Regulation of Calreticulin Gene by bHLH and GATA-1 Proteins

By
Lise Marie Lamoureux

A Thesis
Submitted to the Faculty of Graduate Studies
In Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

Department of Biochemistry and Medical Genetics
Faculty of Medicine
University of Manitoba
and Division of Stroke and Vascular Disease
St. Boniface General Hospital Research Centre
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Regulation of Calreticulin Gene by bHLH and GATA-1 Proteins

BY

Lise Marie Lamoureux

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

TABLE OF CONTENTS

ABSTRACT

The regulation of blood vessel development is critical in the process of embryonic development, cancer progression and pathogenesis of cardiovascular disease. Transcription factors are key molecules involved in the regulation of genes including those that encode proteins implicated in angiogenesis. Two transcription factors, Tal-1 and GATA-1, were found to be expressed in endothelial cell precursors and required for the formation of functioning blood vessels. The E47 protein is ubiquitously expressed and found to be involved in smooth muscle cell differentiation. These three proteins have been previously shown to form a complex when their binding sites are within 8-10bp of each other and work synergistically to regulate the transcription of downstream genes.

Calreticulin (CRT) is a ubiquitous endoplasmic reticulum resident protein involved in mediating chaperone function and the regulation of calcium homeostasis. Recently, CRT has been shown to be highly expressed in the cardiovascular system of the mouse during development. Furthermore, application of exogenous CRT at the site of a tumour has been shown to inhibit the onset of tumour progression by halting the growth of blood vessels. This finding suggests a possible role for this protein in angiogenesis. Analysis of the CRT promoter revealed 4 putative binding sites for Tal-1/E47 which were in close proximity to GATA-1 binding sites. Thus, the aim of this project was to study the regulation of the CRT gene by a complex of Tal-1, E47 and GATA-1 transcription factors. The current study is the first to demonstrate a role for these transcription factors in the regulation of the CRT gene.

Our results demonstrated the presence of an enhancer region between -330 to -1 nucleotides of the promoter. We also showed that the addition of each of GATA-1, E47

and Tal-1 alone can enhance the activity of the CRT promoter. Furthermore, these transcription factors bind to the CRT promoter in complexes of more than 4 proteins at a time. We also demonstrated that GATA-1 and Tal-1 act synergistically to activate the CRT promoter. Using electrophoretic mobility shift assays (EMSA) and supershift assays, we concluded that a complex containing GATA-1, Tal-1 and/or E47 is binding to the CRT promoter and maximally regulating its activity. Point mutation in either the E box, the GATA-1 binding site or both sites was able to alter the complexes formed on these sites. These data collectively show a role for the bHLH and GATA-1 proteins in the regulation of the CRT promoter and suggest their importance in regulating CRT genes during blood vessel development.

ACKNOWLEDGMENTS

My first thank you goes out to Dr. Nasrin Mesaeli who has taught me so much including her love of science and enthusiasm for solving different research "puzzles". She instilled in me the drive to continue when things were rough saying it is not called "search" but "research". My time at the lab was invaluable and I owe all my success to Nasrin's supervision, assistance and friendship.

I would also like to acknowledge my committee members Dr. Jeff Wigle and Dr. Elissavet Kardami who have listened to me endlessly talk about my work and still seem interested in the research.

I could not have gotten through all these years without the support of my fellow lab members, both past and present. Mel, Shahrzad, Tina, Mike, Clark, Maliheh, Min, Anton and Jay have all supported my antics and craziness in the lab and have become great friends. Jean-Sebastien, Azadeh, Lori and May-Ting have done so much grunt work for me that I cannot thank them enough. The members of the Wigle and Pierce labs have also been there to help and support me throughout my time at the research centre.

Most importantly, I would like to acknowledge the support of my family and friends who have been indispensable to me during this process; especially to my parents and grandparents who are always asking how my research is going, who took the time to listen to me rant on about my project and who finally ended up shaking their heads in confusion. They have definitely kept me grounded. I could not have done it without you too Rudy, your support is indescribable! ☺

LIST OF FIGURES

LIST OF TABLES

A. REVIEW OF LITERATURE

I. Blood vessel development

1. Introduction

Blood vessel development is not only crucial in embryonic development, but it is also a necessary process in the adult. The ovarian cycle, wound healing and tumour vascularization are all dependant on new blood vessel growth. By understanding the mechanism of blood vessel branching (angiogenesis), it is possible to treat diseases or conditions based on the regulation of different steps in the angiogenic pathway. To further understand blood vessel development, we will start with a description of the structure of blood vessels.

2. Structure of blood vessels

The first and most important organ to form during development is the circulatory system. The blood vessels are required to circulate the blood even before the heart starts pumping. The structure of a blood vessel consists of multiple cell layers of different origins. The vessel wall can be divided into 3 sections: tunica intima, tunica media and tunica adventitia. The inner most part of the vessel (tunica intima) is comprised of a layer of endothelial cells that faces the lumen. The endothelial cells are covered by a thick layer of smooth muscle cells in the media layer of the vessel (tunica media). This structure is then covered by pericytes and collagen and elastic fibers that interact with the extracellular matrix (tunica adventitia). The development of this structure is a stepwise process originating from the mesoderm.

3. Mesoderm development

Both the heart and blood vessels originate through initial differentiation of the lateral mesoderm. Essentially, during early development, cells pass through the primitive streak and depending on their location specialize into 3 different cell lineages [1, 2]. The first to pass through are the endoderm cells and the second to pass through consist of the mesoderm cells [3]. The mesoderm is specialized into distinct regions. The precursor cells for the circulatory system originate from specialized cells within the lateral plate mesoderm. The lateral plate mesoderm splits into the somatic mesoderm and splanchnic mesoderm which differ by position – dorsal and ventral respectively [4]. A group of cells found in the splanchnic mesoderm form a structure known as the hemangioblast.

4. Hemangioblast

Both hematopoietic cells (blood cell precursors) and angioblasts (endothelial cell precursors) arise from a common precursor cells known as hemangioblasts [5, 6]. This common precursor was first verified by using the Flk-1 knockout mouse where there were defects in both endothelial cell and blood cell development indicating a common precursor [7]. This is an interesting stage in development and requires many factors to specialize a common precursor to become mature endothelial or blood cell. In vitro, a single blast colony (equivalent to hemangioblast) can specialize into endothelial or hematopoietic cells [5]. Once the hemangioblasts are produced, they group together to form blood islands [8, 9]. The inner cells in these islands are destined to become blood cells, whereas the cells on the surface of these islands are destined to become endothelial cells [9].

5. *Mechanism of blood vessel development*

Once cells have been specialized to become angioblasts (or endothelial cell precursors), blood vessel development can commence. This is a tightly controlled process involving multiple different proteins and signaling molecules responsible for proper formation of the vascular network. Endothelial cell formation involves the expression of different tyrosine kinase receptors on the surface of the cell at different points in development [10-14]. The VEGF receptors and their ligands will be discussed in detail as follows as the binding of the ligand to the receptor induces kinase cascades that signal the angioblast to differentiate into endothelial cells [10, 11, 13, 14].

i. Vascular development

Vascular development can be divided into four different steps that are outlined here (see Figure 1 for schematic diagram). The first stage of vascular development is known as vasculogenesis (stage 1). This process is the step by which new blood vessels are formed from angioblasts and produces the early vascular plexus. Once the primary network is formed, the next step involves the remodelling of these vessels to form a more solid vessel (stage 2). This is followed by maturation and stabilization of vessels (stage 2). It is at this stage that smooth muscle cells and pericytes are recruited and integrated with endothelial cells and the extracellular matrix (ECM) to form a more functional vessel. The last critical step in blood vessel development involves the formation of vessels from pre-existing vascular network known as angiogenesis (stage 3). There are 2 types of angiogenesis: sprouting and non-sprouting [10]. Sprouting angiogenesis involves

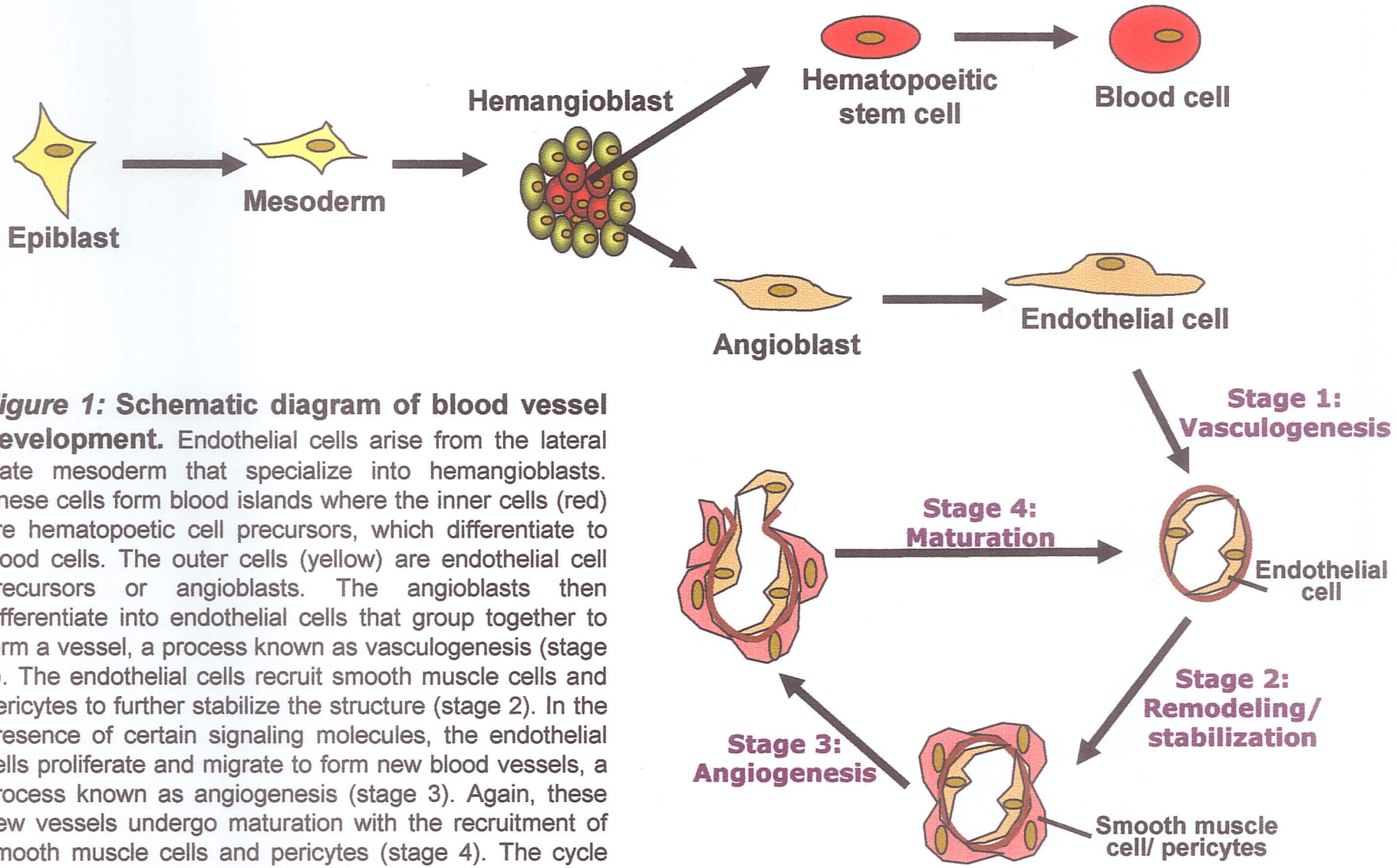


***Figure 1:* Schematic diagram of blood vessel development.** Endothelial cells arise from the lateral plate mesoderm that specialize into hemangioblasts. These cells form blood islands where the inner cells (red) are hematopoetic cell precursors, which differentiate to blood cells. The outer cells (yellow) are endothelial cell precursors or angioblasts. The angioblasts then differentiate into endothelial cells that group together to form a vessel, a process known as vasculogenesis (stage 1). The endothelial cells recruit smooth muscle cells and pericytes to further stabilize the structure (stage 2). In the presence of certain signaling molecules, the endothelial cells proliferate and migrate to form new blood vessels, a process known as angiogenesis (stage 3). Again, these new vessels undergo maturation with the recruitment of smooth muscle cells and pericytes (stage 4). The cycle continues based on a signaling cascade triggered by different signaling molecules and transcription factors (see text for more details).

the degradation of the ECM by matrix metalloproteinases (MMPs) that allow endothelial cells to migrate and proliferate into new blood vessels. Non-sprouting angiogenesis involves the splitting of preexisting blood vessels. A structure known as transcapillary pillars would form to split a blood vessel containing a wide lumen into separate vessels [10]. After angiogenesis, it is critical to stabilize the structure of these newly formed vessels to more mature network by recruiting smooth muscle cells and pericytes (stage 4). The cycle can continue during many processes in the adult [10, 11, 13].

ii. Tyrosine kinase receptors and ligands

There are many signaling proteins involved in the process of vessel development. Blood vessel development is controlled by a series of transmembrane tyrosine kinase receptors (RTK) that are stimulated through various ligands [12-14]. For the sake of simplicity, we will group the ligands into 2 groups the Vascular Endothelial Growth Factors (VEGF) and the angiopoietins (Ang). There are 5 different VEGF proteins known to date which are able to bind to two different receptors expressed early in development: the Flk-1 receptor and Flt-1 receptor [10, 11, 13]. It has been through the use of knockout mice that the importance of these different receptors was elucidated in vascular development. Flk-1 was found to be the first and most important receptor expressed during vascular development [7]. Flk-1 knockout mice die at E8.5 with the failure to develop blood islands and as a result are characterized by the absence of hematopoietic and endothelial cells [7]. The next receptor involved in vascular development is Flt-1 indicated by knockout mice that die also at E8.5 [15]. In this case, there is the presence of both endothelial and hematopoietic cells, but the endothelial cells do not assemble into

functioning tubes [15]. This indicates that Flk-1 is involved in the primary formation of endothelial cells from angioblasts, whereas Flt-1 is required for the process of remodeling and formation of mature blood vessel.

The second group of ligands, angiopoetins (Ang), is able to bind to another group of receptors: Tie 2 and Tie1 [10, 11, 13]. Tie1 and Tie2 are two tyrosine kinase receptors similar in structure, but different in function. Tie2 knockout mice die at E9.5-10.5 with hemorrhaging because of a lack of mature vessels [16]. There are no stromal cells present that normally stabilize the tube rendering it "leaky" which allows blood to leak into body cavities resulting in the hemorrhaging [16]. What is interesting about this receptor is that it can bind to different ligands: Ang1 and Ang2 [17]. Ang1 is a positive effector and helps produce a more stable mature blood vessel. Activation of Tie2 by Ang1 enhances the capacity of endothelial cells to recruit stromal cells that help stabilize the tube structure and function of blood vessels [17]. Ang2 on the other hand acts as a negative effector by inhibiting the effect of Tie2 resulting in unstable leaky vessels [17]. The binding of Ang2 to Tie2 stimulates the process of angiogenesis by destabilizing the blood vessel wall in the areas of branching vessels [17]. Transgenic mice overexpressing Ang2 are characterized by angiogenic defects due to an increase number of leaky vessels [17]. The leaky vessel leads to the detachment of smooth muscle cells and loosening of the matrix between vessel and ECM. The MMPs degrade the ECM which allows room for the endothelial cells to migrate and proliferate. The same affect is seen with nitric oxide (NO) which dilates blood vessels by increasing the level of VEGF which in turn cause leaky vessels. Homozygous knockout mice of the second Tie receptor, Tie1, die at E14.5 to postnatal [18]. Hemorrhaging and abdominal edema are characteristics of these

animals and indicates fluid exchange across capillaries [18]. Because the homozygous mice appear normal up until E13.0, it appears that Tie1 is involved in the proliferation and survival of endothelial cells. It is also involved in the formation of capillaries through angiogenic sprouting [18].

iii. Transcriptional control of blood vessel development

The elucidation of the different tyrosine kinase receptors and their respective ligands revealed a complex regulatory network ultimately leading to blood vessel development. To understand the control of this mechanism at the molecular level, scientists have studied the promoter regions of the different proteins involved in this signaling pathway. The promoter region is defined as the regulatory region of the gene containing binding sites for transcription factors that either activate or inhibit the expression of the corresponding gene. In 1995, the promoter regions of the four main endothelial tyrosine kinase receptors were elucidated and several putative transcription factor binding sites were found. We will focus on a few families of transcription factors important in blood vessels including: zinc-finger proteins, bHLH protein, ETS proteins, Hypoxia-Inducible-Factor proteins and Homeobox proteins.

a. bHLH and zinc-finger proteins

The bHLH (basic helix-loop-helix) protein Tal-1 is co-expressed with Flk1 in blood islands of embryos in early development [8, 9, 19]. Similarly, the GATA family of zinc-finger proteins and expressed in the blood islands of the developing mesodermal cells [8]. GATA-2 co-localizes with the early expression of Flk-1 in the primitive blood

islands [8]. GATA-2 induces the expression of GATA-1 and is co-expressed with Tal-1 in the blood islands [8]. The co-expression of GATA proteins and Tal-1 with Flk-1 indicates a role for these transcription factors in the regulation of Flk-1 expression. Two Tal-1 binding sites and one GATA binding site in the Flk-1 promoter were found to be essential for endothelial cell-restricted expression of Flk-1 [20]. *In vitro* interaction of these proteins with these three different binding sites were verified and seemed to be essential for in vivo expression of Flk-1 [20]. Further studies of potential gene targets for GATA and Tal-1 must be elucidated in order to understand their role in angiogenesis.

b. ETS Proteins

One of the most prominent and widely studied transcription factors in blood vessel development is the ETS (E-twenty-six) family [21]. This group of regulatory proteins is characterized by an ETS domain containing four β-sheets and three α-helixes [22]. One of these α-helixes is responsible for DNA-binding to the GGAA/T core consensus sequence [22]. The ETS proteins can be divided into different subgroups based on their overall protein structure and DNA-binding properties. There are 3 groups of ETS proteins that are expressed in endothelial cells during embryonic development. The first group includes ETS-1 and ETS-2 that both contain an activitory region and a "pointed domain" involved in protein-protein interactions [22]. The second group encompasses Erg and Fli-1 which contain an additional activitory region [22]. The final group expressed in endothelial cells is Elf-1 and NERF-2 which lack the pointed protein-protein interaction domain [22-24].

Promoter sequences of the four tyrosine kinase receptors, Flk-1 [20], Flt-1 [25], Tie1 [26] and Tie2 [27], all revealed putative binding sites for ETS proteins. Analysis of the Flt-1 promoter found 5 putative ETS binding sites [28]. However, the discovery of a minimal promoter of 90bp revealed the importance of only one ETS site essential for expression of this gene [28]. ETS-1, ETS-2 and Erg were all shown to increase the activity of the promoter 5-8 fold [28]. Similarly, the Flk-1 promoter contains six putative ETS sites, only one of which is involved in endothelial cell expression of Flk-1 during embryonic development [20]. ETS-1, ETS-2, NERF2 and Elf-1 increased the activity of an Flk-1 promoter construct [20, 23]. The Tie1 and Tie2 promoters contain conserved putative ETS binding sites that are conserved between human and mouse [24, 26, 29]. Seven different ETS factors were tested on both promoter and only NERF2 and ELF-1 upregulated the promoters [23, 24, 29].

ETS binding sites were also present on the promoters of MMPs and integrins. As mentioned, MMP proteins are necessary for ECM degradation to allow endothelial cells to migrate during angiogenesis. MMP-1, MMP-3 and MMP-9 are all expressed in endothelial cells and contain ETS binding sites within their promoter region [30]. Upregulation of MMP expression was seen in the presence of ETS-1 [30]. Similarly, ETS-1 enhanced the expression of $\alpha V\beta 3$ integrin expression [30]. Integrins are transmembrane proteins involved in cell adhesion by binding to different proteins present in the ECM [31, 32]. Thus the ETS proteins are essential in the correct expression of the tyrosine kinase receptors as well as MMP and integrin expression.

c. Hypoxia-Inducible Factors

Another important family of transcription factors involved in blood vessel development is the Hypoxia-Inducible-Factor (HIF) family. There are two different proteins in this family: HIF-1α and HIF-2α that share regions of high amino acid sequence conservation [33]. The HIF proteins contain a DNA-binding domain, dimerization domain, PAS domain and transactivation domains [33]. A PAS domain contains two regions of high hydrophobicity that is involved in protein-protein interactions [33]. There is also the presence of two transactivation domains one N-terminal and one C-terminal [33]. An important function of these proteins is their response to hypoxia. During normal conditions, the HIF-α proteins are ubiquitinated and degraded by the proteosome-mediated pathway [33]. During hypoxic conditions, the HIF-α proteins translocate to the nucleus and dimerize with the Aryl Hydrocarbon Nulcear Translocator (ARNT) protein and activate target genes through interaction with CBP/P300 [33]. CBP/P300 is a coactivator that aids in the transcriptional activation of genes.

The VEGF protein was found to be upregulated in response to hypoxia [34]. Interestingly, the VEGF promoter contains two putative Hypoxia-responsive elements (HRE, sequence TACGTG(G/C)G) [34, 35]. HIF-1α was able to bind and increase the activity of the VEGF promoter through one of these sites [35]. It was then suggested that HIF-α proteins may be involved in the regulation of the different tyrosine kinase receptors. The transcriptional regulation of Flk-1 and Flt-1 was tested in response to hypoxic conditions. Flt-1 was found to be significantly upregulated in hypoxia, whereas the expression of Flk-1 was not altered by low oxygen [36]. Sequence analysis of the

promoter region of Flt-1 revealed a putative HRE which was found to be critical in the upregulation of the gene in response to hypoxic conditions [36]. Interestingly, the HIF-2α isoform is expressed in endothelial cells during embryonic development and colocalizes with Flk-1 in blood vessels [37, 38]. There are 2 HRE-like motifs (GGCGTG) found in proximity to 2 ETS binding sites on the Flk-1 promoter [38]. These four sites are crucial for the correct expression of Flk-1 [38]. ETS-1 and HIF-2α activate the promoter individually, but they can act synergistically to activate the promoter when both proteins are present [37, 38]. The fact that the ETS proteins contain a "pointed" protein-protein interaction domain suggested that ETS-1 can form a complex with HIF-2α which was confirmed by pull-down assays [38]. Contrarily, the HIF-1α isoform did not have any significant effect on Flk-1 expression suggesting a different role for these proteins during development and hypoxic conditions [37]. The addition of HIF-1α to primary endothelial cells resulted in upregulation of angiopoietin expression [39]. This suggests that the hypoxic response that stimulates HIF-1α may exert some of its effects through the Tie2 receptor tyrosine kinase signaling pathway as well as VEGF and Flk-1 pathways [34, 35, 39].

d. Homeobox proteins

The transcription factors in the homeobox family all contain a 60 amino-acid helix-turn-helix homeodomain that is involved in DNA binding [40]. The homeobox genes are clustered together in four different groups labeled A to D on the genome [40]. A limited amount of other members of this family are divergent genes not found within this cluster [40]. Four main homeobox proteins are expressed in endothelial cells: the

cluster genes HOXD3 and HOXB3 and the divergent genes Oct-1 and HEX [40, 41]. The HoxD3 protein was found to bind to the promoter regions and upregulate αVβ3 and α5β1 integrin expression [41, 42]. These two integrin proteins are expressed in low levels in quiescent vessels, but are upregulated in angiogenic vessels and downregulated in late stages of angiogenesis [31, 32]. The overexpression of HoxD3 resulted in increased amounts of these proteins and created an abnormal vasculature containing endotheliomas [41]. Endotehliomas are tumours containing a large lining of endothelial cells filled with hematopoietic cells which suggests HoxD3 is involved in angiogenesis [41]. Similarly, HoxB3 was found to have an effect on angiogenesis by activating the ephrin 1 protein which is expressed on endothelial cells and involved in capillary formation [43]. The absence of HoxB3 produces capillaries that do not function properly [43]. Thus HoxD3 and HoxB3 are both involved in angiogenesis by affecting different processes (invasion and capillary morphogenesis respectively).

The third homeobox protein expressed in blood islands and mature endothelial cells is HEX which was found to negatively regulate angiogenesis by inhibiting the transcription of the different tyrosine kinase receptors [44]. The expression levels of Flk-1, Flt-1, Tie1, Tie2 and integrin αV were all downregulated in the presence of HEX [44]. In opposition, TIMPS (tissue-inhibitors of matrix metalloproteinases) were upregulated in the presence of HEX [44]. The expression of these proteins affected by HEX demonstrates a role for this protein in the inhibition of angiogenesis. Recently, Hex was found to interact with the GATA proteins and inhibit their DNA binding properties [45]. The activation of the Flk-1 promoter seen by GATA-2 was abolished by the addition of the HEX protein, further indicating a role for HEX in the repression of angiogenesis [45].

The final homeobox enhancer protein Oct-1 binds to the octamer site (ATGCAAAT) on the promoters of both Tie1 and Tie2 [46]. Mutational analysis confirmed the importance of the octamer site in the expression of Tie1 and Tie2 in endothelial cells of the embryo [46].

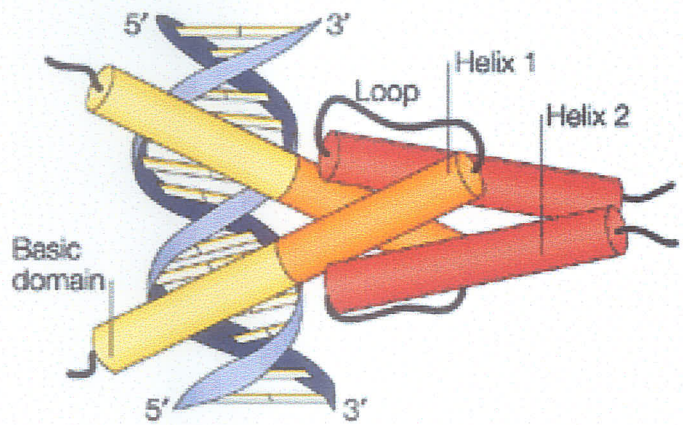
Many other factors are expressed in endothelial cells at different points in development and are crucial in the regulation of the different tyrosine kinase receptors and ligands as well as MMPs and integrins. Not much research has been focused on gene targets for the bHLH proteins and GATA proteins in endothelial cells. We will focus on the effect of Tal-1 and GATA on the regulation of blood vessel development based on their importance in the regulation of Flk-1 gene expression [20].

II. bHLH proteins

1. Structure

The helix-loop-helix family (HLH) is a specific group of DNA binding proteins with a similar motif. As their name suggests, the structure consists of a sequence of 2 short hydrophobic helical segments separated by a loop (helix-loop-helix structure) (see Figure 2A) [47]. These hydrophobic residues are conserved through all of the HLH proteins [48]. On the other hand, the loop region varies between proteins indicating that this region is not crucial for function, but necessary for physically separating the helixes [49]. Comparison analysis of the sequence of different HLH proteins revealed similarity to lamins [48]. Lamins utilize a four-three repeat of hydrophobic residues (amino acid sequence a,b,c,d,e,f,g where a and e are hydrophobic and would be placed on the same side of the helix) [48]. Similarly, HLH proteins interact through the same amphipathic helices to form dimmers [48, 50]. Dimerization cannot occur without intact helix domains in both proteins [48-50]. Disulphide bonds help stabilize the proteins during dimerization [51]. Dimerization is essential for DNA binding – these proteins bind DNA without only as a dimer with another HLH protein [48]. The dimer structure resembles a four helix bundle that recognizes a consensus sequence first identified in the immunoglobulin heavy chain (IgH) gene enhancer region (see Figure 2A). This sequence is known as an E box whose consensus sequence is CANNTG (where N is any nucleotide) [52].

A.



B.

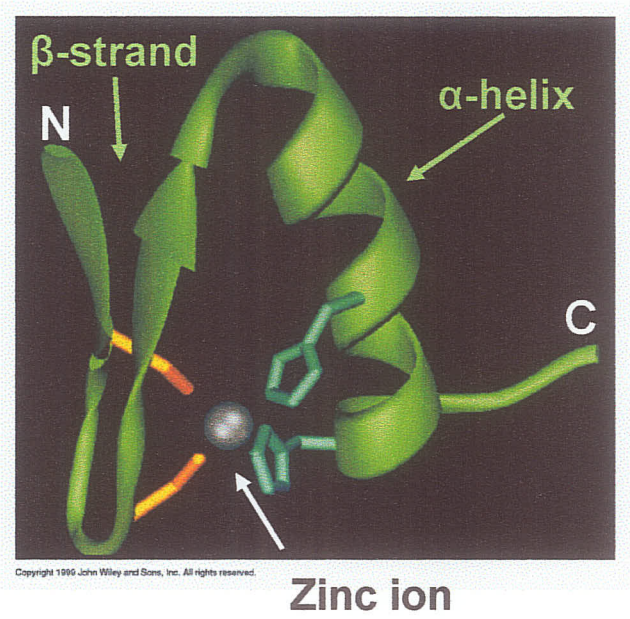


Figure 2: **Structure of bHLH and GATA transcription factors.** ***A***. Four-helix bundle structure of a dimer of bHLH proteins. The basic region (yellow) of each protein binds DNA. The Helix-loop-helix structure (orange and red) of each protein is involved in dimerization. **B**. Zinc-finger structure of GATA binding protein GATA-1. Contains 4 cysteine residues to chelate zinc: two from the α-helix (green) and two from the β-strand (orange)

2. *Classification*

There have been more than 240 proteins in this superfamily identified to date [53, 54]. Many different methods to classify these proteins have been established. The classification is based on three different criteria: tissue distribution, dimerization abilities and DNA specificity. The classification system most widely used involves the differing flanking sequences on either side of the helix-loop-helix motif [47, 55, 56]. This classification system involves 2 main groups: bHLH proteins (class A and class B) and HLH proteins [47, 50, 56]. The first group is called the basic-helix-loop-helix (bHLH) proteins and contains an additional basic region preceding the HLH sequence which is necessary for correct DNA binding. Different sequences of this basic region recognize different E box sequences and render these factors ubiquitous (class A bHLH proteins) or tissue-specific (class B bHLH proteins). Interestingly, the majority of class A and class B bHLH proteins exhibit transactivation of the gene of interest. The bHLH proteins that are repressors found in other groups. E12 and E47 are ubiquitous class A bHLH proteins that can homodimerize, but preferentially heterodimerize with tissue-specific class B bHLH proteins – a type of concentration-dependent regulation [50, 57, 58]. These heterodimers can form posttranslationally and can be assembled in the absence of DNA [48]. If one protein contains a mutation in the basic region, the proteins would dimerize, but could not interact with DNA [49]. There must be two intact bHLH regions for the two proteins to bind DNA. Surprisingly, evolution has generated a group of HLH proteins that lack the basic region rendering them unable to bind DNA – known as the Id proteins that are the second group [47, 59, 60]. They can act as negative regulators of gene expression by forming heterodimers with a bHLH protein [47, 59, 60]. The dominant negative effect of

these proteins inhibits normal DNA binding resulting in erroneous expression of the gene of interest.

