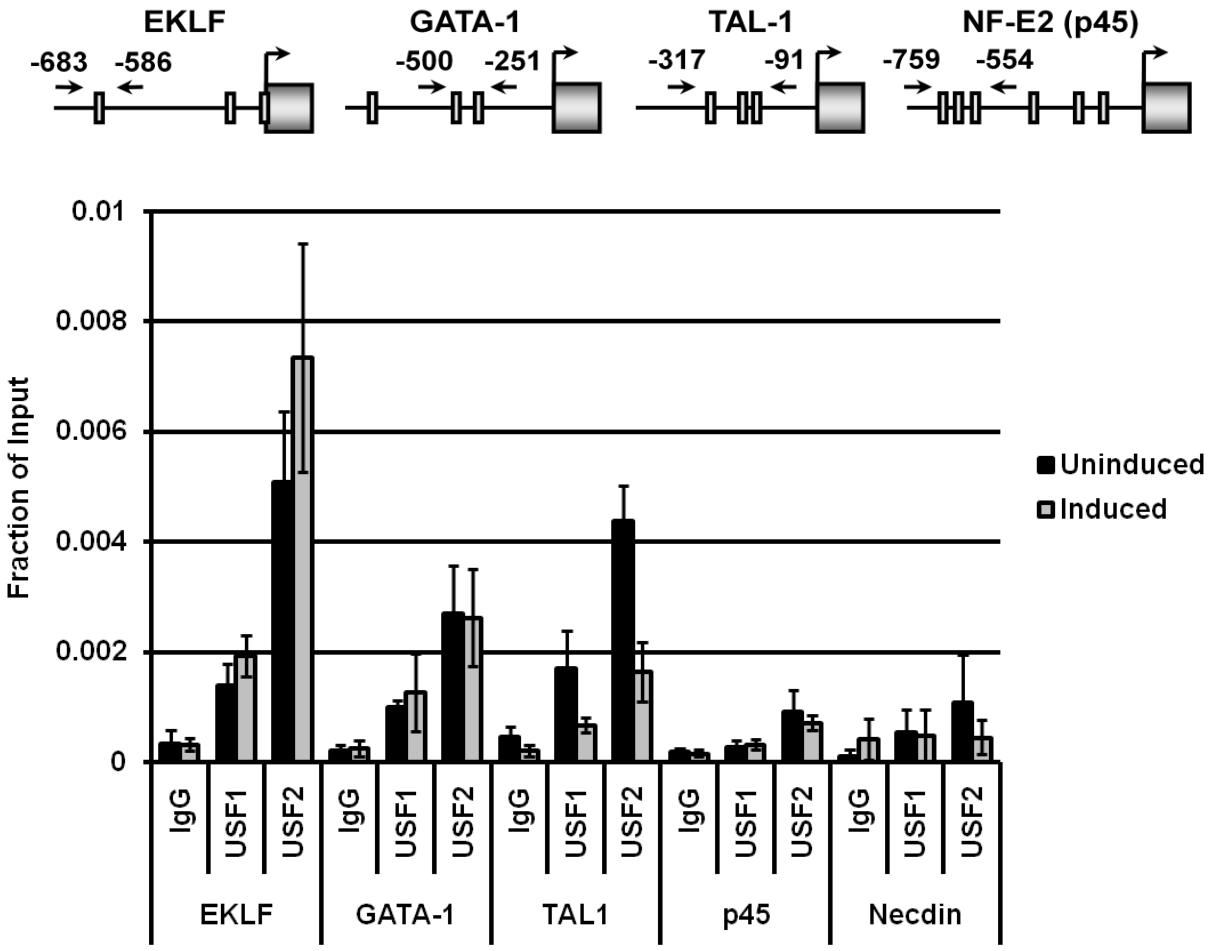

Figure 3-6. Interaction of USF with regulatory elements of genes encoding hematopoietic-specific transcription factors in MEL cells. ChIP analysis of the interaction of USF1 and USF2 with regulatory elements of the EKLF, GATA-1, TAL1, and NF-E2 (p45) genes as well as with the Necdin promoter serving as a negative control. The diagrams at the top indicate the position of E-boxes with respect to the transcription start site of each individual gene, with arrows indicating the location of primers used to amplify each region. ChIP was performed on uninduced or induced MEL cells. Cells were induced to differentiate for 3 days in the presence of 1.5% DMSO. Uninduced and induced cells were incubated with 1% formaldehyde. After being quenched with 125 mM glycine, the cells were lysed and chromatin was fragmented by sonication prior to precipitation with antibodies specific for IgG, USF1, or USF2. The isolated DNA was analyzed by qPCR with primers specific for the EKLF, GATA-1, TAL1, NF-E2 (p45), and Necdin gene promoters, as indicated. Results are represented as means ± standard errors of the means of three independent experiments, with each PCR performed in duplicate.

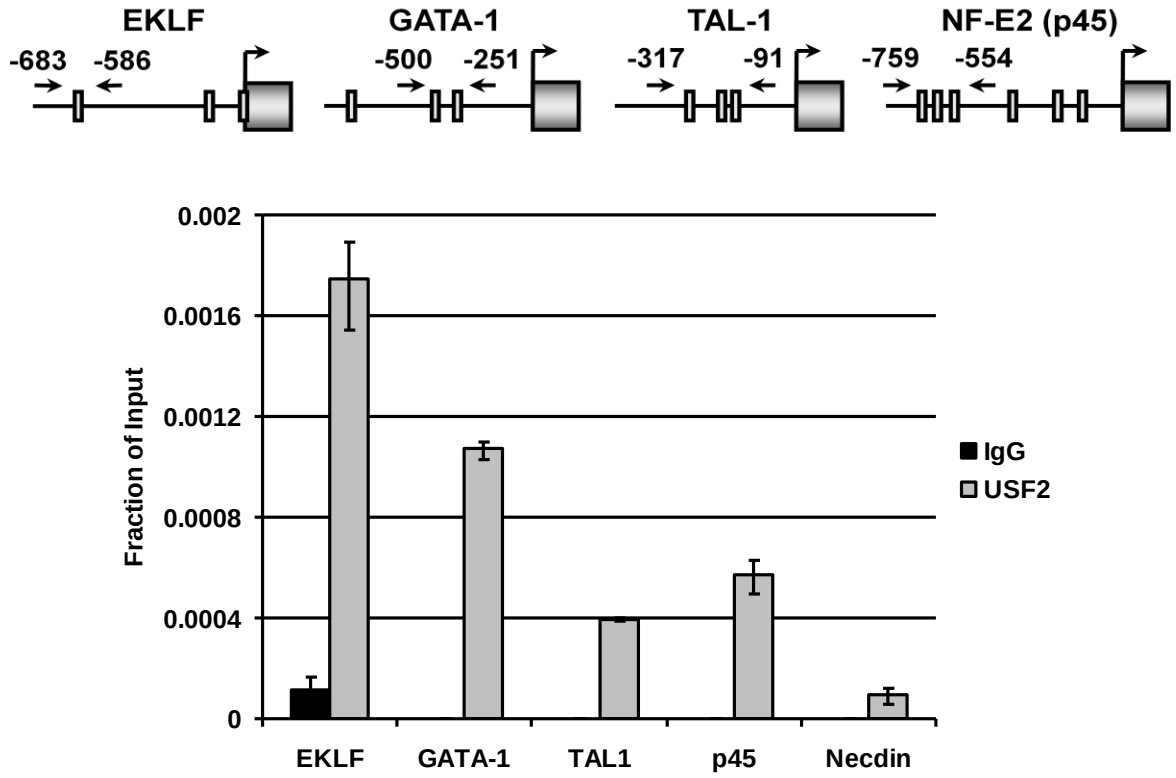

Figure 3-7. Interaction of USF with regulatory elements of genes encoding hematopoietic-specific transcription factors in fetal liver cells. ChIP was performed on 16.5 dpc liver cells, examining the interaction of USF2 with the same regions examined in Figure 3-6. Results are represented as means ± standard errors of the means of two independent experiments with PCRs performed in duplicate.

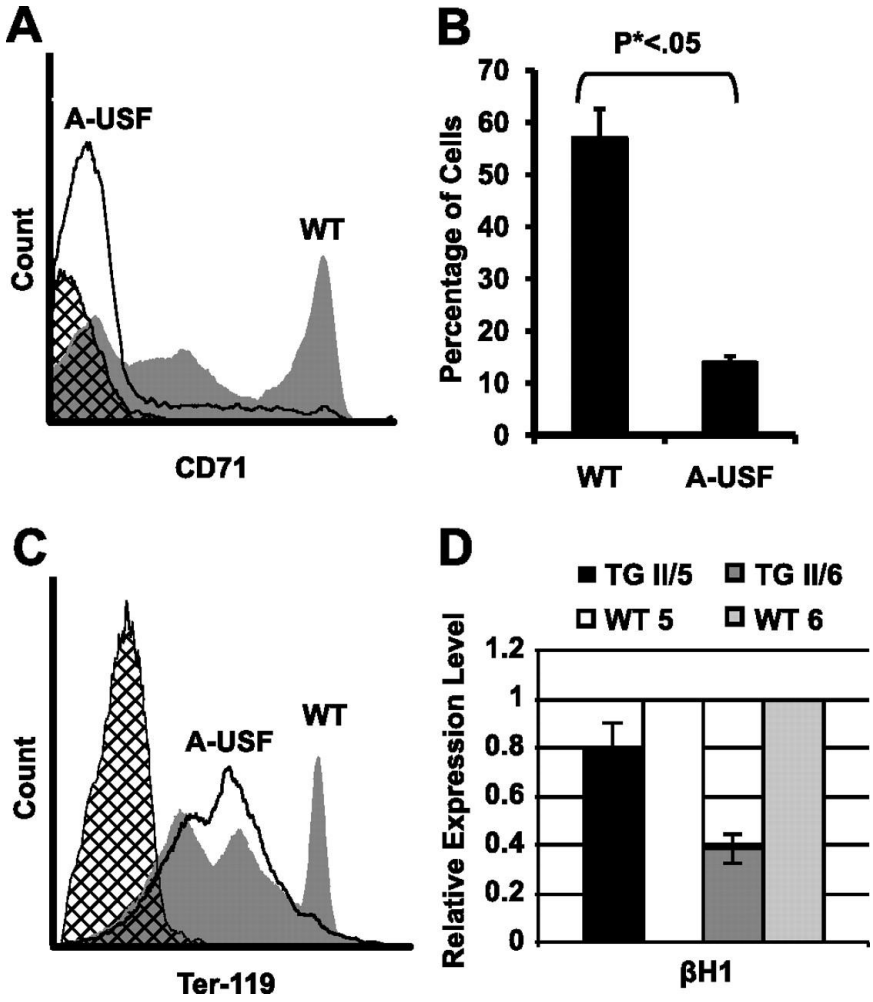

Figure 3-8. Transgenic A-USF embryos reveal a reduction in the number of $CD71^+$ and $Ter\text{-}119^+$ erythroid cells. Yolk sac cells from 10.5 dpc male embryos were isolated and subjected to flow cytometry using antibodies against CD71 or Ter-119. Hatched areas indicate unstained yolk sac cells analyzed separately. Solid lines represent the analysis of cells from A-USF-expressing transgenic embryos, while shaded gray areas represent cells from wild-type embryos. A) FACS analysis using antibodies against CD71. B) Number of CD71-positive cells in the 10.5 dpc yolk sac of wild-type (WT) and A-USF-expressing (A-USF) embryos. C) FACS analysis using antibodies against Ter-119. D) β_{h1} gene expression in $Ter\text{-}119^+$ embryonic yolk sac cells. Yolk sac cells from 10.5 dpc embryos were sorted using Ter-119 antibodies, and a subset of $Ter\text{-}119^+$ cells was collected and subjected to RNA extraction and analysis. Data were analyzed as described in the legend to Figure 3-3 and are represented as the means ± standard errors of the means of two qRT-PCRs performed in duplicate. Data from panels A, B, and C were generated by Babak Moghimi.

CHAPTER 4
THE ASSOCIATION OF USF AND BRG1

Introduction

Because BRG1$^{-/-}$ (Brahma-Related Gene 1) knockout mice do not survive to birth, a hypomorphic mutation of BRG1 (BRG1^{ENU1}) was generated and identified in a mutant mouse screen in which mice were subjected to *N*-ethyl-*N*-nitrosourea (ENU)-induced mutagenesis. In the BRG1 mutant mice, a single amino acid in a highly conserved region of the catalytic adenosine triphosphatase (ATPase) domain was mutated from a negatively charged glutamic acid to a nonpolar glycine (E1083G) (170). The partial loss-of-function BRG1^{ENU1} is stable and able to assemble into SWI/SNF-related complexes, but has diminished nucleosome remodeling capabilities; it is unable to establish DNase I hypersensitive sites.

