# ANALYZING BETA-GLOBIN *CIS*-REGULATORY ELEMENTS BY USING ARTIFICIAL DNA-BINDING DOMAINS

By

JOEVA JADE BARROW

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

2013

To my wonderful family who has always provided a constant source of perspective and inspiration

ACKNOWLEDGMENTS

I would first like to thank my mother. Through her own strength, ambition and independence, she taught me to never back down from a challenge and to always dance. I would also like to express my deep love and gratitude to my husband who never wavered in his support and was always there for me in both the good and heart-wrenching challenging times. It is only because of his support and encouragement that I have made it here today. Eternal gratitude to my beautiful son. Motherhood just breaks you open in every way. I never knew I could love so much, be so tired, and cherish absolutely every single tiny experience. No matter what type of day I would have, his face always made me smile and kept everything in wonderful perspective.

I would like to thank my committee members especially to Dr. Mitchell Knutson who has mentored me since my master's studies. I know I can and always turn to him for advice. I thank my lab members both past and present (Drs. Shermi Liang, Zhuo Zhou, I-Ju Lin as well as Jared Stees, Pamela Chamales, Stephanie Morton, and our lab technician Blanca Ostmark) for their assistance, encouragement, and friendship. Finally, the most significant acknowledgement goes to my advisor Dr. Jorg Bungert for his mentorship. Jorg has always been eternally optimistic and has encouraged me over the years to always try to the best of my ability to stay positive even when things look bleak. He has taught me to limit my desire to strive for perfection and exquisite control and allow myself to drift so that imagination can take hold, after all that's where most of the discoveries are often made. I also thank him for his support especially where family matters are concerned. I truly believe that it is only because I had such a strong supportive network at work as well as at home that I was able to embrace my full

potential. There is no greater mentor alive. I admire his gentle spirit and his positive outlook on everything. I will always be grateful to him.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| BHLH | Basic Helix Loop Helix |
| ChIP | Chromatin Immunoprecipitation |
| CPSF | Cleavage PolyA Specificity Factor |
| DMSO | Dimethylsulfoxide |
| DNaseI | Deoxyribonuclease I |
| D.p.c | Days *post coitum* |
| DSIF | DRB Sensitivity Inducing Factor |
| HAT | Histone Acetyltransferase |
| HS | Hypersensitivity Site |
| KLF1 | Krüppel Like Factor 1 |
| LCR | Locus Control Region |
| LZ | Leucine Zipper |
| MEL Cells | Mouse Erythroleukemia Cells |
| NELF | Negative Elongation Factor |
| Pol II | RNA Polymerase II |
| PTEFb | Positive Transcription Factor b |
| RT-PCR | Reverse Transcriptase Polymerase Chain Reaction |
| TAL1 | T-Cell Acute Lymphocytic Leukemia Protein 1 |
| qRT-PCR | Quantitative (Real Time) Reverse Transcription-Polymerase Chain Reaction |
| USF | Upstream Stimulatory Factor |
| ZF-DBD | Zinc Finger DNA-Binding Domain |

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

# ANALYZING BETA-GLOBIN *CIS*-REGULATORY ELEMENTS BY USING ARTIFICIAL DNA-BINDING DOMAINS

By

Joeva Jade Barrow

May 2013

Chair: Jörg Bungert
Major: Medical Sciences--Biochemistry and Molecular Biology

One of the primary methods to regulate gene expression is through the interaction of *trans*-factors with *cis*-regulatory DNA elements. Transcription factors bind in a DNA sequence–specific manner and recruit activities that modulate the association and activity of transcription complexes at specific genes. Often, transcription factors belong to families of related proteins that interact with similar DNA sequences. Furthermore, genes are regulated by multiple, sometimes redundant, *cis*-regulatory elements. Thus, the analysis of the role of a specific DNA regulatory sequence and the interacting proteins in the context of intact cells is challenging. In this study, we designed and functionally characterized several artificial DNA-binding domains that neutralize the function of a putative *cis*-regulatory DNA element associated with the adult β-globin locus. These artificial proteins are comprised of zinc finger DNA-binding domains (ZF-DBD), each composed of six ZFs tailored to interact with desired sequences. I designed a series of artificial DNA-binding domains (DBD) comprised of modular zinc finger DNA binding domains (ZF-DBD) that targeted putative cis-regulatory sites of interests: the -90 and +60 elements relative to the start of the β-globin promoter. The ZF-DBDs bound to these sites with high affinity and specificity and rendered the

target sites inaccessible to endogenous transcription factors. I was able to successfully validate each target *cis*-element and assessed its requirement for globin gene function. Based on the success of our studies, we conclude that ZF-DBDs are effective tools to identify and characterize putative functional *cis*-regulatory elements *in vivo*

# CHAPTER 1
# INTRODUCTION

## Hemoglobin Function and Composition

Hemoglobin is the major protein constituent of red blood cells and functions together to reversibly bind and transport oxygen around the body [1]. In humans, hemoglobin is comprised of a pair of protein heterodimers derived from the alpha and beta-globin loci on chromosome 16 and 11, respectively. The alpha and beta-globin chains are arranged in a tetrahedral configuration each with a prosthetic ferroprotoporphyrin IX heme moiety embedded in a hydrophobic pocket that is covalently linked to a histidine residue [1]. The structure is stabilized by non-covalent bonds to allow for changes in configuration depending on the oxygen binding status of the protein [2]. Throughout development, different types of proteins expressed from the alpha and beta-globin gene loci combine to give rise to different forms of hemoglobin each having a specific set of characteristics and function throughout development [3].

## The β-Globin Gene Locus

The human beta globin gene locus is located on the short arm of chromosome 11 in humans (11p15.15) that is only active in erythroid cells [4]. The entire locus spans a ~80 KB region and is comprised of five distinct genes (5'-ε-Gγ-Aγ-δ-β-3') that are arranged in a manner consistent with their expression during development (Figure 1-1). All five genes encode for the varying forms of the globin protein (all about 140 amino acids) which is a critical subunit of the hemoglobin molecule (64 kDa) [3].

During the first 6 weeks of gestation in primitive erythroid cells, the ε-globin gene is the first to be activated and transcribed [5]. The ε-globin protein combines with the α-globin proteins to form embryonic hemoglobin (HbE- ζ2ε2). This takes place in the

embryonic yolk sac which serves as the primary site for hematopoiesis in the embryo [6]. As development progresses to the fetal stage, the ε-globin gene is silenced, and the two γ-globin genes (Aγ and Gγ) are expressed as erythropoiesis switches from the yolk sac to the fetal liver [7]. The γ-globin proteins pair with α-globin proteins to form fetal hemoglobin (HbF- α2γ2) [8]. After birth, there is a final switch in the site of hematopoiesis from the fetal liver to the bone marrow concomitant with the silencing of the fetal γ-globin genes. The adult β-globin gene is then activated to form the functional adult hemoglobin (HbA- α2β2). The δ-globin gene is also activated (HbA2- δ 2β2) but is expressed at levels 5% of β-globin due to a mutation in the promoter region [3].

The murine β-globin gene locus is located on chromosome 7 and is highly homologous to humans [9]. It contains four globin genes loosely arranged in the order in which they are expressed (5'-εγ-βH1-βmaj-βmin-3'). The order of the genes reflects developmental expression. The fetal gene βH1 is expressed in the yolk sac followed closely by the expression of εγ globin until approximately day E11.5 after which there is a single switch to the expression of the adult β-type globin genes βmaj and βmin which are co-expressed in the fetal liver and subsequently the bone marrow after birth [10,11].

Developmental stage-specific high level activation of all β-type globin genes both in humans and mice requires an upstream element located 6-22 kb away from the embryonic genes known as the locus control region (LCR) [12,13]. The LCR is a unique enhancer element that has been shown to also possess chromatin opening properties. It can confer high level expression of linked genes in a position-independent manner in transgenic assays [14,15]. Insight into the importance of this element was detailed by examining the classic Hispanic thalassemic patient who developed a de novo deletion

on a maternally inherited chromosome involving approximately 30 kb of sequences 5' to the epsilon gene that emcompassed the majority of the LCR. This resulted in the heterochromatization of the globin locus and silencing of all the downstream genes [16]. Subsequent work showed that the LCR is a powerful DNA regulatory region characterized by five DNase I hypersensitive sites (HS) (in mice there are six HS) [13,17]. Sensitivity to the enzyme DNase I often denotes areas of strong regulatory potential because it indicates that the region of chromatin is open and therefore accessible to *trans*-acting factors. Each HS contains a core region spanning 200-400 bp that is enriched for *cis*-elements such as GATA and MARE motifs that serve as binding sites for regulatory *trans*-acting factors—both ubiquitous and erythroid-specific, that mediate the sequential activation of the globin genes [17]. HS 1-4 exhibit erythroid-specific hypersensitivity to the enzyme DNase I [18]. The current function of HS1 is unknown while HS2 is the only element that functions as a classical enhancer meaning that it can activate heterologous genes in transient expression assays [19,20]. HS3 and 4 also display enhancer activities; however, they require chromatin integration before this property can be observed [21]. HS3 has also been shown to have chromatin opening activities [22]. HS5 has been reported to possess insulator activity in enhancer blocking assays in K562 and murine erythroleukemia (MEL) cells and has been postulated to have the ability to prevent the spread of heterochromatic marks from the upstream olfactory genes that is maintained in a closed chromatin configuration in erythroid cells [23]. The exact mechanism of this process as well as how the activation of the globin genes in a sequential manner is achieved is not completely understood and is currently an area of study.

## Hemoglobinopathies

Regulation of the globin genes is very complex and disruption to the system can lead to hemoglobinopathies. It is estimated that about 7% of the world population are carriers for one of the inherited hemoglobin disorders, making them the most common monogenic diseases in the world [24]. Hemoglobinopathies can be divided into two groups, structural hemoglobin variants such as sickle cell anemia and the thalassaemias, which result from imbalanced and defective synthesis of the globin chains.

### Sickle Cell Anemia

Sickle cell anemia (SCA) was first characterized as a molecular disorder by Linus Pauling and colleagues in 1949 to occur due to an abnormality in the hemoglobin molecule. It is currently classified as the most common inherited genetic disease based on the Newborn Screening Initiative [25,26]. The disorder occurs due to a genetically inherited A→ T mutation at residue 334 in the beta-globin gene that causes a single amino acid change at position 6 of the β-globin protein chain [27,28]. The result is a substitution the normal polar glutamic acid for a mutant non-polar valine. This is a significant problem because upon de-oxygenation of hemoglobin to deliver oxygen to the peripheral tissues, hemoglobin undergoes a conformational change which exposes the substituted hydrophobic valine on the molecular surface resulting in the rapid polymerization of the HbS molecules into long fibers. This causes the red blood cells to adopt a crescent or “sickle” shape resulting in both the inability to efficiently bind oxygen as well as microvascular occlusion leading to tissue ischemia [25]. Individuals afflicted with this disease face a life-long experience of chronic pain, organ failure due to iron overload, anemia, splenic sequestration, and an increased risk for stroke. In the

heterozygote state in which an individual only inherits 1 copy of the mutant allele, the patients are asymptomatic and are known as carriers, or to have a sickle cell trait. The carrier state for SCA is asymptomatic and has been shown to be protective against malaria and as such SCA is concentrated in those countries where malaria has exerted selective evolutionary pressure, such as Africa [25] [29]. According to the National Heart Lung and Blood Institute (NHLBI), the current national statistic estimates the probability to be 1 in every 500 African Americans that have the full blown disease and 1 in every 12 African Americans that have the sickle cell trait [26].

**Thalassemia**

The other major disorder is thalassemia which is an inherited disorder characterized by the absence of or reduction in levels of one or more globin chains. Thalassemia is the most common monogenetic disease worldwide. The disorder can be classified into two major groups, beta and alpha thalassemia depending on the globin chain that is affected [24].

Beta-thalassemia occurs due to various point mutations or deletions within the β-globin gene that result in complete dysfunction and/or decreased synthesis of the β-globin protein [24,30]. The main cause of anemia in β-thalassemia is imbalanced globin chain synthesis and the deleterious effects of excess α-globin chains on erythroid maturation and survival [31]. Alpha-globin chains are unable to form a functional hemoglobin tetramer and precipitate as inclusion bodies in red cell precursors that accumulate in the bone marrow as well as all other stages of the erythroid maturation pathway [32]. In the most severe forms of the disease (thalassemia major), patients present with failure to thrive, anemia, chronic infection, and bone abnormalities due to the expansion of the bone marrow [33,34,35]. In milder cases (thalassemia minor),

patients present with few, if any clinical symptoms. In thalassemia intermedia, the pathogenesis of the patient is variable depending on the individual case and the extent of β-globin disruption [31].

**Therapies for Hemoglobinopathies**

Hematopoietic stem cell transplantation represents the only current long term cure for hemoglobinopathies [36]. The process involves depleting a patient's diseased bone marrow by radiation or chemotherapy and then introducing hematopoietic stem cells from a healthy donor by filtration into the patient's bloodstream. Donor stem cells then migrate to the bone marrow and repopulate the region. Patients after transplantation require lengthy hospital stays where they are given immunosuppressive drugs to minimize rejection of the foreign cells [37]. Stem cell transplantation is extremely promising yet still has significant risks—including death [38]. Case study reports indicate that transplantation with matched donors still prove to have too high of a mortality rate to be acceptable. Yet despite these risks, it still remains the only current cure for the disease.

Other forms of therapy are aimed at relieving symptoms, preventing complications, and avoiding crisis (in the case for sickle cell disease). These therapies such as blood transfusions, and/or the administration of oral medications to increase fetal globin synthesis, require lifelong application for effectiveness.

Blood transfusions involve obtaining healthy red blood cells from a donor and delivering them to the patient intravenously. However, iron is deposited along with the blood and with each successive blood transfusion, iron levels saturate and exhaust the transferrin carrying capacity and lead to chronic iron overload [39,40]. Free non-bound iron accumulates in the plasma and tissues of the body where it then reacts with oxygen

generating free radicals which causes damage particularly in heart and liver [33,41,42]. As there is no effective excretory pathway for iron apart from intestinal shedding and bleeding, iron-chelation therapy such as the oral medication deferasirox is mandatory for these patients [43]. The daily administration requirement of the drug however is very expensive and is not available to many patients worldwide.

Oral medications, the most common of which is hydroxyurea, are targeted towards increasing levels of fetal hemoglobin (HbF) in the blood [44]. Although the natural distribution of HbF cells in an adult is approximately 1-2%, studies have shown that these cells are able to resist sickling *in vitro* and inhibit polymerization of sickled hemoglobin [45,46]. Furthermore, sickle cell patients that genetically have higher levels of HbF—a phenomenon known as hereditary persistence of fetal hemoglobin (HPFH), often have milder clinical pathogenesis [47].

Hydroxyurea is a cytotoxic agent originally used for decades to reduce abnormally high hematocrit and platelet counts for patients with polycythemia vera—a disorder where the bone marrow synthesizes too many red blood cells [48]. It wasn't until a clinical trial testing the effects of hydroxyurea in non-human primates that it was discovered that hydroxyurea increases fetal hemoglobin [49]. Although the details regarding the mechanism of action remain unclear, data suggest that the cytotoxic nature of the drug targets HbS cells because they are derived from highly proliferative precursors. These precursors are targeted by the drug and are subsequently destroyed which will boost the relative percentage of HbF cells which stem from precursors that do not divide as rapidly [50]. Clinical trials demonstrated that hydroxyurea could increase HbF levels in patients with sickle cell disease as early as 72 hours of administration

[51,52]. Hydroxyurea also stimulates NO production which improves vasodilation and minimizes crisis in sickle cell patients [53]. Although this drug has extremely promising outcomes, there are significant complications as well. Hydroxyurea is a myleosuppressive agent and white blood cell levels often drop, requiring patients to withdraw from treatment until levels return back to baseline. Low white blood cell counts pose an increased risk for infection and long-term use of the drug may lead to leukemia as well as other types of cancers [54]. Furthermore, hydroxyurea while effective for patients with sickle cell disease is not as effective for patients with thalassemia.

Based on the limitations of these treatments, there is a strong need to either improve upon or to develop novel therapies. A proposal that has received much interest as a potential therapeutic cure is to consider re-activating the expression of the fetal γ-globin genes that are normally silenced in the adult. In this way the γ-globin proteins could replace the dysfunctional β-globin protein in hemoglobin and oxygen-carrying-potential and red blood cell morphology can be restored [55].

Several gene therapy trials have been attempted but delivery methods are a significant limiting factor and need to be improved. Retroviral vectors which integrate permanently into the host genome have exhibited complications such as vector instability, low titers over time, and variable expression [56,57]. Self-inactivating lentiviral vectors that can integrate into non-dividing cells have also shown some promise. But there remains the risk that the virus can integrate into a region of DNA that can activate some oncogenic potential [58]. Therefore, it is crucial to design vectors which are lineage- and differentiation stage-restricted. The incorporation of tissue-specific promoters and enhancers in addition to genetic elements with enhancer-

blocking properties will aid in reducing the toxicity and increasing the efficiency and stage- and tissue- specific expression of the β-globin gene and preliminary studies have demonstrated promise [59].

The need therefore to understand the intricacies of globin gene regulation is of the utmost importance so that new therapies can be developed.

## Chromatin

Chromatin organization mediates the dynamic and structural compaction of genetic information within the nucleus of an eukaryotic cell. The fundamental units of packaging, commonly known as “beads on a string”, are repeating structural elements known as nucleosomes which consist of a nucleosomal core particle, linker DNA, and the histone linker protein H1 [60]. The core particle is composed of two copies of histone proteins H2A, H2B, H3, and H4 that are arranged in an octameric configuration [61,62,63]. Approximately 146 bp of DNA interacts with the superhelical ridge structure of the nucleosome that is stabilized by electrostatic interactions [63]. Connecting two nucleosomal particles is the linker DNA which varies between different the organisms, ranging from 10-60 bp. Bound to the linker DNA is histone protein H1 which, while not required for the formation of the nucleosome, is essential for higher level structures [64,65]. Successive compaction of the chromatin fiber initially goes through a 30 nm fiber intermediate, which has been proposed to consist of a two-start helix solenoid structure, followed by condensation into higher order looping chromatin structures [66,67,68,69].

Chromatin can be broadly categorized as heterochromatin or euchromatin. In general, heterochromatin is condensed and therefore not accessible for transcription factor interactions typically resulting in the suppression of the genes associated with that

region although, transcription of some genes has been reported to occur in the context of heterochromatin [70]. Euchromatin represents a more open structure, allowing *trans*-acting factors to bind to regulate gene expression [71]. Relative compaction of the genome can be assessed by various methods of which the most classical is the endonuclease deoxyribonuclease I that makes single strand cuts along the phosphate backbone of DNA [72]. Regions of the genome are described based on the interaction with DNAseI and can be insensitive as is the case with heterochromatic regions since the DNA is inaccessible at those sites, or sensitive with a gradient ranging to hypersensitivity [73]. Hypersensitive regions are thought to be devoid of nucleosomes and contain many *cis*-regulatory elements that bind to *trans*-acting factors [74].

As the accessibility of chromatin represents a critical area of control for gene expression, it is therefore essential that chromatin be dynamic. As such, there are several processes to alter the accessibility of chromatin including ATP-dependent chromatin remodelers, and histone modifications [75,76].

## Gene Transcription

Transcription is the process of transcribing DNA to generate a complementary strand of RNA which in general is then used as a template for translation into a functional protein. Transcription is mediated through the action of the RNA polymerase as well as a host of accessory proteins. There are three major RNA polymerases in higher order mammals. RNA polymerase I (Pol I) transcribes the ribosomal RNAs involved in ribosome synthesis and enzymatic activity. RNA polymerase II (Pol II) transcribes the majority of messenger RNAs which typically result in a functional protein while RNA polymerase III (Pol III) transcribes small RNAs and tRNAs used in translation or splicing machineries [77,78,79,80]. Active transcription requires the recruitment of the

polymerase as well as the general transcription factors TFII-A, -B, -D, -E and -F which through coordinated mechanisms form what is known as the pre-initiation complex (PIC) [81,82]. DNA is melted by the helicase activity of TFIIH and the C-terminal tail (CTD) of the polymerase—a 5'-YSPTSPS-3'heptapeptide that is repeated 52 times in humans and can be phosphorylated at distinct serine residues [83,84,85]. Phosphorylation of the fifth serine by the kinase domain of TFII-H promotes transcription initiation [85]. After a few short abortive transcripts, DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF) associates with the serine 5 phosphorylated form of Pol II to halt the polymerase until the capping reaction, which introduces a 7 methyl guanosine at the 5' end of the nascent RNA, can take place [86,87,88]. Positive transcription-elongation factor b (P-TEFb) is then recruited to phosphorylate both DSIF, causing the dissociation of NELF, and the serine 2 position of the CTD of Pol II [89]. Pol II then becomes processive and engaged in active elongation. The RNA is then processed by the transcription termination factor cleavage polyA specificity factor (CSPF) upon the recognition of the polyadenylation sequence [90].

### *Cis*-Regulatory DNA-Elements within the Globin locus

*Cis*-regulatory DNA elements contain binding sites which recruit *trans*-acting factors that modulate transcription. There are a large number of *cis*-regulatory elements in the beta-globin gene locus that can be characterized broadly as promoters, enhancers, insulators, or locus control regions. Promoters are in a position immediately 5' or overlapping the transcription initiation site, while enhancers, insulators and LCRs are distal regulatory elements [14]. An enhancer is defined as having the ability to augment the expression of a gene, exert its function in a position and orientation-independent manner, and be able to interact with heterologous promoters. Locus

control regions are similar to enhancers in that they augment transcription, but they are typically tissue-specific and are able to provide copy number-dependent and position-independent expression to linked genes [14,91]. Insulators are sequences that recruit proteins that either block the function of enhancers or provide a boundary between open and closed chromatin. All of the above described *cis*-regulatory elements are present in the beta-globin gene locus and as an example of the breath and diversity of *cis*-elements, I will highlight elements surrounding the human beta-globin gene.

The *cis*-elements proximal and within the beta-globin gene include: two erythroid transcription factor GATA binding sites located at -200 and -120 relative to the transcription start site (TSS) that have been shown to increase transcription [92,93]. At position-90 resides a CACCC box that has been shown to interact with Krüppel Like Factor 1 (KLF1) [94,95]. This region has been well characterized and demonstrated to be essential in humans due to a patient harboring a C→T mutation at position -88 relative to the TSS which resulted in thalassemia due to the inability to bind KLF1 [95,96]. The site in mice however is not characterized and will be one of the objectives of this study. The CCAAT box at -70 also binds to GATA1 and is a positive regulator of globin gene expression [97]. Position -25 holds a non-canonical TATA box sequence 5'-CATAAA-3' that is believed to function in a coordinated fashion with the initiator—a pyrimidine rich sequence located at +1 of the TSS to recruit the TFIID complex of the PIC [98,99]. Downstream there are two conserved EBOX sequences (+20 and +60) that interact with helix-loop-helix proteins such as TAL1, USF1/2, and TFII-I heterodimers that, depending on the cell type, would decrease or increase transcription, respectively [100]. Mutagenesis studies *in vitro* and in vivo indicated that the +60 EBOX is important

for globin expression. A MARE *cis*-element is located at +24 and has been shown to interact with the transcription factor NF-E2 and regulates beta-globin transcription in a positive manner [101]. Finally, there is a tissue and developmental stage-specific 3' enhancer located approximately 500 bp downstream of the polyadenylation signal of the β-globin gene. This enhancer contains four binding sites for GATA1 and is important for β-globin expression [98,102].

The β-globin LCR as mentioned previously 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. It is comprised of several erythroid specific DNase I hypersensitive sites (HS) each spanning approximately 200-400 bp with each HS insulated by 2-4 kb of flanking DNA [103]. These HSs contain several putative and confirmed binding sites for various regulators and co-regulators. As a representative example, LCR HS2 contains binding sites for the hematopoietic specific proteins GATA-1, KLF1, NF-E2, TAL1, as well as the ubiquitously expressed transcription factor USF (Figure 1-2) [104]. These transcription factors have been shown to be involved in LCR-mediated activation of adult β-globin gene transcription and it is hypothesized that these *trans*-acting factors co-localize to the LCR to be subsequently transferred to the appropriate gene for transcription initiation. Several mechanisms for this interaction have been postulated, such as tracking and looping models [105,106].