Taken together, the bHLH proteins have been shown to contribute to the regulation of the commitment and differentiation of various cell lines – most notably MyoD in myogenesis, Tal-1 in hematopoiesis and NeuroD in neurogenesis [55, 61, 62].

3. E PROTEINS

i. E2A proteins

a. Isolation and characterization of E2A gene

E47/E12 was first isolated in 1989 when they were first categorized as class I bHLH proteins or E proteins. Five different sequences on the immunoglobulin heavy chain gene enhancer region containing the sequence CANNTG were found to be crucial for gene expression [47]. Using the fifth site on this enhancer region (kE2) as a probe, Murre et al. screened for specific binding protein interactors present in a human B cell lymphoma cell line (BJAB) [47, 50]. The E12 protein was discovered because it bound specifically to this probe. The cDNA for E12 was used to screen other proteins and E47 was found to hybridize to this clone. These 2 genes were in vitro transcribed and were found to bind specifically to the kE2 probe. Interestingly, the affinity of E47 binding to the probe was greater than E12 [48]. Sequence analysis of the two genes revealed that the proteins arose from differential splicing of the same gene [47-49, 63]. The promoter region of the E2A gene did not have a TATA region [64]. Most housekeeping genes and genes that are ubiquitously expressed contain a TATA-less promoter [64]. This study suggested a widespread expression pattern that has been confirmed in *in-situ*

hybridization studies. E47 and E12 are ubiquitously expressed throughout the embryo, except for the absence of E47 expression in the heart [48, 65, 66]. Interestingly, expression of these proteins was found mostly in regions of rapid cell division indicating a role for them in cell proliferation and differentiation [65, 66].

b. Protein characteristics

The E2A proteins contain two transcriptional activatory domains [67]. The first site (AD1) is near the N-terminal domain and the second site (AD2) is located in proximity to the bHLH region [67, 68]. Recently a region of 30 highly acidic amino acids known as a Rep domain was discovered to be a repression domain present on E2A proteins [68]. This region located between the AD regions and the bHLH region is well conserved between species of the E proteins only, indicating a regulatory mechanism specific for E proteins [68]. Studies conducted on the muscle creatine kinase enhancer revealed that when a homodimer of E proteins is bound to the promoter region, the Rep domain inhibits the effect of AD2 [68]. Structurally the AD1 cannot activate the promoter and the target gene remains silent. Heterodimerization between MyoD and an E protein allow the Rep domain to inhibit AD1 domain, but allows the AD2 domain to activate transcription of muscle specific genes [68]. This is an interesting regulatory mechanism of E proteins to limit the effect of the ubiquitous expression of the E proteins in transcription of genes in a given cell-type. It would be of interest to look at the role of the Rep domain in other cell types.

c. Other E proteins

Other E proteins were identified soon after the cloning of the E2A gene products. Two proteins, ITF-1 and ITF-2 (immunoglobulin transcription factor-1 and -2), were identified on the immunoglobulin heavy chain enhancer region which are crucial in the activation of gene transcription [69]. Both of these proteins presented transcription factor characteristics because they were found to activate transcription of the immunoglobulin heavy chain gene [69]. ITF-2 has two splice variants (ITF-2a and ITF-2b) that differ in their N terminal regions [70]. The nucleotide sequence of these proteins is 61-80% identical to other E proteins with the highest homology being between the bHLH regions [70].

ii. Function of E2A gene products

a. B cell development

The majority of the research conducted on E2A proteins encompasses their role in B cell development. These proteins have recently been found to be involved in smooth muscle cells differentiation and will be discussed later. The first *in vivo* functional studies on the E2A gene products were conducted by generating knockout mice [63, 71]. The homozygous knockout mice lacking both E12 and E47 appear normal, but most die within the first few weeks of birth [63]. These mice are growth retarded due to an extreme weight loss and dehydration [63]. Further analysis revealed that the knockout mice contain no B cells due to impaired early B cell differentiation [63]. The heterozygous mice contain half as many B cells as the wild-type littermates [63]. In E2A homozygous knockout mice, both the T and B-cell differentiation are blocked at a similar

step that prevents progenitor cells from committing to the definitive cell lines [72]. Unlike B cells however, some mature T cells are present in E2A homozygous knockout mice albeit at lower concentrations, probably due to the other E proteins present in T cells [72]. Because of the ubiquitous expression of E12 and E47, it is of interest that other members of the E proteins family compensate for the E2A protein deficiency in all cell types except B cells [63, 71]. This indicates an important and unique role for E2A proteins in the development of B cells [63]. Further molecular analysis of the E2A protein deficiency proved that in knockout mice, the expression of RAG-1 (a transposase) and Pax-5 (B-cell specific transcription factor) were affected [71]. Pax-5 regulates genes involved in B cell development such as CD19 and lambda 5 [73]. It is interesting that Pax-5 and E2A proteins are found in overlapping patterns of expression during embryonic development, suggesting that Pax-5 may be a downstream target of E2A proteins [71]. Similarly, Rag-1 is an essential transposase involved in excision and integration of different immunoglobulin genes required for the rearrangement process [74, 75]. This implies that the E2A proteins may either directly or indirectly affect these genes and in turn affect the development of B cells.

b. Smooth muscle cell differentiation

Smooth muscle alpha actin is an ideal marker to study smooth muscle differentiation because it is temporally expressed in cardiac and skeletal muscle during development, but exclusively expressed in smooth muscle in the adult [76]. Jung et al. (1999) focused on transcription factors that could be involved in the regulation of protein expression during proliferation and differentiation of smooth muscle cells. After conditions such as vascular

injury, there is proliferation and differentiation of smooth muscle cells to compensate for the injury. The promoter of the smooth muscle alpha actin gene contains a CArG site and 2 E boxes important for the regulation of expression of these proteins [76]. The CArG sequence entails a CC(A-T rich)$_6$GG site involved in SRF binding. Mutations of the CArG site or in the two E boxes decreased overall activity of the gene emphasizing the importance for these sites during smooth muscle cell differentiation [76]. To further characterize the importance of the two E boxes, transgenic mice with one E box mutation or a double mutated E box in the smooth muscle alpha actin gene were generated [77]. The double mutant homozygous mice had no expression of smooth muscle alpha actin in any of the 3 muscle cells types [77]. Interestingly, single mutant embryos had a significant reduction in the level of smooth muscle alpha actin, yet there was still expression present indicating a redundant role for these 2 E boxes in development [77]. Furthermore, all four E proteins E12, E47, ITF-1 and ITF-2 enhanced the activity of the smooth muscle alpha-actin promoter, albeit at different levels [77]. Mutation of both or one the E box binding sites abolished the effect of these proteins indicating importance of both sites in the regulation by bHLH class I proteins [77]. The interaction between the E2A protein and the protein that binds the CArG site, SRF, was studied to look at possible complex formation. The combination of SRF and E2A proteins enhanced the activity of the promoter and was found to interact only in smooth muscle cells [77]. These studies and the fact that E47 is expressed in embryonic blood vessels, mainly cerebral ventricles, suggest a role for the E2A proteins to be master regulators of smooth muscle differentiation [65, 76, 77].

4. Tal-1 protein

i. Isolation and characterization of Tal-1 gene

Tal-1 was first discovered due to its involvement in chromosomal translocations producing T-cell acute lymphoblastic leukemia (T-ALL) [78, 79]. Further analysis of Tal-1 amino acid sequence indicated 84% identity between another protein involved in chromosomal rearrangements causing leukemia – Lyl1 [79]. These 2 proteins were grouped together as hematopoietic-specific family of class B bHLH transcription factors. Class B proteins are characterized by their ability to heterodimerize with class A proteins. Further analysis revealed that Tal-1 heterodimerizes with the class A E proteins E47 and E12 confirming the fact that Tal-1 is a class B bHLH protein [57, 80-82]. Tal-1 can homodimerize, but the spatial positioning of the basic region is incompatible with DNA binding rendering homodimers unable to regulate target genes [58]. Thus, Tal-1/E protein heterodimers are necessary for the transcriptional regulation of genes of interest.

ii. Tal-1 protein expression

Expression studies indicated the presence of Tal-1 in hematopoietic cells including erythroid, megakaryocytic and mastocytic lineages [83-85]. In addition, Tal-1 has been found to be present in endothelial cells suggesting a role for this protein in the development of the endothelium[83-87].

iii. Function of Tal-1 protein

a. Hematopoietic development

As mentioned, Tal-1 was first discovered by its involvement in chromosomal translocations producing T-ALL leukemia [78, 79]. This indicated an important function for Tal-1 in hematopoiesis (blood cell development). The phenotype of mice containing targeted deletion of the Tal-1 gene was characterized by the absence of blood islands thus the absence of red blood cells [88-90]. These Tal-1 knockout mice die *in utero* at E10.5 probably from anemia with a total lack of erythropoiesis [88, 89]. Interestingly, these mice did have blood vessels and a myocardium indicating that Tal-1 may not be involved in blood vessel development. It wasn't until 1998 that this theory was opposed. Gering et al. 1998 studied Tal-1 overexpression in zebrafish. Using 4-cell stage zebrafish embryos, Tal-1 mRNA was microinjected in these cells to study the function of this protein during early lateral mesoderm development. These embryos had a normal heart, but had an increase in blood cells in the intracellular matrix. There was a lack of normal blood circulation probably due to the overproduction of hematopoietic progenitor cells [91]. It was thought that because both blood and endothelial cells share a common precursor (hemangioblast), the development of early endothelial cells might be affected [91]. Endothelial cell formation involves the expression of different tyrosine kinase receptors on the surface of the cell at different points in development. The receptors Flk1 and Flt1 are expressed in early development and require VEGF as a ligand to initiate the kinase cascade [10, 11]. Both of these tyrosine kinase receptors were overexpressed in the Tal-1 transgenic animals indicating overproduction of endothelial cells. The increase in both

endothelial and blood cells indicated that a common precursor (or hemangioblast) is affected by Tal-1 function [91].

b. Blood and endothelial precursor

Characterization of the Flk1 expression pattern indicated an important function in hemangioblast development. Flk1 knockout mice fail to develop blood islands and as a result lack hematopoietic and endothelial cells [7]. A recent study found that Flk1 and Tal-1 are coexpressed in the blood islands of embryos in early development [9, 19]. It was thought that the expression of these 2 proteins may delineate the fate of hemangioblasts to become either blood or endothelial cells [9, 19]. Development originates from mesodermal precursor cells that express Flk1 [9]. These precursor cells only develop into hemangioblasts by the coexpression of Flk1 and Tal-1. At this point, the expression pattern of Flk1 and Tal-1 may possibly affect the fate of the hemangioblasts to become hematopoietic, endothelial and even smooth muscle cells [9]. Hematopoietic stem cells arise from hemangioblasts, but are characterized by the expression of Tal-1 and absence of Flk1. Endothelial cells on the other hand are characterized by the expression of Flk1 and vary in the expression of Tal-1. It is thought that endothelial cells arise from 2 precursor pools: hemangioblasts (express both proteins) and angioblasts (express only Flk1) [9, 19]. It is of interest that Tal-1 may not be expressed in early endothelial cell development, but is involved in characterizing the cells later in development.

c. Vascular development

The functional role of Tal-1 in later development was studied in the vasculature by Visvader et al. 1998. Ectopic expression of Tal-1 cDNA under the regulation of the GATA-1 promoter resulted in its expression only in hematopoietic cells and was able to rescue Tal-1 phenotype [92]. The rescued transgenic mice die *in utero* at E9.5, but had nominal amount of hematopoietic cells and primary blood vessels. However, these blood vessels did not remodel and do not undergo angiogenesis [92]. This is consistent with the fact that Tal-1 is expressed in angioblasts as well as mature endothelial cells [83, 85, 87]. Thus, Tal-1 has a critical role in angiogenesis (reviewed in [93]).

d. Complex formation

The expression pattern of Tal-1 and the importance of correct timing of transcription is a good indication that Tal-1 itself is tightly controlled by other transcription factors expressed at specific times in development. Several reports have demonstrated formation of complex between Tal-1 and other factors expressed in the blood cell lineage such as GATA-1 and Lmo2 [94, 95].

III. GATA proteins

1. Introduction

Transcription factors have evolved to include cofactors in their structure for stabilization and diversity. The proteins that chelate a zinc ion are known as zinc finger proteins[96, 97]. These proteins are one of the most abundant classes of proteins present in the organism, encompassing about 2% of the mammalian genomes [98-100].

2. Classes of zinc-finger proteins

The zinc finger proteins can be divided into different classes based on the organization and spacing between the different zinc chelating structures. The first group includes proteins that contain 2 cysteine residues and 2 histidine residues involved in zinc chelation. The first zinc-chelating protein identified was the TFIIIA protein [96]. The protein structure of TFIIIA and others from this group are characterized by 2 beta sheets followed by an alpha helix [96, 100]. The alpha helix contains basic and hydrophobic residues involved in DNA binding. Usually the proteins in this first group contain C2H2 zinc fingers in series of 3 or more separated by linker sequences that recognize DNA motifs (such as the sequence TGEKP) [96, 100]. The second group is characterized by 4 cysteine residues chelating zinc. The proteins in this group are folded differently into a type of treble clef structure that contains 2 beta hairpins and an alpha helix [100, 101]. GATA proteins and nuclear hormone receptors are members of this family of transcription factors. Whereas the first 2 groups of zinc-finger proteins are involved in DNA binding, the proteins in the last group are more involved in protein-protein interactions, ubiquitination and lipid binding. This third group contains a similar treble

clef structure as the 2nd group but differ by the binding of 2 zinc atoms in alternating Cys/His residues [100]. The BRCA1 protein is a member of this group and contains RING fingers that are involved in ubiquitination and are mutated in breast cancer. Another member, CBP/P300, contains LIM and PHD domains involved in protein-protein interactions favoring the formation of a multi-protein complex [100]. The class of zinc-finger protein most involved in the development of the circulatory system has been the GATA zinc-finger proteins.

3. GATA family

The GATA family of proteins (GATA-1, -2, -3, -4, -5 and -6) as their name suggests bind to a GATA sequence in DNA. All the GATA proteins were found to bind to the WGATAR (W=A or T, R= A or G) site with a little variation [102]. GATA-1 was the first protein discovered in this group of DNA binding proteins. Interestingly, the GATA-1 gene is located on the X chromosome, so most developmental studies were easily conducted in male mice [103-105]. The GATA proteins have been well conserved through all eukaryotes [106, 107]. There is 89% similarity between human and mouse amino-acid sequences, most notably through the cysteine finger region [103-105, 108]. Soon after GATA-1 was discovered, different members of the GATA family were elucidated. There is high sequence conservation of the zinc finger domains between GATA-1,-2 and -3 proteins. GATA-2 and -3 are more similar in sequence than GATA-1 [109]. A recent review describes in more detail the GATA family of transcription factors 1-6 [107]. It has also been shown that GATA proteins act in cohort with other proteins forming complexes that together regulate the activity of a gene [107].

i. Expression pattern

The expression patterns of the GATA family members differ, but have some overlapping expression. GATA-1 is found in erythroid, hematopoietic cells and progenitor cells [107, 110]. GATA-2 is similarly expressed in progenitor cells and hematopoietic cell lines as well as embryonic brain cells and endothelial cells [107, 110]. GATA-3 is expressed solely in T-lymphoid cells as well as embryonic brain cells and endothelial cells [107, 110]. GATA-4, -5 and -6 proteins are localized to different sections of the cardiovascular system. GATA-4 is restricted to embryonic and adult heart, ovary, testis and liver, among others [107, 111]. Contrarily, GATA-5 is expressed only in embryonic heart, but is expressed in adult intestine, stomach, bladder and lungs [107, 111]. The GATA-6 protein is expressed in embryonic and adult heart as well as vascular smooth muscle cells and other organs [107, 111].

ii. GATA-1 protein structure

GATA-1 protein only binds DNA in the presence of zinc indicating the indispensability of zinc for the function of this protein [97]. The structure of a zinc-finger protein is characterized by an antiparallel beta-sheet and an alpha helix [101]. The coordination of the zinc ion is maintained through 2 residues from the β-sheet and 2 residues from the α-helix (see Figure 2B) [101]. The alpha helix of the protein inserts into the major groove of DNA and residues, both basic and hydrophobic, make contact with DNA [98, 112]. It is also thought that the linker sequences between fingers such as TGEKP may recognize DNA motifs [100]. The structure that chelates the zinc ion is known as the zinc finger motif. Usually zinc-finger proteins contain a series of 2 or more

zinc finger motifs in series each separated by linker sequences [100]. One advantage of a single protein containing different zinc-finger motifs is the ability of each motif to specialize in different functions [98]. The GATA-1 protein contains 2 zinc-finger motifs: an N-terminal finger (N-finger) and a C-terminal finger (C-finger). Mutational analysis of the two zinc fingers identified different functions for the two motifs [113]. The C-finger is critical for the binding of GATA-1 to the DNA consensus sequence (A/T)GATA(A/G) and is the most conserved domain between GATA proteins of different organisms [98, 100, 106, 113, 114]. On the other hand, the N-finger is important for protein-protein interactions and involved in the stability of the DNA binding complex [98, 114]. Yeast-2-hybrid analysis identified FOG-1 (Friend of GATA-1) as a binding partner for GATA-1 [114]. Mutation of the N-finger abolishes the interaction with FOG-1 indicating the importance of this motif in the protein-protein interaction of these two factors [98, 114]. Recently, both GATA-1 fingers have been found to interact with SP1 and Lmo2 [100]. Interestingly, the N-finger has recently been shown to bind to a TGATCT consensus sequence in vitro [100, 114]. The variability of the GATA-1 protein to bind to different sequences could indicate a regulatory mechanism dependant on which finger is involved in DNA binding by affecting the GATA-1 protein dimerization partner. Thus, using one finger over the other in DNA binding would eliminate certain protein-protein interactions. Additionally, the GATA-1 protein contains 3 activitory domains indicative of a positive effector [113].

iii. GATA-1 developmental studies

The GATA-1 protein was first discovered by its expression in human erythroleukemia cells (K562) and mouse erythroleukemia cells (MEL) [103]. Studies of GATA-1 deficient ES cells showed that these cells could contribute to the formation of all tissues in the chimeric mouse except the mature red blood cells (hematopoietic origin) [115]. In 1996, Fujiwara et al. used gene-targeted deletion of the GATA-1 initiator codon and showed that the F1 heterozygote females were pale neonatally, but appeared normal as adults [116]. Male GATA-1 knockout embryos die at E10.5-E11.5 with severe anemia. Knockout embryos E9.5-10.5 were pale, had normal vasculature, but no red blood cells. Further analysis revealed that red blood cell precursors developed, but could not mature into adult cells [116]. The globin genes were also activated in knockout cells, yet not to the same degree as wt cells. This would indicate that GATA-1 is not the sole factor involved in erythropoiesis, but may be compensated by GATA-2 or other factors expressed in early erythroid development [116]. An ES cell line derived from GATA-1 knockout erythroid cells was transfected with GATA-1 and found to mature normally [117]. This indicates that GATA-1 is one of the major erythroid factor involved in hematopoietic development.

iv. GATA-1 and hemangioblast

The GATA family of zinc-finger proteins is expressed in the blood islands of the developing mesodermal cells [8]. GATA-2, which is localized with Flk-1 in primitive blood islands, induces the expression of GATA-1 and is coexpressed with Tal-1 in the hemangioblasts [8]. The coexpression of GATA proteins with Flk-1 indicates a role for

these transcription factors in the regulation of Flk-1 expression. One GATA binding site in the Flk-1 promoter was found to be essential for endothelial cell-restricted expression of Flk-1 [20]. *In vitro* interaction of this protein with the binding site was verified and seemed to be essential for *in vivo* expression of Flk-1 [20]. To date, no further analysis of the GATA-1 protein on the development of endothelial cells has been elucidated. Recently, artificial zinc finger proteins have been created to target specific promoters thus regulating a gene of interest [118-120]. Since different zinc finger motifs can recognize unique DNA sequences, different combinations of these motifs are able to target different DNA sequences. Rebar et al. (2002) have designed zinc finger proteins to target VEGFa activity so as to increase blood vessel development. Normally overexpression of VEGFa induces "leaky" vessels, but injection of these synthetic transcription factors reduces the case of "leaky" vessel. Using artificial transcription factors to mimic endogenous regulatory mechanisms is a novel approach to treat diseases and conditions [120]. A company has recently patented their synthetic zinc finger proteins to target specific proteins for gene therapy [121]. Cardiovascular disease, congestive heart failure, cancer and sickle cell anemia are some of the conditions being investigated [121]. Considering the importance of designing transcription factors based on our knowledge of how they work, it is of interest to look at the role of GATA-1 in the regulation of genes involved in angiogenesis.

IV. Complex formation between bHLH and GATA-1

As was previously mentioned, Tal-1, Lmo2 and GATA-1 are all essential for hematopoietic development. It was thought then, that these 3 proteins could act synergistically in regulating genes by forming a complex. It was previously shown that Lmo2 can interact with Tal-1 and GATA-1 individually, but not E47 [122, 123]. In 1995, Osada *et al.* used immunoprecipitation experiments and found that these three proteins can indeed form a trimer. Lmo2 acts as a bridging molecule to link together GATA-1 and Tal-1. Interestingly, Lmo2 is a zinc-finger protein that does not bind DNA. Rather, this protein utilizes the zinc-finger domain to contact other proteins and act more as a cofactor in transcription factor complexes. In 1997, Visvader et al. found by yeast-2-hybrid screen that the protein Ldb1 interacts with Lmo2. They further characterized this interaction and showed that Lmo2 can bind both Tal-1 and Ldb-1 and that a complex can form between Tal-1, E47, Ldb-1 and Lmo2 in the absence of DNA [124]. To find the consensus sequence to which these proteins bind, CASTing experiments (cyclic amplification and selection of target sequences) were conducted in MEL cells (erythroid cell line expressing all four proteins) using Lmo2, Tal-1, E2A and GATA-1 antibodies [94]. These four experiments resulted in the discovery of the same consensus binding site where the E box was separated by 8-11 nucleotides from the GATA binding site [94]. The complex formation on this consensus site was verified by EMSA using supershifts for the five specific proteins [94]. Activation of a 30 nucleotide probe containing the consensus Ebox/GATA binding sites was seen when each protein was present separately, but there was synergistic activation when all 5 proteins were present [94]. Similarly, Grutz et al (1998) studied a complex formation in thymocyte cells that are affected in T-

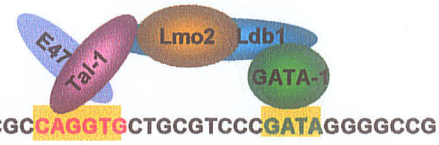
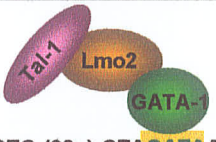
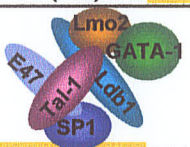
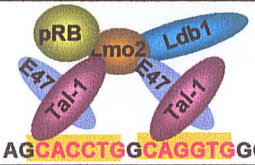
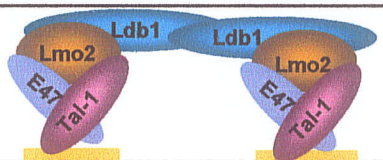
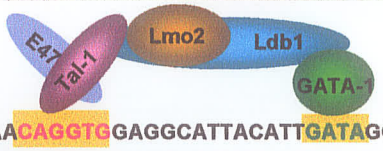
ALL leukemia. A CASTing experiment was conducted on a thymocyte cell line whose consensus binding sequence with Lmo2 antibody was detected as two E boxes in proximity in the same sequence. In this case, GATA was not involved, but it seemed that Lmo2 was a bridging molecule between Tal-1/E47 and Ldb1 [125].

Using Tal-1 antibody in chromatin immunoprecipitation assays, Cohen-Kaminsky et al. (1998) showed that the target gene of Tal-1 contains 2 E box binding sites in proximity (13 nucleotides) from a GATA binding site. This site was also sensitive to micrococcal nuclease indicating it was found in active chromatin [126]. They concluded that an intact Ebox-Ebox-GATA site is crucial for complex formation including Tal-1, GATA-1 and Lmo2 [126]. The gene target to which this sequence belongs is still unknown and thought to be found in an intron of an unknown gene. The authors postulated that this gene is involved in vascular development [126]. In 1998, two genes were identified as targets for differential complex formation between GATA-1, Tal-1 and Lmo2. Retinaldehyde dehydrogenase 2 (RALDH2) is a protein expressed in T-ALL cell lines that express ectopic levels of Lmo2 and Tal-1. The promoter region of this gene contained a consensus binding site for Ebox and GATA binding site in proximity to each other [95]. Interestingly, only the GATA site was found to be crucial for the effect of GATA-1, Tal-1 and Lmo2 [95]. Similarly, the erythroid Kruppel-like factor (EKLF) is a transcription factor important for β-globin expression. Considering the similar expression pattern to GATA-1, Tal-1 and Lmo2, it was thought that the expression of this gene is regulated by a complex of these factors. The EKLF promoter contained a GATA-Ebox-GATA motif that is critical for the regulation of this gene [127].

Vitelli et al. (2000) found two E box consensus sequences on the c-kit promoter separated by 1 nucleotide. c-kit is a transmembrane receptor protein involved in hematopoietic proliferation. In this case a complex of Tal-1, E2A, Lmo2 and Ldb-1 were found to form a complex utilizing the E box consensus site as DNA binding region. Interestingly, the retinoblastoma protein (pRB) was found to be involved in the complex. pRB is a tumour suppressor protein involved in the development of the hematopoietic system and found to be expressed at similar time points in development as Tal-1 [128]. Immunoprecipitation experiments revealed that pRB can bind to E2A, Tal-1, Lmo2 and Ldb1 proteins which support the idea that these proteins form a pentamer complex. Interestingly, it was found that early in development these proteins form a complex and negatively regulate the activity of c-kit, but later in development Tal-1 forms complexes with different proteins to increase the expression of c-kit [128]. The combinatorial assembly of E47, Tal-1, LMO2, Ldb1, and GATA-1/2 had a synergistic activation of c-kit promoter [129]. Each protein alone did not induce the promoter significantly. Deletion mutations of the promoter revealed an activitory site from -122 to -83 which lacked E boxes and GATA sites. A GC-box was found in this region which normally binds SP1. SP1 site is necessary for activation by the Tal-1 complex suggesting that SP1 recruits these factors to the promoter region. In this case, the HLH region of Tal-1 is required for activation and complex formation, but the DNA binding and transactivation domain are dispensable. In vitro pull-down assay revealed that SP1 binds to Tal-1, GATA-1, Lmo2 and Ldb-1, but not E47. The *in vivo* ChIP assay revealed that all these proteins forma a complex in vivo. Tal-1, Lmo2, Ldb-1 and E2A form a core complex and probably recruit GATA-1 through SP1 to act as coactivators [129]. These studies indicate

that not only the sites to which these complexes bind are important for gene regulation, but also the nature of the proteins involved in the complex. Table I summarizes the current research involving different complex formation between GATA-1 and Tal-1 proteins [94, 95, 125-130].

Table 1: Review of previous research involving complex formation between GATA-1, Tal-1 and E47

Reference	Target Gene	Schematic diagram
Wadman, I. A., H. Osada, et al. (1997). Embo J **16**(11): 3145-57.	unknown (erythroid cells)	
Lahlil, R., E. Lecuyer, et al. (2004) Mol Cell Biol **24**(4): 1439-52.	GPA	
Ono, Y., N. Fukuhara, et al. (1998). Mol Cell Biol **18**(12): 6939-50.	RALDH2	
Lecuyer, E., S. Herblot, et al. (2002). Blood **100**(7): 2430-40.	C-kit	
Vitelli, L., G. Condorelli, et al. (2000). Mol Cell Biol **20**(14): 5330-42.	C-kit	
Anderson, K. P., S. C. Crable, et al. (1998). J Biol Chem **273**(23): 14347-54.	EKLF	
Grutz, G. G., K. Bucher, et al. (1998). Embo J **17**(16): 4594-605.	unknown (thymocyte cells)	
Cohen-Kaminsky, S., L. Maouche-Chretien, et al. (1998) Embo J **17**(17): 5151-60.	unknown (MEL cells)	

V. Calreticulin

1. Calreticulin gene

Calreticulin was first isolated in 1974 by Ostwald and Maclennan. The protein was first purified based on its high affinity for calcium and thus first entitled HABCP (high-affinity calcium binding protein) [131]. It wasn't until 1989 that the cDNA was first cloned [132, 133]. The CRT gene consists of 9 exons and 8 introns and is 70% identical between the human (3.6kb) and mouse genes (4.6kb) [134, 135]. Introns 3 and 6 are the regions with the highest degree of difference between the mouse and the human CRT gene [135]. There is no alternate splicing, so all the exons are represented in the mRNA of 1.9kb [134]. A 1.8kb sequence of the CRT promoter was cloned from the mouse gene [135]. The human CRT gene is located on chromosome 19p30.2 and the murine CRT gene is located on chromosome 8 [134, 136]. The gene structure, specifically the promoter region, is very similar in structure to heat shock proteins [134]. This regulatory region will be discussed in more detail later.

2. Calreticulin protein

CRT contains 400 amino acids with a relative molecular weight of 46kDa [133, 137, 138]. Due to the high number of acidic residues in the CRT carboxy terminal, it runs between 55-60kDa on SDS-PAGE gels [139]. The CRT amino-acid sequence is highly conserved between species having a 90% identity between human, rabbit, rat and mouse [138, 140]. Based on the amino-acid sequence and structure, CRT has been divided into three domains termed the globular N-domain, proline-rich P-domain and acidic C-domain (see Figure 3) [132, 138, 141].

