HEME BIOSYNTHESIS IN ERYTHROID CELLS:

Transcriptional Regulation of the

Human 5-Aminolevulinate Synthase 2 Gene

A thesis submitted to the University of Adelaide

for the degree of Doctor of Philosophy

by

Katharina Helen Surinya B.Sc. Hons (University of Adelaide)

Department of Biochemistry

University of Adelaide

Adelaide, South Australia.   August, 1997

TABLE OF CONTENTS

CHAPTER ONE : GENERAL INTRODUCTION

CHAPTER THREE: EVALUATION OF ERYTHROID CELL LINES SUITABLE
FOR TRANSIENT TRANSFECTION STUDIES AND
SYNTHESIS OF A STABLY TRANSFECTED MEL CELL LINE

CHAPTER FOUR : TRANSIENT TRANSFECTION STUDIES OF THE ALAS2
5'-FLANKING AND INTRONIC REGIONS ON THE
TISSUE-SPECIFIC EXPRESSION OF THE ALAS2 GENE

CHAPTER FIVE : IDENTIFICATION OF REGULATORY ELEMENTS AND CORRESPONDING TRANSCRIPTION FACTORS THAT DRIVE EXPRESSION OF THE ALAS2 PROMOTER

PUBLICATIONS

FORMAL PRESENTATIONS

THESIS SUMMARY

The focus of research in our laboratory has involved the elucidation of the regulatory mechanisms that control heme biosynthesis in eukaryotic cells. 5-aminolevulinate synthase (ALAS) is a nuclear encoded mitochondrial matrix enzyme that catalyses the formation of 5-aminolevulinate from glycine and succinyl CoA, the first step in the heme biosynthetic pathway. This enzyme is of particular interest since it is the rate-controlling enzyme. There are two closely related isozymes of ALAS designated ALAS1 and ALAS2 which are encoded by separate genes located on different chromosomes. The housekeeping enzyme, ALAS1 is probably expressed in all tissues to provide heme for respiratory cytochromes and other heme containing proteins. In contrast, ALAS2, is an erythroid-specific enzyme, the synthesis of which is developmentally regulated and is markedly increased during erythropoiesis to accommodate the increased demand for heme during hemoglobin production. Previous studies have demonstrated that expression of the ALAS2 gene is regulated at both the transcriptional and post-transcriptional levels.

During erythropoiesis, transcription of the ALAS2 gene is markedly upregulated together with the genes for the other heme pathway enzymes and for globin. The human ALAS2 gene consisting of 11 exons spanning 22 kb, and extensive 5' (10 kb) and 3' flanking (3 kb) sequence was previously isolated and sequenced in our laboratory. Comparison of the genomic organisation of the human and murine ALAS2 genes exhibit remarkable similarity. The work presented in this thesis is concerned with the identification of regulatory regions responsible for the transcriptional control of the human ALAS2 gene in erythroid cells. A number of regions of the human ALAS2 gene which corresponded to DNase I hypersensitivity sites in the murine ALAS2 gene were examined in transient expression studies performed in erythroid cell lines.

Deletion analysis of the 5'-flanking region of the human ALAS2 gene demonstrated that the first 300 bp upstream of the transcription initiation site could function as a promoter in erythroid cells. Putative binding sites for the erythroid-enriched transcription factors, GATA-1, NF-E2 and the CACCC box binding protein, EKLF, and ubiquitous transcription factors were identified in the ALAS2 promoter. Gel shift assays were performed to determine the proteins binding to these sites and the functional contribution of these sites evaluated by site-directed mutagenesis and transient expression analysis. However, ALAS2

promoter constructs containing -293 bp to -10.3 kb of 5'- flanking sequence also expressed efficiently in non-erythroid COS-1 cells indicating that additional sequence or native chromatin structure may be required for the erythroid-specific expression of the human ALAS2 gene *in vivo*.

Examination of other regions of the human ALAS2 gene which corresponded to DNase I hypersensitivity sites in the murine ALAS2 gene, resulted in the identification of an erythroid-specific enhancer element within intron 8, located approximately 16 kb downstream from the transcription initiation site. This enhancer element was capable of transactivating both the human ALAS2 promoter and the heterologous thymidine kinase promoter in erythroid cells. Sequence analysis of intron 8 revealed putative GATA and CACCC box motifs which are highly conserved between the human, murine and canine ALAS2 genes. Mutational analysis established that a combination of two CACCC boxes and one GATA-1 binding site was essential for activity of the erythroid-specific enhancer.

The association of GATA-1 and CACCC box binding sites have previously been reported to play a critical role in the tissue-specific expression of many other erythroid-specific genes, including the enzymes of the heme biosynthetic pathway and the globin gene cluster. The identification of functional GATA-1 and CACCC box binding sites in the human ALAS2 promoter and in the erythroid-specific enhancer element located within intron 8, highlights the importance of these transcription factors in the erythroid-specific regulation of erythroid-cell expressed genes. Consequently, an interaction of a common group of transcription factors within the regulatory regions of the genes encoding the enzymes of the heme biosynthetic pathway and the globin gene cluster would ensure the coordinated synthesis of heme molecules and globin chains during erythropoiesis.

DECLARATION

This thesis contains no material that has been accepted for the award of any degree or diploma by any other University. To the best of my knowledge it contains no material that has been previously published by any other person, except where due reference has been made in the text. I consent to this thesis, when deposited in the University library, being available for photocopying and loan.

Katharina Helen Surinya

ACKNOWLEDGEMENTS

ABBREVIATIONS

Abbreviations used throughout this thesis are in accordance with those described in The Journal of Biological Chemistry (1994). Additional abbreviations are listed below.

ALA:	5-aminolevulinate
ALAS:	5-aminolevulinate synthase
ALAS1:	housekeeping 5-aminolevulinate synthase
ALAS2:	erythroid 5-aminolevulinate synthase
BKLF:	Basic Krüppel-like factor
bp:	base pair(s)
Ci:	Curie(s)
cpm:	counts per minute
DMEM:	Dulbecco's Modified Eagles Medium
dNTP:	deoxyribonucleoside triphosphate
DTT:	dithiothreitol
EKLF:	Erythroid Krüppel-like factor
Epo:	erythropoietin
GST:	glutathione S-transferase
Hb:	Hemoglobin
Hepes:	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HS:	hypersensitivity site
IRE:	iron responsive element
IRE-BP:	iron responsive element binding protein
kb:	kilobase(s)
LCR:	Locus Control Region
MEL:	murine erythroleukemia
Me_2SO:	dimethyl sulfoxide
PBGD:	Porphobilinogen deaminase
PBS:	phosphate-buffered saline
PLP:	pyridoxal phosphate
PMSF:	phenylmethylsulphonyl fluoride

RNase: ribonuclease

rpm: revolutions per minute

RSV: Rous sarcoma virus

TBP: TATA binding protein

TfR: transferrin receptor

TK: thymidine kinase

Tris: Tris(hydroxymethyl) aminomethane

UTR: untranslated region

V: Volt

XLSA: X-linked sideroblastic anemia

CHAPTER ONE

INTRODUCTION

CHAPTER ONE : GENERAL INTRODUCTION

1.1   INTRODUCTION

Heme is comprised of a ringed tetrapyrrole, protoporphyrin IX, with a central iron atom. This central ferrous iron atom (Fe^{2+}) of heme can be reversibly oxidised to the ferric state (Fe^{3+}) by the transfer of a single electron. This characteristic of heme confers its ability to function as an active cofactor in both the electron transport chain and in reactions involving oxygen-containing compounds. Heme, in the ferrous state, has a high affinity for oxygen allowing it to function as a carrier in the transport of oxygen. In erythroid cells, hemoglobin is the oxygen transport protein and is comprised of a tetramer of two α- and two β- globin polypeptide chains incorporating a single heme molecule.

The requirement for heme by the respiratory cytochromes and other heme-containing proteins necessitates its synthesis in all tissues. The largest quantities of heme are required during erythropoiesis for assembly into hemoglobin. Consequently, both intracellular levels of iron, *de novo* protoporphyrin IX synthesis and the synthesis of globin chains must be coordinately regulated. This coordinated expression is believed to be mediated at two levels: transcriptionally at the onset of erythroid differentiation, and translationally, that predominates following enucleation (maturation) of the reticulocyte. Therefore a common regulatory mechanism is thought to be involved in regulating the expression of the heme biosynthetic enzymes and globin polypeptides. Although, the regulation of the globin genes and the transcription factors responsible for their erythroid-specific expression have been extensively characterised, relatively little is known about the factors controlling expression of the heme biosynthetic enzymes.

5-Aminolevulinate synthase (ALAS) is the first enzyme of the heme biosynthetic pathway and exhibits the lowest relative enzyme activity in both erythroid and non-erythroid tissues in comparison to the other enzymes of this pathway (May *et al.*, 1995). Therefore ALAS is generally accepted to be rate-limiting and represents the key regulatory enzyme in the synthesis of heme in all tissues (May *et al.*, 1995). There are two isozymes for ALAS: a housekeeping form expressed in all cells (ALAS1), and an erythroid-specific form (ALAS2) responsible for the higher protoporphyrin synthesis in these cells (reviewed by May *et al.*,

1995) and each is encoded by a separate gene (Sutherland *et al.*, 1988; Cox *et al.*, 1990; Bishop *et al.*, 1990). The aim of this thesis is to investigate how expression of the ALAS2 gene is specifically regulated at the transcriptional level in erythroid cells. In order to appreciate the importance of ALAS in the heme biosynthetic pathway and therefore the reasons for pursuing these studies, an overview of erythroid differentiation and the current understanding of the regulation of heme and globin gene synthesis is presented and discussed.

1.2   HEMATOPOIESIS

Hematopoiesis is the process by which mature blood cells of distinct cell lineages, including lymphoid cells, erythrocytes, monocytes/macrophages, neutrophils, basophil/mast cells and megakaryocytes, are produced from pluripotent stem cells (Godin *et al.*, 1993). Hematopoiesis is controlled by the stimulation of pluripotent stem cells by growth factors and regulatory nuclear transcription factors involved in the activation of lineage-specific genes (Orkin, 1996). These stem cells must be produced continuously throughout the life span of vertebrates and must be capable of self-maintenance, self-renewal to generate additional hematopoietic stem cells and the production of progenitor cell types (Orkin, 1996). The distinct cell lineages reflect the genes expressed within these cell types. Examination of the transcription factors responsible for the regulation of these genes has provided a useful model for investigating the regulatory mechanisms governing the generation of distinct differentiated cell types from a common progenitor cell. Critical regulatory factors have been identified through the investigation of nuclear factors binding to *cis*-acting regulatory elements in lineage-specific expressed genes, analysis of genes aberrantly activated in leukemias and experimental models established by targeted gene disruption.

The location of hematopoiesis changes during vertebrate embryogenesis. In the mammalian embryo, hematopoietic stem cells initially appear in the blood islands of the yolk sac known as primitive hematopoiesis and in humans are detected at approximately 3 weeks of gestation. The site of hematopoiesis changes from the yolk sac in the early embryo to the fetal liver at approximately 5 weeks of development. Although it was initially thought that

the hematopoietic stem cells migrated from the yolk sac to the fetal liver, recent evidence has suggested that these cells may be derived from an intra-embryonic origin analogous to the mesodermal region in the avian and amphibian systems which gives rise to definitive adult hematopoietic stem cells (Godin *et al.*, 1993; Medvinsky *et al.*, 1993). The final site of hematopoiesis is located in the bone marrow where hematopoietic stem cells migrate from the fetal liver to the bone marrow and is the major site of hematopoiesis in the adult (definitive hematopoiesis) and in humans, this final shift occurs during the third and final trimester (Karlsson & Nienhuis, 1985).

1.3   ERYTHROPOIESIS

The differentiation and maturation of circulating erythrocytes from pluripotent hematopoietic progenitor cells is a process termed "erythropoiesis". The erythroid progenitor cells and circulating erythrocytes have been termed the "erythron" to suggest that they function as an organ (Bessis *et al.*, 1983). As described in Section 1.2, the bone marrow is the major site of erythropoiesis in the adult. Within the marrow, stromal macrophages provide discrete niches known as "erythroblastic islands" where the complete process of erythroid differentiation occurs (Bessis *et al.*, 1983).

Following the commitment of pluripotent stem cells to the erythroid lineage, erythroid progenitor cells progress through a series of morphologically distinct replicative stages (see Fig. 1.1). Throughout this progression of maturation, these cells become increasing specialised in their function to the point where the primary role of the erythrocyte (the end stage of erythropoiesis) is the transport of oxygen around the body.

The glycoprotein erythropoietin, Epo, (MW 30.4 kDa) is synthesised predominantly by the fetal liver and the adult kidney and is the major regulator of erythropoiesis (Jelkmann, 1992; Porter *et al.*, 1993). Epo regulates the synthesis of erythrocytes by stimulating the proliferation and differentiation of erythroid progenitor cells by a specific cell-surface receptor known as the erythropoietin receptor (EpoR), a member of the cytokine receptor family. As stimulation of progenitor cells by Epo must occur to enable differentiation along the erythroid lineage it is not surprising that expression of this receptor is one of the earliest markers of erythropoiesis (Penny & Forget, 1991; Wickrema *et al.*, 1992). Commitment to

the erythroid lineage is associated with a dramatic increase in expression of the EpoR gene resulting in an increase in the number of EpoRs on the cell surface (Herberlein *et al.*, 1992). The binding of Epo to its receptor induces the formation of homodimers (Watowich *et al.*, 1992) and is an essential step in signal transduction (Watowich *et al.*, 1994). The activation of the EpoR results in its rapid auto-phosphorylation (Dusanter-Fourt *et al.*, 1992) as well as that of the cytoplasmic tyrosine kinase JAK2, a member of the Janus kinase (JAK) family of tyrosine kinases (Witthuhn *et al.*, 1993; Barber & D'Andrea, 1994). Activated JAK2 phosphorylates both the EpoR and cellular substrates involved in the regulation of gene transcription (Ihle *et al.*, 1994). The binding of Epo to its receptor also induces the tyrosine phosphorylation of mitogen-activated protein (MAP) kinases (Miura *et al.*, 1994) which are involved in the phosphorylation of transcription factors, membrane proteins, cytoskeletal and other protein kinases (Blenis, 1993; Crews & Erikson, 1993).

The earliest progenitor cells committed to the erythroid lineage are the erythroid burst forming units (BFU-E) and are defined by their ability to create a "burst" consisting of several hundred cells when plated on semi-solid media (Bull *et al.*, 1983). BFUs-E are stimulated by interleukin-3 (IL-3) and granulocyte-macrophage colony-stimulating factor (GM-CSF) leading to their proliferation and differentiation into erythroid progenitor cells (Metcalf *et al.*, 1986; Lopez *et al.*, 1987). As differentiation proceeds, the late progenitor, the colony forming unit-erythroid (CFU-E), develops and is stimulated to differentiate into the proerythroblast by Epo (Fig. 1.1). Epo has been shown to be vital for the viability, proliferation and differentiation of the early erythroid progenitor BFUs-E and CFUs-E (Jelkmann, 1992; Lim & Wintour, 1993). Subsequent to these steps, erythroid maturation occurs by a series of four mitotic divisions during which the proerythroblast passes through the morphologically defined stages of the basophilic erythroblast, the polychromatophilic erythroblast and the orthochromatic erythroblast after which ejection of the nucleus and emergence of the reticulocyte occurs in a period of 72 hours (Fig. 1.1). During the erythroid maturation period, the specialised structures required for the synthesis of hemoglobin, enzymes and structural proteins including the mitochondria, Golgi apparatus, polyribosomes and siderosomes are detected. All of the enzymes of the heme biosynthetic pathway are detectable during this period (Fig. 1.1) and, as discussed in Section 1.11, there is evidence suggesting that heme itself, plays an essential role in the differentiation of the erythroid lineage (Fukuda *et al.*, 1993; 1994). Towards the end of the erythroid maturation period,

Figure 1.1   **Synthesis of Heme and Hemoglobin during erythropoiesis in the adult bone marrow**

Erythropoietin (Epo) stimulates the differentiation of the proerythroblast from BFU-E and CFU-E, as well as the progression from the proerythroblast stage through four mitotic divisions leading to the formation of the reticulocyte which is released into the circulation by "diapedesis". Following the stimulation of the proerythroblast by Epo, there is an appearance in all of the enzymes of the heme biosynthetic pathway. The induction of heme synthesis promotes the synthesis and assembly of hemoglobin.

BFU-e/CFU-e

Erythropoietin

Proerythroblast

appearance of heme
pathway enzymes

globin mRNA synthesis
and translation begins

hemoglobin assembly

MARROW

diapedesis

Reticulocyte

BLOOD

Erythrocyte
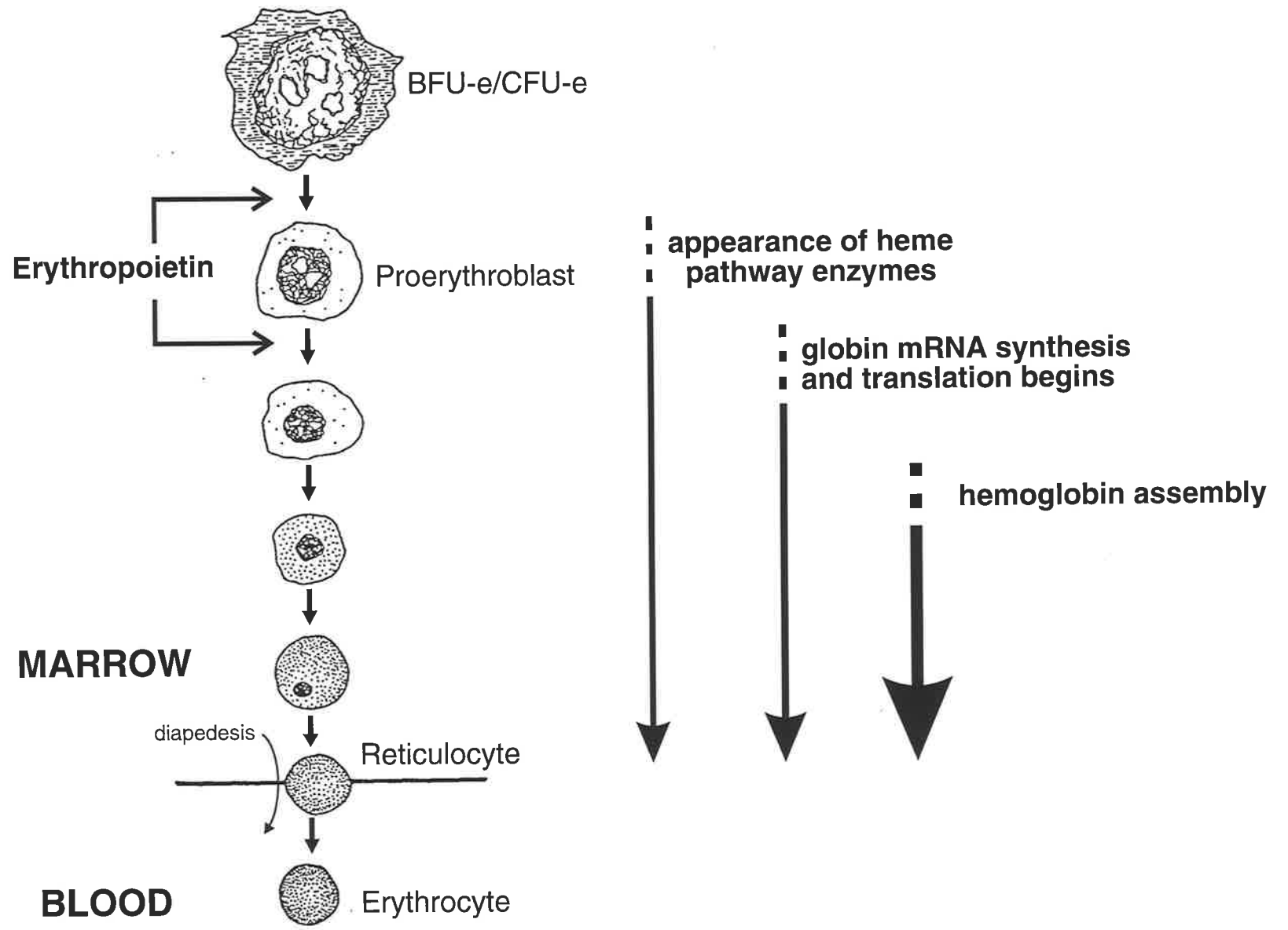
the nuclei of the erythroblast cells condense, overall protein synthesis decreases, organelles (including mitochondria) begin to degenerate and the diameter of the erythroblast is significantly reduced as it develops from the proerythroblast (25µm) to the orthochromatic erythroblast (9µm). Progressively, these cells exhibit a decreased dependence for stimulation by Epo and this is also accompanied by a down-regulation in the expression of the EpoR (Landschulz *et al.*, 1989). As a result of unequal cell division of the erythroblast, the nucleus is phagocytosed and degraded by perisinal marrow macrophages, and the reticulocyte is then released into the marrow capillary and the circulation in a process known as "diapedesis" (Fig. 1.1). The final stage of erythroid maturation occur in the circulation over a period of approximately 48 hours and involves the synthesis of the final 20% of hemoglobin in the erythrocyte and the ejection of the remaining organelles, including mitochondria. Therefore as a result of the four mitotic divisions, 16 circulating reticulocytes are generated per original proerythroblast (Bessis *et al.*, 1983; Lewis, 1990). Approximately 90% of the total protein in the mature erythrocyte is comprised of hemoglobin highlighting the primary functional role of this cell in the transport of oxygen. It has been estimated that approximately $2x10^{11}$ erythrocytes enter the circulation each day, where they remain for a period of up to 120 days and form the major proportion of cells in the circulation (Bessis *et al.*, 1983; Eckardt & Bauer, 1989).

1.3.1   *In vitro* Models of Erythroid Differentiation

The development of transformed erythroid cell lines which can be induced to differentiate *in vitro* and mimic the changes seen during erythropoiesis, provide an excellent model system in which to investigate the regulatory processes that occur during the latter stages of erythroid differentiation. The most commonly used *in vitro* models are the Friend virus-transformed murine erythroleukemic (MEL) cells (Friend *et al.*, 1971) which are considered to represent immature erythroid cells arrested at the proerythroblast stage. These cells can be induced to undergo terminal erythroid differentiation by the addition of a variety of agents including dimethyl sulfoxide (Me$_2$SO) (Friend *et al.*, 1971), hexamethylenebisacetamide (HMBA) (Reuben *et al.*, 1976) and hemin (ferric protoporphyrin IX) (Granick *et al.*, 1978). Treatment of MEL cells with Me$_2$SO exhibits many characteristics of the natural erythropoietic differentiation process over a period of

3 to 6 days. These include a sequential increase in the enzymatic activities of the heme biosynthetic pathway (Sassa, 1975) with genes being transcriptionally activated (Beaumont *et al.*, 1984; Romeo *et al.*, 1986; Elferink *et al.*, 1988; Raich *et al.*, 1989) and the accumulation of α- and β- globin mRNAs following an increase in their rates of gene transcription (Ganguly & Skoultchi, 1985) resulting in hemoglobin synthesis.

In contrast to the MEL cell line which differentiates in response to chemical agents, the murine J2E erythroid cell line can be induced to differentiate with the hormone erythropoietin (Epo), the natural stimulator of erythroid differentiation *in vivo* and closely resembles the natural process of erythropoiesis. This cell line was generated following the *in vitro* infection of murine fetal liver cells with a retrovirus containing *raf* and *myc* oncogenes, which inhibit the spontaneous maturation of the erythroid cells (Klinken *et al.*, 1988). These J2E cells are immortalized at the proerythroblast stage of erythroid differentiation and, following exposure to Epo, undergo a series of morphological and biochemical changes including the transcriptional activation of the genes encoding enzymes of the heme biosynthetic pathway and the globin genes, resulting in the synthesis of hemoglobin, a reduction in cell volume and the spontaneous ejection of the condensed nucleus over a 3 day period (Busfield & Klinken, 1992).

The human leukemic cell line K562, originally established from the pleural effusion of a patient with chronic myelogenous leukemia (CML) in blast crisis (Lozzio *et al.*, 1975), has the potential to differentiate along the erythroid, megakaryocytic or myeloid cell lineages depending on the chemical inducer. Erythroid differentiation of the K562 cell line can be induced by sodium butyrate (Andersson *et al.*, 1979) or hemin (Rutherford *et al.*, 1979; Benz *et al.*, 1980) and cells stained positive for hemoglobin synthesis.

1.3.2   Regulation of the Synthesis of Erythropoietin (Epo)

As mentioned previously, Epo is synthesised predominantly by the fetal liver and the adult kidney, and is produced in response to tissue hypoxia; that is, low oxygen levels (Graber & Krantz, 1989) and can also be induced in animals treated with cobalt (II) chloride (Goldwasser *et al.*, 1958). An undefined "oxygen sensory mechanism" is thought to be located within the kidney which is responsible for monitoring the levels of oxygen, influencing the synthesis of Epo and ultimately determining the extent of erythropoiesis

(reviewed by Eckardt & Bauer, 1989). Goldberg *et al.* (1988) proposed that the oxygen sensing protein may be a hemoprotein, since attempts to interfere with the synthesis of heme affects the synthesis of Epo. They have proposed that the binding of oxygen to the heme moiety of the hemoprotein would cause a conformational change and alter activity of the hemoprotein. Therefore at low oxygen levels, the hemoprotein would be functionally active resulting in increased Epo gene transcription. Conversely, at high oxygen levels, the hemoprotein would be in an inactive conformation and Therefore would not activate Epo gene transcription. Although a hemoprotein which is involved in controlling oxygen-related gene expression in the nitrogen-fixing bacteria, *Rhizobium meliloti* (Gilles-Gonzalez *et al.* 1991) and in yeast (Hodge *et al.*, 1989) have been isolated, a functionally equivalent protein has not yet been identified in eukaryotes.

The synthesis of Epo is induced in the human hepatoma cell lines, HepG2 and Hep3B in response to hypoxia or treatment with cobalt chloride (Goldberg *et al.*, 1987; 1988) and is regulated primarily at the level of gene transcription (Schuster *et al.*, 1989; Goldberg *et al.*, 1991). The use of these cell lines has provided an extremely valuable *in vitro* system in which to investigate the mechanism responsible for the activation of Epo gene expression. The promoter (Blanchard *et al.*, 1992) and the 3'-flanking sequence (Pugh *et al.*, 1991; Blanchard *et al.*, 1992; Semenza *et al.*, 1992) of the Epo gene have been implicated to play a role in the hypoxia-mediated induction of the Epo gene. Transient expression studies performed by Blanchard *et al.* (1992) demonstrated a cooperative interaction between these two regions of the Epo gene. The minimal sequence of the Epo promoter required for hypoxia-mediated induction of a reporter gene was localised to a region -118 bp to -65 bp upstream of the transcription initiation site (Blanchard *et al.*, 1992). Deletion analysis of the 3'-flanking sequence of the Epo gene resulted in the identification of a 43 bp enhancer element located 120 bp downstream of the poly(A) tail site of human Epo gene (Blanchard *et al.*, 1992). A hypoxia-induced DNA binding protein designated HIF-1 (120 kDa) binds specifically to the 3'-enhancer of the Epo gene (Semenza *et al.*, 1992) and is activated in a range of hypoxia-induced cell types (Wang *et al.*, 1993). HIF-1 binding sites have also been identified in several hypoxia-inducible genes encoding glycolytic enzymes and DNA sequences encompassing these binding sites confer hypoxia-inducible reporter gene transcription (Semenza *et al.*, 1994). Therefore these results suggest that HIF-1 not only functions as a physiological regulator of Epo gene

transcription but may play a central role in the transcriptional regulation of hypoxia-responsive genes.

The Epo promoter and 3'-enhancer also contain functional hexanucleotide consensus nuclear receptor binding half sites. Transient transfection studies have demonstrated that the orphan nuclear receptor, HNF-4, which binds an unknown ligand, plays a critical role in the hypoxia-inducible expression of the Epo gene and that intact binding sites in both the promoter and 3'-enhancer are required (Galson *et al.*, 1995). Interestingly, mutational analysis suggested a possible protein-protein interaction between HIF-1 and HNF-4, both of these sites are required for enhancer activity and this is currently under investigation (Galson *et al.*, 1995).

In vivo analysis of the DNA-protein interactions on the human Epo enhancer showed that both HIF-1 and HNF-4 were bound under both normal and hypoxia-induced conditions (Hu *et al.*, 1997). However, under hypoxia conditions, a conformational change in the topography of the Epo enhancer was observed, suggesting a possible modification of the HIF-1 complex or a hormone-induced alteration of the receptor bound to the enhancer (Hu *et al.*, 1997).

1.4   SYNTHESIS OF HEMOGLOBIN

During erythropoiesis, large quantities of hemoglobin are synthesised and this requires the coordinated production of heme molecules and globin polypeptides. Similarly, the production of heme itself requires the coordinated regulation of *de novo* protoporphyrin synthesis with the modulation of intracellular levels of iron. The mechanisms responsible for iron uptake, storage and the maintenance of appropriate levels of intracellular iron are discussed here prior to the regulation of globin gene expression.

1.4.1   Regulation of the Intracellular Levels of Iron

Iron is critical for all living systems since it is an essential component of many important iron-containing proteins involved in oxidative respiration, DNA synthesis and cellular proliferation (Klausner *et al.*, 1993). Iron in conjunction with oxygen potentially results in the generation of harmful oxygen radicals that damage cells. The effect of long

term iron excess in the body results in massive cell and tissue damage as is evident in patients with hemochromatosis (Edwards *et al.*, 1991). Therefore it is crucial to control cellular levels of free iron yet remain capable of responding to the intracellular iron needs of the various specialised cell types in the body.

Environmental iron in the ferrous state (Fe^{2+}) is absorbed by the intestinal mucosal cells and then oxidised to ferric iron (Fe^{3+}) whereupon it is bound by transferrin, the major iron-binding protein found in the serum. The association of iron with transferrin overcomes the low solubility of iron and the generation of harmful oxygen radicals. The level of expression of specific receptors for the iron-transferrin complex (transferrin receptors) on the cell surface dictates the level of targeting those cells with a requirement for iron uptake and utilisation (reviewed by Harford *et al.*, 1991). This iron-transferrin-receptor complex is then internalised in coated vesicles by receptor-mediated endocytosis, the iron released within the endosomes following reduction to ferrous iron (Fe^{2+}) and then transported into the cytosol. Following the release of iron from the complexes, transferrin remains stably bound to the transferrin receptor. This complex is then exocytosed and returned to the cell surface where both transferrin and its receptor are recycled. Ferrous iron in the cytosol is either transported to the mitochondria by an unknown mechanism and utilised in the synthesis of heme or is stored in the cytosol within the multisubunit protein, ferritin. Ferritin shells are composed of 24 subunits of heavy (H) and light (L) chains which are involved in iron sequestration, detoxification and the mobilisation of excess intracellular iron (Aziz & Munro, 1987; Hentze *et al.*, 1987; Leibold & Munro, 1988; Casey *et al.*, 1988; Müllner *et al.*, 1989). The H and L chains of ferritin differ in their ability to mobilise the stored iron. Consequently, the proportion of H and L chains in the shells varies depending on the physiological conditions and the type of tissue (Arosio *et al.*, 1978). In erythroid cells, ferritin shells are comprised of a higher proportion of the H chain of ferritin which are thought to permit an increase in the traffic of intracellular iron (Beaumont *et al.*, 1989) whereas, the L chain of ferritin is utilised for the long term storage of iron and accumulates in the liver and spleen (Boyd *et al.*, 1985).

As mentioned, the number of transferrin receptors expressed on the cell surface regulates the amount of iron imported into cells. These transmembrane glycoproteins are expressed in almost all cell types, with a higher number of transferrin receptors expressed in differentiating erythroid cells because of their large requirement for iron during the synthesis

of heme. Intracellular levels of iron play an important role in regulating of the numbers of transferrin receptors expressed on the cell surface, such that uptake is only proportional to cellular demand. Treatment of cells with heme (Pelicci *et al.*, 1982; Ward *et al.*, 1984) or ferric ammonium citrate (Ward *et al.*, 1982) results in elevated levels of iron and a decrease in the number of transferrin receptors. Conversely, depletion of intracellular iron with iron chelators such as desferrioxamine results in a rapid increase in the number of transferrin receptors (Bridges *et al.*, 1984; Mattia *et al.*, 1984; Rao *et al.*, 1985). Therefore by inversely regulating the expression of the transferrin receptor (involved in iron uptake) and positively that of ferritin (involved in mobilising excess intracellular iron), it is obvious that cells would have the ability of rapidly responding to the fluctuating levels of intracellular iron (Harford *et al.*, 1991).