**Structural properties of DNA-Binding Proteins**

There is an abundance of *trans*-acting proteins that interact with the *cis*-elements within the globin locus yet they all contain similar structural motifs that allow these proteins to bind to DNA and dimerize with other proteins. These motifs include but are not limited to: the zinc finger (ZF) domains, the leucine zipper (LZ) and the basic helix-

loop-helix (bHLH) domain which are primarily structural motifs involved in protein-protein interactions but they also contain basic regions to allow for DNA-interactions [107]. These protein domains are not mutually exclusive; there are many *trans*-acting factors that contain several of these domains. For example, the transcription factors cMyc and USF1/2 contain both LZ and HLH domains. Listed below are the detailed properties of each structural domain.

**Leucine Zipper**

The leucine zipper is a common motif found in dimeric proteins. It was first discovered by McKnight et al. based on protein sequence alignments and secondary structure predictions [108]. The hallmark of the motif is the heptad a repeat of 7 amino acids designated (abcdefg)n that is arranged in a coiled-coiled structure spanning a 35 amino acid stretch [109]. Positions a and d are normally occupied by a leucine residue which provides the major portion of the hydrophobic dimerization domain and is critical for the packing of the oligomerization interface [110]. The leucines of one helix interdigitate with those of the parallel helix in a non-covalent linkage causing the two regions to dimerize. Residues e and g are involved in maintaining this hydrophobic core interface as well, while the remaining residues: b, c and f are charged polar moieties that ensure the solubility of the protein [111,112]. The two most common leucine zipper motifs are the basic leucine zipper (bLZ) and the basic helix-loop-helix leucine zipper (bHLH-LZ) which are found in many eukaryotic transcription factors such as the upstream stimulatory factor (USF) [113]. The basic region interacts with the major groove of DNA through hydrogen bonding allowing leucine zipper-containing transcription factors to recognize and bind to specific DNA motifs.

**Helix-Loop-Helix Domain**

The helix-loop-helix domain is common in many eukaryotic transcription factors that regulate a large array of cellular processes ranging from differentiation to cellular proliferation. The motif was first identified by Murre et al. in the E12 and E47 murine transcription factors and spans approximately 60 amino acids composed of a basic region that mediates DNA binding and two successive amphipathic alpha helices juxtaposed at a 90 degree angle separated by a variable loop [114,115]. Proteins that contain the HLH motif are generally dimeric proteins due to the hydrophobic interactions of the alpha helices which align the basic region to bind to palindromic CANNTG sequences, known as the E-box, in the major groove of DNA through hydrogen bonding [116,117].

**Zinc Finger Domain**

The zinc finger domain is composed of 30 amino acids with two antiparallel β-pleated sheets and an alpha helix fold that is structurally stabilized by both hydrophobic interactions and the tetrahedral coordination of a zinc ion by the conserved cysteine and histidine residues (Figure 1-3). In the most common ZF domain two cysteines and two histidines coordinate the Zn-atom (Cys2-His2 class) [118]. The Cys2-His2 ZF-domain was first discovered in the Xenopus *laevis* transcription factor TFIIIA that contained nine tandem repeats of the zinc finger motif [119,120]. Since then, many other proteins were found to have a similar configuration such as the murine early response protein ZIF268 and human Sp1 [121]. To date, the Cys2-His2 zinc finger domain represents the most common binding motif found in DNA binding proteins in eukaryotes. ZF-proteins are the second most frequently encoded protein in the human genome [122]. Each 30 amino acid zinc finger domain can recognize and bind to 3 bp of B form-DNA through

hydrogen bonding [118,123]. This, in addition to the fact that these domains can be arranged as covalent tandem repeats allows for the recognition of extended asymmetrical DNA sequences. Crystal studies of the murine transcription factor Zif268 detail that base-specific interactions are achieved by hydrogen bonding between the amino acids present on the α-helix reading head and the DNA (Figure 1-3) [118,124]. The α-helix reading head is inserted into the major groove of B form-DNA and contacts are primarily made with the amino acids located at positions -1, 3 and 6 with respect to the start of the α-helix [125]. These contacts are made in an anti-parallel fashion so that the amino acids of the α-helix reading head makes base-specific contacts with the 3', middle, and 5' nucleotides of the 3bp DNA segment (Figure 1-3). One of the β-pleated sheets also makes contacts with the phosphodiester oxygen of the sugar phosphate backbone which further stabilizes the binding through interaction with the first of the conserved histidines (that also coordinate the zinc ion) and a conserved arginine [118,126].

## *Trans*-Acting Factors

### Krüppel-Like Factor 1

Krüppel-Like Factor 1 (KLF1) is a small 3-zinc fingered-transcription factor whose expression is restricted to the erythroid lineage [95]. It belongs to a family of transcription factors called the Krüppel-like factors, which derives its name from the high degree of similarity to the *Drosophila* transcription factor, Krüppel. KLF1 was originally identified by a subtractive erythroid hybridization screen using J774 macrophages as a control with MEL cells as the test sample to determine which genes were erythroid specific [95]. Binding studies indicated that KLF1 occupies 5'-CCACACCCT-3' motifs (CAC box) [127]. Shortly after its identification, its significance in hematopoiesis was

established from patients harboring mutations in a KLF1 binding site, the -87/88 CAC box upstream of the beta-globin promoter [128]. These patients had reduced levels of β-globin expression and developed thalassemia. Studies in knock- out mice subsequently supported the critical role of KLF1. Homozygous null mice for KLF1 are normal during the embryonic stage (10.5 d.p.c) however, when hematopoiesis switches from the yolk sac to the fetal liver, embryonic lethality is observed at 14.5 days post coitum (d.p.c.) due to severe anemia resulting from ineffective erythropoiesis [129]. The fetal livers of these embryos are pale, and they fail to express high levels of adult β-major transcripts. This suggests that KLF1 is critical for mechanisms involved in globin switching [130].

Functionally, KLF1 has many defined roles in hematopoiesis. It is the most developmentally specific gene and is crucial for human β-globin expression [131]. It activates β-globin transcription in a sequence-specific manner and is believed to be involved with globin switching because under-expression of KLF1 in heterozygous mice results in delayed switching (alters the γ/β ratio) and overexpression results in premature switching [132,133]. KLF1 is also important for the silencing of the fetal globin genes. In KLF1 null mice, levels of γ-globin expression are elevated 5 fold compared to control mice [130]. This occurs as a result of the role of KLF1 in the regulation and activation of the BCL11A gene, which encodes a potent silencer for the embryonic genes. In KLF1 null mice, KLF1 is not present to activate BCL11A and as a result, the fetal globin genes are not efficiently silenced. This mechanism is currently an attractive target for therapy for hemoglobinopathies [134].

ChIP studies demonstrated that KLF1 binds to LCR elements HS2 and HS3 in addition to the beta-globin promoter where it is postulated to assist with chromatin-

opening activity [135,136]. This hypothesis was formed due to the association of KLF1 with histone acetyltransferases (HATs) p300 and CBP as well as the ATpase of the SWI/SNF chromatin remodeler complex BRG1 [137,138,139]. The chromatin-opening activity of KLF1 is further supported by the observation that disruption of EKLF results in a reduction in DNase I hypersensitivity in the β-globin promoter and the HS2 and HS3 of the LCR [131]. The chromatin opening activity is likely mediated by KLF1 mediated recruitment of ATP-dependent chromatin remodeling complexes [138]. Furthermore, KLF1 has also been shown to interact with components of the basal transcription machinery, including the TBP associated factor 9 (TAF9) [140]. Chromatin Capture Conformation (3C) studies also indicates that KLF1 is required for mediating proximity between the LCR and the adult β-globin gene promoter [141]. In cells and mice with disruption of KLF1, the looping association between beta-globin and the LCR is not observed. All these studies illustrate that KLF1 binding to the -90 CACCC site is crucial for the recruitment of transcription complexes to the adult β-globin gene promoter.

**Upstream Stimulatory Factor**

USF is a ubiquitously expressed transcription factor that was first identified as a cellular activity that binds to an upstream element of the adenovirus major late promoter and stimulated transcription *in vitro* [142,143]. DNaseI footprinting identified the consensus USF DNA-binding CANNTG (E-box) motifs and its binding has been associated with the transcription of many cellular and viral genes [144]. USF belongs to a family of transcription factors characterized by the basic helix-loop-helix leucine zipper (bHLH-LZ) DNA binding domains. The protein was purified to homogeneity and the activity was shown to be contributed by two polypeptides named: USF1 (44 kDa) and USF2 (43 kDa) [145,146]. The predominant form of USF is a USF1/USF2 heterodimer,

although homodimers were also found in different cell-types [147]. Structurally, dimerization and DNA-binding activity of both proteins is mediated by the highly homologous (>75% sequence identify) C-terminal region composed of a USF specific region (USR), a basic region that mediates DNA-binding, followed by the HLH-LZ region required for dimerization. The N-terminal region is divergent between the two proteins and could contribute to the differences in transcription activity of the two proteins [146,148]. Indeed, USF1 or USF2 null mice, despite being both viable and fertile, display very different phenotypes. The USF1 null mouse showed slight developmental abnormalities and had high levels of USF2—possibly a compensatory mechanism. Conversely, the USF2 null mouse had extreme growth defects and had low levels of USF1. The double knock out of USF1 and USF2 is embryonic lethal [149].

Despite the fact that the USF proteins are ubiquitously expressed, they appear to primarily regulate genes that are expressed in a differentiation and tissue-specific manner [150]. Most genes activated by USF are expressed at high levels in differentiated cells, including the β-globin gene. A genome-wide mapping study of USF interaction sites in hepatocytes revealed that USF preferentially binds DNA in close proximity to transcription start sites, and that this interaction correlates with increased levels of H3 acetylation suggesting that USF may be involved in the recruitment of transcription complexes [151].

**USF in β-Globin Gene Regulation.**

Previous studies have shown that USF interacts with conserved E-box elements located in Hypersensitive Site 2 (HS2) of the LCR, as well as in the adult β-globin downstream promoter region [152,153]. USF interacts with co-activators and histone modifiers such as the HAT CBP/p300, as well as with large co-regulator complexes

such as the histone methyltransferases PRMT1 and SET1in erythroid cells, suggesting that it functions through chromatin remodeling and RNA Pol II recruitment [154,155]. 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 [156]. 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 [157].

**T-cell Acute Lymphocytic Leukemia Protein 1**

T-cell acute lymphocytic leukemia protein 1 (TAL1) is also known as stem cell leukemia (SCL). Its name was derived by the observation that TAL1 is highly aberrantly expressed in approximately 30% of patients with T cell acute lymphoblastic leukemia (T-ALL) [158,159]. TAL1 is a 47 kDa helix-loop-helix protein that heterodimerizes with ubiquitously expressed E proteins (E12/E47) and binds to E-box sequences of which the preferred sequence is CAGATG, though additional preferences have been observed as dictated by the associated E protein [160]. In erythroid cells, it is a part of a large multimeric protein complex comprised of TAL1/E2A/LMO2/GATA1 and LDB1 that binds to composite GATA-EBOX sequences [161].

Despite its association with leukemia, the normal function of TAL1 has not been completely resolved. TAL1 possesses activator as well as repressor capacities depending on cell type and associated proteins [162]. Gene knock out studies established that TAL1 is critical during embryonic development. Erythropoiesis in the yolk sac is absent in TAL1 null mice due to the failure to produce hematopoietic progenitor cells. Embryonic lethality is observed at 9.5 d.p.c [163]. In the adult, TAL is expressed in hematopoietic stem cells (HSC) and multipotent progenitors of the

erythroid, myeloid and mast cell lineages [164]. Knock outs in these systems result in lack of burst forming units (BFU-E) and colony forming cells (CFC-Mk) [165]. TAL1 has been postulated to interact *in vitro* and *in vivo* with an E-box located in the LCR element HS2 as well as the β-globin gene promoter and the interaction with NL1/Ldb1 is required for conformational changes that bring the LCR and the adult β-globin gene into close proximity [166,167]. However, further studies are required to detail the complete mechanism of TAL1 mediated regulated beta-globin gene expression.

## Redundancy of *cis*- and *trans*-Regulatory Components of the β-globin Gene Locus

Given that the majority of hemoglobinopathies result from DNA alterations within the β-globin gene locus, investigators over the years have focused on examining *cis*-elements to identify DNA regions that carry strong regulatory potential and are essential for globin gene function. Employing traditional techniques such as DNaseI hypersensitivity and targeted site disruptions, many *cis*-regulatory DNA elements were identified such as the LCR and binding sites for several *trans*-acting factors. Although significant progress has been made to identify these elements, numerous studies detailing *trans*-factor binding motifs need to be refined in order to identify DNA sequences that truly serve as functional binding sites. This often presents a challenge given the redundant nature of many *trans*-factor binding motifs [168]. Furthermore, there remain many potential *cis*-regulatory elements that have yet to be discovered.

## Zinc Finger Artificial DNA Binding Domains

An alternate approach to evaluate the importance of potential *cis*-regulatory elements *in vivo* is to engineer and employ artificial DNA-binding domains (DBD) composed of zinc finger motifs that will bind with high affinity and specificity to a region

of interest on DNA [169]. Once targeted, the DBD will occupy that region and render the *cis*-element inactive by preventing other *trans*-acting factors from binding. This will allow the regulatory potential of the *cis*-element to be evaluated.

As mentioned previously, a single zinc finger domain can bind to and recognize 3 base pairs of DNA in a sequence-specific manner [118]. What became immediately attractive was that zinc finger domains possess a distinct modular organization and it was postulated that it may be possible to mix and match domains to allow the recognition of extended asymmetric DNA sequences [170]. The potential of ZF proteins to serve as a backbone for the generation of artificial DNA binding domains has early on been recognized by Aaron Klug and colleagues [169]. This was taken a step further by the Pabo group and others who used phage display technology to express varying forms of the ZIF268 zinc finger protein on the filamentous phage (fd). The authors mutated the 4 key amino acids of finger 1 of ZIF268 at positions -1,+2,+3, and +6 relative to the amino terminus of the alpha helix that were shown to make base-specific contacts based on crystal studies. The mutations replaced those sites with amino acids that were randomly generated. Successful alternate ZIF268 proteins were created that could recognize additional DNA sites with high affinity by simply re-arranging the key amino acids [170]. The study served as a platform that inspired several groups to attempt to derive a one to one amino acid-nucleic acid code. The Pabo group improved on the selection process slightly by creating a “Sequential Selection Strategy” that introduced amino acid substitutions to select fingers but in a context-dependent manner [171]. Studies were also performed varying the zinc finger backbone and a zinc finger consensus sequence. The resulting protein Sp1C showed the most promise [172].

Sp1C is derived from the DNA binding domain of human Sp1, a zinc finger protein belonging to the Cys2His2 class that binds to GC boxes 5'-GGGCGG-3'. The Sp1 sequence was modified with "consensus" amino acids that occurred the most frequently from a database of 131 zinc finger domain sequences. The rational was that amino acids involved in stabilizing the zinc finger would be present in relatively high frequency. Overall the differences between the two proteins encompass 25 point mutations and 4 residue deletions [173].

Studies confirmed that Sp1C was able to adopt 3D structures that are similar to its natural counterpart Sp1 based on NMR and DNA binding studies. Furthermore, Sp1C was noted to bind more tightly to its consensus sequence (lower Kd) compared to Sp1 and binding to zinc was also discovered to be stronger. The authors inferred that increased metal ion binding is reflective of more structural stability [172,174].

And thus a "minimalist" zinc finger protein that could be tailored to recognize a desired sequence with high specificity was created. Sp1C is composed of an N-terminal backbone that is composed of two β-strands and associated sequence (amino acid sequence: YKCPECGKSFS) and a C-terminal backbone comprised of the C-terminal portion of the α-helix (amino acid sequence: HQRTH). It also includes a single fixed conserved N-terminus portion (amino acid sequence: LEPGEKP) and C-terminus (amino acid sequence: TGKKTS) to structurally complete the protein. Finally, a linker sequence with amino acid sequence TGEKP fuses the individual zinc finger domains in an array [175,176].

The zinc fingers designed in this work are based on the Barbas modules. The Barbas group successfully designed, expressed, and characterized ZF proteins with 6

contiguous ZF domains that will bind to 18bp of DNA. This target size is sufficient to generate a unique signature in the genome ($3.0x10^9$bp) [175]. The method involved synthesizing two 3-zinc finger proteins based on the Sp1C sequence and linking them together using a linker sequence TGEKP (described above) [176]. The linker, termed the Krüppel type linker peptide, was selected because the pentapeptide sequence represented the consensus sequence that was most commonly found in linking zinc finger domains [177]. Furthermore, extensive computer modeling demonstrated that this linker sequence was sufficient to maintain positioning and H-bond characteristics that are observed in natural zinc finger proteins. Computer modeling also determined that the 6-ZF proteins remain bound to the major groove of DNA and follow it for two full helical turns. Electrophoretic mobility shift assays (EMSA), DNA footprinting, and *in vivo* reporter studies confirmed these results and demonstrated that both the ZIF268 and Sp1C based 6-ZF proteins were able to bind to their target sequences with enhanced affinity (68-74 fold) compared to their 3-finger counterparts [174,175]. Interestingly, the *in vivo* results gave higher affinities (300 fold) than *in vitro* and the authors attribute this to the unstable nature of the proteins *in vitro*. Naturally occurring three-finger proteins, such as Zif268 and SP1, bind their preferred sequences with Kd values between 10nM and 10 pM [178]. Observed DNA affinities tend to improve as the number of fingers increases from one to three, but affinity plateaus beyond three fingers, and only modest improvements in affinity (~70-fold) are seen for six-finger proteins over corresponding three-finger proteins [125]. Zinc fingers have also been successfully fused to a wide array of functional domains such as transcription activation or repression domains, endonucleases, and integrase etc. Remarkably in each case, the DNA binding

properties of these designed proteins were not altered by adding extra domains on either the N-terminus or C-terminus of the protein [125].

To date, 49 ZF domains have been experimentally characterized and validated in the context of polydactyl proteins and are used for protein design by the Barbas laboratory. The zinc finger proteins recognize all 16 GNN DNA triplets, 15 ANN triplets (ATC not represented), 15 CNN triplets (CTC not represented), and only a few TNN triplets (TGA, TGG and TAG) giving rise to a large variability in target site selection [179,180].

### Challenges with Artificial DNA-Binding Domains

The artificial DNA binding domains based on zinc finger technology is not without limitations. Although investigators have been able to design fingers with exquisite specificity, the composition of the target is of extreme importance. Zinc finger proteins possess a stronger affinity and specificity for GNN triplets. Therefore, sequences enriched in GNN are desired. As the composition of the target sequence changes to be less GNN rich, the specificity will decrease slightly as well [181]. Some ZF domains can also recognize a four-base subsite—termed target site overlap (TSO) [182]. This results when an aspartate residue is located at position +2 of the α-helix making a contact outside of the targeted triplet. There are now databases that flag potential TSO so that alternate sites can be selected [176]. Another common critique of the system is that as additional zinc finger domains are introduced, especially in the case of 6-zinc fingered proteins, the specificity towards the target site can be altered. This is particularly problematic when functional domains are fused to zinc finger proteins that could potentially give rise to unwanted activities. The exact mechanism as to how this could occur is unclear, however a possible explanation could be that the increased number of

contacts in the 6-finger protein elevates the binding energy to a point where individual residue:base mismatches are insufficient to prevent binding [181].

Due to the possible off-target effects, validation of specificity and affinity for these proteins becomes of extreme importance. In the current study, no functional domains were used to reduce the possibility of off target activity. It is important to also consider that the binding sites in the genome will be somewhat lower than the theoretical total because many of the sites will be inaccessible due to chromatin structure. Furthermore, given that less than 1% of the human genome is coding, most binding sites will occur in regions that should not affect the transcription of any gene [175]. In addition, kinetic studies have determined that only proteins that bind their target with an affinity of 10 nM or better are productive regulators [178]. Therefore, even if a protein binds a site in a regulatory region that is related but non-consensus, it may not have sufficient affinity to elicit a biological response.

## Summation

The β-globin locus is one of the most studied eukaryotic loci due to its role in β-thalassemia and sickle cell disease, the most common monogenetic diseases world-wide. In order to develop effective treatments that would lead towards a cure for hemoglobinopathies, it is crucial to understand the regulation of the β-globin genes. Given that the majority of hemoglobinopathies result from DNA alterations within the β-globin gene locus, identifying novel *cis*-elements that possess strong regulatory potential would be beneficial towards the development of novel therapies.

The goal of the proposed project is to employ a novel technique to study the *cis*-regulatory elements of the β-globin gene locus in order to identify DNA sequences that carry strong regulatory potential. Idesigned artificial DNA-binding domains (DBD)

comprised of modular zinc finger DNA binding domains (ZF-DBD) that will target a DNA region of interest and bind to this site with high affinity and specificity. Binding of the ZF-DBD renders the target site inaccessible to endogenous transcription factors and allow for the validation of the target *cis*-element to assess its requirement for globin gene function. This technique provides an advantage over traditional approaches because it can be rapidly employed while maintaining the specificity and efficacy that parallels that of traditional methods.

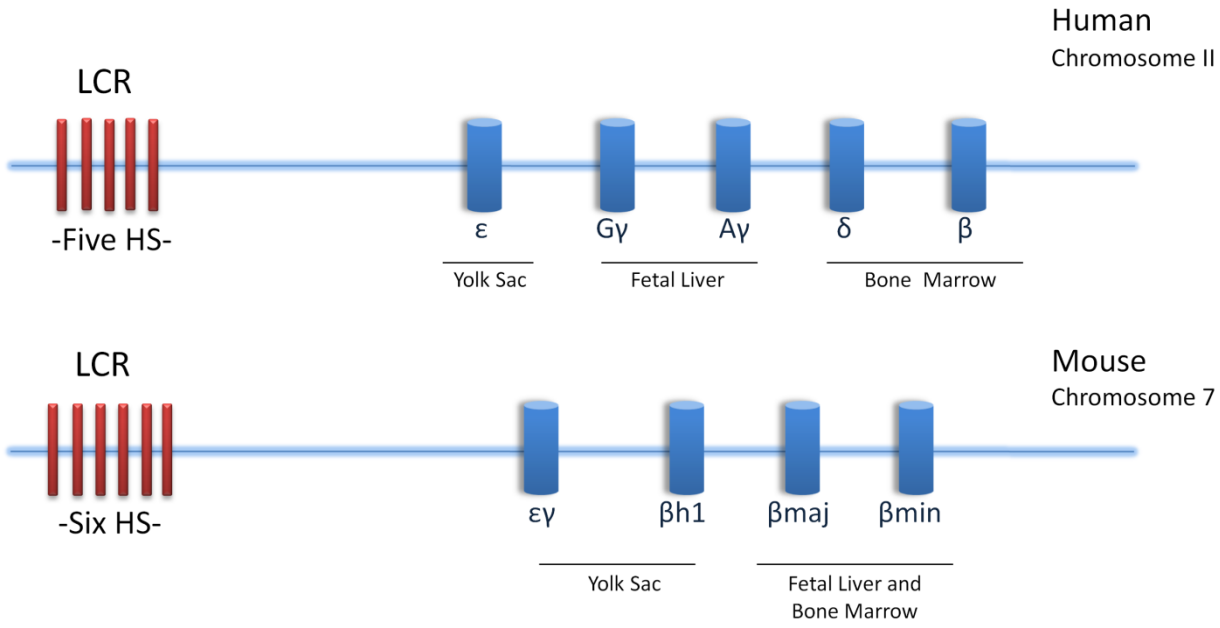


Figure 1-1. Schematic representation of the structural organization of the human and murine β-globin loci located on chromosomes 11 and 7 respectively.