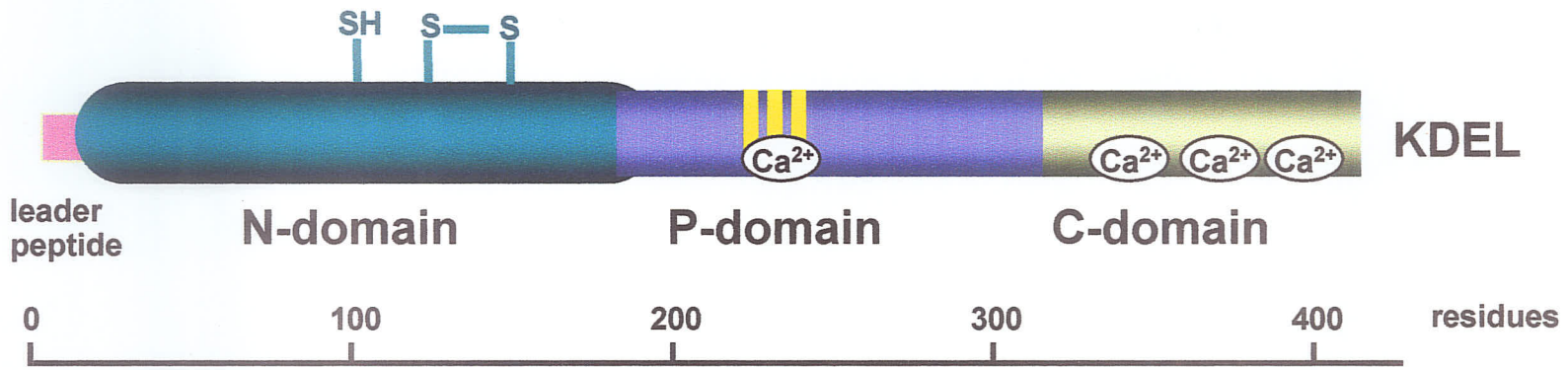


Figure 3: **Structure of CRT protein.** The protein is divided into 3 putative domains: N-domain, P-domain and C-domain. The highly hydrophobic N-domain forms a globular structure that contains 3 cysteine residues. The proline rich P–domain forms a finger-like extension composed of anti-parallel β-sheets. This domain contains a low capacity, high affinity calcium binding site. The C-domain contains a high capacity, low affinity calcium binding site and a KDEL ER retention sequence. This region probably has an extended conformation due to its high negative charge.

i. N domain

The N-domain has the highest amino-acid conservation between different species [140]. This region is highly hydrophobic and predicted to form a globular structure containing a helix-turn-helix motif followed by 8 anti-parallel β-sheets [132, 138, 141]. This globular structure is similar to calnexin and predicted to be involved in glucose binding [141]. Important histidine and cysteine residues within this region were found to be crucial for zinc binding [142]. These same residues may be involved in the interaction with protein disulphide isomerase (PDI) [143]. The N-domain also contains 3 cysteine residues, two of which form an intramolecular disulphide bridge: Cys120 and Cys146 [137]. Finally, this region is found to inhibit proliferation of endothelial cells and suppress angiogenesis, as will be discussed later [144, 145].

ii. P-domain

The P-domain, as the name suggests, is characterized by a high concentration of proline residues [132]. The elevated concentration of proline, serine and threonine residues in this region make the P-domain highly charged [138]. This region contains a high-affinity, low capacity calcium binding site but does not contain EF hands typical for calcium binding proteins [138, 146]. The NMR analysis of the P-domain reveals a hairpin fold resembling a finger-like extension that connects to the globular N-domain [132, 141, 147]. The two sides of the hairpin run anti-parallel to each other [132, 147]. The anti-parallel β-sheets are formed from 6 β-strands of which the first 3 are separated by regular intervals of 17 residues and the last 3 are separated by 14 residues [147]. This region is thought to be important for protein-protein interactions because of the similar amino-acid

sequence to other chaperones [134, 140, 147]. The N and P domains together encompass the chaperone region of the protein [141].

iii. C domain

The C-terminal region is highly acidic containing a high number of negatively charged amino acids such as aspartic and glutamic acid [132, 133, 138]. CRT is known to have a high capacity low affinity calcium binding site in this region [132, 138, 141, 146]. The lack of EF-hand structure typical for calcium binding implicates that the acidic region is responsible for calcium binding [132, 133, 146]. The tertiary structure of the C-domain is unknown, but is thought to have an extended conformation because of the charge repulsion from the acidic residues [132]. A KDEL ER-retention sequence and a possible N-linked glycosylation site (Asp-327) are found at the C-terminal of the protein [132, 133, 138, 140].

iv. Localization

CRT contains an ER signal sequence to direct it to the ER. Immunolocalization techniques confirmed this observation revealing that CRT is localized to the sarcoplasmic reticulum (SR) in skeletal muscle and ER in non-muscle [139, 148]. There was some controversy as to nuclear localization because of the ability of CRT to bind to nuclear steroid receptors *in vitro* [149, 150]. However, this finding has been shown to be due to non-specific antibody binding. CRT is ubiquitously expressed in most tissues in the body and has been localized to the vessel walls of vascular tissue [132, 151]. This suggests the necessity for CRT in normal processes in the system as a whole.

3. Calreticulin functions

CRT has many functions in the cell. Acting as an ER chaperone and regulating calcium homeostasis are the two major functions described for CRT. Other functions such as nuclear receptor gene regulation and cell adhesion have also been demonstrated for CRT and will be discussed below.

i. Chaperone function

Newly synthesized proteins are processed in the ER to be folded or modified into their correct tertiary structure. Chaperones bind to these newly synthesized unfolded proteins to assist in their processing to produce the native protein. The binding of chaperones to the unfolded proteins also inhibit their aggregation in the ER. Jorgensen et al. (2000) showed that CRT was able to bind to unfolded ovalbumin protein better than to native ovalbumin. Similarly, CRT was shown to bind to glycosylated and non-glycosylated proteins favoring the unfolded immature conformation of both types of proteins [152-154]. The ability of CRT to bind to unfolded, immature proteins indicates a chaperone function for this protein. CRT inhibited the aggregation of both glycosylated and non-glycosylated proteins [153]. In the past 10 years, a number of proteins have been shown to be chaperoned by CRT including: laminin [152], IgY antibody [153], ovalbumin [154], citrate synthase [153], malate dehydrogenase [153], HIV-1 envelope glycoprotein [155], MPO [156] and vinculin [157]. ATP and zinc have been shown to enhance the chaperone effect of CRT [153]. These ligands probably change the conformation of CRT resulting in an increase in the surface exposure of hydrophobic residues that aid in protein-protein interactions [153].

ii. Calcium homeostasis

Calcium release from the lumen of the ER increases the cytosolic calcium concentrations and induces multiple processes in the cell such as cell motility, proliferation, contraction, signal transduction and gene expression [158]. There are 3 factors that play a role in the regulation of cytosolic calcium concentration: the SERCA pump (SR/ER Ca^{2+} ATPase) involved in calcium uptake to the ER lumen, IP_3 receptor channels involved in calcium release from the ER and calcium-binding proteins (such as CRT) [158]. As mentioned, CRT was first identified by its interaction with calcium through a high-affinity low capacity binding site and a low-affinity high capacity binding site [131, 132, 138, 141, 146]. CRT knockdown by antisense oligonucleotide in NG-108-15 neuroblastoma X glioma cells induced a decrease in the calcium capacity of the ER [159]. On the other hand, overexpression of CRT in Hela and mouse L fibroblast cells increased the calcium capacity of the ER [160, 161]. This indicates that CRT has a direct effect on the amount of calcium stores available in the ER. The role of CRT in calcium homeostasis may also involve the interaction with the glycosylated residue in the intraluminal loop of the SERCA pump and the IP_3 receptor [162, 163]. CRT and IP_3 receptors are colocalized to the same compartments in the ER suggesting a role for CRT in the regulation of IP_3 activity [159]. Bradykinin is known to induce the activity of the IP_3 channel to cause transient cytosolic calcium elevation [159, 164]. Treatment of Mouse Embryonic Fibroblast (MEF) cells from wild-type and knockout mice with bradykinin increased cytosolic calcium of wild-type cells, however there were no significant changes in the cytosolic calcium of the knock-out cells [164]. Additionally, analysis in CRT knockout embryonic stem (ES) cell-derived cardiomyocytes revealed a

lack of proper myofibril formation [165]. This impaired development is due to the absence of an IP_3-mediated calcium spike that would normally stimulate MEF2C to translocate to the nucleus and activate MLC2v involved in cardiomyocyte development [165]. Thus, CRT affects the quick calcium release from the ER to the cytoplasm via the IP_3 pathway.

iii. Nuclear receptor gene regulation

CRT has been shown to bind to the peptide KLGFFKR [149]. This sequence is similar to a superfamily of DNA-binding nuclear receptors, thus it was suggested that CRT is involved in the regulation of these proteins. CRT has been shown to inhibit the ability of the glucocorticoid receptor and androgen receptor to bind to their cognate DNA binding sites [149, 150]. Overexpression of CRT in cells affected the level of the glucocorticoid-sensitive gene P450 by significantly increasing its mRNA and protein expression [149]. *In vitro* studies showed that the N-domain of CRT binds to the DNA-binding domain between the two zinc fingers of the glucocorticoid receptor, thus preventing the receptor from binding to the glucocorticoid response element (GRE) [149]. Similarly, overexpression of CRT decreased the level of retinoic-acid-responsive genes: CRABPII and RAR-B by binding to the androgen receptor DNA-binding site [150]. However, overexpression of a deletion mutant of CRT (deleting the KDEL or the C domain) resulted in a protein which was localized to the cytoplasm (not the nucleus) and failed to inhibit glucocorticoid receptor function [166]. Thus, only the ER form of CRT has the inhibitory effect on nuclear receptor binding to their target sequence. Yeast-2-hybrid analysis of the N-domain of CRT confirmed the finding that there was no

interaction between the DNA-binding domain of glucocorticoid receptor and CRT [166]. Furthermore, nuclear fractionation showed the presence of CRT in the outer nuclear membrane, not the inner membrane or the nuclear matrix [166].

iv. Cell adhesion

Overexpression of CRT has been shown to enhance cell adhesion, whereas a decrease of the CRT by antisense oligonucleotides resulted in a decrease in cell adhesion [157]. Integrins are transmembrane proteins formed of α and β subunits involved in cell adhesion by binding to different proteins present in the ECM [31, 32]. In vitro, CRT has been shown to interact with the α subunit of integrin [167, 168]. This interaction would require CRT localization outside of the ER. As mentioned, the presence of CRT is restricted to the ER, therefore the effect of CRT on cell adhesion must be indirect [157, 166, 169]. CRT overexpression induces a higher expression of the adhesion-related protein vinculin [157]. This allows CRT to exert its function from within the ER through a signaling mechanism. Fadel et al (1999) revealed that the cytoplasmic mediator tyrosine phosphatase was increased when the level of CRT was decreased. This suggests that the effect of CRT in adhesiveness is probably due to modulation of protein-tyrosine kinases or phosphatases.

4. Calreticulin gene expression studies

As mentioned above, CRT is conserved between species from plants, insects and fish to complicated eukaryotes. This indicates that CRT has been evolutionarily

conserved and is required for normal developmental process. To date, few studies are available about the mechanism of regulation of CRT gene expression.

i. AP-1

The importance of AP-1 in the regulation of CRT has only been studied in the schistosome parasite. The human homologue to CRT was cloned from the schistosome parasite in 1993 [170]. The promoter region of this gene was found to contain an AP-1 binding site. N-acetyl cysteine (NAC) is an inducer of AP-1 activity whose mechanism consists of eliminating cysteine residues that augment the DNA binding and subsequent activation of AP-1. NAC was found to enhance the activity of the CRT promoter and that mutation of the AP-1 site abolished this activation [171]. This suggests a critical role for AP-1 in the regulation of the CRT promoter.

ii. Heat-shock

Analysis of the 5' flanking region (or promoter region) of CRT showed that it contains similar sequences to other heat-shock proteins such as Glucose-Response Protein 78 (GRP78), GRP94 and PDI [134]. These three proteins have been shown to be upregulated in cases of stress such as heat-shock and depletion of intracellular calcium stores [135]. Because of the similarity in promoter structure to other heat-shock protein, CRT was postulated to be upregulated in cases of heat-shock. Indeed, heat shock of 42°C-43°C induced a 4-5-fold increase in human CRT expression as well as the expression of a known heat-shock protein 70 [172-174]. There is a putative heat-shock element (HSE) consisting of 3 inverted repeats of an nGAAn sequence in the human CRT promoter from

-172 to -158 [172]. A CRT promoter fragment containing this region was successfully upregulated after heat shock treatment [172]. *In vivo*, exposure of rats to 41°C heat for 60-90 minutes resulted in an increase in CRT level in lung tissue compared to control rats [172]. This upregulation of CRT in heat shock was seen in other cell types where CRT was expressed [172]. These data indicate a crucial function for CRT in heat shock sensitivity.

iii. Calcium

Increasing intracellular calcium by either thapsigargin (an inhibitor of ER calcium-ATPase) or the calcium ionophore A23187 (causes depletion of intracellular calcium stores) were shown to upregulate CRT expression [135]. The CCAAT sequence is the response element in the CRT promoter which is important for calcium effect. CCAAT, E boxes and GC-rich regions are three regulatory elements shared in the promoters of CRT, GRP78, GRP94 and PDI [134, 173]. The CCAAT site appeared every 30-50 bases regularly in these four related proteins, indicative of a site involved in calcium response [134]. Similar results were seen with the human CRT promoter. 504bp of human CRT promoter was found to be upregulated with the addition of the calcium ionophore A23187 [173]. This upregulation was specific for calcium because the potassium ionophore valinomycin had no effect on the promoter activity [173].

iv. Heavy metal

The fact that CRT has been known to be affected by a divalent cation (calcium), it was postulated that other cations can induce CRT expression. Indeed, the human CRT

promoter was shown to be upregulated 6-fold with the addition of 240uM of $ZnCl_2$ to the media of transfected cells [173]. In addition, the zinc chelator EDTA was found to inhibit this activation indicating that it is a specific effect of zinc ion [173]. Zinc interacts with the promoter through a metal response element (MRE) with a sequence of TGCTCXC on the promoter of genes. The CRT promoter contains 4 putative MRE elements indicating a possible mechanism by which CRT is stimulated by zinc [173].

v. Regulation in heart development

Studies on developmental regulation of CRT promoter in mouse (using GFP reporter gene) showed high CRT promoter activity in the heart, vascular tissue and the brain of the developing mouse embryo [164]. After birth, the expression of CRT in the heart decreases significantly [164]. Furthermore, deletion of the CRT gene results in a defect in heart development causing embryonic lethality of CRT null embryos [164, 175]. Overexpression of CRT in the heart of transgenic mice caused sudden death of mice between 1-2 weeks old [176]. The hearts of these mice were normal in size, but had thinning of the ventricular walls resulting in enlargement of the atria and ventricles [176]. Thus, the precise regulation of CRT in heart development is crucial. Two transcription factors were found to have profound importance in the regulation of CRT expression in the heart: Nkx2.5 and COUP-TF1 [177]. Nkx2.5 is a homeobox transcription factor important for early heart development [178] and has similar expression patterns to CRT during development [177]. Three Nkx2.5 response elements are found on the mouse CRT promoter and were found to increase CRT expression in transfected fibroblast cells and cardiomyocytes [177]. Interestingly, this response element CTCAAGTGT has similar

structure to the COUP-TF1 binding site AGTTCA. COUP-TF1 was found to have an inhibitory effect on the CRT gene [177]. The mechanism of regulation of CRT during heart development is thought to be controlled by the spatial and temporal expression of these two transcription factors. Early in heart development Nkx2.5 is expressed at its highest level, this protein binds preferentially to the binding site and blocks COUP-TF1, thus activating CRT expression in the heart. In the postnatal heart, when Nkx2.5 is downregulated, COUP-TF1 is free to bind to the consensus sequence and represses CRT expression [177]. The role of other transcription factors important in heart development has not yet been studied with respect to the regulation of CRT.

5. Calreticulin and angiogenesis

The developmental regulation of CRT expression was studied using the reporter gene construct containing the Green Fluorescent Protein (GFP) under the control of the CRT promoter [164]. In early stages of embryogenesis, CRT is highly expressed in the heart. The activation of the CRT promoter was also detected in the developing blood vessels including intersomitic vessels, dorsal aorta and smaller arteries [164]. In addition, the CRT reporter gene was detected in the heart, eye, brain, liver and midgut at different embryonic stages [164]. Levels of CRT in the heart are negligible after birth, whereas CRT expression in the blood vessels is maintained throughout life [164, 179, 180]. To date, no data is available on the mechanisms regulating CRT gene expression in vascular wall.

A role for CRT in vessel development was postulated due to the high expression of this protein in the vessel wall during embryonic development [164]. CRT is also

expressed in adult endothelial and vascular smooth muscle cells [181]. Dai et al (1997) were the first to show a role for CRT in the regulation of blood vessel wall [182]. These authors demonstrated that exogenously added CRT decreases intimal hyperplasia after balloon angioplasty [151]. Subsequently, CRT and truncated forms of CRT containing the amino-terminal end (called vasostatin) was shown to inhibit endothelial proliferation when added to the media [144]. Both CRT and vasostatin were able to inhibit tumor angiogenesis when injected with the tumor cells in the mice model of Burkitt lymphoma tumors [144, 145, 183]. Further studies on CRT protein showed that amino-acid residues 120-180 possess the antiangiogenic activity [145]. The mechanism of this effect is not confirmed yet, however, Pike *et al* (1999) postulated that this effect might be due to the release of nitric oxide (NO). Indeed, previous studies have shown that exposure of endothelial cells to CRT results in increase NO production [151]. Another possible mechanism for regulation of vessel wall could be due to downregulation of MMP-2 activity [184]. The suppression of CRT expression in human rhabdomyosarcoma (RD) cells by anti-sense oligonucleotides has been demonstrated to inhibit the activation of MMP-2 by decreasing proMMP-2 secretion [184]. Similarly, overexpression of CRT in RD cells induced a 2.2-fold increase in proMMP-2 secretion [184]. MMP-2 activity is important for the degradation of the extracellular matrix which stimulates endothelial cell migration and proliferation. Similarly, vasostatin was found to bind to laminins which are proteins that allow endothelial cells to attach and migrate during angiogenesis [185]. Vasostatin was able to physically prevent the interaction between laminins and endothelial cells resulting in the eventual inhibition of angiogenesis [185].

Interestingly, vasostatin inhibited tumour growth of the Burkitt lymphoma cells without affecting the wound healing process in mice subjected to skin wounds [186]. This indicates specificity for vasostatin in inhibiting tumour angiogenesis compared to normal angiogenic processes such as wound healing suggesting that vasostatin is a good candidate for gene therapy. Indeed, in 2002, a mouse tumour model was subjected to vasostatin gene therapy. Mice bearing tumours were intramuscularly injected with a vasostatin plasmid which resulted in a decrease in tumour growth without side-affects in these mice after four weeks compared to control mice [183]. Chorioallantoic membrane assay (CAM) indicated that vasostatin inhibits angiogenesis and does not affect larger pre-existing vessels [183]. Also, the vasostatin-treated tumour cells seemed to be undergoing more apoptosis than control cells, indicating a decrease in tumour progression [183]. The addition of interferon-inducible protein-10 (IP-10) in combination with vasostatin reduced significantly the amount of tumour produced in Burkitt lymphoma mice [185]. IP-10 has been shown to induce tumour necrosis in Burkitt lymphoma cells [187, 188]. Thus, vasostatin may require different proteins to induce a higher inhibition of tumour angiogenesis.

The fact that CRT is involved in vascular development and angiogenesis leads us to explore the transcriptional regulation of the CRT protein expression. The understanding of CRT gene regulation may have an importance in the treatment of different diseases such as cancer and cardiovascular disorders.

B. RATIONALE

Blood vessel development requires a complex signaling pathway involving tyrosine kinase receptors and their ligands. Proteins involved in other functions such as cell migration, cell proliferation and extracellular matrix degradation are crucial in this form of development. Calreticulin, an ER resident protein, has been implicated in the regulation of cell migration, adhesion and protein folding. In addition, a recent study of the CRT promoter in transgenic mice suggested a role for this protein in cardiovascular development [164]. Gene targeted deletion of CRT in mice resulted in defects in both cardiac development and wound healing [164, 175]. In the adult, CRT is downregulated in myocardium [164] whereas, it maintains high expression in endothelial and smooth muscle cells [181]. These observations suggest a possible role for CRT in vessel development (angiogenesis). However, the mechanism of regulation of CRT expression in the vascular tissue in not known. Transcription factors are proteins important in the regulation of spatial and temporal expression of genes at different stages of development. To date, no data are available on the transcriptional regulation of CRT gene in blood vessels.

The understanding of the transcriptional regulation of CRT can have therapeutic applications, whether to cease the spread of tumour progression by halting angiogenesis or to accelerate the re-oxygenation of damaged tissues in cardiovascular disease. Different transcription factors are crucial in the molecular regulation of proteins involved in vessel development. Two of these proteins, Tal-1 and GATA-1, were found to be expressed in endothelial cell precursors and necessary for the formation of functioning blood vessels [8]. Similarly, the E47 protein is ubiquitously expressed and found to be

involved in smooth muscle cell differentiation [77]. These three proteins have been previously found to form a complex that facilitates their regulatory role. Analysis of the CRT promoter revealed four putative Tal-1/E47 binding sites and eight GATA-1 binding site (Figure 4). Therefore, we tested the hypothesis that "**CRT is regulated by a complex between Tal-1, E47 and GATA-1 proteins during blood vessel development**". The specific aims of the research were:

(1) to study the transcriptional activity of Tal-1, E47 and GATA-1 alone on the CRT promoter,

(2) to investigate if these three proteins can synergistically upregulate the CRT promoter,

(3) to demonstrate the possible complex formation between these proteins on two sites of the CRT promoter which contain the consensus sequence (Tal-1/E47 binding site 8bp from GATA binding site).

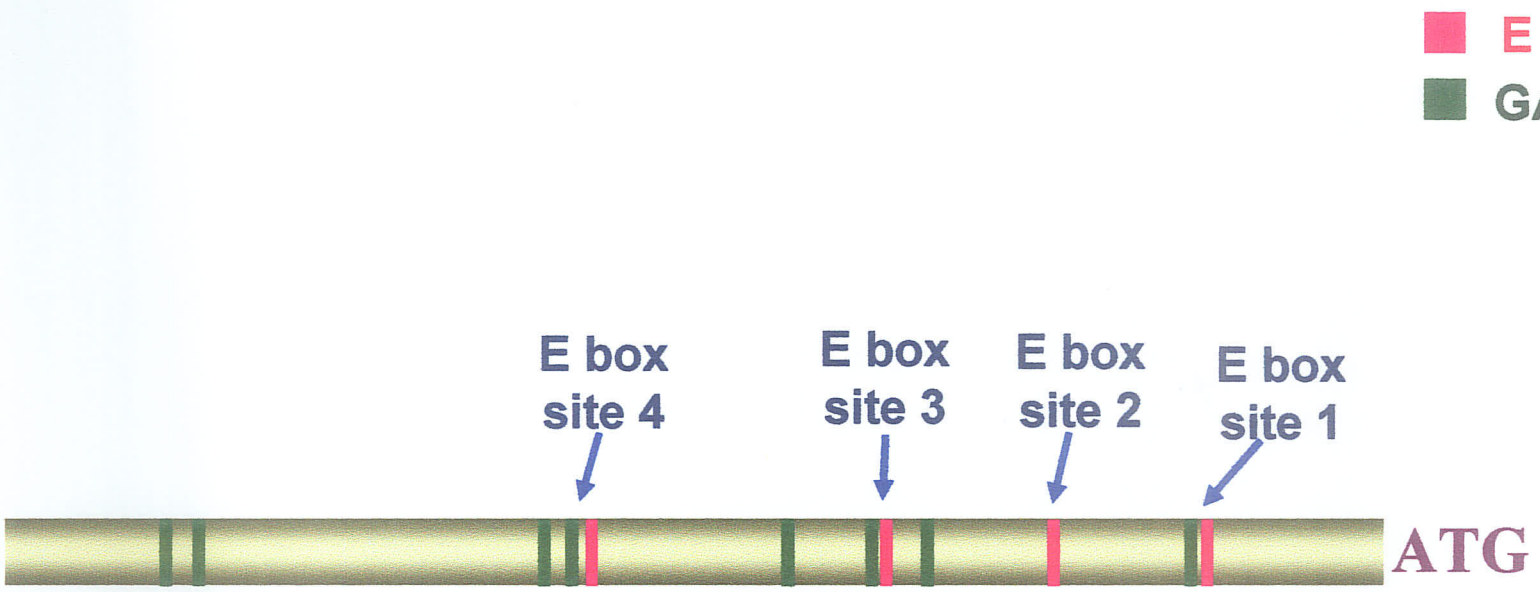


Figure 4: **CRT promoter with 4 putative Tal-1/E47 binding sites.** There are four putative E boxes on the CRT promoter (pink) labeled 1 to 4 in respect to the start site of the gene (ATG). GATA binding sites (green) are in proximity to E box binding sites (pink).

C. MATERIALS AND METHODS

I. Materials

Tissue culture media, Lipofectamine 2000, restriction enzymes, DNA polymerase and T4 kinase were purchased from Invitrogen (Burlington, ON). The DC protein assay kit and the prestained SDS-PAGE low range molecular weight standards used in western blot analysis was purchased from Bio-rad Laboratories (Hercules, CA). The Supersignal West Dura Extended Duration Substrate was purchased from Pierce (Rockford, IL). The pCMV5A and pCMV5B vectors as well as the QuickChange XL Site-Directed Mutagenesis Kit were purchased from Stratagene (La Jolla, CA). The PGL3 vector and the Beetle luciferin potassium salt were purchased from Promega (Madison, WI). The QIAGEN maxi prep plasmid purification was from QIAGEN (Mississauga, ON). The Vectashield mounting medium with DAPI was from Vector Laboratories (Burlingame, CA). The FuGENE 6 reagent was from Roche Diagnostics (Montreal, PQ). All primers used in cloning and EMSA were purchased from Sigma-Aldrich (Oakville, ON). All other chemicals and reagents were of the highest Molecular Biology grade.

II. Plasmids and reporter constructs

1. Plasmids used

PGK-E47 plasmid was a generous gift from Dr. Skerjanc (University of Western Ontario). The Tal-1 cDNA in pBluescript (pB-Tal) was a generous gift from Michelle Kelliher (University of Massachussetts Medical School). The pcDNA3.1-lacZ plasmid was a generous gift from Dr. Hicks (University of Manitoba). The pHBAPneo-E47 plasmid was a generous gift from Dr. Romanow (University of California San Diego).

pXM-GATA-1 plasmid was also a generous gift from Dr. Orkin (Harvard Medical School). The PGL3 luciferase reporter vector was purchased from Promega. The pCMV-5A and pCMV-5B vectors were purchased from Stratagene.

2. Luciferase expression vectors - CRT promoter deletions

The full-length 1.8kb CRT promoter was a gift from Dr. M. Michalak, University of Alberta, as described in Waser et al (1997) [135]. The full-length CRT promoter was cloned into the promotorless PGL3 luciferase reporter vector (hereafter to be referred to as CPF). Deletion constructs of the promoter were used to specify the importance of individual E box sites on the activity of CRT gene. These promoter constructs were generated by sequentially eliminating one E box binding site either from the 5' end or the 3' end (see Figure 5 for these expression vectors). The 5' deletion constructs eliminated the 4^{th} E box (CP3T), the 4^{th} and 3^{rd} E boxes (CP2T) and the 4^{th}, 3^{rd} and 2^{nd} E boxes (CP1T). Similarly, the 3' deletion constructs were generated by eliminating the 1^{st} E box (CPB3T), the 1^{st} and 2^{nd} E boxes (CPB2T) and the 1^{st}, 2^{nd} and 3^{rd} E boxes (CPB1T). (See Figure 5 for details of these expression vectors.) These deletion vectors were constructed using the CPF vector (6603bp) as a template cut with appropriate restriction enzymes and religated: CP3T (5509bp PstI restriction fragment), CP2T (5375bp BlnI blunted and EcoRV restriction fragment), CP1T (5203bp KpnI restriction fragment), CPB3T (6274bp XhoI klenow filled and PmlI restriction fragment), CPB2T (6064bp XhoI/BlnI blunted restriction fragment), and CPB1T (5700bp BfrI/XhoI blunted restriction fragment). To ascertain the importance of a single E box site, CPF-MT3 and CP3T-MT3 were generated

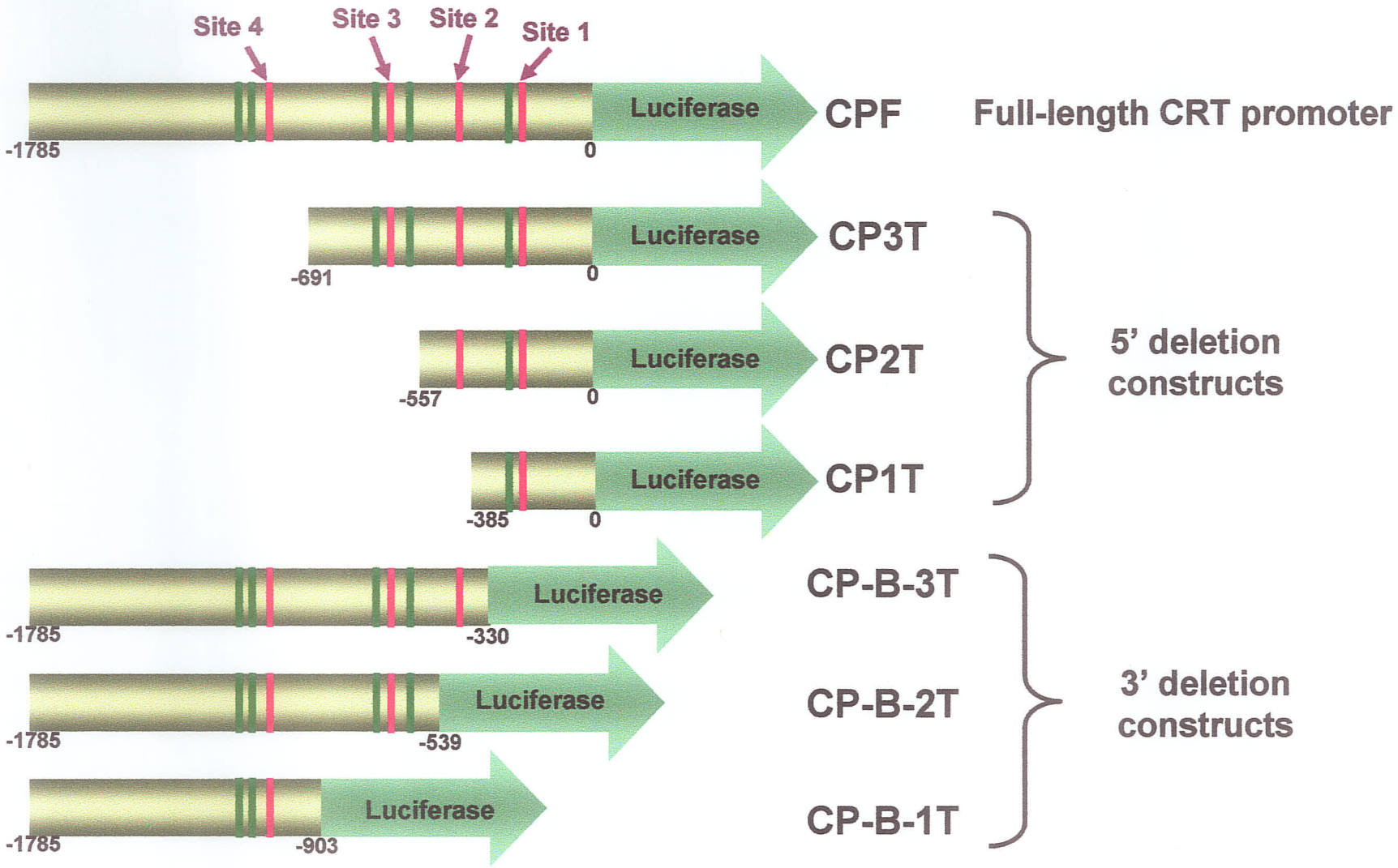


Figure 5: **Schematic diagram of different deletion constructs of the CRT promoter.** CRT promoter deletion constructs were generated from the full-length CRT promoter (CPF) by sequentially eliminating one putative E box binding site from either the 5' or 3' end and cloned upstream of the luciferase gene.

by introducing 2 point mutations in the third E box as described in the site directed mutagenesis section below.