Mice carrying a single copy of the hypomorphic allele (BRG1$^{ENU1/-}$) develop normally until mid-gestation, but begin to exhibit defects in the erythroid lineage at ~11.5 dpc. Ultimately, all mutant embryos die by ~14.5 days *post coitum* (dpc). Although BRG1 is still recruited to the β-globin locus in these mice, chromatin remodeling, transcription, and epigenetic marks such as histone acetylation and DNA methylation are all affected (170). Like the dominant-negative Upstream Stimulatory Factor (A-USF) transgenic male embryos, these BRG1$^{ENU1/-}$ embryos are smaller than their wild-type littermates and exhibit signs of anemia (Figure 4-1).

Additionally, it was recently demonstrated that BRG1 functions to help establish specific chromatin loops in the β-globin locus in G1E-ER4 cells (212). G1E-ER4 cells are GATA-1-null erythroid progenitor cells that express a conditional GATA-1 under control of estrogen receptor ligand binding domain. The pattern of BRG1 occupancy at

both HS2 and the β_{maj} promoter observed upon activation of β_{maj} in G1E-ER4 cells is similar to what has been previously observed with USF in studies using erythroid differentiation of murine embyronic stem (ES) cells (138). Thus, both USF and BRG1 were observed to bind to HS2 and the β_{maj} promoter prior to the binding of RNA polymerase II (RNA Pol II). Due to the similarities in the embryonic phenotypes of both the $BRG1^{ENU1/-}$ mice as well as the transgenic male A-USF mice, it was hypothesized that USF and BRG1 may be associated in a common protein complex and function together to prepare the locus for the recruitment of RNA Pol II. Additionally, the role of USF in loop formation between the murine β-globin locus control region (LCR) and the adult β_{maj}-globin gene was examined as well.

Results

Chromatin immunoprecipitation (ChIP) was performed on uninduced and induced (1.5% dimethylsulfoxide [DMSO], 72 h) murine erythroleukemia (MEL) cells in order to compare the results to the BRG1 binding pattern that was originally observed in G1E-ER4 cells by Kim, et al. (212). G1E-ER4 and MEL cells are both representative of erythroid progenitor cells, and it was anticipated that the MEL ChIP would produce a similar result. Surprisingly, binding of BRG1 to HS2 and β_{maj}-globin in MEL cells appeared to decrease after the 72 h induction when compared to BRG1 binding levels in uninduced cells (Figure 4-2). Based on this result, it was suspected that BRG1 is recruited to the locus early on during induction, but that it may not be required for the continued association of RNA Pol II to the locus. In addition, it was found that RNA Pol II appears to be recruited to the LCR prior to TATA-binding protein (TBP) binding, although low levels of TBP are present at the β_{maj}-globin promoter before induction (Figure 4-2). After induction, TBP levels increase at both HS2 and at the β_{maj}-globin

promoter. The presence of RNA Pol II at HS2 prior to induction would suggest that binding of RNA Pol II to the LCR is independent of TBP, and that other factors, such as BRG1 and CBP, may be involved in its recruitment.

In order to examine the recruitment of BRG1 to the murine β-globin gene locus throughout the duration of the normal 3-day induction period, ChIP was performed on MEL cells incubated with 1.5% DMSO for 24 h, 48 h, and 72 h. As expected, levels of BRG1 binding at both HS2 and the adult β_{maj}-globin gene promoter gradually decreased over each day of the induction, with the highest levels of BRG1 binding occurring 24 h after the addition of DMSO (Figure 4-3). Although this is in agreement with the previous observation of a decrease in BRG1 binding at 72 h after the addition of DMSO, this ChIP has only been performed once and needs to be repeated for confirmation.

In order to determine whether USF and BRG1 are part of a common protein complex, Co-IP was performed on nuclear extract from uninduced, 1-day induced (2% DMSO, 24 h), and 3-day induced (1.5% DMSO, 72 h) MEL cells. In the induced MEL cells incubated with 2% DMSO for 24 h, a band of the expected size of BRG1 (205 kDa) was detected in both the input lane as well as the USF2-pulldown lane (Figure 4-4). These data show that BRG1 interacts with USF2, although the specific nature of this interaction remains uncharacterized. No interaction between BRG1 and USF1 was detected, however. Interestingly, in the induced MEL cells incubated with 1.5% DMSO for 72 h, no interaction between either USF1 or USF2 with BRG1 was detected (data not shown). Based on these results, it is believed that USF2 recruits BRG1 to the globin gene locus early on during induction. This is consistent with data generated by Dr. Zhuo Zhou, in which USF2 is more efficiently recruited to HS2 in undifferentiated MEL cells

compared to USF1 (211). While it could be argued that the lack of detection of BRG1 observed in the USF1 pull-down lane could be due to poor antibody quality, it is unlikely, as the same antibody has been shown to successfully pull down USF1/USF2 complexes in other co-IP studies (211).

Although it is known that the LCR comes into close proximity with the globin genes after DMSO-mediated erythroid induction of MEL cells, it is not known exactly what protein or protein complexes facilitate this interaction. In 2007, Song et al. demonstrated by using chromatin immunoprecipitation-chromosome conformation capture (ChIP-3C) that NLI/Ldb1 is required for loop formation and also that it is a part of the protein scaffold which brings the LCR in close proximity with the β_{maj}-globin gene (160). However, NLI/Ldb1 itself is not a DNA binding protein, and this interaction would require a multi-protein complex to facilitate the interaction between NLI/Ldb1 and the two distal regions of DNA. Because there are binding sites for USF in both the LCR as well as at the adult β-globin gene promoter, and USF binding at both loci increases upon DMSO-mediated erythroid induction in MEL cells, it was hypothesized that USF could be a member of this multi-protein scaffold connecting HS2 to the β_{maj}-globin gene (211). In order to determine USF's role in loop formation, ChIP-3C was performed on uninduced and induced MEL cells, using a modified protocol based on the ChIP-3C technique described in Song, et al. (160). Unfortunately, multiple attempts at the ChIP-3C procedure using antibodies against both USF1 and USF2 yielded no ligation products between HS2 and the εγ, β_{h1}, or β_{maj}-globin gene (data not shown). However, a ligation product between HS2 and the β_{maj}-globin gene was detected when using an antibody against di-methylated H3K4 (Figure 4-5). This band was later subjected to gel extraction

and sequenced for verification, and although the band from this ligation product was positively identified as the correct product, further attempts to reproduce this result did not succeed.

Discussion

The results presented here demonstrate a novel role for the ubiquitously-expressed USF in chromatin remodeling through interaction with the BRG1 subunit of SWI/SNF. USF has previously been implicated in chromatin remodeling, when it was shown that USF interacts with a chromatin boundary element in the chicken β-globin gene locus and recruits the co-regulators p300 and P300/CBP-associated factor (PCAF), which both harbor histone acetyltransferase (HAT) activity (145). The chromatin-modifying enzymes recruited by USF could establish an open and accessible chromatin region, which would counteract the spread of heterochromatin. USF could also play a similar role in establishing accessible chromatin in the LCR and the adult β-globin gene promoter. Other data has shown that the association of p300 with LCR element HS2 and the β_{maj}-globin promoter, both of which associate with USF, is reduced in cells expressing A-USF (90,113). Additionally, USF2, but not USF1, has been shown to interact with CBP in undifferentiated cells (211). It is believed that USF2 interacts with CBP and BRG1 to establish an open chromatin configuration (Figure 4-6). Once this open chromatin is established, however, BRG1 dissociates from the complex, but USF2 and CBP remain bound, possibly for the purposes of recruiting RNA Pol II to the locus. This would be in agreement with the ChIP data, which indicate that RNA Pol II is present at HS2 prior to induction. As MEL cells are representative of an erythroid progenitor cell line, the β-globin locus may be “poised” for expression prior to the DMSO induction, which leads to terminal erythroid differentiation.

Although it could not be conclusively determined whether or not USF plays a role in loop formation between the LCR and the globin genes, the data presented here do not exclude the possibility. Future investigation using the ChIP-3C assay will require more experimentation and controls. Using antibody against Ldb1 in the ChIP-3C assay could serve as a positive control, as it has previously been shown to be a part of the loop complex (160). Additional controls include performing the 3C assay in a non-erythroid cell line, no restriction enzyme digestion, no ligation, and no antibody.

In cell cycle studies using K562 cells, USF was also observed to bind to the LCR and globin genes prior to binding of RNA Pol II (90). The next chapter describes further studies on the conformation of the globin locus, with an emphasis on changes that occur in binding of USF or RNA Pol II to the locus during different stages of the cell cycle.

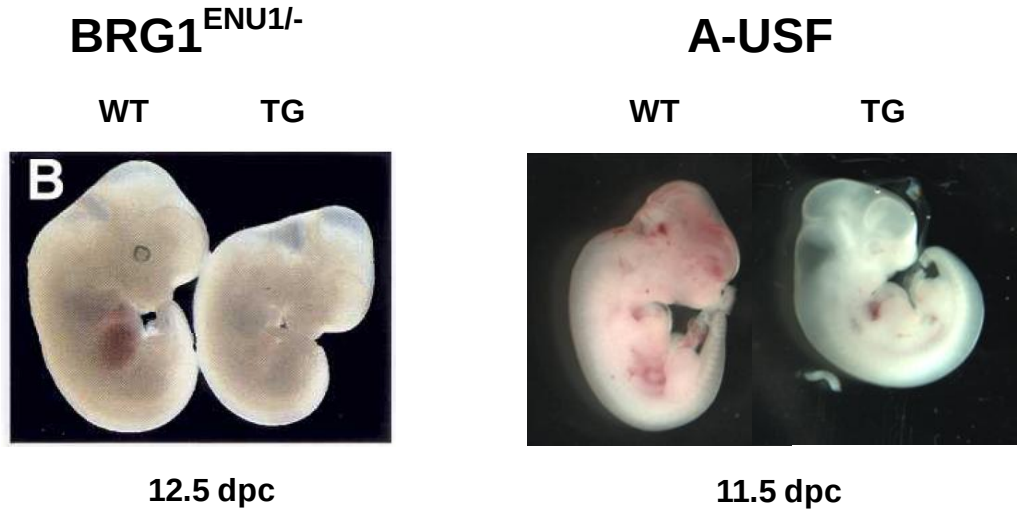

Figure 4-1. Comparison of hypomorphic BRG1$^{ENU1/-}$ 12.5 dpc mutant embryos with A-USF 11.5 dpc transgenic male embryos (170). Note that the transgenic embryos are pale, which is indicative of anemia, and are smaller in size than their wild-type littermates.

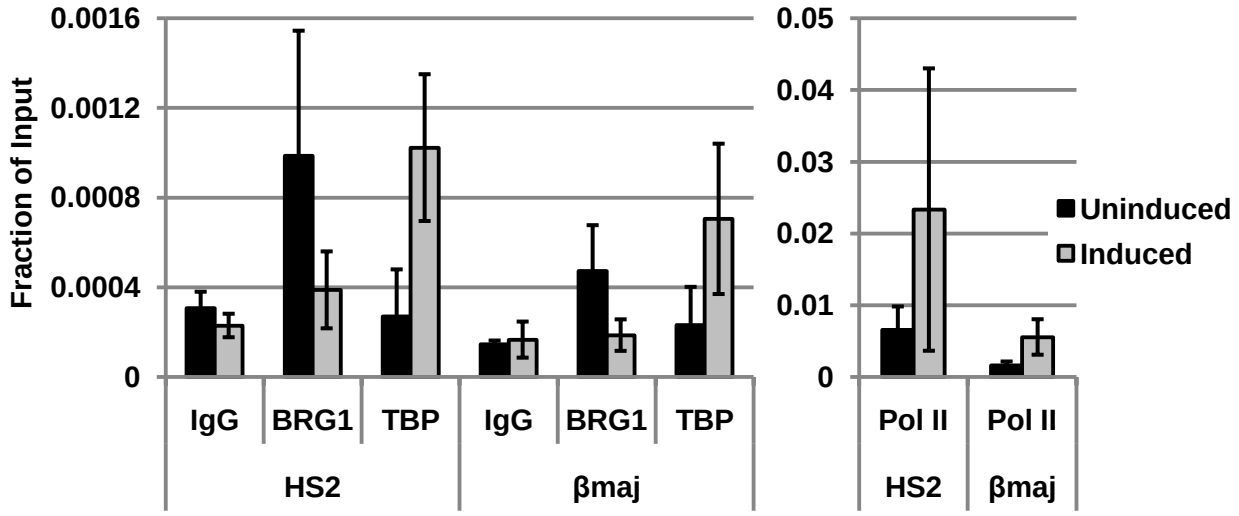

Figure 4-2. Interaction of BRG1, TBP, and RNA Pol II with the murine β-globin locus in uninduced and induced (1.5% DMSO, 72 h) MEL cells. Crosslinked cells were sonicated, and the cell extracts were subjected to immunoprecipitation with antibodies specific for IgG, rabbit polyclonal BRG1, TBP, and RNA Pol II. The isolated DNA was analyzed by qPCR with primers specific for HS2 and the β_{maj}-globin gene promoter. Results are normalized to IgG levels and represented as means ± standard errors of the means of at least three independent experiments, with each PCR performed in duplicate.