In non-erythroid cells, the expression of the transferrin receptors and ferritin are both regulated by the levels of intracellular iron at the post-transcriptional level through iron-responsive elements (IREs) located in the mRNAs. The IRE consists of a secondary stem loop structure with the loop containing the consensus sequence 5'-CAGUGN-3', an unpaired cytosine residue located five nucleotides 5' of this loop and an additional base paired stem structure (Casey *et al.*, 1988). A single IRE has been identified in the 5'-UTR of the ferritin H and L subunit mRNAs whereas five IREs are located in the 3'-UTR of the transferrin receptor mRNA (Klausner *et al.*, 1993).

IREs are bound by the cytosolic iron-responsive element binding protein (IRE-BP, MW 90 kDa) whose affinity for the stem loop structure is directly modulated by the level of free iron (Rouault *et al.*, 1988). The IRE-BP shares sequence homology to the mitochondrial enzyme, aconitase, which catalyses the conversion of citrate to isocitrate in the second step of the tricarboxylic acid (TCA) cycle (Rouault *et al.*, 1991; Hentze & Argos, 1991; Lauble *et al.*, 1992). Like the IRE-BP, aconitase contains an iron-sulfur cluster that plays a critical role in the regulation of the enzyme's activity (Kim *et al.*, 1996a), the labile fourth iron atom of this iron-sulfur clusters acting as the "iron-sensing centre" (Rouault *et al.*, 1991; Lauble *et al.*, 1992; Klausner *et al.*, 1993). At low levels of iron, the cytosolic IRE-BP specifically binds to the single IRE in the 5'-UTR of the ferritin mRNA and the five IREs in the 3'-UTR of the transferrin receptor mRNAs with high affinity, inhibiting translation of ferritin mRNA and increasing the stability of transferrin receptor mRNA (Theil *et al.*, 1990). The binding of the IRE-BP to the IRE in the 5'-UTR

10

of the ferritin mRNA is thought to sterically prevent the formation of a stable translation initiation complex at the 7-methylguanine (m^7G)-cap structure since an IRE only functions when positioned within approximately 60 nucleotides from the cap (Goossen *et al.*, 1990; Goossen & Hentze, 1992; Bhasker *et al.*, 1993). Conversely, at high levels of iron, the IRE-BP exhibits a reduced binding affinity for the IREs in the ferritin and transferrin receptor mRNAs, resulting in an increase in the synthesis of ferritin and a decrease in the synthesis of transferrin receptor (Philpott *et al.*, 1993).

Although, expression of the transferrin receptor and ferritin genes are regulated at the post-transcriptional level in non-erythroid cells, it has been questioned whether there is a similar mechanism responsible for the regulation of these genes in immature erythroid cells as these cells account for approximately 70% of the total iron utilised in the body.

The expression of transferrin receptors is developmentally regulated during erythroid maturation, with receptor numbers markedly increasing at the early phase of erythroid development and then declining at the terminal stages of erythroid differentiation to the point where they are absent in circulating erythrocytes (Nunez *et al.*, 1977; Iacopetta *et al.*, 1982; Horton *et al.*, 1983). Northern blot analysis of Me_2SO-induced MEL and Epo-induced J2E cell lines have also demonstrated a rapid increase in the level of the transferrin receptor mRNA with differentiation (Fujita *et al.*, 1991; Busfield *et al.*, 1992). This increase in transferrin receptors precedes the incorporation of iron into heme yet coincides with the induction of β-globin gene transcription (Bondurant *et al.*, 1985). However, studies performed in Me_2SO-induced MEL cells (Chan *et al.*, 1994), Epo-induced J2E cells (Busfield *et al.*, 1993) and differentiating chick embryonic erythroid cells (Chan *et al.*, 1992) have demonstrated that the increase in the number of transferrin receptors occurred primarily as a result of an increase in gene transcription. Although, an increase in the stability of the transferrin receptor mRNA was observed in both Me_2SO-induced MEL cells (Chan *et al.*, 1994) and J2E cells induced to differentiate with Epo (Klinken *et al.*, 1993), this appeared to be independent of iron and the IRE-BP.

Analysis of the promoter and the intron 1 sequence of the mouse and chicken transferrin receptor genes revealed numerous putative binding sites for GATA-1, CACCC box binding proteins and AP-1 which may be involved in the erythroid-specific upregulation of the transferrin receptor gene (Chan & Shao, 1993) (see Section 1.5). However, the

functional role of these putative binding sites in the transcriptional regulation of the transferrin receptor gene during erythroid differentiation remains to be determined.

In contrast to the transferrin receptor whose expression is regulated primarily at the transcriptional level in erythroid cells, evidence exists to suggest that expression of ferritin is regulated at both the transcriptional and post-transcriptional level. Expression of the ferritin H chain promoter in heme-treated Friend leukemia cells resulted in the identification of a minimal region capable of conferring high level erythroid-specific expression. The elements identified within this region included two binding sites for Sp1 and a CCAAT box binding site (Marziali *et al.*, 1997). The binding site for the heme-responsive factor corresponded to the functional CCAAT box binding site and bound the ubiquitous factor, NF-Y, which is composed of at least three subunits (A, B and C) (Coustry *et al.*, 1995; Kim *et al.*, 1996b). Transient transfection studies performed with a dominant negative mutant of NF-YA that was capable of sequestering the NF-YB subunit and inhibiting binding to DNA, abolished the heme-induced transcription of the ferritin H chain promoter indicating the critical role of NF-Y in the activation of this promoter. Interestingly, expression of NF-YA itself increased during the maturation of heme-induced erythroid cells (Marziali *et al.*, 1997). In addition to the functional role of the CCAAT box binding site in the transcriptional activation of the ferritin H chain gene, an enhancer region has also been identified approximately 4.5 kb upstream from the transcription initiation site which is responsible for the increased transcription of this gene during erythropoiesis but the identity of the *trans*-acting factors that bind to this region have not yet been determined (Beaumont *et al.*, 1994a; 1994b).

The mechanism responsible for the post-transcriptional regulation of ferritin in non-erythroid cells is also thought to occur during erythropoiesis, since translation of the ferritin mRNA can also be regulated by iron in Me_2SO-induced MEL cells (Melefors *et al.*, 1993).

1.4.2 The Globin Gene Cluster

The globin polypeptides are encoded by genes in two clusters, α and β and each cluster is composed of a small family of genes. The expression of the globin genes is selectively activated at a specific developmental stage and then subsequently repressed following that stage in a process known as "developmental switching" involving the

alternate expression of embryonic, fetal and adult globins. As a result, different hemoglobin tetramers are assembled within erythroid cells at different developmental stages (Orkin, 1995a) and this coincides with a change in the location of erythropoiesis during vertebrate embryogenesis.

In mammals and avians, the α- and β- globin gene clusters are encoded on separate chromosomes but are located within a single region in frogs and fish (as reviewed by Orkin, 1995a). The human α- and β- globin gene clusters are located on chromosome 16 and 11, respectively. The α-globin gene cluster is composed of three functional genes, where the embryonic (ζ) gene is located upstream of two duplicated adult α-globin genes encompassing approximately 30 kb (Fig. 1.2). The β-globin gene cluster consists of five functional genes: an embryonic (ε) gene, two duplicated fetal (γ) genes, and the minor δ- and major adult β- globin genes spanning approximately 60 kb (Fig. 1.2).

As mentioned in Section 1.2, the yolk sac is the first site of erythropoiesis and hemoglobin production. These first hemoglobins consist of tetramers composed of two ζ -globin or two α-globin polypeptides and two ε-globin polypeptides, designated $\zeta_2\varepsilon_2$ and $\alpha_2\varepsilon_2$, respectively. Expression of the ζ-globin gene gradually decreases while expression of the α-globin gene increases. The change in the site of erythropoiesis from the yolk sac to the fetal liver is accompanied by a decrease in ε-globin gene expression and the activation of the fetal γ-globin gene, the major hemoglobin synthesised at this stage is fetal hemoglobin ($\alpha_2\gamma_2$). As the final site of erythropoiesis changes from the fetal liver to the bone marrow, this stage is accompanied by a final switch in the expression of the globin genes. Expression of the fetal γ-globin gene gradually decreases and expression of the δ- and β- globin genes increase until approximately 98% of the total hemoglobin synthesised is hemoglobin A ($\alpha_2\beta_2$) and approximately 1% is hemoglobin A2 ($\alpha_2\delta_2$). However, a minor population of erythroid cells which develop in the bone marrow continue to express the γ-globin gene and these represent approximately 1% of total hemoglobin in adult blood (Stamatoyannopoulos *et al.*, 1994; Donze *et al.*, 1995).

Figure 1.2   Structural Organisation of the α- and β- globin gene clusters

The individual globin genes of each locus which are sequentially activated during development are indicated. The erythroid-specific DNase I hypersensitivity site of the α-globin locus, HS-40 is located upstream of the embryonic (ζ) globin gene. The DNase I hypersensitivity sites (HS) (*arrowed*) are located upstream of the embryonic (ε) globin gene in the β-globin locus. HS1-4 are erythroid-specific and comprise the Locus Control Region (LCR) whereas HS5 is constitutive.

α-globin gene cluster

HS-40

ζ_2 α_2 α_1

β-globin gene cluster

LCR

HS5 HS4 HS3 HS2 HS1

Embryonic   Fetal   Adult

ε Gγ Aγ δ β

Expression: Yolk Sac   Fetal liver   Bone marrow
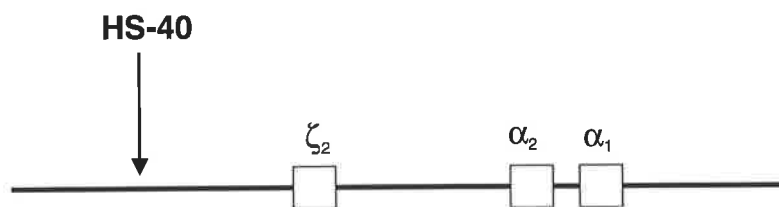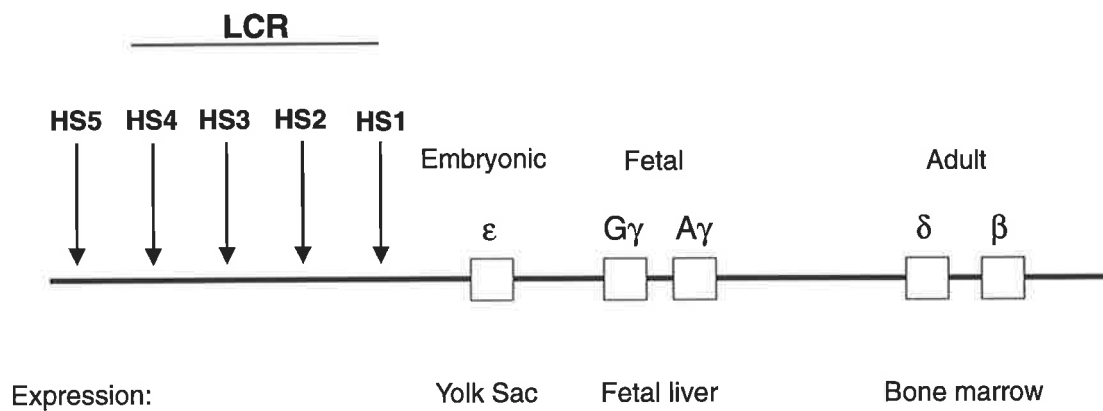
1.4.3 Regulation of Globin Gene Expression

Expression of the genes from the α- and β- globin gene clusters must be coordinately regulated, since an excess of either type of globin chain can be cytotoxic (Weatherall, 1994). Extensive research has therefore concentrated on the molecular mechanisms responsible for the specific transcriptional activation of these genes in developing erythroid cells, and the switching of globin gene expression observed during development. Consequently, the identification of the critical factors responsible for the high level expression of the globin genes during erythroid differentiation has provided a model to understand the mechanisms responsible for both the erythroid- and tissue- specific gene expression in eukaryotes.

Initial studies which attempted to identify the *cis*-elements responsible for erythroid-specific expression of β-globin genes in erythroid progenitor cells focused on DNA sequences residing upstream or downstream of the genes. Preliminary analysis of the human β-globin sequences in transgenic mice showed extremely low but erythroid-specific expression that was subject to extensive variation among transgenic lines. This variation was attributable to the site of integration of the construct (Magram *et al.*, 1985; Townes *et al.*, 1985) and suggested that critical *cis*-regulatory elements were absent from the immediate 5'-flanking region.

DNase I hypersensitivity studies, which serve as an indicator of chromatin structure, were performed in the mid-1980s and resulted in the identification of four erythroid-specific DNase I sensitive regions (5'HS 1-4), located approximately 11 to 18 kb upstream of the ε -globin gene (Tuan *et al.*, 1985; Forrester *et al.*, 1986). These have collectively been termed the Locus Control Region (LCR) (Fig. 1.2). An additional DNase I hypersensitive site present in both erythroid and non-erythroid tissues was identified upstream of these four erythroid-specific sites. Grosveld *et al.* (1987) linked the four erythroid-specific hypersensitivity sites upstream of the human β-globin gene and created transgenic mice. Analysis of the expression of the human β-globin transgene in these mice revealed a high-level of expression which was erythroid-specific, and independent of the site of integration. The LCR is also capable of inducing high level tissue-specific expression of the α-globin gene (Hanscombe *et al.*, 1989) and heterologous promoters such as the Herpes thymidine kinase promoter linked to a G418R gene (Blom-van Assendelft *et al.*, 1989).

Naturally occurring deletions which eliminated this hypersensitivity region but left the globin genes intact, were associated with a failure to activate the intact *cis*-linked genes of the β-globin gene cluster in either fetal or adult erythroid cells (Driscoll *et al.*, 1989; Kioussis *et al.*, 1983), with a dramatic alteration in the chromatin structure spanning at least 200 kb downstream of the deletion and timing of DNA replication (Forrester *et al.*, 1990). In conclusion, the generation of transgenic mice and analysis of naturally occurring deletions of these erythroid-specific hypersensitivity sites, demonstrated the functional importance of these sequences in conferring high level erythroid-specific expression.

The LCR was initially proposed to serve a dual function. Firstly, the LCR "opens" a chromosomal domain of the entire β-globin locus specifically in erythroid cells to permit expression of downstream genes, and overcomes integration site position effects (Caterina *et al.*, 1994a). Secondly, the LCR is thought to function as an extremely powerful erythroid-specific enhancer that stimulates globin gene expression. However, evidence exists to dispute this model (Walters *et al.*, 1995; Martin *et al.*, 1996; Milot *et al.*, 1996; Sutherland *et al.*, 1997). A "binary" model has been proposed suggesting that the primary role of the LCR is to establish and possibly maintain domains permissive for transcription rather than functioning to increase the rate of gene transcription (Walters *et al.*, 1995).

The individual sites of the LCR of the β-globin locus have been assayed in a number of expression systems with varying results suggesting that these have different functional capabilities. The individual elements of the LCR (HS2, HS3 and HS4) are all capable of conferring high-level tissue-specific transcription when linked to human β-globin genes in transgenic mice (Curtin *et al.*, 1989; Ryan *et al.*, 1989; Ney *et al.*, 1990; Talbot *et al.*, 1990; Pruzina *et al.*, 1991; Talbot & Grosveld, 1991; Lloyd *et al.*, 1992; Morley *et al.*, 1992; Caterina *et al.*, 1994a). Fraser *et al.* (1990) demonstrated that the individual sites HS1, HS2 and HS3 provide position-independent, copy-number dependent expression and both HS2 and HS3 are required for full expression of the human β-globin gene in transgenic mice. Expression studies performed in stably transformed MEL cells have demonstrated that HS2 and HS3 are both capable of directing high levels of inducible β-globin expression while HS1 and HS4 give reduced levels of activity (Collis *et al.*, 1990).

The role of the HSs in the switching of globin gene expression during development has also been examined and each site shown to be functionally distinct. Fraser *et al.* (1993)

examined the effect of each HS on the developmental expression pattern of the human γ-
and β- globin genes in transgenic mice. The expression of HS1 is low but both HS1 and
HS2 express either the γ- or β- globin transgenes throughout development. However,
neither site is capable of driving γ-globin expression in the fetal liver. HS4 exhibits low
transcriptional activity during the embryonic and fetal stages but shows the highest level of
activity during the adult stage. In contrast, HS3 is the most active during the embryonic
period and preferentially activates the γ-globin gene rather than the β-globin gene in
erythrocytes derived from the yolk sacs and fetal livers of transgenic mice.

To investigate the functional roles of HS2 and HS3 *in vivo*, each site was deleted by
homologous recombination in murine embryonic stem cells and the phenotype of the mice
examined. The replacement of HS2 with the selectable marker gene, neomycin
phosphotransferase, resulted in a 2-5 fold reduction in the expression of most of the genes
in the locus, with a more dramatic reduction in expression (10-12 fold) observed with the
embryonic ε-globin gene (Fiering *et al.*, 1995). However, this mutation did not alter the
timing of expression of the globin genes during development. Hug *et al.* (1996) examined
the effect of a targeted deletion of the endogenous HS3 region in mice. Only a marginal
reduction in the expression of the embryonic globin gene, bh1 (which is developmentally
equivalent to the human γ-globin gene) was observed and expression of the β-globin gene in
adult animals was reduced by approximately 30% (Hug *et al.*, 1996). This finding suggested
that although HS3 is required for approximately 30% of the total LCR activity associated
with adult β-globin gene expression in adult erythrocytes it may not be be essential for
embryonic ε-globin gene expression or can be rescued by another HS within the LCR.

A positive regulatory element similar to the LCR of the β-globin gene cluster has
also been identified by erythroid-specific DNase I hypersensitivity mapping within the
human α-globin gene cluster (termed HS-40) (Higgs *et al.*, 1990) (Fig. 1.2). This element is
located approximately 40 kb upstream of the human embryonic ζ-globin gene and functions
as an enhancer element in both stable erythroid cell lines and in transiently transfected cells
during erythroid differentiation (Pondel *et al.*, 1992; Ren *et al.*, 1993; Sharpe *et al.*, 1993a).
Analysis of the HS-40 enhancer in transgenic mice has demonstrated that it is also capable
of facilitating expression of linked globin genes (Higgs *et al.*, 1990). In contrast to the β
-globin LCR, the HS-40 does not ensure copy-number dependent and position independent

expression (Sharpe *et al.*, 1993b). Gourdon *et al.* (1994) recently reported the successful introduction of a 70 kb fragment of DNA containing the human ζ- and α- globin genes linked to the HS-40 element in transgenic mice. However, expression of the human α -globin transgene was significantly lower in comparison to the endogenous α-globin gene in these transgenic mice. Consequently, it is possible that *cis*-acting elements located further upstream of HS-40 may be responsible for stable, copy number dependent expression of the α-globin gene cluster and have not been identified due to the technical difficulty of analysing large segments of DNA in transgenic mice.

Studies performed by Bernet *et al.* (1995) in which the HS-40 of the α-globin locus was inactivated by homologous recombination in a murine erythroleukemia hybrid cell line containing a single copy of human chromosome 16, demonstrated that the HS-40 was capable of functioning as a strong inducible enhancer of human α-globin gene expression. However, deletion of HS-40 did not appear to affect chromatin structure, which is in direct contrast to results obtained following deletion of the β-globin LCR.

The functional importance of HS-40 in α-globin gene expression is also apparent in the analysis of patients with α-thalassemias. In these patients, naturally occurring deletions upstream of the α-globin gene cluster which also remove the HS-40 region are responsible for the abolished activity of the α-globin genes (Hatton *et al.*, 1990; Liebhaber *et al.*, 1990).

The differences observed between the functional enhancers of the α- and β- globin gene clusters may be attributable to differences in chromatin context, but remains to be resolved (Vyas *et al.*, 1992; Raich *et al.*, 1993). Extensive research has focused on identifying the specific DNA sequences located within the enhancer elements responsible for the erythroid-specific expression of the α- and β- globin gene cluster. The functional enhancer activity of the LCR elements is thought to be mediated through their interaction with both cell-specific and ubiquitous nuclear proteins. These nuclear regulatory proteins which interact with the DNA sequences located in the LCRs and promoters of the globin genes have been characterised by *in vitro* DNA binding assays. Only three DNA binding sites: GATA, which binds the erythroid-enriched factor GATA-1 (Martin *et al.*, 1990), an AP-1-like motif, which binds the erythroid-enriched factor, NF-E2 (Andrews *et al.*, 1993a); and the CACCC/GTGGG motif; which binds Sp1 (Kadonaga *et al.*, 1987) and Sp1-related proteins (Hartzog & Myers, 1993), have been consistently identified within the core regions

of the LCRs (Ney *et al.*, 1990; Talbot *et al.*, 1990, Philipsen *et al.*, 1990; Strauss *et al.*, 1992; Zhang *et al.*, 1993). Interestingly, the GATA and CACCC sites, but not NF-E2, are also present in the globin promoters (deBoer *et al.*, 1988; Gong *et al.*, 1991) and all three sites are found in only a limited number of other erythroid-expressed gene promoters which have been examined including the genes for human erythroid PBGD (the third enzyme of the heme biosynthetic pathway) (Mignotte *et al.*, 1989a), ferrochelatase (Tugores *et al.*, 1994) and carbonic anhydrase II (Disela *et al.*, 1991).

Studies aimed at the identification of specific elements located within the LCR which are responsible for the erythroid-specific enhancer activity have revealed that an AP-1 motif of HS2 is essential (Ney *et al.*, 1990; Talbot *et al.*, 1991). The position-independent activity observed with HS2 (Ellis *et al.*, 1993) and HS3 (Philipsen *et al.*, 1993) are associated with GATA and CACCC sites. The formation of human HS4 requires the presence of both GATA and AP-1 like binding sites at a specific spacing (Stamatoyannopoulos *et al.*, 1995). Transient transfection studies performed by Walters *et al.* (1992) demonstrated that only a combination of GATA motifs with CACCC or AP-1 sites permitted the erythroid-specific expression of an artificial promoter construct.

Consequently, during erythroid differentiation, a common set of transcription factors may be involved in the coordinated synthesis of the heme pathway enzymes and globin, as well as other erythroid-specific genes. The predominant erythroid-enriched nuclear DNA binding proteins involved in this process (are summarised in Table 1) and are discussed in detail in the following section.

1.5   *TRANS*-ACTING FACTORS INVOLVED IN ERYTHROID-SPECIFIC GENE EXPRESSION

1.5.1   GATA-1

GATA-1 (formerly known as eryf1, NF-E1 and GF-1) was initially identified as a zinc-finger containing nuclear protein that binds the consensus sequence 5' (T/A)GATA(A/G) 3' in the promoters of erythroid-expressed genes (Orkin, 1995a; 1995b). It was the founding member of a family of GATA factors (GATA-1 to GATA-6), many of which were first demonstrated in chickens by Yamamoto *et al.* (1990). These proteins all recognise the core GATA consensus sequence but have distinguishable binding

Table 1   **Transcription factors and other regulatory proteins in erythroid cells**

A summary of the transcription factors involved in the regulation of erythroid cell-specific genes. The class of transcription factor, their DNA recognition motif and their composition are indicated.

Transcription Factor	Class	Binding Site
GATA-1	GATA finger	(A/T) GATA (A/G)
NF-E2	basic leucine zipper	(T/C) TGCTGA (C/G) TCA (T/C)
EKLF	Krüppel zinc finger	CCNCNCCCN
Thyroid hormone receptor	zinc finger	$(AGGT (C/A) A)_2$
Ets	winged helix-turn-helix	(A/C) GGA (A/T)
SCL	basic helix-loop-helix	CAGATG
Rbtn2	LIM domain protein	unknown

specificities (Merika & Orkin, 1993; Ko & Engel, 1993) and distinctive tissue-restricted expression profiles (Yamamoto et al., 1990; Orkin, 1992; Laverriere et al., 1994).

Structural analysis of the members of the GATA family revealed two highly conserved zinc fingers, Cys-X_2-Cys-X_{17}-Cys-X_2-Cys (Evans & Felsenfeld, 1989) which are both involved in DNA recognition but are functionally distinct (Martin & Orkin, 1990; Yang & Evans, 1992). Mutational analysis showed that the C-terminal finger was critical for sequence-specific DNA-binding activity and recognises the consensus GATA motif. In contrast, the highly related N-terminal finger is unable to bind DNA by itself but plays a role in DNA binding by affecting the stability of the DNA-protein complex interaction. GATA proteins identified in *Aspergillus* (Kudla et al., 1990), *Neurospora* (Fu et al., 1990) and *Saccharomyces cerevisiae* (Cunningham et al., 1991) all contain only a single zinc finger that is almost identical to the C-terminal finger of vertebrate GATA proteins demonstrating the critical role of this region in DNA binding.

The GATA-1 mRNA, is highly expressed in erythroid cells, and in two highly related hematopoietic cell lineages, mast and megakaryocyte cells, (Martin et al., 1990; Romeo et al., 1990), as well as in eosinophils (Zon et al., 1993) and at low levels in multipotential hematopoietic progenitor cells. This suggests that initiation of GATA-1 expression occurs at the multipotential progenitor stage and may be selectively switched off in specific cell lineages (Sposi et al., 1992; Cairns et al., 1994). Expression of GATA-1 has also been detected in the Sertoli cells of the testis (Ito et al., 1993), with initiation of transcription from an upstream promoter. In addition, RNA splicing generates a protein identical to GATA-1 present in erythroid cells, inferring a functional role in gene transcription in the Sertoli cells (Ito et al., 1993).

The expression pattern of GATA-2 is extremely similar to GATA-1 in hematopoietic lineages (Martin et al., 1990; Romeo et al., 1990; Yamamoto et al., 1990) but is also found in endothelial cells (Dorfman et al., 1992) and embryonic brain (Yamamoto et al., 1990). Expression of GATA-2 in early erythroid precursor cells is thought to promote cellular proliferation (Briegel et al., 1993) and expression is then subsequently down-regulated during erythroid differentiation (Leonard et al., 1993) possibly by GATA-1 (Weiss et al., 1994). GATA-3 is highly abundant in T lymphocytes (Joulin et al., 1991) and developing neural tissue (Yamamoto et al., 1990). The remaining members of the GATA family, GATA-4 to -6 are expressed in non-hematopoietic lineages.

GATA-4 is expressed in the developing heart, primitive endoderm, intestinal epithelium and gonads (Arceci *et al.*, 1993; Kelley *et al.*, 1993; Laverriere *et al.*, 1994). GATA-5 and GATA-6, isolated from chickens, are also expressed in the heart, gut and gonads (Laverriere *et al.*, 1994).

The distribution of the GATA-binding proteins in various tissues has suggested that these may function in distinct tissue-specific roles and this may be determined by the unique primary structure of each GATA family member. Alternatively, the tissue-specific function of the GATA factors may be solely determined by their timing and tissue-specific expression pattern.

Gene targeting experiments in murine embryonic stem cells and the generation of chimeric animals demonstrated that GATA-1 is essential for normal erythropoiesis (Pevney *et al.*, 1991). GATA-1 deficient embryonic stem cells contributed to all non-hematopoietic tissues examined in the chimeric mice but not to mature erythroid cells. The role of GATA-1 during erythroid development has also been investigated by the *in vitro* differentiation of these GATA-1 deficient embryonic stem cells. The *in vitro* differentiation of embryonic stem cells give rise to embryoid bodies containing various committed cell types including hematopoietic precursors (Doetschman *et al.*, 1985). The formation of embryoid bodies from GATA-1 deficient embryonic stem cells demonstrated that mature erythroid cells were absent and expression of the globin mRNAs of either the primitive (ζ) or definitive lineages (β) were nearly undetectable (Simon *et al.*, 1992). However, the phenotype of the GATA-1 deficient embryonic stem cells could be rescued following the introduction of an intact GATA-1 gene (Simon *et al.*, 1992). Disaggregation of the GATA-1 deficient embryoid bodies and subsequent replating during their maturation, demonstrated that these cells are capable of forming definitive erythroid colonies containing cells blocked at the proerythroblast stage (Weiss *et al.*, 1994). Interestingly, expression of GATA-2 at this stage is markedly elevated in comparison to wild-type, and globin mRNAs are expressed at near normal levels suggesting that GATA-2 can partially compensate for the absence of GATA-1 until the late stage of erythroid development and that GATA-1 may be required to prevent apoptosis (Weiss *et al.*, 1995). In similar experiments, Blobel *et al.* (1995) demonstrated that the chicken GATA-1, which is highly divergent outside of the zinc finger region (Trainor *et al.*, 1990), could rescue GATA-1 deficiency to a similar degree as murine GATA-1. Similar results were obtained using GATA-3 and GATA-4

which are normally expressed in different tissues to GATA-1 (Blobel *et al.*, 1995). These results suggest that the GATA-1 zinc finger does not impart erythroid-specificity *in vivo* although a specific role for other regions unique to GATA-1 cannot be excluded. Pevney *et al.* (1995) used *in vitro* differentiation assays of yolk sac or fetal liver-derived hematopoietic cells isolated from chimeric GATA-1 deficient mice to examine the role of GATA-1 in the differentiation of megakaryocytes and mast cells. In contrast to the essential role of GATA-1 in erythroid differentiation, megakaryocyte and mast cell lineages are able to complete the differentiation process in the absence of GATA-1 (Pevney *et al.*, 1995) suggesting that GATA-2 which is also expressed in these lineages, may be functionally equivalent to GATA-1. Further evidence which supports the view that these GATA proteins are interchangeable was provided by Visvader *et al.* (1992; 1993) who demonstrated that forced expression of GATA-1, GATA-2 or GATA-3 was capable of inducing megakaryocytic differentiation in an early myeloid cell line. In contrast, GATA-2 but not GATA-1 or GATA-3, has been shown to be essential in maintaining chicken erythroid cell precursors in a proliferative state (Briegel *et al.*, 1993). Although *in vitro* studies support the interchangeability of the GATA proteins, GATA-1 is essential for normal erythroid differentiation *in vivo*.

The role of GATA-1 in the regulation of erythroid-specific genes is likely to extend beyond DNA binding and involve protein-protein interactions with other transcription factors to facilitate DNA looping in promoter-enhancer interactions. Crossley *et al.* (1995) demonstrated that GATA-1 is capable of forming homodimers with itself or heterodimers with GATA-2 or GATA-3 and this is mediated through a region located within the C-terminal zinc finger. GATA-1 is also capable of physically interacting with the Krüppel proteins Sp1 and EKLF through this region, as are GATA-2 and GATA-3 with Sp1 (Merika & Orkin, 1995). The physical interaction observed between members of the GATA family with the Krüppel family of transcription factors implies an important and possibly general mechanism which may be utilised in gene regulation involving only a small subset of transcription factors.

1.5.2 NF-E2

The erythroid-enriched factor, NF-E2 which binds the following consensus sequence 5' (T/C)GCTGA(C/G)TCA(T/C) 3' (Andrews *et al.*, 1993a) has been identified in the LCRs of the α- and β- globin gene clusters and is responsible for their enhancer activities (Ney *et al.*, 1990; Talbot *et al.*, 1990; 1991; Strauss *et al.*, 1992; Zhang *et al.*, 1995). Binding sites have also been identified in the erythroid promoter of human PBGD (Mignotte *et al.*, 1989a) and the ferrochelatase promoter (Tugores *et al.*, 1994). NF-E2 was purified from murine erythroleukemia cells and is an obligate heterodimer composed of two novel basic leucine zipper proteins, p45 and p18 (Andrews *et al.*, 1993a; 1993b). The larger subunit, designated p45, is restricted in its expression to cells of the hematopoietic lineage, namely hematopoietic progenitor, erythroid, megakaryocyte and mast cells and is strikingly similar to the tissue distribution of GATA-1 (Andrews *et al.*, 1993a). The DNA binding domain of p45 shares extensive homology to two members of the basic leucine zipper family identified in *Drosophila*, cnc (cap'n collar) and in *Caenorhabditis elegans*, skn-1, involved in head segmentation (Mohler *et al.*, 1991) and blastomere fate determination (Bowerman *et al.*, 1992), respectively. Additional proteins related to NF-E2 in the DNA-binding domain include the yeast factors, Nrf1 (Chan *et al.*, 1993a) and Nrf2 (Moi *et al.*, 1994) and the human LCR-F1 (Caterina *et al.*, 1994b).

In contrast to the larger p45 subunit of NF-E2, p18 is ubiquitously expressed (Andrews *et al.*, 1993b) and is highly related to the chicken oncogene *v-maf* (Nishizawa *et al.*, 1989) and human retinal protein, NRL (Swaroop *et al.*, 1992) and are members of the *maf* subfamily. Additional members of the *maf* subfamily include mafF, mafG and mafK which contain the DNA-binding domain but lack a transactivation domain. These small maf proteins are capable of forming heterodimers with p45 and activate transcription of a reporter gene in transient transfection assays (Igarashi *et al.*, 1994). In contrast, the formation of homodimers of p18 or other small maf proteins capable of binding to the NF-E2 motif, repress transcriptional activation of a reporter gene (Igarashi *et al.*, 1994) and may also be involved in regulating gene transcription *in vivo*.