3. Site-Directed Mutagenesis

Two point mutations (from **C**AAC**T**G to **A**AAC**A**G) in the 3rd E box of CPF and CP3T were generated using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) following their instructions. Briefly, 5-50ng of CPF or CP3T plasmid was mixed with 125ng of MT3-5' primer and 125ng of MT3-3' primer (see Table 2 for sequences), reaction buffer, dNTP and PfuTurbo DNA polymerase. The PCR reaction consisted of 18 cycles including a 50 second denaturing step at 95°C, 50 second annealing step at 60°C and a 2 minute elongation step at 68°C. The PCR products were treated with DpnI restriction enzyme to cleave the template plasmid. The resultant constructs were transformed in to ultracompetent XL-10 cells and positive clones were selected for resistance to ampicillin. CPF-MT3 and CP3T-MT3 were sequenced at the University of Calgary to verify the correct point mutations in the E box.

4. Myc-tag constructs

The myc-tag constructs were constructed by PCR amplification. Using 100ng of pcDNA-hygro-Tal-1 plasmid as a template, a 1kb fragment of DNA was amplified using 0.2μg/μl of Tal-QE9-5' primer and 0.2μg/μl of Tal-myc-3' primer (see Table 2 for sequences). Similarly, a 1.6kb fragment of DNA was amplified using 100ng of pGK-E47 plasmid as template and 0.2μg/μl of each the E47-QE9-5'primer and the E47-3'-MH primer (see Table 2 for sequences). 100ng of pXM-GATA-1 plasmid was used as a

Table 2: Oligonucleotide sequences used for cloning

Name	Oligo Sequence	Restriction site
Myc constructs		
Talmyc3	ATAT**CTCGAG**GGGGCCAGCCCCATC	Xho1
Tal-QE9-5	TATA**GGATCC**ATGACGGAGCGG	BamHI
GATAmyc3	TATA**CTCGAG**TGAACTGAGTGGGG	Xho1
GATA-QE9-5	TATA**GGATCC**ATGGATTTTCCTGG	BamHI
E473MH	TATA**AAGCTT**AGGTGCCCGGCTGGGT	HindIII
E47-QE9-5	TATA**GGATCC**ATGATGAACCAGTC	BamHI
Site-directed mutagenesis		
MT3-3'	GCATTCTGGTTCCACTGTTTTTCAGGCCATAGAGAACG	
MT3-5'	CGTTCTCTATGGCCTGAAAAACAGTGGAACCAGAATGC	

(Note: Restriction enzyme sites of the primers are in bold)

template to produce a 1.2 kb fragment of DNA amplified with GATA-QE9-5' and GATA-myc-3' primers (both 0.2μg/μl, see Table 2 for sequences). The 35 cycle PCR reaction included a 45 second denaturing step at 94°C, a 1 minute annealing step at 55°C and a 1 minute elongation step at 72°C. The resulting PCR products for GATA-1 and Tal-1 were cut with XhoI and BamHI, These DNA fragments were annealed to the pCMV5A and pCMV5B vector (Stratagene) respectively. The E47 PCR product was cut with HindIII and BamHI and annealed to the pCMV5B vector. The resulting plasmids: pCMV5B-Tal (Tal-myc), pCMV5B-E47 (E47-myc), and PCMV5A-GATA (GATA-1-myc) were sequenced at the University of Calgary Core DNA and Protein Services to verify that the coding sequence of each construct was in-frame. Expression of the myc-tagged proteins was verified by Western Blot analysis and immunocytochemistry using the 9E10 monoclonal antibody (hybridoma bank of University of Iowa).

III. Cell lines

HEK-293 (human embryonic kidney) cell line was maintained in Dulbecco's Modified Eagle's Medium (DMEM) plus 10% fetal bovine serum. NIH-3T3 (fibroblast) cells were cultured in DMEM plus 10% calf bovine serum. HEK-293 cell line expresses GATA-1. NIH3T3 was selected because it does not express Tal-1 or GATA-1 proteins (Figure 6)

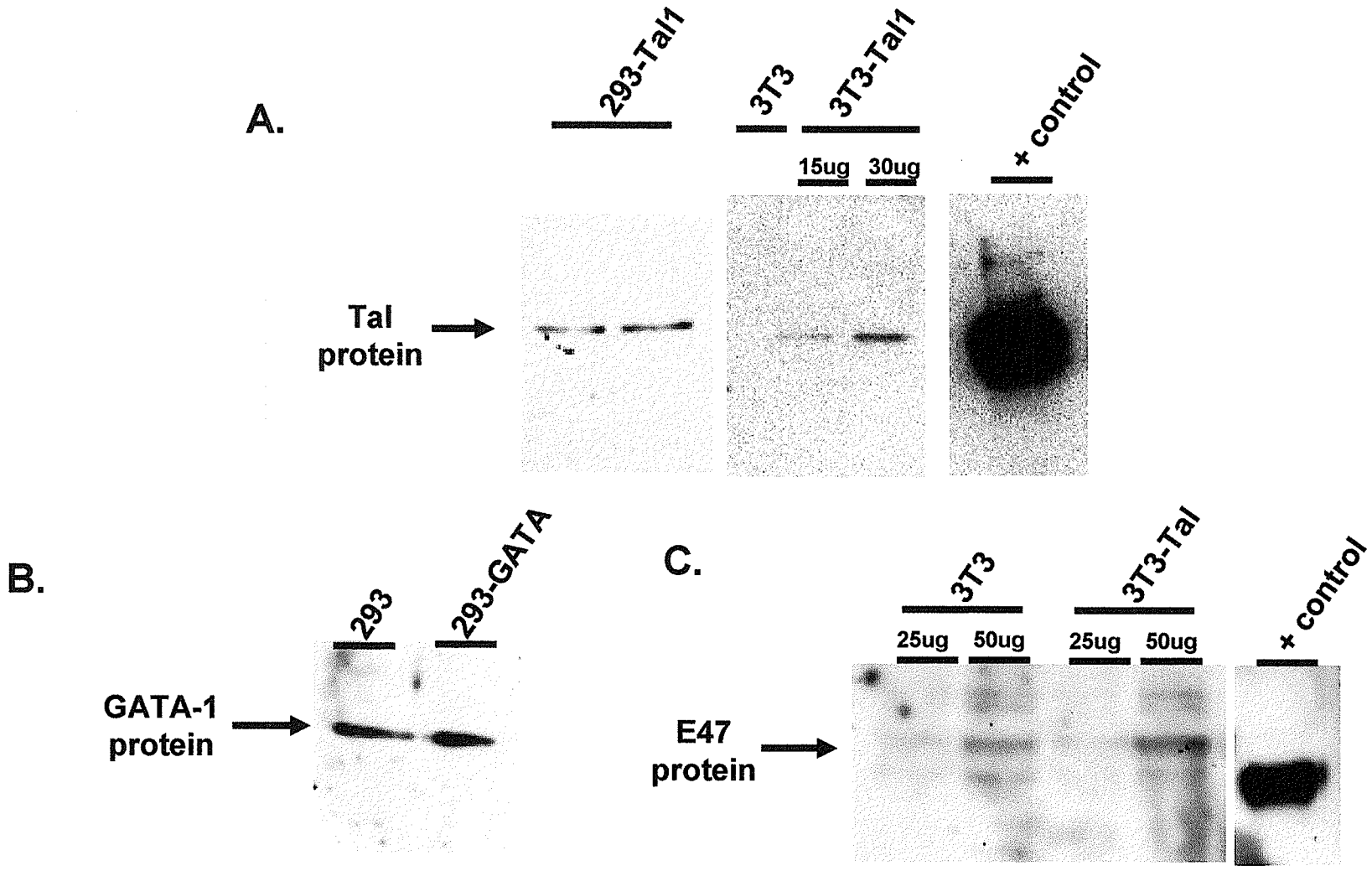


Figure 6: **Expression of Tal-1, GATA-1 and E47 in different cell lines.** ***A***. Western of cell lines with Tal-1 antibody (Geneka). ***B***. Western of HEK-293 cells with GATA-1 antibody (Geneka). **C**. Western of cell lines with E47 antibody (Geneka).

IV. Transfection and generation of stable cell lines

Plasmid DNA from various constructs described above was purified by Qiagen column chromatography. The following three methods of transfection were used according to cell line and efficiency of transfection required.

1. Generation of stable cell lines

For the generation of stable cell lines, NIH-3T3 cells were transfected using Lipofectamine 2000 (from Invitrogen) with 10μg of pcDNA-Tal plasmid and treated for 2 weeks with the antibiotic G418 to select for positive clones. Similarly, 293 cells were transfected using the calcium-phosphate method with 10ug of each construct: Tal-myc, E47-myc or GATA-myc. Single clones of cells expressing each of the 3 different constructs were selected for resistance to G418 after a 2 week treatment. Nuclear extracts (see below) from these stable cell lines were resolved on SDS-PAGE gels followed by western blot analysis. A primary antibody to myc (Santa Cruz) was used to identify clones expressing a high level of the myc-tagged protein.

2. Transient transfection

Transient transfection was used for reporter gene assays. HEK-293 and NIH-3T3 cells were seeded onto 6-well plates (50 000 cells per well) 24hrs prior to transfection. Each sample was transfected with 2μg of reporter plasmid (CPF, CP3T, CP2T, CP1T, CPB3T, CPB2T, CPB1T, Tal2X or PGL3), 2-3μg total of expression plasmids and 2μg of B-gal plasmid for normalization of transfection efficiency. The amount of total plasmid was kept constant at 6-7μg to avoid mass-dependant affects of each transfection.

GATA-myc, Tal-myc, E47-myc or a combination of these expression vectors were added to each transfection sample. If the amount of the vectors was less than the total plasmid volume, an empty vector (pCMV5B) was added to keep the volume constant.

V. Nuclear Extract Preparation

Nuclear extract preparation was important for Electrophoretic Mobility Shift Assay (EMSA). Stable cell lines and their respective control cell lines (HEK-293 and NIH-3T3) were plated on 20cm plates and left to grow until 80% confluency. Cells were washed with phosphate-buffered saline (PBS) twice at room temperature and 1ml of cold lysis buffer (PBS+1mM Sodium orthovanadate) was added to the cells. The plate was scraped and the cells were spun at 9000rpm for 1min at 4°C. 1 volume of hypotonic buffer (20mM HEPES pH 7.9, 20mM NaF, 1mM Sodium-pyrophosphate, 1mM Sodium-orthovanadate, 0.25mM Sodium-molybdate, 40mM B-glycerophosphate, 1mM EDTA, 1mM EGTA, 2mM DTT, 0.5mM PMSF, 10ug/ml SL inhibitors) was added to the pellet followed by the addition of $1/10^{th}$ volume of 0.5% NP40 buffer (diluted in 1X hypotonic buffer). Cells were spun at 13,000rpm at 4°C for 1 min and $1/5^{th}$ volume of high salt buffer (hypotonic buffer + 20% glycerol, 0.4M NaCl) was added to the pellet and the resulting mixture was left to shake for 1 hour at 4°C. Cell debris was removed by centrifugation at 13,000rpm at 4°C for 20 minutes. The protein concentration of the supernatant was measured using the DC Protein Assay kit (Bio-rad).

VI. Western blotting

To analyze protein expression, 25-50μg of proteins from the nuclear extracts were loaded on a 10% Sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE). The proteins were separated and the gel was transferred to nitrocellulose membrane by semi-dry transfer system. The membrane was blocked with 5% milk-powder diluted in 1X PBS for 1 hour, then incubated with primary antibody overnight at 4°C [goat anti-c-myc (1:200 Santa Cruz), rabbit anti-Tal-1 (1:500 Geneka), rabbit anti-E47 (1:500 Geneka), or rabbit anti-GATA (1:500 Active Motif)]. The following day, the membrane was incubated with an HRP-secondary antibody (for respective species) for 30 minutes (1:10000 Jackson Laboratories). The secondary antibodies were detected using enhanced chemiluminescence detection system (SuperSignal West Dura) and visualized on a Fluor-S-Max MultiImager (Bio-Rad).

VII. Immunocytochemistry

Immunocytochemistry was utilized to localize myc fusion proteins to the nucleus. Stable cell lines were plated on glass cover slips. The next day, cells were fixed in 4% formaldehyde in PBS for 15 minutes, followed by 3 subsequent washes with 0.1% saponin in PBS. The coverslips were transferred to holding container and blocked for 30 minutes with 0.1% saponin in PBS and 2% milk powder. The primary antibody (mouse anti-myc 9E10 1:50, from the hybridoma bank of University of Iowa) was incubated on the coverslips for 1 hour followed by 30 minute incubation with goat anti-mouse FITC (1:70 Jackson Laboratories). Coverslips were then mounted on slide using Vectashield

mounting media with DAPI (Vector Laboratories) to stain the nucleus. The fluorescence was visualized with a Zeiss Axiovert Fluorescent microscope.

VIII. Reporter Gene Assay

NIH-3T3 and HEK-293 cells were plated at $5X10^5$ cells/well in 2cm plates. The following day, cells were transiently co-transfected with 2μg of luciferase reporter plasmid, 2μg of β-gal plasmid and 1μg of each expression vectors (as described above). 48 hours following transfection, cells were washed with PBS and lysed with 100μl of NP40 lysis buffer (100mM Tris pH 7.8, 0.5% NP40, 50mM DTT). The cell lysate was centrifuged briefly to precipitate insoluble material. 20μl of the cell lysate was aliquoted in duplicate into tubes and mixed with 100μl of luciferase assay buffer (20mM Tricine, 1.07mM $MgCO_3$, 2.67mM $MgSO_4$, 0.1mM EDTA, 33.3mM DTT, 270μM coenzyme A, 470μM luciferin, 530μM ATP). For β-galactosidase assay, 20μl of the cell lysate was aliquoted in duplicate into a 96-well plate and mixed with 80μl ddH2O and 20μl of 4mg/ml of ONPG. After incubation at 37°C, the absorbance at OD420 of this mixture was determined on a microplate reader. For normalization of transfection efficiency, the average luciferase value was divided by the average β-gal value. The results are presented as bar graphs generated using Microsoft Excel.

IX. Electrophoretic Mobility Shift Assay (EMSA)

1. Probe labeling

Probes were selected from the CRT promoter, site 1 and site 3, both of which contain GATA-1 and E box binding sites. Both sense and antisense oligos representing 32

nucleotides of the specified sequences were purchased from Sigma (see Table 3 for a list of oligos). The probes were generated by annealing corresponding 3' and 5' oligonucleotides listed in Table 3. An equal volume of each oligonucleotide was mixed together and heated to 100°C in a heating block, followed by gradual cooling to room temperature. The probes were then labeled by kinase assay using [γ-^{32}P]-ATP. Briefly, 100ng of probe was mixed with T4 polynucleotide kinase (Invitrogen), 5X buffer and 3,000Ci/mM of [γ-^{32}P]-ATP. The mixture was incubated at 37°C for 30 minutes after which a Stop solution was added (1% SDS, 100mM EDTA). 25µl of sample was centrifuged at 3,000rpm for 1 minute in a Sephadex G50 column and washed with 75µl of ddH2O to purify the probe from unincorporated ^{32}P. A 1µl aliquot of each sample was tested for activity (cpm/µl), each probe activity was at least 5X10^5 cpm/µl.

2. EMSA

Nuclear extracts and 2.5X binding buffer (50mM Hepes pH 7.6, 125mM KCl, 2.5mM EDTA, 65.5% glycerol, 2.5 mM DTT) were mixed with ~5x10^{5}cpm/µl of probe and incubated at room temperature for 30 minutes. All samples were loaded on a pre-run 5% non-denaturing polyacrylamide gel (0.5X TBE buffer, 38:2 acrylamide:bisacrylamide) and run at 200V for 2-3 hrs at room temperature in 0.5X TBE buffer (10X: 890mM Tris, 890mM Boric Acid, 20mM EDTA pH 8.0). The gel was then vacuum dried on a Bio-Rad gel dryer for 1.5 hours. Gels were wrapped in seran wrap and visualized on a phosphoimager screen or X-ray film. To specify the protein interactions, the nuclear extracts and binding buffer was premixed with antibodies [mouse monoclonal anti-Tal-1 (Active Motif), rabbit polyclonal anti-GATA-1 (Santa Cruz), rabbit polyclonal anti-E47

Table 3: Oligonucleotide sequences used as probes for EMSA

Name	Description	Oligo Sequence
TLG3-5	32nucl (3rd site)	ACGTTCT**CTAT**GGCCTGAA**CAACTG**TGGAACC
TLG3-3	32nucl (3rd site)	GGTTCCA**CAGTTG**TTCAGGCC**ATAG**AGAACGT
TLG1-5	32nucl (1st site)	GCGACCA**ATAG**AAATCAGC**CATCTG**GGATCCC
TLG1-3	32nucl (1st site)	GGGATCC**CAGATG**GCTGATTT**CTAT**TGGTCGC
TLG3M-5	32nucl (3rd site)-mutated Ebox	ACGTTCT**CTAT**GGCCTGAA**CAACAA**TGGAACC
TLG3M-3	32nucl (3rd site)-mutated Ebox	GGTTCCA**TTGTTG**TTCAGGCC**ATAG**AGAACGT
TLG1M-5	32nucl (1st site)-mutated Ebox	GCGACCA**ATAG**AAATCAGC**CATCAA**GGATCCC
TLG1M-3	32nucl (1st site)-mutated Ebox	GGGATCC**TTGATG**GCTGATTT**CTAT**TGGTCGC
TG1MM5	32nucl (1st site)-mutated GATA + Ebox	GCGACCA**CTCG**AAATCAGC**CATCAA**GGATCCC
TG1MM3	32nucl (1st site)-mutated GATA + Ebox	GGGATCC**TTGATG**GCTGATTT**CGAG**TGGTCGC
TG3MM5	32nucl (3rd site)-mutated GATA + Ebox	ACGTTCT**CGAG**GGCCTGAA**CAACAA**TGGAAC
TG3MM3	32nucl (3rd site)-mutated GATA + Ebox	GGTTCCA**TTGTTG**TTCAGGCC**CTCG**AGAACGT
TG1GM5	32nucl (1st site)-mutated GATA site	GCGACCA**CTCG**AAATCAGC**CATCTG**GGATCCC
TG1GM3	32nucl (1st site)-mutated GATA site	GGGATCC**CAGATG**GCTGATTT**CGAG**TGGTCGC
TG3GM5	32nucl (3rd site)-mutated GATA site	ACGTTCT**CGAG**GGCCTGAA**CAACTG**TGGAACC
TG3GM3	32nucl (3rd site)-mutated GATA site	GGTTCCA**CAGTTG**TTCAGGCC**CTCG**AGAACGT

(Note: The GATA binding sites and E boxes of each oligonucleotide are in bold.)

(Santa Cruz) and goat anti-c-myc (Santa Cruz)] and incubated at room temperature for 30 minutes prior to the addition of the ^{32}P-labelled probe. To locate unspecific bands, an excess of cold probe (10-, 100- or 1000-fold excess) was added to the mixture at the same point as the ^{32}P-labelled probe.

X. Statistics

Mean and standard error were plotted on a graph for each individual experiment. The statistical differences between two values were assessed using Student's T-test and a P value <0.05 was considered as significant difference.

D. RESULTS

I.. CRT promoter analysis

The 1.8kb mouse CRT promoter sequence was analyzed using the TRANSFAC database to search for potential transcription factor binding sites [189]. This sequence analysis revealed putative binding sites for many different transcription factors. Of interest were 4 different putative E boxes in proximity to GATA binding sites (see Figure 4). As mentioned, previous studies have indicated complex formation between bHLH proteins and GATA transcription factors when their binding sites are in proximity on the promoter (8-10 nucleotides apart) [94]. Sites 1 and 3 on the CRT promoter contain this exact consensus sequence. To determine if these putative E box and GATA binding sites are conserved in different species, the DNA sequences of the CRT promoter of mouse, rat and human were compared using the Clustal-Wallis program (http://www.ebicac.uk/clustalw/). This Clustal-Wallis alignment of the different promoters revealed perfect conservation of site 1 E box and GATA binding sites as well as the 8bp linker region between the sites (see Figure 7A), while site 3 had differences in the E box and GATA binding site nucleotides between species (see Figure 7B).

II. Myc-tag protein expression vectors

To study the role of Tal-1, E47 and GATA-1 transcription factors in regulation of CRT promoter, we generated myc-tagged proteins as described in the Materials and Methods. The mammalian expression vectors pCMV-5A and pCMV-5B were used to clone the above transcription factors in frame with the myc-tag. These expression vectors

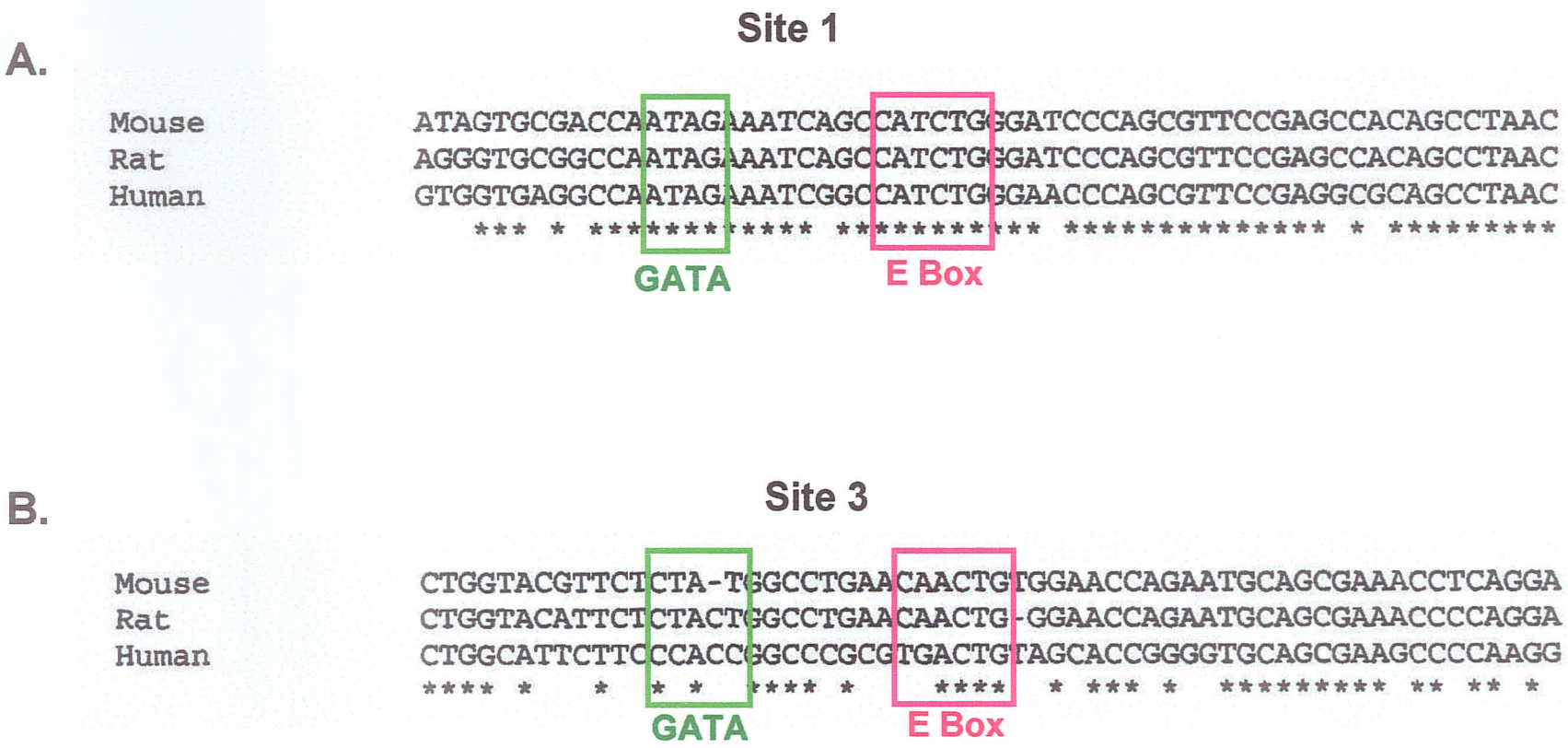


Figure 7: **Clustal-Wallis alignment of CRT promoter between rat, human and mouse.** The sequences for CRT promoter of rat, human and mouse were compared on clustal-wallis database. **A**. The sequence encompassing the first E box (site 1) and neighboring nucleotides shows high conservation of both E box, GATA site and 8bp linker sequence among the three species. **B**. The sequence encompassing the third E box (site 3) and neighboring nucleotides shows low conservation of both E box and the GATA site among the three species

were then transfected into HEK-293 cells and expression of the myc-tagged protein was checked by immunocytochemistry using myc-antibody (Figure 8). The expressed Tal-1-myc, E47-myc and GATA-1-myc proteins co-localized with the DAPI nuclear stain (Figure 8). These three myc-tag expression constructs were next stably transfected into HEK-293 cells and colonies expressing these proteins were isolated. Western analysis with anti-c-myc antibody demonstrated that Tal-myc cell line #7 (Figure 9A, lane 4) and GATA-myc cell line #5 (Figure 9B, lane 5) had the highest expression of the transfected proteins. Screening of the E47-myc clones did not result in any colonies expressing the protein more than the E47-myc stable pool of cells (Figure 9C, lane 2). Therefore, the E47-myc stock as well as Tal-myc cell line #7 and GATA-myc cell line #5 were used for later experiments. The endogenous expression of E47, Tal-1 and GATA-1 were tested on the cell lines used in our study. As seen in Figure 6, NIH-3T3 cells do not express Tal-1 (Figure 6A) and GATA-1 proteins (Figure 6B) endogenously. HEK-293 cells express GATA-1, but do not have endogenous Tal-1 protein (Figure 6A, Figure 6B). However, both NIH-3T3 and HEK-293 cells have endogenous expression of E47 (Figure 6C).

III. Reporter Gene Assays

To test the role of each E box of the CRT promoter in the regulation of the CRT gene, full-length CRT promoter and truncated CRT promoter fragments were cloned upstream of the luciferase reporter gene (see Figure 5). The truncated CRT promoter constructs were generated by sequentially deleting one E box from the 5' or 3' as described in the Materials and Methods.

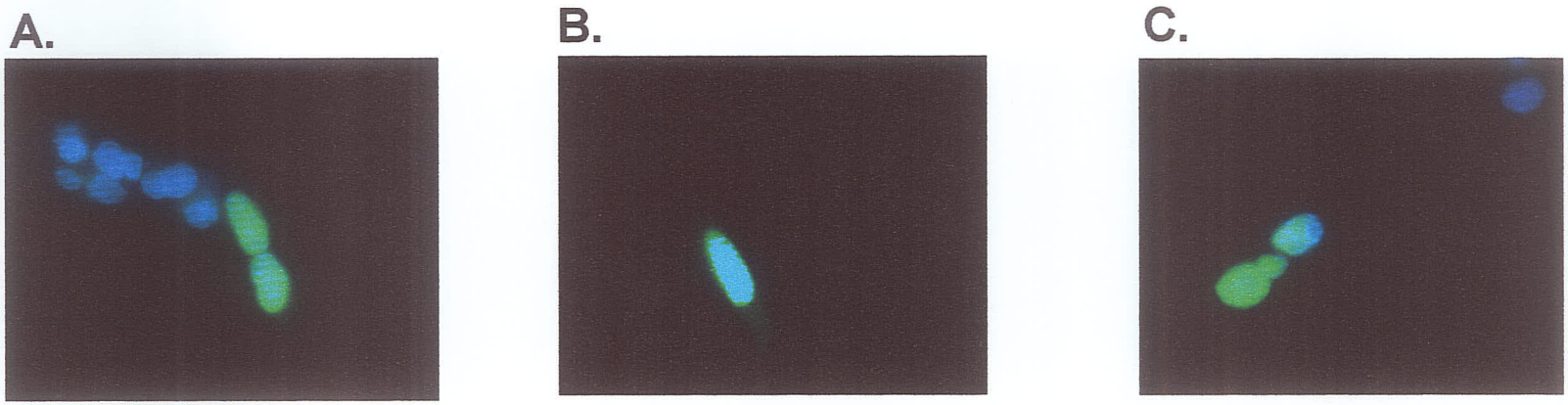


Figure 8: **Immunocytochemical localization of myc-tag proteins in transiently transfected HEK-293 cells.** HEK-293 cells were transfected with expression vectors for Tal-myc (A), GATA-myc (B) or E47-myc (C). Monoclonal antibody to myc (9E10) followed by FITC-labeled secondary was used to detect myc-tag protein expression (green fluorescence). The nuclei was stained with DAPI (blue fluorescence). All myc-tag proteins localize to the nucleus as seen by overlapping patterns of green and blue fluorescence.