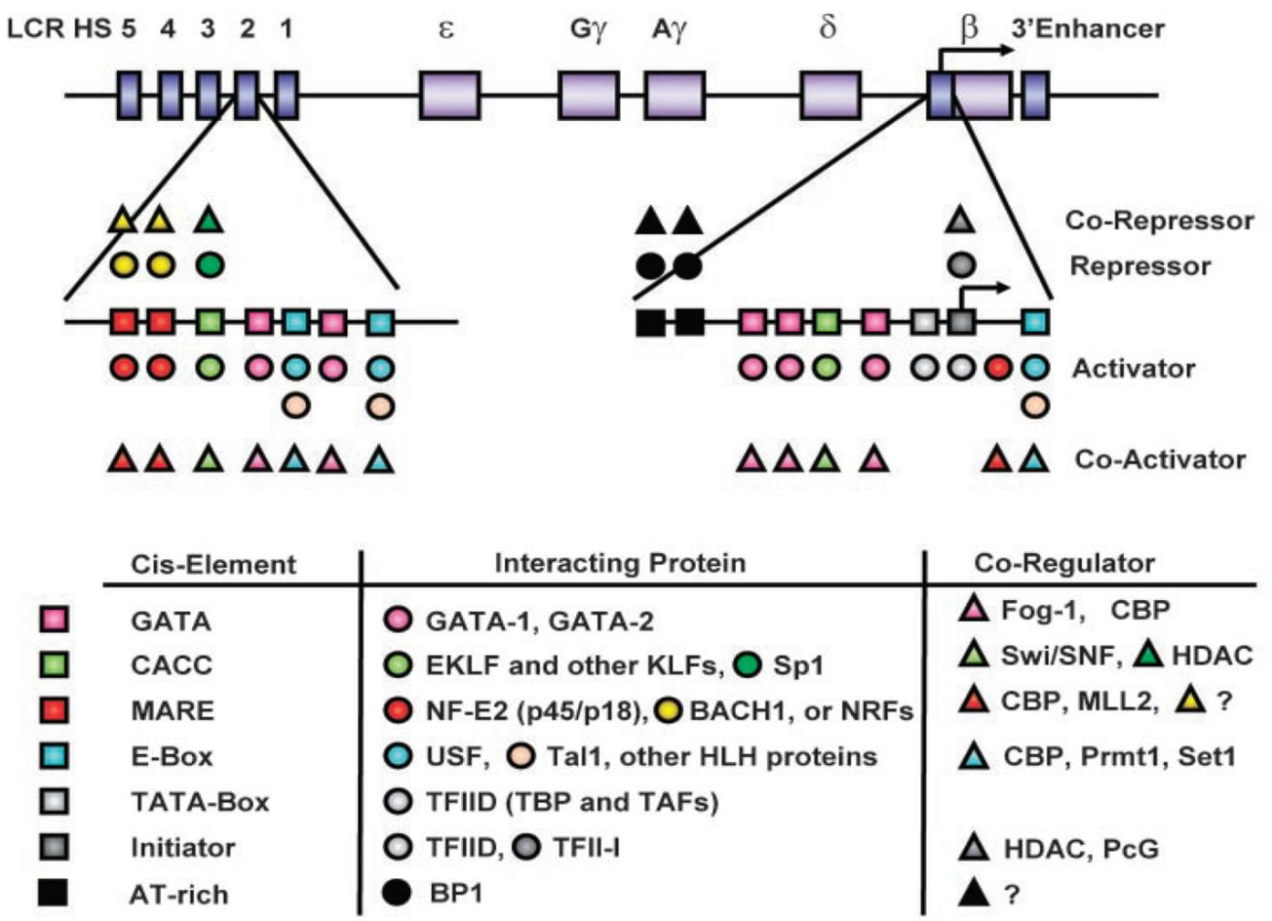


Figure 1-2. A highlight of the *cis*-regulatory elements located within the human β-globin locus control region HS2 and β-globin promoter region. The sequence of the *cis*-elements as well as the target interacting proteins and co-regulators are listed below [183].

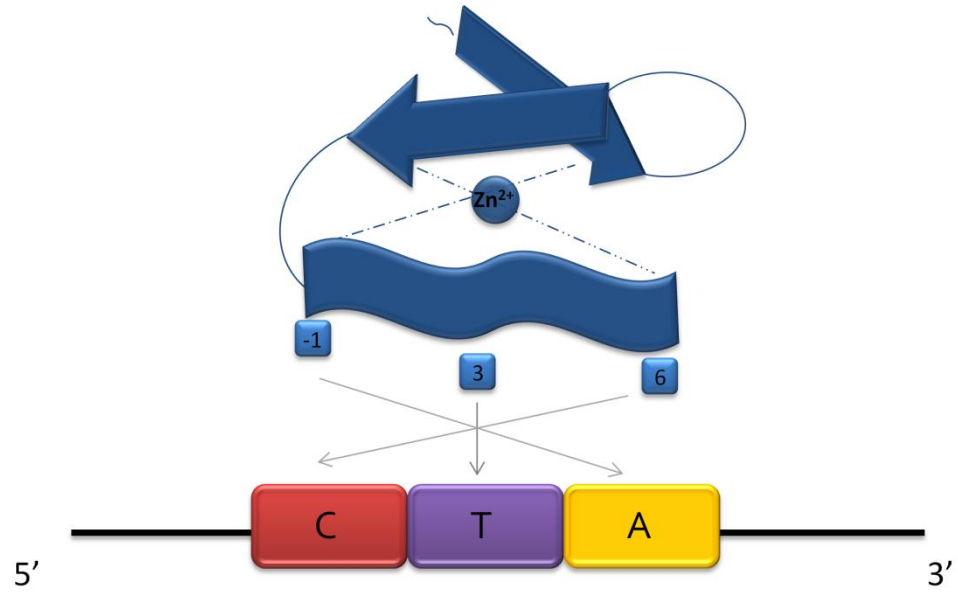


Figure 1-3. Schematic of Zinc Finger binding to DNA. Base-specific contacts are made with amino acids located at positions -1, 3, and 6 relative to the start of the alpha helix. The zinc finger domain binds to DNA in an anti-parallel fashion such that the amino acid at position 6 makes base-specific contacts with the 5’ nucleotide of the triplet.

# CHAPTER 2
# MATERIAL AND METHODS

## Zinc Finger Design and Construction

Figure 2-1 provides the overall schematic of the 6-fingered zinc finger DNA-binding domains (ZF-DBD) derived from the Barbas modules and Zinc Finger Tools database [176]. Detailed recognition helices and oligos are listed in Tables 2-1 to 2-3. The ZF-DBD was constructed as described previously with some modifications to incorporate the Flag system (Sigma) and a viral approach was used for eukaryotic studies. Briefly, the 6-fingered ZF-DBD was split into two reactions that would each encode a 3-fingered ZF domain termed ZF 1-3 and ZF 4-6. A series of oligos coding for either the zinc finger constant backbone or the variable oligos tailored to recognize desired target sequences were hybridized by overlap polymerase chain reaction (PCR) and gaps annealed in a phase I modular PCR assembly using a high fidelity Pfu polymerase (Aligent). PCR conditions were performed according to the manufacturer's protocol (Aligent) using the following programs. Assembly PCR: 1) 95 $^{o}$C for 2 min, 2) 95 $^{o}$C for 30 s, 3) 60 $^{o}$C for 30 s, 4) 72 $^{o}$C for 30 s, repeating cycles (2-4) 12X followed by a final extension for 5 min at 72 $^{o}$C. The program for the subsequent PCR termed amplification is as follows: 1) 95 $^{o}$C for 2 min, 2) 95 $^{o}$C for 30 s, 3) 60 $^{o}$C for 30 s, 4) 72 $^{o}$C for 30 s, repeating cycles (2-4) 25X followed by a final extension for 5 min at 72 $^{o}$C. The ZF fragments ZF 1-3 and ZF 4-6 were purified using a PCR purification kit (Qiagen) along with a canonical linker which was gel purified (Qiagen). Fragments were digested with the appropriate restriction enzymes (Table 2-1) and ligated with T4 ligase (NEB) in frame to generate the coding sequence of the complete 6-fingered ZF protein. The complete ZF fragment was then PCR amplified according to manufacture's instruction

with Pfu polymerase. The product was then isolated by gel extraction and subcloned into the digested, CIP treated (NEB), pT7-Flag2 vector (Sigma) using EcoRI/KpnI (NEB) restriction sites. Sequence was verified by Sanger Sequencing before use in *in vitro* studies. To construct the eukaryotic viral vector, the coding region for the ZF-DBDs were isolated by PCR from pT7-Flag with modified oligos (Table 2-1) to incorporate a NLS as well as novel restriction sites and introduced into the pMSCV-neo vector (clontech) at the HpaI site.

**Expression and Purification of the ZF-DBD**

10 ng of the pT7-Flag2 vector containing the desired ZF-DBD of interest was used to transform BL21-DE3 cells (Invitrogen). Cells were cultured to log phase (OD 0.5) in Luria broth medium supplemented with 100 µg/mL ampicillin and expression of the ZF-DBD was induced upon the addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma) and 100 µM $ZnCl_2$ (Sigma) following a 4 hour incubation at 37 $^{o}$C. Cells were then centrifuged at 1900xg at 4 $^{o}$C for 20 min and pellets were resuspended in storage buffer as described in Cathomen et al. [184]. Cells were lysed twice by French press, treated with 200µg DNase I (Sigma) and centrifuged at 44,000xg for 30 min at 4 $^{o}$C to remove cell debris. Retained supernatant was passed through a 0.2 µM filter (Corning) and immunopurified using anti-Flag M2 magnetic beads (Sigma) according to the manufacturer's protocol. Flag-tagged ZF-DBDs were eluted using 3X Flag peptide (Sigma) in 10 mM Tris pH 8.0, 90 mM KCl, 100 µM $ZnCl_2$, 5 mM DTT (Sigma), 0.1% Triton X-100 (Sigma) and 30% glycerol (Fisher) and stored at -80 $^{o}$C until further analysis. Protein concentrations were determined by Bradford method and protein purity was assessed by separating the eluted proteins on a 4-15% SDS PAGE (Bio-Rad) and staining with Coomassie Blue (Bio-Rad).

## Electrophoretic Mobility Shift Assay

Electrophoretic Mobility Shift Assay (EMSA) was performed using the Lightshift Chemiluminescent Kit (Thermo Scientific) according to manufacturer's protocol. To prepare DNA, double stranded oligonucleotides representing either the murine -90 CAC box (5'-GGATCCGAATTCCTGCAGGGTAACACCCTGGCATTGGCCAA-3') or a mutant sequence (5'-GGATCCGAATTCAGTACTTTGCCTGTTTCAATGCCTTAACC-3') were annealed in 250 mM Tris-HCl, pH 7.7 to their complement antisense sequences by heating to 95 $^{o}$C and cooling in 0.5 $^{o}$C increments to 4 $^{o}$C for several hours. Oligonucleotides were then digested with BamHI (NEB), passed through a G-25 column (GE Healthcare) to purify the probe, and blunted with Klenow polymerase I (NEB) using biotinlyated nucleotides: bio-dATP and bio-dCTP (Invitrogen) along with unlabeled nucleotides dTTP and dGTP. For binding reactions, 1 µg of purified recombinant ZF-DBD protein was incubated with 2 ng of either biotinlyated WT or mutant oligos. Binding was challenged with 1 µg of excess unlabeled WT and Mut DNA and 1 µg of mouse M2 anti-Flag antibody (Sigma).

## Immunoblot Analysis

Protein isolation was performed as previously described by Leach et al. with modifications to include a mechanical lysis step with a micro-grinder (Radnoti) prior to centrifugation at 20,800 x g for 15 min at 4 $^{o}$C [100]. Proteins were quantified by Bradford Assay (BioRad) and a range of 10-20 µg was loaded on a 4-15% TGX Tris-HCl gel (Biorad) and separated by SDS-PAGE. Resolved proteins were transferred onto a PVDF membrane using the *Trans*-Blot Turbo (Bio-rad). Proteins were blocked in 5% nonfat dried milk in TBST prior to incubation with the following antibodies: mouse anti-Flag (F-3165, Sigma), rabbit anti-ZF sera (gift from Dr. Carlos Barbas, Scripps, CA),

mouse anti-BRG1 (gift from Dr. Reissman, University of Florida), and mouse anti-Tubulin (sc-55529, Santa-Cruz). Anti-mouse and anti-rabbit secondary antibodies were purchased from Santa Cruz. Proteins were detected by ECL reagent (Millipore) and visualized on X-ray film (Kodak). Compartmentalization immunoblot was performed using the NE-PER® kit (Thermo Scientific) according to the manufacturer's instruction and isolated proteins were analyzed as described above.

**Cell Culture and Transfections/Transductions**

Murine erythroleukemia (MEL) and Phoenix A cells were cultured in Dulbecco's modified Eagle's medium (Cellgro) supplemented with 10% (vol/vol) fetal bovine serum (FBS) and 1% penicillin/streptomycin (Cellgro). Cells were grown at 37 $^{o}$C with 5% $CO_2$ and maintained at a density of 2.0x10$^5$ cells/mL. Induction of MEL erythroid differentiation was achieved by the addition of 2% (vol/vol) dimethylsulfoxide (DMSO) to the media following a 72 hour incubation. Retroviral-mediated creation of stable MEL cell lines was achieved by transfecting the packaging cell line phoenix A with pMSCV-neo containing the target ZF-DBD or empty vector via Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. After 48 hours, supernatant containing live replication-deficient virus was harvested and centrifuged at 913 x g for 5 min to remove cellular debris. Retained supernatant was treated with 2 µg/mL polybrene and added to MEL cells. Cells were incubated for 48 hours before the addition of 800 µg/mL geneticin for selection for 2 weeks before the concentration was lowered to 100 µg/mL geneticin for maintenance.

**Immunofluorescence Microscopy**

A total of 1.0×10$^6$ induced MEL cells were plated on poly-lysine (Sigma) coated plates and cultured overnight at 37°C with 5% $CO_2$ before being fixed with 4% (wt/vol)

paraformaldehyde (Sigma) for 10 min. Cells were rinsed thoroughly with PBS and permeabilized with 0.5%Triton X-100 for 20 min, followed by a blocking step with 3% (wt/vol) BSA (Sigma). Cells were probed with a ZF-specific antibody, washed with 4% (vol/vol) Tween (Sigma) and incubated with FITC-conjugated secondary antibody (sc-2777, Santa Cruz). Cells were washed again in 4% (vol/vol) Tween before being placed on slides with mounting media containing DAPI (Vecta Shield). Fluorescence was visualized using a fluorescence microscope (Leica).

### RNA Isolation and Analysis

RNA was isolated from MEL cells and mouse fetal liver by using the RNeasy kit (Quiagen) according to the manufacturer's instructions and was reverse transcribed with the iScript cDNA synthesis kit (Bio-Rad). RNA analysis was performed as previously described in Liang et al. [185]. The following primers were used to amplify cDNA: mouse β-minor, 5′-TACGTTTGCTTCTGATTCT-3′ (upstream [US]) and 5′-CAGAGGCAGAGGATAGGTC-3′ (downstream [DS]); mouse dematin (band 4.9), 5′-CCGCATGAGGCTTGAGAGG-3′ (US) and 5′-TCTTCTTAAGTTCGTTCCGCTTCC-3′ (DS); mouse β-spectrin, 5′-GCTTAAGGAACGCCAGACTCCAG-3′ (US) and 5′-ATTTCTCCTGCTCGTCTTTGT-3′; and mouse β-actin, 5′-GTGGGCCGCTCTAGGCACCA-3′ (US) and 5′-TGGCCTTAGGGTGCAGGGGG-3′ (DS). All RT-qPCR data were normalized to mouse β-actin expression levels.

### Chromatin Immunoprecipitation

ChIP assays were performed as described previously with the following modifications [185]. A minimum of $2.0\times10^7$ cells were isolated, crosslinked, and sonicated to produce 200–400 bp chromatin fragments. Lysates were pre-cleared with mouse IgG (sc-2025, Santa Cruz) for 2 h at 4°C with gentle rotation following a

subsequent pre-clearing step with protein A Sepharose beads (GE Healthcare; CL-4B). Lysates were centrifuged at 1,700xg for 10 min at 4°C to pellet beads and the supernatant was retained and incubated with specific antibodies overnight at 4°C with gentle rotation. Sheared chromatin was incubated with 10 µL of ZF antibody. Alternatively, 2-10 µg of the following antibodies were used: anti-mouse RNA-Pol II (CTD45H8; Upstate Biotechnology, Inc.), anti-mouse RNA Poll II serine 2 (ab5095; Abcam), anti-mouse USF2 (sc-862X; Santa-Cruz), anti-mouse TAL1 (gift from the Huang lab, University of Florida), anti-mouse NF-E2 (sc-291X; Santa Cruz), and anti-mouse KLF1 (ab2483; Abcam, and a gift from James Bieker, Mount Sinai Hospital, New York, NY). DNA was purified by phenol/ chloroform/isoamyl alcohol and subsequent chloroform extractions and precipitated by adding 2.5X volume of 100% (vol/vol) ethanol in the presence 10µg glycogen (Invitrogen) overnight at −20°C. DNA precipitates were washed in 70% ethanol, resuspended in 10mM Tris-HCl pH 8.5, and analyzed by RT-qPCR as previously described (24). The following primer pairs were used: mouse β-major promoter 5'-AAGCCTGATTCCGTAGAGCCACAC-3' (US), 5'-CCCACAGGCCAGAGACAGCAGC-3' (DS); mouse β-minor promoter, 5′-GCCATAGCCACCCTGTGTAG-3′ (US), 5′GAGACAGCAGCCTTCTCAGA-3′ (DS); dematin (band 4.9) promoter, 5′-AATGACGGCAGGGGTCAG (US), 5′-CTTGGTCATGCCCATAGCTT-3′ (DS); mouse HS2, 5′-TGCAGTACCACTGTCCAAGG-3′ (US), 5′-ATCTGGCCACACACCCTAAG-3′ (DS) and mouse 3' β-major region, 5'-GCTCTTGCCTGTGAACAATG-3' (US), 5'-TGCTTTTTATTTGTCAGAAGACAG-3' (DS).

## Benzidine Staining

Benzidine (Sigma) stock solution was prepared as follows: 3% Benzidine (w/v) in 90% glacial acetic acid and 10% ddH20 from which fresh working solution was then prepared comprised of a ratio of 1:1:5 for benzidine stock solution, $H_20_2$, and $ddH_20$ respectively. Induced MEL ($1.0x10^6$) cells were centrifuged at 2000 rpm for 5 min and resuspended in 500 µl PBS and 100 µl of the benzidine working solution was added to the cell suspension. After a 5 min incubation period, cells were centrifuged at 2000 rpm for 5 min and resuspended again in 500 µl PBS. Cells were counted under a light microscope (Van guard, Kirkland, WA) using a hemacytometer (Bright-Line, Horsham, PA) and blue cells were calculated as a percentage of total cells. Cells were then imaged using the Scope Image 9.0 software (Life Scientz Bio-tech, China).

## Generation of Transgenic Mice

The plasmid pMSCV-neo containing the -90 β-ZF-DBD was linearized with Acl I (NEB), gel purified using the Qiagen gel extraction kit, resuspended in injection buffer at a concentration of 2ng/µl, and injected into FvB oocytes as described previously [186]. After transfer to pseudopregnant recipients, embryos were taken at day 13.5 dpc and imaged using the Leica microscope (Leica), and Q-capture software (Q-Imaging, B.C, Canada). RNA and DNA was extracted and analyzed as described previously. The cDNA was analyzed by qPCR using primers specific for the βmin-globin gene as described above and primers specific for the -90 β-ZF-DBD cDNA: 5'-CTCGAGCCCGGGGAGAAAC-3' (US), 5'-TCACTTGTCATCGTCGTCCT-3' (DS). The animal work was approved by the UF IACUC committee.

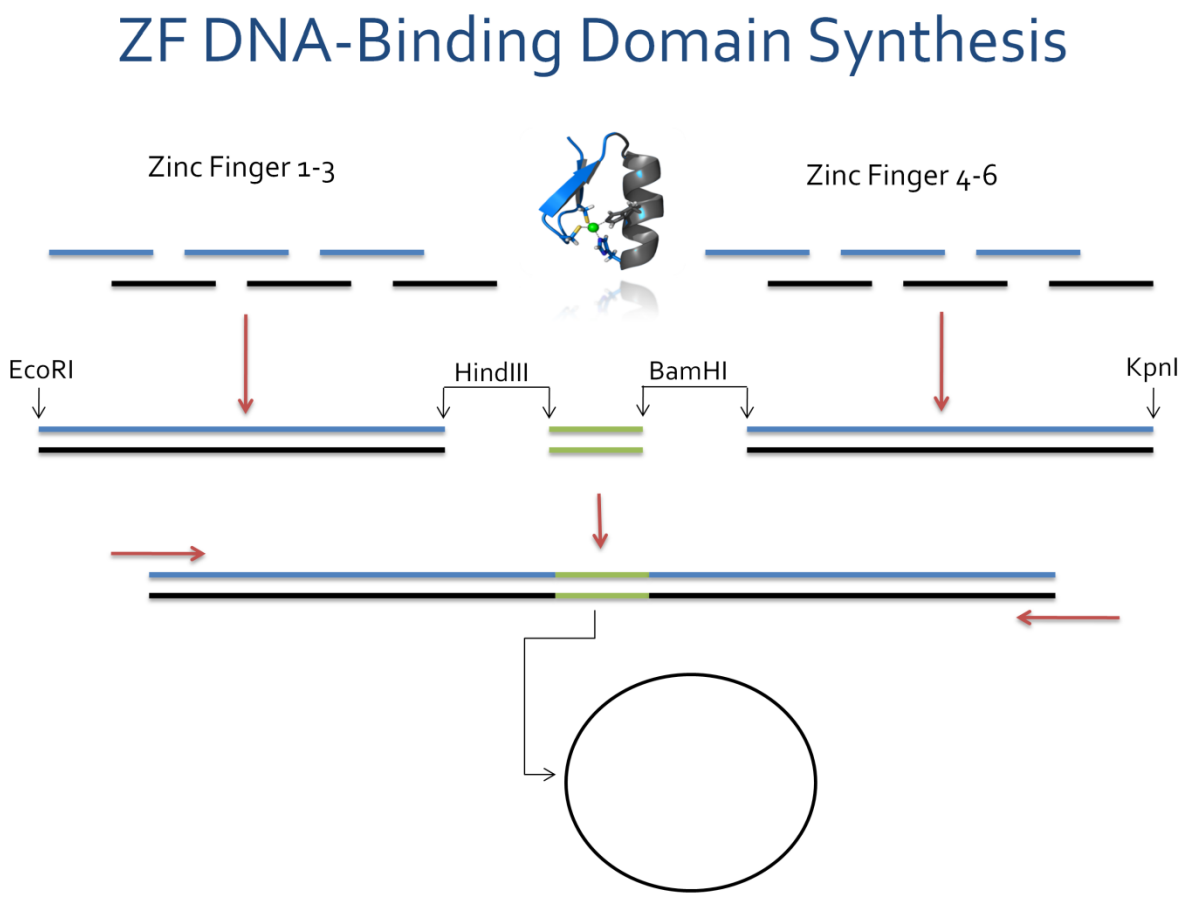


Figure 2-1. Overall design schematic of the ZF-DBD. In order to maximize the efficiency of the reaction, the 6-fingered ZF-DBD is separated into two domains representing ZF 1-3 and ZF 4-6. These domains are assembled separately using a series of overlapping oligonucleotides coding for either the constant ZF backbone (blue) or the variable α-helix reading heads (black). Domains are annealed and gaps sealed in a phase I PCR assembly step. The two domains are then ligated together incorporating a non-canonical linker in the process to yield the complete 6-fingered ZF-DBD which is then amplified by PCR (Phase II) using flanking forward and reverse primers and cloned into the appropriate expression vector

Table 2-1. Tailored ZF-DBD recognition helices.

| 90 β-ZF-DBD Target Sequence: CTGCAGGGTAACACCCTG | | | | | | | |
|---|---|---|---|---|---|---|---|
| ZF Domains | ZF1 | ZF2 | ZF3 | LINKER | ZF4 | ZF5 | ZF6 |
| Target Recognition Triplet | CTG | ACC | AAC | | GGT | CAG | CTG |
| ZF Helix | RNDALTE | DKKDLTR | DSGNLRV | TGEKP | TSGHLVR | RADNLTE | RNDALTE |
| -2KB NC ZF-DBD Target Sequence: CTAGAGACCCATGATTGA | | | | | | | |
| ZF Domains | ZF1 | ZF2 | ZF3 | LINKER | ZF4 | ZF5 | ZF6 |
| Target Recognition Triplet | TGA | GAT | CAT | | ACC | GAG | CTA |
| ZF Helix | QAGHLAS | TSGNLVR | TSGNLTE | TGEKP | DKKDLTR | RSDNLVR | QNSTLTE |
| +60 ZF-DBD Target Sequence: CACCTGACTGATGCTGAG | | | | | | | |
| ZF Domains | ZF1 | ZF2 | ZF3 | LINKER | ZF4 | ZF5 | ZF6 |
| Target Recognition Triplet | GAG | GCT | GAT | | ACT | CTG | CAC |
| ZF Helix | RSDNLVR | TSGELVR | TSGNLVR | TGEKP | THLDLIR | RNDALTE | SKKALTE |

Target triplets with the corresponding amino acids shown were derived from the ZF Tools website [176]. ZF helices are positioned in the antiparallel orientation relative to the DNA target sequence. Amino acids reflecting positions -1 to +6 relative to the start of the alpha helix reading head of the ZF protein are shown. All sequences are displayed in a 5'-3' orientation.