Evidence supporting the role of NF-E2 in enhancer-dependent globin transcription has been obtained from *in vitro* studies. The MEL cell line, CB3, transformed by the Friend murine leukemia virus, does not express the NF-E2 p45 protein due to the integration of viral sequences within the NF-E2 p45 gene (Lu *et al.*, 1994). Expression of the α- and β-

globin genes are also markedly reduced in these cells but can be partially rescued following the introduction of a retroviral vector containing the NF-E2 p45 cDNA. Kotkow *et al.* (1995) created a dominant negative mutant of p18 unable to bind DNA but capable of forming heterodimers with p45. When stably transfected in MEL cells, this mutant resulted in reduced levels of functional NF-E2 protein and globin gene expression. The introduction of a tethered p45-p18 NF-E2 construct into these cells (or the MEL CB3 cell line lacking endogenous NF-E2) returned expression of the globin genes to wild-type levels (Kotkow *et al.*, 1995).

To investigate the role that NF-E2 plays in the formation of HS2 in the β-globin LCR, Armstrong and Emerson (1996) used an *in vitro Drosophila* chromatin assembly system. Purified recombinant NF-E2 (p45 and p18) was shown to be capable of forming a DNase I hypersensitive site at HS2 *in vitro* as observed *in vivo*. NF-E2 is therefore apparently able to disrupt local nucleosome organisation in a process that is ATP dependent (Armstrong & Emerson, 1996). Other ATP-dependent factors capable of remodelling chromatin *in vitro* have been reported, including the human (Kwon *et al.*, 1994) and yeast (Cote *et al.*, 1994) SWI/SNF, and *Drosophila* NURF (Tsukiyama *et al.*, 1995). However, whether NF-E2 can interact with these chromatin remodelling factors remains to be elucidated. The disruption of nucleosomes by NF-E2 also facilitates the binding of GATA-1 to inverted GATA motifs located approximately 60 bp downstream of the NF-E2 site (Armstrong & Emerson, 1996). Interestingly, the presence of both NF-E2 and GATA motifs are required for the formation of HS4 in the human β-globin LCR (Stamatoyannopoulos *et al.*, 1995). Similarly positioned NF-E2 and GATA motifs have also been identified in the human β-globin HS1 and HS3, murine HS2, HS3 and HS4 and chicken β-globin enhancer (Stamatoyannopoulos *et al.*, 1995) suggesting that a mechanism conserved among species may be utilised to create these hypersensitivity sites.

To investigate the *in vivo* role of p45 and p18, the genes encoding these proteins were each disrupted by homologous recombination in embryonic stem cells. Examination of mice lacking p45 revealed that these mice failed to produce platelets (Shivdasani *et al.*, 1995c) but, surprisingly, erythroid development proceeded in a similar manner to that observed in wild-type mice. In p45 knockout mice, appropriate switching of fetal to adult globin was observed during development, expression of the α- and β- globin genes were similar to wild-type and hemoglobin content was only marginally reduced (Shivdasani *et al.*,

1995b). Similarly, mice lacking p18 exhibited normal erythroid and megakaryocytic development and were indistinguishable from wild-type (Kotkow *et al.*, 1996). Thus, in contrast to *in vitro* studies which established that NF-E2 was essential for globin gene expression, examination of the role of p45 and p18 *in vivo* demonstrated that these factors were not essential for erythroid development and suggested that other related factors possibly AP-1, may compensate in their absence.

1.5.3 EKLF

The erythroid Krüppel-like transcription factor (EKLF) was isolated from a murine erythroleukemia cell line, using the method of subtractive hybridisation (Miller & Bieker, 1993). Expression of EKLF mRNA is restricted to tissues of the erythroid lineage and is barely detectable in mast cells in contrast to that of GATA-1 mRNA (Miller & Bieker, 1993).

EKLF protein consists of a proline-rich amino acid domain at the amino-terminus and contains three TFIIIA-like zinc fingers at the carboxyl terminus, corresponding to the activation and DNA-binding domains, respectively (Miller & Bieker, 1993; Bieker & Southwood, 1995). As its name suggests, these zinc fingers also exhibit extensive homology to members of the Krüppel family of transcription factors (Miller & Bieker, 1993). EKLF is capable of binding to, and activating transcription from, the CACCC site (5' CCACACCCT 3') located at -90 bp from the transcription initiation site in the human and murine adult β-globin promoters (Miller & Bieker, 1993). Functional CACCC sites have also been identified in the promoters of many erythroid-expressed genes, including PBGD (Mignotte *et al.*, 1989b), GATA-1 (Tsai *et al.*, 1991) and the EpoR (Zon *et al.*, 1991), and EKLF may represent a candidate nuclear factor involved in the transcriptional activation of these erythroid-specific genes. However, other members of the Krüppel family of transcription factors, including the ubiquitously expressed Sp1 (Kadonaga *et al.*, 1987), and the basic Krüppel-like factor, BKLF (Crossley *et al.*, 1996) are also capable of binding to the same target sequence *in vitro* (Hartzog & Myers, 1993; Crossley *et al.*, 1996).

In support of a major role in the erythroid-specific transcriptional activation through these CACCC sequences, EKLF is expressed throughout erythroid development even as the site of erythropoiesis changes from the yolk sac to fetal liver and finally the bone marrow

(Southwood *et al.*, 1996). Additional *in vivo* and *in vitro* evidence has been determine from naturally occurring mutations in the CACCC site in the human β-globin promoter. These mutations result in a mild to severe degree of β-thalassemia due to a reduction in β-globin gene expression (Orkin *et al.*, 1984; Kulozik *et al.*, 1991) and the ability of EKLF to transactivate a reporter plasmid containing these mutations is also markedly reduced (Feng *et al.*, 1994).

Bieker and Southwood (1995) demonstrated that the proline-rich region of EKLF when fused to a GAL4 DNA-binding domain efficiently transactivated a reporter plasmid containing a GAL4 DNA binding site in non-erythroid cells and was critical for the activation of the β-globin promoter by the LCR in erythroid cells. Co-transfection assays performed with exogenously expressed EKLF demonstrated that EKLF preferentially activated the human β-globin promoter (1000-fold) in comparison to the human fetal γ -globin promoter (3-fold) which contains a mismatch in the EKLF binding site. These studies suggested a possible role for EKLF in the switching of globin gene expression during development (Donze *et al.*, 1995).

To investigate the role of EKLF *in vivo*, mice were created which were deficient in EKLF by gene targeting experiments (Nuez *et al.*, 1995; Perkins *et al.*, 1995). Erythropoiesis in the yolk sac appeared normal in homozygous EKLF knockout mice. However, as the site of erythropoiesis switched to the fetal liver at day 11.5, these embryos failed to express the adult β-globin genes and died from anemia at approximately day 15. Examination of fetal liver cells from these mice revealed increased levels of iron in Küpffer cells and erythroblasts and a reduced level of β-globin transcripts in comparison to α-globin transcripts (Nuez *et al.*, 1995; Perkins *et al.*, 1995). These are characteristic features observed in β-thalassemia patients. Wijgerde *et al.* (1996) investigated the role of EKLF in the switching of the human β-globin genes in EKLF null mice containing the entire human β -globin locus. In the absence of EKLF, expression of human β-globin gene was undetectable and was associated with a decrease in the sensitivity of the β-globin promoter to DNase I, strikingly the LCR was unaffected by the lack of EKLF (Wijgerde *et al.*, 1996). In contrast, expression of the γ-globin gene increased approximately 3-5 -fold in comparison to heterozygous or wild-type mice and the DNase I hypersensitivity of the γ-globin promoter

was increased in EKLF null mice. These studies suggested that EKLF may be involved in the both the formation of chromatin structure and expression of the β-globin promoter.

An intriguing problem associated with the EKLF protein which is expressed throughout development, is understanding how this protein is capable of functioning at a specific time during erythropoiesis.

Chen and Bieker (1996) examined the proline-rich activation region of EKLF and identified an activation domain and an inhibitory domain located at the amino- and carboxyl-termini, respectively. The inhibitory domain of EKLF may be capable of exerting its function by interfering with efficient DNA binding, however, following activation of the protein, may undergo a conformational change resulting in an ability to bind DNA. A number of putative phosphorylation sites within EKLF have been identified which may play a role in the activation of EKLF (Chen & Bieker, 1996).

EKLF may also function in a cooperative fashion with other transcription factors. Merika and Orkin (1995) demonstrated that the zinc finger domain of EKLF is capable of interacting with the zinc finger domain of GATA-1. The physical association of EKLF and GATA-1 results in the synergistic activation of a promoter containing binding sites for these proteins *in vitro* (Merika & Orkin, 1995; Gregory *et al.*, 1996) suggesting an important role in erythroid-specific gene activation and a mechanism by which sequences located within the promoter and enhancer regions of erythroid-specific genes such as CACCC and GATA sites may interact.

1.5.4   Thyroid Hormone Receptor:

The thyroid hormone receptor (Sap *et al.*, 1986; Weinberger *et al.*, 1986) is a member of the nuclear receptor family of ligand-dependent transcription factors that control development and homeostasis through the activation and repression of gene transcription. The transcriptional activation of target genes by steroid hormones require the binding of a ligand-activated receptor to the steroid-hormone response element located within the regulatory region of the target gene. The binding of the steroid hormone receptor to the DNA binding element is mediated by a highly conserved double zinc finger region identified in all members of the steroid hormone receptor family (Miller *et al.*, 1985). Steroid hormone

receptors contain distinct functional domains responsible for DNA binding, transcriptional activation as well as heterodimerisation and homodimerisation.

There are two major classes of the thyroid hormone receptor, TRα and TRβ encoded by the human *c-erbAα* gene and the *c-erbAβ* gene, respectively, located on different chromosomes (Weinberger *et al.*, 1986; Thompson *et al.*, 1987). Multiple isoforms of TRα are generated by alternative splicing; α1 is a thyroid hormone receptor whereas α2 is unable to bind thyroid hormone (Izumo *et al.*, 1988; Lazar *et al.*, 1988; 1989a; Mitsuhashi *et al.*, 1988). Multiple isoforms of TRβ are also generated from the c-erbAβ gene, β1 and β2 (Weinberger *et al.*, 1986; Koenig *et al.*, 1988; Hodin *et al.*, 1989). Another member of the thyroid hormone receptor family, Rev-ErbAα (Rev) is encoded at the c-erbA α locus and is transcribed in the opposite direction to the c-erbAα mRNA (Lazar *et al.*, 1989b).

The thyroid hormone receptor is capable of forming homodimers or heterodimers with retinoid X receptors (Berrodin *et al.*, 1992; Sugawara *et al.*, 1993) and the retinoic acid receptor (Glass *et al.*, 1989). The formation of heterodimers enhance the binding of the thyroid hormone receptor to the TRE and may play an important role in regulation of gene transcription by thyroid hormone. Mutational analysis of TREs and comparison of TREs from thyroid hormone responsive genes have suggested that thyroid hormone receptors bind to a hexamer half-site sequence of 5' AGGT(C/A)A 3' (Brent *et al.*, 1989) but this can vary slightly in nucleotide sequence. Thyroid hormone receptors can bind to TREs in which the half sites are arranged as palindromes, direct repeats or inverted palindromes. The distance between such half sites can also vary and may influence transcriptional activation (Yen *et al.*, 1994).

The thyroid hormone receptor is capable of binding to a TRE in the absence of ligand to repress gene transcription, and the addition of thyroid hormone relieves repression (Chambon, 1994; Mangelsdorf *et al.*, 1995). The repressive activity of the thyroid hormone receptor is dependent on the binding of the corepressor, N-CoR to the CoR-box located in the ligand binding domain of the thyroid hormone receptor (Horlein *et al.*, 1995). Mutations in the receptor that inhibit the binding of N-CoR do not affect receptor DNA binding or ligand-dependent transactivation but inhibit repressor function (Horlein *et al.*, 1995). These findings suggest that N-CoR may be involved in mediating ligand-independent

transcriptional repression. N-CoR is associated with two other corepressor factors, mSIN3 and the mammalian histone deacetylase, mRPD3 (Alland *et al.*, 1997; Heinzel *et al.* 1997), and a model has been proposed by Wolfe (1997) to explain how repression of gene transcription is relieved by histone deacetylases. In the absence of thyroid hormone, the thyroid hormone receptor binds to nucleosomal DNA and targets histone deacetylation resulting in the repression of gene transcription. However, histone acetylases are recruited in the presence of thyroid hormone resulting in the alleviation of repressed gene transcription (Wolfe, 1997).

The oncoprotein, v-erbA is a highly mutated version of its cellular homologue, TRα 1 and is one of two oncogenes of the avian erythroblastosis virus (AEV) that causes erythroleukemia and fibrosarcomas in chickens (Debuire *et al.*, 1984). Comparison of the amino acid sequence of the chicken TRα1 and v-erbA proteins have revealed multiple differences; two changes are located within the DNA binding domain of the protein, nine in the hormone binding domain and small deletions in the amino- and carboxyl- termini (Sap *et al.*, 1986; Munoz *et al.*, 1988). As a result, the v-erbA oncoprotein expressed in mammalian cells is unable to bind thyroid hormone but is still capable of binding DNA in a sequence-specific manner (Damm *et al.*, 1989; Sap *et al.*, 1989).

Expression of the v-erbA oncoprotein in erythroblasts blocks erythroid differentiation and suppresses erythrocyte specific gene transcription in chicks, including transcription of the band 3, carbonic anhydrase II (CAII) and ALAS2 genes which are normally upregulated during erythroid differentiation (Zenke *et al.*, 1988). Overexpression of the thyroid hormone receptor by a retrovirus in erythroid cells induces these cells to differentiate into erythrocytes in the presence of thyroid hormone but erythroid differentiation and expression of erythrocyte specific genes are inhibited in the absence of thyroid hormone (Zenke *et al.*, 1990). Consequently, v-erbA has lost the ability to regulate gene transcription in the presence of thyroid hormone and instead constitutively binds DNA and represses transcription. Disela *et al.* (1991) identified a motif at -660 bp in the carbonic anhydrase II promoter capable of binding either c-erbA or v-erbA, and which was responsible for the down-regulation of carbonic anhydrase II gene transcription by the v-erbA oncoprotein. Interestingly, this erbA binding site exhibited sequence homology to the binding site for NF-E2 suggesting a possible role for these two factors in the erythroid-specific regulation of the carbonic anhydrase II gene (Disela *et al.*, 1991).

1.5.5 Ets Gene Family of Transcription Factors

The founding member of the Ets family of transcription factors is the oncoprotein v-ets, discovered as part of a fusion protein with v-myb in the E26 avian leukemia retrovirus which can transform cells of both erythroid and myeloid cell lineages *in vitro* (LePrince *et al.*, 1983; Nunn *et al.*, 1983; Beug *et al.*, 1984). The cellular homologue of the v-ets protein is ets-1 (Leprince *et al.*, 1983; Watson *et al.*, 1988) and numerous ets-related proteins have since been identified in species ranging from humans to *Drosophila*. Each of these proteins share sequence similarity to what is known as the 'ets domain', an 85 amino acid sequence that is responsible for DNA binding (Nye *et al.*, 1992). Ets proteins recognise the purine rich core motif 5' (A/C)GGA(A/T) 3' identified in the control regions of specific target genes and can function as transcriptional activators in cooperation with other transcription factors (Wasylyk *et al.*, 1993).

The expression of Ets proteins has been demonstrated in cells of the hematopoietic lineage and several lines of evidence suggest that Ets proteins may be involved in erythropoiesis. Examination of various Ets family mRNAs in erythroid cells by Northern blot analysis have shown that significant levels of Fli-1 and Spi-1/PU.1 mRNAs are expressed in both fetal liver and MEL cells, with negligible levels of ets-1 or ets-2 mRNAs. In contrast, K562 cells express very low levels of ets-2 and Spi-1/PU.1 mRNAs whereas the human erythroleukemic (HEL) cell line expresses detectable levels of all four ets mRNAs (O'Prey *et al.*, 1993).

The E26 virus causes erythroid and myeloid leukemias. In contrast, E26 viruses containing mutations in the v-ets region are only capable of transforming myeloid cells but not erythroid cells *in vitro* (Nunn *et al.*, 1989). A temperature-sensitive mutant of E26, *ts*1.1, that contains a point mutation in the DNA binding domain of the v-ets protein, shows reduced DNA binding at the non-permissive temperature of 42°C (Golay *et al.*, 1988; Kraut *et al.*, 1994). Multipotent hematopoietic progenitor cells transformed with *ts*1.1 can be induced to differentiate along the erythroid and myeloid lineages following a shift to the non-permissive temperature suggesting that the ets portion of the vMyb-vEts fusion protein of E26 may inhibit the lineage commitment of multipotent hematopoietic progenitor cells (Kraut *et al.*, 1994).

Studies performed by Metz and Graf (1991a; 1991b) have also demonstrated the requirement for a functional interaction between v-myb and v-ets in the transformation of erythroblast-like cells. Retroviral constructs expressing either v-myb or v-ets only weakly transform erythroblast-like cells *in vitro*. In contrast, viruses encoding a single molecule containing the DNA binding domains of v-myb and v-ets and the transactivation domain of v-Myb are sufficient to induce leukemia. The v-ets oncoprotein is also capable of cooperating with the oncoprotein, v-erbA, resulting in erythroleukemia in chick embryos (Metz & Graf, 1992). Consequently, v-ets is capable of transforming erythroid cells by itself but when in cooperation with v-myb or v-erbA can enhance the transformation potential of these oncoproteins.

This cooperation of v-ets with v-myb and v-erbA suggests that they may regulate specific target genes (Metz & Graf, 1991a; 1991b; 1992). Aurigemma *et al.*(1992) demonstrated that the Myb-Ets fusion protein of the ME26 virus is capable of directly transactivating the GATA-1 promoter, suggesting that these proteins may be involved in the activation of GATA-1 *in vivo*. Overexpression of MafB, an AP-1-like protein, in erythroblasts inhibits transactivation of the transferrin receptor gene by ets-1 and this is mediated by the interaction of the DNA binding domains of these two proteins (Sieweke *et al.*, 1996). Interestingly, over-expression of the MafB protein, which is usually restricted to myeloid cells, can inhibit erythroid differentiation and repress expression of the genes encoding PBGD, transferrin receptor, and the α- and β- globins. These findings suggest that the interaction of MafB with ets-1 may represent a possible mechanism for repressing erythroid-specific genes in myeloid cells (Sieweke *et al.*, 1996).

Studies of murine erythroleukemia cells transformed by the Friend virus, which is composed of a complex of a replication defective spleen focus-forming virus (SFFV) and the replication competent Friend murine leukemia virus (F-MuLV), have indicated a possible role for two members of the ets family, Fli-1 and Spi-1/PU.1, in erythropoiesis. The *Fli-1* locus is rearranged in approximately 75% of erythroleukemias induced by the Friend murine leukemia virus (Ben-David *et al.*, 1991). Similarly, Moreau-Gachelin *et al.* (1988) demonstrated that the SFFV has integrated upstream of the *Spi-1/PU.1* locus in at least 90% of the MEL cell lines examined, resulting in elevated levels of Spi-1 mRNA in comparison to the very low levels detected in normal differentiating erythroblasts (Galson *et al.*, 1993). Conversely, Northern blot analysis of Me$_2$SO-inducible MEL cells has shown

that the level of Spi-1 mRNA is rapidly reduced to approximately 20% of the uninduced level before cells become committed to differentiate (Galson *et al.*, 1993). Retroviral constructs expressing the ets factor, PU.1, can also immortalise erythroblasts (Schuetze *et al.*, 1993). These results suggest that PU.1 may interfere with the commitment of erythroblasts to differentiate and chemicals that reduce expression of PU.1 can reinstate the erythropoietic program. Therefore Ets may be an inhibitor of erythroid-specific gene expression.

1.5.6   SCL and Rbtn2

The SCL and Rbtn2 oncogenes were initially identified on the basis of chromosomal translocations in human T cell leukemias (Boehm *et al.*, 1988; Begley *et al.*, 1989; Chen *et al.*, 1990; Sanchez-Garcia & Rabbitts, 1993). Gene targeting studies have demonstrated that expression of both SCL and Rbtn2 are essential for erythroid development (Warren *et al.*, 1994; Robb *et al.*, 1995; Shivdasani *et al.*, 1995a).

SCL is a member of the basic-helix-loop-helix (bHLH) family of transcription factors (Begley *et al.*, 1989; Chen *et al.*, 1990). Expression of this factor is largely restricted to hematopoietic tissues where its expression precedes or parallels the sites of hematopoiesis (Kallianpur *et al.*, 1994). SCL mRNA is detected in multipotential myeloid and erythroid precursor cells, mast and megakaryocytic cells but is absent in T and B cells (Green *et al.*, 1991; Visvader *et al.*, 1991; Mouthon *et al.*, 1993; Kallianpur *et al.*, 1994).

Several lines of evidence have demonstrated a role for SCL in erythroid development. Expression of SCL is transcriptionally upregulated in differentiating Me_2SO-induced MEL (Visvader *et al.*, 1991; Aplan *et al.*, 1992) and Epo-induced J2E cell lines (Prasad *et al.*, 1995). In addition, the introduction of antisense SCL constructs into MEL cells can inhibit erythroid differentiation (Aplan *et al.*, 1992). In turn, the murine proximal SCL promoter contains a functional GATA-1 binding site and responds to exogenously expressed GATA-1 suggesting that GATA-1 may be involved in the transcriptional upregulation of SCL during erythroid differentiation (Aplan *et al.*, 1992). Gene targeting experiments have demonstrated that SCL is essential for embryonic erythropoiesis (Robb *et al.*, 1995; Shivdasani *et al.*, 1995a). *In vitro* differentiation experiments performed with SCL null embryonic stem cells have also demonstrated that

31

these cells do not give rise to precursor or mature hematopoietic cells of any lineage but could be phenotypically rescued following the introduction of a cDNA encoding SCL (Porcher *et al.*, 1996).

Rbtn2 contains a cysteine-rich zinc finger-like LIM domain (Sanchez-Garcia & Rabbitts, 1993). Expression of Rbtn2 is restricted to the erythroid lineage and gene targeting studies have demonstrated that Rbtn2 is essential for erythroid development (Warren *et al.*, 1994). Interestingly, the absence of Rbtn2 mimics the effects observed with the loss of either SCL or GATA-1 (Pevney *et al.*, 1991; Robb *et al.*, 1995; Shivdasani *et al.*, 1995a). Complexes of Rbtn2 and SCL have been detected in the nuclei of erythroid cells (Valge-Archer *et al.*, 1994) and this is thought to be mediated by an association of the bHLH portion of SCL and the LIM domain of Rbtn2 (Wadman *et al.*, 1994). Rbtn2 can also physically interact with GATA-1 and GATA-2 (Osada *et al.*, 1995). Although Rbtn2 has not been demonstrated to bind DNA with sequence-specificity, the possible interaction of Rbtn2 with SCL or GATA-1 may result in the generation of a transcriptional complex responsible for the activation of erythroid-specific target genes. The target genes that are regulated by SCL and Rbtn2 are currently under investigation and the relative levels of these erythroid regulatory complexes within these cells may influence erythroid differentiation.

1.6   THE HEME BIOSYNTHETIC PATHWAY

In eukaryotic cells, the heme biosynthetic pathway is comprised of eight nuclear encoded enzymes (Fig. 1.3). The first enzyme and the last three enzymes of the pathway, are localised in the mitochondria while the intermediate steps of the heme biosynthetic pathway are performed by enzymes that are located in the cytosol. The reason for the compartmentalisation of these enzymes is unknown but may reflect substrate availability, for example, one of the substrates for 5-aminolevulinate synthase (ALAS), the first enzyme of the heme biosynthetic pathway is succinyl-coenzyme A (succinyl-CoA) which is only located in the mitochondria.

ALAS is synthesised as a precursor protein which is processed into the mature enzyme as it traverses the mitochondrial membrane, to its site of action on the matrix side of

Figure 1.3   The Heme Biosynthetic Pathway in Non-Plant Eukaryotes

The enzymes catalysing each of the eight steps of the heme biosynthetic pathway and their subcellular compartmentalisation, and the intermediates of heme biosynthesis are indicated. The following abbreviations have been used: **ALAS**, 5-aminolevulinate synthase; **ALA**, 5-aminolevulinic acid; **ALAD**, 5-aminolevulinate dehydratase; **PBG**, porphobilinogen; **PBGD**, porphobilinogen deaminase; **UROS**, uroporphyrinogen III synthase; **UROGEN**, uroporphyrinogen III; **UROD**, uroporphyrinogen III decarboxylase; **COPROGEN**, coproporphyrinogen III; **CPX**, coproporphyrinogen III oxidase; **PROTO'GEN III**, protoporphyrinogen III; **PPOX**, protoporphyrinogen oxidase; **A**, acetate; **M**, methyl; **P**, propionate and **V**, vinyl.

GLYCINE
$NH_2 - CH_2 - COO^-$

+

$^-OOC - CH_2 - CH_2 - C = O$
SCoA

SUCCINYL CoA

ALAS

ALA
$^-OOC - CH_2 - CH_2 - C - CH_2 - NH_2$
O

ALAD

PBG

PBGD

HYDROXYMETHYLBILANE

UROS

UROGEN

UROD

COPROGEN

CPX

PROTO'GEN III

PPOX

PROTOPORPHYRIN IX

Ferrochelatase Fe^{2+}

HEME

CYTOPLASM
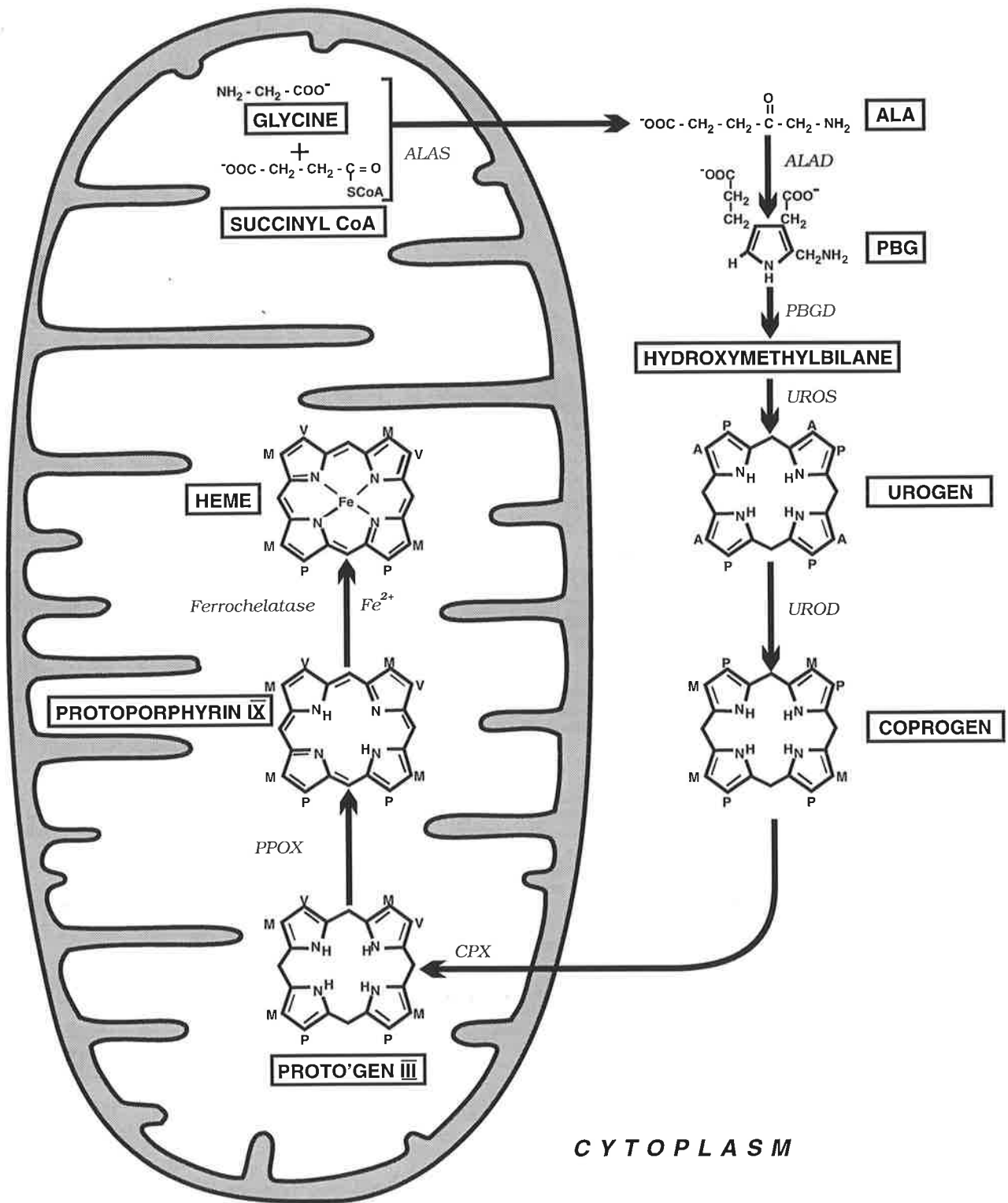
the inner mitochondrial membrane. ALAS is a pyridoxal phosphate-dependent enzyme that catalyses the condensation of glycine and succinyl-CoA to form 5-aminolevulinic acid (ALA) which is then transported to the cytosol. The first cytosolic enzyme of the heme biosynthetic pathway, 5-aminolevulinate dehydratase (ALAD), then catalyses the condensation of two ALA molecules to form the pyrrole porphobilinogen (PBG). The enzyme porphobilinogen deaminase (PBGD) polymerises four PBG molecules to form a highly unstable linear tetrapyrrole, hydroxymethylbilane which is then cyclised by uroporphyrinogen III synthase (UROS) to generate the ring tetrapyrrole uroporphyrinogen III (UROGEN). In the absence of this enzyme, the unstable intermediate cyclises non-enzymatically to give the isomer uroporphyrinogen I which has no known biological function. The decarboxylation of UROGEN to coproporphyrinogen III (COPROGEN) by uroporphrinogen decarboxylase (UROD) is the final cytosolic step of the heme biosynthetic pathway. COPROGEN is transported to the mitochondria where coproporphyrinogen III oxidase (CPX), located in the intermembrane space of the mitochondria, converts COPROGEN to protoporphyrinogen III. Protoporphyrinogen III is then oxidised by the enzyme, protoporphyrinogen III oxidase (PPOX). Unusually, PPOX does not contain an apparent leader sequence for mitochondrial targeting (Nishimura *et al.*, 1995) and import, yet is located in the inner mitochondrial membrane where it forms protoporphyrin IX. The final step in the heme biosynthetic pathway is the insertion of ferrous iron into protoporphyrin IX by ferrochelatase, also located in the inner mitochondrial membrane, with the resulting compound being heme. As previously mentioned, heme is either utilised for mitochondrial respiratory cytochromes and P450s, or is transported out of the mitochondrion to other cellular locations, for example lysosomes, or into a cytosolic pool for assembly into hemoglobin. Although, the mechanism responsible for the transport of heme and its precursors from the mitochondrion to the cytosol is unknown (Müller-Eberhard & Fraig, 1993; Smith, 1990), a candidate mitochondrial membrane protein has been recently identified and may be involved in the transport process (Taketani *et al.*, 1994).

1.7 ALAS

1.7.1 Structure and Properties of ALAS

ALAS activity was initially described in bacterial (Kikuchi *et al.*, 1958) and avian erythrocytic (Gibson *et al.*, 1958) extracts. Although the ALAS enzyme has since been purified from a number of prokaryotic and eukaryotic sources, molecular studies of the enzyme have been directed towards the eukaryotic ALAS enzymes (Warnick *et al.*, 1971; Srivastava *et al.*, 1982; Volland & Felix, 1984; Borthwick *et al.*, 1986).

Eukaryotic ALAS is a nuclear-encoded enzyme, localised to the matrix side of the inner mitochondrial membrane by immuno-gold labelling studies (Rohde *et al.*, 1990). ALAS is synthesised in the cytosol as a precursor protein which contains a signal sequence responsible for targeting the precursor protein to the mitochondria (Kikuchi & Hayashi, 1981). This signal sequence is then cleaved to generate the mature ALAS protein (Pirola *et al.*, 1986). In contrast, the bacterial ALAS protein lacks a signal sequence (May *et al.*, 1995). Electron microscopy and cross-linking studies established that the purified chicken liver mitochondrial ALAS enzyme is probably composed of a homodimer of two identical subunits (Pirola *et al.*, 1984).