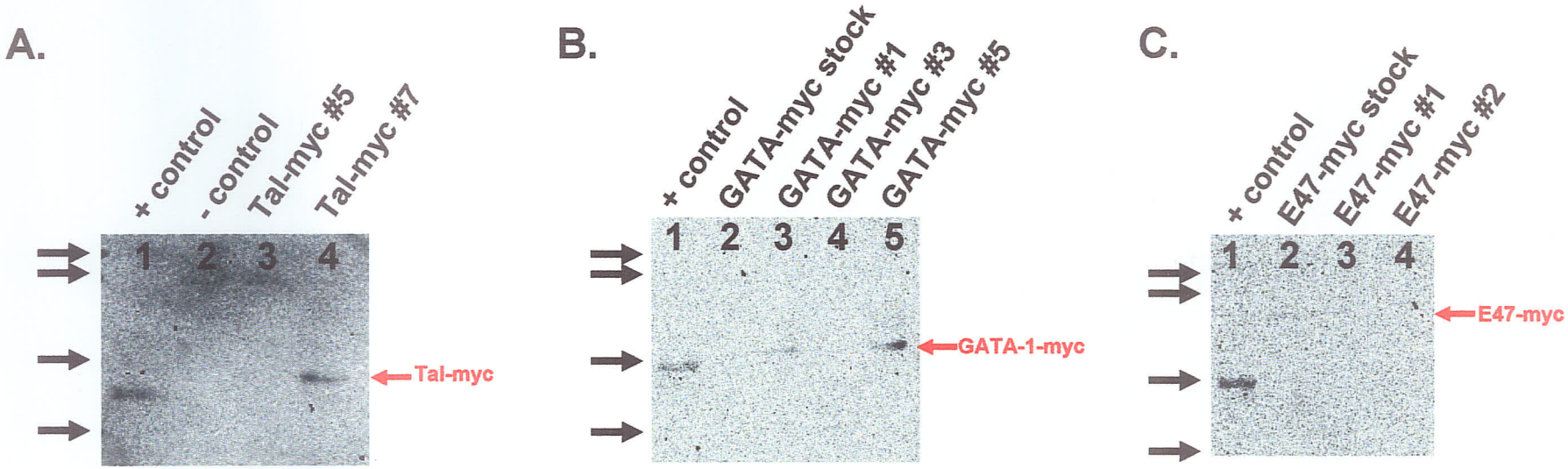


Figure 9: **Western of myc-tag stable HEK-293 cell lines.** HEK-293 cells were stably transfected with one of three different myc-tag constructs: Tal-myc (A), GATA-myc (B) and E47-myc (C). The four arrows to the left of each figure correspond to protein standards with sizes, from top to bottom, 114kDa, 88kDa, 50.7kDa and 35.5kDa. ***A***. The highest Tal-myc overexpressing cell line is #7 (lane 4) ***B***. The highest GATA-myc overexpressing cell line is #5 (lane 5) ***C***. The highest E47-myc overexpressing cell line is the stock (lane 2).

Basal activity of the 7 CRT promoter constructs was tested in HEK-293 cells and NIH-3T3 cells. The full-length promoter (CPF) and the 6 deletion constructs (CP3T, CP2T, CP1T, CPB3T, CPB2T, CPB1T) were transfected individually into HEK-293 cells and NIH-3T3 cells with the β-gal plasmid as a control for transfection efficiency. As seen in Figure 10 (HEK-293 cells) and Figure 11 (NIH-3T3 cells), there is a significant decrease in the CRT promoter activity when the 3' end of the CRT promoter containing site 1 was deleted. Eliminating this region (CPB3T, CPB2T and CPB1T) resulted in a 100-fold reduction in luciferase activity. This indicates the presence of a possible enhancer in this region of the promoter.

1. E47 protein

To study the role of E47 on the CRT promoter, both HEK-293 (Figure 12) and NIH-3T3 cells (data not shown) were transfected with each of the individual CRT-luciferase constructs, β-gal and E47-myc expression vectors. As seen in Figure 12, E47 resulted in higher luciferase activity (with all 7 CRT promoter constructs) in HEK-293 cells versus controls. There was a 1.5 fold increase in the full-length CRT promoter and CP3T activity in the presence of E47 (Figure 12A). Furthermore, E47 resulted in a 2.5 fold increase in the activity of CP3T (deletion of site 1), while there was no significant increase after deletion of site 2 and 3 respectively (Figure 12B). As seen in Figure 12, these 3 constructs (CPF, CP3T and CPB3T) which contain the 3rd putative binding site showed significant increase in activity by E47 which suggests the importance of this site for E47 function. Interestingly, in the NIH-3T3 cells, E47 induced 1-fold increase in the CPF activity; however these differences were not significant (data not shown). The

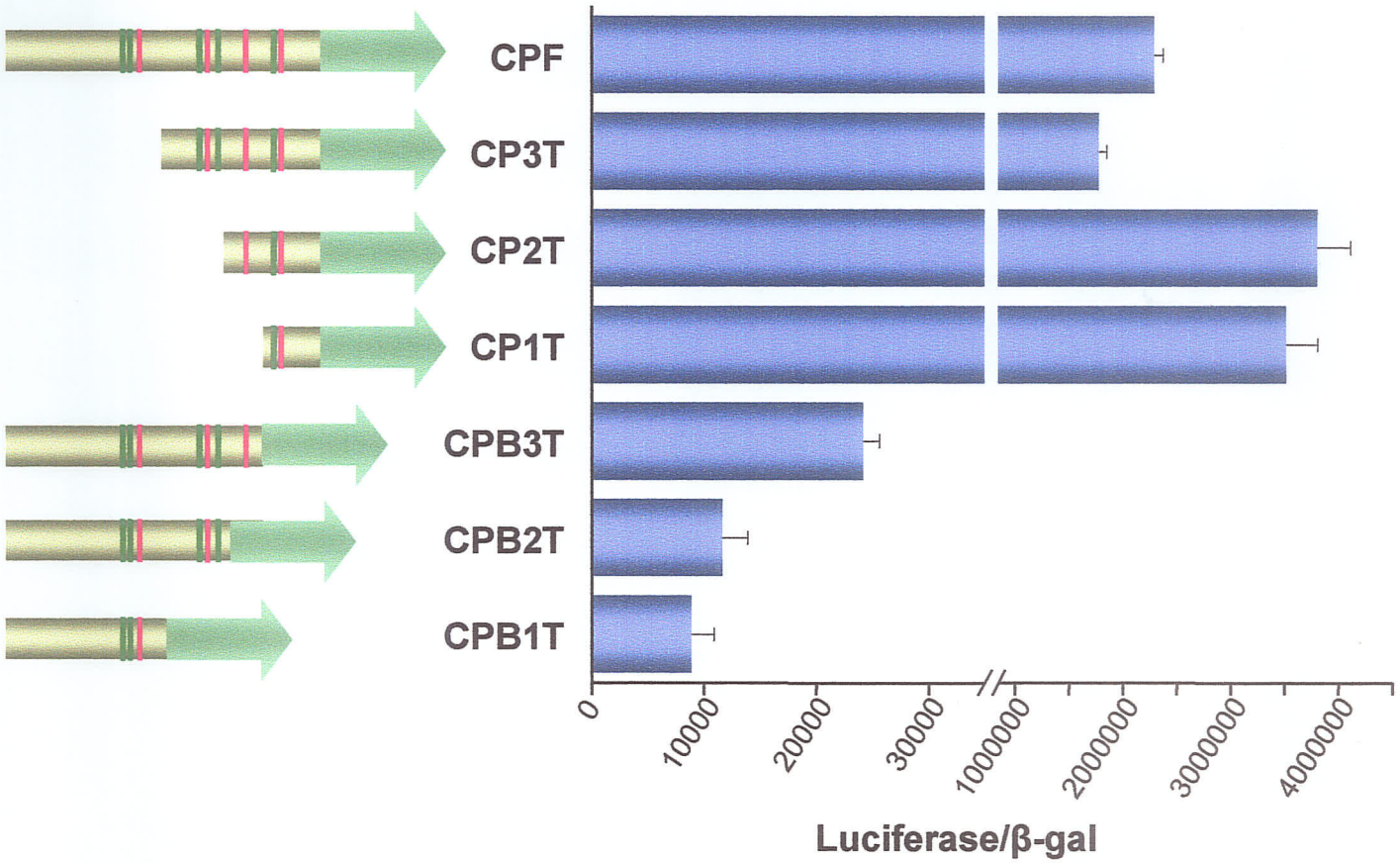


Figure 10: **Basal CRT promoter activity in HEK-293 cells.** HEK-293 cells were cotransfected with one of the 7 CRT promoter constructs and β-gal plasmid. There is 100-fold decrease in activity of the CRT promoter with the elimination of the first sequence encompassing putative E box site 1. Values are mean ± SE of 5 experiments carried out in triplicate.

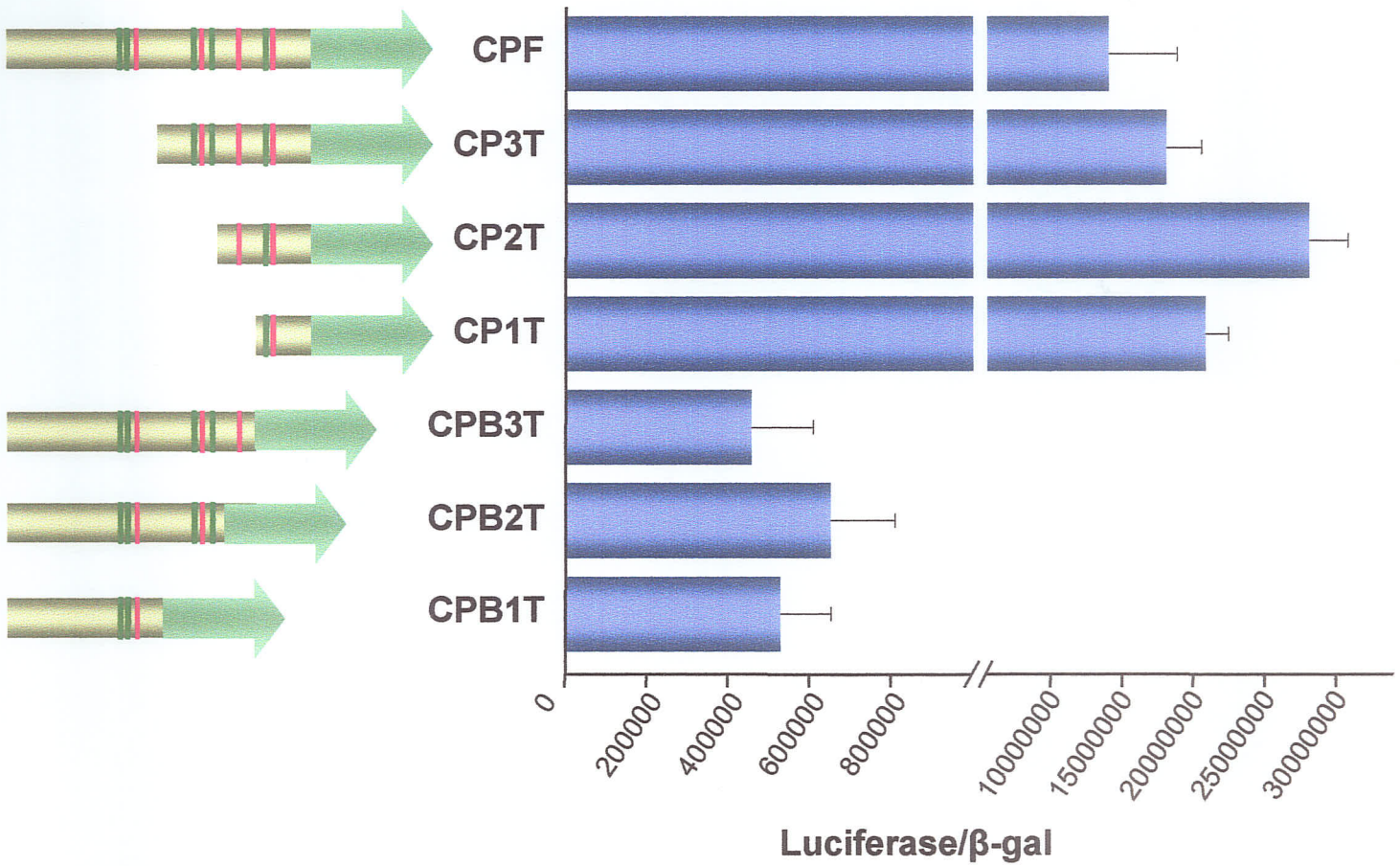


Figure 11: **Basal CRT promoter activity in NIH-3T3 cells.** NIH-3T3 cells were cotransfected with one of the 7 CRT promoter constructs and β-gal plasmid. There is 100-fold decrease in activity of the CRT promoter with the elimination of the first sequence encompassing putative E box site 1. Values are mean ± SE of 5 experiments carried out in triplicate.

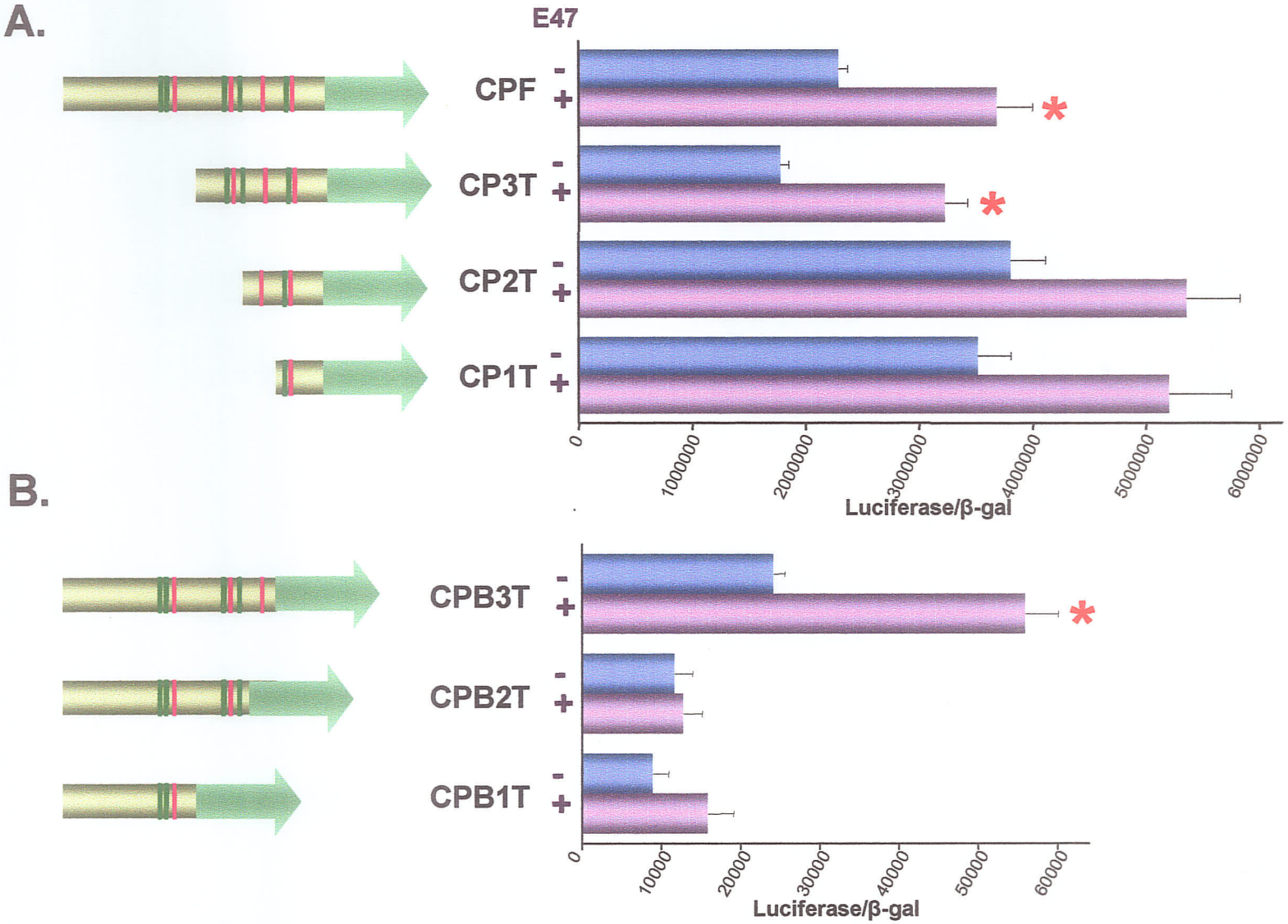


Figure 12: **Effect of E47 on the activity of different constructs of CRT promoter.** HEK-293 cells were cotransfected with one of the 7 CRT promoter constructs (schematic diagram on the left) and β-gal plasmid. Each sample was cotransfected with either an empty plasmid (control, purple bars) or an E47 expression plasmid (pink bars). Luciferase values were normalized with β-galactosidase. There was significant increase in the activity of the CRT promoter in CPF, CP3T and CPB3T with the addition of E47. Values are mean ± SE of 4 experiments carried out in triplicate. * $P<0.05$, significantly different from control (empty plasmid).

difference in E47 function in the different cell types could be due to the presence or absence of different co-factors that interfere or aid in the binding of E47 to the CRT promoter. As shown in Figure 6, HEK-293 expresses GATA-1 endogenously while NIH-3T3 does not express GATA-1. The ability of E47 to act on the CRT promoter in HEK-293 cells therefore could be due to this factor or alternatively the NIH-3T3 cells might have a higher endogenous expression of E47, thus the activation of CRT promoter is maximal without the exongenously overexpressed E47 protein.

2. Tal-1 protein

As discussed earlier, E47 can form heterodimers with other bHLH proteins such as Tal-1. Therefore, we tested a possible role for Tal-1 in forming heterodimers with E47 to regulate CRT gene expression. The Tal-1 expression vector (Tal-1-myc) was transfected into HEK-293 cells and NIH-3T3 cells along with the β-gal plasmid and one of the seven CRT promoter constructs individually. Expression of Tal-1 in HEK-293 cells did not alter CRT promoter activity (Figure 13). Contrarily, the addition of Tal-1 to NIH-3T3 cells increased the activity in most of the constructs (Figure 14). The 5' deletion constructs showed the most significant difference with a 1.5-fold increase in full-length CRT promoter (CPF) activity and 1.6-fold increase in the 5' deletion construct CP1T (Figure 14). The 3' deletion constructs do not have significant differences with the addition of Tal-1 (Figure 14). Thus the elimination of the site 1 (CPB3T, CPB2T, CPB1T) indicates the importance of this region in mediating Tal-1 activity. The differential effects seen between HEK-293 cells and NIH-3T3 cells could be due to

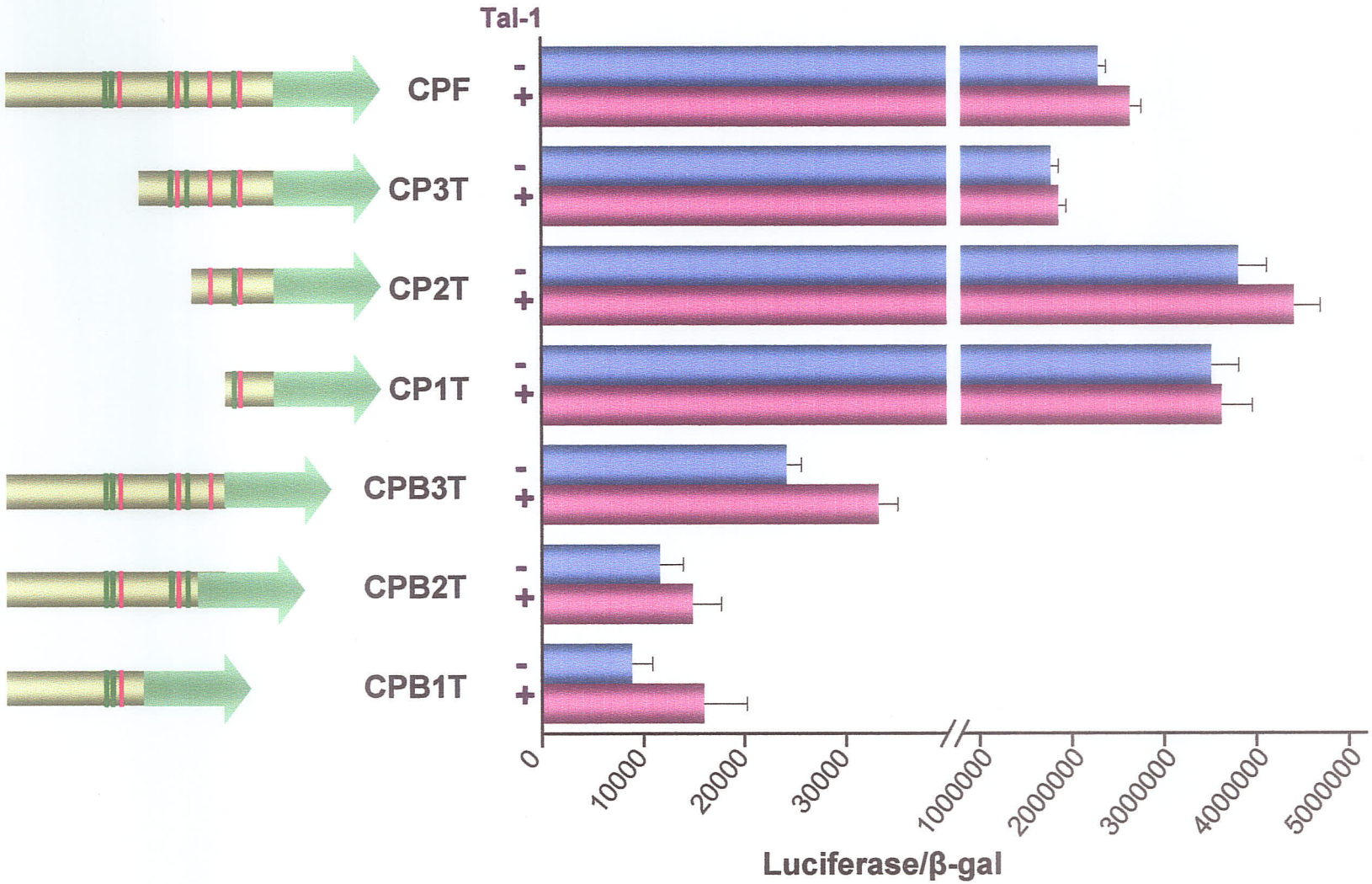


Figure 13: **Effect of Tal-1 on the activity of different constructs of CRT promoter.** HEK-293 cells (purple bars) or HEK-293 Tal-myc stable cells (pink bars) were cotransfected with one of the 7 CRT promoter constructs (schematic diagram on the left) and β-gal plasmid. Luciferase values were normalized with β-galactosidase. Values are mean ± SE of 4 experiments carried out in triplicate.

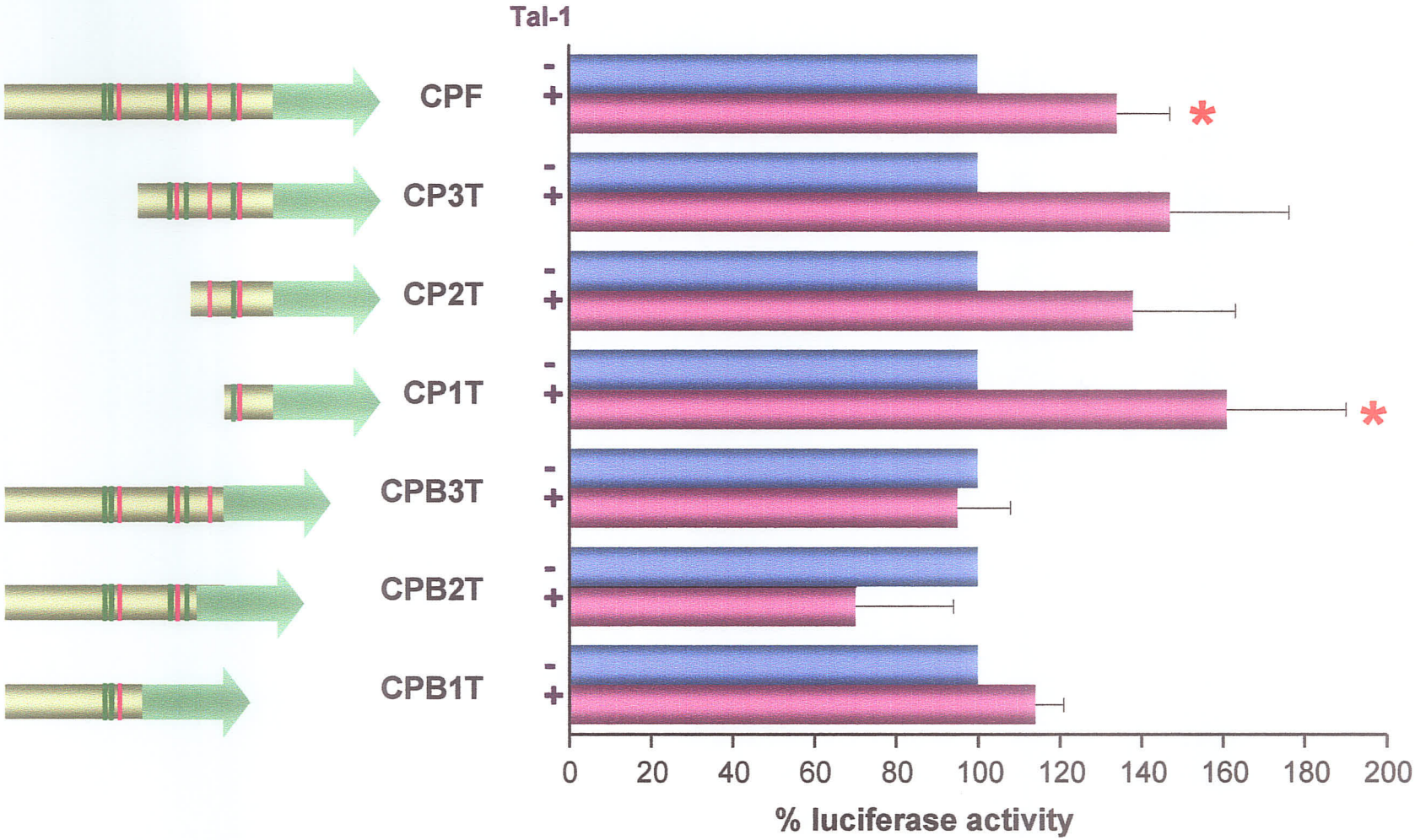


Figure 14: **Effect of Tal-1 on the activity of different constructs of CRT promoter.** NIH-3T3 cells were cotransfected with one of the 7 CRT promoter constructs (schematic diagram on the left) and β-gal plasmid. Each sample was cotransfected with either an empty plasmid (purple bars) or a Tal-1 expression plasmid (pink bars). Luciferase values were normalized with β-galactosidase. Values are mean ± SE of 6 experiments carried out in triplicate. * $P<0.05$, significantly different from control (empty plasmid).

endogenous expression of other factors such as GATA-1 (see Figure 6B). Thus, there could be maximal expression of the vector without the addition of other proteins.

3. GATA-1

Since GATA-1 has been previously shown to interact with Tal-1 and E47, it was of interest to study GATA-1 in the regulation of the CRT promoter. The expression of GATA-1 (GATA-1-myc) increased the activity of three of the four CRT promoter constructs tested (CPF, CP2T and CP1T – see Figure 15). The significant decrease in the activity of the 5' deletion construct CP3T could be due to GATA-1 binding to other accessory proteins and repressing the expression of the luciferase gene. Interestingly, when GATA-1 and Tal-1 expression vectors were transfected together in NIH-3T3 cells, there was a synergistic activation of the CRT promoter (Figure 15). There was a 3.5-fold in the case of the full-length CRT promoter (CPF) and a 2.5-fold increase in the case of the CP2T promoter (Figure 15). These differences were significantly different from the control cells. This indicates that the CRT promoter is regulated by a cooperation between GATA-1 and Tal-1 proteins.