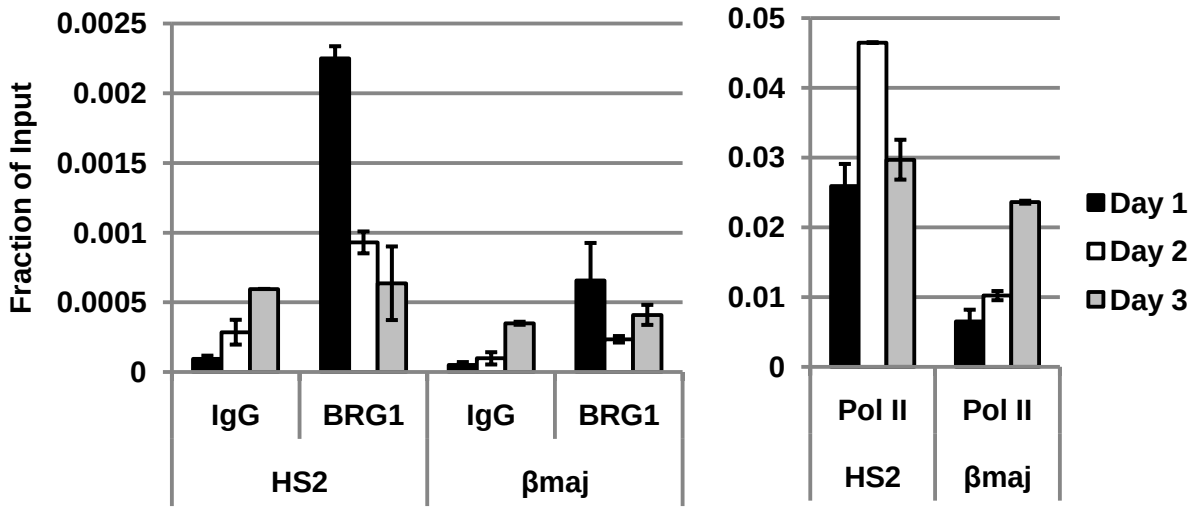

Figure 4-3. Interaction of BRG1 with the murine β-globin locus in uninduced and induced (1.5% DMSO; 24, 48, or 72 h) MEL cells. ChIP was performed on uninduced or induced MEL cells. Crosslinked cells were sonicated, and the cell extracts were subjected to immunoprecipitation with antibodies specific for IgG and rabbit polyclonal BRG1. The isolated DNA was analyzed by qPCR with primers specific for HS2 and the β_{maj}-globin gene promoter. Results are represented as means ± standard deviation of a single ChIP experiment with qPCR performed in duplicate.

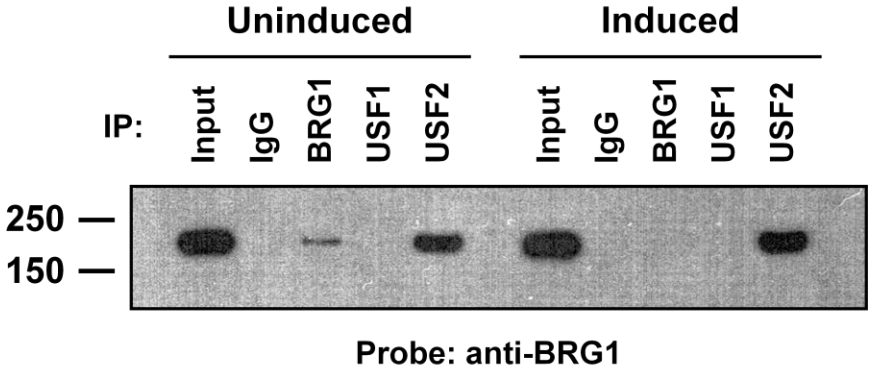

Figure 4-4. Immunoblot analysis of Co-IP on nuclear extract from uninduced and induced (2% DMSO, 24 h) MEL cells. Protein complexes were isolated from nuclear extract using antibodies against rabbit IgG, mouse monoclonal BRG1, USF1, and USF2, while the membrane was probed using mouse polyclonal antibody against BRG1.

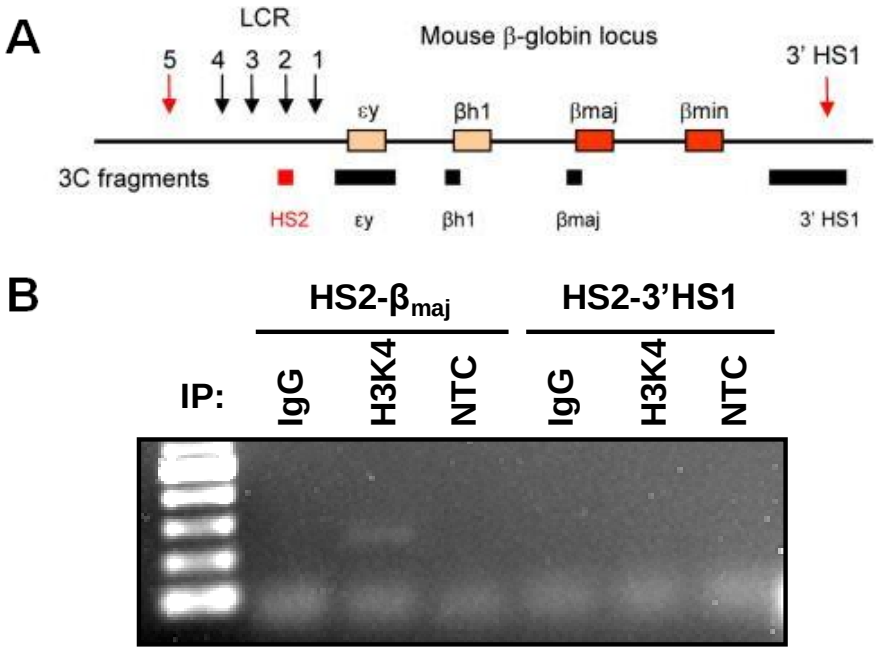

Figure 4-5. Analysis of ChIP-3C ligation products from induced (1.5% DMSO, 72 h) MEL cells. A) Schematic of the murine β-globin gene locus; red or black bars below indicate Hind III digestion fragments within which the primers bind (160). B) Agarose gel analysis of ligation products using primers for the regions indicated at the top. HS2 served as an anchor primer to either the β_{maj}-globin gene or the 3' HS1. Protein-DNA complexes were pulled down with the antibodies listed below, and eluted DNA was subjected to PCR. The band observed in HS2-β_{maj} using antibody against di-methylated H3K4 was excised and sequenced for verification. NTC indicates no template control.

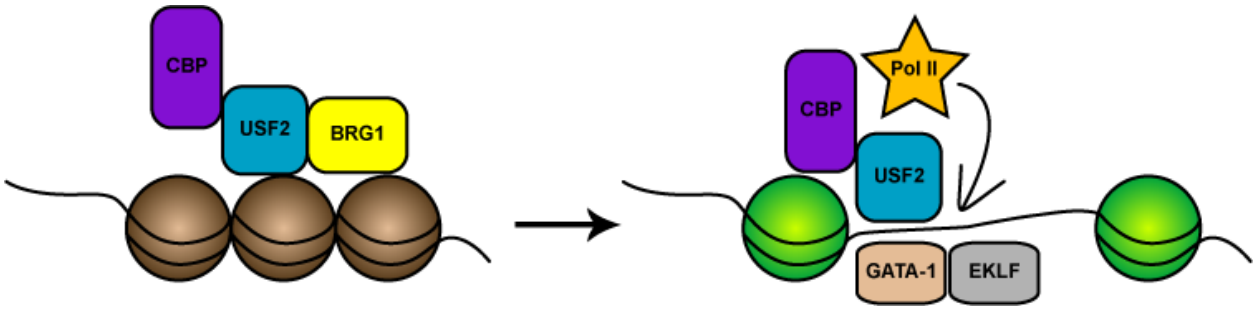

Figure 4-6. Model of USF2/CBP/BRG1-mediated recruitment of RNA Pol II to the β-globin LCR. USF2 and CBP are recruited to the LCR prior to induction and recruit the chromatin modifying enzyme BRG1. After remodeling has occurred and nucleosomes have been shifted or slid to create a more open chromatin structure, BRG1 is no longer needed and dissociates from the LCR. USF2 and CBP remain bound, and may serve to recruit RNA Pol II. The open chromatin also allows for other transcription factors, such as GATA-1 and EKLF, to bind and enhance β-type globin gene expression.

CHAPTER 5
EXAMINING THE CONFORMATION OF THE BETA-GLOBIN LOCUS AT DIFFERENT CELL CYCLE STAGES

Introduction

All organisms consist of cells that multiply through cell division. An adult human has approximately 1×10^{14} cells, all of which originated from a single fertilized egg cell (59). An enormous number of cells in the human body are also continuously dividing to replace dying cells. Before a cell can divide, it must grow in size, duplicate its chromosomes, and separate those chromosomes for exact distribution between two daughter cells (3). These different processes are coordinated in the cell cycle. During the first phase (G_1, first gap) of the cell cycle, the cell grows and becomes larger. Once it has reached a certain size, the cell enters S-phase (synthesis), in which it duplicates its genome and a copy of each chromosome is formed. During the next phase (G_2, second gap), the cell ensures that DNA replication is completed and prepares for cell division. The chromosomes are separated during M-phase (mitosis) and the cell divides, yielding two daughter cells that each carry a copy of the original genome (3).

The transcriptional activity of genes may be affected by replication and other changes that alter the DNA structure during these different stages of the cell cycle (213). In general, the replication of many tissue specific genes occurs late in most tissues during S-phase but early in the tissue of expression (214). In humans, the β-globin gene locus replicates early in erythroid cells, but late in nearly all other cell types (84). Studies show that in the β-globin gene locus, this early replication timing is correlated with an open chromatin structure but not gene transcription (215). In eukaryotic cells entering mitosis, transcription is silenced and most transcription factors, as well as RNA polymerase II (RNA Pol II), are displaced (216). However, other studies

have shown that transcription factors TFII-D and TFII-B can remain associated with active gene promoters during mitosis (217). These mitotic promoter complexes may maintain the transcriptional competence of genes or facilitate the rapid reactivation of transcription upon exit from mitosis.

In a previous study by Dr. Karen Vieira, the interaction of transcription complexes with the human β-globin locus was analyzed using chromatin immunoprecipitation (ChIP) in synchronized human erythroleukemia (K562) cells (90). The analysis of protein/DNA interactions at the globin locus during S-phase in synchronized cells revealed a dynamic pattern of dissociation and re-association for both transcription factors and RNA Pol II. Synchronized samples were analyzed directly after double thymidine block (time 0), as well as cells harvested at 15 min, 45 min, 2 h, and 6 h after release from the block. At time 0, RNA Pol II is observed to associate with HS2 as well as with the ε- and γ-globin promoters. After 15 min, NF-E2 (p45) is also observed to colocalize to HS2, but is not present at the γ-globin promoter until after 45 min. After 2 h, RNA Pol II dissociates from both HS2 as well as the γ-globin promoter. Interestingly, USF2 is observed to associate with HS2 and the γ-globin promoter at this time. By performing semi-quantitative PCR using primers specific for the human globin locus, it was determined that replication of the β-globin locus is completed by 2 h in K562 cells (90). After 6 h, USF2 is observed to dissociate from the locus, which is accompanied by the reappearance of RNA Pol II at HS2 and the γ-globin promoter.

In the study presented here, the synchronization of K562 cells by double thymidine block was repeated in order to verify the previously observed results. Additionally, K562 cells released from block were subsequently incubated in the presence of nocodazole;

this reversibly arrests cells in M-phase by chemically interfering with the polymerization of microtubules, which prevents the mitotic spindles from forming (218,219). The interaction of transcription complexes with specific regions of the human β-globin locus in these mitotically-arrested K562 cells was examined by ChIP and PCR. Finally, synchronized K652 cells were also harvested for chromatin conformation capture (3C) analysis of the human β-globin locus at specific time points after the release from double thymidine block in order to examine the conformation of the locus at those times.