Table 2-2. Constant primers and oligonucleotides used in zinc finger construction

| Primer Nomenclature | Constant Oligos (5'-3') |
|---|---|
| C1 (ZF 1-3) | CGGGGAGAAACCCTATAAGTGTCCGGAGTGTGGCAAGTCGTTCTC |
| C1 (ZF 4-6) | TATAAGTGTCCGGAGTGTGGCAAGTCGTTCTC |
| C2 | GCGTACCCACACGGGCGAAAAGCCGTACAAATGCCCAGAATGCGGTAAATCCTTCAGC |
| C3 | CGGGGAGAAACCCTATAAGTGTCCGGAGTGTGGCAAGTCGTTCTC |
| | General Primers (5'-3') |
| Forward Primer (ZF1-3) | CAGGACGAATTCACTCGAGCCCGGGGAGAAACCCTATAAG |
| Forward Primer (ZF4-6) | CAGGACGGATCCTATAAGTGTCCGGAGTGTGGCAA |
| Reverse Primer (ZF1-3) | CAGGACAAGCTTGTGAGTGCGCTGGTG |
| Reverse Primer (ZF4-6) | CAGGACGGTACCGCTGGTTTTTTTGCCGGTGTGAGTGCGCTGGTG |
| Eukaryotic Forward Primer | ACGGTTAACATGCCGAAAAAAAAACGCAAAGTGCTCGAGCCCGGGGAGAAAC |
| Eukaryotic Reverse Primer | GCAGTTAACTCACTTGTCATCGTCGTCCT |

Oligonucleotides and primers used for the construction of the ZF-DBDs. Sequences are listed in the 5'-3' orientation.

Table 2-3. Variable oligonucleotides used in zinc finger construction

| Primer Nomenclature | Variable Oligos (5'-3') |
|---|---|
| | Oligonucleotides for -90 ZF-DBD |
| V1 (CTG) (ZF1-3) | GCCCGTGTGGGTACGCTGATGTTCGGTCAGCGCATCGTTGCGCGAG AACGACTTGCCACAC |
| V2 (ACC) (ZF1-3) | CCTGTATGCGTCCGTTGATGGCGGGTCAGATCTTTTTTATCGCTGAAG GATTTACCGCAT |
| V3 (AAC)(ZF1-3) | TTGTGAGTGCGCTGGTGCACGCGCAGGTTGCCGCTATCACTAAAACT CTTCCCACATTCG |
| V1 (GGT) (ZF4-6) | GCCCGTGTGGGTACGCTGATGGCGCACCAGATGGCCGCTGGTCGAG AACGACTTGCCACAC |
| V2 (CAG) (ZF4-6) | CCTGTATGCGTCCGTTGATGTTCGGTCAGGTTATCCGCGCGGCTGAA GGATTTACCGCAT |
| V3 (CTG)(ZF4-6) | CGGTGTGAGTGCGCTGGTGTTCGGTCAGCGCATCGTTGCGACTAAAA CTCTTCCCACATTCG |
| | Oligonucleotides for +60 ZF-DBD |
| V1 (GAG) (ZF1-3): | GCCCGTGTGGGTACGCTGATGGCGCACCAGGTTATCGCTGCGCGAG AACGACTTGCCACAC |
| V2 (GCT) (ZF1-3): | CCTGTATGCGTCCGTTGATGGCGCACCAGTTCGCCGCTGGTGCTGAA GGATTTACCGCAT |
| V3 (GAT)(ZF1-3): | TTGTGAGTGCGCTGGTGGCGCACCAGGTTGCCGCTGGTACTAAAACT CTTCCCACATTCG |
| V1 (ACT) (ZF4-6): | GCCCGTGTGGGTACGCTGATGGCGAATCAGATCCAGATGGGTCGAG AACGACTTGCCACAC |
| V2 (CTG) (ZF4-6): | CCTGTATGCGTCCGTTGATGTTCGGTCAGCGCATCGTTGCGGCTGAA GGATTTACCGCAT |
| V3 (CAC)(ZF4-6) | CGGTGTGAGTGCGCTGGTGTTCGGTCAGCGCTTTTTTGCTACTAAAA CTCTTCCCACATTCG |
| | Oligonucleotides for NC ZF-DBD |
| V1 (GAG) (ZF1-3): | GCCCGTGTGGGTACGCTGATGGCTCGCCAGATGGCCCGCCTGCGAG AACGACTTGCCACAC |
| V2 (GCT) (ZF1-3): | CCTGTATGCGTCCGTTGATGGCGCACCAGGTTGCCGCTGGTGCTGAA GGATTTACCGCAT |
| V3 (GAT)(ZF1-3): | TTGTGAGTGCGCTGGTGTTCGGTCAGGTTGCCGCTGGTACTAAAACT CTTCCCACATTCG |
| V1 (ACT) (ZF4-6): | GCCCGTGTGGGTACGCTGATGGCGGGTCAGATCTTTTTTATCCGAGA ACGACTTGCCACAC |
| V2 (CTG) (ZF4-6): | CCTGTATGCGTCCGTTGATGGCGCACCAGGTTATCGCTGCGGCTGAA GGATTTACCGCAT |
| V3 (CAC)(ZF4-6): | CGGTGTGAGTGCGCTGGTGTTCGGTCAGGGTGCTGTTCTGACTAAAA CTCTTCCCACATTCG |

Oligonucleotides encoding the variable regions for ZF1-3 and ZF 4-6 of the ZF-DBDs targeted towards the -90, +60 and -2KB *cis*-elements are shown. All sequences are displayed in the 5'-3' orientation.

# CHAPTER 3
# CHARACTERIZING THE -90 β-ZF-DNA BINDING DOMAIN

## Background

Transcription is regulated by proteins that interact in a sequence-specific manner with the DNA at promoters or other regulatory elements. Most transcription factors belong to families of proteins that share characteristic DNA binding or protein/protein interaction domains and often recruit co-regulator complexes that modify chomatin structure, recruit transcription complexes, or modulate transcription elongation rates [187,188]. Because multiple members of transcription factor families are usually expressed in any given cell-type, it is difficult to assess in the context of the cell which protein mediates the effect of a *cis*-regulatory DNA element. Furthermore, transcription factors that regulate expression of genes often bind to multiple sites in promoter or enhancer regions [168]. The analysis of the functional role of *cis*-regulatory elements in the context of intact cells or *in vivo* therefore is challenging. Transgenic or reporter gene assays are powerful but limited by potential position-effects. Genetic manipulation of *cis*-regulatory elements is technically challenging and time consuming.

The zinc finger domain is the most commonly found DNA binding domain in eukaryotic transcription factors [122]. This domain is characterized by a DNA binding alpha helix which is stabilized by an adjacent finger-like structure in which histidine or cysteine residues coordinate a zinc atom [118]. The mode of DNA binding by zinc finger proteins is very well understood. This knowledge led to the development of artificial proteins containing a defined zinc finger DNA binding domain (ZF-DBD) that interacts with a specific sequence of interest [170]. Each alpha-helix reading head of a ZF-DBD recognizes 3 to 4 specific DNA base pairs and such reading heads can be designed to

essentially recognize most triplets of DNA base pairs [118]. Furthermore, these interactions are modular in nature and therefore arranging these zinc finger domains in tandem provides the recognition of asymmetrical DNA sequences [175]. Previous studies often link these artificial ZF-DBDs to effector domains that either enhance or repress transcription. Furthermore, ZF-DBDs have been quite extensively utilized to target nucleases to specific genomic sites to induce recombination [180,189]. Most of the artificial zinc finger proteins studied to date appear to interact with desired target sequences with high-specificity, however, off-targets have been detected [190]. The presence of effector domains can lead to additional unwarranted effects. For example, activation or repression domains can interact with proteins to modulate transcription. Thus expression of these effector domains fused to ZF-DBDs could sequester proteins and change expression of multiple genes, not just those targeted by the artificial transcription factor [190].

In the present study we wished to examine if ZF-DBDs without effector domains could be utilized to compete with endogenous transcription factors at specific *cis*-regulator elements thus neutralizing this site. Our purpose is to assess applicability of using ZF-DBDs as a screening tool to assess the function of potential *cis*-regulatory elements *in vivo*. To address this question, I designed and expressed a ZF-DBD harboring 6 zinc finger domains that specifically interacts with a known critical *cis*-regulatory element associated with adult β-globin gene expression—the -90 CAC box [191]. I termed this artificial binding protein −90 β-ZF-DBD. The CACCC site, located about 90bp upstream of the β-globin transcription start site, interacts with transcription factor KLF1, a zinc finger protein related to Sp1 [127]. KLF1 deficiency leads to

embryonic lethality in mice due to anemia [130]. Mutations of the CACCC site in the human population have been associated with β-thalassaemia [96]. I designed and expressed a ZF-DBD targeting 18bp flanking the adult β-minor globin -90 CACCC site. Statistically, an 18bp sequence occurs only once per mouse and human genome [175]. Binding assays indicated that the -90 β-ZF-DBD specifically interacted with DNA fragments containing the β-minor globin -90 CACCC site. Expression of this ZF-DBD reduced expression of the β-minor globin gene but not that of other KLF1-regulated genes. Transient transgenic embryos expressing the -90 β-ZF-DBD are phenotypically normal but exhibited reduced expression of the adult β-minor globin gene. The data demonstrate that ZF-DBDs without effector domains can be used to neutralize and thereby assess the function of specific *cis*-regulatory DNA elements *in vivo* [191].

## Results

To demonstrate the feasibility of using artificial DNA binding domains to neutralize the function of a *cis*-regulatory DNA element, I focused on the -90 CACCC box sequence located upstream of the murine adult β-minor globin gene promoter. This *cis*-element is known to interact with the transcription factor KLF1 and is critical for high-level β-minor globin gene expression in erythroid cells [192]. I hypothesized that neutralization of this site by an artificial DNA binding domain would prevent KLF1 from interacting with its target *cis*-element and would therefore reduce β-minor globin gene expression. Thus, I designed a ZF-DBD comprised of six zinc finger domains (herein referred to as the -90 β-ZF-DBD) that would bind to 18bp of DNA flanking the target -90 CACCC box (Fig. 3-1A). I designed the -90 β-ZF-DBD using the Zinc Finger Tools website and amino acid-nucleic acid recognition sequences were derived from the Barbas modules [176]. ZF-DBD proteins were then assembled and generated using a

modified protocol from Muller Lerch [184]. The -90 β-ZF-DBD was cloned in the pT7-Flag2 vector and expressed in and purified from *E.coli* (Fig.3-1B and C). The -90 β-ZF-DBD protein migrated at an expected size of 24 kDa. I next examined the DNA association activity of the purified recombinant-90 β-ZF-DBD using electrophoretic mobility shift assays (EMSAs) shown in Fig.3-1D. The -90 β-ZF-DBD specifically interacted with an oligonucleotide containing the 18 bp target WT sequence harboring the -90 CACCC site but not with mutant oligonucleotides. An excess of unlabeled WT oligonucleotides efficiently competed for the binding while mutant oligonucleotides did not perturb binding. The addition of a Flag-antibody to the binding reaction eliminated formation of the protein DNA complex indicating that this interaction was specific. These data demonstrate that the -90 β-ZF-DBD interacts with the target 18bp sequence encompassing the -90 CACCC box in a sequence-specific manner.

For eukaryotic *in vivo* studies, the coding sequence for the -90 β-ZF-DBD was cloned into the pMSCV-neo plasmid and modified to include a NLS. After packaging, viruses either harboring the vector encoding the -90 β-ZF-DBD or an empty vector were used to transduce mouse erythroleukemia (MEL) cells. Single cell clones were generated from transduced MEL population pools and clonal variability was minimal with respect to expression of the -90 β-ZF-DBD and similar globin gene expression profiles were observed. MEL cells are erythroid progenitor cells that express low levels of the adult α- and β-globin genes. Several compounds have been shown to induce differentiation of MEL cells including Dimethylsulfoxide (DMSO). Incubation of MEL cells for 3 days in the presence of 2% DMSO results in a dramatic increase in globin gene expression [193]. To examine the cellular localization of the -90 β-ZF-DBD I first

performed immunofluorescence microscopic analysis in induced MEL cells using an antibody against the backbone of the ZF-DBD (Fig. 3-2A). The data demonstrate that the -90 β-ZF-DBD localized to the nucleus. I next fractionated the MEL cells into cytosolic and nuclear compartments and performed an immunoblot analysis (Fig. 3-2B). Brg1, a nuclear chromatin regulatory protein, and α-tubulin, a predominantly cytoplasmic protein were used as controls to indicate complete cellular fractionation. The -90 β-ZF-DBD was only present in the nuclear fraction in induced MEL cells confirming immunofluorescence results.

I analyzed KLF-1-dependent gene expression in single cell clones from MEL cells that stably expressed either the -90 β-ZF-DBD or empty vector. Upon phenotypic analysis MEL cell containing the -90 β-ZF-DBD when pelleted were pale compared to vehicle control cells suggesting that globin synthesis was impaired. (Fig. 3-3A). This was confirmed by staining induced MEL cells with benzidine—a stain that reacts with intact hemoglobin for form a blue color indicative of adequate globin synthesis, and found that there was a marked reduction in benzidine positive cells that stably expressed the -90 β-ZF-DBD to levels of approximately 30% compared to control cells (Fig. 3-3B and C). I next extracted RNA and analyzed the expression of β-minor globin. Vehicle-treated control cells exhibited a significant increase in β minor globin expression upon induction as expected. Cells expressing the -90 β-ZF-DBD however revealed a dramatic reduction in of β-minor globin expression in both uninduced and induced MEL cells (Fig. 3-3D). To ensure that this effect was specific to only β-minor globin and to validate our hypothesis that the -90 β-ZF-DBD was interfering with KLF1 occupancy at only β-minor globin thus causing the reduction in expression, I analyzed expression of

other KLF1-regulated genes in erythroid cells. Dematin (also known as band 4.9) and β-spectrin (Fig. 3-3E and F) are genes both induced upon erythroid differentiation and positively regulated by KLF1. Importantly, the sequences of the KLF1 binding sites in the β-spectrin and dematin promoter regions are very similar to the -90 CACCC site in the β-minor globin gene promoter. There were no differences in expression levels of β-spectrin and dematin in induced and uninduced MEL cells expressing the -90 β-ZF-DBD when compared to vehicle-control cells demonstrating that expression of the -90 β-ZF-DBD did not affect expression of these genes (Fig. 3-3E and F). This demonstrates that the -90 β-ZF-DBD specifically reduces expression of the β-globin genes but did not affect expression of other KLF1-target genes in erythroid cells.

I next wanted to map the *in vivo* occupancy of the -90 β-ZF-DBD to chromatin by chromatin immunoprecipitation (ChIP). Our original attempts using an anti-flag antibody did not yield conclusive results. I obtained antisera from Dr. Carlos Barbas (Scripps, La Jolla, CA), which recognizes the backbone of the ZF-DBD. The specificity of the ZF antibody was verified by immuno-blot analysis and was consistent with results seen with the anti-Flag antibody. I first examined the occupancy of the -90 β-ZF-DBD to the β-minor globin gene promoter, to the promoter of the dematin gene, and to locus control region element HS2, which all contain KLF1 binding sites (Fig. 3-4A). The data demonstrate that the -90 β-ZF-DBD binds specifically to the β-minor globin promoter but not with LCR element HS2. Furthermore, expression of the -90 β-ZF-DBD drastically reduced the binding of KLF1 to the β-minor globin promoter but not with LCR HS2 (Fig. 3-4B). I also observed transient binding of the -90 β-ZF-DBD to the dematin promoter but it was not as pronounced as the interaction with the β-minor globin promoter region.

Transient binding possibly occurred due to the target sequence similarities in the promoter regions of these genes (Fig. 3-4E). Yet interestingly, KLF1 binding to this region was not affected (Fig. 4-4A and B).

To address the question if binding of the -90 β-ZF-DBD affects the recruitment of transcription complexes to the β-minor globin gene promoter, I performed ChIP using antibodies specific for RNA polymerase II (Pol II). Previous studies have shown that Pol II interacts with LCR HS2 and with the adult β-globin gene promoter and that the binding increases upon induction of MEL cell differentiation [194]. As shown in Fig 3-4C., Pol II bound efficiently to the β-minor globin promoter and to a lesser extent to LCR element HS2 in induced MEL cells harboring the empty vector. The interaction of Pol II was reduced at the β-globin promoter in MEL cells expressing the -90 β-ZF-DBD. Interestingly, the binding of Pol II at LCR HS2 was also somewhat reduced in these cells despite the fact that KLF1 binding was not perturbed at this site which could indicate a reduction in interaction frequencies between the LCR and the β-globin promoter in cells expressing the −90 β-ZF-DBD. The data demonstrate that occupancy of the -90 β-ZF-DBD to the β-minor promoter prevents KLF1 from binding to the -90 CACCC site thereby inhibiting the recruitment of transcription complexes (Fig. 3-4D).

To test the effect and applicability of introducing an artificial DNA binding domain in a complete *in vivo* system, I generated and analyzed transient transgenic embryos ubiquitously expressing the -90 β-ZF-DBD. There was the possibility that transgenic embryos expressing the -90 β-ZF-DBD would die in utero due to defects in hemoglobin synthesis. Thus, we decided to analyze transient transgenic and wild-type embryos at 13.5 dpc. The switch from the embryonic to the fetal/adult globin expression program

occurs around day 11.5 dpc [3]. I analyzed three transgenic and 3 wild-type mice. Two out of the three transgenic embryos revealed reduced expression of adult β-minor globin gene expression (Fig. 3-5B). Fig. 3-5A illustrates a representative transgenic -90 β-ZF-DBD-expressing embryo and as can be seen, these embryos were paler in comparison to non-transgenic littermates. Importantly, transgenic embryos did not reveal any obvious developmental phenotype indicating that expression of the -90 β-ZF-DBD did not affect development and differentiation. This confirms the possibility that ZF-DBD without effector domains to minimize non-specific interactions can be expressed in *in vivo* systems.

## Discussion

The functional role of *cis*-regulatory DNA elements and interacting proteins is difficult to assess *in vivo*. Transgenic studies or stable reporter gene assays are hampered by the fact that sequences are taken out of their natural environment and may be influenced by neighboring chromatin at the site of integration. Artificial DNA-binding domains represent promising tools for studying and modulating gene regulation *in vivo*. As shown in this study, ZF-DBDs without effector domains can be used to alter expression of genes by preventing transcription factors from accessing DNA at specific sites. These synthetic proteins can be used to assess the function of a *cis*-regulatory DNA element and the identity of interacting proteins *in vivo*. Several different strategies have been developed to generate artificial binding domains; the majority are based on DNA-binding domains found in either ZF or transcription activator-like effector (TALE) proteins [178,195]. TALE proteins, identified in prokaryotic plant pathogens, are modular in nature and like ZF proteins can be combined to recognize extended DNA sequences [196]. The advantage of ZF proteins over TALE proteins is that they are relatively small,

minimizing effects resulting from potential protein interactions and perhaps in the long term allowing delivery by protein transduction [197]. Much is known about the configuration of amino acids in the ZF α-helix reading head that interact with specific DNA base pair triplets, and this led to the development of algorithms that have been used to design artificial ZF proteins that bind to specific DNA sequences [198]. Klug and colleagues first described the use of a synthetic 3-ZF protein that blocked the activating sequence that arose from a chromosomal translocation [169]. In this study, I used the Zinc Finger Tools web site to design a ZF-DBD targeting the −90 CACCC site in the murine adult β-minor globin gene promoter [176]. The −90 β-ZF-DBD interacted with the β-minor globin CACCC box and efficiently blocked association of KLF1 with this site. As a consequence, β-minor globin gene expression was reduced. In previous studies, Bieker and colleagues over-expressed the KLF1 DNA-binding domain in erythroid cells, which led to reduced β-major- globin expression levels [199]. However, in contrast to the −90 β-ZF-DBD, the KLF1-binding domain interacts with many gene regulatory elements in erythroid cells. The −90 β-ZF-DBD did not interact with LCR element HS2, which also contains a KLF1-binding site. I detected a weak interaction of the −90 β-ZF-DBD with the dematin promoter in induced MEL cells. However, this binding event did not affect binding of KLF1 or dematin expression, strongly suggesting that the −90 β-ZF-DBD failed to interfere with KLF1 function at this particular site. I also observed a reduction in expression of the βmaj-globin gene in cells expressing the −90 β-ZF-DBD. The reason for this is not known but could indicate that the adult β-globin promoters cooperatively regulate each other. This would be consistent with a recent study demonstrating extensive promoter–promoter interactions and cross-regulation in

mammalian cells [200]. Alternatively, the artificial zinc finger protein could bind to both adult globin gene promoters due to close proximity. A recent study by Lam et al. demonstrated that the majority of modular artificial C2H2 ZF proteins bind DNA with high sequence specificity and preferentially associate with sites to which they were targeted [201]. Although we did not perform a global analysis of the interaction of the −90 β-ZF-DBD in MEL cells, the analysis of selected sites that resemble the targeted site suggests that the −90 β-ZF-DBD preferentially associates with the β-minor globin gene promoter. This is further supported by our observation that transient transgenic embryos expressing the −90 β-ZF-DBD ubiquitously had reduced β-globin gene expression but did not reveal any other developmental abnormalities. The study presented here demonstrates that ZF-DBDs without effector domains can be used to alter expression patterns of genes in vivo. These artificial proteins thus provide a tool for modulating and analyzing *cis*-regulatory DNA elements and may also impact therapeutic approaches aimed at altering expression of specific genes.