4. Mutated 3rd E box

Analysis of the sequence of the 3rd E box and its flanking region revealed that it was identical to the published sequence which was important in formation of complexes between the bHLH and GATA-1 transcription factors (Table I) [94]. Thus, to test the importance of this site in the regulation of the CRT promoter, this E box was mutated from CAACTG to AAACAG in the full-length promoter (CPF-M) and the 5' deletion

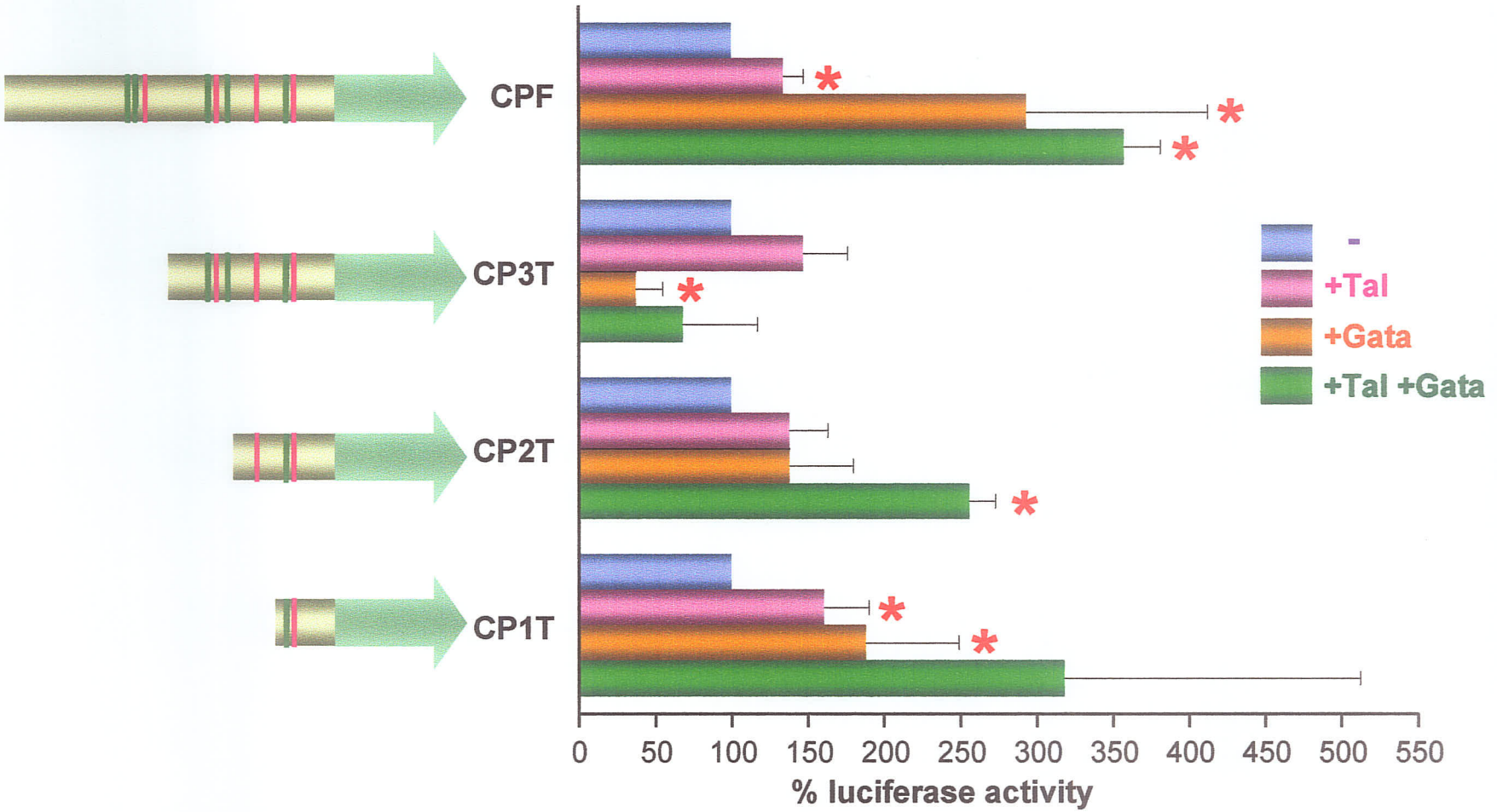


***Figure 15:* Effect of Tal-1 and GATA-1 on the activity of CRT promoter.** NIH-3T3 cells were cotransfected with a luciferase reporter construct containing either a full-length CRT promoter construct (CPF) or a forward deletion construct (CP3T, CP2T, CP1T, refer to schematic diagram to the left of the bar graph). Each construct was transfected with either 2µg of empty plasmid (purple bars), 1µg of Tal-1-myc plasmid (pink bars), 1µg of GATA-1-myc plasmid (orange bars) or both 1µg Tal-1-myc and 1µg GATA-1-myc expression plasmids (green bars). Luciferase values were normalized with β-galactosidase. Tal-1 alone or in combination with GATA-1 enhances the activity of all constructs. Whereas GATA-1 has stimulatory effect on most constructs except the CP3T where it significantly inhibits the luciferase activity. Values are mean ± SE of 4 experiments carried out in triplicate. * $P<0.05$, significantly different from control (empty plasmid).

construct CP3T (CP3T-M). The wild-type CRT promoter sequence and the mutated CRT promoter sequence were transfected individually into HEK-293 cells with β-gal for control of transfection efficiency. There was a 2-fold increase in the activity of full-length CRT promoter with the mutation of the 3rd site (Figure 16 – CPF). This indicates that the 3rd site E box is a repressory region, because when mutated, the activity of the promoter increases. Contrarily, there was a decrease in activity of the CP3T promoter constructs with the mutation of the 3rd site E box (Figure 16 – CP3T). This indicates that the 3rd site E box has variable importance depending on the presence of the 4th site. If the 4th site is intact, the repression by the 3rd site E box does not take place. Mutating the 3rd E box in the absence of the 4th E box indicates that this site is not as important in the repression as another possible binding site.

IV. DNA-transcription factor interaction

The reporter gene assays with GATA-1, Tal-1 and E47 showed that these factors can regulate the CRT gene. However, no data is available on the interaction of these proteins individually or in a complex with the CRT promoter. We utilized Electrophoretic Mobility Shift Assays (EMSA) to determine if these factors bind to the putative E box and GATA-1 sites in the CRT promoter. Analysis of the CRT promoter showed that the E box and GATA site at site 1 and 3 are similar to that seen by Wadman et al. 1997, which allows the formation of a complex between GATA and bHLH proteins when their binding sites are separated by 8-11bp [94]. We have shown (Figures 12, 14 and 15) that these two sites are important in the regulation of the CRT promoter by GATA-1 and Tal-1. Therefore, 32 nucleotides encompassing the E box and GATA site of either site 1 or

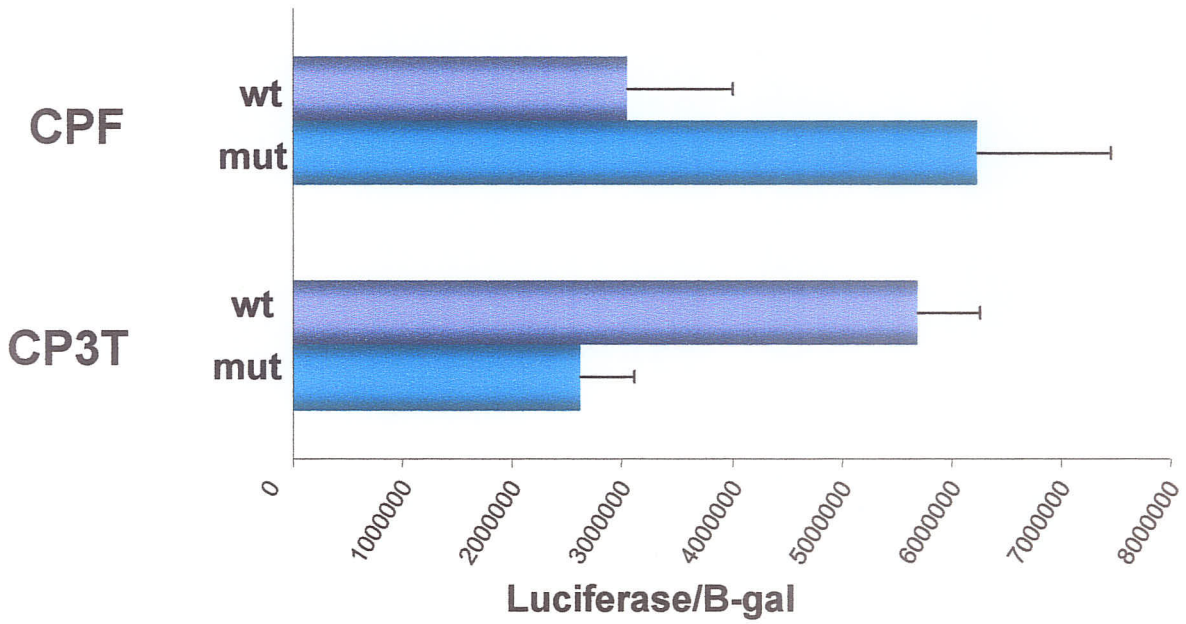


Figure 16: **Effect of mutated 3rd E box on the expression of the CRT promoter.** HEK-293 cells were transfected with the wild-type promoter constructs (wt: CPF or CP3T) or the mutated 3rd E box promoter constructs (mut: CPF-M or CP3-M) as described in the Materials and Methods. Values are mean ± SE of 5 experiments carried out in triplicate.

site 3 were used as probes in the EMSA experiments (TLG-1 and TLG-3 respectively). To differentiate between the role of the E box and GATA binding sites in binding of transcription factors, mutated probes containing either a mutated E box (TLG1-M, TLG3-M), a mutated GATA site (TLG1-GM, TLG3-GM) or both mutated E box and GATA site (TLG1-MM, TLG3-MM) were also used in the EMSA experiments. Figure 17 shows a schematic presentation of the two different wild type probes used in the EMSA experiments as well as the single and double mutant versions of these probes.

1. GATA-1 protein binds to both site 1 and site 3 of the CRT promoter

To examine the binding of GATA-1 to the CRT promoter, nuclear extracts were prepared from HEK-293 cells overexpressing GATA-myc protein. Figure 18 shows a representative EMSA with site 1 (TLG1) of the CRT promoter. As seen in Figure 18B (lane 4) and Figure 18B (lane 1), the TLG-1 resulted in mobility shifts of 4 bands, emphasizing the formation of different complexes of transcription factors. The top band appeared in all lanes indicating that it was not specific. In contrast, the lower three bands were competed when excess unlabeled TLG-1 was used in the EMSA reaction (Figure 18A lane 5, Figure 18B lane 2). To test the specificity of the GATA-1 bands, a supershift assay was carried out with myc and GATA-1 antibodies. Both GATA-1 (Figure 18A - lane 7 and 8, and inset) and myc (Figure 18B - lane 3) antibodies resulted in supershift of the GATA-1 band (red arrow supershift to orange arrow), whereas a non-specific antibody (Figure 18A - lane 6) did not supershift this band. Interestingly, there is a dose-dependency in the ability to supershift of the complex with the GATA-1 antibody. As seen in Figure 18A – lane 7 and 8, increasing the amount of GATA-1 antibody by 1.5-

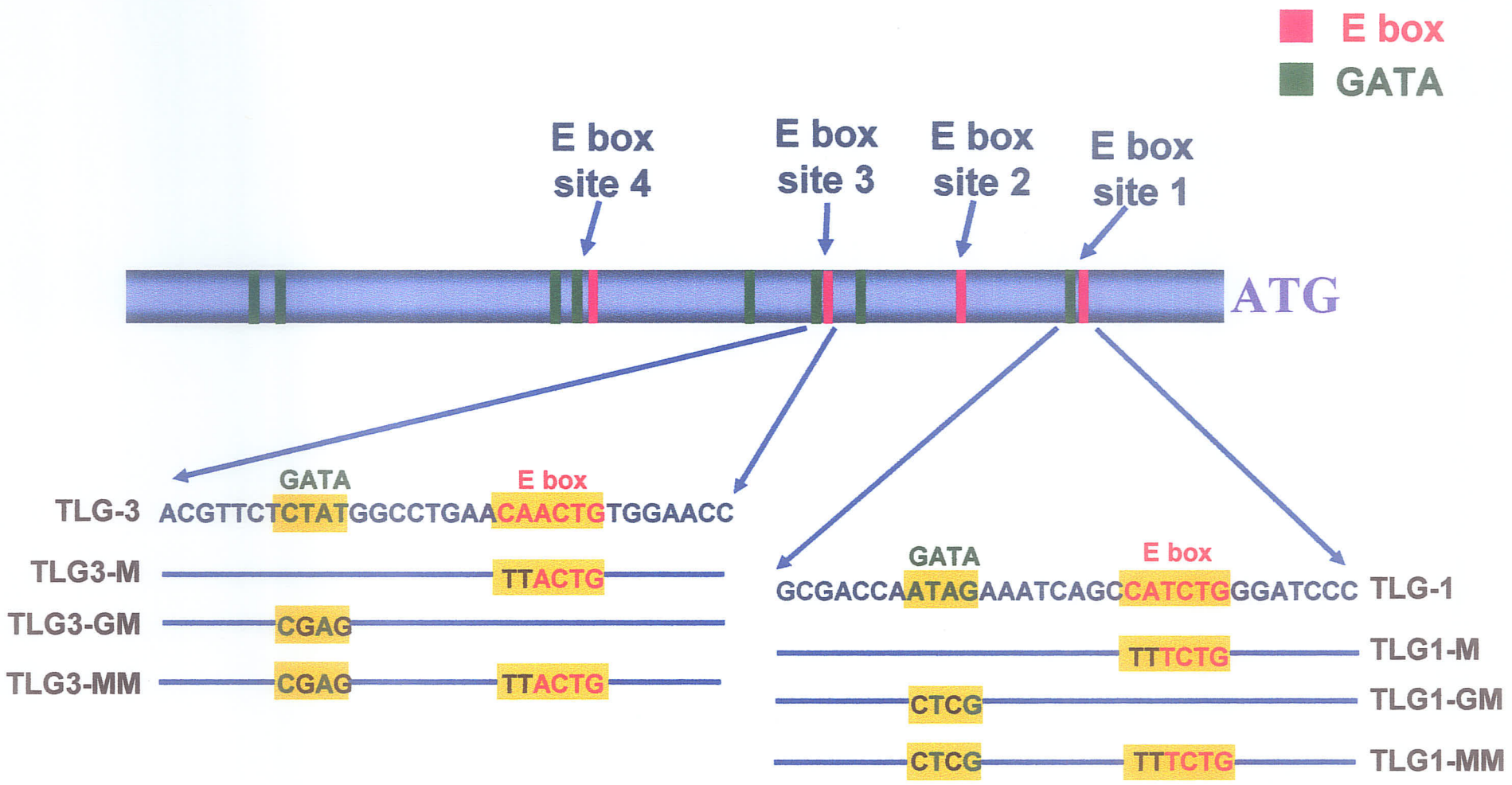


Figure 17: **Schematic drawing of the CRT promoter and sequences of two probes and their mutated forms used in EMSA.** GATA binding sites (green) are separated by 8 nucleotides from the Tal-1/E47 binding sites (pink).

fold (lane 8) resulted in a stronger intensity of the supershifted band indicating more of the complex was shifted (see inset). To further test the specificity of this band to GATA-1, 100-fold excess unlabeled TLG1-GM probe (TLG1 mutated in GATA site) was used to compete for E box binding. As shown in Figure 18A - lane 9, this mutant probe did not compete with the TLG1 probe binding to the GATA1 band (red arrow). However, other bands were competed out (Figure 18A - lane 9, see star). To determine if the E box plays any role in GATA-1 binding to the TLG1, a probe with mutated E box (TLG1-M) and a probe with mutated E box and GATA site (TLG1-MM) were generated. As seen in Figure 18A – lane 2, mutation of the E box (TLG1-M) resulted in the appearance of an extra band with fast mobility (see arrowhead). Interestingly, this band was not supershifted with the GATA-1 antibody, whereas the GATA-1 band did supershift in this lane (Figure 18A – lane 3). Mutation in both the GATA-1 and E box site of TLG-1 (TLG1-MM) resulted in loss of this newly observed band (arrowhead) as well as another second band (double arrowhead, Figure 18A – lane 1).

To test whether GATA-1 binds to site 3 of the CRT promoter, nuclear extracts of HEK-293 overexpressing GATA-1-myc protein were used with probes for site 3 (TLG-3). This EMSA shows 4 bands shifted (Figure 19 – lane 1). Similar to the TLG-3 probe, the top band is non-specific and appears in all lanes, whereas the other three bands were competed by adding 1000-fold excess of unlabeled TLG-3 (Figure 19 – lane 2). To test the specificity of the bands, GATA-1 antibody was used for supershifting (Figure 19 – lane 3). Two bands corresponding to GATA-1 protein complexes are labeled with a red arrow and were supershifted by the GATA-1 antibody (Figure 19). To find out whether the GATA site is involved in the binding of these complexes, a probe with a mutated

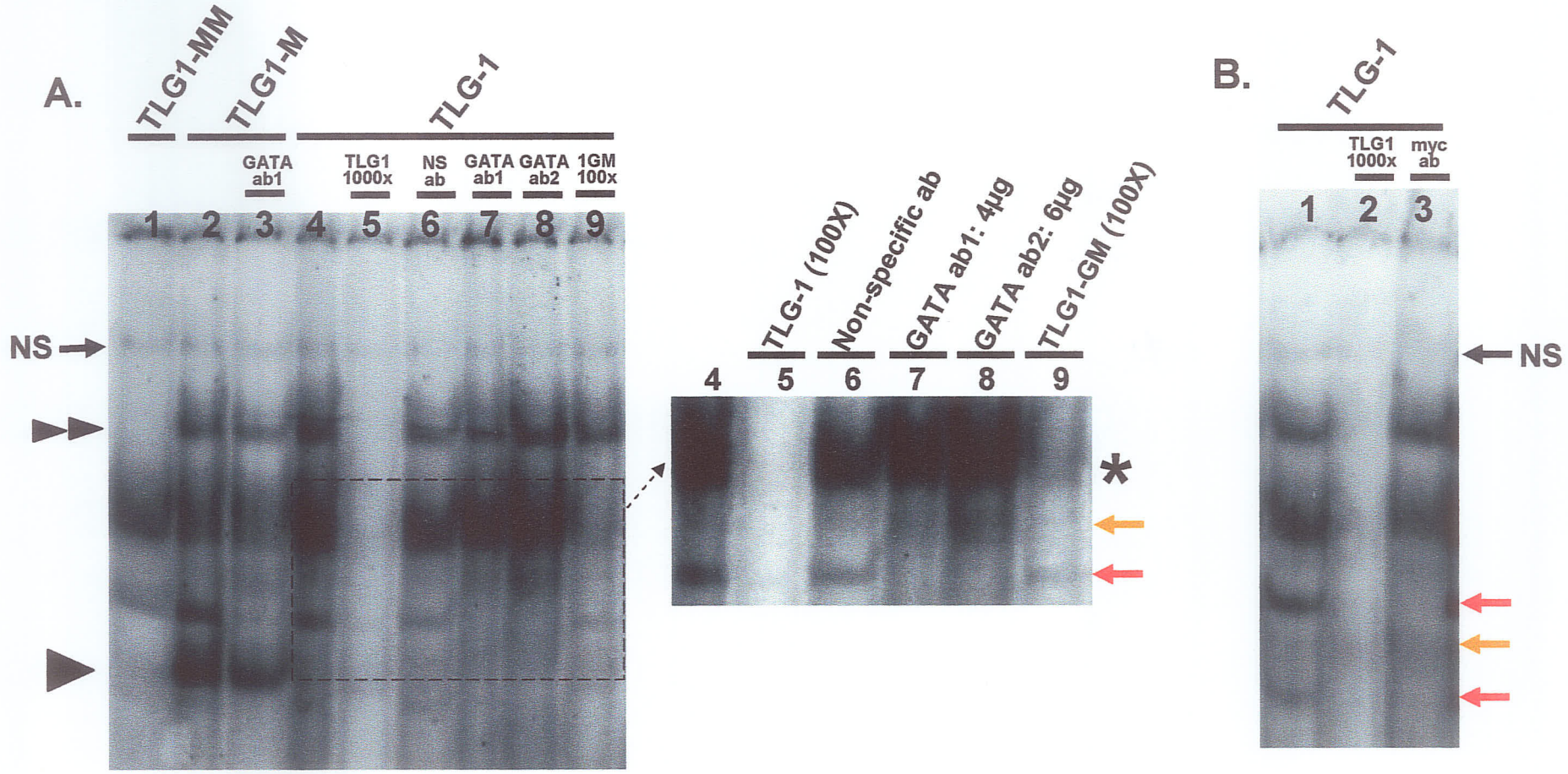


***Figure 18:* GATA-1 binds to site 1 on CRT promoter.** EMSA was performed using nuclear extracts from HEK-293 cells overexpressing GATA-1 myc protein. Synthetic oligos (32bp) representing the wild-type CRT promoter sequence of site 1 (TLG-1), or mutated CRT promoter sequences (TLG1-M and TLG1-MM) were labeled with ^{32}P. 1000-fold excess unlabeled TLG-1 oligo (A - lane 5 and B - lane 2) and 100-fold excess TLG1-GM (A - lane 9) were used for competition studies. Mutating the GATA site (TLG1-GM) resulted in competition for a low mobility band (see star, inset). Antibodies to GATA-1 (either 4µg – lane 7 or 6µg – lane 8) and myc (B – lane 3) were used for supershift assays. The GATA-1 protein is marked with a red arrow and the supershifted complexes are labeled with an orange arrow. Mutation of E box (TLG1-M) resulted in shifting of an extra band (arrowhead). Mutation of E box and GATA site (TLG1-MM) resulted in the absence of the high-mobility band (arrowhead) and a low mobility band (double arrowhead). (NS = non-specific, NS ab = non-specific antibody, GATAab1 = 4µg of GATA-1 antibody, GATA ab2 = 6µg of GATA-1 antibody, 1GM 100X = competition with 100-fold excess TLG1-GM)

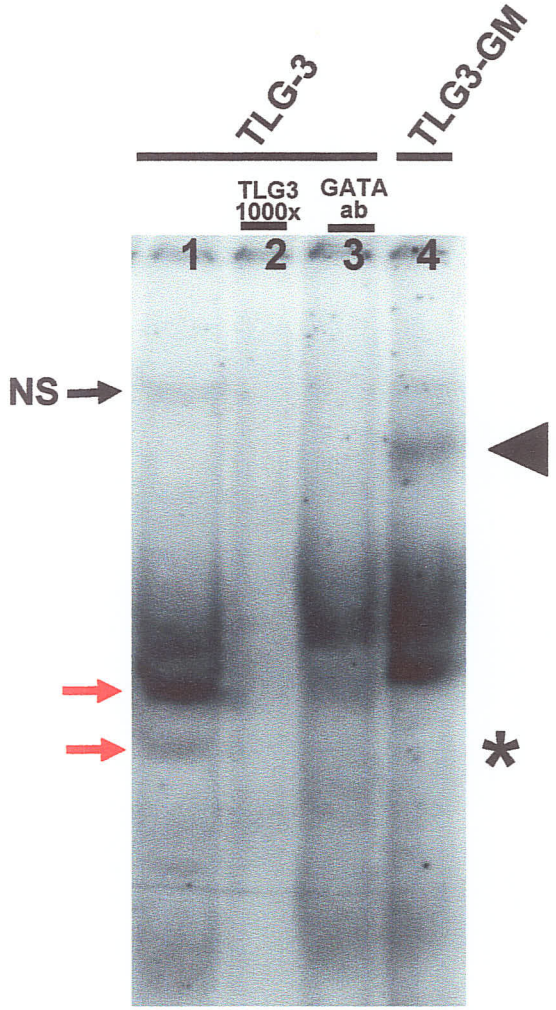


***Figure 19:* GATA-1 binds to site 3 of the CRT promoter.** EMSA was performed with nuclear extracts of HEK-293 overexpressing GATA-1 myc protein. Synthetic oligos (32bp) representing wild-type CRT promoter sequence on site 3 (TLG-3) or mutated GATA-1 site 3 of the CRT promoter sequence (TLG3-GM) were labeled with ^{32}P. 1000-fold excess unlabeled TLG-3 oligo (lane 2) was used in competition studies. Antibodies to GATA-1 were used for supershift assays (lane 3). The GATA-1 protein is marked with a red arrow. The mutation of the GATA site (TLG3-GM) generates an extra band (arrowhead), but does not produce the GATA-1 corresponding band (star). (NS = non-specific)

GATA site was generated (TLG3-GM). Mutation of this GATA site resulted in the loss of the high mobility supershifted band (Figure 19 – lane 4, see star). Interestingly, there is an extra band with low mobility generated by this mutation of the GATA site (Figure 19, see arrowhead).

2. Tal-1 protein binds to both site 1 and site 3 of the CRT promoter

Since Tal-1 has been shown to activate the CRT promoter alone or in combination with GATA-1, it is of interest to look at Tal-1 binding to either site 1 or site3. Nuclear extracts of the HEK-293 overexpressing Tal-1-myc protein were used in these experiments. Figure 20 represents an EMSA with site 1 of the CRT promoter and Figure 21 represents an EMSA with site 3. The EMSA with TLG-1 shows the presence of 5 bands (Figure 20 – lane 2). The top band was proven to be non-specific because it was not competed by 1000-fold excess of unlabeled TLG-1 in contrast to the lower bands (Figure 20 – lane 3). These bands were tested for further specificity by supershift with myc and Tal-1 antibodies. The myc antibody resulted in supershift of the Tal-1 bands labeled with red arrows (Figure 20 – lane 4). The Tal-1 antibody resulted in a decrease in the intensity of the Tal-1 band and appearance of a shifted band, but the concentration of the antibody was not sufficient enough to produce a complete supershift of the Tal-1 band (Figure 20 – lane 5). To determine if the E box plays any role in Tal-1 binding to the TLG1 oligo, a probe with a mutated E box (TLG1-M) was generated. Surprisingly, the mutation of this E box did not affect the band shifts (Figure 20 – lane 1).

Analysis of site 3 revealed differences in binding. The EMSA with TLG-3 shows the presence of 5 bands (Figure 21 – lane 2). Similar to TLG-1, the top band as well as

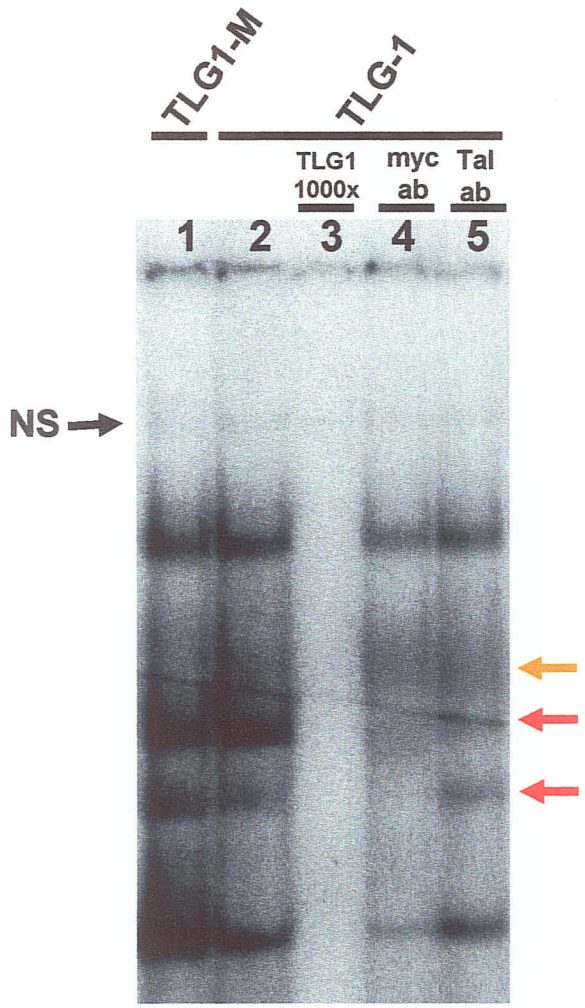


Figure 20: **Tal-1 binds to site 1 of the CRT promoter.** EMSA was performed with nuclear extracts of HEK-293 overexpressing Tal-1 myc protein. Synthetic oligos (32bp) representing wild-type CRT promoter sequence on site 1 (TLG-1) or mutated CRT promoter sequence (TLG1-M) were labeled with ^{32}P. Addition of 1000-fold excess unlabeled TLG-1 oligo (lane 3) was used for competition studies. Antibodies to myc (lane 4) and Tal-1 (lane 5) were used for supershift assays. The Tal-1 protein is marked by red arrows and the supershifted complex is labeled by an orange arrow. The mutation of the E box (TLG1-M) does not affect the mobility of the bands (lane 1). (NS = non-specific)

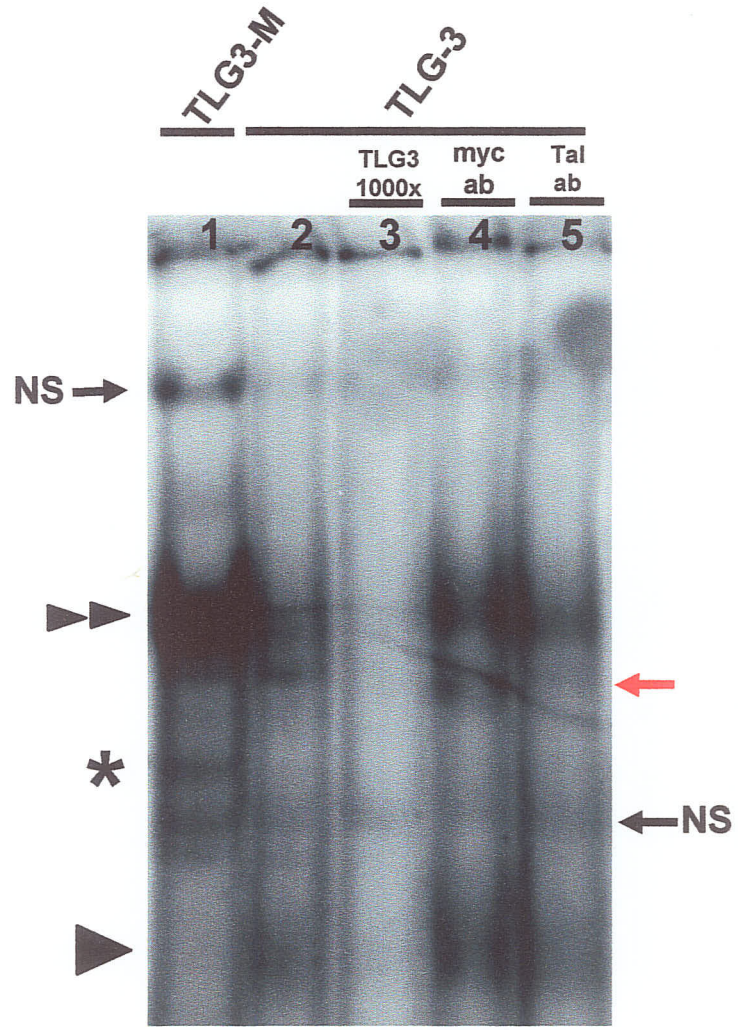


Figure 21: **Tal-1 binds to site 3 of CRT promoter.** EMSA was performed with nuclear extracts of HEK-293 overexpressing Tal-1 myc protein. Synthetic oligos (32bp) representing wild-type CRT promoter sequence on site 3 (TLG-3) or mutated CRT promoter sequence (TLG3-M) were labeled with ^{32}P. Addition of 1000-fold excess unlabeled TLG-3 oligo (lane 3) was used for competition studies. Antibodies to myc (lane 4) and Tal-1 (lane 5) were used for supershift assays. The Tal-1 protein is marked by a red arrow. The mutation of the E box (TLG1-M) generates an extra band shift (star), as well as a more intense low-mobility band (double arrowhead). The E box mutation does not produce a high mobility band (arrowhead). (NS = non-specific)

the second fast-mobility band appears in all the lanes indicating that they were non-specific. The Tal-1 protein is labeled with a red arrow. To test the specificity of this band, myc and Tal-1 antibodies were used for supershifting. The band labeled with the red arrow was supershifted by the Tal-1 antibody (Figure 21 – lane 5). On the other hand, the myc antibody was not able to supershift any bands (Figure 21 – lane 4). To find out whether the E box is involved in the binding of this complex, a probe with a mutated E box site was used (TLG3-M). Mutation of this E box resulted in the loss of the high mobility supershifted band (Figure 21 – lane 1, see arrowhead). Interestingly, there is an extra band with lower mobility generated by the mutation of the E box (Figure 21 – lane 1, see star). There appears to be more proteins binding to TLG1-M by the presence of a more intense low-mobility band (Figure 21 – lane 1, see double-arrowhead).