Results

Both untreated and drug-treated K562 cells were subjected to flow cytometry analysis in order to verify synchronization and arrest (Figure 5-1). Cells harvested for analysis directly after the double thymidine block were considered time 0, and are blocked at the boundary of G_1/S-phase. These cells are labeled as “synchronized” in Figures 5-1 and 5-2. As observed in Dr. Karen Vieira’s work, RNA Pol II binds to both HS2 and the γ-globin gene promoter in synchronized cells harvested at time 0 (Figure 5-2).

Originally, it was anticipated that the extensive condensation of chromatin during mitosis would disrupt the globin locus conformation required for transcription of the genes. However, results from PCRs performed on ChIP samples from mitotically arrested K562 cells consistently showed association of RNA Pol II with the γ-globin gene promoter region during M-phase (Figure 5-2). While RNA Pol II appears to associate with the γ-globin gene promoter region in these cells, it is not known whether transcription is taking place, or if RNA Pol II is merely associated but remaining idle until completion of mitosis. Since K562 cells express high levels of γ-globin, it is possible that the continuous association of RNA Pol II may assist in the transition to active

transcription immediately upon exit from M-phase. Interestingly, CBP is also seen to bind to HS2 during M-phase, and may serve to re-recruit RNA Pol II to HS2 (Figure 5-2). Interestingly, during M-phase, lower levels of RNA Pol II are observed to associate with HS2 than with the γ-globin promoter (Figure 5-2), which was confirmed by qPCR (data not shown).

Additional PCRs using primers designed to amplify the 3' end of the ^{A}γ-globin were performed in order to determine whether RNA Pol II was also present in this region. Levels of RNA Pol II binding appear to be reduced compared to that of the 5' end, suggesting that any RNA Pol II bound to the ^{A}γ-globin gene promoter during the M-phase arrest are stalled and not engaged in transcription elongation (Figure 5-2). The reduction of RNA Pol II loading at the 3' end was also confirmed by quantitative RT-PCR (data not shown).

In order to determine if RNA Pol II was stalled at the 5' end of the γ-globin gene, ChIP was performed on both synchronized (time 0) and mitotically arrested K562 cells using antibodies for total RNA Pol II as well as antibodies that discriminate between the transcribing and non-transcribing forms of RNA Pol II. Phosphorylation of serine 5 (Ser-5) of the carboxy-terminal domain (CTD) of RNA Pol II is indicative of transcriptional initiation, while serine 2 (Ser-2) phosphorylation of the CTD is indicative of transcriptional elongation. In comparison to total RNA Pol II levels, both Ser-2 and Ser-5 RNA Pol II levels at the γ-globin promoter are low (Figure 5-3). This would indicate that in both synchronized and mitotically arrested cells, any RNA Pol II binding to HS2 and the γ-globin promoter are unphosphorylated and therefore inactive. However, this ChIP

has only been performed once and needs to be repeated before any conclusions can be made regarding the data.

In order to ascertain the chromatin conformation at specific times during S-phase, 3C was performed on synchronized K562 cells. The 3C assay allows for the study of higher order chromatin structures; specifically, where two or more pieces of chromatin are brought together in close spatial proximity (194). It is thought that these associations between different chromatin domains may mediate transcriptional regulation and other cellular functions. In the β-globin gene locus, 3C has been used extensively in mammalian cells to study of long-range chromatin interactions between the LCR and the globin genes downstream (220). In this study, K562 cells were first synchronized by double thymidine block, harvested at specific times after block (0, 45 min, 2 h), and then analyzed by both ChIP and 3C. The ChIP was performed in order to verify what was previously observed in Dr. Karen Viera's work, but the resulting timing of dissociation of RNA Pol II and USF2 are not the same. Nevertheless, the trend is still similar, as RNA Pol II is seen to dissociate from HS2 and the γ-globin gene promoter at 45 min after release from block, and observed to re-appear at 2 h (Figure 5-4).

The same samples harvested at 0, 45 min, and 2 h after release were also subjected to 3C analysis followed by subsequent PCR analyses using primers that would detect a ligation product between LCR element HS2 and either HS3, the ^{A}γ-globin gene, or the 3'HS1 (Figure 5-5). A band the expected size (~226 bp) was observed in the lane using primers amplifying a ligation product between HS2 and the ^{A}γ-globin gene for all time points. While this could indicate a possible interaction between HS2 and the ^{A}γ-globin gene during these times, and therefore suggest that the

loop formation remains intact during the dissociation of RNA Pol II, various larger and nonspecific bands also apparent on the gel (Figure 5-5). The ~226 bp band, which matches the expected size of a ligation product between HS2 and the the $^{A}\gamma$-globin gene, was excised and subjected to gel extraction for verification by sequencing, but unfortunately a clear verifiable sequence was unable to be obtained. A PCR using the same primers on genomic DNA extracted directly from K562 cells yielded no product, indicating that the observed band is not a byproduct of genomic amplification using these primers (data not shown).

Discussion

It could be argued that the binding of RNA Pol II observed in the mitotically arrested K562 cells is attributable to the small population of cells escaping arrest, and the non-arrested cells would continue to transcribe γ-globin. However, this gene would not, in theory, undergo active transcription in arrested cells. In order to test this hypothesis, heterogeneous nuclear RNA (hnRNA) levels of the gene could be measured in both unsynchronized and arrested cells. The short-lived hnRNA, also known as precursor mRNA or pre-mRNA, is an incompletely processed single strand of mRNA synthesized from a DNA template (221). Thus, aside from containing uracil in place of thymine, the hnRNA sequence is identical to the original genomic DNA because it still contains introns, but also contains the 5' cap and poly-A tail. While mRNA dictates steady state transcription levels, the rapid processing and turnover of hnRNA makes this transient RNA form ideal for determining active transcription levels, since eukaryotic hnRNA exists only briefly before it is fully processed into mature mRNA. It has previously been shown that both chicken and mouse globin hnRNA have half-lives of around 3-5 minutes (222,223). In comparison, it is estimated that the

structural half-lives of both γ- and β-globin mRNA are approximately 27-29 hours, while the functional half-lives are approximately 7-9 hours, based on studies using cultured human adult and neonatal reticulocytes (224). If the level of reduction in the rapidly processed form of newly transcribed γ-globin hnRNA between the unsynchronized and the mitotically-arrested K562 cells is proportionally comparable to the reduction in the total number of cells in G_1-phase, then it could be assumed that the RNA Pol II binding is attributable to arrested cells. This would indicate that the mitotically arrested K562 cells have RNA Pol II bound to the γ-globin gene promoter even in M-phase, but are not actively transcribing. In addition to these possible future hnRNA studies, more ChIP on the Ser-2 and Ser-5 phosphorylated forms of RNA Pol II need to be performed on the synchronized and mitotically arrested cells. Additionally, ChIP using antibodies against Ser-2 and Ser-5 RNA Pol II needs to be performed on unsynchronized K562 cells to serve as a comparison.

The 3C assay presented here is only preliminary and although it is reproducible, the amplified ligation product between HS2 and ^{A}γ-globin was unable to be sequenced for verification. Based on this as well as the additional bands observed on the gel, it is likely the primer may not be entirely specific for the intended region of amplification. Adjustments to the annealing temperature in the 3C PCR protocol may need to be made to reduce background levels. Melting curve analysis on the PCR-amplified DNA may also help determine whether the primers for amplifying a ligation product between HS2 and the ^{A}γ-globin gene have sufficient specificity. Additionally, different primers for 3C analysis of the human β-globin gene locus were designed, but no 3C ligation products were detected in PCR analyses using these new primers (data not shown).

Immediately, it is imperative that viable 3C primers for the human globin locus be obtained before any more studies of the chromatin conformation in these synchronized K562 cells can be done. In the future, additional important 3C controls must also be implemented, including but not limited to performing 3C on non-crosslinked cells as well as non-erythroid cells, a no restriction enzyme control, and a no ligation control. Furthermore, the HS3-HS2 ligation fragment was originally intended to serve as a positive control in the 3C assay due to the fact that these two regions reside in close proximity within the LCR. However, no ligation between HS3 and HS2 was detected by PCR. It is possible that a fragment may become apparent with additional cycles in the PCR.

Nevertheless, if the PCR band observed in the ligation product between HS2 and the γ-globin gene promoter is not a mere artifact, it poses an interesting scenario. This would suggest that LCR remains in contact with the globin genes downstream even when RNA Pol II is dissociated, and that the association of other factors, such as USF2, may mediate the continued association between the LCR and the genes. These factors may serve to re-recruit RNA Pol II upon completion of replication. It is unknown how replication factors affect the conformation of the locus, and unfortunately, in the context of the nonspecific bands and lack of sequence verification, no definitive conclusions can be made regarding these 3C data. In addition to the controls mentioned above, more time points after release from block should be analyzed as well, such as 0, 15 min, 45 min, 2 h, 4 h, and 6 h after release from block.

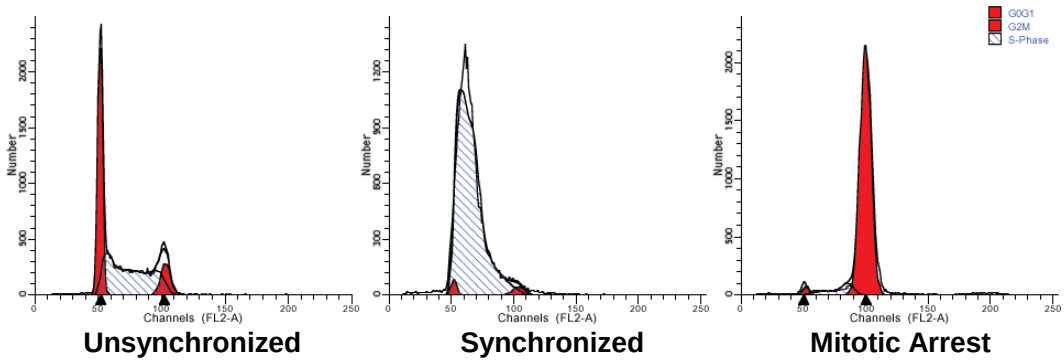

Figure 5-1. Representative flow cytometry analysis of untreated (unsynchronized) and drug-treated (synchronized, mitotic arrest) K562 cells stained with propidium iodide. Cells treated with double thymidine block and harvested at time 0 are labeled as synchronized, while mitotic arrest indicates cells incubated with nocodozole after release from block.

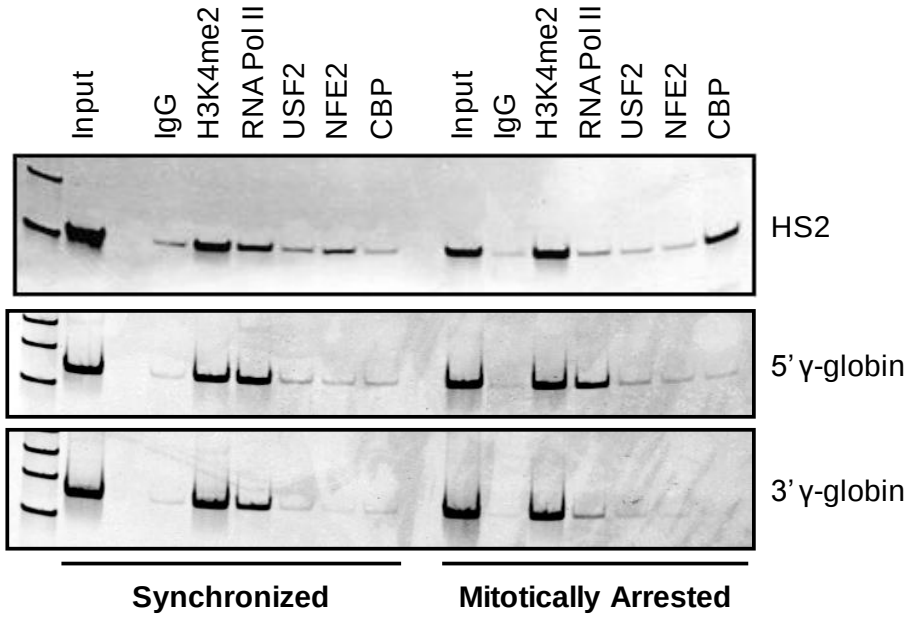

Figure 5-2. Analysis of ChIP from synchronized (harvested at time 0) or mitotically arrested K562 cells. ChIP was performed with the indicated antibodies, and samples were analyzed by PCR using primers for LCR element HS2 (top), the 5' γ-globin promoter region (middle), or the 3' end of γ-globin (bottom). PCR products were electrophoresed on a 5% TBE gel and stained with SybrGreen for analysis.