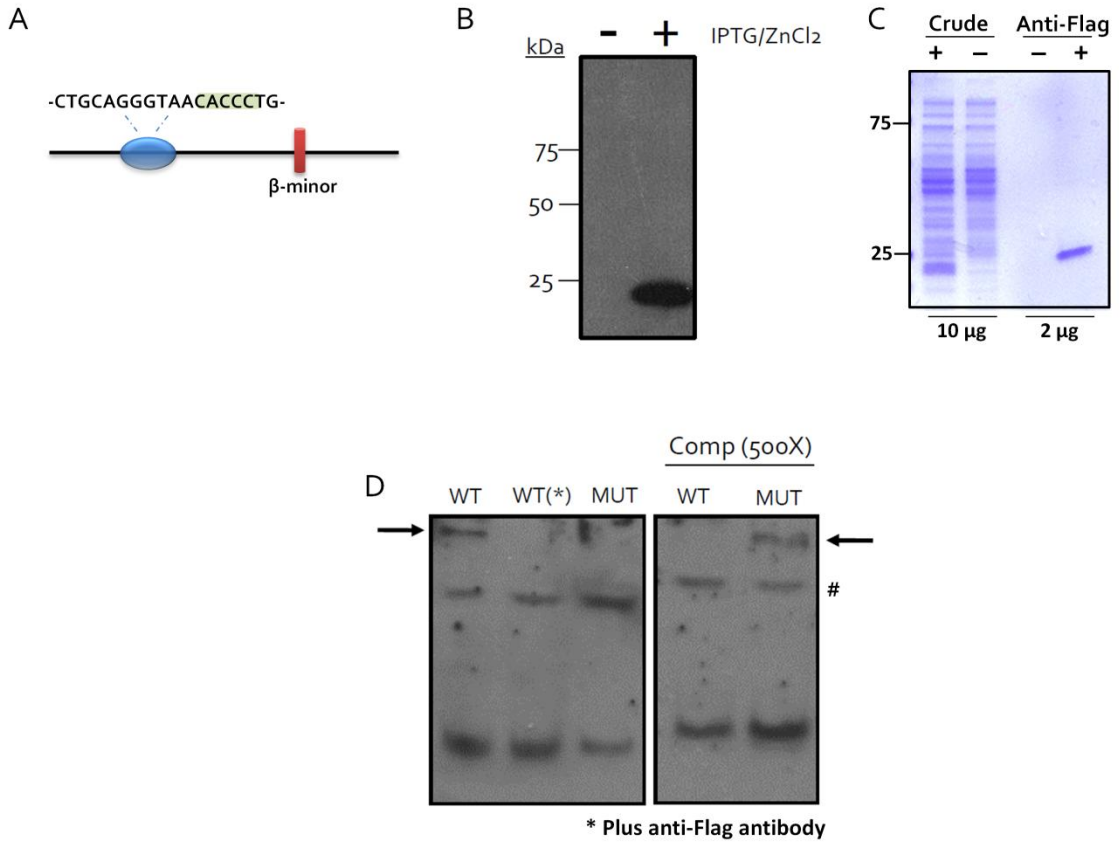


Figure 3-1. *In vitro* characterization of the -90 β-ZF-DBD. (A) Representation of the adult murine β-minor globin gene with the -90 β-ZF-DBD (oval) binding to the -90 KLF1 binding site upstream of the β-minor globin gene. The 18 bp sequence targeted by the -90 β-ZF-DBD is shown with the CACCC box highlighted in green. (B) Immunoblot using a flag-specific antibody. The -90 β-ZF-DBD migrated with an apparent molecular weight of 24 kDa. (C) SDS-PAGE and subsequent Coomassie stain of induced and uninduced E.coli (lanes 1 and 2), respectively. The -90 β-ZF-DBD was immunopurified from crude protein lysates using anti-flag magnetic beads (lane 4). (D) Electrophoretic mobility shift analysis (EMSA) of the purified recombinant -90 β-ZF-DBD using oligonucleotides representing the -90 KLF1 site (WT) or a mutant oligonucleotide (left panel lanes 1 and 3). Binding of the -90 β-ZF-DBD to the labeled WT oligonucleotide was abolished in the presence of 500 molar excess of unlabeled WT competitor but was unaffected by excess of unlabeled mutant oligonucleotides (right panel). The lane labeled WT (*) included a Flag-specific antibody during the binding reaction (left panel).

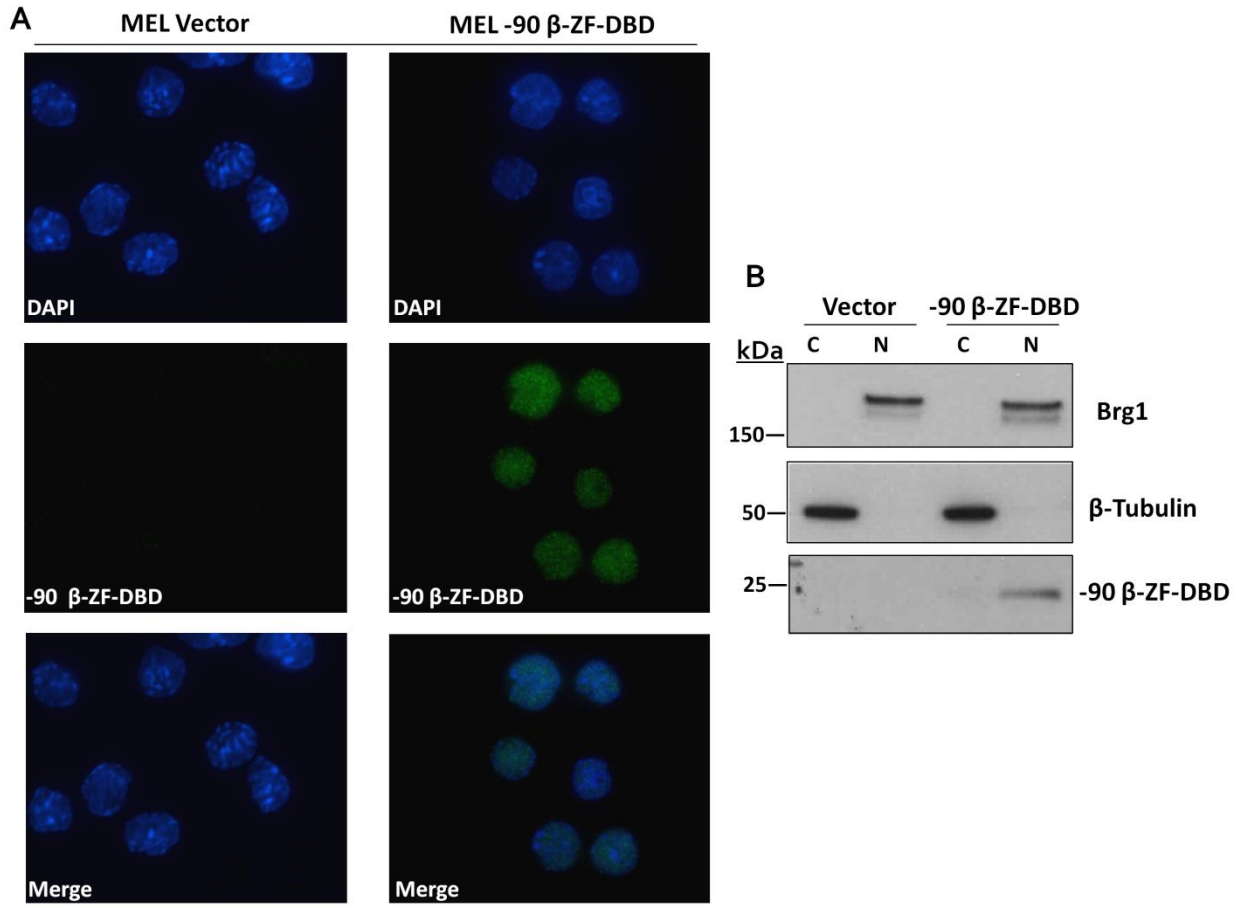


Figure 3-2. The -90 β-ZF-DBD localizes to the nucleus in induced MEL cells. (A) To visualize the location of the ZF-DBD, MEL cells were stably transduced, fixed and stained. DAPI stains chromatin in the nucleus (blue) and the -90 β-ZF-DBD was visualized using antisera against the ZF backbone (green). The bottom panel represents the merge of the two previous panels. (B) Induced stable MEL cells were fractionated into cytoplasmic [C] and nuclear fractions [N] which were subjected to immunoblot analyses using antibodies specific for Brg1, α-tubulin, and anti-Flag, as indicated.

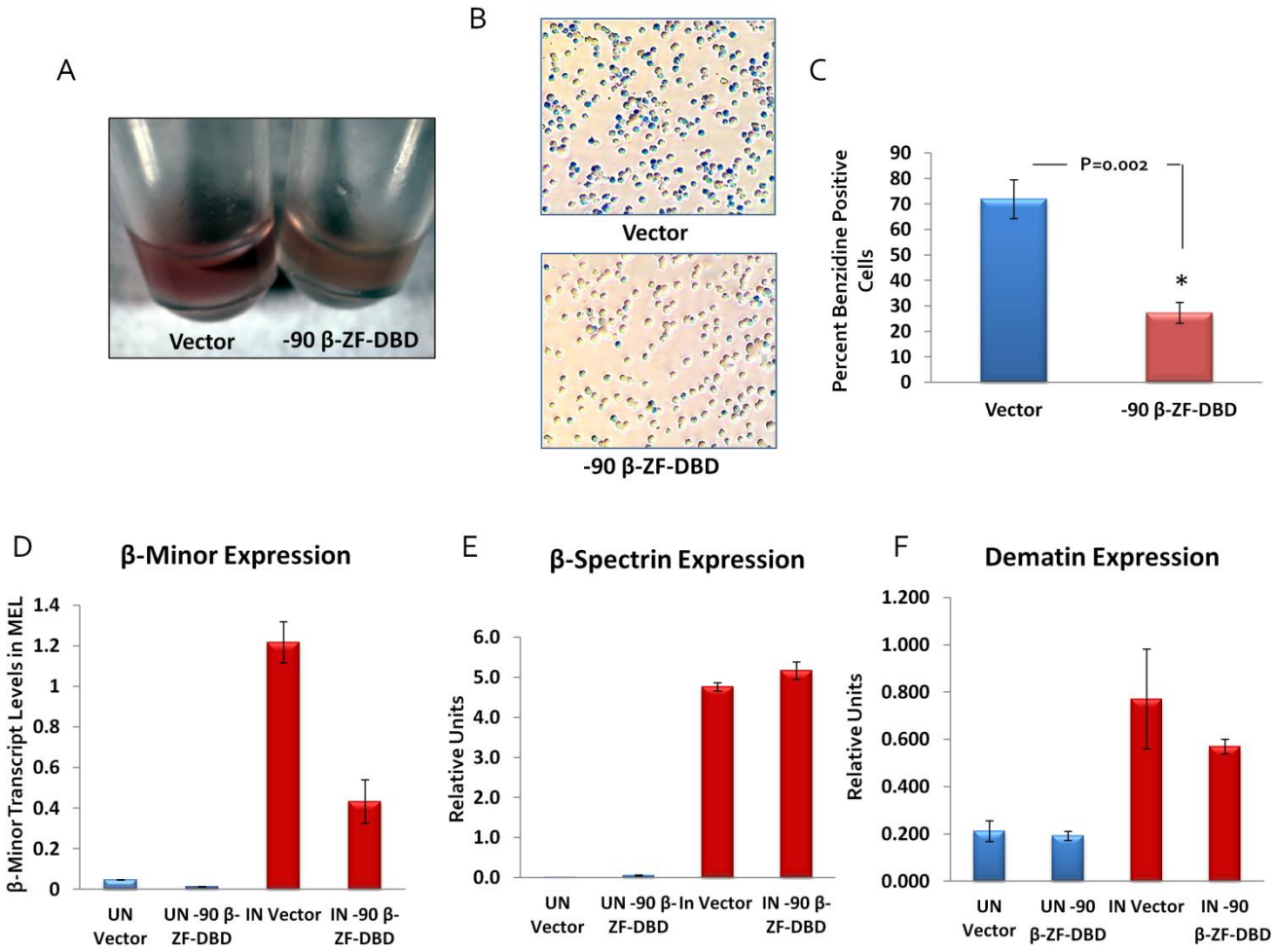


Figure 3-3. Specific reduction in βmin-globin gene transcription in MEL cells expressing the -90 β-ZF-DBD. (A) Phenotypic analyses of induced MEL cells that were chemically induced for 72h with 2.5% DMSO. (B) Benzidine staining of MEL cells expressing the -90 β-ZF-DBD or harboring the empty vector. (C) Percent benzidine positive MEL cells expressing the -90 β-ZF-DBD with number of benzidine positive cells in vehicle control MEL cells set at 100%. D-F. Quantitative RT-PCR analysis of βmin-globin (D), β-spectrin (E), and dematin (F) transcripts in uninduced (UN) or induced (IN) MEL cells expressing either the -90 β-ZF-DBD or harboring the empty vector. Data presented is the average of 3 independent experiments ± SEM.

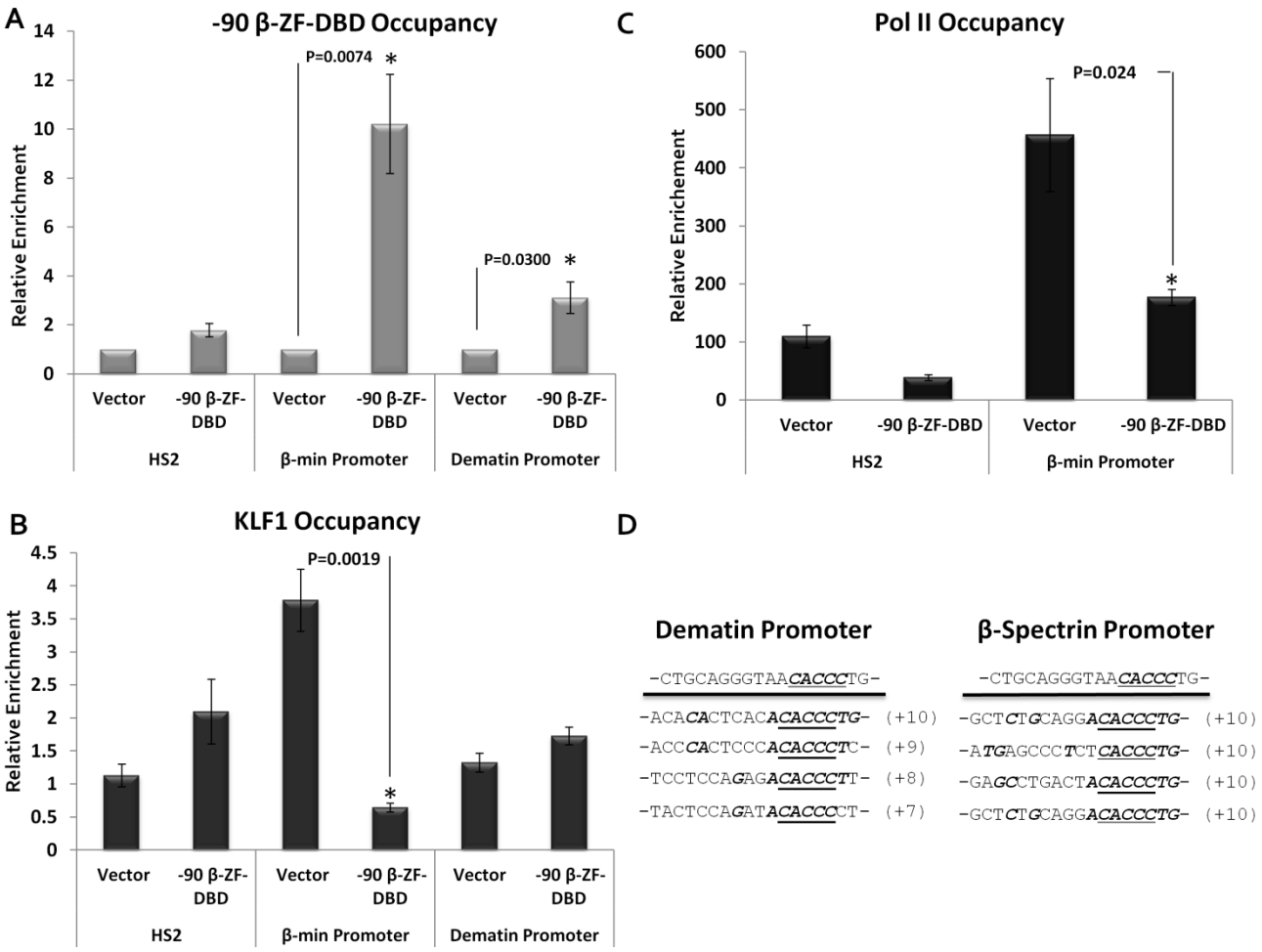


Figure 1-4. Displacement of KLF1 and RNA polymerase II from the βmaj-globin gene promoter by the -90 β-ZF-DBD in MEL cells. ChIP analysis of MEL cells expressing the -90 β-ZF-DBD or harboring the empty vector. (A and B) ChIP analysis of the interaction of the -90 β-ZF-DBD (A) or KLF1 (B) with LCR element HS2, the β-minor globin gene promoter, and the dematin promoter, as indicated. (C) ChIP analysis of the interaction of RNA polymerase II (Pol II) with LCR element HS2 and the β-minor globin gene promoter. Data represent two independent ChIP experiments with PCRs run in triplicate ± SEM. (D) Four sequences derived from either the dematin or β-spectrin promoter harboring potential CAC boxes were aligned with the -90 β-ZF-DBD target site (18bp sequence shown above the line). Sequence homology is indicated by the number in parentheses.

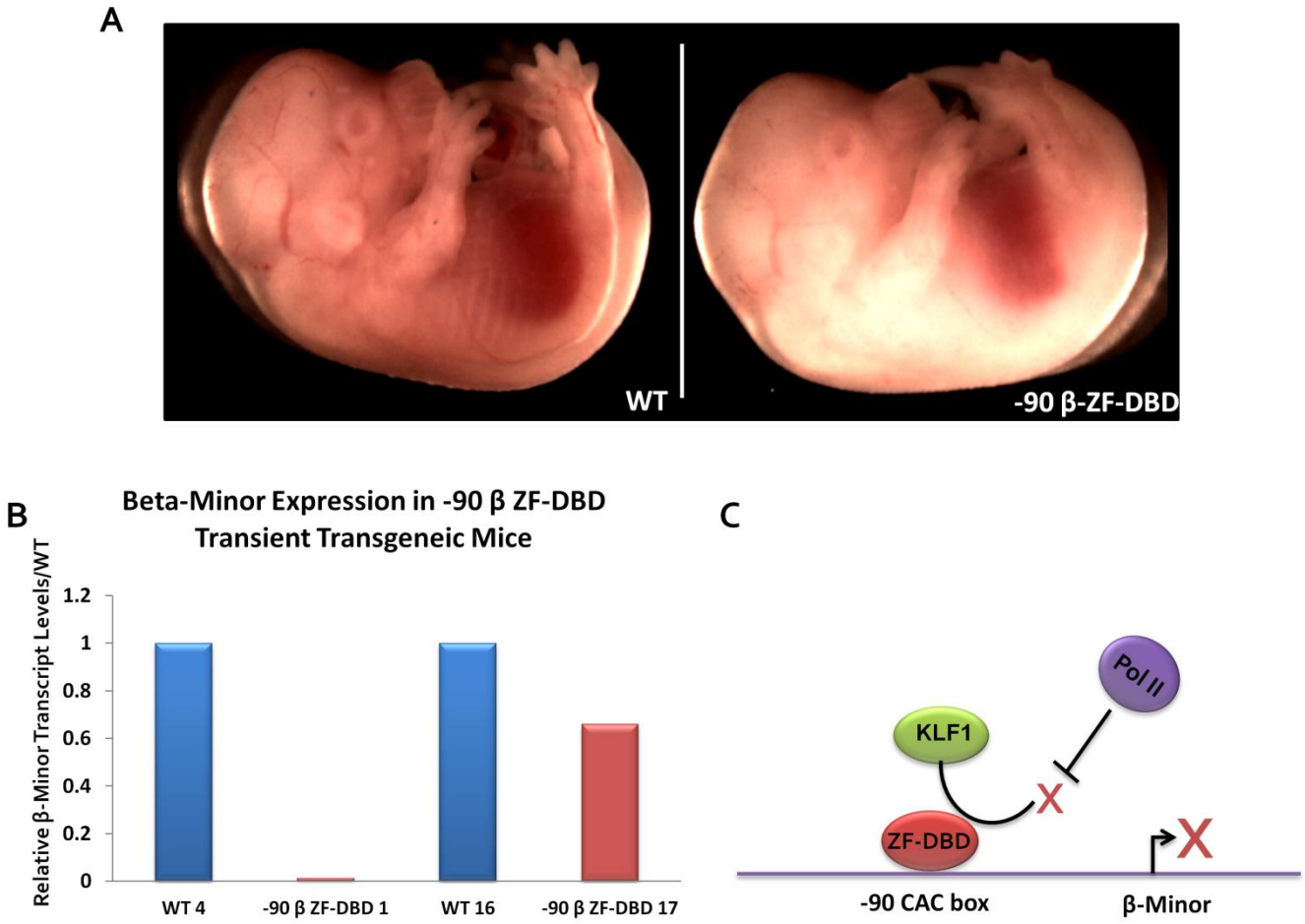


Figure 3-5. Transient transgenic mouse embryos expressing the -90 β-ZF-DBD exhibit reduced expression of the βmin-globin gene. (A) Images of 13.5 dpc wild-type (WT) and -90 β-ZF-DBD transgenic embryos. (B) Quantitative RT-PCR analysis of βmin-globin gene expression in WT and -90 β-ZF-DBD transgenic embryos. Expression levels in WT were set at 1. (C) Model illustrating ZF-DBD mediated disruption of KLF1 binding and Pol II recruitment at the adult βmin-globin gene promoter.

# CHAPTER 4
# CHARACTERIZING THE +60 EBOX ZF-DNA BINDING DOMAIN

## Background

*Cis*-element redundancy often presents a challenge in molecular biology especially when attempting to interrogate the functional potential of a *cis*-element *in vivo* [168]. To date, techniques such as ChIP sequencing can provide a means to address this problem; however, given that ChIP sequencing analyses reflect 100 bp reading frames that often contain numerous *cis*-elements, it remains difficult to identify which element truly harbors the function [202].

Artificial DNA-binding proteins comprised of zinc finger domains can surmount this challenge. Zinc finger DNA-binding domains (ZF-DBD) bind to DNA in a sequence-specific manner with high affinity and neutralize a target site. A single zinc finger domain can recognize 3 base pairs of DNA—an interaction that can be extended by adding additional zinc finger domains in tandem to allow the recognition of extended asymmetric DNA sequences [118,181]. The benefit of employing ZF-DBDs includes rapid delivery and more importantly, enhanced resolution for *cis*-element detection. The enhanced resolution allows targeting of sequences ranging from 6 to18 base pairs with 18 being optimal as this sequence length confers a unique signature within the genome [175].

We employ the ZF-DBD technology to evaluate the putative function of E-box elements, comprised of the sequence 5’-CANNTG-3’, within the murine beta-globin promoter region. Helix-loop-helix proteins such as the heterodimeric transcription factor upstream stimulatory factor (USF), which is composed of the closely related proteins USF1 and USF2, binds to E-box elements and has been shown through ChIP analysis

to occupy the beta-globin promoter region where it is involved in mediating high level expression of the adult beta-globin gene [147,156]. Given that there are two predominant E-box elements located at positions +20 and +60 relative to the murine beta-major transcription start site, it is difficult by conventional methods to discern which E-box element truly harbors the function *in vivo* [100]. Understanding the intricacies of beta-globin regulation is warranted to improve therapies for individuals with hemoglobinopathies and detailing which of the two E-box elements bind functional USF is critical and could provide insight as to a potential target site for therapy.

Given that *in vitro* mutagenesis studies indicate that the +60 E-box elements likely carries the functional relevance, I designed a ZF-DBD to target and neutralize 18bp encompassing the +60 E-box to determine functional relevance *in vivo* [194]. Neutralizing this element will prevent endogenous transcription factors from interacting with the target *cis*-element and beta-major expression can be evaluated. In our study, targeting the +60 E-box led to a reduction in β-major globin expression in a dose-dependent manner relative to the amount of ZF-DBD protein expressed in the murine erythroleukemia cell line. I also observed a dose-dependent reduction in RNA polyermase II (Pol II) recruitment to the β-major promoter region relative to vehicle control cells. This suggests that the +60 E-box is critical for high level expression as well as recruitment of Pol II to the beta-globin gene.