The condensation of glycine and succinyl-CoA by ALAS is critically dependent on the cofactor, pyridoxal phosphate (PLP) which is a derivative of pyridoxine (vitamin B6). Studies performed with purified bacterial ALAS (Nandi, 1978) and chicken liver ALAS (Pirola, 1986) demonstrated that PLP forms a Schiff base with a lysine residue located in the catalytically active domain of the ALAS protein (Ferreira *et al.*, 1993). A glycine-rich motif has also been identified in all ALAS proteins examined and resembles the GXGXXG sequence corresponding to the active site of other PLP-dependent enzymes and may interact with the phosphate moiety of the cofactor (Marceau *et al.*, 1990).

1.7.2 Isozymes of ALAS

There are two closely related isozymes of ALAS in vertebrates, designated ALAS1 and ALAS2 which are encoded by separate genes located on different chromosomes (Sutherland *et al.*, 1988; Cox *et al.*, 1990; Bishop *et al.*, 1990). The housekeeping enzyme, ALAS1, is probably expressed in all tissues to provide heme for respiratory cytochromes,

whereas the erythroid-specific enzyme, ALAS2, is involved in synthesising the large amounts of heme required during erythropoiesis (Elliott *et al.*, 1989; Cox *et al.*, 1991; May *et al.*, 1995).

The synthesis of ALAS1 can be induced by porphyrinogenic drugs including phenobarbital and repressed by heme in the liver (Granick, 1966; Srivastava *et al.*, 1988; 1990). In contrast, the synthesis of ALAS2 is unresponsive to porphyrinogenic drugs, is developmentally regulated and markedly increased during erythropoiesis to meet the demand for heme during hemoglobin production (May *et al.*, 1995). The mechanisms responsible for the regulation of the ALAS1 and ALAS2 isozymes are described in detail in Section 1.8 and 1.9, respectively.

The ALAS isozymes can be divided into three regions based on their functional role. To highlight these regions, the deduced amino acid sequence of the human ALAS1 (Bishop *et al.*, 1990) and ALAS2 (Cox *et al.*, 1990) precursor proteins were aligned with that of the chicken (Borthwick *et al.*, 1985) and rat (Srivastava *et al.*, 1988) ALAS1, chicken (Riddle *et al.*, 1989) and mouse (Schoenhaut & Curtis, 1989) ALAS2 precursor proteins and ALAS proteins from yeast (Urban-Grimal *et al.*, 1986) and prokaryotic (McClung *et al.*, 1987; Leong *et al.*, 1985; Elliot *et al.*, 1989) sources (Fig. 1.4).

Region 1 corresponds to the signal sequence responsible for the transport of the precursor ALAS enzyme from the cytoplasm to the mitochondria. The chicken and rat ALAS1 precursor proteins have N-terminal sequences of 56 amino acids, with the proteolytic cleavage site located between two glutamine residues at positions 61 and 62 (Maguire *et al.*, 1986; Srivastava *et al.*, 1988). Therefore the corresponding proteolytic cleavage site in the human ALAS1 precursor can also be predicted to be located between these two glutamine residues resulting in a signal sequence of 56 amino acids (Cox *et al.*, 1991). From the sequence alignment, the cleavage site for the ALAS2 precursor protein is predicted to reside between a serine and glutamine residue (as indicated in Fig. 1.4), with a signal sequence of 49 amino acids (Cox *et al.*, 1991). Therefore the size of the mature mitochondrial forms of the human ALAS1 and ALAS2 proteins are 64.4 kDa and 59.5 kDa, respectively (Cox *et al.*, 1991). Although a different cleavage site has been postulated for the mouse ALAS2 precursor protein (Schoenhaut & Curtis, 1986), the signal sequence of both the human and mouse ALAS2 enzymes exhibit significant amino acid similarity. In contrast, the signal sequence of the chicken ALAS2 enzyme (Riddle *et al.*, 1989) more

Figure 1.4   Comparison of Amino Acid Sequence of Eukaryotic and Prokaryotic ALAS proteins

Comparison of the amino acid sequence of eukaryotic and prokaryotic ALAS proteins from cDNA clones or direct sequencing. Region 2 found in the mature eukaryotic ALAS protein but not in the prokaryotic ALAS protein is shaded. The amino acid sequences used for this comparison were obtained from the following sources: **1)** rat ALAS1, Srivastava *et al.* (1988); **2)** human ALAS2, Cox *et al.* (1991); **3)** human ALAS1, Bawden *et al.* (1987); **4)** chicken ALAS1, Borthwick *et al.* (1985); **5)** chicken ALAS2, Riddle *et al.* (1989); **6)** mouse ALAS2, Schoenhaut & Curtis (1986); **7)** *Saccharomyces cerevisiae*, Urban-Grimal *et al.* (1986); **8)** *Bradyrhizobium japonicum*, McClung *et al.* (1987); **9)** *Rhizobium meliloti*, Leong *et al.* (1985); **10)** *Rhodopseudomonas spheroides*, Elliott *et al.* (1989). The predicted proteolytic cleavage site in the precursor ALAS2 isozyme is located between position 61 and 62 and is *arrowed*. The lysine residue involved in the Schiff base linkage with the PLP cofactor is shaded and indicated with a star.

```
            1         10        20        30        40        50        60        70        80        90
1.     METVVRRCPFLSRVPQAFLQKAGK**SLLFYAQNCPKMMEVGAKPAPRTVSTSAAQCQQVKETPPANEKEKTAKAAVQQ*APDESQM
2.   MVTAAMLLQCCPVLARGPTSLLGKVVKTHQFLFGIGRCPI***********LATQGPNCSQI*******HLKATKA***********
3.     MESVVRRCPFLSRVPQAFLQKAGK**SLLFYAQNCPKMMEVGAKPAPRALSTAAVHYQQIKETPPASEKDKTAKAKVQQ*TPDGSQQ
4.     MEAVVRRCPFLARVSQAFLQKAGP**SLLFYAQHCPKMMEAAPPAAARGLATSAARGQQVEETPAAQPEAKKAKEVAQQ********
5.     MAAFLR*CPLLARHP**PLARAFAT******GARCPFMGFA****************************************
6.   MVAAAMLLRSCPVLSNGPTGLLGKVAKTYQFLFSIGRCPI***********LATQGPTCSQI*******HLKATKA***********
7.                     MQRSIFARFGNSSAAVSTLNRL****STTAA*PHAKNGYA***********

            91        100       110       120       130       140       150       160       170       180
1.   AQTPDGTQLPPGHPSPSTSQS*SGSKCPFLAAQLSQTGSSVFRKASLELQEDVQEMHAVRTEVAQSPVLPS**LVNAKRDGEG*PSPLLK
2.   ***********GGDSPSWAKG****HCPFMLSELQDGKSKIVQKAAPEVQEDVKAFKTD*******LPS*SLVSVS********LRK
3.   **SPDGTQLPSGHPLPATSQG*TASKCPFLAAQMNQRGSSVFCKASLELQEDVQEMNAVRKEVAETSAGPS**VVSVKTDGGD*PSGLLK
4.   **NTDGSQPPAGHPPAAAVQS*SATKCPFLAAQMNHKSSNVFCKASLELQEDVKEMQVDRKGKEFAK*IPTNSVVRNTEAEGEEQSGLLK
5.   *******************************HRAAPELQEDVERPQIPAVEVLEEL********LRDGGA**ALNR
6.   ***********GGASPSWAKS****HCPFMLSELQDRKSKIVQRAAPQVQEDVKTFKTD********LLS*TMDSTT********RSH
7.                    TATG*ATAAAATATASSTHAAAAAAAAA***********************

            181       190       200       210       220       230       240   .  250       260       270
1.   NF*QDTMRKQRPERVSHLLQDNLPKVVSTFQYDHFFEKKIDEKKNDHT***YRVFKTVNRRAQIFPMADDYTDSLITKKQVSVWCSNDYL
2.   PFSGPQEQEQISGKVTHLIQNNMPGN*YVFSYDLFFRDKIMEKKQDHT***YRVFKTVNRWADAYPFAQHFSEASVASKDVSVWCSNDYL
3.   NF*QDIMQKQRPERVSHLLQDNLPKSVSTFQYDRFFEKKIDEKKNDHT***YRVFKTVNRRAHIFPMADDYSDSLITKKQVSVWCSNDYL
4.   KF*KDTMLKQRPESVSHLLQDNLPKSVSTFQYDQFFEKKIDEKKKDHT***YRVFKTVNRKAQIFPMADDYSDSLITKKEVSVWCSNDYL
5.   TV*RDCMDE***************DAFPYEEQFQAQLGALRRTHT***YRVFTAVGRRADAPPLGTRGTAPH*TS**VELWCSSDYL
6.   SFPFSQEPEQTEGAVPHLIQNNMTGS*QAFGYDQFFRDKIMGKKQDHT***YRYFNNINRLAKEFPLA****HRQREADKVTVWCSNDYL
7.   ****************************NHSTQESGFDYEGLIDSELQKKRLDKS***YRYFNNINRLAKEFPLA****HRQREADKVTVWCSNDYL
8.                     MDYSQFFNSALDRI***HTERRYRVFADLERMAGRFPHA**IWHSPKGKRDVVIWCSNDYL
9.                     MDFESFFKNELDGL***HEEGRYRVFADLARHRGSFPKATRYTADG***EEVTVWCSNDYL
10.                    MDYNLALDTALNRL***HTEGRYRTFIDIERRKGAFPKAMWRKPDGSE*KEITVWCSND

            271       280       290       300       310       320       330       340       350       360
1.   GMSRHPRVCGAVIETVKQHGAGAGGTRNISGTSKFHVELEQELADLHGKDAALLFSSCFVANDSTLFTLAKMMPGCEIYSDSGNHASMIQ
2.   GMSRHPQVLQATQETLQRHGAGAGGTRNISGTSKFHVELEQELAELHQKDSALLFSSCFVANDSTLFTLAKILPGCEIYSDAGNHASMIQ
3.   GMSRHPRVCGAVMDTLKQHGAGAGGTRNISGTSKFHVDLERELADLHGKDAALLFSSCFVANDSTLFTLAKMLPGCEIYSDSGNHASMIQ
4.   GMSRHPRVCGAVMDTLKQHGAGAGGTRNISGTSKFHVDLEKELADLHGKDAALLFSSCFVANDSTLFTLAKMLPGCEIYSDSGNHASMIQ
5.   GLSRHPAVLRAARAALDAHGLGAGGTRNIGGTSPLHGALERALALLHRQPRAALFSSCFAANDTALDTLARILPGCQVYSDAGNHASMIQ
6.   GISRHPRVLQAIEETLKNHGAGAGGTRNISGTSKFHVGLEQELAELHQKDSALLFSSCFVANDSTLFTLAKLLPGCEIYSDAGNHASMIQ
7.   ALSKHPEVLDAMHKTIDKYGCGAGGTRNIAGHNIPTLNLEAELATLHKKEGALVFSSCYVANDAVLSLLGQKMKDLVIFSDELNHASMIV
8.   GMGQHPKVVGAMVETATRVGTGAGGTRNIAGTHHPLVQLEAELADLHGKEAALLFTSGYVSNQTGIATIAKLIPNCLILSDELNHNSMIE
9.   GMGQCPILTEAMKNAIDE
10.               RLEAELADLHGK                               LNHASMIE

            361       370       380       390       400       410       420       430       440       450
1.   GIRNSRVPKYIFRHNDVNHLRELLQRSDPSVPKIVAFETVHSMDGAVCPLEELCDVAHEFGAITFVDEVHAVGLYGASGGGI*********
2.   GIRNSGAAKFVFRHNDPDHLKKLLEKSNPKIPKIVAFETVHSMDGAICPLEELCDVSHQYGALTFVDEVHAVGLYGSRGAGI********
3.   GIRNSRVPKYIFRHNDVSHLRELLQRSDPSVPKIVAFETVHSMDGAVCPLEELCDVAHEFGAITFVDEVHAVGLYGARGGGI********
4.   GIRNSRVPKHIFRHNDVHLKKSDPSTPKIVAFETVHSMDGAICPLEELCDVAHEGAITFVDEVHAVGLYGARGAGI********
5.   GIRRRGVPKFIFRHNDPHHLEQLLGRSPPGVPKIVAFESLHSMDGSIAPLEELCDVAHAYGALTFVDEVHAVGLYGARGAGI********
6.   GIRNSGAAKFVFRHNDPGHLKKLLEKSDPKTPKIVAFETVHSMDGAICPLEELCDVAHQYGALTFVDEVHAVGLYGARGAGI********
7.   GIKHANVKKHIFKHNDLNELEQLLQSYPKSVPKLIAFESVYSMAGSVADIEKICDLADKYGALTFLDEVHAVGLYGPHGAGVAEHCDFES
8.   GIRQSGCERQVFRHNDLADLEALLKAAGANRPKLIACESLYSMDGDVAPLAKICDLAEKYNAMTYVDEVHAVGMYGPRGGGI********

10.  GIRRSGTEK

            451       460       470       480 *     490       500       510       520       530       540
1.   ***************GDRDGVMPKMDIISGTLGKAFGCVGGYIASTSLLIDTVRSYAAGFIFTTSLPPMLLAGALESVRILKSNEGRALRR
2.   ***************GERDGIMHKIDIISGTLGKAFGCVGGYIASTRDLVDMVRSYAAGFIFTTSLPPMVLSGALESVRLLKGEEGQALRR
3.   ***************GDRDGVMPKMDIISGTLGKAFGCVGGYIASTSSLIDTVRSYAAGFIFTTSLPPMLLAGALESVRILKSAEGRVLRR
4.   ***************GDRDGVMPKMDIISGTLGKAFGCVGGYIASTSSLIDTVRSHAAGFIFTTSLPPMLLAGALESVRTLKSAEGQVLRR
5.   ***************AERDGVQHKVDVVSGTLGKALGAVGGYIAGSEALVDAVRSLGPGFIFTTALPPQRGGGALAALQVVGSAEGAALRR
6.   ***************GERDGIMHKLDIISGTLGKAFGCVGGYIASTRDLVDMVRSYAAGFIFTTSLPPMLSGALESVRLLKGEEGQALRR
7.   HRASGIATPKTNDKGGAKTVMDRVDMITGTLGKSFGSVGGYVAASRKLIDWFRSFAPGFIFTTTLPPSVMAGATAAIRYQR**CHIDLRT
8.   ***************AERDGVMHRIDILEGTLAKAFGCLGGYIAANGRIIDAVRSYAPGFIFTTALPPAICSAATAAIKHLKTSSWERER*

10.                AYGVFGGYIAASSK

            541       550       560       570       580       590       600       610       620       630
1.   QHQRNVKLMRQMLMDAGLPVIHCPSHIIPVRVADAAKNTEICDELMTRHNIYVQAINYPTVPRGEELL*RIAPTPHHTPQMMNYFLEKLL
2.   AHQRNVKHMRQLLMDRGLPVIPCPSHIIPIRVGNAALNSKLCDLLLSKHGIYVQAINYPTVPRGEELL*RLAPSPHHSPQMMEDFVEKLL
3.   QHQRNVKLMRQMLMDAGLPVVHCPSHIIPVRVADAAKNTEVCDELMSRHNIYVQAINYPTVPRGEELL*RIAPTPHHTPQMMNYFLENLL
4.   QHQRNVKLMRQMLMDAGLPVVHCPSHIIPIRVADAAKNTEICDKLMSQHSIYVQAINYPTVPRGEELL*RIAPTPHHTPQMMSYFLEKLL
5.   AHQRHAKHLRVLLRDRGLPAL**PSHIVPVRW*DAEANTRLSRALLEEHGLYVQAINHPTVPRGQELLLRIAPTPHHSPPMLENLADKLS
6.   AHQRNVKHMRQLLMDRGFPVIPCPSHIIPIRVGNAALNSKICDLLLSKHGIYVQAINYPTVPRGEELL*RLAPSPHHSPQMMENFVEKLL
7.   SQQKHTMYVKKAFHELGIPVIPNPSHIVPVLIGNADLAKQASDILINKHQIYVQAINFPTVARGTERL*RITPTPGHT****NDLSDILI
8.   *HQDRAARVKAILNAAGLPVMSSDTHIVPLFIGDAEKCKQASDLLLEEHGIYIQPINYPTVAKGSERL*RITPSPYHDDGLIDQLAEALL

            631       640       650       660       670       680       690
1.   LTWKRVGLELKPHSSAECNFCRRPLHFEVMSEREKAYFSGMSKM*VSAQA
2.   LAWTAVGLPLQDVSVAACNFCRRPVHFELMSEWERSYFGNMGPQYVTTYA
3.   VTWKQVGLELKPHSSAECNFCRRPLHFEVMSEREKSYFSGLSKL*VSAQA
4.   ATWKDVGLELKPHSSAECNFCRRPLHFEVMSERERSYFSGMSKL*LSVSA
5.   ECWGAVGLPREDPPGPSCSSCHRPLHLSLLSPLERDQFG******VRGAAAG
6.   LAWTGVGLPLQDVSVAACNFCHRPVHFELMSEWERSYFGNMGPQYVTTYA
7.   NAVDDVFNELQLPRVRDWESQGGLLGVGESGFVEESNLWTSSQLSLTNDDLNPNVRDPIVKQLEVSSGIKQ
8.   QVWDRLGLPLKQKSLAA
```

closely resembles the chicken ALAS1 enzyme (Borthwick *et al.*, 1985) rather than the mouse or human ALAS2 enzymes (Cox *et al.*, 1991) (Fig. 1.4).

The amino acid sequence comparison reveals a significant degree of similarity between the C-terminal region of the eukaryotic ALAS protein and the complete bacterial ALAS proteins which lack a signal sequence (Fig. 1.4). The N-terminus of this highly conserved domain corresponds to the N-termini of the bacterial ALAS enzymes (Elliott *et al.*, 1989) and therefore must contain the catalytically active domain of the eukaryotic ALAS enzyme (designated as Region 3). This C-terminal region of the vertebrate protein is comprised of approximately 440 amino acids and represents 75% of the mature protein. Amino acid sequence comparison of region 3 from the human ALAS1 and ALAS2 enzymes and the mouse ALAS2 enzyme revealed 73% identity between the human ALAS1 and ALAS2 proteins and 94% identity between the human and mouse ALAS2 proteins (Cox *et al.*, 1991) (Fig. 1.4).

The remaining region of the mature ALAS protein, located between the signal sequence and the catalytic domain is designated region 2. Region 2 of the mature human ALAS2 enzyme is 47 amino acids shorter than that for human ALAS1 and exhibits limited sequence similarity (Fig. 1.4). Although the human and mouse ALAS2 proteins closely resemble each other in this region, the chicken ALAS2 protein exhibits only limited sequence similarity to either of these erythroid isozymes. The function of region 2 remains to be determined although it has been postulated to serve a regulatory role because of its conservation (May *et al.*, 1995).

1.7.3   Structural Organisation of the ALAS1 and ALAS2 Genes

The exon-intron organisation of the chicken (Maguire *et al.*, 1986) and rat (Yomogida *et al.*, 1993) ALAS1 genes and the chicken (Lim *et al.*, 1994), mouse (Schoenhaut & Curtis, 1986) and human (Conboy *et al.*, 1992) ALAS2 genes have been determined (Fig. 1.5). Comparison of the genomic organisation of the ALAS genes from various species revealed that they are remarkably conserved and are composed of 10 coding exons and one 5'-untranslated exon. The exception is the chicken ALAS1 gene which is comprised of only 10 coding exons, lacking the intron located in the 5'-untranslated region of the other ALAS genes (May *et al.*, 1995).

Figure 1.5   Structural Organisation of the Genes for ALAS1 and ALAS2

Structural Organisation of the genes for ALAS1 and ALAS2. Exons are numbered. The ATG initiation codon and the TAA termination codon are shown. The proposed functional roles of the exons are indicated. The alternative splicing pattern for human ALAS2 is also shown.
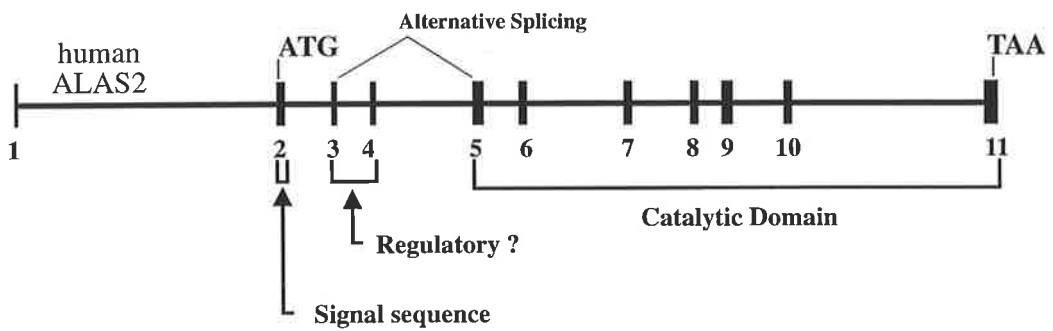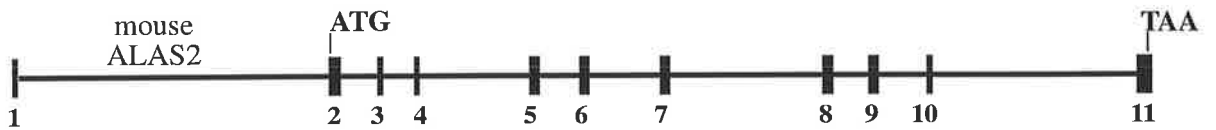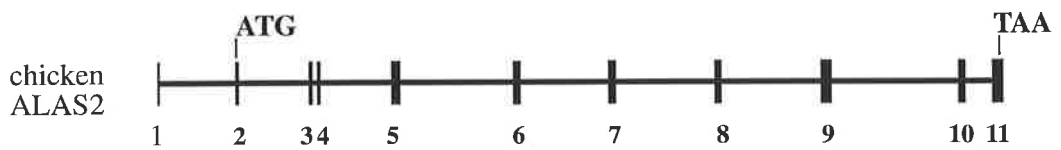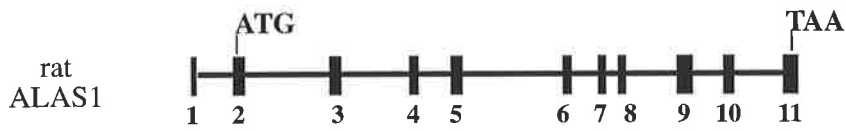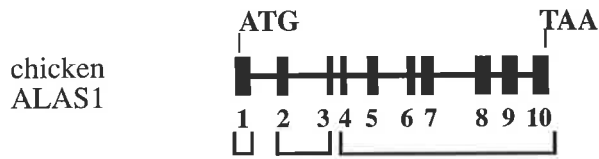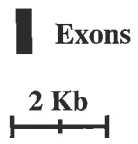
The genomic organisation of these ALAS genes also correlates well with the putative functional domains of the ALAS protein as described in Section 1.7.2. The first exon of the ALAS2 gene is non-coding and is highly conserved in all three species. This exon contains an iron-responsive element involved in the post-transcriptional regulation of the ALAS2 gene (See Section 1.9.2). As seen in Fig. 1.5, region 1, corresponding to the mitochondrial signal sequence of the ALAS enzyme is encoded by exon 2. Region 2, comprising the N-terminal region of the mature ALAS2 protein, is encoded by exons 3 and 4. The proposed catalytic domain of the mature ALAS enzyme (region 3) is encoded by exons 5 to 11.

The structural conservation of the housekeeping and erythroid ALAS genes indicate that these two genes have evolved from a common ancestral gene, possibly with the loss of an intron in the 5'-untranslated region of the chicken ALAS1 gene. The evolution of two genes encoding ALAS would accommodate the distinct regulatory requirements for the synthesis of the ALAS enzyme in erythroid and non-erythroid cells.

Interestingly, two isoforms of mouse ALAS transcripts are generated by alternative splicing of exon 3 (Schoenhaut & Curtis, 1989). Two potential splice acceptor sites (dinucleotide AG) are present within the mouse exon 3 (Schoenhaut & Curtis, 1989). The major upstream site constitutes approximately 85% of total ALAS2 transcripts whereas the minor downstream site comprises the remaining 15% of transcripts (Schoenhaut & Curtis, 1989). In the human ALAS2 sequence, an A to G substitution at the downstream splice acceptor site eliminates the possibility of alternative splicing within exon 3 while the remaining upstream splice acceptor site is functional (Conboy et al., 1992). An A to T substitution (ATAG) identified at the upstream splice acceptor in the human ALAS2 sequence also more closely resembles the consensus N(T/C)AG splice site in comparison to the mouse (TAAG) ALAS2 sequence. In support of these molecular observations, analysis of the human erythroid ALAS mRNA by polymerase chain reaction did not reveal the isoforms detected in the mouse, instead alternative splicing of exon 4 was observed (Conboy et al., 1992). The two isoforms of the human ALAS2 mRNA are expressed throughout erythroid cell differentiation including the early and late erythroid precursor cells and are also detected in the human fetal liver, adult bone marrow and peripheral blood reticulocytes (Conboy et al., 1992). Notably, the alternative splicing of exon 4 was restricted to the human ALAS2 transcript and was not observed in the ALAS2 transcripts

of the mouse or dog. This finding demonstrates species-specific alternative splicing of the ALAS2 transcript, indicating that this mechanism has occurred following the divergence of the human, mouse and dog (Conboy *et al.*, 1992).

Although, the functional implications of the alternative splicing of exon 4 in the human ALAS2 transcript remain unclear, it is interesting to note that exon 4 resides within region 2 whose functional role has not yet been defined (Fig. 1.5). If this region is functionally indepensable, the species-specific splicing may represent neutral intermediates in ALAS gene evolution (Cox *et al.*, 1991). Alternatively, region 2 may function to modulate regions 1 or 3 involved in mitochondrial transport and catalysis since the presence or absence of exon 4 may alter the tertiary structure of the ALAS protein (Cox *et al.*, 1991).

1.8   REGULATION OF ALAS1 GENE EXPRESSION IN HEPATIC TISSUE

ALAS1, initially isolated from the liver, is probably expressed in all tissues to supply heme for mitochondrial respiratory cytochromes and other cellular heme containing proteins (May *et al.*, 1995). The regulation of heme biosynthesis has been extensively studied in hepatic tissue, where activity of ALAS can be induced by porphyrinogenic drugs and repressed by heme (Granick, 1966; Srivastava *et al.*, 1988; 1990). Therefore expression of the ALAS1 gene is subjected to a negative feedback control mechanism where its synthesis is regulated by the end product of the heme biosynthetic pathway, heme. The transcription factors responsible for the basal expression of this gene in various tissues and the mechanisms responsible for drug induction and heme-mediated repression of the ALAS1 gene are discussed.

1.8.1   Transcriptional Regulation of the ALAS1 Gene

To investigate the regulatory sequences required for the basal expression of the rat ALAS1 gene, transient expression studies were performed in different cell types with various lengths of the rat ALAS1 5'-flanking sequence and the first intron which is located in the 5'-UTR of the rat ALAS1 gene (Braidotti *et al.*, 1993). Maximal expression was obtained with sequences located in the promoter region to -479 bp upstream of the

transcription initiation site and the inclusion of intron 1 substantially increased this expression. Sequence analysis of the rat ALAS1 promoter resulted in the identification of two putative binding sites for nuclear respiratory factor 1 (NRF-1) located at -59 bp and -88 bp (Braidotti *et al.*, 1993). Functional binding sites for NRF-1 have been identified in the promoters of genes encoding mitochondrial proteins involved in oxidative phosphorylation, including cytochrome *c*, cytochrome *c* oxidase subunit VIc, ubiquinone-binding protein and the ATP synthase γ-subunit (Virbasius *et al.*, 1993). Therefore to ensure an adequate supply of heme for respiratory cytochromes, NRF-1 may be involved in coordinating expression of these respiratory chain subunits with the rate-limiting enzyme of the heme biosynthetic pathway, ALAS. Indeed, gel shift assays and mutational analysis demonstrated the functional importance of both these NRF-1 binding sites in ALAS1 promoter expression. Binding sites for NRF-1 have also been identified in the promoters of the human and chicken ALAS1 genes highlighting a likely conserved role for NRF-1 in basal expression. In contrast to the identification of functional binding sites for NRF-1 in the promoters of the rat, human and chicken ALAS1 genes, binding sites for NRF-1 have not been identified in the promoter of the ALAS2 gene or in the promoters of ALAD (Kaya *et al.*, 1994), PBGD (Chretien *et al.*, 1988), uroporphyrinogen decarboxylase (Romana *et al.*, 1987) or ferrochelatase (Taketani *et al.*, 1992).

The induction of hepatic ALAS observed in chick embryo liver (Hansen *et al.*, 1989), chick embryo hepatocytes (Hamilton *et al.*, 1992; Dogra *et al.*, 1993) and rat liver (Srivastava *et al.*, 1988) following treatment with porphyrinogenic drugs occurs at the level of transcription. However, the mechanism responsible for the increase in ALAS1 gene transcription remains to be elucidated. Transient expression studies performed in chick embryo hepatocytes with 10 kb of 5'-flanking and 9 kb of 3'-flanking chicken ALAS1 sequence fused to a reporter gene were not induced with phenobarbital (Sadlon, 1997). This result indicated that the drug-responsive enhancer element either does not reside within the 5'- or 3'- flanking sequence or may require the appropriate chromatin configuration or is located elsewhere within the ALAS1 gene. Although the identity of the transcription factors involved in the transcriptional activation of the ALAS1 gene by drugs are unknown, recent evidence has suggested a possible role for phosphorylation. Studies performed by Dogra *et al.* (1996) with 2-aminopurine, an inhibitor of serine/threonine kinases, interfered with the phenobarbital induction of the chicken ALAS1 gene in chick embryo hepatocytes. Therefore

it has been proposed that phenobarbital may be involved in the activation of a kinase resulting in the phosphorylation of a transcription factor involved in the activation of the ALAS1 gene in hepatocytes (Dogra *et al.*, 1996).

In the liver, drugs that induce expression of ALAS1 also induce specific cytochrome P450s which are responsible for the "solubilisation" of drugs to facilitate their excretion in the urine (Gonzalez, 1989). Cytochrome P450s are heme-containing proteins and it has been estimated that at least half of the heme synthesised in the liver is utilised by these cytochrome P450s. Therefore the coordinated regulation of ALAS1 and the genes encoding the drug-inducible cytochrome P450s would ensure the continued supply of heme for these heme-containing proteins. It was initially proposed that pre-existing levels of heme were utilised by the drug-induced cytochrome P450s and thereby reducing the level of heme which in turn resulted in derepression of ALAS1 transcription and an increase in the synthesis of the enzyme (May *et al.*, 1995). However, overexpression of cytochrome P450 cDNAs in chick embryo hepatocytes does not affect the level of ALAS1 mRNA suggesting that the induction of ALAS1 by drugs may not be a direct consequence of an increased synthesis of P450s and utilisation of heme (Jover *et al.*, 1996).

1.8.2   Regulation of the ALAS1 Gene by Heme

Heme can be toxic to cells by interacting with hydrogen peroxide, with the resultant generated reactive oxygen species damaging membrane lipids and proteins (Müller-Eberhard & Fraig, 1993). Porphyrinogens, when oxidised to porphyrins, absorb light also generating dangerous free radicals that cause cutaneous injury (Darr & Fridovich, 1994). In addition, evidence also exists to suggest that excessive synthesis of ALA and PBG may also contribute to neurological abnormalities associated with attacks of porphyria (reviewed by May *et al.*, 1995). Therefore it is critical to maintain appropriate levels of heme and intermediates and this is achieved through the end-product inhibition of ALAS1. Heme is capable of regulating the synthesis of ALAS1 in the liver at a number of levels by directly inhibiting transcription of the ALAS1 gene, affecting the stability of the ALAS1 mRNA and the transport of the ALAS1 precursor into the mitochondria. It has therefore been proposed that there are two pools of heme: one located in the cytosol and the other located in the nucleus which are responsible for monitoring and regulating expression of

both ALAS1 and heme oxygenase, which is involved in heme degradation (Fig. 1.6) (May *et al.*, 1995).