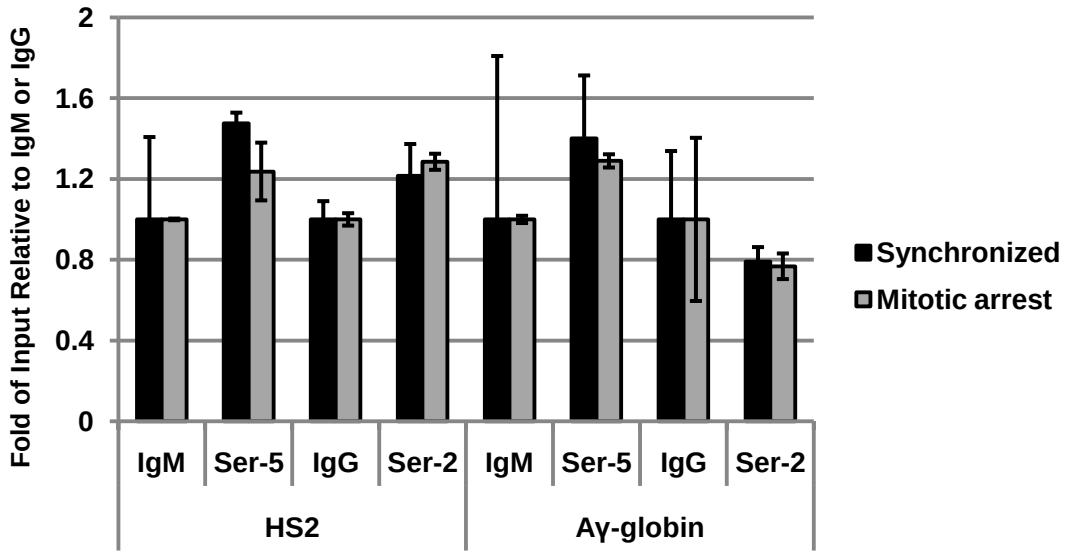

Figure 5-3. qPCR analysis of ChIP on various phosphorylation states of RNA Pol II in synchronized and mitotically arrested K562 cells. Data are normalized to IgM (Ser-5) or IgM (Ser-2), and error bars reflect means ± standard error of a single ChIP experiment with qPCR performed in duplicate.

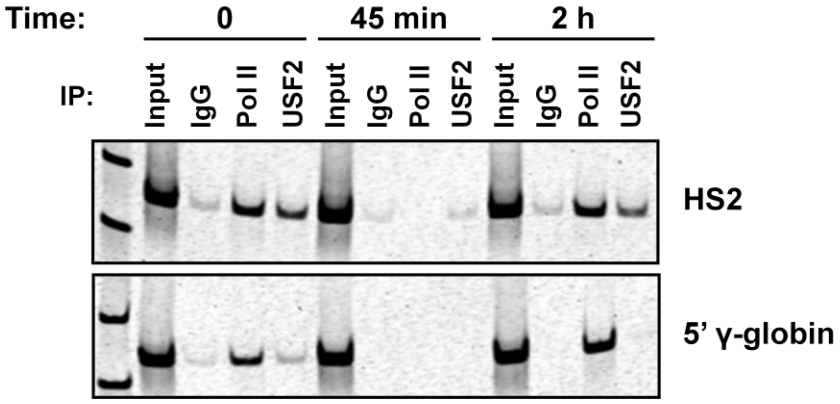

Figure 5-4. Analysis of ChIP from synchronized K562 cells harvested at the indicated time points (0, 45 min, 2 h). ChIP was performed with the indicated antibodies, and samples were analyzed by PCR using primers for LCR element HS2. PCR products were electrophoresed on a 5% TBE gel and stained with SybrGreen for analysis.

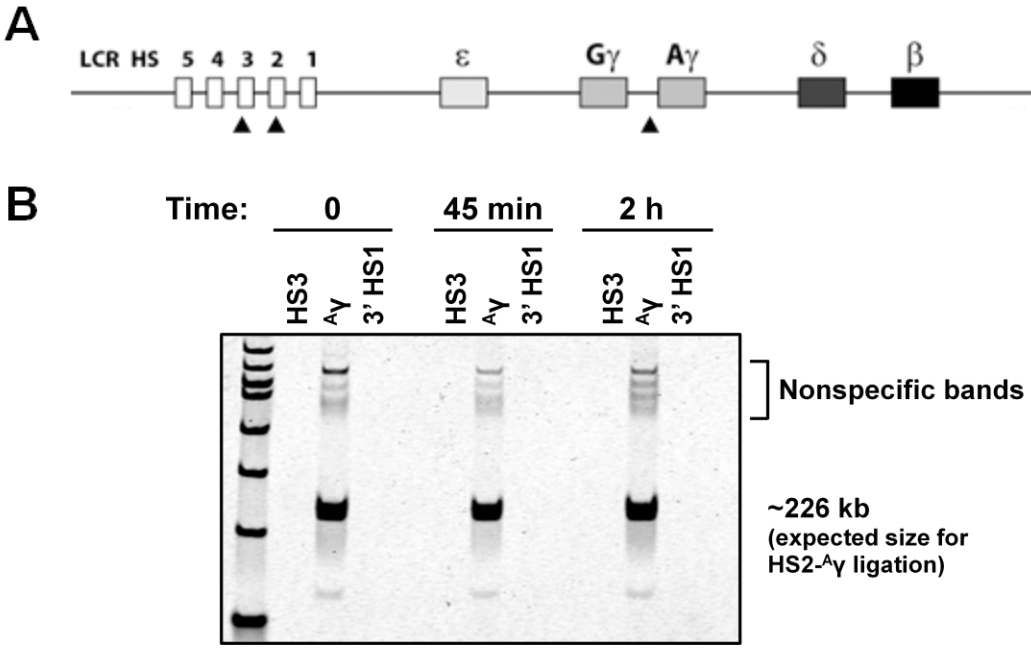

Figure 5-5. 3C Analysis of synchronized K562 cells. A) Schematic of the human β-globin locus, with black triangles indicating approximate location of primers used. B) PCR analyses for cells harvested at 0, 45 min, and 2 h after double thymidine block, as indicated. The HS2 primer was used as an anchor to examine ligation products between the LCR and other regions in the locus, including HS3, $^{A}\gamma$-globin, and the 3' HS1. PCR products were electrophoresed on a 5% TBE gel and stained with SybrGreen for analysis.

CHAPTER 6
DISCUSSION AND FUTURE DIRECTIONS

Regulation of Mouse and Human β-globin Gene Expression by USF

It currently is unknown how ubiquitously expressed and tissue-specific transcription factors coordinate the activation of highly expressed genes during differentiation. Tissue-specific factors may mediate the accessibility of regulatory sites, while ubiquitously expressed proteins perform basic functions involved in the local remodeling of nucleosomes and the recruitment of transcription complexes. The data presented here demonstrate that Upstream Stimulatory Factor (USF) is required for high-level expression of the murine adult β_{maj}-globin gene. Based on the homology of the E-boxes in the β-globin promoter region (Figure 1-3), it is likely that USF also regulates the adult human β-globin gene. Additionally, it was shown previously that mutations of the +60 E-box in the adult β-globin promoter reduced transcription to a level comparable to that observed after mutation of the initiator sequence (113). It is hypothesized that in adult erythroid cells, USF interacts with the β-globin promoter and recruits coactivator complexes which modify the chromatin structure to increase accessibility for the transcriptional machinery.

In order to study the role of USF in adult human β-globin expression, transgenic mice expressing both dominant-negative USF (A-USF) and the human β-globin YAC (β-YAC) have been generated. This was done by crossing transgenic females from the A-USF mouse line with transgenic males carrying β-YAC. The β-YAC mice contain the entire human β-globin gene locus, including the locus control region (LCR) as well as areas upstream and downstream of the locus (Figure 6-1) (225,226). Studies of β-YAC mice have shown that in these mice, expression of the human β-type genes is also

limited to erythroid cells, and that the developmental regulation remains similar to that of the murine β-globin locus (226). Figure 6-1 illustrates differences in gene and tissue expression when analyzed in the context of transgenic mice (227). Any changes in recruitment of RNA polymerase II (RNA Pol II) or USF to the human β-globin gene in these A-USF/β-YAC mice can be examined by ChIP and subsequent real-time PCR. Unsurprisingly, no adult male mice carrying both A-USF and β-YAC have been generated as of yet. This is likely due to the fact that erythropoiesis in these mice is disturbed. Because the production of adult A-USF/β-YAC males is not anticipated to be possible, embryonic A-USF/β-YAC males should be analyzed in addition to adult A-USF/β-YAC females.

Recruitment of Transcription Complexes to the LCR

Previous studies have shown that Erythroid Krüppel-Like Factor (EKLF) is required for the recruitment of chromatin remodeling complexes to the LCR and to the adult β-globin gene promoter (165,228,229). Interestingly, a recent report from Sengupta et al. revealed a requirement for EKLF in the recruitment of TAF9, which interacts with a basal promoter element in the β-globin gene, suggesting that EKLF is directly involved in recruiting transcription complexes to the β-globin gene promoter (230). NF-E2 (p45) also is required for the efficient recruitment of RNA Pol II to the adult β-globin gene promoter, but it is dispensable for its initial recruitment to the LCR (231). Since both proteins appear to be regulated by USF, the data presented here suggest that USF regulates β-globin gene expression indirectly by enhancing the expression of erythroid-specific transcription factors and directly by cooperating with these factors in the recruitment of transcription complexes to the globin gene locus.

While the transcription tracking mechanism may explain gene activation, it is not known how looping can lead to the enhanced expression of the globin genes. LCR-mediated gene activation either results in enhanced recruitment of transcription complexes to the globin genes and/or in the conversion of transcription initiation complexes to elongation active complexes (26,108). This could be achieved by providing activities that are first recruited to the LCR and subsequently transferred to globin genes. For example, transcription complexes could first be recruited to the LCR and looping would mediate the transfer to the globin gene promoters. The active chromatin hub (ACH) would provide a high local concentration of transcription factors that could efficiently capture transcription complexes which would then be recruited to and positioned at the basal promoters of the globin genes to engage in productive transcription. Alternatively, or additionally, elongation incompetent transcription complexes may be recruited to the genes while the LCR provides activities necessary for activation, for example, kinases that phosphorylate serine 2 (Ser-2) at the RNA Pol II carboxy-terminal domain (CTD) or co-regulator complexes that modify chromatin structure at the globin gene promoters either to enable recruitment of transcription complexes and/or to allow elongation.

Additionally, the data presented here also suggest that in MEL cells, RNA Pol II is recruited to the LCR prior to recruitment of TBP (TATA-binding protein), although TBP is observed to be present at the β_{maj} promoter before induction (Figure 4-4). A TBP-independent mechanism of recruitment of RNA Pol II to the HS2 may exist, which would involve other factors. CBP (CREB binding protein) is one likely candidate; it is known to bind to enhancers, and was observed to bind to HS2 in mitotically-arrested K562 cells

(Figure 5-2). Studies regarding the TBP-independent recruitment of RNA Pol II to the LCR can be performed on linearized LCR constructs immobilized on streptavidin-coated magnetic beads, as described previously (90,144,211). Nuclear extract from uninduced and induced MEL cells would be immunodepleted using antibodies against TBP and/or CBP, and the immobilized LCR-streptavidin bead constructs would then be incubated in this nuclear extract. This would allow the observation of any changes in the recruitment of RNA Pol II to the LCR in the absence of TBP or CBP, or both.

USF, BRG1, and Transcription Factories

USF was one of the first transcription factors shown to activate transcription mediated by RNA Pol II, and it plays a role in the high-level expression of many genes in differentiated cells (130,139). Accumulating evidence points to the possibility that highly expressed genes are transcribed in specialized nuclear domains enriched for splicing factors and RNA Pol II, often referred to as transcription factories or transcription domains (207-209). It is possible that the LCR nucleates such a transcription domain in erythroid cells (184). USF and Brahma-Related Gene 1 (BRG1) are possible candidate proteins that could mediate the association of genes or regulatory elements with transcription factories in the nucleus. In future studies, immunofluorescence and florescent *in situ* hybridization (FISH) can be used on yolk sac cells isolated from A-USF transgenic embryos in order to determine if recruitment of the β-globin locus to transcription factories is hindered by expression of A-USF.