In order to ensure that the introduction of a ZF-DBD to the beta-globin locus is not the causative factor for altering gene expression, I stably expressed a ZF-DBD targeted towards a region -2KB upstream of the beta-major promoter that is predicted to be inert based on global ChIP data from UCSC genome tracks. Neutralizing this site did

not alter beta-globin expression indicating that the effects seen with the +60 ZF-DBD is due to specific neutralization of the *cis*-element.

## Results

ZF-DBDs derived from the Barbas modules were designed to neutralize the +60 E-box and -2KB *cis*-elements [176]. The designed ZF-DBD is composed of 6 zinc finger domains to target 18bp regions flanking the desired target sites (Figure 4-1A). The coding region for the ZF-DBD was modified to include a Flag and nuclear localization signal (NLS) tags and placed into the retroviral pMSCV vector. Murine erythroleukemia cells (MEL) were then transduced to generate stable populations for the -2KB ZF-DBD (denoted as the negative control (NC)-ZF-DBD)) through geneticin sulfate selection. Single cell clones stably expressing either low levels, or medium levels of the +60 E-box ZF-DBD protein, were selected to evaluate and titrate the concentration of ZF-DBD needed to provide optimal neutralization (Figure 4-1C).

To confirm nuclear localization, MEL cells harboring the -2KB and the +60 ZF-DBDs were fractionated into cytoplasmic and nuclear fractions (Figure 4-1B and C respectively). The nuclear protein BRG1 and the cytoplasmic protein β-tubulin were used as controls to indicate complete fractionation. The ZF-DBDs localize to the nuclear fractions in MEL cells while displaying no staining for vehicle treated cells.

To determine the effect on adult β-major expression, RNA was extracted from uninduced and DMSO-induced MEL cells. Quantitative RT-PCR analysis in stable MEL cells harboring the +60-ZF-DBD indicated that β-major globin expression levels were significantly decreased compared to vehicle control cells. Furthermore, this decrease occurred in a dose-dependent manner relative to the concentration of ZF-DBD protein

present (Figure 4-2B). To confirm the specificity of targeting, additional E-box containing erythroid specific genes such as Dematin and β-Spectrin were evaluated and displayed no change in expression compared to control cells (Figure 4-2B). As another form of control, β-major globin transcripts from MEL cells harboring the NC ZF-DBD targeting a region a few kilobases upstream did not change compared to control MEL cells (Figure 4-2A). This indicates that the effect observed with the +60 ZF-DBD containing cells was due to the specific targeting and neutralization of the +60 E-box cis element by the ZF-DBD and not simply due to the introduction of an artificial DNA-binding protein to the β-major globin promoter region.

Next to confirm the specificity of the ZF-DBDs, I examined target occupancy by chromatin immunoprecipitation (ChIP) in induced MEL cells. Both the +60 and NC-ZF-DBDs occupied the designated target regions at the β-major promoter and -2KB regions respectively (Figure 4-3A and B).To evaluate specificity of targeting, other regions within the beta-globin locus were examined.  MEL cells harboring low levels of the +60 ZF-DBD did not display significant binding to other select regions within the beta-globin gene locus highlighting the specificity of the ZF-DBD. MEL cells harboring higher levels of the +60 ZF-DBD however did have significant off-target binding at the -2KB and 3’ β-major regions. This could be due to excessive levels of protein or saturation of the binding to the target site causing it to bind to other regions. The NC ZF-DBD binding to the -2KB site though significant, was weak over non-specific controls. This was likely attributed to the fact that ChIP was performed on a population level with heterogeneous expression of the NC ZF-DBD resulting in increased variability and less evaluative power relative to the homogeneous expression in the single cell clones.

To determine the effect of neutralization the target sites, I examined candidate erythroid-specific transcription factors that are known to bind to E-box sequences such as the Upstream Stimulatory Factor heterodimer complex (USF1 and USF2) as well as the T-cell acute lymphocytic leukemia protein 1 (TAL1). Both USF and to a lesser extent TAL1, occupied both HS2 of the LCR and the β-major promoter in induced vehicle treated control MEL cells. Interestingly, USF2 levels displayed reduced occupancy at the β-major promoter while TAL1 displayed increased occupancy. Though these findings did not cross the threshold of significance, it suggests that the USF1/USF2 complex may be the transcription factor that predominately interacts with the +60 E-box *cis*-element. The increased binding of TAL1 could be a compensatory mechanism perhaps through the interaction with the +20 E-box element. Although this hypothesis cannot be verified by ChIP as the two regions are too close in proximity, it is an attractive explanation of why I see a trending increase in TAL1 occupancy at the β-major promoter region. Curiously, a significant and dramatic reduction in USF2 occupancy was also observed at the LCR HS2 element. This is an intriguing finding and may suggest co-regulation between HS2 and β-major promoter regions. Indeed, HS2 of the LCR and the β-major promoter are known to come into close proximity as determined by 3C analysis and this interaction is critical for the high level expression of the adult globin genes. TAL1 occupancy at HS2 did not display significant differences compared to control MEL cells. Of further note, both USF2 and TAL1 displayed altered occupancy at the 3' end of the β-major gene in MEL harboring the +60 ZF-DBD—a phenomenon that warrants further investigation.

Since it is hypothesized that USF is involved in the recruitment of the RNA polymerase II (Pol II) to the LCR for subsequent transfer to the β-major promoter, the reduction in USF2 occupancy due to the neutralization of the +60 E-box element observed may also lead to a reduction in Pol II binding at the LCR and the β-major promoter. To test this hypothesis, I mapped Pol II occupancy in induced MEL cells harboring the +60 ZF-DBD (Figure 4-5 A and B). Vehicle control cells display the characteristic binding pattern with accumulation of Pol II both at the LCR and the β-major promoter with levels tailoring slightly as the 3' end of the β-major gene is approached. In MEL cells harboring the +60 ZF-DBD however, total Pol II levels, as well as the serine 2 modified, representing the initiating form of Pol II, displayed a significant and surprising increase in occupancy at HS2 of the LCR rather than a decrease as expected. This increase occurred in a dose-dependent manner with respect to +60 ZF-DBD expression and suggests that Pol II is being stalled and therefore accumulating at HS2. Indeed, upon examination of Pol II occupancy at the β-major promoter and within the body of the gene, levels were significantly reduced compared to vector control cells. The level of reduction also displayed a dose-responsive association. I also observed an increase in occupancy of serine 2 phosphorylated Poll II at the 3' end of the β-major gene relative to vehicle control cells that warrants further investigation (Figure 4-5B).

Poll II as well as USF2 occupancy in the NC ZF-DBD-containing induced MEL cells displayed no change from vector treated cells at the -2KB site as well as the β-major promoter regions consistent with the prediction that neutralizing the -2KB element should not perturb β-major globin regulation. I did however observe reduced occupancy of Pol II at HS2 of the LCR as well as an increase in USF2 levels which was attributed

to the variability typically observed when evaluating factor occupancy in a heterogeneous population (Figure 4-5C).

## Discussion

To assess whether the +60 E-box *cis*-element located 60 bp downstream of the β-major globin gene carried functional relevance *in vivo*, I designed an artificial zinc finger DNA binding domain (ZF-DBD) to neutralize 18bp encompassing the target site. Successful stable expression and localization of the ZF-DBD into murine erythroleukemia cells (MEL) was confirmed by compartmentalization immunoblot analysis. β-major globin expression was reduced in MEL cells harboring the +60 ZF-DBD in a dose-dependent manner while no changes to other selected erythroid-specific genes were affected indicating the specificity of targeting. Specificity to the target site was confirmed by ChIP analysis which indicated that the +60 ZF-DBD occupied the β-major promoter region in a specific dose-responsive manner. Non-specific occupancy was detected at the -2KB site as well as the 3' end of the β-major globin gene in a single cell clone that expressed higher levels of the +60 ZF-DBD. This could be due to over saturation of binding to the target site due to the increased levels of ZF-DBD protein causing transient occupancy at other regions. This indicates that it may be beneficial to select clones with a more moderate "physiological-like" expression pattern to prevent non-specific interactions.

Neutralizing the +60 E-box element led to a decrease in USF2 occupancy at the β-major promoter and curiously also at HS2 of the LCR as well as at the 3' end of the β-major globin gene. This was an intriguing finding and could be explained by the complex 3-dimensional interaction between the LCR and the β-major gene that has been extensively documented by 3C analysis [105]. Regardless, significant reductions in

USF2 recruitment were observed which suggests that this transcription factor may be the likely candidate that interacts with the +60 E-box element. Occupancy of TAL1 in comparison was largely unaffected, particularly at the LCR and the β-major promoter, where dramatic reductions in USF2 occupancy were seen. I did note a significant decrease in TAL1 occupancy compared to vehicle treated cells at the 3' end of the β-major gene which could indicate a potential role for TAL1 but this warrants further investigation. Clone 15 that harbored higher levels of the ZF-DBD protein exhibited increased TAL1 occupancy at the 3' end of the β-major gene. This may be challenging to interpret given the over-saturation of the protein producing non-physiological effects.

The most remarkable finding was the mapping of Pol II. Our hypothesis was that Poll II levels should be reduced at the β-major promoter as well as the LCR consistent with the reduction in USF2 occupancy since it has been shown that USF is involved in the recruitment of the polymerase first to the LCR for subsequent delivery to the β-major globin promoter. Conversely, I observed an increase in both total Poll II levels as well as the serine 2 phosphorylated form—a marker for active Pol II. ). A possible explanation for this finding is that the inhibition of USF binding to the +60 Ebox by the +60 ZFDBD interferes with the transfer of Pol II to the promoter. This is consistent with a previous study from the Bresnick laboratory [203]. The authors demonstrated that deficiency of transcription factor NF-E2, which can be crosslinked to the LCR and the adult β-globin gene promoter caused a reduction of Pol II binding at the promoter but not at the LCR. The data thus support a model according to which NF-E2 and USF collaborate in the transfer of Pol II from the LCR to the globin gene promoter [204]. The observation that USF binding at LCR HS2 decreased in cells expressing the +60 ZF-DBD suggests that

the binding of USF at the LCR is indirectly mediated through its association with the β-globin gene promoter.

In contrast, the NC ZF-DBD, through its specific binding at the -2KB element—a sequence predicted to be inert, revealed no change in β-major globin expression, and no difference in recruitment or occupancy of USF2 or Pol II at the β-major promoter. I did observe changes in Pol II and USF2 occupancy at HS2 of the LCR and the level of binding of NC ZF-DBD was relatively weak which is attributed to heterologous population variation. It will be beneficial to examine the effects of the NC ZF-DBD on a single cell level since the results on the population scale are promising.

In our study, I successfully demonstrate that ZF-DBDs can be used to effectively target and neutralize *cis*-regulatory elements in vivo. Due to the specificity of binding however as well as the high affinity for the target site over most endogenous factors, it may be only feasible to examine *cis*-regulatory elements outside of gene bodies. The targeting of ZF-DBDs to the transcribed region of a gene could be problematic due to the possibility that they may interfere with the elongation properties of Pol II. An indication for this is the reduced levels of serine 2 phosphorylated Pol II at the LCR and the β-globin gene in cells expressing the + 60 ZF-DBD. Barring this cautionary statement, artificial ZF-DBDs still remain extremely effective tools for examining the functional potential of select *cis*-regulatory elements *in vivo*.

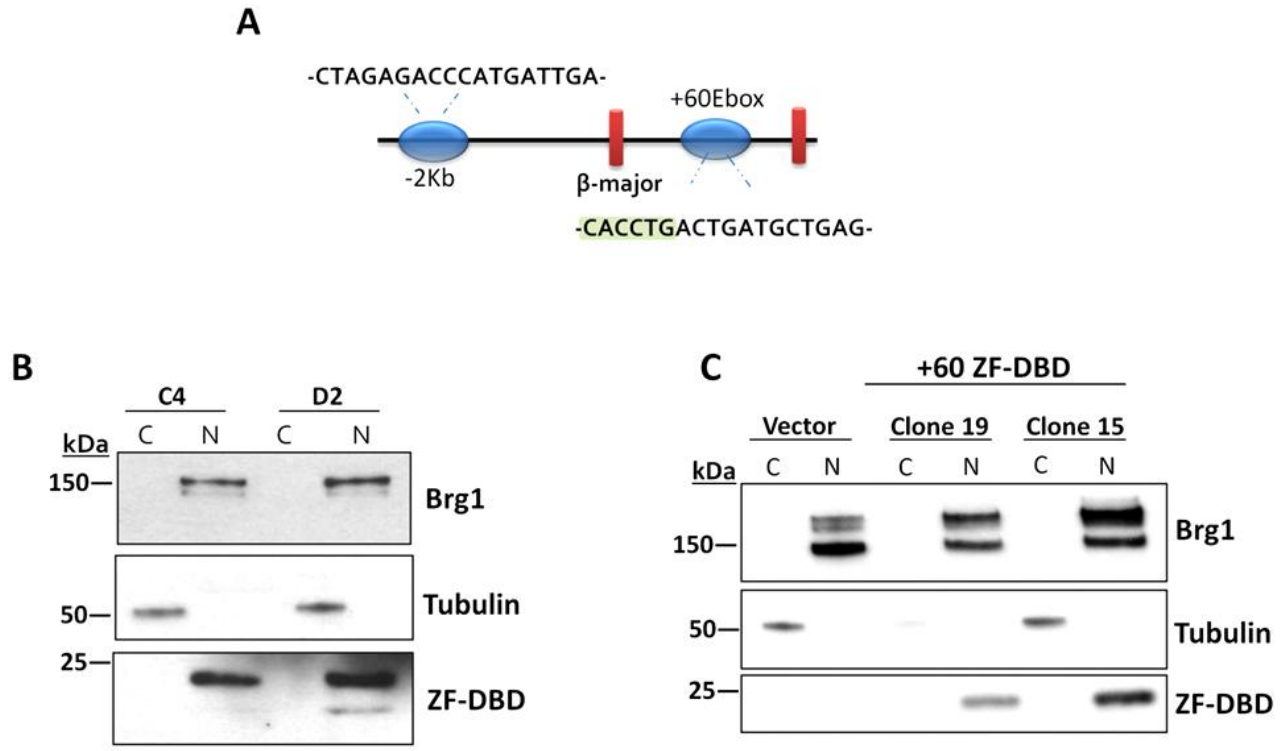


Figure 4-1. Zinc Finger DNA-binding domains (ZF-DBD) stably localize to the nucleus in MEL cells. A) Schematic of the target sites for the -2KB and +60 ZF-DBDs are shown relative to the start of the β-major globin gene. B) Compartmentalization immunoblot where cells were fractionated into cytosolic (C) and nuclear (N) fractions. Brg1 and Tubulin (top and middle panel respectively) represent controls. The -2KB ZF-DBD protein in MEL cell population C4 and D2 are shown and indicates nuclear localization (bottom panel). C) Two single cell clones reflecting low and medium protein expressing levels (clones 19 and 15 respectively) both localize the ZF-DBD to the nuclear fraction in MEL cells.

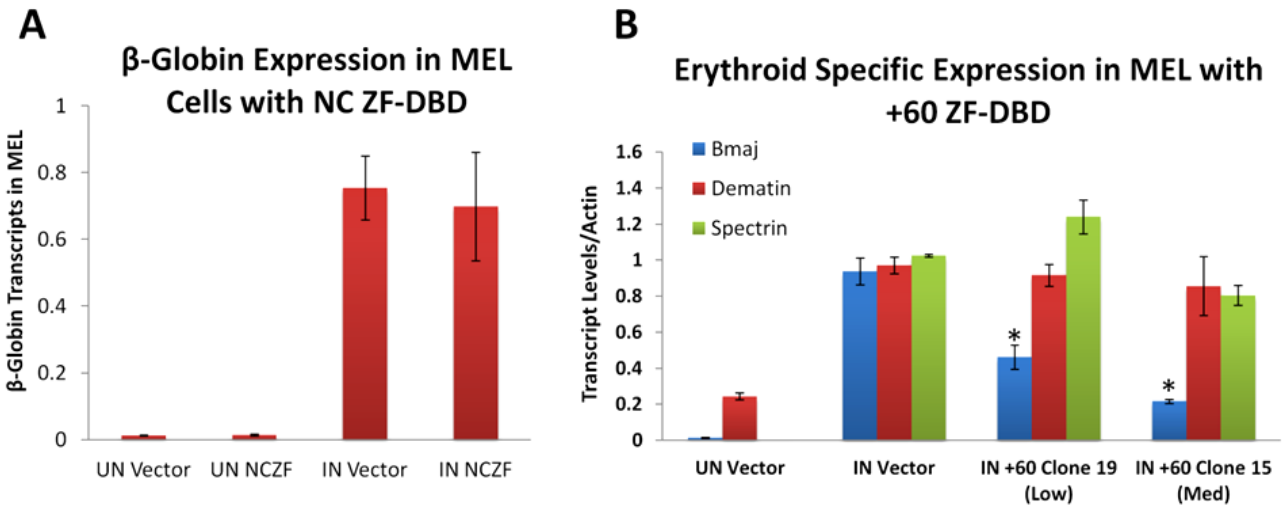


Figure 4-2. Specific reduction in β-major-globin gene transcription in MEL cells expressing the +60 ZF-DBD or empty vector. (A) RT-qPCR examining β-globin transcripts in uninduced and induced MEL cells harboring the NC ZF-DBD (B) Quantitative RT-PCR analysis of β-major globin, β-spectrin, and dematin transcripts in uninduced and induced MEL cells. expressing either the +60 ZF-DBD or harboring the empty vector. Data presented is the average of 2 independent RNA extractions with qPCR performed in triplicate ± SEM. Statistical analysis were based on student T test $p \leq 0.05$.

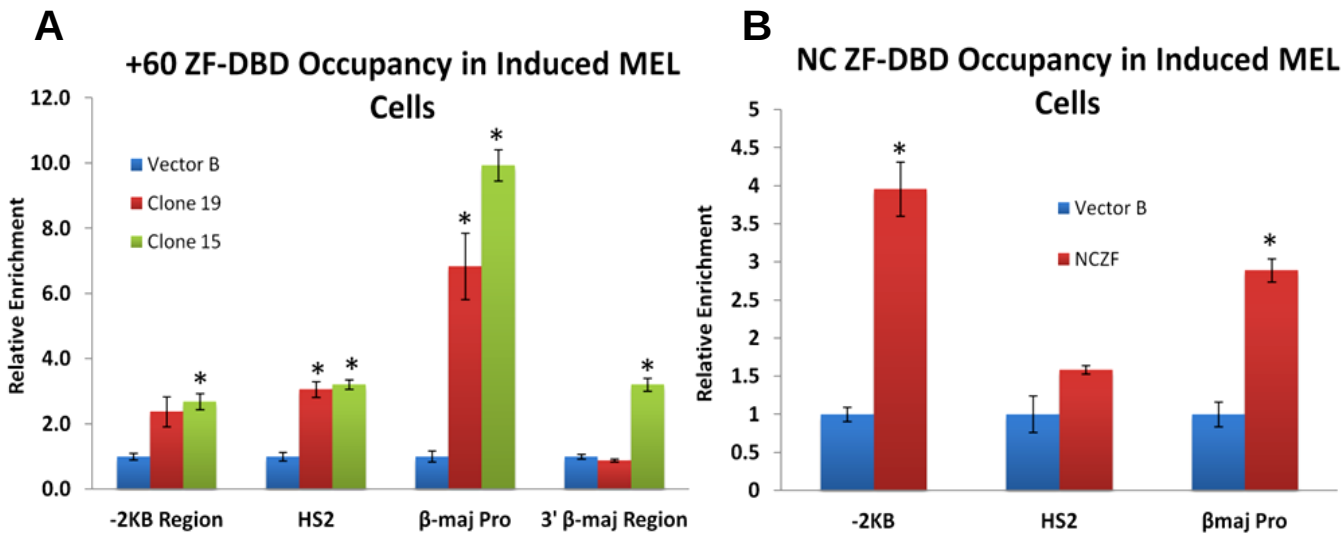


Figure 4-3. ZF-DBDs bind to the intended target sites. ChIP analysis of induced MEL cells expressing the +60 ZF-DBD, NC ZF-DBD, or harboring the empty vector control. (A) ChIP analysis of the occupancy of the +60 ZF-DBD to select sites in the β-globin locus. Two single cell clones with low (clone 19) or high (clone 15) expressing protein levels of the +60 ZF-DBD were analyzed. (B) Mapping of the NC ZF-DBD to select sites in the β-globin gene locus. Data represent two independent ChIP experiments with PCRs run in triplicate ± SEM. Statistical analysis were based on student T test $p \leq 0.05$.

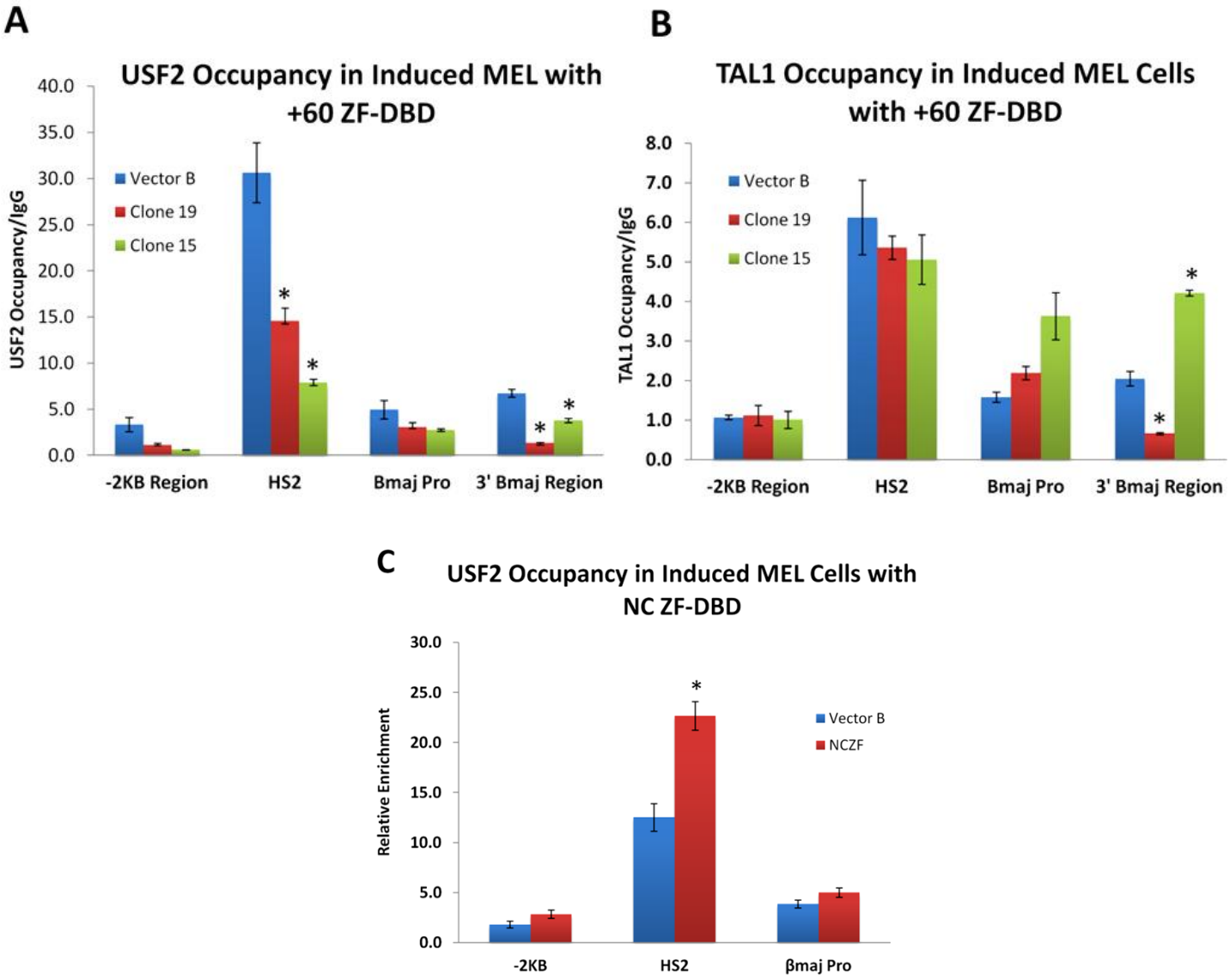


Figure 4-4. USF2 occupancy is reduced at HS2 of the LCR and the β-major promoter in induced MEL cells harboring the +60 ZF-DBD. ChIP analysis of induced MEL cells expressing the +60 ZF-DBD, NC ZF-DBD, or harboring the empty vector control. ChIP analysis of the occupancy of the USF2 (A) and TAL1 (B) to select sites in the β-globin locus. Two single cell clones with low (clone 19) or med (clone 15) expressing protein levels of the +60 ZF-DBD were analyzed. (C) Mapping of the USF2 to select sites in the β-globin gene locus in induced MEL cells expressing the NC ZF-DBD. Data represent two independent ChIP experiments with PCRs run in triplicate ± SEM. Statistical analyses were based on student T test $p \leq 0.05$.