Following the injection of rats with drugs such as 2-allyl-2-isopropylacetamide, there is a marked increased in the level of ALAS1 mRNA in the liver and kidney (Srivastava *et al.*, 1988). However, administration of heme (Srivastava *et al.*, 1990; 1990) or the heme precursor, 5-aminolevulinate (ALA) (Srivastava *et al.*, 1988) prevented this induction by altering the rate of ALAS1 gene transcription as determined by nuclear run-off experiments. Heme also reduced the basal level of ALAS1 expression in all tissues examined, except the erythroid spleen where levels of ALAS were elevated (Srivastava *et al.*, 1988). Consequently, heme is thought to mediate repression of ALAS1 gene expression probably in all non-erythroid tissues. However, the identification of the elements involved in mediating repression of ALAS1 activity at the transcriptional level have been hampered due to the inability to find a transformed cell line in which endogenous ALAS1 gene transcription is repressed by heme (Healy, 1990; Braidotti, 1992).

Treatment of rat hepatoma cells with heme resulted in a significant reduction in the level of ALAS1 mRNA (Sadlon, 1995). The transcription rate of the rat ALAS1 gene was unaffected by heme whereas the half-life of the ALAS1 mRNA was substantially reduced. Treatment of hepatoma cells with actinomycin D to inhibit transcription demonstrated that the half-life of the ALAS1 mRNA was specifically reduced from 130 minutes to 42 minutes by heme. This effect occurred independently of translation since treatment of these cells with cycloheximide did not inhibit the heme-mediated destabilisation of the ALAS1 mRNA (Sadlon, 1995). The region responsible for the heme-mediated instability of the ALAS1 mRNA has been localised to the coding region of the ALAS1 mRNA and this region is currently being investigated further in our laboratory to identify the elements responsible for the heme-mediated destabilisation of the ALAS1 mRNA.

Similar studies have also been performed in chick embryo hepatocytes where heme reduces the half-life of the ALAS1 mRNA from 3.5 hours to 1.2 hours without affecting the transcription rate of the ALAS1 gene (Drew & Ades, 1989; Hamilton *et al.*, 1991). However, in contrast to the rat ALAS1 mRNA, treatment of chick embryo hepatocytes with cycloheximide inhibited the heme-mediated instability of the ALAS1 mRNA suggesting that a labile protein may be involved in this process (Drew & Ades, 1989; Hamilton *et al.*, 1991).

Figure 1.6   Proposed sites of transcriptional and post-transcriptional regulation of ALAS1 by heme in hepatic tissue

The proposed model assumes that there are two intracellular regulatory pools of heme; one located in the nucleus and the second in the cytosol.

NUCLEUS

ALAS1
gene transcription

Heme Pool

ALAS1
mRNA
stability

ALAS1
precursor
transport

Heme Pool

CYTOPLASM
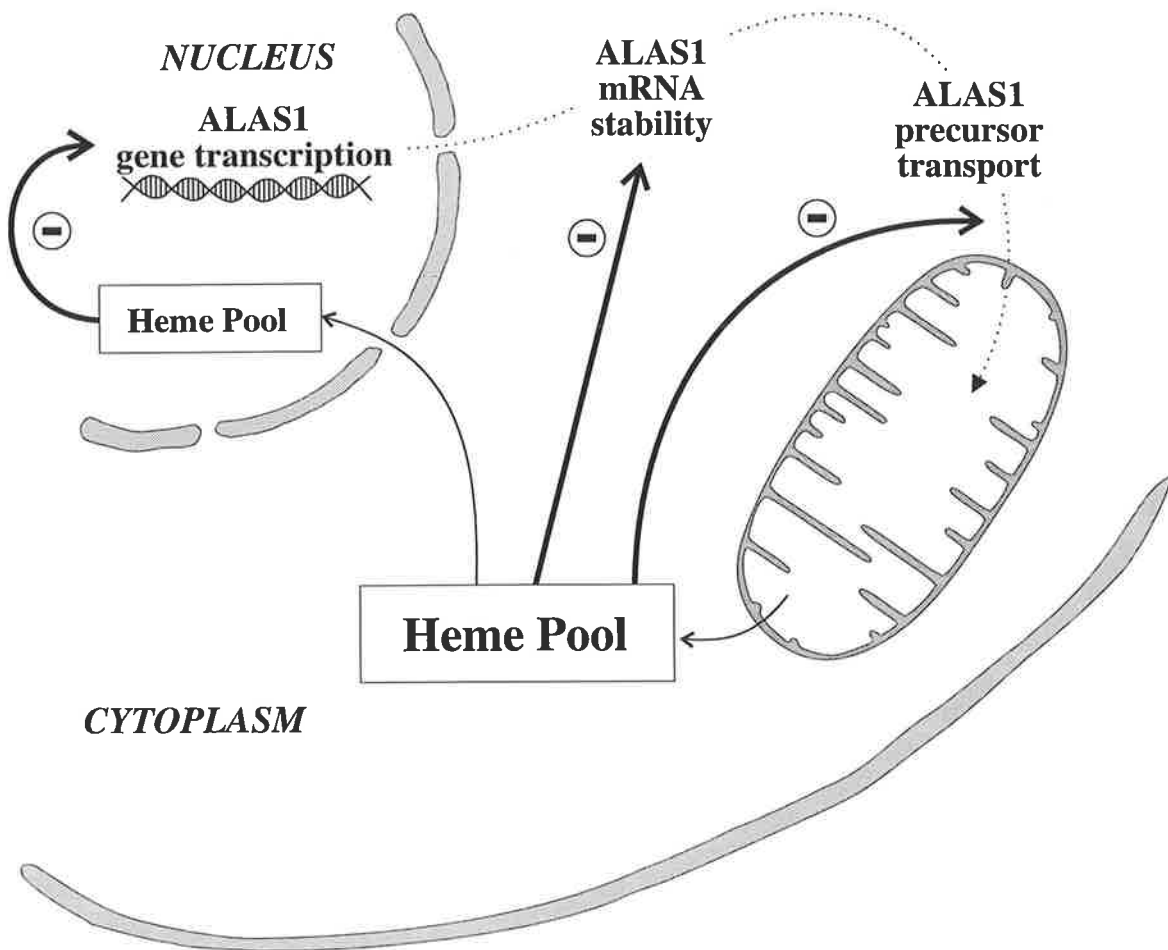
Studies performed in rat (Hayashi *et al.*, 1972) and chick embryo (Srivastava *et al.*, 1983) livers treated with heme demonstrated that the level of ALAS in the mitochondria is rapidly reduced whereas an increase in the level of the ALAS protein is observed in the cytosol. Therefore it is believed that heme is also capable of specifically inhibiting the import of the precursor ALAS1 protein from the cytosol to the mitochondria and would provide an extremely effective control mechanism to reduce the levels of ALAS1 in the mitochondria and consequently the synthesis of heme.

The yeast heme-activated protein (HAP1) is involved in the transcriptional regulation of a variety of nuclear encoded respiratory proteins whose synthesis is coordinated by heme (Pfeifer *et al.*, 1989). Examination of the amino acid sequences of the liver and erythroid ALAS precursor proteins resulted in the identification of three Cys-Pro motifs that resemble the seven Cys-Pro motifs of the heme-binding domain of HAP1. Lathrop & Timko (1993) demonstrated that these Cys-Pro motifs located in the mouse ALAS2 precursor protein were responsible for inhibiting the transport of the precursor into the mitochondria by heme. Similar motifs have also been identified in the signal sequence of the liver ALAS proteins (Dierks, 1990; May *et al.*, 1995) but are not present in the yeast ALAS precursor protein whose import into the mitochondria is not affected by heme (Urban-Grimal *et al.*, 1986).

The precise mechanism responsible for inhibiting the transport of the precursor ALAS protein from the cytosol to the mitochondria remains to be elucidated. However, several mechanisms have been proposed based on information obtained from studying the transport of other mitochondrial matrix proteins (Kiebler *et al.*, 1993). The signal sequence of the precursor is postulated to specifically interact with a surface receptor complex located on the outer mitochondrial membrane and then migrate across the mitochondrial membrane through a protein transport pore. Heme could either directly interfere with the interaction of the signal sequence with the mitochondrial surface receptor or prevent the transfer across the mitochondrial membrane (May *et al.*, 1995). Although, heme is thought to directly bind to the precursor protein and affect mitochondrial transport of the precursor protein, the possible involvement of a heme-binding protein cannot be ignored.

In conclusion, regulation of the rate-limiting enzyme of the heme biosynthetic pathway, ALAS1, by heme would ensure that appropriate levels of heme are synthesised and avoid the toxicity effect observed with heme. To highlight this, at least three

mechanisms have been identified whereby heme is capable of regulating expression of ALAS1 (Fig. 1.6).

1.9   REGULATION OF ALAS2 GENE EXPRESSION IN ERYTHROID TISSUE

At the onset of erythroid differentiation, the marked increase in ALAS2 gene expression (Elferink *et al.*, 1988) together with the genes encoding the other heme pathway enzymes (Beaumont *et al.*, 1984; Romeo *et al.*, 1986; Raich *et al.*, 1989) and for globin (Karlsson & Nienhuis, 1985) occurs primarily at the level of transcription. A number of transcription factors involved in the erythroid-specific regulation of the globin genes may also serve a role in the transcriptional regulation of the ALAS2 gene in erythroid cells. This would ensure that the genes encoding enzymes of the heme biosynthetic pathway and the globin polypeptides are coordinately regulated. Expression of the ALAS2 gene is also regulated at the post-transcriptional level through an iron-responsive element located in the 5'-untranslated region of the ALAS2 mRNA (Cox *et al.*, 1991). Therefore expression of the ALAS2 gene is also tightly coupled to the availability of iron. In addition, ALAS2 and globin expression are also regulated by heme (as described in Section 1.11).

1.9.1   Transcriptional Regulation of the ALAS2 Gene

Studies directed at investigating the *trans*-acting factors responsible for the regulation of the globin and other erythroid-expressed genes have revealed that only a small number of erythroid cell restricted transcription factors are involved and these include GATA-1, NF-E2 and the CACCC box binding protein, EKLF (reviewed by Orkin, 1995a; 1995b).

Although the promoters of the murine (Schoenhaut & Curtis, 1989), human (Cox *et al.*, 1991), and chicken (Lim *et al.*, 1994) ALAS2 genes have been isolated, sequence comparison of their promoters has revealed only minor sequence conservation in putative *cis*-acting elements. In contrast to both the murine (Schoenhaut & Curtis, 1989) and chicken (Lim *et al.*, 1994) ALAS2 promoters which lack a canonical "TATA" box, the human ALAS2 promoter contains a putative GATA-1 binding motif at the TATA box location (Cox *et al.*, 1991). A GATA motif at the TATA box has also been identified in the

promoters of other erythroid-specific genes including the chicken β-globin (Fong & Emerson, 1992), human glycophorin B (Rahuel *et al.*, 1992) and rat pyruvate kinase (Max-Audit *et al.*, 1993) and these non-canonical TATA boxes bind GATA-1 in preference to TFIID. Interestingly, the binding of GATA-1 to the -30 GATA motif in the chicken β-globin promoter prevents the assembly of a repressive nucleosome (Barton *et al.*, 1993) and also imparts erythroid-specificity through the interaction with another GATA-1 molecule bound to the 3'-enhancer (Fong & Emerson, 1992). Therefore GATA-1 is postulated to play a role in the transcriptional initiation of erythroid-cell expressed genes by inhibiting the assembly of a nucleosome at the TATA box. Although the mechanism remains unclear, GATA-1 is then thought to be displaced from the TATA box allowing the binding of TFIID and the transcriptional initiation machinery.

The human ALAS2 promoter also contains two putative GATA-1 binding sites located at -124 bp and -100 bp from the transcription initiation site (Cox, 1993). Interestingly, the -100 GATA site in the human ALAS2 promoter is also present at the corresponding location in the murine ALAS2 promoter and is one of only two putative *cis*-acting elements which is conserved between the human ALAS2 promoter and the published 102 bp of murine ALAS2 promoter sequence. GATA-1 binding sites have been identified in the regulatory regions of many other erythroid-specific genes including the erythroid promoter of human PBGD gene (Mignotte *et al.*, 1989a), the human EpoR (Zon *et al.*, 1991) and the globins (deBoer *et al.*, 1988; Talbot *et al.*, 1990; Philipsen *et al.*, 1990; Pruzina *et al.*, 1991; Gong *et al.*, 1991).

A putative binding site for the erythroid-enriched factor, NF-E2 was identified at -48 bp in the human ALAS2 promoter (Cox *et al.*, 1991) with a mismatch at both extremities of the 11 bp consensus sequence for NF-E2 (Andrews *et al.*, 1993a). Although functional NF-E2 binding sites have been identified in the erythroid promoter of the human PBGD gene (Mignotte *et al.*, 1989b) and in the α- (Strauss *et al.*, 1992; Zhang *et al.*, 1995) and β- (Talbot *et al.*, 1990; 1991; Ney *et al.*, 1990) globin enhancers, binding sites for NF-E2 are not a common occurrence in erythroid cell-specific promoters.

A partially duplicated CACCC box binding site consisting of two overlapping CACCC sequences was identified at -54 bp in the human ALAS2 promoter (Cox *et al.*, 1991). Perhaps importantly, this site also exhibits sequence similarity to the functional CACCC site of the adult β-globin that binds the erythroid CACCC box binding protein,

EKLF (Miller & Bieker, 1993). In contrast, the murine ALAS2 promoter contains a duplicated CACCC box (Schoenhaut & Curtis, 1989), whereas the chicken ALAS2 promoter (Lim *et al.*, 1994) contains numerous CACCC/GC box binding sites presumably for the ubiquitous transcription factor, Sp1 (Kadonaga *et al.*, 1987). The association of GATA motifs with CACCC box binding sites has been reported in many erythroid promoters (deBoer *et al.*, 1988; Mignotte *et al.*, 1989b; Tsai *et al.*, 1991; Zon *et al.*, 1991; Rahuel *et al.*, 1992; Max-Audit *et al.*, 1993) and evidence exists to suggest that the factors binding to these sites functionally interact and play a critical role in the transcriptional activation of erythroid-specific genes (Merika & Orkin, 1995; Gregory *et al.*, 1996).

A putative thyroid hormone responsive element located at -21 bp to -6 bp immediately downstream of the non-canonical TATA box was also identified in the human ALAS2 promoter (Cox *et al.*, 1991). As mentioned in Section 1.5.4, *in vitro* studies performed in chick erythroblasts have demonstrated that the viral oncogene, *v-erbA*, encoding a refractory thyroid hormone receptor can suppress the erythroid-specific transcription of the ALAS2 gene (Zenke *et al.*, 1988) implicating a possible role for thyroid hormone in the regulation of the ALAS2 gene.

A putative binding site for the Ets family of transcription factors is also present in the human ALAS2 promoter, located between the putative -124 and -100 GATA sites (Cox, 1993). As described in Section 1.5.5, several lines of evidence have implicated a role for Ets proteins in the regulation of erythroid-specific genes.

A consensus CCAAT box binding site for the transcription factor, CP1 (Chodosh *et al.*, 1988), identical to the functional CCAAT box binding site identified in both the human (deBoer *et al.*, 1988) and murine (Cowie & Myers, 1988) β-globin promoters, was identified at -84 bp to -90 bp in the human ALAS2 promoter (Cox *et al.*, 1991). A similarly positioned CCAAT box binding site is not present in either the murine (Schoenhaut & Curtis, 1989) or the chicken ALAS2 promoters (Lim *et al.*, 1994).

Transient expression studies performed with the chicken ALAS2 5'-flanking sequence resulted in the identification of two regions required for high level erythroid-specific expression, designated the proximal (-155 bp to +21 bp) and distal (-784 bp to -505 bp) promoter region (Lim *et al.*, 1994). DNase I footprint analysis of the proximal promoter revealed seven protected regions corresponding to a GC-rich binding motif for the ubiquitous transcription factor, Sp1. Another site located at -130 bp was also

identified by footprinting studies. This site exhibited sequence similarity to the consensus AP-1 binding motif and possibly NF-E2, as well as the retinoic acid responsive element (Lim *et al.*, 1994). However, gel shift assays failed to detect binding of either AP-1 or NF-E2 to this site and mutagenesis of this sequence did not affect the erythroid-specific expression of the chicken ALAS2 promoter (Lim *et al.*, 1994). A putative binding site for GATA-1 (-514) was identified in the distal promoter region of the chicken ALAS2 gene but deletion of this element did not affect expression (Lim *et al.*, 1994). Although the chicken ALAS2 promoter contains numerous binding sites for Sp1, the erythroid factor EKLF can also bind related sequences (Hartzog & Myers, 1993; Crossley *et al.*, 1996) and may be a possible candidate transcription factor involved in the erythroid-specific expression of the chicken ALAS2 gene.

As described in Section 1.7.3, the observed remarkable conservation in the structural organisation of the human and murine ALAS2 genes has led to the use of "phylogenetic footprinting" (deBoer *et al.*, 1988) to identify potentially significant regulatory regions in the human ALAS2 gene. DNase I hypersensitivity studies performed on the murine ALAS2 gene resulted in the identification of five hypersensitivity sites present in both uninduced and Me_2SO-induced MEL cells but absent in non-erythroid cells (Schoenhaut & Curtis, 1989). These hypersensitivity sites were located in the immediate promoter, two sites located at the 5' end of intron 1, one site within intron 3 and another site at the 3' end of intron 8 (Schoenhaut & Curtis, 1989). These DNase I hypersensitive sites are indicative of nucleosome-free regions of DNA and usually correspond to regions that interact with *trans*-acting factors critical for transcriptional regulation (Elgin, 1988; Gross & Garrard, 1988). Based on the manageable size of intron 8, sequence comparison of the human (562 bp) and murine (1067 bp) ALAS2 intron 8 sequences was performed (Cox, 1993). This comparison revealed extensive sequence conservation between the human and the 3' end of the murine intron 8 sequences. Interestingly, a cluster of putative binding sites for GATA-1 and CACCC box binding proteins were identified within the intron 8 sequence which were conserved between the two species thus alluding to a possible functional role for these sequences (Cox, 1993). Additional sequence located at the 5' end of the murine ALAS2 intron 8 was not present in the human ALAS2 gene, and showed similarity to a murine-specific repetitive element and therefore likely to have been added to this intron following the divergence of the human and mouse (Cox, 1993). Intron 8 of the human

ALAS2 gene is located approximately 16 kb from the transcription initiation site and may function as a erythroid-specific enhancer responsible for the transcriptional activation of the ALAS2 gene during erythropoiesis.

The primary aims of the work described in this thesis was to investigate the role of these putative *cis*-acting elements identified in the human ALAS2 promoter and to investigate the possible functional role of the putative erythroid-enhancer element identified within intron 8 of the human ALAS2 gene in the transcriptional regulation of the human ALAS2 gene.

1.9.2   Post-transcriptional Regulation of the ALAS2 Gene

As described earlier, around 70% of the total iron in the body is required for the synthesis of hemoglobin during erythropoiesis. Therefore it is essential that there is a coordinated regulation between iron uptake and the synthesis of heme. ALAS2 has been demonstrated to play a key regulatory role in this process (reviewed by May *et al.*, 1995). The role of ALAS2 was initially proposed following the observations that activity of ALAS in erythroid cells was low in iron-deficient patients and markedly increased in patients suffering iron-overload. In addition, following enucleation and the release of the reticulocyte into the circulation, there is continued synthesis of heme and globin which accounts for the final 20% of hemoglobin synthesis. Consequently, regulation of this final hemoglobin synthesis can only occur post-transcriptionally.

As described in Section 1.4.1, expression of transferrin receptors and ferritin are regulated at the post-transcriptional level by an iron-responsive element (IRE) located in their mRNAs. An IRE was identified in the 52 bp 5'-UTR of the human ALAS2 gene (Cox *et al.*, 1991) and exhibits both sequence and structural similarity to the IRE of the transferrin receptor and ferritin mRNAs. As seen in Fig. 1.7, the highly conserved loop is identical between the IREs of the ferritin mRNA and the IRE identified in the human ALAS2 5'-UTR. The 5'-UTR of both the mouse (Dierks, 1990) and chicken (Lim *et al.*, 1994) ALAS2 mRNAs also contain an IRE which is highly conserved between all three species (Fig. 1.7). In contrast, the human (Bawden *et al.*, 1987; Bishop *et al.*, 1990), rat (Srivastava *et al.*, 1988; Yamamoto *et al.*, 1988) and chicken (Borthwick *et al.*, 1985; Maguire *et al.*, 1986) ALAS1 mRNAs do not contain an IRE motif (Cox, 1993).

Figure 1.7   Sequence and proposed structure of an iron-responsive element (IRE)

A   The consensus IRE motif is shown. The critical features of functional IREs include the hexanucleotide loop of sequence 5'-CAGUGN-3', an unpaired cytosine residue located five nucleotide pairs 5' from this loop and a lower stem of variable length.

B   Comparison of the proposed IRE structures in the 5'-untranslated regions of the chicken, mouse and human ALAS2 mRNAs and the human ferritin mRNA. The transcription initiation site (+1), the translation initiation site (AUG) and the conserved bases are shaded. The sequence of the hexanucleotide loop of the chicken ALAS2 IRE differs by a single nucleotide from the consensus.

A

$[N:N]_5$

B

ALAS2 · · · Ferritin

chicken · · · mouse · · · human · · · human
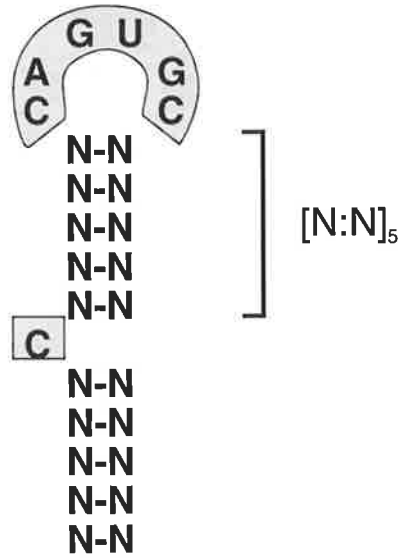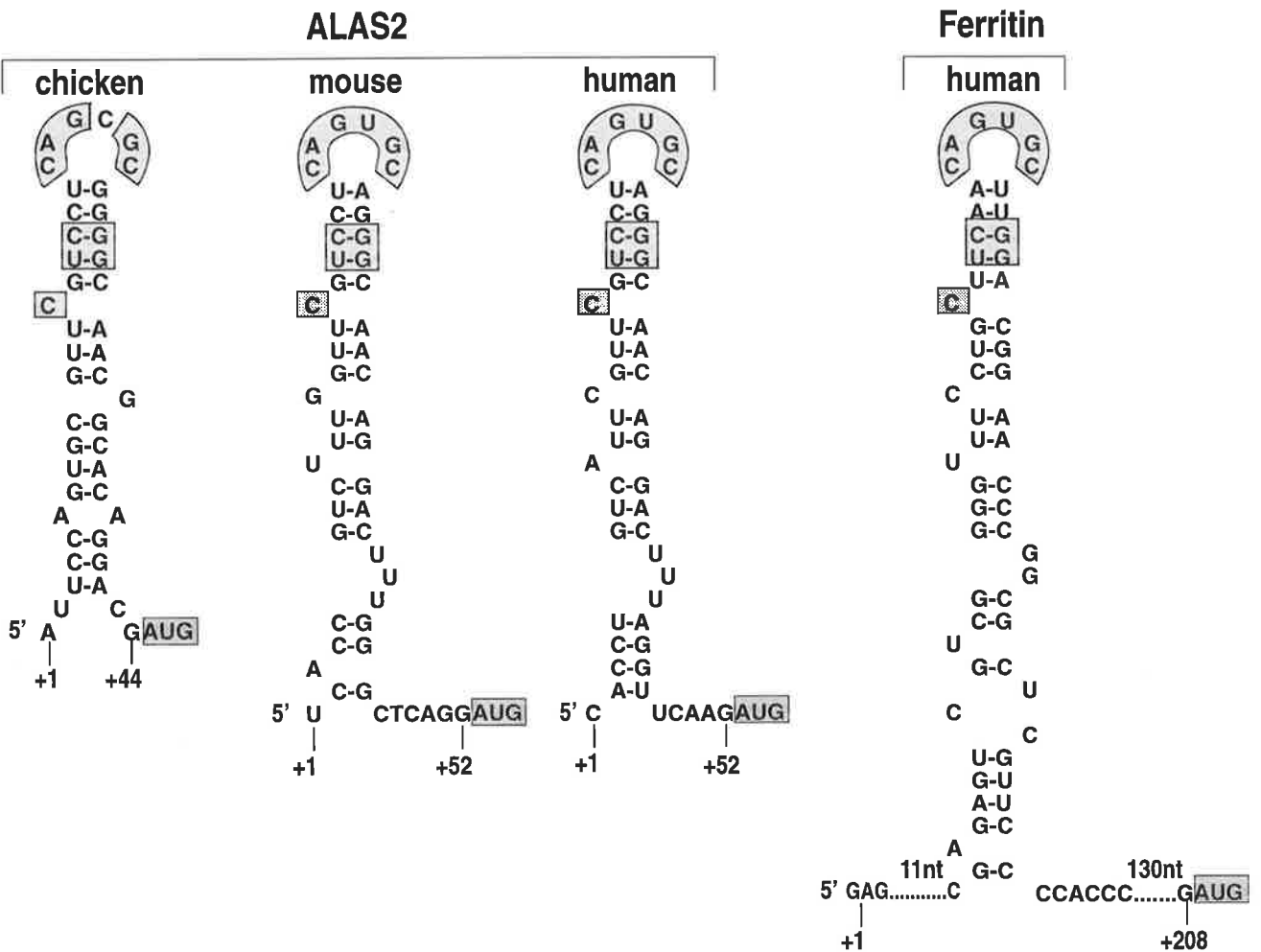
RNA gel shift experiments demonstrated the binding of a cytosolic protein to the human ALAS2 IRE which was most likely identical or at least highly related to the protein that binds to the IREs of the ferritin and transferrin receptor mRNAs (Cox *et al.*, 1991). *In vitro* transcription/translation experiments performed in a cell-free system demonstrated that the human ALAS2 IRE can confer iron-dependent translational control to the heterologous human growth hormone cDNA sequence (Bhasker *et al.*, 1993). The addition of purified IRE-BP to the chimeric transcript could inhibit *in vitro* translation and this was abolished following mutagenesis of the conserved loop (Bhasker *et al.*, 1993). In addition, translation of the ALAS2 mRNA in MEL cells is iron-dependent (Melefors *et al.*, 1993). Although, it is uncertain whether the binding of the IRE-BP to the IRE directly prevents the binding of the small ribosomal subunit with the mRNA or the binding of cap-site initiation factors required for attachment of the small ribosomal subunit, the location of the IRE is critical since both the ferritin (Goossen *et al.*, 1990; Goossen & Henzte, 1992) and ALAS2 (Bhasker *et al.*, 1993) IREs only function when located near the cap site.

As described in Section 1.3, the final stages of erythroid maturation occurs in the circulation where the final 20% of hemoglobin synthesis occurs in the enucleated erythrocyte. Therefore the post-transcriptional regulation of ALAS2 expression would permit the continued synthesis of heme in the absence of gene transcription. The binding activity of the IRE-BP from Epo-induced J2E cells progressively increases during erythropoiesis (Cox, 1993) and is consistent with an important role at the later stages of the differentiation process.

As described in Section 1.7.3, two isoforms of the human ALAS2 transcript are generated by alternative splicing and expressed throughout erythroid differentiation at approximately equal levels (Conboy *et al.*, 1992). However, the functional significance of these two ALAS2 transcripts remain to be elucidated but may represent an additional level of post-transcriptional regulation of the ALAS2 gene.

1.10 OTHER ENZYMES OF THE HEME BIOSYNTHETIC PATHWAY

In contrast to ALAS, there is only one structural gene for the other enzymes of the heme biosynthetic pathway. These genes either have a composite promoter which contain

binding sites for both ubiquitous and erythroid-specific transcription factors or, alternatively, have two separate promoters; one with a housekeeping function and the other that is erythroid cell-specific.

The second and third enzymes of the heme biosynthetic pathway, ALAD and PBGD, respectively, share a common mechanism for expression in erythroid cells. These enzymes are encoded by a single structural gene with two separate promoters that generate two transcripts by 5'-alternate splicing.

The ALAD gene contains two first non-coding exons, 1A and 1B, that are differentially spliced to the second exon containing the translation initiation codon resulting in both transcripts encoding the same amino acid sequence (Bishop et al., 1991; Kaya et al., 1994; Bishop et al., 1996). Tissue distribution studies demonstrated that ALAD mRNA containing exon 1A is ubiquitous, whereas mRNA containing exon 1B is only expressed in erythroid cells. Each of these first exons contain their own promoter and transient expression studies demonstrated that the housekeeping promoter was active in both erythroid and non-erythroid cells, whereas the erythroid-specific promoter was active in only erythroid cells (Kaya et al., 1994).

The human (Kaya et al., 1994) and murine (Bishop et al., 1996) housekeeping promoters located upstream of exon 1A both lack a canonical TATA box but contains many GC boxes reminiscent of other housekeeping gene promoters (Dynan, 1986). In contrast, erythroid-specific promoters of both genes (Kaya et al., 1994; Bishop et al., 1996) located upstream of exon 1B, contain putative binding sites for GATA-1, and CACCC box binding proteins including Sp1 and possibly EKLF. A single putative binding site for the erythroid-enriched factor, NF-E2, was identified 2.3 kb upstream from the murine ALAD housekeeping promoter (Bishop et al., 1996). By analogy to the role of NF-E2 in the LCR of the β-globin gene cluster (see Section 1.5.2), the putative NF-E2 site identified in the murine ALAD promoter may represent a distal enhancer region (Bishop et al., 1996).

In addition to the role that ALAD serves in the synthesis of heme, evidence also exists suggesting other possible functional roles for ALAD. Guo et al. (1994) reported that ALAD is the 240 kDa proteasome inhibitor, CF-2, and functions as an inhibitor of protein degradation via the proteasome. ALAD can bind reversibly to lead and may function as a scavenger to remove lead from the body (Jaffe et al., 1991). Although, the precise functional roles of ALAD remains to be elucidated, it is interesting that only four mutations

in the ALAD gene have been reported, suggesting that mutations in this gene are only rarely viable (Ishida *et al.*, 1992) or show no obvious phenotype.

PBGD also exists in two isoforms, one expressed in all tissues while the second is expressed only in erythroid cells. These isoforms are derived from different mRNAs transcribed from a single gene (Grandchamp *et al.*, 1987). As is the case with ALAD, the two transcripts arise from two distinct promoters, the upstream promoter being active in all tissues whereas the downstream promoter is erythroid-specific (Chretien *et al.*, 1988). Following differential splicing, two mRNAs are produced possessing a specific first exon and thirteen common exons (Chretien *et al.*, 1988). However, in contrast to ALAD, translation is initiated at different sites in the two transcripts. An additional AUG translation initiation codon located in the alternatively spliced exon of the ubiquitously expressed transcript results in the housekeeping isozyme containing an additional 17 N-terminal amino acids not present in the erythroid-specific isozyme (Chretien *et al.*, 1988).

The erythroid-specific promoter of the human PBGD gene contains a functional binding site for GATA-1 (-70), NF-E2 (-160) and a CACCC box (Mignotte *et al.*, 1989b). Intriguingly, differences between the erythroid-specific promoters of the human and murine PBGD genes have been identified. The functional NF-E2 site in the human erythroid promoter is absent in the corresponding region of the murine erythroid promoter and a functional double CACCC box rather than a single CACCC box is present in the murine erythroid promoter of the PBGD gene (Porcher *et al.*, 1991). A functional binding site for GATA-1 has also been identified adjacent to the functional CACCC box in the promoter of the murine PBGD gene and together these sites contribute to erythroid-specific expression (Porcher *et al.*, 1991).

The remaining five enzymes of the heme biosynthetic pathway: uroporphyrinogen III synthase (UROS) (Warner *et al.*, 1990), uroporphyrinogen III decarboxylase (UROD) (Romana *et al.*, 1987), coproporphyrinogen oxidase (CPX) (Delfau-Larue *et al.*, 1994), protoporphyrinogen oxidase (PPOX) (Taketani *et al.*, 1995) and ferrochelatase (Taketani *et al.*, 1992) are all transcribed from a single gene promoter.

Expression of the human UROD gene, the fifth enzyme of the heme biosynthetic pathway is transcriptionally upregulated during erythropoiesis but only a single mRNA species is detected in both erythroid and non-erythroid cells (Romana *et al.*, 1987). Although the promoter contains a TATA box (-21) and a GC box binding site for Sp1

(-60), no erythroid-specific transcription factor binding sites have been identified in this promoter. Therefore it has been suggested that erythroid-specific enhancer elements located elsewhere may be responsible for the increased transcription of the UROD gene during erythropoiesis (Romana *et al.*, 1987).