Therapeutic Strategies

The β-globin genes are competitively regulated by the LCR, and activation of the β-type globin genes requires the LCR to come in close proximity to the genes (104,127). During the differentiation of erythroid cells, it appears that certain factors and protein

complexes, including RNA Pol II, first associate with the LCR before they interact with the globin gene promoters (90,212,232). The LCR could serve as the primary site of recruitment for activities involved in globin gene regulation, and these activities could be transferred to the globin genes by looping mechanisms (26,105,144). If mechanisms are known that mediate the stage-specific association of the globin genes with the LCR, strategies could be developed to change these association patterns. This could lead to novel therapies for the treatment of sickle cell anemia or other hemoglobinopathies, e.g., by favoring interactions of the LCR with therapeutic γ-globin genes over those with the mutant β-globin genes.

Summary

In summary, these data show that USF is required for high-level globin gene expression, and that it enhances the expression of erythropoietic transcription factors. Additionally, USF mediates the recruitment of transcription complexes to the β-globin gene locus. USF also appears to form a complex with the BRG1 subunit of the SWI/SNF chromatin remodeling complex. It is hoped that this increased understanding of USF's role in β-globin gene expression and erythropoiesis will lead to better therapeutic strategies for individuals suffering from hemoglobinopathies.

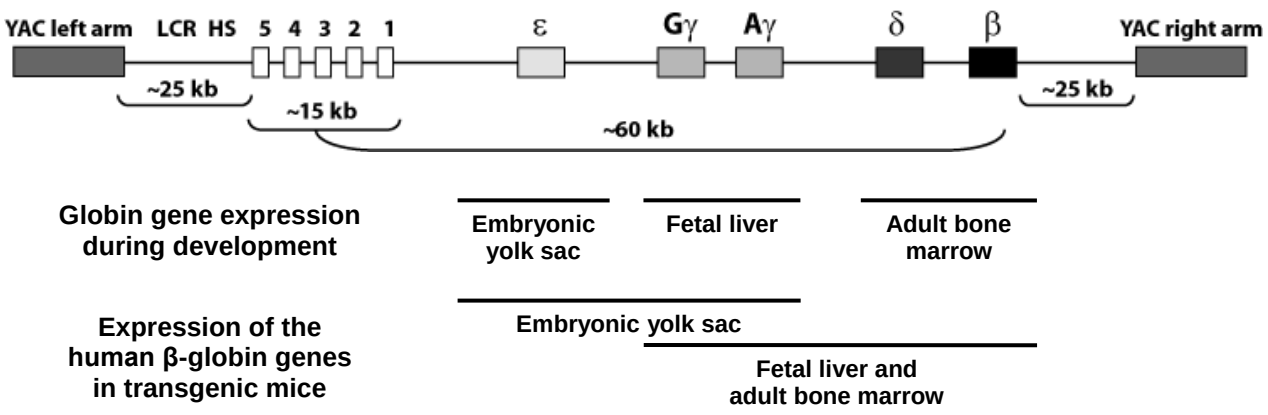

Figure 6-1. Schematic representation of the human β-globin gene locus YAC (β-YAC) and changes in expression pattern when expressed in transgenic mice (26). In β-YAC mice, the ε-globin and γ-globin genes are co-expressed in the embryonic yolk sac, while the β-globin gene is expressed at high levels in fetal liver and circulating erythroid cells from bone marrow.

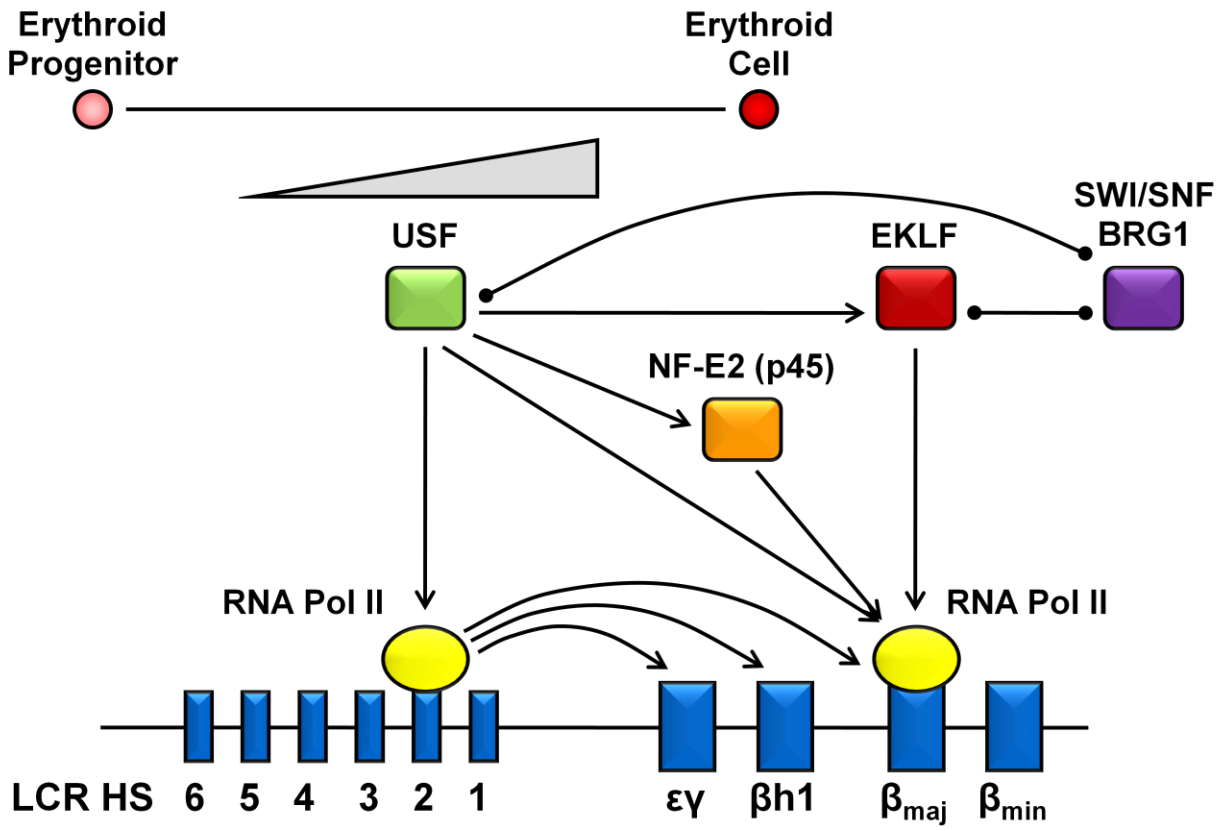

Figure 6-2. Model depicting USF and BRG1-mediated regulation of β-globin gene expression (210). The expression of USF increases during the differentiation of erythroid cells. Both USF and EKLF interact physically with the BRG1 subunit of SWI/SNF, and may serve to modify the chromatin structure to a transcriptionally permissive state. USF also regulates the recruitment of transcription complexes to the β-globin gene locus by interacting with E-boxes located in LCR element HS2 and in the adult β-globin gene promoter. Through the LCR, USF regulates the expression of the embryonic genes. USF further regulates the expression of the globin genes indirectly by enhancing the expression of erythroid cell-specific transcription factors with which it cooperates in mediating the recruitment of transcription complexes to the globin gene locus.

LIST OF REFERENCES

1. Stamatoyannopoulos, G., and Nienhuis, A. (1994) *The Molecular Basis of Blood Diseases*, 2nd ed., W. B. Saunders, Philadelphia, PA

2. Norberg, B., Bandmann, U., and Rydgren, L. (1977) *J Mechanochem Cell Motil* **4**, 37-53

3. Campbell, N. A., Reece, J. B., and Mitchell, L. G. (1999) *Biology*, 5th ed., Benjamin/Cummings, an imprint of Addison Wesley Longman, Inc., Menlo Park, CA

4. Tsiftsoglou, A. S., Vizirianakis, I. S., and Strouboulis, J. (2009) *IUBMB Life* **61**, 800-830

5. Fried, W. (2009) *Exp Hematol* **37**, 1007-1015

6. Perutz, M. F. (1960) *Brookhaven Symp Biol* **13**, 165-183

7. Huehns, E. R., Flynn, F. V., Butler, E. A., and Beaven, G. H. (1961) *Nature* **189**, 496-497

8. Huisman, T. H., and Prins, H. K. (1955) *J Lab Clin Med* **46**, 255-262

9. Allen, D. W., Wyman, J., Jr., and Smith, C. A. (1953) *J Biol Chem* **203**, 81-87

10. Perutz, M. F. (1978) *Sci Am* **239**, 92-125

11. Weatherall, D. J. (1997) *BMJ* **314**, 1675-1678

12. Trampuz, A., Jereb, M., Muzlovic, I., and Prabhu, R. M. (2003) *Crit Care* **7**, 315-323

13. Williams, T. N., Maitland, K., Bennett, S., Ganczakowski, M., Peto, T. E., Newbold, C. I., Bowden, D. K., Weatherall, D. J., and Clegg, J. B. (1996) *Nature* **383**, 522-525

14. Aidoo, M., Terlouw, D. J., Kolczak, M. S., McElroy, P. D., ter Kuile, F. O., Kariuki, S., Nahlen, B. L., Lal, A. A., and Udhayakumar, V. (2002) *Lancet* **359**, 1311-1312

15. Murayama, M. (1967) *Clin Chem* **13**, 578-588

16. Voet, D., Voet, J. G., and Pratt, C. W. (2002) *Fundamentals of Biochemistry*, Upgrade ed., John Wiley & Sons, Inc., Hoboken, NJ

17. Galanello, R., and Cao, A. (1998) *Ann N Y Acad Sci* **850**, 325-333

18. Weatherall, D. J. (2004) *Nat Rev Genet* **5**, 625-631

19. Urbinati, F., Madigan, C., and Malik, P. (2006) *Expert Rev Mol Med* **8**, 1-26

20. Nathan, D. G., and Gunn, R. B. (1966) *Am J Med* **41**, 815-830

21. Hershko, C., Graham, G., Bates, G. W., and Rachmilewitz, E. A. (1978) *Br J Haematol* **40**, 255-263

22. Porter, J. B., Abeysinghe, R. D., Marshall, L., Hider, R. C., and Singh, S. (1996) *Blood* **88**, 705-713

23. Modell, B., Khan, M., and Darlison, M. (2000) *Lancet* **355**, 2051-2052

24. Spivak, J. L., Gascon, P., and Ludwig, H. (2009) *Oncologist* **14 Suppl 1**, 43-56

25. Bulger, M., and Groudine, M. (1999) *Genes Dev* **13**, 2465-2477

26. Levings, P. P., and Bungert, J. (2002) *Eur J Biochem* **269**, 1589-1599

27. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389**, 251-260

28. Bishop, T. C. (2008) *Biophys J* **95**, 1007-1017

29. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L. (1998) *Proc Natl Acad Sci U S A* **95**, 14173-14178

30. Thoma, F., Koller, T., and Klug, A. (1979) *J Cell Biol* **83**, 403-427

31. Luger, K., and Richmond, T. J. (1998) *Curr Opin Genet Dev* **8**, 140-146

32. Lennartsson, A., and Ekwall, K. (2009) *Biochim Biophys Acta* **1790**, 863-868

33. Munshi, A., Shafi, G., Aliya, N., and Jyothy, A. (2009) *J Genet Genomics* **36**, 75-88

34. Strahl, B. D., and Allis, C. D. (2000) *Nature* **403**, 41-45

35. Jenuwein, T., and Allis, C. D. (2001) *Science* **293**, 1074-1080

36. Pal, S., and Sif, S. (2007) *J Cell Physiol* **213**, 306-315

37. Tian, X., and Fang, J. (2007) *Acta Biochim Biophys Sin (Shanghai)* **39**, 81-88

38. Shi, Y. (2007) *Nat Rev Genet* **8**, 829-833

39. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., and Casero, R. A. (2004) *Cell* **119**, 941-953

40. Anand, R., and Marmorstein, R. (2007) *J Biol Chem* **282**, 35425-35429

41. Kuo, M. H., and Allis, C. D. (1998) *Bioessays* **20**, 615-626

42. Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999) *Nature* **399**, 491-496

43. Vignali, M., Hassan, A. H., Neely, K. E., and Workman, J. L. (2000) *Mol Cell Biol* **20**, 1899-1910

44. Gangaraju, V. K., and Bartholomew, B. (2007) *Mutat Res* **618**, 3-17

45. Ho, L., and Crabtree, G. R. (2010) *Nature* **463**, 474-484

46. Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., and Workman, J. L. (2002) *Cell* **111**, 369-379

47. Martens, J. A., and Winston, F. (2003) *Curr Opin Genet Dev* **13**, 136-142

48. Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M., Tsukiyama, T., Wu, C., Pimpinelli, S., and Tamkun, J. W. (2000) *Mol Cell* **5**, 355-365