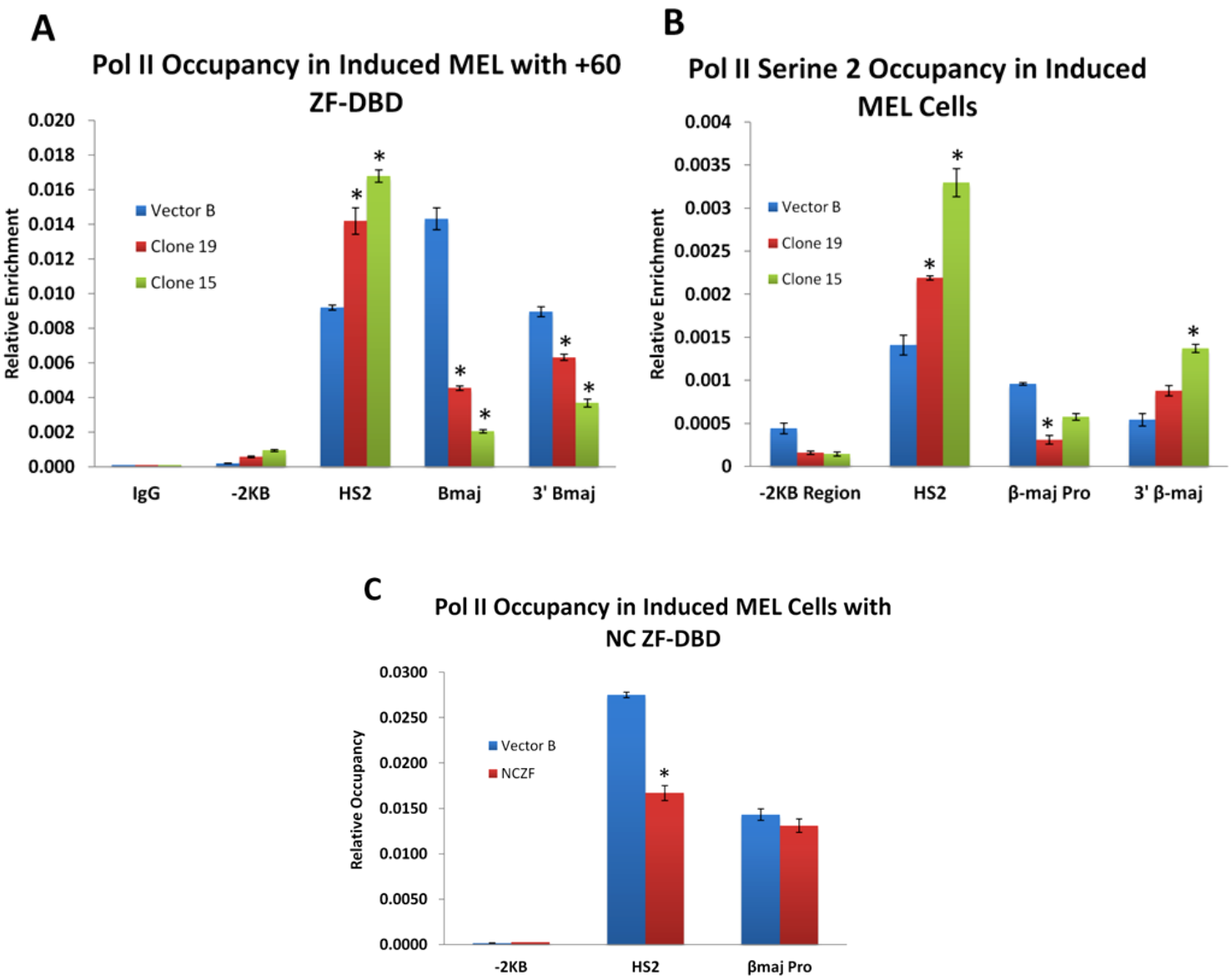


Figure 4-5. Pol II occupancy levels accumulate at HS2 of the LCR and dramatic reductions in β-major promoter occupancy is observed in induced MEL cells harboring the +60 ZF-DBD. ChIP analysis of induced MEL cells expressing the +60 ZF-DBD, NC ZF-DBD, or harboring the empty vector control. ChIP analysis of the occupancy of the RNA Pol II (A) and the RNA Pol II serine 2 phosphorylated form (B) to select sites in the β-globin locus. Two single cell clones with low (clone 19) or high (clone 15) expressing protein levels of the +60 ZF-DBD were analyzed. (C) Mapping of the Pol II to select sites in the β-globin gene locus in induced MEL cells expressing the NC ZF-DBD. Data represent two independent ChIP experiments with PCRs run in triplicate ± SEM. Statistical analyses were based on student T test $p \leq 0.05$.

# CHAPTER 5
# CONCLUSIONS AND FUTURE STUDIES

*Cis*-element redundancy is a significant challenge in molecular biology [168]. The question of how a *trans*-acting factor can direct its functional activity to a subset of 5-mer sequences while displaying no affinity for the other over-represented identical sites is still a mystery that remains to be elucidated. Several lines of evidence suggests that the cell and tissue specific expression of these *trans*-factors as well as the 3-dimensional structure of chromatin introducing the interplay of eu- and heterochromatin play significant roles in the segregation and availability of certain sequences over others [187,205,206]. Yet, techniques such as ChIP-sequencing and DNAseI *in vivo* footprinting indicate that there still remains a significant degree of *trans*-factor binding to *cis*-elements to which only a subset truly harbors function [207]. The question then is simple: how can investigators assess functional relevance *in vivo*.

This phenomenon holds true in the beta-globin gene locus. The locus is enriched for a large number of *cis*-regulatory elements that, through precise regulation and binding of *trans*-acting factors, coordinate the developmental and stage-specific expression of the β-globin genes [3]. Many *cis*-elements have been resolved and functionally defined, yet many more remain to be characterized [183]. Classical approaches to identify the *in vivo* functional potential of a *cis*-element such as knock out or gene deletion studies are labor-intensive, expensive, and often times result in alterations of local architecture of the surrounding chromatin which could confound results obtained [208].

In our study, we employed the properties of the zinc finger motif to create artificial DNA-binding domains. The zinc finger DNA-binding domains (ZF-DBD) bind and

neutralize 3 base pairs of DNA in a sequence-specific manner [118]. I designed ZF-DBDs each comprised of 6 zinc finger motifs to bind and neutralize 18 base pair DNA arrays. I successfully directed the ZF-DBDs to 3 target sites within the murine β-globin gene locus: the -90 CACC box predicted to interact with KLF1, the +60 E-box believed to interact with the USF1/USF2 heterodimer complex, and the -2KB site as a control region that is predicted to be inert. ChIP studies indicated that all ZF-DBDs bound to their target sites and successful neutralization of these sites by the ZF-DBDs allowed for *in vivo* functional characterization of the putative *cis*-elements. Both neutralization of the -90 CACC box as well as the +60 E-box resulted in dramatic and significant reductions in β-globin expression indicating that these sites are critical for high level expression of the adult β-globin gene [191]. The -2KB site was predicted to be inert and accordingly, a ZF-DBD targeting this site did not change β-globin expression. In order to determine the mechanism of action, I mapped the occupancy of trans-acting factors predicted to interact with the target *cis*-element. In the -90 CACC box study, KLF1 the transcription factor predicted to interact with that site revealed significantly reduced occupancy compared to vector-treated controls. This also led to a decrease in Pol II levels. USF2, the predicted transcription factor to interact the target *cis*-element also displayed reduced occupancy towards the neutralized site (the +60 E-box). The -2KB study confirmed that factors associated with either the -2KB site or the β-major promoter were unchanged.

Based on the success of our studies, we conclude that ZF-DBDs are effective tools to identify and characterize putative functional *cis*-regulatory elements *in vivo*.

## Future Studies

In future work, we would like to continue to evaluate putative functional *cis*-regulatory elements. Our primary focus will be to evaluate *cis*-elements that could increase the expression of fetal globin genes. This stems from patients with hemoglobinopathies of varying levels of disease severity that because of genetic mutations that give rise to high persistence of fetal hemoglobin, exhibit a clinically benign phenotype [47]. The fetal form of hemoglobin provides compensation for the impaired adult globin chain synthesis and/or structure. As such, identifying mechanisms that will increase fetal hemoglobin levels is the ideal form of therapy. Neutralizing key *cis*-regulatory elements, particular DNA sequences that contribute to the silencing of the fetal globin genes such as those that regulate BCL11A protein expression—a key silencer for fetal globin expression, would lead to potential reactivation of the fetal globin genes [209]. Indeed this has been observed in a recent study in which BCL11A was conditionally ablated by genetic knock out under the control of the erythropoetin receptor (EPOR)-GFP Cre in sickle cell disease mice (SCD). Conditional knockout re-activated the fetal globin genes and increase levels to therapeutic effects [210]. Given that the regulation of fetal hemoglobin expression is highly complex, there are likely other sequences that are important for re-activation of this gene. These sequences will be identified by examining genome wide association studies (GWAS) and neutralized using ZF-DBDs.

Following this line of study, we would like to continue to evaluate the effects of employing ZF-DBDs *in vivo*. Our *in vivo* results generated by globally expressing the ZF-DBD in a transient manner in transgenic gave promising results and indicated that the ZF-DBDs can be used in a live animal model. Indeed, other groups have reported

similar successful findings in different disease models [211]. To refine our studies, we are interested in generating stable mouse lines to allow the integration of ZF-DBDs into the germ line. We could then examine the efficacy and specificity of ZF-DBD-mediated neutralization of target *cis*-elements.

Inducible promoters such as the tetracycline responsive promoter system will also be beneficial in this case to allow for the control of ZF-DBD expression. We could then evaluate the importance of putative *cis*-regulatory elements during different stages of development. Cell type-specific expression will also be desired to minimize any potential non-specific binding. In previous studies in the Bungert lab, constructs were designed to confer erythroid-specific expression [152]. These constructs involved the placement of the target gene, in our case the ZF-DBD, under the control of the human β-globin promoter and followed immediately by the β-globin 3' enhancer. Only *trans*-acting factors in erythroid cells will activate this promoter. Furthermore, sequences for HS2 to HS3 including the intervening linker region will be included to confer high level expression of the ZF-DBD. Finally, the construct contains chicken HS4 sequences that flank the target gene which are insulator elements to prevent position effects. This will be an ideal construct for the cell type-specific expression of ZF-DBDs.

**Therapeutic Delivery of the ZF-DBDs**

The ultimate goal for our study is to treat hemoglobinopathies. Once ZF-DBDs are thoroughly characterized in *in vivo* animal models, the next stage would be to administer the protein to the human population in a safe, specific, and minimally invasive manner. Given the reported challenges surrounding the gold standard AAV-mediated viral delivery system which has displayed problems with host-immune

responses and the maintenance of robust titers; alternative therapeutic delivery systems have been proposed [212]. One such system is to employ protein-based delivery mechanisms of which there are many classes, but one that shows the most promise are the nano-particles-based delivery systems. These are synthetic nanocarriers in which the desired protein, in our case, the ZF-DBD would be adhered to the surface of the particle for incorporation into the target cell [213]. The nano-particles range from 20-200nm in size which is optimal for *in vivo* delivery and they are typically composed of a combination of liposomes, polymers and inorganic materials. Mechanisms of nano-particle incorporation vary depending on the type and composition of liposomes and polymers used, but it typically employs the endocytotic pathway or receptor-mediated incorporation [213]. The unique composition of the lipids allows for efficient entry into the cell and release from the lysosomes to yield maximal delivery of the target protein to the cell. Popular lipid compositions include the use of trifluoroacetylated lipopolyamines (TFA-DODAPL) in combination with dioleoyl phopsophtidylethanolamine (DOPE) [214].

In contrast to gene delivery, intracellular protein delivery avoids permanently altering genomic information which could lead to mutagenesis. They display high internalization efficiency, long circulation and minimal internalization by the liver and renal system for clearance [215,216,217]. The synthetic nature of these carriers allow for precise design to optimize delivery as a function of desired cell type, immune response, and circulization time [218]. Some optimization techniques include introducing cell penetrating peptides (CPP) such as the TAT sequence derived from HIV, or the use of gold for the nano-particles since it is bioinert which will minimize toxicity and improve penetration [219,220]. Currently the FDA has approved over 140

nano-particle based delivery systems and will be a promising system to deliver ZF-DBDs.

## Challenges and Alternative Strategies

There are several documented challenges with the zinc finger technology. These include only a 49 out of 64 triplet recognition capacity and varying affinity for target triplets with TNN and ANN being weaker than CNN and GNN [221]. As mentioned previously, this preference for GNN could be because the original zinc finger proteins used as platforms for generating the amino acid-nucleic acid interaction algorithm were Sp1 and Zif268 which bind to GC rich regions. If the target region desired is a GNN rich sequence, zinc finger based artificial proteins are ideal [221]. Other sequence variations require extensive validation to ensure the affinity and specificity of the ZF is maintained. Furthermore, little is known about the context effects on individual fingers in an array which could alter the predicted specificity [222]. In our studies, we did not encounter most of these challenges mainly because we had a large degree of flexibility in our target sequence. However, if neutralization of a fixed sequence is desired, then the limitation associated with ZF-DBDs may be a significant concern.

Another alternative approach that has received much interest are the transcription-like activator (TAL) effectors proteins [223]. TAL effector proteins are DNA binding proteins that were first identified from the Xanthomonas class of plant pathogenic bacteria. The bacteria inserted these proteins into a plant cell by the bacterial type III secretory system to translocate to the nucleus and bind to DNA and alter gene expression—presumably to create a more favorable condition to the bacteria [224]. The TAL effector proteins can be arranged in an array of tandem polymorphic amino acid repeats that recognize DNA in a modular fashion through single contiguous

nucleotide recognition [225]. The DNA binding domain is composed of 34 amino acids and can be repeated 13 to 28 times. Each DNA binding domain has select polymorphisms known as repeat-variable-di-residue (RVD) that can occur at position 12 and 13 of the protein domain [195]. There are 4 different RVD to recognize each nucleotide and the algorithm is as follows: HD →C, NG→T, NI→A, and the finally polymorphism NN preferentially recognizes G but has been reported to recognize the adenine nucleotide as well. The number of repeats, as well as the string of RVDs is what determine the sequence recognition length and nucleotide composition [226].

To date custom TAL effector arrays are being used to target varying length of DNA sequences. The benefits over ZF system is the single nucleotide recognition instead of triplet recognition, which again, some triplets are not recognized. In addition, TAL effector design and customization are free from context-dependent arrays that could lead to changes in specificity and off-target effects that have been seen with zinc fingers [223,227]. The disadvantages to the technique however, is that the exact mechanism of DNA binding remains to be elucidated as a crystal structure is not yet available. Furthermore, these proteins are extremely large which could present problems downstream especially in terms of the feasibility of packaging for therapeutics. To address the extent of off-target binding for both ZFDBDs and TALE proteins genome wide protein/DNA interaction maps should be generated. This could be achieved by ChIP-sequencing experiments if appropriate antibodies are available.

In final concluding remarks, we employed a novel technique to study the *cis*-regulatory elements of the β-globin gene locus in order to identify DNA sequences that carry strong regulatory potential. ZF-DBDs provide an advantage over traditional

approaches because they can be rapidly employed while maintaining the specificity and efficacy that parallels that of traditional methods. We designed a series of artificial DNA-binding domains (DBD) comprised of modular zinc finger DNA binding domains (ZF-DBD) that targeted putative cis-regulatory sites of interests: the -2KB, -90 and +60 elements relative to the start of the β-globin promoter. The ZF-DBDs bound to these sites with high affinity and specificity and rendered the target sites inaccessible to endogenous transcription factors. We were able to successfully validate each target *cis*-element and assessed its requirement for globin gene function. Based on the success of our studies, we conclude that ZF-DBDs are effective tools to identify and characterize putative functional *cis*-regulatory elements *in vivo.*

LIST OF REFERENCES

1. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, et al. (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature 185: 416-422.

2. Perutz MF (1978) Hemoglobin structure and respiratory transport. Sci Am 239: 92-125.

3. Stamatoyannopoulos G, Nienhuis AW, Majerus P, Varmus H (1994) The Molecular Basis of Blood Diseases; Company WBS, editor. Philadelphia, PA.

4. Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, et al. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci U S A 96: 5129-5134.

5. Hecht F, Motulsky AG, Lemire RJ, Shepard TE (1966) Predominance of hemoglobin Gower 1 in early human embryonic development. Science 152: 91-92.

6. Huehns ER, Dance N, Beaven GH, Keil JV, Hecht F, et al. (1964) Human Embryonic Haemoglobins. Nature 201: 1095-1097.

7. Schroeder WA, Huisman TH, Shelton JR, Shelton JB, Kleihauer EF, et al. (1968) Evidence for multiple structural genes for the gamma chain of human fetal hemoglobin. Proc Natl Acad Sci U S A 60: 537-544.

8. Peschle C, Mavilio F, Care A, Migliaccio G, Migliaccio AR, et al. (1985) Haemoglobin switching in human embryos: asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. Nature 313: 235-238.

9. Hutton JJ, Bishop J, Schweet R, Russell ES (1962) Hemoglobin inheritance in inbred mouse strains. II. Genetic studies. Proc Natl Acad Sci U S A 48: 1718-1724.

10. Farace MG, Brown BA, Raschella G, Alexander J, Gambari R, et al. (1984) The mouse beta h1 gene codes for the z chain of embryonic hemoglobin. J Biol Chem 259: 7123-7128.

11. Jahn CL, Hutchison CA, 3rd, Phillips SJ, Weaver S, Haigwood NL, et al. (1980) DNA sequence organization of the beta-globin complex in the BALB/c mouse. Cell 21: 159-168.

12. Tuan D, Solomon W, Li Q, London IM (1985) The "beta-like-globin" gene domain in human erythroid cells. Proc Natl Acad Sci U S A 82: 6384-6388.

13. Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucleic Acids Res 15: 10159-10177.

14. Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51: 975-985.

15. Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, et al. (1989) A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature 338: 352-355.

16. Driscoll MC, Dobkin CS, Alter BP (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. Proc Natl Acad Sci U S A 86: 7470-7474.

17. Lowrey CH, Bodine DM, Nienhuis AW (1992) Mechanism of DNase I hypersensitive site formation within the human globin locus control region. Proc Natl Acad Sci U S A 89: 1143-1147.

18. Dhar V, Nandi A, Schildkraut CL, Skoultchi AI (1990) Erythroid-specific nuclease-hypersensitive sites flanking the human beta-globin domain. Mol Cell Biol 10: 4324-4333.

19. Bungert J, Tanimoto K, Patel S, Liu Q, Fear M, et al. (1999) Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription. Mol Cell Biol 19: 3062-3072.

20. Philipsen S, Talbot D, Fraser P, Grosveld F (1990) The beta-globin dominant control region: hypersensitive site 2. EMBO J 9: 2159-2167.

21. Hug BA, Wesselschmidt RL, Fiering S, Bender MA, Epner E, et al. (1996) Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. Mol Cell Biol 16: 2906-2912.

22. Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, et al. (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. EMBO J 15: 562-568.

23. Li Q, Stamatoyannopoulos G (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. Blood 84: 1399-1401.

24. Weatherall DJ (2001) Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2: 245-255.

25. Pauling L, Itano HA, et al. (1949) Sickle cell anemia a molecular disease. Science 110: 543-548.

26. Institute NHLaB (2012) What is Sickle Cell Anemia.

27. Ingram VM (1957) Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. Nature 180: 326-328.

28. Orkin SH, Little PF, Kazazian HH, Jr., Boehm CD (1982) Improved detection of the sickle mutation by DNA analysis: application to prenatal diagnosis. N Engl J Med 307: 32-36.

29. Allison AC (1954) Protection afforded by sickle-cell trait against subtertian malareal infection. Br Med J 1: 290-294.

30. Nathan DG, Gunn RB (1966) Thalassemia: the consequences of unbalanced hemoglobin synthesis. Am J Med 41: 815-830.

31. Weatherall DJ (1998) Pathophysiology of thalassaemia. Baillieres Clin Haematol 11: 127-146.

32. Fessas P (1963) Inclusions of hemoglobin erythroblasts and erythrocytes of thalassemia. Blood 21: 21-32.

33. Zurlo MG, De Stefano P, Borgna-Pignatti C, Di Palma A, Piga A, et al. (1989) Survival and causes of death in thalassaemia major. Lancet 2: 27-30.

34. Wickramasinghe SN, Bush V (1975) Observations on the ultrastructure of erythropoietic cells and reticulum cells in the bone marrow of patients with homozygous beta-thalassaemia. Br J Haematol 30: 395-399.

35. Cooley T, Witwer ER, Lee P (1927) ANEMIA IN CHILDREN WITH SPLENOMEGALY AND PECULIAR CHANGES IN THE BONES REPORT OF CASES. Am J Dis Child 3: 347-363.

36. Urbinati F, Madigan C, Malik P (2006) Pathophysiology and therapy for haemoglobinopathies. Part II: thalassaemias. Expert Rev Mol Med 8: 1-26.

37. Giardini C, Lucarelli G (1999) Bone marrow transplantation for beta-thalassemia. Hematol Oncol Clin North Am 13: 1059-1064, viii.

38. Lucarelli G, Clift RA, Galimberti M, Angelucci E, Giardini C, et al. (1999) Bone marrow transplantation in adult thalassemic patients. Blood 93: 1164-1167.

39. Brittenham GM, Cohen AR, McLaren CE, Martin MB, Griffith PM, et al. (1993) Hepatic iron stores and plasma ferritin concentration in patients with sickle cell anemia and thalassemia major. Am J Hematol 42: 81-85.

40. Gardenghi S, Ramos P, Follenzi A, Rao N, Rachmilewitz EA, et al. (2010) Hepcidin and Hfe in iron overload in beta-thalassemia. Ann N Y Acad Sci 1202: 221-225.

41. Rachmilewitz EA, Weizer-Stern O, Adamsky K, Amariglio N, Rechavi G, et al. (2005) Role of iron in inducing oxidative stress in thalassemia: Can it be prevented by inhibition of absorption and by antioxidants? Ann N Y Acad Sci 1054: 118-123.

42. Ramos P, Melchiori L, Gardenghi S, Van-Roijen N, Grady RW, et al. (2010) Iron metabolism and ineffective erythropoiesis in beta-thalassemia mouse models. Ann N Y Acad Sci 1202: 24-30.

43. Wolfe L, Olivieri N, Sallan D, Colan S, Rose V, et al. (1985) Prevention of cardiac disease by subcutaneous deferoxamine in patients with thalassemia major. N Engl J Med 312: 1600-1603.