Expression of the CPX gene is transcriptionally upregulated during the erythroid induction of MEL cells (Kohno *et al.*, 1993). The level of CPX is also higher in erythroid cells in comparison to non-erythroid cells (Martasek *et al.*, 1994). These findings suggest that expression of CPX is differentially regulated in erythroid and non-erythroid cells at the level of transcription. The promoter of the human CPX gene lacks consensus TATA and CCAAT boxes but contains two putative CACCC box binding sites for Sp1 and four putative GATA-1 binding sites (Delfau-Larue *et al.*, 1994) although the functionality of these sites in this promoter remains to be determined.

The CPX gene does, however, contain two polyadenylation signals which probably give rise to the two cDNAs differing at their 3' end isolated from placenta (Taketani *et al.*, 1994) and fibroblast (Martasek *et al.*, 1994) libraries. Whether these transcripts are regulated at the post-transcriptional level, that is, through instability elements present in the 3'-UTR affecting the stability of the CPX mRNA, remains to be determined. It is possible that a similar mechanism as described for the ferrochelatase gene (Ponka & Schulman, 1993) may be acting whereby the stability of the CPX transcript is increased during erythroid differentiation in comparison to non-erythroid cells.

PPOX is synthesised from a single mRNA expressed in both erythroid and non-erythroid cells. The PPOX promoter contains binding sites for both ubiquitous transcription factors including CCAAT and CACCC box binding proteins and two putative binding sites for the erythroid-enriched factor, GATA-1 (Taketani *et al.*, 1995).

cDNA clones encoding the human (Nakahashi *et al.*, 1990) and murine (Taketani *et al.*, 1990; Brenner *et al.*, 1991) ferrochelatase, the last enzyme of the heme biosynthetic pathway, resulted in the identification of two mRNAs differing in their 3' end which are transcribed from a single gene (Brenner *et al.*, 1992). Two putative polyadenylation sites identified in the murine ferrochelatase cDNA may account for the presence of these two transcripts (Taketani *et al.*, 1990).

Northern blot analysis of murine ferrochelatase mRNA demonstrated that both transcripts (2.9 kb and 2.2 kb in size) were expressed in all tissues. However, the 2.9 kb

transcript was more abundantly expressed in non-erythroid tissues in comparison to the 2.2 kb transcript which was predominantly expressed in the spleen, a site of erythropoiesis in the mouse (Ponka & Schulman, 1993). Interestingly, levels of the 2.2 kb transcript was preferentially increased in MEL cells induced to differentiate with Me_2SO and was also the predominant transcript detected in mouse reticulocytes (Chan *et al.*, 1993b; Ponka & Schulman, 1993).

The 2.9 kb transcript contains two AUUUA pentamers and many (A+U)-rich sequences present in the 3' UTR (Taketani *et al.*, 1990) which are absent in the 2.2 kb transcript, suggesting that these ferrochelatase transcripts may be differentially regulated at the post-transcriptional level. Ponka and Schulman (1993) demonstrated that the 2.9 kb transcript was undetectable in uninduced MEL cells maintained in the presence of actinomycin D, an inhibitor of RNA synthesis after 8 hours. In contrast, the 2.2 kb transcript was more stable in Me_2SO-induced MEL cells. The half life of the 2.9 kb transcript was estimated to be 4 hours in uninduced MEL cells, whereas the half-life of the 2.2 kb transcript in Me_2SO-induced MEL cells was greater than 10 hours. Therefore removal of destabilising sequences from the 2.2 kb ferrochelatase transcript resulted in an increased stability of this transcript in differentiating erythroid cells, and this may represent an important mechanism for ensuring the continued synthesis of ferrrochelatase in erythroid cells.

The finding that there are no other differences observed between ferrochelatase species from erythroid and non-erythroid cells is consistent with immunological studies with antisera against purified ferrochelatase enzyme. These studies concluded that the enzymes isolated from erythroid and non-erythroid tissues were indistinguishable (Nakahashi *et al.*, 1990).

The human ferrochelatase promoter which is transcribed in both erythroid and non-erythroid cells, lacks a TATA box and contains putative binding sites for GATA-1, NF-E2 and CACCC box binding proteins (Tugores *et al.*, 1994). A negative regulatory region postulated to bind a repressor protein was identified downstream of the binding sites for GATA-1 and NF-E2 and deletion of this region resulted in a significant increase in erythroid-specific expression in ferrochelatase promoter constructs containing intact binding sequences for GATA-1 and NF-E2 (Tugores *et al.*, 1994). However, the factor binding to this site has not been characterised further.

The regulatory elements responsible for expression of the human gene encoding UROS, the fourth enzyme of the heme biosynthetic pathway have not been characterised (Warner et al., 1990).

1.11   ROLE OF HEME IN THE REGULATION OF HEMOGLOBIN SYNTHESIS

Evidence obtained from studies performed in erythroid cell lines has implicated a possible role for heme in the regulation of ALAS2, other enzymes of the heme biosynthetic pathway and globin.

The MEL cell line, DR-1, which fails to differentiate following treatment with Me_2SO does not express ALAS2 mRNA but transcripts encoding ALAD, PBGD and ferrochelatase are detectable in these cells (Fujita et al., 1991). Genomic Southern blot analysis of the ALAS2 gene in these cells indicated that there is neither a major deletion nor rearrangement in the ALAS2 gene in DR-1 cells in comparison to the parental MEL cell line (Fujita et al., 1991). Northern blot analysis of the Me_2SO-treated DR-1 cell line demonstrated that the level of β-globin mRNA is down-regulated and that treatment with hemin marginally increases this level (Fukuda et al., 1994). These findings suggest that the level of β-globin mRNA is partially dependent on the availability of intracellular heme. In addition, hemin is also necessary for a sustained increase in the level of mRNA encoding ferrochelatase in differentiating Me_2SO-induced MEL cells (Fukuda et al., 1993). Suppression of ALAS2 expression in MEL cells using an antisense oligonucleotide strategy correlated with a decrease in the levels of transcripts encoding other enzymes of the heme pathway, including ALAD, PBGD and ferrochelatase, as well for β-globin and p45, the erythroid-specific subunit of NF-E2 (Meguro et al., 1995). Interestingly, the addition of hemin to K562 cells can stimulate expression of globin genes containing functional NF-E2 binding sites (Palma et al., 1994) suggesting that regulation of some erythroid-specific genes by heme may be mediated via NF-E2. However, the mechanism responsible for this is unclear.

1.11.1 Regulation of ALAS2 by Heme

As described in Section 1.8.2, heme is capable of regulating ALAS1 gene transcription (Srivastava *et al.*, 1988; 1990), ALAS1 mRNA stability (Sadlon, 1995) and inhibiting the transport of the ALAS1 precursor protein from the cytosol to the mitochondria (Hayashi *et al.*, 1972; Srivastava *et al.*, 1983). There is no evidence to suggest that ALAS2 is regulated by heme at the transcriptional level or by an instability element located in the ALAS2 mRNA, since it is an extremely stable message in differentiating erythroid cells (Cox, 1993; Klinken, 1996). However, as described in Section 1.8.2, putative heme-binding motifs have been identified in the signal sequence of the mouse ALAS2 precursor protein and are responsible for the heme-mediated inhibition of transport into the mitochondria (Lathrop & Timko, 1993). In agreement with this, the addition of succinylacetone, an inhibitor of protoporphyrin (and thus heme) synthesis, to Me_2SO-induced MEL cells significantly increases ALAS activity (approximately 10-fold) (Beaumont *et al.*, 1984; Elferink *et al.*, 1988) but not levels of ALAS2 mRNA in comparison to control Me_2SO-induced MEL cells (Fujita *et al.*, 1991). Conversely, the addition of heme prevents the increase in ALAS2 activity (Beaumont *et al.*, 1984; Elferink *et al.*, 1988).

1.11.2 Translation of Globin is Regulated by Heme

Evidence obtained from studies performed nearly 30 years ago implicated a possible role for heme in the translational regulation of the globins where maintenance of globin synthesis in rabbit reticulocytes was demonstrated to be dependent on the availability of hemin (Kruh & Borsook, 1956; Bruns & London, 1965). In the absence of iron for heme synthesis (Waxman & Rabinovitz, 1965) or hemin itself (Waxman & Rabinovitz, 1965; Grayzel *et al.*, 1966; Waxman & Rabinovitz, 1966), the cellular polyribosomes disaggregate reducing globin synthesis (Waxman *et al.*, 1967) but can be reversed with the addition of hemin (Gross & Rabinovitz, 1972). Therefore the regulation of globin synthesis by heme is mediated by a heme-sensitive translational repressor that functions at the level of peptide chain initiation (Gross & Rabinovitz, 1972). Despite these early observations, the mechanism responsible for the heme-mediated translational control of globin synthesis has only been elucidated during the last decade.

During initiation of protein synthesis, eukaryotic initiation factor 2 (eIF-2) binds GTP and Met-tRNA as a ternary complex and then forms the 43 S pre-initiation complex with the 40 S ribosomal subunit (Chen & London, 1995). In heme deficiency, the heme-regulated inhibitor (HR1), a cyclic AMP-independent protein kinase phosphorylates the α-subunit of eIF-2 resulting in the inhibition of translation of globin mRNA and other erythroid mRNAs. Hemin has been shown to bind to HR1 (Fagard & London, 1981) and promote intersubunit bond formation between two HR1 molecules to form an inactive dimer (Yang et al., 1992). In heme-deficient cells, HR1 exists as an active dimer capable of phosphorylating eIF-2α and inhibiting the initiation of translation (Yang et al., 1992). In conclusion, the phosphorylation of eIF-2α by HR1 is the principal mechanism responsible for the regulation of protein synthesis in heme deficiency.

HRI is expressed in the bone marrow and peripheral blood and in both uninduced erythroid MEL and K562 cell lines and is increased during erythroid differentiation (Crosby et al., 1994). Treatment of Me$_2$SO-induced MEL cells with 3-amino-1,2,4-triazole (AT), an inhibitor of heme synthesis, significantly reduced the accumulation of the HR1 mRNA in differentiating MEL cells (Crosby et al., 1994). Similar effects were observed with globin mRNA in Me$_2$SO-induced MEL cells following AT treatment (Mager & Bernstein, 1979). Therefore the increased levels of both HRI and globin mRNAs during Me$_2$SO-induced MEL cells are dependent on heme and suggests a direct relationship between intracellular heme levels and the accumulation of HR1 in MEL cells.

In conclusion, heme is capable of regulating its own synthesis possibly by preventing import of the precursor ALAS2 protein into the mitochondria, and also translation of globin mRNA in erythroid cells. A model has been proposed (May et al., 1995) and is represented in Fig. 1.8.

1.12   DISEASES RELATED TO HEME BIOSYNTHESIS

The "porphyrias" represent a diverse group of inherited disorders where there is an accumulation of various porphyrins or porphyrin precursors (Kappas et al., 1989). Since the primary sites of heme synthesis are the liver and the erythron, porphyrias are classified as either hepatic or erythropoietic depending on the organ primarily affected by the

Figure 1.8   Proposed model for the translational regulation of ALAS2 and globin mRNAs by heme in erythroid cells

Epo activates the transcription of the genes encoding ALAS2, transferrin receptor (TfR) and globin. The pool of intracellular iron regulates translation of ALAS2 mRNA and is incorporated into protoporphyrin to generate heme. The cytosolic pool of heme may regulate its own synthesis, possibly by preventing the import of precursor ALAS2 into the mitochondria and the translation of globin mRNA.

NUCLEUS

Epo action
via GATA-1, etc

DNA

Globin mRNA

CYTOPLASM

TfR

Iron-Tf

IRE-BP

Tf

IRE-BP

ALAS2 mRNA

TfR

translation

Iron
Pool

TfR

pre ALAS2

translation
Globin mRNA

ALAS2

Heme
Pool
+
Globin

Heme

MITOCHONDRION

Hemoglobin
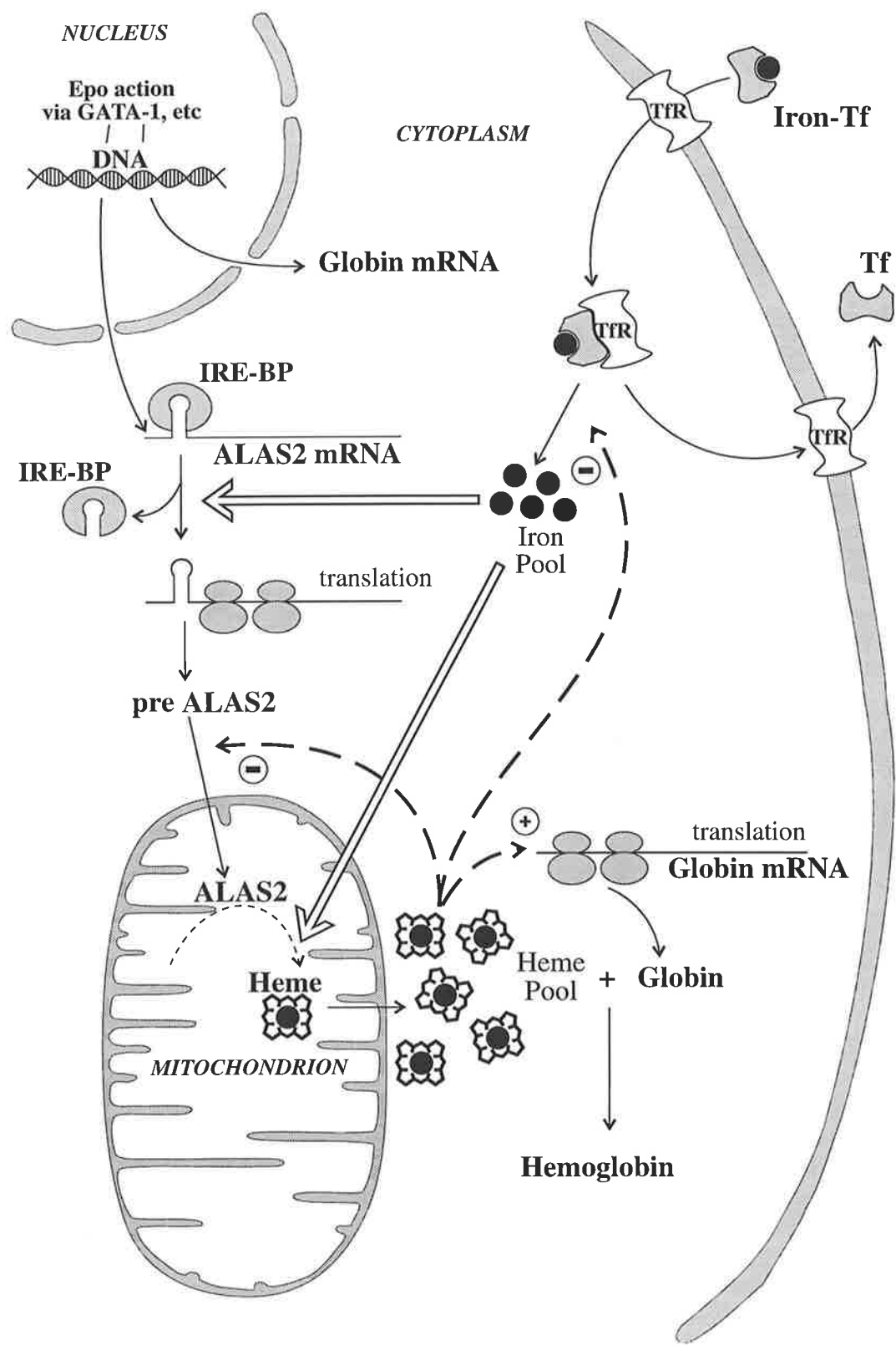
accumulation of porphyrins or precursors. Porphyrias are characterised by a decrease in the activity of one or more of the heme biosynthetic pathway enzymes with the exception of ALAS (for a review, the reader is referred to Bottomley *et al.*, 1995; May *et al.*, 1995). There have been no reports of a disorder associated with a defect in ALAS1 probably because the normal ALAS allele may be able to compensate for the mutant allele in the heterozygous state and/or it is lethal in the homozygous individual.

1.12.1 Sideroblastic Anemias

The sideroblastic anemias comprise a heterogenous group of erythropoietic disorders that are either inherited or acquired. These anemias are characterised by defective circulating erythrocytes which are small (microcytic) and lack hemoglobin (hypochromic) and reflect a defect in the synthesis of hemoglobin (Bottomley, 1993). A characteristic feature identified in all sideroblastic anemias is the presence of bone marrow erythroblasts with large, perinuclear, Prussian blue-positive granules representing amorphous iron deposits in the mitochondria. These iron-laden erythroblasts are termed sideroblasts (Bottomley *et al.*, 1995). In severe cases of sideroblastic anemia, residual iron-laden mitochondria (Pappenheimer bodies) are also detected within circulating erythrocytes. Although the synthesis of globin is also impaired in erythroid cells of these patients, *in vitro* studies have demonstrated that their reticulocytes retain the capacity to synthesise hemoglobin with the addition of exogenous heme and is therefore thought to be secondary to an underlying defect in heme synthesis (Bottomley, 1993). In addition, the delivery of iron to erythroid cells via the iron-transferrin complex is also unaffected (Bottomley, 1993). Consequently, the accumulation of mitochondrial iron in sideroblasts is thought to be due to a defect in protoporphyrin synthesis, where iron is imported into the cell but is not incorporated into protoporphyrin IX.

The most common form of inherited sideroblastic anemia follows an X-linked mode of inheritance and is manifested almost exclusively in males. A defect in the first enzyme of the heme biosynthetic pathway, ALAS, has been implicated as the cause of X-linked sideroblastic anemia (XLSA) since the gene encoding ALAS2 is located on the X chromosome (Cox *et al.*, 1990), and approximately one-third of patients respond to treatment with pyridoxine, the precursor of pyridoxal phosphate which is the essential

cofactor for ALAS activity (Bottomley, 1993). Those patients with XLSA who are unresponsive to treatment with pyridoxine may require blood transfusion. However, while this alleviates the symptoms of the anemia it only adds to the iron-overload and in turn causes tissue damage. Therefore the iron chelating agent desferrioxamine may be administered or regular phlebotomies performed throughout life.

To investigate whether mutations in the ALAS2 gene indeed underly XLSA, ALAS2 mRNA was isolated from reticulocytes from a number of XLSA patients and sequenced. In one patient whose level of ALAS activity was 50% of the normal level, and following administration of pyridoxine the enzyme activity was restored to normal levels (Cox et al., 1994), sequence analysis was performed and resulted in the identification of a single base change (C to G) at position 388 in exon 8 (Cox et al., 1994). The C to G mutation resulted in the replacement of an invariant threonine residue located three residues from the lysine that binds pyridoxal phosphate by a serine residue (Cox et al., 1994). Expression studies performed in E. coli demonstrated that the mutant ALAS2 enzyme has lower activity when compared to the wild-type (Cox et al., 1994). Further support for the causative role was obtained by the demonstration that all similarly affected relatives of the patient also carried this altered ALAS2 allele (Cox et al., 1994). ALAS kinetic assays have shown that both the mutant ALAS and wild-type ALAS enzymes respond to pyridoxal phosphate indicating that the level of the cofactor for ALAS, pyridoxal phosphate is rate-limiting (Chandler, 1996; Matthews, 1997). A number of other single base changes have been identified in the ALAS2 gene in patients with XLSA and these are located in exons 5, 7 and 9 (May et al., 1995). However, whether these mutations affect the catalytic activity of ALAS or reduce protein stability is currently under investigation in our laboratory.

A model for the role of a defective ALAS2 protein in inherited XLSA has been proposed (Fig. 1.9) (May et al., 1995). As a result of a mutation in the ALAS2 gene, ALAS activity is reduced leading to a decrease in the synthesis of protoporphyrin IX and hence heme, and globin chains in erythroid cells. However, in response to low heme levels, iron continues to enter into the mitochondria resulting in its accumulation and the characteristic feature associated with sideroblastic anemia: ring sideroblasts (sideroblasts with iron-laden mitochondria surrounding the nucleus).

Figure 1.9   Proposed model for reduced hemoglobin synthesis in X-linked Sideroblastic Anemia

A mutation in the ALAS2 enzyme resulting in reduced catalytic activity, in turn leads to a decrease in the synthesis of protoporphryin and hence heme, and globin chains in erythroid cells. However, iron continues to enter into the erythroid cell, where it accumulates within the mitochondria. The abbreviations used are as follows: **Tf,** transferrin; **TfR,** transferrin receptor; **ALA,** 5-aminolevulinic acid; **PP,** protoporphyrin.

Tf

Iron-Tf

TfR

TfR

TfR

Plasma Membrane

Cytosolic Iron

Globin mRNA

Iron accumulates

mutant ALAS2 → ALA → → PP → Heme

Globin
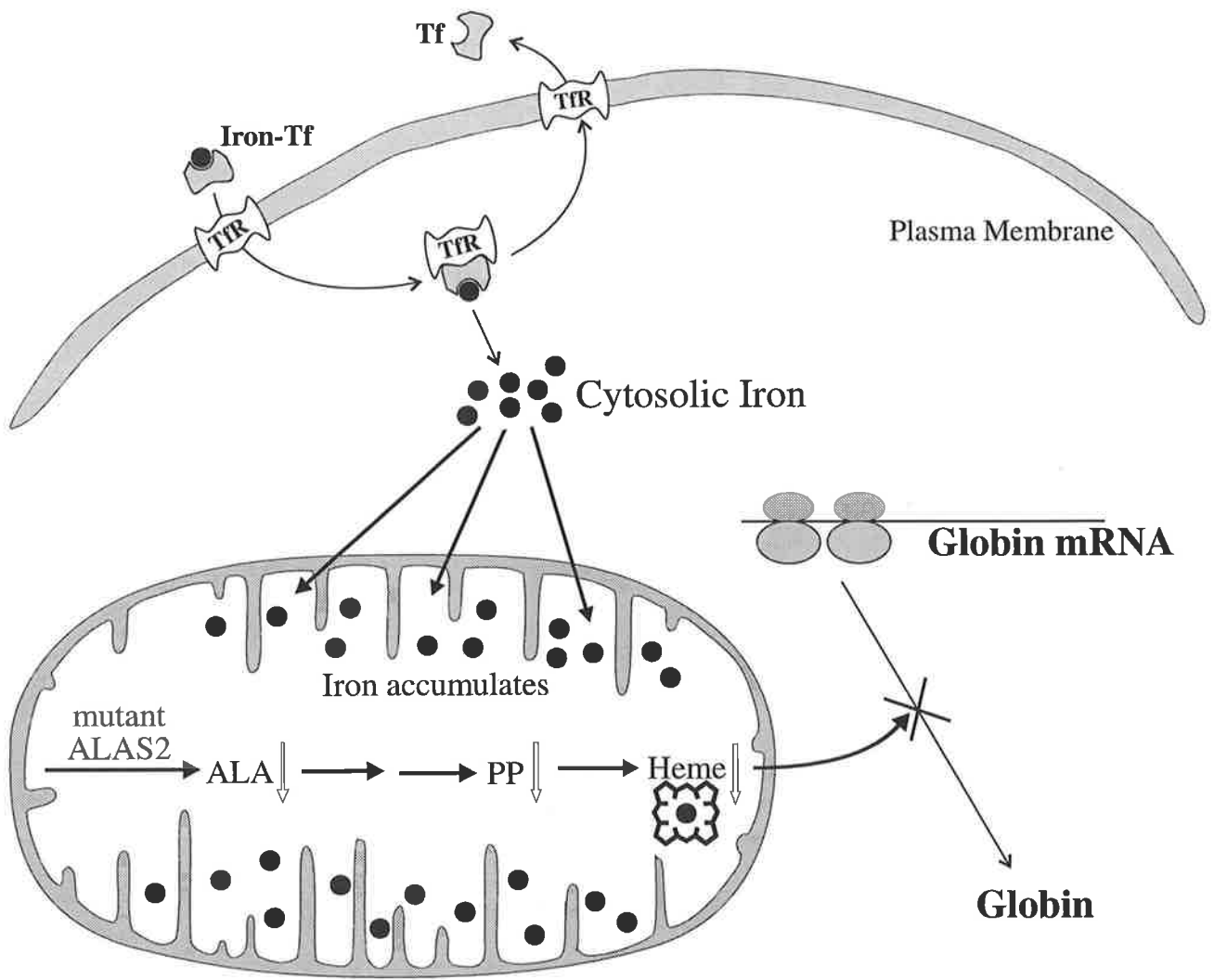
Treatment of rabbit reticulocytes with succinylacetone, an inhibitor of heme synthesis also results in an accumulation of iron in the mitochondria (Adams *et al.*, 1989). Therefore it is believed that iron by an unknown mechanism continues to enter the cell in the absence of its utilisation. Although it has been demonstrated that heme is capable of preventing import of the ALAS2 precursor protein from the cytosol to the mitochondria (Lathrop & Timko, 1993), an increase in the import of the precursor ALAS2 protein would be expected in the presence of reduced levels of heme. However, high levels of iron may inhibit ALAS2 activity or the import of ALAS2 into the mitochondria, resulting in inadequate levels of ALAS2 (May *et al.*, 1995).

1.13   AN *IN VITRO* MODEL TO INVESTIGATE ERYTHROID-SPECIFIC GENE EXPRESSION

The development of transformed erythroid cell lines which can be induced to differentiate *in vitro* and mimic the changes seen during erythropoiesis has provided the opportunity to investigate the regulatory processes which occur during the latter stages of erythroid differentiation.

ALAS1 and ALAS2 mRNAs are differentially expressed during Me_2SO induced erythroid differentiation of MEL cells. The level of ALAS1 mRNA rapidly decreases while ALAS2 mRNA and transferrin receptor (TfR) mRNA levels are markedly elevated and this apparently precedes globin gene transcription (Fujita *et al.*, 1991). Similarly, the levels of ALAS2 and TfR mRNAs rapidly increase in Epo-induced J2E cells in contrast to the transcription of the globin genes which exhibit a gradual increase (Busfield *et al.*, 1992).

The erythroid-enriched transcription factor, GATA-1 plays a critical role in the transcriptional activation of many erythroid cell specific genes including the enzymes of the heme biosynthetic pathway (Mignotte *et al.*, 1989b; Tugores *et al.*, 1994) and the globins (deBoer *et al.*, 1988; Talbot *et al.*, 1990; Philipsen *et al.*, 1990; Pruzina *et al.*, 1991; Gong *et al.*, 1991). However, in contrast to the significant increase in globin mRNA observed in MEL cells chemically induced to differentiate, the level of GATA-1 mRNA marginally decreases although the level of GATA-1 protein remains similar throughout erythroid differentiation (Busfield *et al.*, 1995). The levels of GATA-1 mRNA in Epo-induced J2E

cells also increase rapidly prior to the increase in the transcripts of the enzymes of the heme pathway and globin (Klinken *et al.*, 1993; Busfield *et al.*, 1995). Similar results have been obtained following addition of Epo to erythroid progenitor cells where there is a dramatic increase in the levels of GATA-1 mRNA immediately preceding the major rise in globin gene mRNA (Dalyot *et al.*, 1993). However, the level of GATA-1 transcripts decrease at approximately 72 hour in Epo-induced J2E cells whereas the levels of β-globin transcripts remain elevated suggesting that the continued synthesis of GATA-1 is not required (Busfield *et al.*, 1995).

The K562 cell line has been extensively used to investigate erythroid and megakaryocytic differentiation and provides an *in vitro* model system to delineate the activation and downregulation of lineage restricted transcription factors during hematopoietic commitment (Green *et al.*, 1993). However, the K562 cell line only expresses the embryonic ε- and fetal γ- globin genes and the gene coding for β-globin is not expressed in these cells even following induction by hemin (Rutherford *et al.*, 1979; Benz *et al.*, 1980). Examination of the β-globin gene in K562 cells has revealed that it is not transcribed (Charnay & Maniatis, 1983) although restriction mapping revealed no major insertions, deletions or rearrangements in or adjacent to the gene (Dean *et al.*, 1983; Mueller *et al.*, 1983). Retroviral-mediated transfer studies of the human β-globin gene from K562 cells into MEL cells performed by Weber-Benarous *et al.* (1988) demonstrated that the β-globin gene was transcribed and induced following treatment of the MEL cells with Me$_2$SO and is regulated in an identical manner to a control human β-globin gene construct transferred into the MEL cells using the same procedure. Interestingly, Northern blot analysis using mRNA isolated from K562 and MEL cells and a probe for the erythroid specific transcription factor, EKLF, revealed that K562 cells synthesise lower levels of EKLF mRNA in comparison to MEL cells (Donze *et al.*, 1995). However, K562 cells permanently transfected with an EKLF expression vector failed to activate the endogenous β-globin gene (Donze *et al.*, 1995) indicating that other factors in addition to EKLF may be required.

In conclusion, the development of transformed erythroid cell lines which can be induced to differentiate *in vitro* and closely resemble the morphological and biochemical processes observed during erythropoiesis provide an extremely suitable *in vitro* model system to investigate the regulatory processes occurring during erythroid differentiation.

These cell lines provide an abundant source of material and can be readily manipulated. The MEL and K562 cell lines have been successfully used to identify and characterise the regulatory regions of erythroid-specific genes including the human β-globin (deBoer et al., 1988) and γ-globin (Fischer et al., 1993), EpoR (Zon et al., 1991) and rat pyruvate kinase (Max-Audit et al., 1993).

1.14 AIMS OF THIS THESIS

ALAS is considered to catalyse the rate-limiting step of the heme biosynthetic pathway. During erythropoiesis, large quantities of heme are required and its synthesis must be coordinately regulated with that of globin for subsequent hemoglobin assembly. The ALAS enzyme is encoded by two genes; ALAS1 is ubiquitously expressed (Elliott et al., 1989) whereas ALAS2 is specifically expressed in erythroid cells (Cox et al., 1991). Therefore distinct regulatory mechanisms are postulated to be responsible for the tissue-specific regulation of the ALAS2 gene, in contrast to the ALAS1 gene.

Expression of the ALAS2 gene is regulated at both the transcriptional and post-transcriptional level. Translation of the ALAS2 mRNA in erythroid cells is controlled by intracellular iron levels and is mediated by an iron-responsive element (Cox et al., 1991) to ensure that the production of protoporphyrin IX is coordinated to iron availability. The transcriptional upregulation of the ALAS2 gene during erythropoiesis is accompanied by an increase in the transcription of the other genes of the heme biosynthetic pathway and globin (Beaumont et al., 1984; Karlsson & Nienhuis, 1985; Romeo et al., 1986; Elferink et al., 1988; Raich et al., 1989). Three critical DNA binding motifs responsible for the erythroid-specific expression of the globin gene clusters have been identified and bind the erythroid-enriched factors, GATA-1, NF-E2 and the CACCC box binding protein, EKLF (reviewed by Orkin, 1995a). Binding sites for GATA-1, NF-E2 and CACCC box binding proteins have been identified in the regulatory regions of many other erythroid-specific genes (Mignotte et al., 1989b; Tugores et al., 1994). Therefore during erythropoiesis a common set of transcription factors are thought to be involved in the coordinated synthesis of heme and globin, as well as other erythroid-specific genes.

The work presented in this thesis was directed at investigating the molecular mechanism responsible for the transcriptional activation of the ALAS2 gene during erythropoiesis. The determination of the complete sequence of the human ALAS2 locus by our laboratory (Cox *et al.*, 1991; 1993) has facilitated these studies. Regions of the human ALAS2 gene corresponding to DNase I hypersensitivity sites in the mouse ALAS2 gene were initially investigated. Erythroid transformed cell lines were utilised to identify the critical *cis*-acting elements and the transcription factors that bind to these sites which are responsible for the erythroid-specific expression of the ALAS2 gene. Numerous putative transcription factor binding sites located within the immediate ALAS2 promoter were characterised. These studies also resulted in the demonstration of an erythroid-specific enhancer element located within intron 8 of the human ALAS2 gene, and the subsequent characterisation of *cis*-acting elements that constitute this enhancer element. Comparison of the human ALAS2 intron 8 enhancer with the murine and canine ALAS2 intron 8 sequences revealed a high degree of sequence conservation and this erythroid-specific enhancer element is also shown to be functional in the context of the murine ALAS2 gene.

The elucidation of the transcription factors responsible for the transcriptional upregulation of the human ALAS2 gene during erythropoiesis ultimately contributes to our understanding of how expression of other erythroid-specific genes are both activated during erythroid differentiation and indeed how their expression is restricted to cells of the erythroid lineage.

CHAPTER TWO

MATERIALS AND METHODS

CHAPTER TWO: MATERIALS AND METHODS

2.1   MATERIALS

2.1.1   Drugs, Chemicals and Reagents

Hemin (Ferriprotoporphyrin IX chloride) was purchased from Porphyrin Products Inc. USA.