3. E47 protein binds to both site 1 and site 3 of the CRT promoter

To examine the binding of E47 to site 1 and site 3, nuclear extracts of HEK-293 overexpressing the E47-myc protein were used in these experiments. Figure 22 represents an EMSA with site 1 of the CRT promoter and Figure 23 represents an EMSA with site 3. The EMSA with TLG-1 resulted in the mobility shifts of four bands (Figure 22 – lane 3). The top band appeared in all lanes indicating its non-specificity. The lower three bands were competed when 1000-fold excess unlabeled TLG-1 was used in the EMSA reaction (Figure 22 - lane 4). The location of The E47 protein is labeled by a red arrow (Figure 22 – inset). The specificity of this band was tested by supershift with an unspecific control antibody and a specific myc antibody. The non-specific antibody, as expected, did not supershift any bands (Figure 22 - lane 5). On the other hand, myc

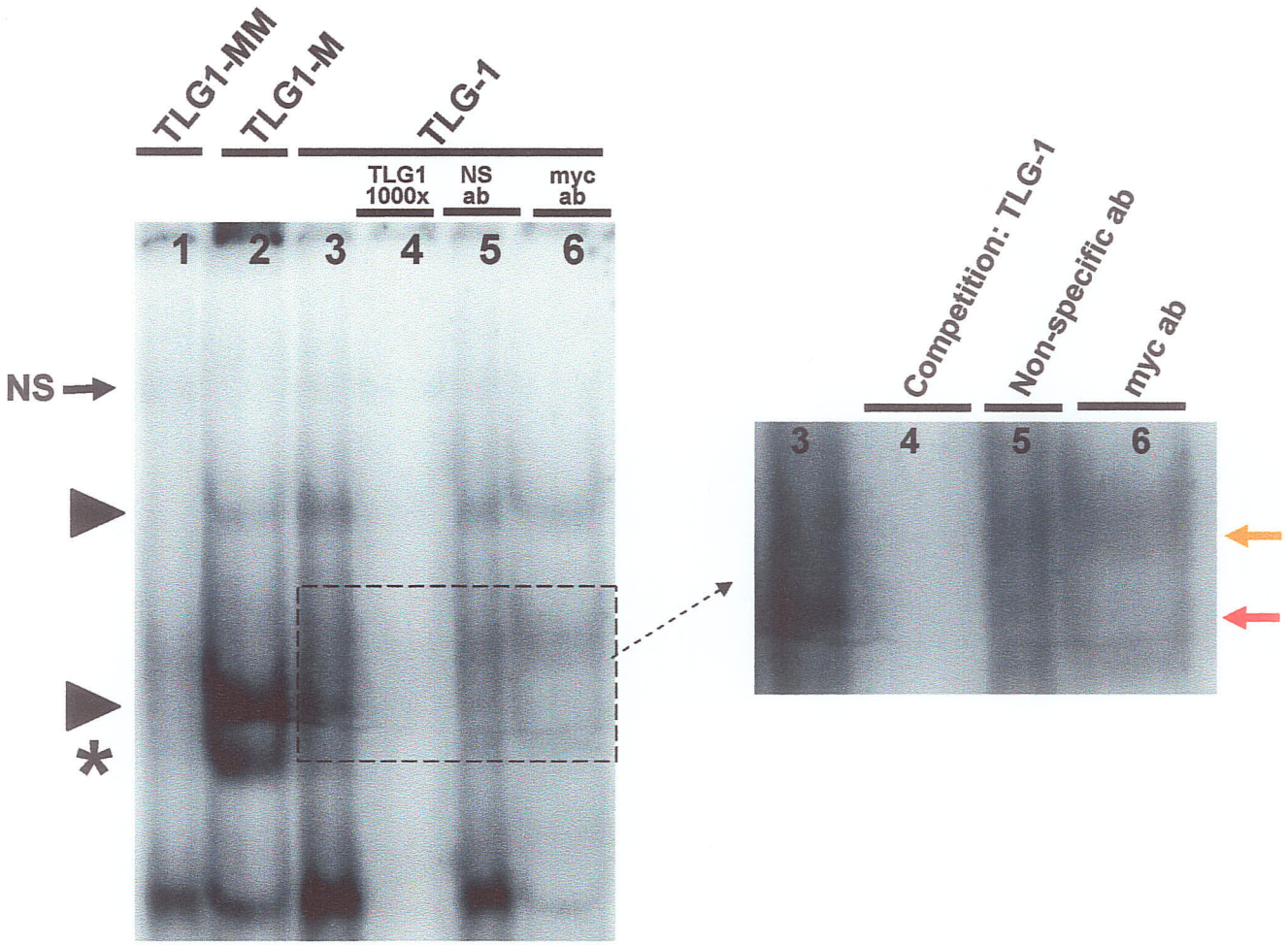


***Figure 22:* E47 binds to site 1 of CRT promoter.** EMSA was performed using nuclear extracts from HEK-293 cells overexpressing E47 myc protein. Synthetic oligos (32bp) representing the wild-type CRT promoter sequence of site 1 (TLG-1) or mutated CRT promoter sequences (TLG1-M and TLG1-MM) were labeled with ^{32}P. 1000-fold excess unlabeled TLG-1 oligo (lane 4) was used for competition studies. Antibodies to myc (lane 6) and an unspecific antibody (lane 5) were used for supershift assays. The E47 protein is marked with a red arrow and the supershifted complex is labeled with an orange arrow (see inset). Mutation of E box (TLG1-M) resulted in shifting of an extra band (lane 2 - star). Mutation of both E box and GATA site (TLG1-MM) did not generate 2 bandshifts seen with wild-type TLG-1 (lane 1 - arrowheads). (NS = non-specific)

antibody resulted in the supershift of the Tal-1 band (Figure 22 – lane 6, see orange arrow in inset). To determine if the E box plays any role in E47 binding to the TLG1, two probes were generated with either a mutated E box (TLG1-M) or a mutated E box and GATA site (TLG1-MM). Mutation of the E box (TLG1-M) resulted in shifting of an extra band with faster mobility (Figure 22 – lane 2, see star). Mutation in both the GATA-1 and E box site of TLG-1 (TLG1-MM) resulted in loss of this newly observed band as well as the Tal-1 band and an additional lower mobility band (Figure 22 – lane 1, see arrowheads and star).

The binding of E47 on site 3 of the CRT promoter was analyzed by using nuclear extracts of HEK-293 overexpressing E47-myc protein. This EMSA shows five mobility shifted bands (Figure 23 – lane 3). Similar to the TLG-1 probe, the top band appears in all lanes and indicates non-specificity. Adding a 1000-fold excess of unlabeled TLG-3 oligo competed the four lower bands (Figure 23 – lane 4). A non-specific antibody and a myc antibody were used to detect specific band shifts. As expected, the non-specific antibody did not change the mobility of any of the bands (Figure 23 – lane 5). The myc antibody shifted the mobility of the highest mobility band which is marked by a red arrow (Figure 23 – lane 6, see inset). The exact binding site for E47 was tested by using two different probes, one with an E box mutation (TLG3-M) and one with both an E box and GATA site mutation (TLG3-MM). Mutation of the E box did not have any differential effects to that seen in wild-type probe (Figure 23 – lane 2). However, the mutation of both E box and GATA site revealed the disappearance of the high mobility band corresponding to the E47 protein (Figure 23 – lane 1, see arrowhead).

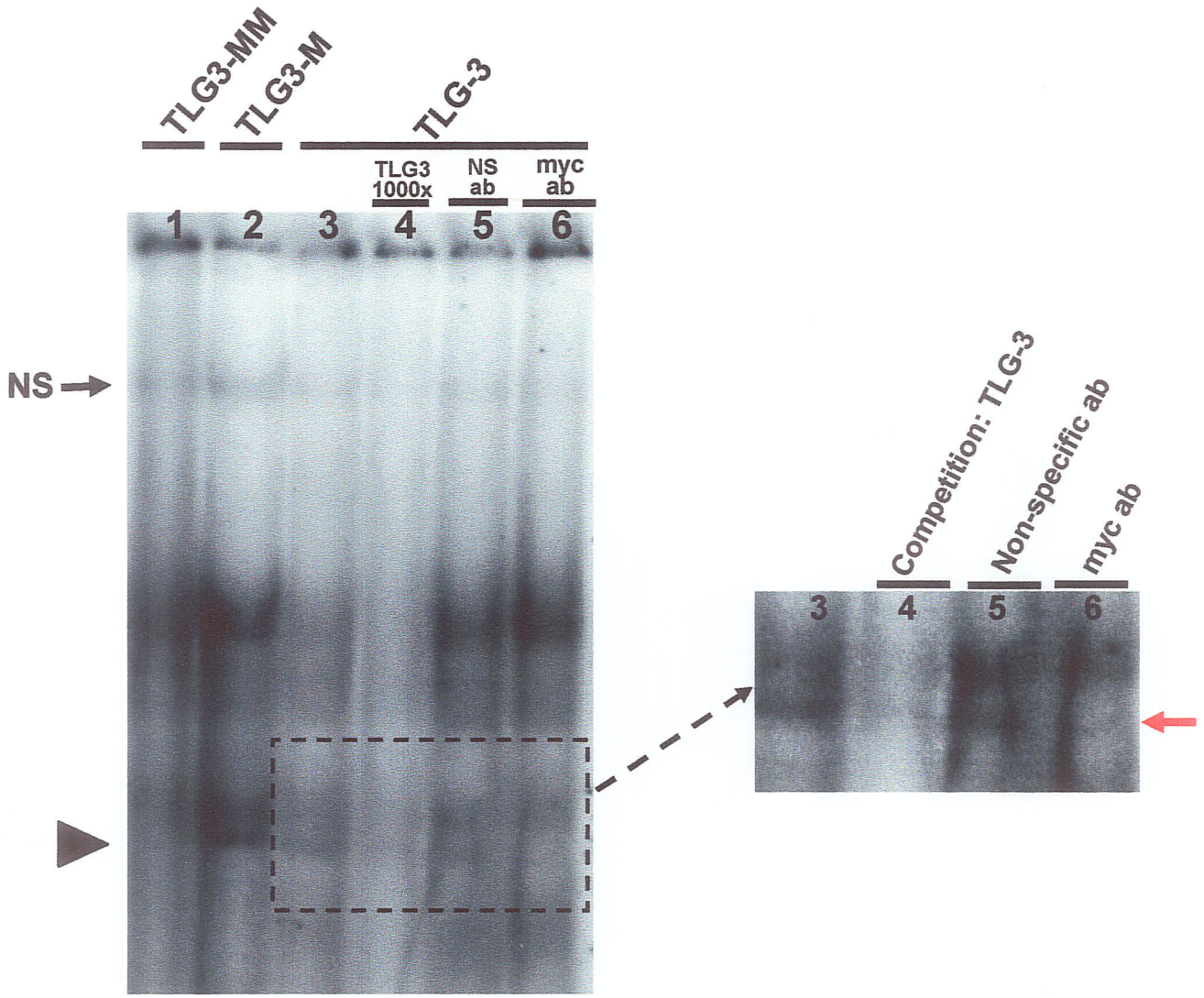


Figure 23: **E47 binds to site 3 of the CRT promoter.** EMSA was performed using nuclear extracts from HEK-293 cells overexpressing E47 myc protein. Synthetic oligos (32bp) representing the wild-type CRT promoter sequence of site 3 (TLG-3) or mutated CRT promoter sequences (TLG3-M and TLG3-MM) were labeled with ^{32}P. 1000-fold excess unlabeled TLG-3 oligo (lane 4) was used for competition studies. Antibodies to myc (lane 6) and an unspecific antibody (lane 5) were used for supershift assays. The E47 protein is marked with a red arrow (see inset). Mutation of E box (TLG1-M) did not change any band shifting (lane 2). Mutation of both E box and GATA site (TLG3-MM) did not generate band corresponding to E47 protein (lane 1 - arrowhead). (NS = non-specific)

E. DISCUSSION

The aim of this study was to investigate the role of the bHLH and GATA-1 proteins in the regulation of CRT gene expression during vascular development. To our knowledge, this is the first study on the regulation of CRT by bHLH proteins. Here we demonstrated that the CRT promoter contains four E boxes flanked by GATA-1 binding sites. Using EMSA and supershift assays, we showed that Tal-1, E47 and GATA-1 bind to the response elements found on the CRT promoter. Furthermore, E47, Tal-1 and GATA-1 enhanced the activity of the CRT promoter individually. We also demonstrated that Tal-1 and GATA-1 can function synergistically to increase expression of the CRT gene.

CRT is a ubiquitous ER resident protein with two main functions: regulation of calcium homeostasis and chaperone of newly synthesized proteins [140]. Recently, the role of CRT in cardiovascular development has been demonstrated using gene targeted deletion [164, 175]. In addition, the CRT promoter was shown to be highly active in the heart, vessel wall and brain of the developing mouse embryo [164]. Interestingly, the expression of CRT diminishes in the heart after birth, while it remains unchanged in the vessel wall of the adult [164, 180, 181, 190]. Furthermore, application of exogenous CRT at the site of a tumour has been shown to inhibit the onset of tumour progression by halting the growth of blood vessels [144, 145, 183]. This suggests a role for CRT in angiogenesis. Blood vessel development requires the coordination of many different signaling and accessory proteins at different stages in development. The regulation of these proteins is crucial in building a mature vascular network. Many different transcription factors are involved in the expression of these proteins during

vasculogenesis and angiogenesis. The understanding of the transcriptional regulation of angiogenic proteins can have therapeutic applications, whether to limit tumour progression by halting angiogenesis or to accelerate the re-oxygenation of damaged tissues in cardiovascular disease. Two proteins, Tal-1 and GATA-1, were found to be expressed in angioblasts and necessary for the maturation of these precursor cells into endothelial cells which leads to the formation of functioning blood vessels [8]. Similarly, the E47 protein is ubiquitously expressed and recently found to be involved in smooth muscle cell differentiation [77]. These three transcription factors form a complex that was shown to interact with the promoters of genes involved in blood development [94]. Since blood cells and endothelial cells arise from the same precursor cell (the hemangioblast), it was thought that these proteins were able to form a complex that activates genes involved in endothelial cell differentiation. One important requirement for this complex formation is that the binding sites for Tal-1 and E47 (E box) and GATA-1 (GATA site) must be in close proximity on the promoter (see Table 1) [94, 95, 125-130]. A consensus sequence of a 6 bp E box site separated by 8-10bp from the 4 bp GATA binding site was found to be optimal for complex formation between these transcription factors [94]. Indeed the CRT promoter contains four putative E box binding sites in proximity to GATA-1 binding sites (Figure 1), which suggests a possible role for its regulation by complex formation between these transcription factors. Two of these E boxes (site 1 and site 3) contain the consensus sequence similar to that reported earlier (Figure 7) and listed in Table I, further supporting this complex formation.

Deletion of the 3' end of the CRT promoter, nucleotides -330 to -1 which contain site 1 E box and GATA, resulted in a 100-fold decrease in the promoter activity (Figure

10 and Figure 11). This indicates that this region of the CRT promoter contains an enhancer region. Furthermore, deletion mutant containing site I (CP1T) alone was sufficient to result in a significant increase in luciferase activity when cells were co-transfected with E47 (Figure 12), Tal-1 (Figure 14), GATA-1 (Figure 15) or Tal-1 and GATA-1 (Figure 15). The site 1 consensus E box/GATA binding site is conserved between CRT promoter of human, rat and mouse indicating an evolutionary importance for this site in the expression of the CRT gene (Figure 7A). A role for site 3 E box in the regulation of the CRT gene was demonstrated by introducing a point mutation in this E box in the full-length promoter (CPF) and deletion mutant (CP3T) (Figure 16). This mutation of the E box in site 3 E box/GATA consensus sequence altered the basal activity of CPF and CP3T in HEK-293 cells. There was an increase in the activity of the full-length promoter and a decrease in activity in CP3T with the mutated E box (Figure 16). This indicates the importance of both site 1 and site 3 in the regulation of the CRT promoter, thus we focused our studies on the effect of E47, Tal-1 and GATA-1 on these two sites.

As seen in Figure 12, expression of E47 alone in HEK-293 cells resulted in a significant increase in the activity of the CRT promoter (and its deletion mutants). However, the effect of E47 expression on CRT was not significant in NIH-3T3 cells (data not shown). This difference between cell lines could be due to the presence of endogenous proteins that differ between NIH-3T3 and HEK-293 cells. Indeed, as seen in Figure 6B, HEK-293 cells express the GATA-1 protein, whereas NIH-3T3 cells do not express GATA-1 [103]. Therefore, the presence of GATA-1 (or other proteins) could interfere or aid in the binding of this protein to the CRT promoter. The effect of E47 in

HEK-293 cells could also be due to the fact that NIH-3T3 cells might have a higher endogenous expression of E47, thus the expression of the CRT promoter is maximal without the overexpression of the E47 vector. An interesting finding with E47 is the role of the third site of the CRT promoter. The three constructs that showed the highest activation of the promoter activity upon the addition of E47 were CPF (2-fold induction), CP3T (2-fold induction) and CPB3T (3-fold induction) (Figure 12). These three constructs all contain the 3rd putative binding site suggesting the importance of this site in E47 function. Furthermore, EMSA experiments using site 3 probe showed that E47 binds to this region in the CRT promoter (Figure 23 – lane 3 and 6). Mutation of the E box site did not affect this binding, whereas mutation of both the E box and GATA-1 site abolished the complex formation (Figure 23 – lane 1 and 2). This indicates that the E box is not essential for E47 binding to the CRT promoter, but it may interact through cofactors dependant on the GATA site and E box. Similarly, E47 was found to bind to site 1 of the CRT promoter (Figure 22 – lane 3 and 6) and mutation of the E box alone did not affect the E47-DNA complex formation (Figure 22 – lane 2). Mutation of both the GATA site and E box inhibited the formation of the E47 complex as well as the formation of two novel complexes (Figure 22- lane 1). This suggests that the binding of E47 protein, as well as other transcription factors present in HEK-293 cells, is dependant on the interaction and formation of complexes with other proteins that bind to the GATA and E box response element of the CRT promoter. Furthermore, the interaction was E47 specific because myc-antibody was able to supershift the band corresponding to E47 (E47-myc). E47 may still be present in the other complexes, but the antibody may not have access to the peptide sequence because of the spatial positioning of the other

proteins. Thus, E47 binds in vitro to site 1 and site 3 of the CRT promoter and has been found to activate the expression of CRT probably by interacting with other transcription factors.

Interestingly, ectopic Tal-1 was found to be involved in the regulation of the CRT promoter in NIH-3T3 cells, but not in HEK-293 cells (Figure 13 and Figure 14). As mentioned, there are different proteins expressed between these cell types such as the GATA-1 protein. The interaction between Tal-1 and GATA-1 may be necessary for correct function of the Tal-1 protein (as will be discussed later). An interesting finding is that Tal-1 had an effect on the promoter only when nucleotides -330 to -1 of the CRT promoter, containing site 1, was intact (Figure 14). The fact that this region contains site 1 indicates the importance for this region in the regulation of Tal-1 activity. This first region was found to be essential for CRT expression (containing an activitory sequence), so it is not surprising that this region is crucial for Tal-1 activity. The fact that site 1 is conserved through human, rat and mouse indicates the need for this site in the regulation of CRT promoter during evolution. In fact, using EMSA we showed that Tal-1 can interact with site 1 of the CRT promoter and it is involved in the formation of two different complexes (Figure 20 – red arrows). These complexes do not depend on the E box for their binding indicating a possible role of the GATA site in Tal-1 activity (Figure 20 – lane 1, 4, 5). Previously, Tal-1 has been shown to exert its effect on vascular development without directly binding to DNA [191]. Thus, it is physiologically possible that Tal-1 may not utilize the E box for the regulation of the CRT promoter. Interaction of Tal-1 to the site 3 of the CRT promoter seems to be weaker, as seen in Figure 21 (lane 2), since there is a shift in only a few bands and with lower intensity than site 1.

Given the fact that Tal-1 must heterodimerize with a ubiquitous bHLH protein is suggestive of the importance of these proteins together in the regulation the CRT promoter. This was in fact not the case, because the addition of Tal-1 and E47 did not induce significant changes in the CRT promoter activity (data not shown). It is possible that Tal-1 requires another E protein family member (such as E12, ITF-1 or ITF-2) for heterodimerization in the maximal activation of the CRT promoter. Potentially, the presence of cofactors may be necessary to significantly up-regulate promoter activity. Interestingly, overexpression of both GATA-1 and Tal-1 in NIH-3T3 cells induced synergistic activation of the CRT promoter (Figure 15). This suggests that Tal-1 and GATA-1 are forming a complex and optimally up-regulating the transcription of the CRT promoter. Tal-1 has been shown to bind to site 1 and site 3 of the CRT promoter, similarly GATA-1 also binds to both of these sites (Figure 18 and 19). This suggests that GATA-1 has a stronger interaction with the CRT promoter and forms complexes with several proteins. However, GATA antibody was able to supershift only the band corresponding to the GATA-1 protein alone. Mutation of GATA binding sequence in site 3 probe (TLG3-GM) resulted in loss of GATA-1 binding (Figure 19 – lane 4, see star). Similarly, competition using the mutated GATA binding sequence in site 1 probe (TLG1-GM) did not affect the mobility of the GATA-myc protein, indicating the importance of an intact GATA site for the binding of this protein to site 1 (Figure 18 – lane 9, see red arrow). Furthermore, we showed that mutation of E box alone (in site 1) does not interfere with GATA-1 binding as was predicted (Figure 18 – lane 2). This mutation resulted in binding of a new protein to the complex. The binding of this new protein to the TLG1-M could be due to the creation of a binding site for a different transcription

factor. However, analysis of the 32 nucleotide sequence TLG1-M by TRANSFAC, did not reveal the presence of any new transcription factor binding site (data not shown). The other possibility is that there is a novel transcription factor binding to the mutated probe. In addition, this new band can not be GATA-1 because GATA-1 specific antibody was not able to supershift this protein (Figure 18 – lane 13), while the band corresponding to GATA-1 was supershifted (Figure 19 – lane 3, see arrow). The other possibility is that GATA-1 antibody may not have access to the peptide sequence because of the spatial positioning of the other proteins in the complex. There is also potentially another GATA family protein member involved in this complex that is not recognized by the GATA-1 specific antibody. It is also probably that the GATA-1 complex formation on TLG-1 may require flanking regions between the E box and GATA binding sites for binding. Finally, mutation of both GATA-1 and the E box in TLG-1 resulted in loss of 2 proteins from the shifted complex. One was the newly seen protein and a lower mobility protein (Figure 18 – lane 1, see double arrowhead). This indicates that the GATA site is important for the binding of the new protein.

The complex formation between GATA-1, Tal-1 and E47 has been well documented (Table 1). It is known that GATA-1 and Tal-1 cannot interact directly [122, 123], however we have shown a synergistic activation of the CRT promoter by these transcription factors. Thus, there must be an accessory protein involved in this synergistic activation that has yet to be determined. Other groups have found that Tal-1, E47 and GATA-1 alone could not significantly regulate the promoter of interest, but with the addition of cofactors such as Lmo2, Ldb-1, SRF and SP1 could induce synergistic activation of the promoter [130]. Lmo2, which is primarily involved in hematopoietic

development, is also expressed in the endothelium [192]. In addition, Lmo2 knockout mice have a primitive vascular network, but no organization of these vessels [192]. Lmo2 requires the Ldb-1 cofactor during normal processes in development. These studies indicate a role for Lmo2 in angiogenesis and suggest a possible role for this protein in the regulation of the CRT promoter during vessel development, but there are no reports on the role of Lmo2 in the regulation of the CRT promoter. This is an interesting area which will be addressed in future studies. Serum-response factor (SRF) is a transcription factor that binds to CArG elements ($CC(A/T_{rich})GG$) in smooth muscle specific promoters. SRF is an important factor in recruiting coactivators such as myocardin [193]. SRF has been shown to interact with E proteins for the synergistic activation of smooth-muscle specific genes and involved in the VEGF signaling mechanism [77, 194]. The interaction of the bHLH proteins with SRF could be important in the regulation of the CRT promoter in blood vessel wall. Finally, as seen in Table 1, SP1 has been demonstrated to be involved in complex formation with Tal-1, E47 and GATA-1 [129, 130]. SP1 is a "house-keeping" gene responsible for recruiting TFIID to TATA-less promoters to help initiate transcription [195, 196]. SP1 has been shown to bind to the Flk-1 promoter and activate its expression in endothelial cells indicating the importance of SP1 in blood vessel development [195, 196]. Thus, much like Lmo2 and SRF, SP1 may be involved in the complex formation regulating the CRT promoter activity.

The fact that there are multiple complexes formed on both site 1 and site 3 of the CRT promoter suggest that there are other proteins involved in the regulation of CRT. Other proteins that bind to the E box and GATA sites are expressed in endothelial cells, thus future research could be directed toward the effect of GATA-2 and Id proteins in the

regulation of the CRT promoter. GATA-2 is highly expressed in endothelial cells and has been shown to have an effect on the Flk-1 gene regulation [20]. Similarly, the Id bHLH proteins, Id-1 and Id-3, were found to be essential for vasculogenesis and angiogenesis in embryonic development and tumour progression [197-199]. Further work is necessary to decipher the precise cofactors that are important to maximally regulate the CRT promoter during vessel development.

F. CONCLUSIONS

1. There are 4 E box binding sites on the CRT promoter which are in proximity to GATA-1 binding sites. Two of these E box sites show conserved motif GATA-8bp-Ebox that has been shown for complex formation between Tal-1 and GATA.
2. There is an enhancer region between -330 to -1 nucleotides of the CRT promoter.
3. The overexpression of E47 increases the activity of the CRT promoter in HEK-293 cells (CPF, CP3T, and CPB3T) suggesting the involvement of site 3 in this regulation. However, ectopic E47 had no significant effect in NIH-3T3 cells.
4. The expression of Tal-1 increases the activity of the CRT promoter in NIH-3T3 cells (CPF, CP1T). On the other hand, Tal-1 does not have a significant effect in HEK-293 cells.
5. The expression of GATA-1 increases the activity of the CRT promoter in NIH-3T3 cells (CPF, CP1T). There is also selective inhibition with the addition of GATA-1 to one construct (CP3T).
6. The addition of GATA-1 and Tal-1 together induce synergistic activation of the CRT promoter in NIH-3T3 cells (CPF, CP2T, CP1T).
7. GATA-1, Tal-1 and E47 all bind to the CRT promoter.

G. REFERENCES

1. Bellairs, R., *The primitive streak.* Anat Embryol (Berl), 1986. **174**(1): p. 1-14.
2. Schoenwolf, G.C., V. Garcia-Martinez, and M.S. Dias, *Mesoderm movement and fate during avian gastrulation and neurulation.* Dev Dyn, 1992. **193**(3): p. 235-48.
3. Gilbert, S.F., *The early development of vertebrates: fish, birds and mammals*, in *Developmental Biology*. 2000, Sinauer associates, Inc. p. 339-376.
4. Gilbert, S.F., *Lateral plate mesoderm and endoderm*, in *Development Biology*. 2000, Sinauer associates, Inc. p. 471-501.
5. Choi, K., *Hemangioblast development and regulation.* Biochem Cell Biol, 1998. **76**(6): p. 947-56.
6. Caprioli, A., et al., *Hemangioblast commitment in the avian allantois: cellular and molecular aspects.* Dev Biol, 2001. **238**(1): p. 64-78.
7. Shalaby, F., et al., *Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice.* Nature, 1995. **376**(6535): p. 62-6.
8. Minko, K., et al., *From mesoderm to blood islands: patterns of key molecules during yolk sac erythropoiesis.* Gene Expr Patterns, 2003. **3**(3): p. 261-72.
9. Ema, M. and J. Rossant, *Cell fate decisions in early blood vessel formation.* Trends Cardiovasc Med, 2003. **13**(6): p. 254-9.
10. Risau, W., *Mechanisms of angiogenesis.* Nature, 1997. **386**(6626): p. 671-4.
11. Hanahan, D., *Signaling vascular morphogenesis and maintenance.* Science, 1997. **277**(5322): p. 48-50.

12. Petrova, T.V., T. Makinen, and K. Alitalo, *Signaling via vascular endothelial growth factor receptors.* Exp Cell Res, 1999. **253**(1): p. 117-30.

13. Yancopoulos, G.D., et al., *Vascular-specific growth factors and blood vessel formation.* Nature, 2000. **407**(6801): p. 242-8.

14. Zachary, I. and G. Gliki, *Signaling transduction mechanisms mediating biological actions of the vascular endothelial growth factor family.* Cardiovasc Res, 2001. **49**(3): p. 568-81.

15. Fong, G.H., et al., *Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium.* Nature, 1995. **376**(6535): p. 66-70.

16. Dumont, D.J., et al., *Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo.* Genes Dev, 1994. **8**(16): p. 1897-909.