44. Loukopoulos D, Voskaridou E, Stamoulakatou A, Papassotiriou Y, Kalotychou V, et al. (1998) Hydroxyurea therapy in thalassemia. Ann N Y Acad Sci 850: 120-128.

45. Boyer SH, Belding TK, Margolet L, Noyes AN (1975) Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. Science 188: 361-363.

46. Wood WG, Stamatoyannopoulos G, Lim G, Nute PE (1975) F-cells in the adult: normal values and levels in individuals with hereditary and acquired elevations of Hb F. Blood 46: 671-682.

47. Nagel RL, Fabry ME, Pagnier J, Zohoun I, Wajcman H, et al. (1985) Hematologically and genetically distinct forms of sickle cell anemia in Africa. The Senegal type and the Benin type. N Engl J Med 312: 880-884.

48. Donovan PB, Kaplan ME, Goldberg JD, Tatarsky I, Najean Y, et al. (1984) Treatment of polycythemia vera with hydroxyurea. Am J Hematol 17: 329-334.

49. Letvin NL, Linch DC, Beardsley GP, McIntyre KW, Nathan DG (1984) Augmentation of fetal-hemoglobin production in anemic monkeys by hydroxyurea. N Engl J Med 310: 869-873.

50. Platt OS (2008) Hydroxyurea for the treatment of sickle cell anemia. N Engl J Med 358: 1362-1369.

51. Charache S, Dover GJ, Moore RD, Eckert S, Ballas SK, et al. (1992) Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia. Blood 79: 2555-2565.

52. Rodgers GP, Dover GJ, Noguchi CT, Schechter AN, Nienhuis AW (1990) Hematologic responses of patients with sickle cell disease to treatment with hydroxyurea. N Engl J Med 322: 1037-1045.

53. Cokic VP, Smith RD, Beleslin-Cokic BB, Njoroge JM, Miller JL, et al. (2003) Hydroxyurea induces fetal hemoglobin by the nitric oxide-dependent activation of soluble guanylyl cyclase. J Clin Invest 111: 231-239.

54. Health NIo, Medicine UNLo (2013) Hydroxyurea.

55. Tuan D, Murnane MJ, deRiel JL, Forget BG (1980) Heterogeneity in the molecular basis of hereditary persistence of fetal haemoglobin. Nature 285: 335-337.

56. Baum C, Dullmann J, Li Z, Fehse B, Meyer J, et al. (2003) Side effects of retroviral gene transfer into hematopoietic stem cells. Blood 101: 2099-2114.

57. Sadelain M, Lisowski L, Samakoglu S, Rivella S, May C, et al. (2005) Progress toward the genetic treatment of the beta-thalassemias. Ann N Y Acad Sci 1054: 78-91.

58. May C, Rivella S, Callegari J, Heller G, Gaensler KM, et al. (2000) Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. Nature 406: 82-86.

59. Emery DW, Yannaki E, Tubb J, Nishino T, Li Q, et al. (2002) Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. Blood 100: 2012-2019.

60. Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). Science 183: 330-332.

61. Kornberg RD, Thomas JO (1974) Chromatin structure; oligomers of the histones. Science 184: 865-868.

62. Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A (1984) Structure of the nucleosome core particle at 7 A resolution. Nature 311: 532-537.

63. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251-260.

64. Thoma F, Koller T (1977) Influence of histone H1 on chromatin structure. Cell 12: 101-107.

65. Thoma F, Koller T, Klug A (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 83: 403-427.

66. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, et al. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci U S A 95: 14173-14178.

67. McGhee JD, Rau DC, Charney E, Felsenfeld G (1980) Orientation of the nucleosome within the higher order structure of chromatin. Cell 22: 87-96.

68. Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, et al. (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306: 1571-1573.

69. Marsden MP, Laemmli UK (1979) Metaphase chromosome structure: evidence for a radial loop model. Cell 17: 849-858.

70. Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434: 400-404.

71. Horvath JE, Bailey JA, Locke DP, Eichler EE (2001) Lessons from the human genome: transitions between euchromatin and heterochromatin. Hum Mol Genet 10: 2215-2223.

72. Lutter LC (1978) Kinetic analysis of deoxyribonuclease I cleavages in the nucleosome core: evidence for a DNA superhelix. J Mol Biol 124: 391-420.

73. Clark RJ, Felsenfeld G (1971) Structure of chromatin. Nat New Biol 229: 101-106.

74. Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, et al. (1980) Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell 20: 451-460.

75. Kosak ST, Groudine M (2004) Gene order and dynamic domains. Science 306: 644-647.

76. Yodh J (2013) ATP-Dependent Chromatin Remodeling. Adv Exp Med Biol 767: 263-295.

77. Roeder RG, Rutter WJ (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224: 234-237.

78. Roeder RG, Rutter WJ (1970) Specific nucleolar and nucleoplasmic RNA polymerases. Proc Natl Acad Sci U S A 65: 675-682.

79. Weinmann R, Roeder RG (1974) Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A 71: 1790-1794.

80. Weinmann R, Raskas HJ, Roeder RG (1974) Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc Natl Acad Sci U S A 71: 3426-3439.

81. Buratowski S, Hahn S, Guarente L, Sharp PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56: 549-561.

82. Van Dyke MW, Roeder RG, Sawadogo M (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science 241: 1335-1338.

83. Allison LA, Moyle M, Shales M, Ingles CJ (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell 42: 599-610.

84. Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A 82: 7934-7938.

85. Goodrich JA, Tjian R (1994) Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell 77: 145-156.

86. Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, et al. (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev 12: 343-356.

87. Yamaguchi Y, Inukai N, Narita T, Wada T, Handa H (2002) Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol 22: 2918-2927.

88. Shatkin AJ (1976) Capping of eucaryotic mRNAs. Cell 9: 645-653.

89. Wada T, Orphanides G, Hasegawa J, Kim DK, Shima D, et al. (2000) FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. Mol Cell 5: 1067-1072.

90. Murthy KG, Manley JL (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J Biol Chem 267: 14804-14811.

91. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G (2002) Locus control regions. Blood 100: 3077-3086.

92. Delvoye NL, Destroismaisons NM, Wall LA (1993) Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. Mol Cell Biol 13: 6969-6983.

93. Wall L, Destroismaisons N, Delvoye N, Guy LG (1996) CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells. J Biol Chem 271: 16477-16484.

94. Hartzog GA, Myers RM (1993) Discrimination among potential activators of the beta-globin CACCC element by correlation of binding and transcriptional properties. Mol Cell Biol 13: 44-56.

95. Miller IJ, Bieker JJ (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol Cell Biol 13: 2776-2786.

96. Orkin SH, Antonarakis SE, Kazazian HH, Jr. (1984) Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. J Biol Chem 259: 8679-8681.

97. Antoniou M, deBoer E, Habets G, Grosveld F (1988) The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. EMBO J 7: 377-384.

98. deBoer E, Antoniou M, Mignotte V, Wall L, Grosveld F (1988) The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. EMBO J 7: 4203-4212.

99. Lewis BA, Kim TK, Orkin SH (2000) A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. Proc Natl Acad Sci U S A 97: 7172-7177.

100. Leach KM, Vieira KF, Kang SH, Aslanian A, Teichmann M, et al. (2003) Characterization of the human beta-globin downstream promoter region. Nucleic Acids Res 31: 1292-1301.

101. Kang SH, Vieira K, Bungert J (2002) Combining chromatin immunoprecipitation and DNA footprinting: a novel method to analyze protein-DNA interactions in vivo. Nucleic Acids Res 30: e44.

102. Behringer RR, Hammer RE, Brinster RL, Palmiter RD, Townes TM (1987) Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. Proc Natl Acad Sci U S A 84: 7056-7060.

103. Grosveld F, Greaves D, Philipsen S, Talbot D, Pruzina S, et al. (1990) The dominant control region of the human beta-globin domain. Ann N Y Acad Sci 612: 152-159.

104. Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, et al. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene 205: 73-94.

105. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10: 1453-1465.

106. Zhu X, Ling J, Zhang L, Pi W, Wu M, et al. (2007) A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res 35: 5532-5544.

107. Busch SJ, Sassone-Corsi P (1990) Dimers, leucine zippers and DNA-binding domains. Trends Genet 6: 36-40.

108. Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240: 1759-1764.

109. McLachlan AD, Stewart M (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol 98: 293-304.

110. Hu JC, O'Shea EK, Kim PS, Sauer RT (1990) Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science 250: 1400-1403.

111. Hu JC, Newell NE, Tidor B, Sauer RT (1993) Probing the roles of residues at the e and g positions of the GCN4 leucine zipper by combinatorial mutagenesis. Protein Sci 2: 1072-1084.

112. Spek EJ, Bui AH, Lu M, Kallenbach NR (1998) Surface salt bridges stabilize the GCN4 leucine zipper. Protein Sci 7: 2431-2437.

113. Hurst HC (1994) Transcription factors. 1: bZIP proteins. Protein Profile 1: 123-168.

114. Henthorn P, Kiledjian M, Kadesch T (1990) Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science 247: 467-470.

115. Murre C, McCaw PS, Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56: 777-783.

116. Ellenberger T, Fass D, Arnaud M, Harrison SC (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev 8: 970-980.

117. Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature 363: 38-45.

118. Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252: 809-817.

119. Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 4: 1609-1614.

120. Frankel AD, Berg JM, Pabo CO (1987) Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc Natl Acad Sci U S A 84: 4841-4845.

121. Kadonaga JT, Carner KR, Masiarz FR, Tjian R (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51: 1079-1090.

122. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304-1351.

123. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245: 635-637.

124. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure 4: 1171-1180.

125. Wolfe SA, Nekludova L, Pabo CO (2000) DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct 29: 183-212.

126. Suzuki M, Gerstein M, Yagi N (1994) Stereochemical basis of DNA recognition by Zn fingers. Nucleic Acids Res 22: 3397-3405.

127. Feng WC, Southwood CM, Bieker JJ (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. J Biol Chem 269: 1493-1500.

128. Orkin SH, Kazazian HH, Jr., Antonarakis SE, Goff SC, Boehm CD, et al. (1982) Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster. Nature 296: 627-631.

129. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature 375: 316-318.

130. Perkins AC, Sharpe AH, Orkin SH (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature 375: 318-322.

131. Wijgerde M, Gribnau J, Trimborn T, Nuez B, Philipsen S, et al. (1996) The role of EKLF in human beta-globin gene competition. Genes Dev 10: 2894-2902.

132. Perkins AC, Gaensler KM, Orkin SH (1996) Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF. Proc Natl Acad Sci U S A 93: 12267-12271.

133. Tewari R, Gillemans N, Wijgerde M, Nuez B, von Lindern M, et al. (1998) Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. EMBO J 17: 2334-2341.

134. Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM (2010) KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat Genet 42: 742-744.

135. Lee JS, Ngo H, Kim D, Chung JH (2000) Erythroid Kruppel-like factor is recruited to the CACCC box in the beta-globin promoter but not to the CACCC box in the gamma-globin promoter: the role of the neighboring promoter elements. Proc Natl Acad Sci U S A 97: 2468-2473.

136. Zhou D, Pawlik KM, Ren J, Sun CW, Townes TM (2006) Differential binding of erythroid Krupple-like factor to embryonic/fetal globin gene promoters during development. J Biol Chem 281: 16052-16057.

137. Zhang W, Bieker JJ (1998) Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. Proc Natl Acad Sci U S A 95: 9855-9860.

138. Armstrong JA, Bieker JJ, Emerson BM (1998) A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. Cell 95: 93-104.

139. Kadam S, McAlpine GS, Phelan ML, Kingston RE, Jones KA, et al. (2000) Functional selectivity of recombinant mammalian SWI/SNF subunits. Genes Dev 14: 2441-2451.

140. Sengupta T, Cohet N, Morle F, Bieker JJ (2009) Distinct modes of gene regulation by a cell-specific transcriptional activator. Proc Natl Acad Sci U S A 106: 4213-4218.

141. Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, et al. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 18: 2485-2490.

142. Carthew RW, Chodosh LA, Sharp PA (1985) An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell 43: 439-448.

143. Moncollin V, Miyamoto NG, Zheng XM, Egly JM (1986) Purification of a factor specific for the upstream element of the adenovirus-2 major late promoter. EMBO J 5: 2577-2584.

144. Sawadogo M, Roeder RG (1985) Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell 43: 165-175.

145. Gregor PD, Sawadogo M, Roeder RG (1990) The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev 4: 1730-1740.

146. Sirito M, Lin Q, Maity T, Sawadogo M (1994) Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. Nucleic Acids Res 22: 427-433.

147. Sirito M, Walker S, Lin Q, Kozlowski MT, Klein WH, et al. (1992) Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. Gene Expr 2: 231-240.

148. Qyang Y, Luo X, Lu T, Ismail PM, Krylov D, et al. (1999) Cell-type-dependent activity of the ubiquitous transcription factor USF in cellular proliferation and transcriptional activation. Mol Cell Biol 19: 1508-1517.

149. Sirito M, Lin Q, Deng JM, Behringer RR, Sawadogo M (1998) Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. Proc Natl Acad Sci U S A 95: 3758-3763.

150. Corre S, Galibert MD (2005) Upstream stimulating factors: highly versatile stress-responsive transcription factors. Pigment Cell Res 18: 337-348.

151. Rada-Iglesias A, Ameur A, Kapranov P, Enroth S, Komorowski J, et al. (2008) Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. Genome Res 18: 380-392.

152. Crusselle-Davis VJ, Vieira KF, Zhou Z, Anantharaman A, Bungert J (2006) Antagonistic regulation of beta-globin gene expression by helix-loop-helix proteins USF and TFII-I. Mol Cell Biol 26: 6832-6843.

153. Bresnick EH, Felsenfeld G (1993) Evidence that the transcription factor USF is a component of the human beta-globin locus control region heteromeric protein complex. J Biol Chem 268: 18824-18834.

154. Roy AL, Meisterernst M, Pognonec P, Roeder RG (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature 354: 245-248.

155. Huang S, Li X, Yusufzai TM, Qiu Y, Felsenfeld G (2007) USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. Mol Cell Biol 27: 7991-8002.

156. Crusselle-Davis VJ, Zhou Z, Anantharaman A, Moghimi B, Dodev T, et al. (2007) Recruitment of coregulator complexes to the beta-globin gene locus by TFII-I and upstream stimulatory factor. FEBS J 274: 6065-6073.

157. West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G (2004) Recruitment of histone modifications by USF proteins at a vertebrate barrier element. Mol Cell 16: 453-463.

158. Carroll AJ, Crist WM, Link MP, Amylon MD, Pullen DJ, et al. (1990) The t(1;14)(p34;q11) is nonrandom and restricted to T-cell acute lymphoblastic leukemia: a Pediatric Oncology Group study. Blood 76: 1220-1224.

159. Aplan PD, Lombardi DP, Ginsberg AM, Cossman J, Bertness VL, et al. (1990) Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science 250: 1426-1429.

160. Hsu HL, Cheng JT, Chen Q, Baer R (1991) Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. Mol Cell Biol 11: 3037-3042.

161. Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, et al. (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J 16: 3145-3157.

162. Hsu HL, Wadman I, Tsan JT, Baer R (1994) Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. Proc Natl Acad Sci U S A 91: 5947-5951.

163. Shivdasani RA, Mayer EL, Orkin SH (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature 373: 432-434.

164. Visvader JE, Fujiwara Y, Orkin SH (1998) Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. Genes Dev 12: 473-479.

165. Mikkola HK, Klintman J, Yang H, Hock H, Schlaeger TM, et al. (2003) Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature 421: 547-551.

166. Song SH, Hou C, Dean A (2007) A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28: 810-822.

167. Elnitski L, Miller W, Hardison R (1997) Conserved E boxes function as part of the enhancer in hypersensitive site 2 of the beta-globin locus control region. Role of basic helix-loop-helix proteins. J Biol Chem 272: 369-378.

168. McLean C, Bejerano G (2008) Dispensability of mammalian DNA. Genome Res 18: 1743-1751.

169. Choo Y, Sanchez-Garcia I, Klug A (1994) In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature 372: 642-645.

170. Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263: 671-673.

171. Greisman HA, Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science 275: 657-661.

172. Krizek BA, Amann B, Kilfoil V, Merkle D, Berg J (1991) A Consensus Zinc Finger Peptide: Design, High-Affinity Metal Binding, a pH-Dependent Structure, and a His to Cys Sequence Variant. J Am Chem SOC: 4518-4523.

173. Shi Y, Berg JM (1995) A direct comparison of the properties of natural and designed zinc-finger proteins. Chem Biol 2: 83-89.

174. Michael SF, Kilfoil VJ, Schmidt MH, Amann BT, Berg JM (1992) Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. Proc Natl Acad Sci U S A 89: 4796-4800.

175. Liu Q, Segal DJ, Ghiara JB, Barbas CF, 3rd (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc Natl Acad Sci U S A 94: 5525-5530.

176. Mandell JG, Barbas CF, 3rd (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res 34: W516-523.

177. Ghosh D (1993) Status of the transcription factors database (TFD). Nucleic Acids Res 21: 3117-3118.

178. Segal DJ, Barbas CF, 3rd (2001) Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. Curr Opin Biotechnol 12: 632-637.

179. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF, 3rd (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 276: 29466-29478.

180. Segal DJ, Dreier B, Beerli RR, Barbas CF, 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. Proc Natl Acad Sci U S A 96: 2758-2763.

181. Beerli RR, Barbas CF, 3rd (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20: 135-141.

182. Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, et al. (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry 42: 2137-2148.

183. Liang S, Moghimi B, Yang TP, Strouboulis J, Bungert J (2008) Locus control region mediated regulation of adult beta-globin gene expression. J Cell Biochem 105: 9-16.

184. Cathomen T, Segal DJ, Brondani V, Muller-Lerch F (2008) Generation and functional analysis of zinc finger nucleases. Methods Mol Biol 434: 277-290.

185. Liang SY, Moghimi B, Crusselle-Davis VJ, Lin IJ, Rosenberg MH, et al. (2009) Defective erythropoiesis in transgenic mice expressing dominant-negative upstream stimulatory factor. Mol Cell Biol 29: 5900-5910.

186. Bungert J, Dave U, Lim KC, Lieuw KH, Shavit JA, et al. (1995) Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. Genes Dev 9: 3083-3096.

187. Roeder RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett 579: 909-915.

188. Orphanides G, Reinberg D (2002) A unified theory of gene expression. Cell 108: 439-451.

189. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11: 636-646.

190. Gupta A, Meng X, Zhu LJ, Lawson ND, Wolfe SA (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. Nucleic Acids Res 39: 381-392.

191. Barrow JJ, Masannat J, Bungert J (2012) Neutralizing the function of a beta-globin-associated cis-regulatory DNA element using an artificial zinc finger DNA-binding domain. Proc Natl Acad Sci U S A 109: 17948-17953.

192. Siatecka M, Bieker JJ (2011) The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood 118: 2044-2054.

193. Antoniou M (1991) Induction of Erythroid-Specific Expression in Murine Erythroleukemia (MEL) Cell Lines. Methods Mol Biol 7: 421-434.

194. Vieira KF, Levings PP, Hill MA, Crusselle VJ, Kang SH, et al. (2004) Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro. J Biol Chem 279: 50350-50357.

195. Zhang F, Cong L, Lodato S, Kosuri S, Church GM, et al. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29: 149-153.

196. Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, et al. (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7: 171-192.

197. Raagel H, Saalik P, Pooga M (2010) Peptide-mediated protein delivery-which pathways are penetrable? Biochim Biophys Acta 1798: 2240-2248.

198. Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, et al. (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 280: 35588-35597.

199. Manwani D, Galdass M, Bieker JJ (2007) Altered regulation of beta-like globin genes by a redesigned erythroid transcription factor. Exp Hematol 35: 39-47.

200. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148: 84-98.

201. Lam KN, van Bakel H, Cote AG, van der Ven A, Hughes TR (2011) Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. Nucleic Acids Res 39: 4680-4690.

202. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497-1502.

203. Johnson KD, Christensen HM, Zhao B, Bresnick EH (2001) Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. Mol Cell 8: 465-471.

204. Zhou Z, Li X, Deng C, Ney PA, Huang S, et al. (2010) USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the beta-globin gene locus. J Biol Chem 285: 15894-15905.

205. Paixao T, Azevedo RB (2010) Redundancy and the evolution of cis-regulatory element multiplicity. PLoS Comput Biol 6: e1000848.

206. de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26: 11-24.

207. Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10: 669-680.

208. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6: 507-512.

209. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, et al. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322: 1839-1842.

210. Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, et al. (2011) Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science 334: 993-996.

211. Sera T (2009) Zinc-finger-based artificial transcription factors and their applications. Adv Drug Deliv Rev 61: 513-526.

212. Mingozzi F, High KA (2011) Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 12: 341-355.

213. Du J, Jin J, Yan M, Lu Y (2012) Synthetic nanocarriers for intracellular protein delivery. Curr Drug Metab 13: 82-92.

214. Zelphati O, Wang Y, Kitada S, Reed JC, Felgner PL, et al. (2001) Intracellular delivery of proteins with a new lipid-mediated delivery system. J Biol Chem 276: 35103-35110.

215. Rejman J, Oberle V, Zuhorn IS, Hoekstra D (2004) Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis. Biochem J 377: 159-169.

216. Jorgensen KE, Moller JV (1979) Use of flexible polymers as probes of glomerular pore size. Am J Physiol 236: F103-111.

217. Desai MP, Labhasetwar V, Walter E, Levy RJ, Amidon GL (1997) The mechanism of uptake of biodegradable microparticles in Caco-2 cells is size dependent. Pharm Res 14: 1568-1573.

218. Connor EE, Mwamuka J, Gole A, Murphy CJ, Wyatt MD (2005) Gold nanoparticles are taken up by human cells but do not cause acute cytotoxicity. Small 1: 325-327.

219. Torchilin VP, Rammohan R, Weissig V, Levchenko TS (2001) TAT peptide on the surface of liposomes affords their efficient intracellular delivery even at low temperature and in the presence of metabolic inhibitors. Proc Natl Acad Sci U S A 98: 8786-8791.

220. Verma A, Uzun O, Hu Y, Han HS, Watson N, et al. (2008) Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles. Nat Mater 7: 588-595.

221. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, et al. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294-301.

222. Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5: 374-375.

223. Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333: 1843-1846.

224. Kay S, Hahn S, Marois E, Hause G, Bonas U (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 318: 648-651.

225. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326: 1509-1512.

226. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326: 1501.

227. Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, et al. (2011) Targeted genome editing across species using ZFNs and TALENs. Science 333: 307.

# BIOGRAPHICAL SKETCH

Joeva Barrow was born in Kingston, Jamaica. Joeva completed her Bachelor of Science in food science and human nutrition at the University of Florida in 2005 with a specialization in dietetics and a minor in French. Joeva then obtained her master's degree in food science and human nutrition at the University of Florida in a combined program in which she also fulfilled the dietetic internship requirements (MS-DI) and became certified as a registered dietitian in December 2008. Joeva has always had a profound interest in nutrition especially from the research perspective. In 2006, Joeva worked with Dr. Gail Kauwell in the Food Science and Human Nutrition department studying folate metabolism using the microbiological assay. Joeva then is worked with Dr. John Arthington and Dr. Mitchell Knutson evaluating a copper chaperone protein and its application as a copper status indicator in the bovine system. Her work resulted in a first author publication and she successfully completed her Master of Science degree. With a passion for research, Joeva then pursued her Ph.D. degree in Biochemistry and Molecular Biology in the College of Medicine at the University of Florida working with Dr. Jörg Bungert studying the effects of β-globin regulation using artificial DNA binding domains as a predictive tool to identify novel *cis*-regulatory elements in the β-globin gene locus. She graduated in May 2013 with a Ph.D. in medical sciences with a specialization in biochemistry and molecular biology. Given that the field of nutrition has always been Joeva’s passion, she plans to pursue her post-doctoral studies in nutritional biochemistry with Dr. Pere Puigserver at Harvard University at the Dana Farber Cancer Institute. There, she hopes to unveil the mechanisms that link genetics and epigenetics to nutrition and disease.