The following products were obtained from Sigma Chemical Co.: Acrylamide, adenosine triphosphate (ATP), agarose (Type 1), ampicillin, aprotinin, benzamidine, bestatin, β-glycerophosphate, β-mercaptoethanol, bisacrylamide (N, N'-methylene-bis-acrylamide), bovine serum albumin (BSA), chloramphenicol, deoxyribonucleoside triphosphates (dNTPs), diaminofluorene, dimethyl sulfoxide (Me$_2$SO), dithiothreitol (DTT), ethidium bromide, ethylenediaminetetra-acetic acid (EDTA), glycylglycine, kanamycin, leupeptin, levamisole, N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid (Hepes), o-nitrophenyl-B-D-galactopyranoside (ONPG), pepstatin, phenylmethylsulfonyl fluoride (PMSF), salmon sperm DNA, sodium dodecyl sulphate (SDS), spermidine, Tris-base,

Sources of other reagents were as follows : acetyl coenzyme A: Boehringer Mannheim; Beetle luciferin: Promega; guanidine isothiocyanate: Gibco BRL; phenol: Wako pure chemicals; polyethylene glycol 6000 (PEG$_{6000}$): BDH chemicals; poly(dI-dC): Pharmacia; N,N,N',N'- tetramethethylethenediamine (TEMED): Tokyo Kasei.

Sequenase Version 2.0 sequencing kit was purchased from United States Biochemical Corporation. Oligolabelling of DNA was performed using Amersham Inc. Megaprime DNA labelling kit. All other chemicals and reagents were of analytical grade.

2.1.2   Radiochemicals

D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol (57 mCi/mmol) were purchased from Amersham. [α-^{32}P] dATP (1800 Ci/mmol) and [γ-^{32}P] dATP (>2000 Ci/mmol) and were purchased from Bresatec.

2.1.3 Enzymes

Restriction enzymes were purchased from Pharmacia or New England Biolabs.

Other enzymes were obtained from the following sources: Calf intestinal phosphatase: Boehringer Mannheim; *E. coli* DNA polymerase I (Klenow fragment): Bresatec; lysozyme: Sigma; proteinase K: Boehringer Mannheim; ribonuclease A (RNase A): Sigma (the stock solution (10mg/ml) was incubated at 100°C for 10 minutes to inactivate any DNase activity); T4 DNA ligase and T4 Polynucleotide Kinase: Bresatec; T4 DNA polymerase: Boehringer Mannheim.

2.1.4 Buffers

Denhardt's solution: 0.1%(w/v) Ficoll, 0.1%(w/v) polyvinylpyrrolidine, 0.1%(w/v) BSA.

NET: 100mM NaCl, 1mM EDTA, 10mM Tris-HCl, pH 7.5.

SSC: 150mM NaCl, 15mM sodium citrate.

SSPE: 150mM NaCl, 10mM NaH_2PO_4, 1mM EDTA.

TAE: 40mM Tris-acetate, 20mM sodium acetate, 1mM EDTA, pH 8.2.

TBE: 90mM Tris, 90mM boric acid, 2.5mM EDTA, pH 8.3.

TE: 10mM Tris-HCl, pH 7.5, 0.1mM EDTA.

TES: 25mM Tris-HCl, pH 8.0, 10mM EDTA, 15% sucrose.

3x Urea Loading Buffer: 4M urea, 50% sucrose, 50mM EDTA, 0.1% bromophenol blue

5x Ligation Buffer: 250mM Tris-HCl, pH 7.6, 50mM $MgCl_2$, 5mM DTT, 25%(w/v) PEG-6000

Measuring Buffer: 51.9mM glycylglycine, 31.1mM $MgSO_4.7H_2O$, adjust to pH 7.8 with NaOH.

rATP: 31.1mM rATP, adjust to pH 7.5 with NaOH.

Luciferase Reaction Buffer: Mix 7.5ml of Measuring Buffer with 2.5ml of 31.1mM rATP.

Luciferin: 111μM of Beetle luciferin dissolved in water, pH 7.0.

All buffers were sterilised by autoclaving or filtration through a Sartorius[a] Minisart NML 0.2μm filter where necessary.

2.1.5 Cloning Vectors

pBluescript KS+ was purchased from Stratagene. pSP72 was purchased from Bresatec. pGL2-Basic vector was purchased from Promega. ptk-LUC, pIBICAT and pRSV-CAT were provided by Dr. C. N. Hahn. pRSV-βgal was provided by Dr. H. Healy. pPGKneoβ was provided by Dr. L. Whyatt. pCATOO was kindly provided by Dr. M. Crossley.

2.1.6 Cloned DNA Sequences

The following cloned DNA sequences, used throughout this study, were generous gifts from the following:

pXM/GF-1: murine GATA-1 cDNA expression clone was provided by Dr. S. H. Orkin.

pSG5/EKLF: murine EKLF cDNA expression clone was provided by Dr. J. J. Bieker.

pMT2/BKLF: murine BKLF cDNA expression clone was provided by Dr. M. Crossley.

pMT2/RINFE: murine p45 NF-E2 cDNA expression clone was provided by Dr. N. Andrews.

pMT2/p18w-1: murine p18 cDNA expression clone was provided by Dr. N. Andrews.

hALAS: 2.1kb human cDNA of ALAS1 (Healy, 1990) was provided by Dr. T. Sadlon.

eALAS: 1.9kb human ALAS2 cDNA (Cox et al., 1991) was provided by Dr. T. Cox.

PBGD: 1.2kb rat cDNA of PBGD (Grandchamp et al., 1984) was provided by Dr. B. Grandchamp.

α-globin: 500bp murine cDNA of α-globin (Rougeon & Mach, 1977) was provided by Dr. P. Klinken.

β-globin: 1kb murine cDNA of β-globin (Hofer et al., 1982) was provided by Dr. P. Klinken.

pRGAPDH: a cDNA of rat glyceraldehyde-3-phosphate deydrogenase (Piechaczyk et al., 1984) was provided by Dr. P. Klinken.

2.1.7 Synthetic Oligonucleotides

Synthetic DNA primers were obtained from Bresatec, IDT or Gibco. The primer sequences are listed below:

i) ***Oligonucleotides for sequencing plasmid constructs***

M13 Universal sequencing primer (17mer): 5'-dGTAAAACGACGGCCAGT-3'

M13 Reverse sequencing primer (25mer): 5'-dCACACAGGAAACAGCTATGACCATG-3'

primer #5270 for pGL2-Basic (16mer): 5'-dGTATCTTATGGTACTG-3'

ii) ***Oligonucleotides for the generation of ALAS2/reporter gene constructs***

The introduced restriction enzyme sites and the mutations are underlined.

Human ALAS2 promoter:

Primer 1: 5'-dCCC_AAGCTT_GCACTGAGGACGAACG-3';
_Hind_III site introduced at +12/+28

Primer 2: 5'-dGGGTTCTGTAACTACATTGCC-3'; at -718/-699

RW-142: 5'-dGAACGAATGACAGGTGG_AAGCTT_GGGTTAGGACCTGG-3';
_Hind_III site introduced at -7/-2

RW-143: 5'-dCATGGCTGCAGTCAC_AAGCTT_GAACCTAAAGTCCTG-3';
_Hind_III site introduced at the ATG in exon 2 of the human ALAS2 gene

Primer 3: 5'-dGGTTT_AGATCT_TAGCAAGGAAGGGA-3';
_Bgl_II site introduced at -125/-120

Primer 4:   5'-dCCC<u>AAGCTT</u>GCACTGAGGACGAACGAATGACA<u>AAC</u>GGGTACTTGGGT-3';
               mutation introduced at +2/+4 and a *Hin*dIII site at +28

Human ALAS2 intron 8:

I8 primer 1:   5'-dGAA<u>GAATTC</u>GTAAGTGAATGCTTTGGGCCT-3'; *Eco*RI site underlined

I8 primer 2:   5'-dAGG<u>GAATTC</u>CTGGAGACAGAAAGGAATAAG-3'; *Eco*RI site underlined

I8 primer 3:   5'-dCACCCT<u>CTCGAG</u>AAGCTTCATCTTAGCTCC-3'; *Xho*I site underlined

I8 primer 4:   5'-dGTCGTAGC<u>GTCGAC</u>TTCTGCTGCTTTGAGATA-3'; *Sal*I site underlined

I8 primer 5:   5'-dAAAGTC<u>CTCGAG</u>CAAAGCAGCAGAATTATC-3'; *Xho*I site underlined

I8 primer 6:   5'-dGCCAAAGG<u>GTCGAC</u>CTGGAGACAGAAAGGAAT-3'; *Sal*I site underlined

iii)   ***Oligonucleotides used for Site-directed Mutagenesis***

         The introduced mutations are underlined.

Human ALAS2 promoter:

-124 GATA: 5'-dACTTTGGGTTT<u>CAGCTG</u>TTAGCAAGGAA-3'

-100 GATA: 5'-dGGAAGGGACTGG<u>CAGCTG</u>TTTGGGGCCA-3'

-27 GATA: 5'-dAGAGGAGAAGG<u>CAGCTG</u>TGCCAGGTCCT-3'

TATA: 5'-dAGAGGAGAAGG<u>TATAAAT</u>GCCAGG-3'

66

GATA-27G: 5'-dTCAGAGGA<u>GACATGATAAG</u>TGCCAGGTCCT-3'

-54 CACCC: 5'-dCAGAAGGCAGG<u>CAGCTG</u>GGGGGCTGAGTC-3'

-44 NF-E2:   5'-dGTGGGTGGGG<u>CAGCTG</u>CAGAGGAGAAG-3'

-110 Ets: 5'-dATCTCTAGCAA<u>CCAA</u>GGGACTGAG-3'

-13 TRE: 5'-dGGGATAAATGC<u>AAGCTT</u>CTAACCCAAG-3'

Human ALAS2 intron 8:

I8 CACCC site A: 5'-dTAAACCCCTCCT<u>CAGCTG</u>TAGCCCCAAGCTT-3'

I8 CACCC site B: 5'-dCAGCTAAAGGTT<u>CAGCTG</u>AGCTACTGCCT-3'

I8 GATA site A: 5'-dCCAGCTACTGC<u>CAGCTG</u>AGTCATTGCAT-3'

I8 GATA site B: 5'-dACTTGAAAGTC<u>CAGCTG</u>CAAAGCAGCAG-3'

iv)   *Sequences of the Sense Strand of Synthetic Oligonucleotides used in Gel Shift Assays*

The binding motifs are underlined.

Consensus DNA Binding Sites:

β-globin GATA-cons (Wall *et al.*, 1988):   5'-dTTGGCTCCC<u>TTATCA</u>TGTCCCTG-3'

β-globin CACCC (Miller & Bieker, 1993):   5'-dAGCTAG<u>CCACACCCT</u>GAAGCT-3'

67

Sp1 consensus (Promega):   5'-dATTCGATCGGGGCGGGGCGAGC-3'

AP-1(Promega):   5'-dCGCTTGATGAGTCAGCCGGAA-3'

TRE consensus (Glass *et al.*, 1988):

    5'-dGATCTAGATTCAGGTCATGACCTGAGGAGAA-3'

ETS CYP$_{24}$ (ets site in the rat CYP$_{24}$ promoter, Dr. Prem Dwivedi, personal communication):   5'-dTCGACGCTGACTCCATCCTCTTCC-3'

non-specific (NS) competitor (NRF-1 site in the rat somatic cytochrome *c* promoter, Evans & Scarpulla, 1990):   5'-dCAGAAGGCAGGGTGGGTGGGGCTGAGTC-3'

Human ALAS2 promoter:

GATA-124:   5'-dCTTTGGGTTTTATCTCTAGCAAGG-3'

GATA-100:   5'-dAAGGGACTGAGATACCTTTGGGGC-3'

GATA-27:   5'-dGAGGAGAAGGGATAAATGCCAGGT-3'

TATA: 5'-dGAGGAGAAGGTATAAATGCCAGGT-3'

GATA-27G: 5'-dTCAGAGGAGACATGATAAGTGCCAGGTCCT-3'

CACCC-54:   5'-dCAGAAGGCAGGGTGGGTGGGGCTGAGTC-3'

NF-E2 -44:   5'-dTGGGTGGGGCTGAGTCAGAGGAGAA-3'

ETS-110:   5'-dTTGGGTTTTATCTCTAGCAAGGAAGGGACTGAGATACCTTTGGG-3'

TRE-13:   5'-dGATAAATGCCAGGTCCTAACCCAAGTACCCAC-3'

CAP-S :   5'-dCCCAAGTACCCACCTGTCATTC-3'

Human ALAS2 intron 8:

I8 CACCC site A-S: 5'-dCTAGTCCCCCACCCTAGCGAA-3'

I8 CACCC site B-S: 5'-dAAAGGTCCCCACCCAGCTACT-3'

I8 GATA site A-S: 5'-dAGCTACTGCCTATCTAGTCATTGC-3'

I8 GATA site B-S: 5'-dTTGAAAGTCCTATCTCAAAGCAGC-3'

2.1.8   Bacterial Strains

The following *E.coli* K12 strains were used as hosts for recombinant plasmids in recombinant DNA procedures:

E.coli DH5α: supE44 ΔlacU169 (p80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1 host for recombinant plasmids, obtained from the *E. coli* Genetic Stock Centre, Yale University, New Haven.

E. coli CJ236 : dut, ung, thi, rel A; pCJ105 (Cm^r), obtained from Dr. J. C. Wallace.

Long term storage of stock cultures of these and plasmid transformed bacteria were prepared by dilution of an overnight culture with an equal volume of 80% glycerol and then stored at -80°C. Single colonies of bacteria, obtained by streaking the glycerol stock onto agar plates (Section 2.1.9) were used to inoculate liquid growth medium, and the bacterial cultures were grown at 37°C with continuous shaking to provide adequate aeration. The agar plates of the liquid growth medium was supplemented with the appropriate antibody where required.

2.1.9 Bacterial Growth Media

Growth media were prepared in double-distilled water and sterilised by autoclaving. Antibiotics and other labile chemicals were added after the solution had cooled to $50^{\circ}C$.

Luria (L) broth: contained 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH.

Agar plates were prepared by adding 1.5% (w/v) Bacto-agar (Difco) to the L broth.

2YT Medium: 1.6% (w/v) Bacto-tryptone (Difco), 1% (w/v) yeast extract (Difco), 0.5% NaCl, adjusted to pH 7.5 with NaOH.

Psi (ψ) Broth : 2% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 0.5% $MgSO_4$, adjusted to pH 7.6 with 0.1M KOH.

FTB: 1.2% (w/v) Bacto-tryptone (Difco), 2.4% (w/v) yeast extract (Difco), 0.4% glycerol, autoclaved, with 17mM KH_2PO_4 and 72mM K_2HPO_4.

SOC: 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose.

YENB Medium: 0.75% (w/v) Bacto yeast extract (Difco) and 0.8% (w/v) Bacto-Nutrient Broth (Difco).

Ampicillin (100μg/ml), kanamycin (70μg/ml) or chloramphenicol (33μg/ml) were added where appropriate to maintain selective pressure.

2.1.10 Tissue Culture Cell Lines and Media

i) *Cell lines*

The sources of the following cell lines used throughout the course of this work are as indicated: Human erythroleukemia K562 cells (Prof. N. G. Abraham, USA); murine erythroleukemia cells DS-19 (Dr. S. Sassa, Rockefeller University, New York, USA), MELC 707T (Dr. G. Yeoh, Nedlands, Western Australia) and MEL F4-12B2 (Dr. G. Bergholz, Heinrich Pette Institute, Hamburg, Germany); murine erythropoietin-responsive J2E cells (Dr. S. P. Klinken, W. Australia); COS-1 (American Type Cell Culture (ATCC) Laboratory) and CV-1 cell lines (Dr. O. Bernard, WEHI, Melbourne, Australia).

ii) *Solutions*

Phosphate buffered saline (PBS): 136mM NaCl, 2.6mM KCl, 1.5mM KH_2PO_4 and 8mM Na_2HPO_4, pH 7.4, was sterilised by autoclaving (20 psi for 25 minutes at 140°C).

Trypsin/EDTA solution: 0.1% trypsin (Difco) and 1xEDTA Versene buffer solution (CSL), was sterilised by filtration through a 0.2µm filter (Whatman).

iii) *Media*

Dulbecco's Modified Eagle's Medium, (DMEM) (Gibco), 28mmol/l $NaHCO_3$, 19mmol/l glucose, and 20mmol/l Hepes, pH 7.3.

RPMI 1640 with 28mmol/l $NaHCO_3$.

Ham's F12 with L-glutamine (Gibco) and 28mmol/l $NaHCO_3$, pH 7.4.

All media was supplemented with 50,000 Units/l of Gentamycin (Gibco) and filter sterilised prior to use.

Fetal Calf Serum : CSL

2.1.11 Miscellaneous

X-Omat AR diagnostic and Biomax MR film: Kodak, USA

Nytran 0.45μm: Schleicher and Schuell

3MM paper : Whatman Ltd.

2.2 RECOMBINANT DNA METHODS

2.2.1 General DNA Methods

The following methods were performed essentially as described in *"Molecular Cloning: A Laboratory Manual"* Sambrook *et al.* (1989) : Growth, maintenance and preservation of bacteria; quantitation of DNA and RNA; autoradiography; agarose and polyacrylamide gel electrophoresis; precipitation of DNA and RNA; phenol/chloroform extractions; end-filling or end-labelling of DNA fragments using the Klenow fragment of *E.coli* DNA polymerase I.

All manipulations involving viable organisms which contained recombinant DNA were carried out in accordance with the regulations and approval of the Australian Academy of Science Committee on Recombinant DNA and the University Council of the University of Adelaide.

2.2.2 Plasmid DNA Preparation

The rapid alkaline hydrolysis procedure of Birnboim and Doly (1979) was used for the isolation of plasmid DNA from 2ml overnight cultures for analytical restriction digests and for the preparation of plasmid cDNAs for use as probes (Section 2.2.10) in Northern and Southern hybridisation analysis.

Plasmid DNA used for the transfection of tissue culture cell lines was routinely grown up in 250ml cultures of FTB inoculated with 100μl from a 2ml overnight culture. Plasmid DNA was extracted using the alkaline lysis procedure described above and further purified by the cesium-chloride/ethidium bromide density gradient procedure (as described in Sambrook *et al.*, 1989) in a Beckman TL-100 benchtop ultracentrifuge and TLA-100.2 rotor. Plasmid DNA was then butanol extracted to remove ethidium bromide, dialysed in TE buffer at 4°C with three buffer changes and precipitated with 0.1 volume of 3M sodium

acetate, pH 4.6 and 2.5 volumes of 100% ethanol. Plasmid DNA was then washed with 70% ethanol, resuspended in TE buffer, quantified by spectrophotometry and analysed by agarose gel electrophoresis to confirm concentration and supercoiling.

2.2.3 Restriction Enzyme Digestions of DNA

In analytical digests, 0.5-1μg of DNA was incubated with 2-5 units each of the appropriate restriction enzyme(s) for a minimum of 2 hours in the buffer conditions specified by the manufacturer. Reactions were terminated with the addition of a 1/3 volume of urea load buffer and electrophoresed on 1% or 3% mini-agarose gels depending on the size of the restricted DNA fragment(s) in TBE buffer.

In preparative digests, 5-10μg of DNA was restricted in a reaction volume of 30μl, and the desired DNA fragments were isolated as detailed below.

2.2.4 Preparation of Cloning Vectors

Plasmids were linearised with the appropriate restriction enzyme(s). To prevent self-ligation of the vector, 5' terminal phosphate groups were removed by incubation in 50mM Tris-HCl, pH 9.0, 1mM $MgCl_2$, 0.1mM $ZnCl_2$, with 1 unit of calf intestinal phosphatase (CIP), in a final volume of 50μl for 15 minutes at 37°C followed by 15 minutes at 56°C. Another 1 unit of CIP was added and the incubations repeated. The vector DNA was electrophoresed on a 1% agarose TBE gel and the gel stained with ethidium bromide. The linearised vector DNA was visualised under UV light, excised and purified using the Qiaex II gel extraction kit according to the manufacturer's instructions or purified from a low melting point agarose gel by phenol/chloroform extraction, followed by DNA precipitation.

2.2.5 Preparation of DNA Restriction Fragments

DNA was incubated with the appropriate restriction enzyme(s) as described in Section 2.2.3, and restriction fragments were isolated from either a horizontal 0.8%-2.0% agarose TBE gel or a vertical 8% polyacrylamide gel, depending on the size of the restriction fragment(s). Bands representing restriction fragments were visualised under UV

light following staining with ethidium bromide, and the appropriate fragment(s) excised from the gel. DNA fragments were isolated after electrophoresis from agarose gels using either the Qiaex II gel extraction kit according to the manufacturer's instructions or purified from a low melting point agarose gel by phenol/chloroform extraction, followed by DNA precipitation. Fragments isolated from polyacrylamide gels were eluted from the gel slice by incubation in 400μl of TE buffer at 37°C for 16 hours. The DNA was precipitated by the addition of 0.1 volume of 3M sodium acetate, pH 4.6 and 2.5 volumes of 100% ethanol, washed in 70% ethanol, air dried and resuspended in 10-20μl of TE buffer. The DNA was resuspended at a concentration of 20-50ng/μl, for use in ligation reactions.

2.2.6 Ligation of DNA

A 10μl reaction contained 20-50ng of vector DNA, a 3 molar excess of the insert DNA, 50mM Tris-HCl, pH 7.6, 10mM $MgCl_2$, 1mM DTT, 5% (w/v) PEG_{6000}, 1mM ATP, and 1 unit of T4 DNA ligase. The reactions were incubated for either 4 hours at room temperature or overnight at 4°C. A control ligation with vector only was set up and included in the subsequent transformation to determine background levels of uncut or recircularised vector DNA.

2.2.7 Transformation of *E. coli* with Recombinant Plasmids

i) *Preparation of Competent E. coli*

E. coli cells were made competent by either the rubidium chloride method or by growing the cells in salt-free medium for bacterial electroporation.

For the rubidium chloride method, a single colony of the *E. coli* host strain was inoculated into 5ml of Psi (ψ) broth (where appropriate the ψ-broth was supplemented with an antibiotic) and the culture incubated with continuous shaking at 37°C overnight. 1.6ml of the overnight culture was subcultured into 50ml of ψ–broth and incubated at 37°C until the culture reached an absorbance at 600nm (OD_{600}) of 0.4-0.6. 25ml of this culture was then used to inoculate 500ml of ψ-broth and again the culture was grown up to an OD_{600} of 0.4-0.6. The cells were chilled on ice for 5 minutes prior to centrifugation at 4000 rpm for

5 minutes at 4°C. The cells were then resuspended in 0.4 volumes of ice cold Tfb1 (30mM KAc, 100mM RbCl, 10mM CaCl₂, 50mM MnCl, 15% glycerol, adjusted to pH 5.8, with 0.2M acetic acid), incubated on ice for 5 minutes and pelleted at 4000 rpm, at 4°C for 5 minutes. Cells were then resuspended in 0.04 volumes of ice cold Tfb2 (10mM MOPS free acid, 10mM RbCl, 75mM CaCl₂, 15% glycerol adjusted to pH 6.5, with 0.1M KOH), and kept on ice for a further 15 minutes, before being stored at -80°C in 500μl aliquots.

For the preparation of electrocompetent *E. coli*, cells were prepared as described by Sharma and Schimke (1996). Briefly, a single colony was inoculated into 2ml of YENB medium and the culture incubated with continuous shaking at 37°C overnight. 1L of YENB medium was inoculated with 5ml of the overnight culture and grown with shaking at 37°C until the culture reached an OD_{600} of 0.4-0.6. The culture was chilled on ice for 5 minutes and cells were harvested by centrifugation at 4000 rpm at 4°C for 10 minutes. The cells were then washed twice with 100ml of ice cold water and resuspended in 20ml of cold 10% glycerol. The cells were pelleted by centrifugation, then resuspended in 3ml of ice cold 10% glycerol and stored at -80°C in 40μl aliquots.

ii)   ***Transformation of Competent Bacteria***

100-150μl of the cell suspension prepared using the rubidium chloride method was mixed with 5μl of the DNA ligation reaction mix (Section 2.2.6) and left on ice for 30 minutes. The cells were then heat shocked at 42°C for 2 minutes, placed on ice for 2 minutes and 500μl of SOC medium added. The cells were incubated at 37°C for 20 minutes, centrifuged briefly at 6500 rpm for 30 secs and 400μl of supernatant removed.

1-5μl of DNA in TE was added to 40μl of the electrocompetent cells, gently mixed, incubated on ice for 2 minutes and transferred to a cold 0.2-cm electroporation cuvette (Bio-Rad) and exposed to a single 2000-volt pulse at 25 microfarads using the Bio-Rad Gene Pulser. Following electroporation, cells were placed on ice for 2 minutes, gently resuspended in 1ml of SOC medium, transferred to an eppendorf tube and incubated at 37°C for 20 minutes. The cells were then centrifuged briefly at 6500 rpm for 30 secs and 900μl of supernatant removed.

All transformed cells were gently resuspended in the remaining medium and then plated onto L-agar containing 100µg/ml of ampicillin by spreading with a wire spreader. The agar plates were routinely incubated at 37°C overnight.

2.2.8 Dideoxy-Chain Sequencing Analysis

Sequencing was performed by the Sanger *et al.* (1977) dideoxy method, using the sequencing reagents supplied in the USB Sequenase Version 2.0 kit.

Double-stranded sequencing was performed using plasmid DNA purified by the rapid alkaline hydrolysis procedure. The sequencing reaction contained 5-10µg of plasmid DNA or approximately one quarter of the DNA isolated from a 2ml overnight culture was used per reaction. The DNA was denatured in 0.2M NaOH, 0.2mM EDTA for 15 minutes at 37°C. The mixture was then neutralised by passage through a TE equilibrated spin column (Murphy & Kavanagh, 1988). Sequencing reactions were performed with 7µl of the eluant (~2µg), 15ng of primer and [^{35}S]-dATP in accordance with the protocol accompanying the Sequenase Version 2.0 kit.

2.2.9 Gel Electrophoresis of DNA for Sequence Analysis

Sequencing reactions (1-1.5µl) were electrophoresed on 6% polyacrylamide gels containing 7M urea in 1xTBE buffer at 1500-1700 V. The temperature of the gels were monitored and maintained at a temperature ranging between 55-60°C. Following electrophoresis, the gels were transferred to Whatman 3MM chromatography paper, vacuum dried for 60 minutes and the dried gels subjected to autoradiography for 16-24 hours at room temperature.

2.2.10 Preparation of Radiolabelled DNA Probes

i) *Oligo-Labelling of DNA*

Linearised plasmids or isolated cDNA inserts were labelled with [α-^{32}P]-dATP using the Megaprime DNA Labelling kit (Amersham). Briefly 50-100ng of DNA was denatured in the presence of random nonamers in a total volume of 25µl at 100°C for 4 minutes.

Reactions were set up at room temperature as per the instruction manual, using 50μCi of [α-^{32}P] dATP and 2 units of Klenow fragment in a total volume of 50μl. The reactions were then incubated at 37°C for 10 minutes before being terminated by the addition of 2μl of 0.5M EDTA, pH 8.0 (final concentration 20mM). The probes were then denatured by incubation at 100°C for 4 minutes and quenched on ice before being added to the hybridisation solution.

ii)   5' End-Labelling of Synthetic DNA oligonucleotides

The synthetic DNA oligonucleotides used as probes were [^{32}P] labelled at the 5' end using [γ-^{32}P]-ATP and T4 polynucleotide kinase. The reaction mixture contained 100ng of oligonucleotide, 10mM MgCl$_2$, 50mM Tris-HCl, pH 7.4, 5mM DTT, 0.1mM spermidine, 0.1mM EDTA, 100μCi [γ-^{32}P] ATP and 5 units of T4 polynucleotide kinase in a final volume of 10μl. This was incubated at 37°C for 45 minutes and the reaction terminated following the addition of 40μl of water. The radiolabelled oligonucleotide was then precipitated with 2μg of poly(dI-dC), 0.1 volume of 3M sodium acetate, pH 4.6 and 2.5 volumes of cold 100% ethanol at -20°C overnight or placed on dry ice for 5 minutes. The probe was the pelleted at room temperature by centrifugation at 13000 rpm for 20 minutes, washed once with 200μl of 70% ethanol, dried and resuspended in water.

2.2.11 Oligonucleotide Site-Directed Mutagenesis

i)   Preparation of Single-Stranded DNA

Site-directed mutagenesis was performed on single-stranded DNA, using the Bio-Rad Muta-Gene M13 *in vitro* Mutagenesis Kit based on the method of Kunkel *et al.* (1987). Plasmids to be mutagenised were transformed into the *E. coli* strain CJ236. A single colony was inoculated into 2ml of 2YT medium supplemented with 33μg/ml chloramphenicol, 100μg/ml ampicillin and 0.5μg/ml uridine and incubated with shaking overnight at 37°C. 4ml of 2YT medium supplemented with 33μg/ml chloramphenicol, 100μ g/ml ampicillin and 0.5μg/ml uridine was inoculated with 200μl of the overnight culture and

incubated with shaking at 37°C until an OD_{600}~0.3 (~3-4 hours) was obtained. To 1ml of this culture, 20µl of M13K07 helper phage (10^{10} pfu/ml) was added and incubated for 1 hour at 37°C with vigorous shaking. 400µl of the infected culture was inoculated into 10ml of 2YT medium supplemented with 33µg/ml chloramphenicol, 100µg/ml ampicillin, 0.5µg/ml uridine and 70µg/ml of kanamycin and incubated with shaking overnight at 37°C.

Bacterial cells were centrifuged at 10,000 rpm at 4°C for 15 minutes. The supernatant was transferred to a clean tube and this centrifugation step was repeated twice. To the phagemid-containing supernatant, 0.25 volume of 2.5M NaCl/20% PEG-6000 (final concentration of 625mM NaCl and 5% (w/v) PEG) was added, gently mixed and incubated on ice for 30 minutes. The phagemids were collected following centrifugation at 15,000 rpm at 4°C for 15 minutes and resuspended in 300µl of TE. The phagemids were extracted twice with an equal volume of TE-buffered phenol and then once with an equal volume of $CHCl_3$:Isoamyl alcohol (49:1). The single-stranded DNA was then precipitated with 0.1 volume of 3M sodium acetate, pH 4.6 and 2.5 volumes cold 100% ethanol, washed with 90% ethanol and resuspended in 30µl of TE.

ii)   *Synthesis of the Mutagenic DNA strand*

Mutagenic oligonucleotides were phosphorylated by incubating 200pmol of oligo in 100mM Tris-HCl, pH 8.0, 10mM $MgCl_2$, 5mM DTT, 433nM dATP and 4.5 units of T4 polynucleotide kinase in a final volume of 30µl at 37°C for 45 minutes. The reaction was terminated by heating at 65°C for 10 minutes and the phosphorylated oligos diluted to 6pmol/µl.

Approximately 200ng of single-stranded DNA was mixed with 6pmol of the phosphorylated mutagenic oligo in Annealing buffer (20mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 50mM NaCl) and 1mM DTT in a final volume of 10µl. The mixture was heated at 68°C for 3 minutes and allowed to slowly cool to room temperature.

To synthesise the complementary DNA strand, the reaction contained the annealed template, 2.4µl of 10xSynthesis buffer (100 mM Tris-HCl, pH 7.5, 50mM $MgCl_2$, 20mM DTT), 2.4µl of 10mM rATP, 2.4µl of 5mM of each dNTP, 1µg of T4 gene 32 protein (Bresatec), 3 units of T4 DNA polymerase and 5 units of T4 DNA ligase in a final volume

of 24μl. The reaction was placed initially on ice for 5 minutes, incubated at 25°C for 5 minutes and then incubated at 37°C for 2 hours. The reaction was terminated by heating at 65°C for 5 minutes. 5μl of the undiluted mutagenesis reaction was used to transform *E. coli* strain DH5α cells as described in Section 2.2.7, usually yielding 100-200 colonies per plate.

Screening of mutants was performed either by restriction enzyme digestion or by colony screening followed by washing with tetramethylammonium chloride (TMACl).

2.2.12 Colony Screening

Following transformation of recombinant DNA plasmids or the site-directed mutagenesis reactions into *E. coli* DH5α, the colonies were transferred to nylon membranes according to the published "Microwave Lysis" protocol (Buluwela *et al.*, 1989).

Filters were pre-hybridised for 4 hours at 42°C in 50% formamide, 5xSSPE, 5xDenhardt's solution, 0.1% SDS, 0.05% sodium pyrophosphate and 200μg/ml of sonicated salmon sperm DNA in roller bottles. Hybridisations were carried out for 12-18 hours under the same conditions, with the addition of radiolabelled probe. Filters were washed in 2xSSC, 0.1% SDS at room temperature for 10 minutes and then at 65°C for 10-20 minutes. The filters were then air dried and autoradiographed overnight.