17. Maisonpierre, P.C., et al., *Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis.* Science, 1997. **277**(5322): p. 55-60.

18. Puri, M.C., et al., *The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells.* Embo J, 1995. **14**(23): p. 5884-91.

19. Chung, Y.S., et al., *Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression.* Development, 2002. **129**(23): p. 5511-20.

20. Kappel, A., et al., *Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development.* Blood, 2000. **96**(9): p. 3078-85.

21. Buchwalter, G., C. Gross, and B. Wasylyk, *Ets ternary complex transcription factors.* Gene, 2004. **324**: p. 1-14.

22. Oikawa, T. and T. Yamada, *Molecular biology of the Ets family of transcription factors.* Gene, 2003. **303**: p. 11-34.

23. Dube, A., et al., *Role of the Ets transcription factors in the regulation of the vascular-specific Tie2 gene.* Circ Res, 1999. **84**(10): p. 1177-85.

24. Dube, A., et al., *Elf-1 is a transcriptional regulator of the Tie2 gene during vascular development.* Circ Res, 2001. **88**(2): p. 237-44.

25. Morishita, K., D.E. Johnson, and L.T. Williams, *A novel promoter for vascular endothelial growth factor receptor (flt-1) that confers endothelial-specific gene expression.* J Biol Chem, 1995. **270**(46): p. 27948-53.

26. Korhonen, J., et al., *Endothelial-specific gene expression directed by the tie gene promoter in vivo.* Blood, 1995. **86**(5): p. 1828-35.

27. Schlaeger, T.M., et al., *Uniform vascular-endothelial-cell-specific gene expression in both embryonic and adult transgenic mice.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3058-63.

28. Wakiya, K., et al., *A cAMP response element and an Ets motif are involved in the transcriptional regulation of flt-1 tyrosine kinase (vascular endothelial growth factor receptor 1) gene.* J Biol Chem, 1996. **271**(48): p. 30823-8.

29. Iljin, K., et al., *Role of ets factors in the activity and endothelial cell specificity of the mouse Tie gene promoter.* Faseb J, 1999. **13**(2): p. 377-86.

30. Oda, N., M. Abe, and Y. Sato, *ETS-1 converts endothelial cells to the angiogenic phenotype by inducing the expression of matrix metalloproteinases and integrin beta3.* J Cell Physiol, 1999. **178**(2): p. 121-32.

31. Scatena, M. and C. Giachelli, *The alpha(v)beta3 integrin, NF-kappaB, osteoprotegerin endothelial cell survival pathway. Potential role in angiogenesis.* Trends Cardiovasc Med, 2002. **12**(2): p. 83-8.

32. Hodivala-Dilke, K.M., A.R. Reynolds, and L.E. Reynolds, *Integrins in angiogenesis: multitalented molecules in a balancing act.* Cell Tissue Res, 2003. **314**(1): p. 131-44.

33. Bracken, C.P., M.L. Whitelaw, and D.J. Peet, *The hypoxia-inducible factors: key transcriptional regulators of hypoxic responses.* Cell Mol Life Sci, 2003. **60**(7): p. 1376-93.

34. Levy, A.P., et al., *Transcriptional regulation of the rat vascular endothelial growth factor gene by hypoxia.* J Biol Chem, 1995. **270**(22): p. 13333-40.

35. Forsythe, J.A., et al., *Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1.* Mol Cell Biol, 1996. **16**(9): p. 4604-13.

36. Gerber, H.P., et al., *Differential transcriptional regulation of the two vascular endothelial growth factor receptor genes. Flt-1, but not Flk-1/KDR, is up-regulated by hypoxia.* J Biol Chem, 1997. **272**(38): p. 23659-67.

37. Kappel, A., et al., *Identification of vascular endothelial growth factor (VEGF) receptor-2 (Flk-1) promoter/enhancer sequences sufficient for angioblast and endothelial cell-specific transcription in transgenic mice.* Blood, 1999. **93**(12): p. 4284-92.

38. Elvert, G., et al., *Cooperative interaction of hypoxia-inducible factor-2alpha (HIF-2alpha) and Ets-1 in the transcriptional activation of vascular endothelial growth factor receptor-2 (Flk-1).* J Biol Chem, 2003. **278**(9): p. 7520-30.

39. Yamakawa, M., et al., *Hypoxia-inducible factor-1 mediates activation of cultured vascular endothelial cells by inducing multiple angiogenic factors.* Circ Res, 2003. **93**(7): p. 664-73.

40. Gorski, D.H. and K. Walsh, *Control of vascular cell differentiation by homeobox transcription factors.* Trends Cardiovasc Med, 2003. **13**(6): p. 213-20.

41. Boudreau, N., et al., *Induction of the angiogenic phenotype by Hox D3.* J Cell Biol, 1997. **139**(1): p. 257-64.

42. Boudreau, N.J. and J.A. Varner, *The homeobox transcription factor Hox D3 promotes integrin alpha5beta1 expression and function during angiogenesis.* J Biol Chem, 2004. **279**(6): p. 4862-8.

43. Myers, C., A. Charboneau, and N. Boudreau, *Homeobox B3 promotes capillary morphogenesis and angiogenesis.* J Cell Biol, 2000. **148**(2): p. 343-51.

44. Nakagawa, T., et al., *HEX acts as a negative regulator of angiogenesis by modulating the expression of angiogenesis-related gene in endothelial cells in vitro.* Arterioscler Thromb Vasc Biol, 2003. **23**(2): p. 231-7.

45. Minami, T., et al., *Interaction between hex and GATA transcription factors in vascular endothelial cells inhibits flk-1/KDR-mediated vascular endothelial growth factor signaling.* J Biol Chem, 2004. **279**(20): p. 20626-35.

46. Boutet, S.C., T. Quertermous, and B.M. Fadel, *Identification of an octamer element required for in vivo expression of the TIE1 gene in endothelial cells.* Biochem J, 2001. **360**(Pt 1): p. 23-9.

47. Kadesch, T., *Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription.* Immunol Today, 1992. **13**(1): p. 31-6.

48. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins.* Cell, 1989. **56**(5): p. 777-83.

49. Voronova, A. and D. Baltimore, *Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4722-6.

50. Murre, C., et al., *Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence.* Cell, 1989. **58**(3): p. 537-44.

51. Benezra, R., *An intermolecular disulfide bond stabilizes E2A homodimers and is required for DNA binding at physiological temperatures.* Cell, 1994. **79**(6): p. 1057-67.

52. Church, G.M., et al., *Cell-type-specific contacts to immunoglobulin enhancers in nuclei.* Nature, 1985. **313**(6005): p. 798-801.

53. Atchley, W.R. and W.M. Fitch, *A natural classification of the basic helix-loop-helix class of transcription factors.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5172-6.

54. Massari, M.E. and C. Murre, *Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms.* Mol Cell Biol, 2000. **20**(2): p. 429-40.

55. Littlewood, T.D., *Helix-loop-helix transcription factors*. 3 ed. Protein profile. 1998, Oxford ; New York: Oxford University Press.

56. Murre, C., et al., *Structure and function of helix-loop-helix proteins.* Biochim Biophys Acta, 1994. **1218**(2): p. 129-35.

57. Hsu, H.L., I. Wadman, and R. Baer, *Formation of in vivo complexes between the TAL1 and E2A polypeptides of leukemic T cells.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3181-5.

58. Ghosh, I., P. Bishop, and J. Chmielewski, *DNA binding properties of basic helix-loop-helix fusion proteins of Tal and E47.* J Pept Res, 2001. **57**(5): p. 354-60.

59. Benezra, R., et al., *The protein Id: a negative regulator of helix-loop-helix DNA binding proteins.* Cell, 1990. **61**(1): p. 49-59.

60. Ellis, H.M., D.R. Spann, and J.W. Posakony, *extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins.* Cell, 1990. **61**(1): p. 27-38.

61. Molkentin, J.D. and E.N. Olson, *Defining the regulatory networks for muscle development.* Curr Opin Genet Dev, 1996. **6**(4): p. 445-53.

62. Molkentin, J.D. and E.N. Olson, *Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9366-73.

63. Zhuang, Y., P. Soriano, and H. Weintraub, *The helix-loop-helix gene E2A is required for B cell formation.* Cell, 1994. **79**(5): p. 875-84.

64. Hata, K. and J. Mizuguchi, *Genomic organization and characterization of the promoter for the E2A gene.* Gene, 2004. **325**: p. 53-61.

65. Roberts, V.J., R. Steenbergen, and C. Murre, *Localization of E2A mRNA expression in developing and adult rat tissues.* Proc Natl Acad Sci U S A, 1993. **90**(16): p. 7583-7.

66. Rutherford, M.N. and D.P. LeBrun, *Restricted expression of E2A protein in primary human tissues correlates with proliferation and differentiation.* Am J Pathol, 1998. **153**(1): p. 165-73.

67. Aronheim, A., et al., *The E2A gene product contains two separable and functionally distinct transcription activation domains.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8063-7.

68. Markus, M., Z. Du, and R. Benezra, *Enhancer-specific modulation of E protein activity.* J Biol Chem, 2002. **277**(8): p. 6469-77.

69. Henthorn, P., M. Kiledjian, and T. Kadesch, *Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif.* Science, 1990. **247**(4941): p. 467-70.

70. Skerjanc, I.S., et al., *A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity.* J Biol Chem, 1996. **271**(7): p. 3555-61.

71. Bain, G., et al., *E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements.* Cell, 1994. **79**(5): p. 885-92.

72. Bain, G., et al., *E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas.* Mol Cell Biol, 1997. **17**(8): p. 4782-91.

73. Morrison, A.M., et al., *Loss- and gain-of-function mutations reveal an important role of BSAP (Pax-5) at the start and end of B cell differentiation.* Semin Immunol, 1998. **10**(2): p. 133-42.

74. Agrawal, A., Q.M. Eastman, and D.G. Schatz, *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.* Nature, 1998. **394**(6695): p. 744-51.

75. Plasterk, R., *V(D)J recombination. Ragtime jumping.* Nature, 1998. **394**(6695): p. 718-9.

76. Jung, F., et al., *Characterization of an E-box-dependent cis element in the smooth muscle alpha-actin promoter.* Arterioscler Thromb Vasc Biol, 1999. **19**(11): p. 2591-9.

77. Kumar, M.S., et al., *Smooth muscle alpha-actin gene requires two E-boxes for proper expression in vivo and is a target of class I basic helix-loop-helix proteins.* Circ Res, 2003. **92**(8): p. 840-7.

78. Finger, L.R., et al., *Involvement of the TCL5 gene on human chromosome 1 in T-cell leukemia and melanoma.* Proc Natl Acad Sci U S A, 1989. **86**(13): p. 5039-43.

79. Chen, Q., et al., *The tal gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein.* Embo J, 1990. **9**(2): p. 415-24.

80. Hsu, H.L., et al., *Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins.* Mol Cell Biol, 1991. **11**(6): p. 3037-42.

81. Hsu, H.L., et al., *Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins.* Mol Cell Biol, 1994. **14**(2): p. 1256-65.

82. Voronova, A.F. and F. Lee, *The E2A and tal-1 helix-loop-helix proteins associate in vivo and are modulated by Id proteins during interleukin 6-induced myeloid differentiation.* Proc Natl Acad Sci U S A, 1994. **91**(13): p. 5952-6.

83. Kallianpur, A.R., J.E. Jordan, and S.J. Brandt, *The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis.* Blood, 1994. **83**(5): p. 1200-8.

84. Pulford, K., et al., *Expression of TAL-1 proteins in human tissues.* Blood, 1995. **85**(3): p. 675-84.

85. Elefanty, A.G., et al., *SCL expression in the mouse embryo detected with a targeted lacZ reporter gene demonstrates its localization to hematopoietic, vascular, and neural tissues.* Blood, 1999. **94**(11): p. 3754-63.

86. Hwang, L.Y., et al., *Expression of the TAL1 proto-oncogene in cultured endothelial cells and blood vessels of the spleen.* Oncogene, 1993. **8**(11): p. 3043-6.

87. Drake, C.J., et al., *TAL1/SCL is expressed in endothelial progenitor cells/angioblasts and defines a dorsal-to-ventral gradient of vasculogenesis.* Dev Biol, 1997. **192**(1): p. 17-30.

88. Robb, L., et al., *Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7075-9.

89. Shivdasani, R.A., E.L. Mayer, and S.H. Orkin, *Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL.* Nature, 1995. **373**(6513): p. 432-4.

90. Elefanty, A.G., et al., *Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells.* Blood, 1997. **90**(4): p. 1435-47.

91. Gering, M., et al., *The SCL gene specifies haemangioblast development from early mesoderm.* Embo J, 1998. **17**(14): p. 4029-45.

92. Visvader, J.E., Y. Fujiwara, and S.H. Orkin, *Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development.* Genes Dev, 1998. **12**(4): p. 473-9.

93. Bloor, A.J., et al., *The role of the stem cell leukemia (SCL) gene in hematopoietic and endothelial lineage specification.* J Hematother Stem Cell Res, 2002. **11**(2): p. 195-206.

94. Wadman, I.A., et al., *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, 1997. **16**(11): p. 3145-57.

95. Ono, Y., N. Fukuhara, and O. Yoshie, *TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3.* Mol Cell Biol, 1998. **18**(12): p. 6939-50.

96. Hanas, J.S., et al., *Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene.* J Biol Chem, 1983. **258**(23): p. 14120-5.

97. Omichinski, J.G., et al., *A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex.* Proc Natl Acad Sci U S A, 1993. **90**(5): p. 1676-80.

98. Mackay, J.P. and M. Crossley, *Zinc fingers are sticking together.* Trends Biochem Sci, 1998. **23**(1): p. 1-4.

99. Tupler, R., G. Perini, and M.R. Green, *Expressing the human genome.* Nature, 2001. **409**(6822): p. 832-3.

100. Matthews, J.M. and M. Sunde, *Zinc fingers--folds for many occasions.* IUBMB Life, 2002. **54**(6): p. 351-5.

101. Pavletich, N.P. and C.O. Pabo, *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A.* Science, 1991. **252**(5007): p. 809-17.

102. Merika, M. and S.H. Orkin, *DNA-binding specificity of GATA family transcription factors.* Mol Cell Biol, 1993. **13**(7): p. 3999-4010.

103. Tsai, S.F., et al., *Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells.* Nature, 1989. **339**(6224): p. 446-51.

104. Zon, L.I., et al., *The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome.* Proc Natl Acad Sci U S A, 1990. **87**(2): p. 668-72.

105. Trainor, C.D., et al., *Structure and evolution of a human erythroid transcription factor.* Nature, 1990. **343**(6253): p. 92-6.

106. Lowry, J.A. and W.R. Atchley, *Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain.* J Mol Evol, 2000. **50**(2): p. 103-15.

107. Patient, R.K. and J.D. McGhee, *The GATA family (vertebrates and invertebrates).* Curr Opin Genet Dev, 2002. **12**(4): p. 416-22.

108. Trainor, C.D., et al., *A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction.* Mol Cell Biol, 1996. **16**(5): p. 2238-47.

109. Zon, L.I., et al., *Expression of GATA-binding proteins during embryonic development in Xenopus laevis.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10642-6.

110. Orkin, S.H., *GATA-binding transcription factors in hematopoietic cells.* Blood, 1992. **80**(3): p. 575-81.

111. Molkentin, J.D., *The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression.* J Biol Chem, 2000. **275**(50): p. 38949-52.

112. Mott, B.H., J. Bassman, and M.J. Pikaart, *A molecular dissection of the interaction between the transcription factor Gata-1 zinc finger and DNA.* Biochem Biophys Res Commun, 2004. **316**(3): p. 910-7.

113. Hannon, R., et al., *Structure and promoter activity of the gene for the erythroid transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1991. **88**(8): p. 3004-8.

114. Newton, A., J. Mackay, and M. Crossley, *The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA.* J Biol Chem, 2001. **276**(38): p. 35794-801.

115. Pevny, L., et al., *Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1.* Nature, 1991. **349**(6306): p. 257-60.

116. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.

117. Weiss, M.J., C. Yu, and S.H. Orkin, *Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line.* Mol Cell Biol, 1997. **17**(3): p. 1642-51.

118. Desjarlais, J.R. and J.M. Berg, *Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2256-60.

119. Kang, J.S. and J.S. Kim, *Zinc finger proteins as designer transcription factors.* J Biol Chem, 2000. **275**(12): p. 8742-8.

120. Rebar, E.J., et al., *Induction of angiogenesis in a mouse model using engineered transcription factors.* Nat Med, 2002. **8**(12): p. 1427-32.

121. Sangamo, B.I., *http://www.sangamo.com/index.php.* http://www.sangamo.com/index.php, 2004.

122. Osada, H., et al., *Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1.* Proc Natl Acad Sci U S A, 1995. **92**(21): p. 9585-9.

123. Osada, H., et al., *LIM-only protein Lmo2 forms a protein complex with erythroid transcription factor GATA-1.* Leukemia, 1997. **11 Suppl 3**: p. 307-12.

124. Visvader, J.E., et al., *The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13707-12.

125. Grutz, G.G., et al., *The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells.* Embo J, 1998. **17**(16): p. 4594-605.

126. Cohen-Kaminsky, S., et al., *Chromatin immunoselection defines a TAL-1 target gene.* Embo J, 1998. **17**(17): p. 5151-60.

127. Anderson, K.P., S.C. Crable, and J.B. Lingrel, *Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene.* J Biol Chem, 1998. **273**(23): p. 14347-54.

128. Vitelli, L., et al., *A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts.* Mol Cell Biol, 2000. **20**(14): p. 5330-42.

129. Lecuyer, E., et al., *The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1.* Blood, 2002. **100**(7): p. 2430-40.

130. Lahlil, R., et al., *SCL assembles a multifactorial complex that determines glycophorin A expression.* Mol Cell Biol, 2004. **24**(4): p. 1439-52.

131. Ostwald, T.J. and D.H. MacLennan, *Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum.* J Biol Chem, 1974. **249**(3): p. 974-9.

132. Fliegel, L., et al., *Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum.* J Biol Chem, 1989. **264**(36): p. 21522-8.

133. Smith, M.J. and G.L. Koch, *Multiple zones in the sequence of calreticulin (CRP55, calregulin, HACBP), a major calcium binding ER/SR protein.* Embo J, 1989. **8**(12): p. 3581-6.

134. McCauliffe, D.P., et al., *The 5'-flanking region of the human calreticulin gene shares homology with the human GRP78, GRP94, and protein disulfide isomerase promoters.* J Biol Chem, 1992. **267**(4): p. 2557-62.

135. Waser, M., et al., *Regulation of calreticulin gene expression by calcium.* J Cell Biol, 1997. **138**(3): p. 547-57.

136. Rooke, K., et al., *Mapping of the gene for calreticulin (Calr) to mouse chromosome 8.* Mamm Genome, 1997. **8**(11): p. 870-1.

137. Matsuoka, K., et al., *Covalent structure of bovine brain calreticulin.* Biochem J, 1994. **298 (Pt 2)**: p. 435-42.

138. Coppolino, M.G. and S. Dedhar, *Calreticulin.* Int J Biochem Cell Biol, 1998. **30**(5): p. 553-8.

139. Sontheimer, R.D., et al., *The unveiling of calreticulin--a clinically relevant tour of modern cell biology.* J Investig Med, 1995. **43**(4): p. 362-70.

140. Michalak, M., et al., *Calreticulin: one protein, one gene, many functions.* Biochem J, 1999. **344 Pt 2**: p. 281-92.

141. Michalak, M., et al., *Calreticulin in cardiac development and pathology.* Biochim Biophys Acta, 2002. **1600**(1-2): p. 32-7.

142. Baksh, S., et al., *Identification of the Zn2+ binding region in calreticulin.* FEBS Lett, 1995. **376**(1-2): p. 53-7.

143. Baksh, S., et al., *Interaction of calreticulin with protein disulfide isomerase.* J Biol Chem, 1995. **270**(52): p. 31338-44.

144. Pike, S.E., et al., *Vasostatin, a calreticulin fragment, inhibits angiogenesis and suppresses tumor growth.* J Exp Med, 1998. **188**(12): p. 2349-56.

145. Pike, S.E., et al., *Calreticulin and calreticulin fragments are endothelial cell inhibitors that suppress tumor growth.* Blood, 1999. **94**(7): p. 2461-8.

146. Baksh, S. and M. Michalak, *Expression of calreticulin in Escherichia coli and identification of its Ca2+ binding domains.* J Biol Chem, 1991. **266**(32): p. 21458-65.

147. Ellgaard, L., et al., *Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment.* FEBS Lett, 2001. **488**(1-2): p. 69-73.

148. Fliegel, L., et al., *The high-affinity calcium binding protein of sarcoplasmic reticulum. Tissue distribution, and homology with calregulin.* Biochim Biophys Acta, 1989. **982**(1): p. 1-8.

149. Burns, K., et al., *Modulation of gene expression by calreticulin binding to the glucocorticoid receptor.* Nature, 1994. **367**(6462): p. 476-80.

150. Dedhar, S., *Novel functions for calreticulin: interaction with integrins and modulation of gene expression?* Trends Biochem Sci, 1994. **19**(7): p. 269-71.

151. Kuwabara, K., et al., *Calreticulin, an antithrombotic agent which binds to vitamin K-dependent coagulation factors, stimulates endothelial nitric oxide production, and limits thrombosis in canine coronary arteries.* J Biol Chem, 1995. **270**(14): p. 8179-87.

152. McDonnell, J.M., et al., *Calreticulin binding affinity for glycosylated laminin.* J Biol Chem, 1996. **271**(14): p. 7891-4.

153. Saito, Y., et al., *Calreticulin functions in vitro as a molecular chaperone for both glycosylated and non-glycosylated proteins.* Embo J, 1999. **18**(23): p. 6718-29.

154. Jorgensen, C.S., et al., *Polypeptide binding properties of the chaperone calreticulin.* Eur J Biochem, 2000. **267**(10): p. 2945-54.

155. Otteken, A. and B. Moss, *Calreticulin interacts with newly synthesized human immunodeficiency virus type 1 envelope glycoprotein, suggesting a chaperone function similar to that of calnexin.* J Biol Chem, 1996. **271**(1): p. 97-103.

156. Nauseef, W.M., S.J. McCormick, and R.A. Clark, *Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase.* J Biol Chem, 1995. **270**(9): p. 4741-7.

157. Opas, M., et al., *Calreticulin modulates cell adhesiveness via regulation of vinculin expression.* J Cell Biol, 1996. **135**(6 Pt 2): p. 1913-23.

158. Meldolesi, J. and T. Pozzan, *The endoplasmic reticulum Ca2+ store: a view from the lumen.* Trends Biochem Sci, 1998. **23**(1): p. 10-4.

159. Liu, N., et al., *Decreasing calreticulin expression lowers the Ca2+ response to bradykinin and increases sensitivity to ionomycin in NG-108-15 cells.* J Biol Chem, 1994. **269**(46): p. 28635-9.

160. Bastianutto, C., et al., *Overexpression of calreticulin increases the Ca2+ capacity of rapidly exchanging Ca2+ stores and reveals aspects of their lumenal microenvironment and function.* J Cell Biol, 1995. **130**(4): p. 847-55.

161. Mery, L., et al., *Overexpression of calreticulin increases intracellular Ca2+ storage and decreases store-operated Ca2+ influx.* J Biol Chem, 1996. **271**(16): p. 9332-9.

162. Michikawa, T., et al., *Transmembrane topology and sites of N-glycosylation of inositol 1,4,5-trisphosphate receptor.* J Biol Chem, 1994. **269**(12): p. 9184-9.

163. John, L.M., J.D. Lechleiter, and P. Camacho, *Differential modulation of SERCA2 isoforms by calreticulin.* J Cell Biol, 1998. **142**(4): p. 963-73.

164. Mesaeli, N., et al., *Calreticulin is essential for cardiac development.* J Cell Biol, 1999. **144**(5): p. 857-68.

165. Li, J., et al., *Calreticulin reveals a critical Ca(2+) checkpoint in cardiac myofibrillogenesis.* J Cell Biol, 2002. **158**(1): p. 103-13.

166. Michalak, M., et al., *Endoplasmic reticulum form of calreticulin modulates glucocorticoid-sensitive gene expression.* J Biol Chem, 1996. **271**(46): p. 29436-45.

167. Coppolino, M., et al., *Inducible interaction of integrin alpha 2 beta 1 with calreticulin. Dependence on the activation state of the integrin.* J Biol Chem, 1995. **270**(39): p. 23132-8.

168. Coppolino, M.G., et al., *Calreticulin is essential for integrin-mediated calcium signalling and cell adhesion.* Nature, 1997. **386**(6627): p. 843-7.

169. Fadel, M.P., et al., *Calreticulin affects focal contact-dependent but not close contact-dependent cell-substratum adhesion.* J Biol Chem, 1999. **274**(21): p. 15085-94.

170. Khalife, J., et al., *Cloning and sequencing of the gene encoding Schistosoma mansoni calreticulin.* Mol Biochem Parasitol, 1993. **62**(2): p. 313-5.

171. Khalife, J., C. Godin, and A. Capron, *Transcriptional regulation of Schistosoma mansoni calreticulin: possible role of AP-1.* Parasitology, 1995. **111 (Pt 4)**: p. 469-75.

172. Conway, E.M., et al., *Heat shock-sensitive expression of calreticulin. In vitro and in vivo up-regulation.* J Biol Chem, 1995. **270**(28): p. 17011-6.

173. Nguyen, T.O., J.D. Capra, and R.D. Sontheimer, *Calreticulin is transcriptionally upregulated by heat shock, calcium and heavy metals.* Mol Immunol, 1996. **33**(4-5): p. 379-86.

174. Elmileik, H., et al., *Use of bleomycin- and heat shock-induced calreticulin promoter for construction of a mammalian expression vector.* J Biochem (Tokyo), 2001. **129**(5): p. 671-4.

175. Rauch, F., et al., *Heart, brain, and body wall defects in mice lacking calreticulin.* Exp Cell Res, 2000. **256**(1): p. 105-11.

176. Nakamura, K., et al., *Complete heart block and sudden death in mice overexpressing calreticulin.* J Clin Invest, 2001. **107**(10): p. 1245-53.

177. Guo, L., et al., *COUP-TF1 antagonizes Nkx2.5-mediated activation of the calreticulin gene during cardiac development.* J Biol Chem, 2001. **276**(4): p. 2797-801.

178. Lyons, I., et al., *Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5.* Genes Dev, 1995. **9**(13): p. 1654-66.

179. Milner, R.E., et al., *Calreticulin, and not calsequestrin, is the major calcium binding protein of smooth muscle sarcoplasmic reticulum and liver endoplasmic reticulum.* J Biol Chem, 1991. **266**(11): p. 7155-65.

180. Tharin, S., et al., *Widespread tissue distribution of rabbit calreticulin, a non-muscle functional analogue of calsequestrin.* Cell Tissue Res, 1992. **269**(1): p. 29-37.

181. Patton, W.F., et al., *Components of the protein synthesis and folding machinery are induced in vascular smooth muscle cells by hypertrophic and hyperplastic agents. Identification by comparative protein phenotyping and microsequencing.* J Biol Chem, 1995. **270**(36): p. 21404-10.

182. Dai, E., et al., *Calreticulin, a potential vascular regulatory protein, reduces intimal hyperplasia after arterial injury.* Arterioscler Thromb Vasc Biol, 1997. **17**(11): p. 2359-68.

183. Xiao, F., et al., *A gene therapy for cancer based on the angiogenesis inhibitor, vasostatin.* Gene Ther, 2002. **9**(18): p. 1207-13.

184. Ito, H., Y. Seyama, and S. Kubota, *Calreticulin is directly involved in anti-alpha3 integrin antibody-mediated secretion and activation of matrix metalloprotease-2.* Biochem Biophys Res Commun, 2001. **283**(2): p. 297-302.

185. Yao, L., S.E. Pike, and G. Tosato, *Laminin binding to the calreticulin fragment vasostatin regulates endothelial cell function.* J Leukoc Biol, 2002. **71**(1): p. 47-53.

186. Lange-Asschenfeldt, B., et al., *The angiogenesis inhibitor vasostatin does not impair wound healing at tumor-inhibiting doses.* J Invest Dermatol, 2001. **117**(5): p. 1036-41.

187. Angiolillo, A.L., et al., *Human interferon-inducible protein 10 is a potent inhibitor of angiogenesis in vivo.* J Exp Med, 1995. **182**(1): p. 155-62.

188. Sgadari, C., et al., *Interferon-inducible protein-10 identified as a mediator of tumor necrosis in vivo.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13791-6.

189. Wingender, E., et al., *TRANSFAC: an integrated system for gene expression regulation.* Nucleic Acids Res, 2000. **28**(1): p. 316-9.

190. Miller, J., A.D. McLachlan, and A. Klug, *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.* Embo J, 1985. **4**(6): p. 1609-14.

191. Porcher, C., et al., *Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding.* Development, 1999. **126**(20): p. 4603-15.

192. Yamada, Y., et al., *The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice.* Proc Natl Acad Sci U S A, 2000. **97**(1): p. 320-4.

193. Wang, Z., et al., *Myocardin is a master regulator of smooth muscle gene expression.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7129-34.

194. Chai, J., M.K. Jones, and A.S. Tarnawski, *Serum response factor is a critical requirement for VEGF signaling in endothelial cells and VEGF-induced angiogenesis.* Faseb J, 2004.

195. Patterson, C., et al., *Nuclear protein interactions with the human KDR/flk-1 promoter in vivo. Regulation of Sp1 binding is associated with cell type-specific expression.* J Biol Chem, 1997. **272**(13): p. 8410-6.

196. Hata, Y., et al., *Transcription factors Sp1 and Sp3 alter vascular endothelial growth factor receptor expression through a novel recognition sequence.* J Biol Chem, 1998. **273**(30): p. 19294-303.

197. Lyden, D., et al., *Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts.* Nature, 1999. **401**(6754): p. 670-7.

198. Benezra, R., *Role of Id proteins in embryonic and tumor angiogenesis.* Trends Cardiovasc Med, 2001. **11**(6): p. 237-41.

199. Tang, J., et al., *The helix-loop-helix protein id-1 delays onset of replicative senescence in human endothelial cells.* Lab Invest, 2002. **82**(8): p. 1073-9.