49. Corona, D. F., and Tamkun, J. W. (2004) *Biochim Biophys Acta* **1677**, 113-119

50. Goldmark, J. P., Fazzio, T. G., Estep, P. W., Church, G. M., and Tsukiyama, T. (2000) *Cell* **103**, 423-433

51. Grune, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F., Becker, P. B., and Muller, C. W. (2003) *Mol Cell* **12**, 449-460

52. Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000) *Nature* **406**, 541-544

53. Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J., and Wang, W. (1998) *Mol Cell* **2**, 851-861

54. Kobor, M. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., Madhani, H. D., and Rine, J. (2004) *PLoS Biol* **2**, E131

55. Ruhl, D. D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M. P., Conaway, R. C., Conaway, J. W., and Chrivia, J. C. (2006) *Biochemistry* **45**, 5671-5677

56. Adam, M., Robert, F., Larochelle, M., and Gaudreau, L. (2001) *Mol Cell Biol* **21**, 6270-6279

57. Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002) *Curr Opin Genet Dev* **12**, 162-169

58. Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J., and Lyons, I. (2001) *Curr Biol* **11**, 1183-1187

59. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) *Molecular Biology of the Cell*, 4th ed., Garland Science, New York, NY

60. Schardin, M., Cremer, T., Hager, H. D., and Lang, M. (1985) *Hum Genet* **71**, 281-287

61. Ragoczy, T., Bender, M. A., Telling, A., Byron, R., and Groudine, M. (2006) *Genes Dev* **20**, 1447-1457

62. Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., and Bickmore, W. A. (2001) *Hum Mol Genet* **10**, 211-219

63. Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., and Bickmore, W. A. (1999) *J Cell Biol* **145**, 1119-1131

64. Hughes, T. A., Pombo, A., McManus, J., Hozak, P., Jackson, D. A., and Cook, P. R. (1995) *J Cell Sci Suppl* **19**, 59-65

65. Wei, X., Samarabandu, J., Devdhar, R. S., Siegel, A. J., Acharya, R., and Berezney, R. (1998) *Science* **281**, 1502-1506

66. Grande, M. A., van der Kraan, I., de Jong, L., and van Driel, R. (1997) *J Cell Sci* **110 (Pt 15)**, 1781-1791

67. Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996) *J Cell Sci* **109 (Pt 6)**, 1427-1436

68. Hieda, M., Winstanley, H., Maini, P., Iborra, F. J., and Cook, P. R. (2005) *Chromosome Res* **13**, 135-144

69. Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., and Fraser, P. (2004) *Nat Genet* **36**, 1065-1071

70. Bender, M. A., Bulger, M., Close, J., and Groudine, M. (2000) *Mol Cell* **5**, 387-393

71. Smale, S. T., and Kadonaga, J. T. (2003) *Annu Rev Biochem* **72**, 449-479

72. Willis, I. M. (1993) *Eur J Biochem* **212**, 1-11

73. Grummt, I. (1999) *Prog Nucleic Acid Res Mol Biol* **62**, 109-154

74. Sims, R. J., 3rd, Mandal, S. S., and Reinberg, D. (2004) *Curr Opin Cell Biol* **16**, 263-271

75. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989) *Cell* **56**, 549-561

76. Hoiby, T., Zhou, H., Mitsiou, D. J., and Stunnenberg, H. G. (2007) *Biochim Biophys Acta* **1769**, 429-436

77. Robert, F., Douziech, M., Forget, D., Egly, J. M., Greenblatt, J., Burton, Z. F., and Coulombe, B. (1998) *Mol Cell* **2**, 341-351

78. Thomas, M. C., and Chiang, C. M. (2006) *Crit Rev Biochem Mol Biol* **41**, 105-178

79. Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999) *Proc Natl Acad Sci U S A* **96**, 13668-13673

80. Hampsey, M., and Reinberg, D. (2003) *Cell* **113**, 429-432

81. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998) *Genes Dev* **12**, 343-356

82. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) *Cell* **97**, 41-51

83. Orphanides, G., and Reinberg, D. (2002) *Cell* **108**, 439-451

84. Li, Q., Peterson, K. R., Fang, X., and Stamatoyannopoulos, G. (2002) *Blood* **100**, 3077-3086

85. Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987) *Cell* **51**, 975-985

86. Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990) *Genes Dev* **4**, 1637-1649

87. Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M. A., Fiering, S., Enver, T., Martin, D. I., Kennedy, M., Keller, G., and Groudine, M. (1998) *Mol Cell* **2**, 447-455

88. Stamatoyannopoulos, G. (2005) *Exp Hematol* **33**, 259-271

89. Higgs, D. R. (1998) *Cell* **95**, 299-302

90. Vieira, K. F., Levings, P. P., Hill, M. A., Crusselle, V. J., Kang, S. H., Engel, J. D., and Bungert, J. (2004) *J Biol Chem* **279**, 50350-50357

91. Li, Q., Zhang, M., Duan, Z., and Stamatoyannopoulos, G. (1999) *Genomics* **61**, 183-193

92. Li, Q., and Stamatoyannopoulos, G. (1994) *Blood* **84**, 1399-1401

93. Li, Q., Zhang, M., Han, H., Rohde, A., and Stamatoyannopoulos, G. (2002) *Nucleic Acids Res* **30**, 2484-2491

94. Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. (1996) *EMBO J* **15**, 562-568

95. Tuan, D. Y., Solomon, W. B., London, I. M., and Lee, D. P. (1989) *Proc Natl Acad Sci U S A* **86**, 2554-2558

96. Tuan, D., Kong, S., and Hu, K. (1992) *Proc Natl Acad Sci U S A* **89**, 11219-11223

97. Orphanides, G., and Reinberg, D. (2000) *Nature* **407**, 471-475

98. Tanimoto, K., Sugiura, A., Omori, A., Felsenfeld, G., Engel, J. D., and Fukamizu, A. (2003) *Mol Cell Biol* **23**, 8946-8952

99. Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007) *Nucleic Acids Res* **35**, 5532-5544

100. Johnson, K. D., Grass, J. A., Park, C., Im, H., Choi, K., and Bresnick, E. H. (2003) *Mol Cell Biol* **23**, 6484-6493

101. Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008) *PLoS One* **3**, e1661

102. Hosey, A. M., Chaturvedi, C. P., and Brand, M. (2010) *Epigenetics* **5**

103. Haussecker, D., and Proudfoot, N. J. (2005) *Mol Cell Biol* **25**, 9724-9733

104. Choi, O. R., and Engel, J. D. (1988) *Cell* **55**, 17-26

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

106. Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003) *Nat Genet* **35**, 190-194

107. Wijgerde, M., Grosveld, F., and Fraser, P. (1995) *Nature* **377**, 209-213

108. Sawado, T., Halow, J., Bender, M. A., and Groudine, M. (2003) *Genes Dev* **17**, 1009-1018

109. Lewis, B. A., and Orkin, S. H. (1995) *J Biol Chem* **270**, 28139-28144

110. Lewis, B. A., Kim, T. K., and Orkin, S. H. (2000) *Proc Natl Acad Sci U S A* **97**, 7172-7177

111. Hartzog, G. A., and Myers, R. M. (1993) *Mol Cell Biol* **13**, 44-56

112. Miller, I. J., and Bieker, J. J. (1993) *Mol Cell Biol* **13**, 2776-2786

113. Leach, K. M., Vieira, K. F., Kang, S. H., Aslanian, A., Teichmann, M., Roeder, R. G., and Bungert, J. (2003) *Nucleic Acids Res* **31**, 1292-1301

114. Bodine, D. M., and Ley, T. J. (1987) *EMBO J* **6**, 2997-3004

115. Stoeckert, C. J., Jr., and Cheng, H. (1996) *Am J Hematol* **51**, 220-228

116. Liu, Q., Tanimoto, K., Bungert, J., and Engel, J. D. (1998) *Proc Natl Acad Sci U S A* **95**, 9944-9949

117. Donovan-Peluso, M., Acuto, S., O'Neill, D., Hom, A., Maggio, A., and Bank, A. (1991) *Blood* **77**, 855-860

118. Wall, L., deBoer, E., and Grosveld, F. (1988) *Genes Dev* **2**, 1089-1100

119. Liu, Q., Bungert, J., and Engel, J. D. (1997) *Proc Natl Acad Sci U S A* **94**, 169-174

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

121. Kollias, G., Hurst, J., deBoer, E., and Grosveld, F. (1987) *Nucleic Acids Res* **15**, 5739-5747

122. Trudel, M., and Costantini, F. (1987) *Genes Dev* **1**, 954-961

123. Antoniou, M., deBoer, E., Habets, G., and Grosveld, F. (1988) *EMBO J* **7**, 377-384

124. Kollias, G., Wrighton, N., Hurst, J., and Grosveld, F. (1986) *Cell* **46**, 89-94

125. Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A. M., Strouboulis, J., de Laat, W., and Grosveld, F. G. (2004) *Genes Dev* **18**, 1495-1509

126. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L., and Townes, T. M. (1990) *Genes Dev* **4**, 380-389

127. Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Costantini, F., and Stamatoyannopoulos, G. (1990) *Nature* **344**, 309-313

128. Peterson, K. R., and Stamatoyannopoulos, G. (1993) *Mol Cell Biol* **13**, 4836-4843

129. Tanimoto, K., Liu, Q., Bungert, J., and Engel, J. D. (1999) *Nature* **398**, 344-348

130. Corre, S., and Galibert, M. D. (2005) *Pigment Cell Res* **18**, 337-348

131. Sirito, M., Walker, S., Lin, Q., Kozlowski, M. T., Klein, W. H., and Sawadogo, M. (1992) *Gene Expr* **2**, 231-240

132. Luo, X., and Sawadogo, M. (1996) *Mol Cell Biol* **16**, 1367-1375

133. Sirito, M., Lin, Q., Maity, T., and Sawadogo, M. (1994) *Nucleic Acids Res* **22**, 427-433

134. Cogswell, J. P., Godlevski, M. M., Bonham, M., Bisi, J., and Babiss, L. (1995) *Mol Cell Biol* **15**, 2782-2790

135. North, S., Espanel, X., Bantignies, F., Viollet, B., Vallet, V., Jalinot, P., Brun, G., and Gillet, G. (1999) *Oncogene* **18**, 1945-1955

136. Pawar, S. A., Szentirmay, M. N., Hermeking, H., and Sawadogo, M. (2004) *Oncogene* **23**, 6125-6135

137. Sirito, M., Lin, Q., Deng, J. M., Behringer, R. R., and Sawadogo, M. (1998) *Proc Natl Acad Sci U S A* **95**, 3758-3763

138. Crusselle-Davis, V. J., Vieira, K. F., Zhou, Z., Anantharaman, A., and Bungert, J. (2006) *Mol Cell Biol* **26**, 6832-6843

139. Sawadogo, M., and Roeder, R. G. (1985) *Cell* **43**, 165-175

140. Gregor, P. D., Sawadogo, M., and Roeder, R. G. (1990) *Genes Dev* **4**, 1730-1740

141. Rada-Iglesias, A., Ameur, A., Kapranov, P., Enroth, S., Komorowski, J., Gingeras, T. R., and Wadelius, C. (2008) *Genome Res* **18**, 380-392

142. Bresnick, E. H., and Felsenfeld, G. (1993) *J Biol Chem* **268**, 18824-18834

143. Elnitski, L., Miller, W., and Hardison, R. (1997) *J Biol Chem* **272**, 369-378

144. Leach, K. M., Nightingale, K., Igarashi, K., Levings, P. P., Engel, J. D., Becker, P. B., and Bungert, J. (2001) *Mol Cell Biol* **21**, 2629-2640

145. West, A. G., Huang, S., Gaszner, M., Litt, M. D., and Felsenfeld, G. (2004) *Mol Cell* **16**, 453-463

146. Crusselle-Davis, V. J., Zhou, Z., Anantharaman, A., Moghimi, B., Dodev, T., Huang, S., and Bungert, J. (2007) *FEBS J* **274**, 6065-6073

147. Roy, A. L., Meisterernst, M., Pognonec, P., and Roeder, R. G. (1991) *Nature* **354**, 245-248

148. Roy, A. L., Du, H., Gregor, P. D., Novina, C. D., Martinez, E., and Roeder, R. G. (1997) *EMBO J* **16**, 7091-7104

149. Huang, S., Li, X., Yusufzai, T. M., Qiu, Y., and Felsenfeld, G. (2007) *Mol Cell Biol* **27**, 7991-8002

150. Gallagher, P. G., Nilson, D. G., Steiner, L. A., Maksimova, Y. D., Lin, J. Y., and Bodine, D. M. (2009) *Blood* **113**, 1547-1554

151. Kim, S. I., and Bresnick, E. H. (2007) *Oncogene* **26**, 6777-6794

152. Johnson, K. D., Grass, J. A., Boyer, M. E., Kiekhaefer, C. M., Blobel, G. A., Weiss, M. J., and Bresnick, E. H. (2002) *Proc Natl Acad Sci U S A* **99**, 11760-11765

153. Munugalavadla, V., Dore, L. C., Tan, B. L., Hong, L., Vishnu, M., Weiss, M. J., and Kapur, R. (2005) *Mol Cell Biol* **25**, 6747-6759

154. Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J., and Blobel, G. A. (2005) *Mol Cell* **17**, 453-462

155. Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. (2005) *EMBO J* **24**, 2354-2366

156. Kooren, J., Palstra, R. J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F., and de Laat, W. (2007) *J Biol Chem* **282**, 16544-16552

157. Demers, C., Chaturvedi, C. P., Ranish, J. A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F. J., Groudine, M., and Brand, M. (2007) *Mol Cell* **27**, 573-584

158. Ogawa, K., Sun, J., Taketani, S., Nakajima, O., Nishitani, C., Sassa, S., Hayashi, N., Yamamoto, M., Shibahara, S., Fujita, H., and Igarashi, K. (2001) *EMBO J* **20**, 2835-2843

159. Zhao, X. F., and Aplan, P. D. (1999) *J Biol Chem* **274**, 1388-1393

160. Song, S. H., Hou, C., and Dean, A. (2007) *Mol Cell* **28**, 810-822

161. Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004) *Genes Dev* **18**, 2485-2490

162. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995) *Nature* **375**, 316-318

163. Zhou, D., Pawlik, K. M., Ren, J., Sun, C. W., and Townes, T. M. (2006) *J Biol Chem* **281**, 16052-16057

164. Perkins, A. C., Sharpe, A. H., and Orkin, S. H. (1995) *Nature* **375**, 318-322

165. Armstrong, J. A., Bieker, J. J., and Emerson, B. M. (1998) *Cell* **95**, 93-104

166. Kadam, S., McAlpine, G. S., Phelan, M. L., Kingston, R. E., Jones, K. A., and Emerson, B. M. (2000) *Genes Dev* **14**, 2441-2451

167. Muchardt, C., and Yaniv, M. (1999) *Semin Cell Dev Biol* **10**, 189-195

168. Reyes, J. C., Barra, J., Muchardt, C., Camus, A., Babinet, C., and Yaniv, M. (1998) *EMBO J* **17**, 6979-6991

169. Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., and Magnuson, T. (2000) *Mol Cell* **6**, 1287-1295

170. Bultman, S. J., Gebuhr, T. C., and Magnuson, T. (2005) *Genes Dev* **19**, 2849-2861

171. Kadam, S., and Emerson, B. M. (2003) *Mol Cell* **11**, 377-389

172. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M., and Eckner, R. (1995) *Nature* **374**, 81-84

173. Nordheim, A. (1994) *Nature* **370**, 177-178

174. Yao, T. P., Oh, S. P., Fuchs, M., Zhou, N. D., Ch'ng, L. E., Newsome, D., Bronson, R. T., Li, E., Livingston, D. M., and Eckner, R. (1998) *Cell* **93**, 361-372

175. Tanaka, Y., Naruse, I., Hongo, T., Xu, M., Nakahata, T., Maekawa, T., and Ishii, S. (2000) *Mech Dev* **95**, 133-145

176. Bannister, A. J., and Kouzarides, T. (1996) *Nature* **384**, 641-643

177. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H., and Nakatani, Y. (1996) *Cell* **87**, 953-959

178. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) *Nat Genet* **39**, 311-318

179. Xi, H., Shulha, H. P., Lin, J. M., Vales, T. R., Fu, Y., Bodine, D. M., McKay, R. D., Chenoweth, J. G., Tesar, P. J., Furey, T. S., Ren, B., Weng, Z., and Crawford, G. E. (2007) *PLoS Genet* **3**, e136

180. Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A. (2009) *Nature* **457**, 854-858

181. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S. H. (1998) *Proc Natl Acad Sci U S A* **95**, 2061-2066

182. Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998) *Nature* **396**, 594-598

183. Zhang, W., and Bieker, J. J. (1998) *Proc Natl Acad Sci U S A* **95**, 9855-9860

184. Liang, S., Moghimi, B., Yang, T. P., Strouboulis, J., and Bungert, J. (2008) *J Cell Biochem* **105**, 9-16

185. Emery, D. W., Yannaki, E., Tubb, J., and Stamatoyannopoulos, G. (2000) *Proc Natl Acad Sci U S A* **97**, 9150-9155

186. Bungert, J., Dave, U., Lim, K. C., Lieuw, K. H., Shavit, J. A., Liu, Q., and Engel, J. D. (1995) *Genes Dev* **9**, 3083-3096

187. Chomczynski, P., and Sacchi, N. (1987) *Anal Biochem* **162**, 156-159

188. Basu, P., Morris, P. E., Haar, J. L., Wani, M. A., Lingrel, J. B., Gaensler, K. M., and Lloyd, J. A. (2005) *Blood* **106**, 2566-2571

189. Tanabe, O., Shen, Y., Liu, Q., Campbell, A. D., Kuroha, T., Yamamoto, M., and Engel, J. D. (2007) *Genes Dev* **21**, 2832-2844

190. Nilson, D. G., Sabatino, D. E., Bodine, D. M., and Gallagher, P. G. (2006) *Exp Hematol* **34**, 705-712

191. Dahl, J. A., and Collas, P. (2009) *Methods Mol Biol* **567**, 59-74

192. Dahl, J. A., and Collas, P. (2007) *Stem Cells* **25**, 1037-1046

193. Miele, A., Gheldof, N., Tabuchi, T. M., Dostie, J., and Dekker, J. (2006) *Curr Protoc Mol Biol* **Chapter 21**, Unit 21 11

194. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002) *Science* **295**, 1306-1311

195. Chan, P. K., Wai, A., Philipsen, S., and Tan-Un, K. C. (2008) *PLoS One* **3**, e2134

196. Fullwood, M. J., and Ruan, Y. (2009) *J Cell Biochem* **107**, 30-39

197. Kunieda, T., Xian, M., Kobayashi, E., Imamichi, T., Moriwaki, K., and Toyoda, Y. (1992) *Biol Reprod* **46**, 692-697

198. Loken, M. R., Shah, V. O., Dattilio, K. L., and Civin, C. I. (1987) *Blood* **69**, 255-263

199. Qyang, Y., Luo, X., Lu, T., Ismail, P. M., Krylov, D., Vinson, C., and Sawadogo, M. (1999) *Mol Cell Biol* **19**, 1508-1517

200. Rishi, V., and Vinson, C. (2003) *Methods Enzymol* **370**, 454-466

201. Szentirmay, M. N., Yang, H. X., Pawar, S. A., Vinson, C., and Sawadogo, M. (2003) *J Biol Chem* **278**, 37231-37240

202. Giannola, D. M., Shlomchik, W. D., Jegathesan, M., Liebowitz, D., Abrams, C. S., Kadesch, T., Dancis, A., and Emerson, S. G. (2000) *J Exp Med* **192**, 1479-1490

203. Tanner, M. J. (1993) *Semin Hematol* **30**, 34-57

204. Anderson, K. P., Crable, S. C., and Lingrel, J. B. (2000) *Blood* **95**, 1652-1655

205. Anderson, K. P., Crable, S. C., and Lingrel, J. B. (1998) *J Biol Chem* **273**, 14347-14354

206. Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A. S., Weissman, I. L., and Katsura, Y. (2000) *Br J Haematol* **109**, 280-287

207. Iborra, F. J., Pombo, A., McManus, J., Jackson, D. A., and Cook, P. R. (1996) *Exp Cell Res* **229**, 167-173

208. Chakalova, L., Debrand, E., Mitchell, J. A., Osborne, C. S., and Fraser, P. (2005) *Nat Rev Genet* **6**, 669-677

209. Sutherland, H., and Bickmore, W. A. (2009) *Nat Rev Genet* **10**, 457-466

210. Liang, S. Y., Moghimi, B., Crusselle-Davis, V. J., Lin, I. J., Rosenberg, M. H., Li, X., Strouboulis, J., Huang, S., and Bungert, J. (2009) *Mol Cell Biol* **29**, 5900-5910

211. Zhou, Z., Li, X., Deng, C., Ney, P. A., Huang, S., and Bungert, J. (2010) *J Biol Chem* **285**, 15894-15905

212. Kim, S. I., Bultman, S. J., Kiefer, C. M., Dean, A., and Bresnick, E. H. (2009) *Proc Natl Acad Sci U S A* **106**, 2259-2264

213. Gilbert, D. M. (2002) *Curr Opin Cell Biol* **14**, 377-383

214. Simon, I., and Cedar, H. (1996) DNA Replication in Eukaryotic Cells. in *DNA Replication in Eukaryotic Cells* (DePamphilis, M. L. ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 387-408

215. Cimbora, D. M., Schubeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E. M., and Groudine, M. (2000) *Mol Cell Biol* **20**, 5581-5591

216. Parsons, G. G., and Spencer, C. A. (1997) *Mol Cell Biol* **17**, 5791-5802

217. Christova, R., and Oelgeschläger, T. (2002) *Nat Cell Biol* **4**, 79-82

218. De Brabander, M. J., Van de Veire, R. M., Aerts, F. E., Borgers, M., and Janssen, P. A. (1976) *Cancer Res* **36**, 905-916

219. Lee, J. C., Field, D. J., and Lee, L. L. (1980) *Biochemistry* **19**, 6209-6215

220. Palstra, R. J. (2009) *Brief Funct Genomic Proteomic* **8**, 297-309

221. Lewin, B. (1975) *Cell* **4**, 11-20

222. Bastos, R. N., and Aviv, H. (1977) *Cell* **11**, 641-650

223. Crawford, R. J., and Wells, J. R. (1978) *Biochemistry* **17**, 1591-1596

224. Ross, J., and Sullivan, T. D. (1985) *Blood* **66**, 1149-1154

225. Gaensler, K. M., Burmeister, M., Brownstein, B. H., Taillon-Miller, P., and Myers, R. M. (1991) *Genomics* **10**, 976-984

226. Gaensler, K. M., Kitamura, M., and Kan, Y. W. (1993) *Proc Natl Acad Sci U S A* **90**, 11381-11385

227. Strouboulis, J., Dillon, N., and Grosveld, F. (1992) *Genes Dev* **6**, 1857-1864

228. Gillemans, N., Tewari, R., Lindeboom, F., Rottier, R., de Wit, T., Wijgerde, M., Grosveld, F., and Philipsen, S. (1998) *Genes Dev* **12**, 2863-2873

229. Lee, C. H., Murphy, M. R., Lee, J. S., and Chung, J. H. (1999) *Proc Natl Acad Sci U S A* **96**, 12311-12315

230. Sengupta, T., Cohet, N., Morle, F., and Bieker, J. J. (2009) *Proc Natl Acad Sci U S A* **106**, 4213-4218

231. Johnson, K. D., Christensen, H. M., Zhao, B., and Bresnick, E. H. (2001) *Mol Cell* **8**, 465-471

232. Levings, P. P., Zhou, Z., Vieira, K. F., Crusselle-Davis, V. J., and Bungert, J. (2006) *FEBS J* **273**, 746-755

BIOGRAPHICAL SKETCH

Shermi Yen Liang was born in Raleigh, North Carolina, to Dr. Alan Yuh-Lin and Yen Liang, and spent the majority of her childhood in Albuquerque, New Mexico. There, she attended La Cueva High School and participated in school activities such as National Honors Society, Orchestra, and Creative Writing Club. In 2000, she graduated as one of the top 50 students in her class and went on to perform undergraduate studies at the University of Washington (UW) with the intention of majoring in biology. While there, she joined the first year of students to take part in a new exchange student program that formed as a collaborative effort between the UW and Sichuan University (SU). Within her first quarter at the UW, her participation in this program thrust her immediately into the realm of scientific research. As a result, she completed several undergraduate research projects in both botany and genome sciences, and spent her junior year improving her Chinese by studying abroad in Chengdu, China. There, she and along with other members in the UW-SU exchange performed a botanical field-study of the local plant biodiversity in a minority region of Sichuan province under the guidance of Dr. Richard Olmstead (UW) and Dr. Jiemei Xu (SU). After returning to the United States, Shermi joined the lab of Dr. Benjamin D. Hall (UW) and worked with students from the SU side of the exchange, performing genetic research on the phylogeny of various *Rhododendron* species. After graduating from UW in 2004 with a degree in cell & molecular biology, Shermi went on to perform doctoral studies at the University of Florida (UF) under the guidance of Dr. Jörg Bungert (UF), joining his lab in May 2006. She hopes to one day work in the field of forensic science.