In order to discriminate between colonies containing plasmids with only a single or multiple nucleotide differences, the filters were washed twice following hybridisation in 200ml 6xSSC for 5 minutes at room temperature and then rinsed briefly in 30ml of tetramethylammonium chloride (TMACl) solution (3M TMACl, 50mM Tris-HCl, pH 8.0, 2mM EDTA, 0.1%SDS). The filters were then washed in 30ml of TMACl solution at approximately 10°C below the theoretical melting temperature of the mutagenic primer (Wood *et al.*, 1985) for 15-30 minutes and radioactivity monitored during this period. The filters were air dried and autoradiographed overnight.

The recombinant and mutant colonies were identified by aligning the developed autoradiogram with the regrown colonies on the original plate. Plasmid DNA from putative mutagenised colonies was prepared by the rapid alkaline lysis method and used to transform *E. coli* DH5α and this was repeated once. All mutants were sequenced for both the desired mutation as well as for any random mutations introduced during the synthesis reaction.

2.3   POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION OF DNA

The polymerase chain reaction (PCR) was employed to amplify various genomic fragments from the human ALAS2 gene using the human genomic clone, pTC-EA1 (Cox *et al.*, 1991) or for the amplification of ALAS2 fragments from recombinant plasmid DNA. All amplified fragments were cloned and sequenced.

The polymerase chain reactions contained 10ng of cosmid template DNA (pTC-EA1) or 1ng of plasmid DNA, Taq polymerase buffer (50mM KCl, 10mM Tris-HCl, 1.5mM $MgCl_2$, 0.1% gelatin) supplemented with an additional 2mM $MgCl_2$, 25 pmol (~100ng) each of sense and anti-sense strand oligonucleotides, dNTPs to a final concentration of 160μM and 1.25 units of Taq polymerase in a final volume of 50μl. Thirty cycles of amplification were performed using an automated PTC-100 Programmable thermal cycler (MJ Research, Inc.) under the following conditions: Cycle 1: denaturation for 4 minutes 30 seconds at 94°C, annealing for 45 seconds at 62°C, extension for 45 seconds at 72°C; Cycle 2-30: denaturation for 30 seconds at 94°C, annealing for 45 seconds at 62°C, extension for 45 seconds at 72°C.

To eliminate the possible introduction of random errors in the ALAS2 sequence during the amplification of the fragment by Taq, the enzyme Pfu DNA polymerase (Stratagene) was used in later amplifications. The PCR reactions in a final volume of 50μl, contained 10ng of cosmid template DNA or 1ng of plasmid DNA, Pfu reaction buffer (20mM Tris-HCl, pH 8.75, 10mM KCl, 10mM $(NH_4)_2SO_4$, 2mM $MgCl_2$, 0.1% Triton X-100, 0.1mg/ml BSA), 200μM dNTPs, 25 pmol each of the sense and anti-sense oligonucleotides, 1% glycerol and 2.5 units of Pfu DNA polymerase. The amplification was performed under the following conditions: Cycle 1: denaturation for 5 minutes at 95°C, annealing for 2 minutes at 50°C, extension for 2 minutes at 72°C; Cycle 2-30: denaturation for 45 seconds at 95°C, annealing for 50 seconds at 50°C, extension for 1 minute 45 seconds at 72°C; Cycle 31: final extension for 5 minutes at 72°C.

2.4   SYNTHESIS OF PLASMID CONSTRUCTS

The plasmid, pRSV-LUC, containing the Rous sarcoma viral promoter driving expression of the luciferase reporter gene was included as a positive control in transfection experiments. To synthesise this plasmid, a *Nru*I-*Hin*dIII fragment was isolated from pRSV-CAT and ligated into the *Sma*I/*Hin*dIII linearised promoterless firefly luciferase reporter gene vector, pGL2-Basic (Promega).

2.4.1   Construction of ALAS2 Promoter/Reporter Gene Plasmids

A series of 5'-flanking ALAS2 deletion constructs were generated from subcloned fragments isolated from the human genomic clone, pTC-EA1 (Cox *et al.*, 1991) and ligated into pGL2-Basic.

The polymerase chain reaction was performed using pTC-EA1 as the template and the following primers: primer 1, bound at +12/+36 (relative to the initiation site) and introduced a *Hin*dIII site at +28 and primer 2, which bound upstream of an *Avr*II site at -718/-699 and resulted in the amplification of a 730 bp promoter fragment (Fig. 2.1A). The amplified product was digested with *Bgl*II and *Hin*dIII, and a 321 bp fragment ligated into the similarly digested pGL2-Basic vector. The resulting construct contains ALAS2 promoter sequence from -293 bp to +28 bp and is designated pALAS-293-LUC. The amplified product was also digested with *Sac*I and *Hin*dIII and ligated into the similarly digested pBluescript KS+ phagemid to generate pKS-ALAS(420) (Fig. 2.1A). This plasmid, containing ALAS2 promoter sequence from -392 bp to +28 bp was used as a template for site-directed mutagenesis (as described in Section 2.2.11). Following mutagenesis, the *Bgl*II-*Hin*dIII fragment harbouring the mutation was excised from pBluescript KS+ and subcloned into the *Bgl*II/*Hin*dIII digested pGL2-Basic vector.

To synthesise plasmids with promoter lengths of -124 and -27 (Fig. 2.1B), a *Pvu*II site was introduced at these positions in the plasmid pKS-ALAS(420) by site-directed mutagenesis as described in Section 2.2.11. The resulting modified plasmids were digested with *Sma*I (polylinker) and *Pvu*II and religated to form pALAS-124-LUC and pALAS-27-LUC.

The synthesis of the longer promoter constructs was performed in several steps. In separate studies, a *Hind*III site was introduced at -7/-2 in the ALAS2 promoter by site-directed mutagenesis in a subclone containing -6.0 kb to +5.0 kb of contiguous human ALAS2 sequence. An *Xba*I-*Hind*III fragment containing 1.9 kb of ALAS2 5'-flanking sequence was isolated from this subclone, and ligated into pGL2-Basic linearised with *Nhe*I/*Hind*III (Fig. 2.1C). A 5.7 kb *Kpn*I-*Hind*III fragment was also isolated from this subclone, containing 5.7 kb of ALAS2 5'-flanking sequence and was ligated into the similarly digested pGL2-Basic vector (Fig 2.1D). These initial constructs terminated at position -4 and therefore did not contain the native transcription initiation site. To permit a strict comparison with the shorter ALAS2 promoter constructs, the sequence from around the native transcription initiation site was then re-introduced into these constructs as follows: an *Avr*II-*Hind*III fragment (-700 bp to -4 bp) was excised from the -1.9 kb promoter construct and replaced with an *Avr*II-*Hind*III fragment (-700 bp to +28 bp) that was amplified by the polymerase chain reaction, resulting in pALAS-1.9 kb-LUC (Fig. 2.1C). An *Nco*I-*Hind*III fragment (-1.0 kb to -4 bp) was removed from the -5.7 kb promoter construct and replaced with the *Nco*I-*Hind*III fragment (-1.0 kb to +28 bp) isolated from pALAS-1.9 kb to generate pALAS-5.7kb-LUC (Fig. 2.1D).

To synthesise the construct containing 10.3 kb of 5'-flanking region (pALAS-10.3kb-LUC), pTC-EA1 was digested with *Cla*I and *Xho*I and a 5.7 kb fragment cloned into the similarly digested vector pSP72 (Bresatec) (Fig. 2.1E). An *Eco*RV-*Xho*I fragment isolated from this plasmid was used to replace a 1.1 kb *Sma*I-*Xho*I fragment in the construct pALAS-5.7kb-LUC (Fig. 2.1E).

To synthesise an ALAS2 promoter/CAT reporter construct containing approximately 1.9 kb of ALAS2 promoter, an *Xba*I-*Hind*III fragment (-1.9 kb to +28 bp) was isolated from pALAS-5.7kb-LUC, blunted with Klenow enzyme, and ligated into the *Sma*I linearised pIBICAT vector containing the bacterial chloramphenicol acetyltransferase reporter gene (Fig. 2.2). The resulting construct was designated pALAS-1.9kb-CAT.

In summary, six chimeric human ALAS2 promoter/luciferase constructs were synthesised which contained increasing lengths of the 5'-flanking sequence ranging from -27 bp to -10.3 kb from the transcription initiation site. An ALAS2 promoter/CAT construct containing 1.9 kb of human ALAS2 promoter was also synthesised. In addition to

Figure 2.1   Cloning   strategies   used   to   generate   chimeric   human   ALAS2
promoter-luciferase reporter gene plasmids

Details of the cloning procedures are outlined in the text.

A

B

pALAS-293-LUC

*Sma*I *Kpn*I *Sac*I *Mlu*I *Nhe*I *Xho*I

*Bgl*II

*Hind*III

LUC

-293 bp

+1

+28 bp
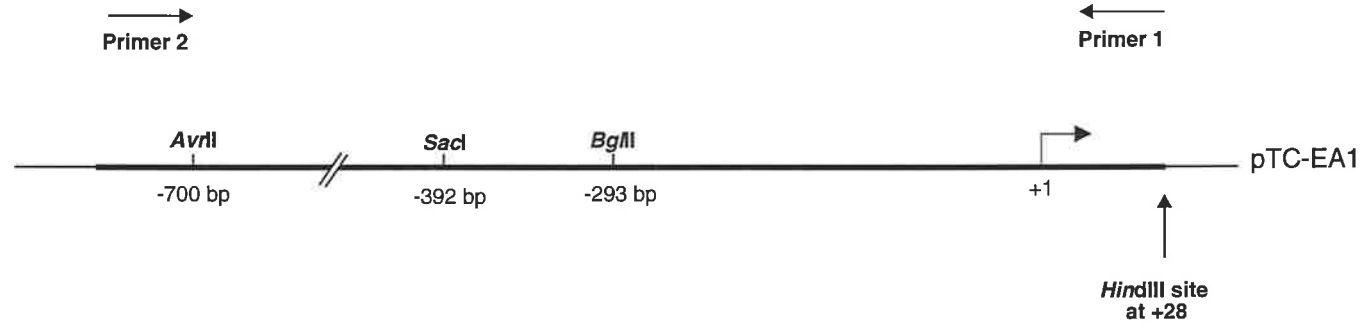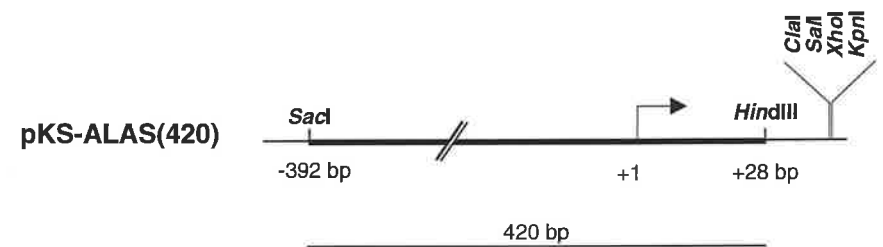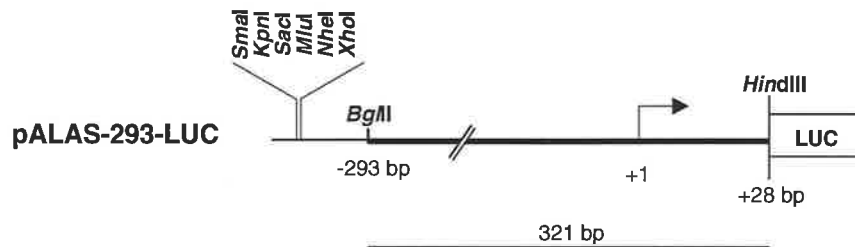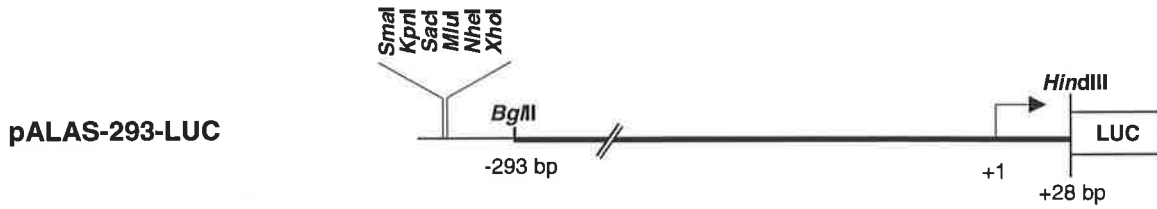
Pvu II site introduced at -124 and -27

by Site-directed Mutagenesis

Digest with *Sma*I and *Pvu*II

and religate

pALAS-124-LUC

*Sma*I/ *Pvu*II

*Hind*III

LUC

-124 bp

+1

+28 bp

pALAS-27-LUC

*Sma*I/ *Pvu*II

*Hind*III

LUC

-27 bp

+1

+28 bp

C

KpnI   XhoI   XbaI   NcoI   AvrII   BglII
-6.0 kb -5.7 kb -4.6 kb   -1.9 kb   -1.0 kb   -700 bp   -293 bp   +1   +5.0 kb

HindIII site
introduced at -7/-2
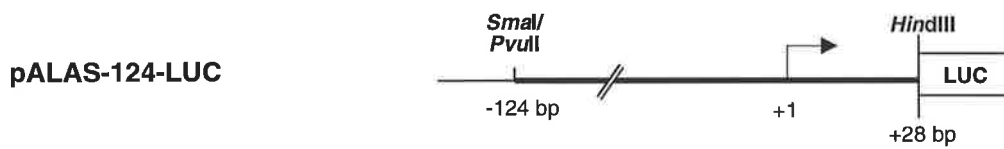
Ligate an XbaI-HindIII fragment into

NheI/XbaI digested pGL2-Basic

XbaI   AvrII   HindIII
                                                  LUC
-1.9 kb   -700 bp
                                          -4 bp
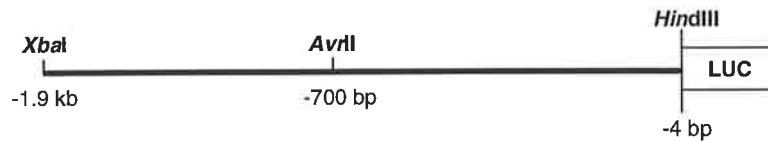
Excise an AvrII-HindIII fragment

Replace with an AvrII-HindIII

digested fragment (-700 to +28)

amplified by PCR

AvrII   HindIII
-700 bp   -4 bp

AvrII   HindIII
-700 bp   +1   +28 bp

NheI/
Xbal   HindIII
                                              LUC
pALAS-1.9kb-LUC
-1.9 kb   +1
                                    +28 bp

D

Kpnl   Xhol   Xbal   Ncol   BglII   +1
-6.0 kb -5.7 kb -4.6 kb   -1.9 kb   -1.0 kb   -293 bp   +5.0 kb

HindIII site
introduced at -7/-2

Ligate a Kpnl-HindIII fragment

into pGL2-Basic

HindIII

Kpnl   Ncol   LUC
-5.7 kb   -1.0 kb   -4 bp
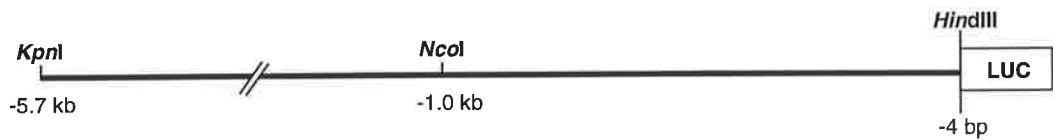
Excise an Ncol-HindIII fragment

Replace with an Ncol-HindIII fragment

(-1.0kb to +28) from pALAS-1.9kb-LUC

Ncol   HindIII   Ncol   HindIII
-1.0 kb   -4 bp   -1.0 kb   +1   +28 bp

pALAS-5.7kb-LUC   Kpnl   HindIII
                   -5.7 kb   +1   LUC
                                                                     +28 bp

E
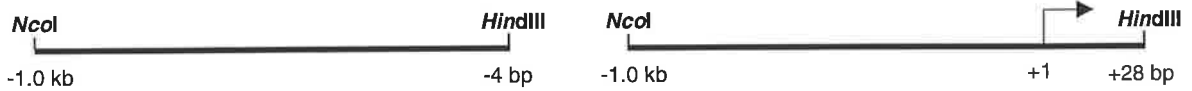
Ligate a *Cla*I-*Xho*I fragment (-10.3 kb to -4.6 kb)

isolated from pTC-EA1 into pSP72

pALAS-5.7kb-LUC

*Sma*I *Kpn*I   *Xho*I   *Hind*III

LUC

-5.7 kb   -4.6 kb   +1

+28 bp

*Eco*RV *Cla*I   *Xho*I   pSP72

-10.3 kb   -4.6 kb

Excise a *Sma*I-*Xho*I fragment

from pALAS-5.7kb-LUC

*Sma*I   *Xho*I

-5.7 kb   -4.6 kb

Excise an *Eco*RV-*Xho*I fragment

Ligate an *Eco*RV-*Xho*I

fragment (-10.3 kb to -4.6 kb)

*Eco*RV   *Xho*I

-10.3 kb   -4.6 kb

pALAS-10.3kb-LUC

*Sma*I/
*Eco*RV   *Xho*I   *Hind*III

LUC

-10.3 kb   -4.6 kb   +1

+28 bp

Figure 2.2   Construction of a chimeric human ALAS2-CAT reporter gene plasmid

To generate an ALAS2 promoter-CAT expression construct, an *Xba*I-*Hin*dIII fragment corresponding to -1.9 kb to +28 bp of the ALAS2 gene was isolated from pALAS-5.7kb-LUC. This fragment was blunted with Klenow enzyme and ligated into the *Sma*I linearised pIBICAT vector to produce pALAS-1.9kb-CAT.

pALAS-5.7kb-LUC

*Kpn*I −5.7 kb

*Xba*I −1.9 kb

+1

*Hind*III +28 bp

LUC
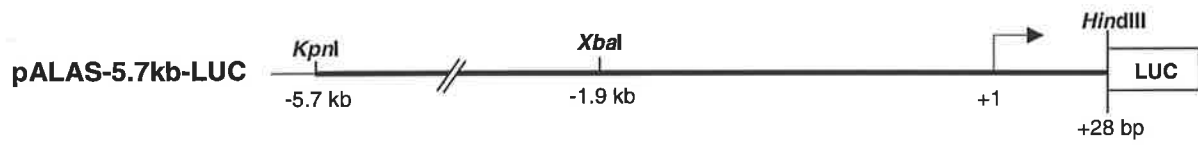
Isolate an *Xba*I-*Hind*III fragment
from pALAS-5.7kb-LUC

*Xba*I −1.9 kb

+1

*Hind*III +28 bp

Blunt with Klenow and
ligate into *Sma*I linearised pIBICAT

pALAS-1.9kb-CAT

*Eco*RI
*Sac*I
*Kpn*I

*Sma*I/
*Xba*I −1.9 kb

+1

*Sma*I/
*Hind*III

CAT
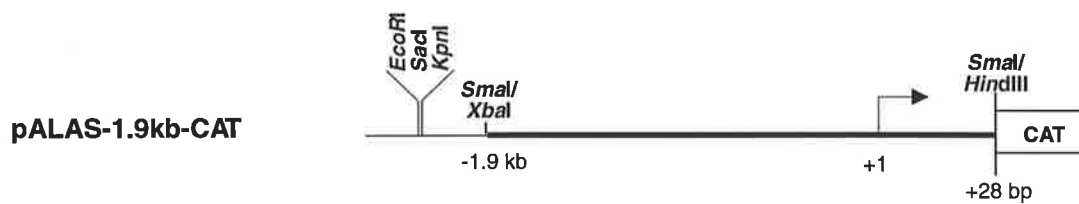
+28 bp

the 5'-flanking sequence, all of these constructs contained 28 bp of the 5'-untranslated region including exon 1.

2.4.2 Construction of ALAS2 Intron and 3'-Flanking/Reporter Gene Plasmids

i)   Synthesis of ALAS2 intron and 3'-flanking constructs fused to the human ALAS2 promoter

A series of constructs containing the human ALAS2 intron 1, intron 3 and intron 8 fragments and the 3'-flanking region were generated from the human genomic clone, pTC-EA1 (Cox *et al.*, 1991) and ligated into the plasmid pALAS-293-LUC. For simplicity, the plasmid pALAS-293-LUC described in Section 2.4.1 containing ALAS2 promoter sequence from -293 bp to +28 bp is referred to as pALASp-LUC in this section so as to distinguish between constructs containing 5'-flanking sequence or intronic sequences.

To synthesise a plasmid construct containing approximately 293 bp of ALAS2 promoter sequence and intron 1, a *Hin*dIII site was introduced at -7/-2 in the ALAS2 promoter and at +4973/4978 at the translation initiation site located in exon 2 by site-directed mutagenesis in a subclone containing -6.0 kb to +5.0 kb of contiguous human ALAS2 sequence. A 289 bp *Bgl*II-*Hin*dIII fragment (-293 bp to -4 bp) and a 4.98 kb *Hin*dIII fragment (-4 bp to +4976 bp) were individually isolated from this subclone. The 289 bp *Bgl*II-*Hin*dIII fragment was ligated into *Bgl*II-*Hin*dIII digested pGL2-Basic vector and then the 4.98 kb *Hin*dIII fragment was subsequently ligated into this *Hin*dIII linearised plasmid construct (Fig. 2.3A). The resulting construct was designated as pALASp-Int1-LUC and contained approximately 5.269 kb of contiguous ALAS2 sequence. Clones were initially screened by restriction enzyme analysis and sequenced to confirm the correct orientation of the intron 1 fragment downstream of the ALAS2 promoter.

The plasmid, pKS-Int3(850) containing the complete sequence of human ALAS2 intron 3 of approximately 850 bp was generated by PCR using the template, pTC-EA1 and ligated into the *Eco*RI site of pBluescript KS+ and was provided by Dr. Tim Cox. A *Sma*I-*Hin*dIII fragment was isolated from this clone, blunted with Klenow enzyme and ligated upstream of the 293 bp ALAS2 promoter, into the *Sma*I linearised pALASp-LUC and the resulting construct designated pALASp-Int3-LUC (Fig. 2.3B). Only clones of this plasmid construct containing intron 3 in the native orientation were isolated.

To isolate sequence of intron 8, the polymerase chain reaction was performed using pTC-EA1 as the template and the following primers: I8 primer 1, which bound at the exon/intron boundary of exon 8 and intron 8, and I8 primer 2, which bound at the intron/exon boundary between intron 8 and exon 9. Each primer contained an EcoRI site and together resulted in the amplification of a 562 bp intron 8 fragment. The amplified product was digested with EcoRI, and a 460 bp fragment (as a result of an internal EcoRI site) ligated into the similarly digested pBluescript KS+ vector and the resulting plasmid designated pKS-Int8(460) (Fig. 2.3C). A KpnI-BamHI fragment was then isolated from this clone, ligated into the KpnI/BglII linearised pALASp-LUC and the resulting plasmid construct designated pALASp-Int8(460)-LUC. However, only clones of this plasmid construct containing intron 8 in the native orientation were isolated and attempts to synthesise a construct containing the human intron 8 fragment located downstream of the luciferase reporter gene were unsuccessful. The KpnI-BamHI fragment was also ligated into the KpnI/BglII linearised construct pALAS-293-mut6-LUC containing a TATA box at the -27 site, introduced by site-directed mutagenesis. This constructs is referred to as pALASt-Int8(460)-LUC and clones containing intron 8 in both orientations were isolated.

To synthesise plasmid constructs containing the 3'-flanking sequence of the human ALAS2 gene, a 2.935 kb BglII-BamHI fragment was isolated from a subclone containing a 9.45 kb HindIII fragment from the genomic clone, pTC-EA1, corresponding to the 3' end of the human ALAS2 gene (Fig. 2.3D). This BglII-BamHI fragment was ligated into the BamHI linearised pALASp-LUC, downstream of the luciferase reporter gene. Plasmid clones containing the 3'-flanking sequence in both the native (pALASp-3'-LUC) and reverse (pALASp-3'R-LUC) orientations were isolated.

Plasmid constructs were also synthesised containing both intron 1 and intron 8 in which a fragment of intron 8 was ligated either upstream of the ALAS2 promoter or downstream of the luciferase reporter gene. To synthesise a plasmid construct containing a 460 bp intron 8 fragment ligated upstream of the ALAS2 promoter and intron 1, an EcoRV-SmaI fragment was isolated from pKS-Int8(460) and cloned into the SmaI linearised pALASp-Int1-LUC and the resulting construct designated pALASp-Int1-5'I8-LUC (Fig. 2.3E). However, only plasmid clones of pALASp-Int1-5'I8-LUC containing intron 8 in the native orientation were isolated.

To synthesise the plasmid construct, pALASp-Int1-3'I8-LUC, in which intron 8 was ligated downstream of the luciferase reporter gene, an *Eco*RV-*Sma*I fragment was isolated from pKS-Int8(460) and ligated into the *Sal*I linearised pALASp-Int1-LUC blunted with Klenow enzyme (Fig. 2.3E). Plasmid clones were isolated which contained the intron 8 fragment in both orientations and these constructs are designated pALASp-Int1-3'I8-LUC and pALASp-Int1-3'I8R-LUC to distinguish between the intron 8 fragment in the native and reverse orientations, respectively.

ii) Synthesis of ALAS2 intron constructs fused to the heterologous thymidine kinase promoter

Plasmid constructs were also synthesised containing the human ALAS2 intron 1 and intron 8 ligated upstream of the thymidine kinase promoter in the plasmid ptk-LUC. The plasmid construct, ptk-LUC contained a 164 bp *Bam*HI-*Bgl*II fragment of the thymidine kinase promoter isolated from pBLCAT2 (Luckow & Schutze, 1987) and ligated into the *Bgl*II linearised pGL2-Basic vector and was kindly provided by Dr. C. Hahn.

To synthesise a plasmid construct containing an intron 1 fragment ligated upstream of the thymidine kinase promoter, a 3.984 kb *Sac*I fragment containing the majority of putative transcription factor binding sites identified within intron 1, was isolated from pALASp-Int1-LUC and ligated into the similarly digested ptk-LUC vector and the resulting construct designated, ptk-Int1-LUC (Fig. 2.3F). Plasmid clones which were isolated only contained the intron 1 fragment in the native orientation.

To synthesise plasmid constructs containing a 460 bp human intron 8 fragment ligated upstream of the thymidine kinase promoter, a *Kpn*I-*Sac*I fragment and a *Kpn*-*Sma*I fragment were individually isolated from pKS-Int8(460) and ligated in the native and reverse orientations, upstream of the thymidine kinase promoter in the similarly digested ptk-LUC to generate ptk-Int8(460)-LUC and ptk-Int8(460)R-LUC, respectively (Fig. 2.3G).

Sequence comparison of the human and murine ALAS2 intron 8 sequences revealed numerous putative transcription factor binding sites which were conserved between the two species (Cox, 1993). Therefore plasmid constructs were synthesised which contained the corresponding fragment of the murine ALAS2 intron 8 ligated upstream of the thymidine

Figure 2.3   Cloning strategies used to generate chimeric ALAS2 intron and 3'-flanking-luciferase reporter gene plasmids

Details of the cloning procedures are outlined in the text.

A

KpnI   BglI   (arrow +1)
-6.0 kb -5.7 kb   -293 bp   +1   +5.0 kb

Isolate a BglII-HindIII fragment (-293 to -4)

HindIII site
introduced at -7/-2

HindIII site introduced at
the ATG in Exon 2 (+4976)

Isolate a HindIII fragment (-4 to +4976)

BglI   HindIII
-293 bp   -4 bp

Ligate into BglII/HindIII
digested pGL2-Basic

HindIII   Human Intron 1
4.976 kb   HindIII
-4 bp   +1   +4976 bp

Ligate HindIII fragment into
HindIII linearised pALAS-(-293/-4)-LUC

pALAS-(-293/-4)-LUC
BglI   HindIII   LUC
-293 bp   -4 bp

Human Intron 1
4.976 kb   HindIII

pALASp-Int1-LUC
BglI   HindIII   LUC
-293 bp   -4   +1   +4976 bp

B

pKS-Int3(850)

KpnI XhoI SaII ClaI HindIII EcoRV

← Human Intron 3 850 bp

PstI SmaI BamHI SpeI XbaI SacI

EcoRI ————————— EcoRI pKS⁺

Ligate a blunted *Sma*I-*Hin*dIII fragment

into *Sma*I linearised pALASp-LUC

↓

pALASp-Int3-LUC

→ Human Intron 3 850 bp

*Sma*I ———//——— HindIII/ SmaI ——— BgIII ——————————— HindIII — | LUC |

-293 bp   +1   +28 bp

C

I8 Primer 1

I8 Primer 2

Exon 8   *Eco*RI   Intron 8   Exon 9   pTC-EA1

introduce *Eco*RI site   introduce *Eco*RI site
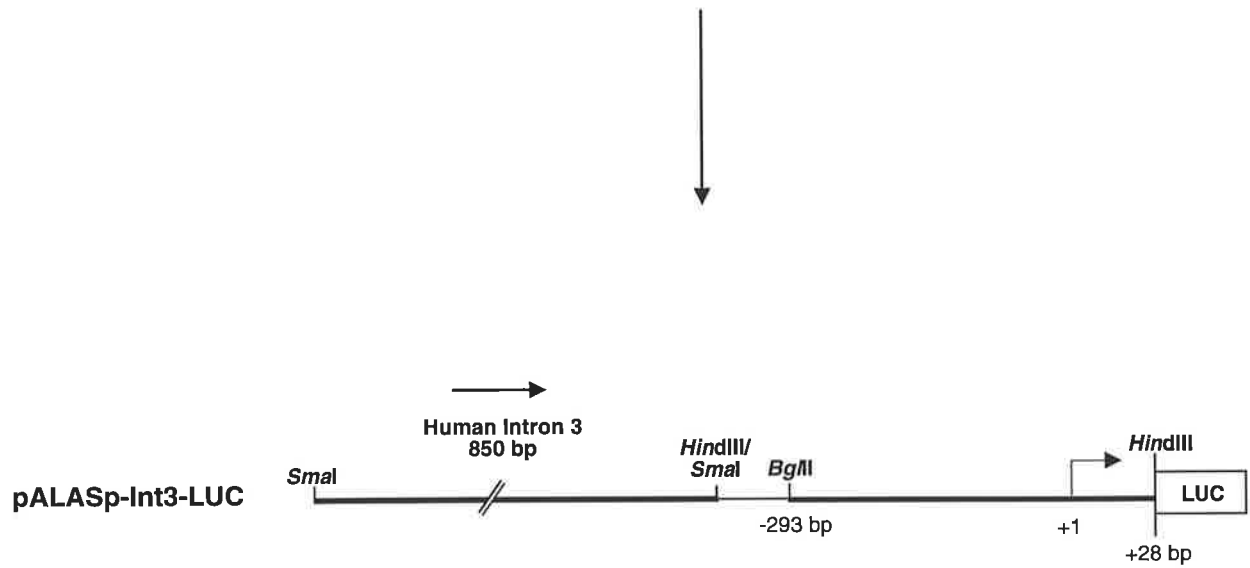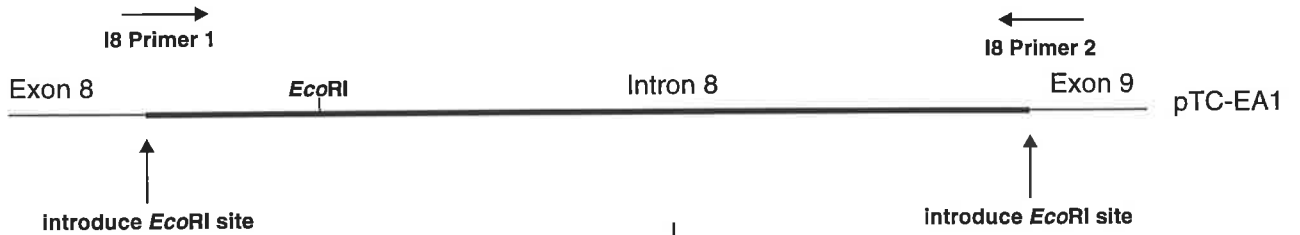
Digest amplified product with *Eco*RI

and ligate into pBluescript KS⁺

*Kpn*I
*Xho*I
*Sal*I
*Cla*I
*Hin*dIII
*Eco*RV

*Pst*I
*Sma*I
*Bam*HI
*Spe*I
*Xba*I
*Sac*I

Human Intron 8
460 bp

pKS-Int8(460)   *Eco*RI   *Eco*RI   pKS⁺
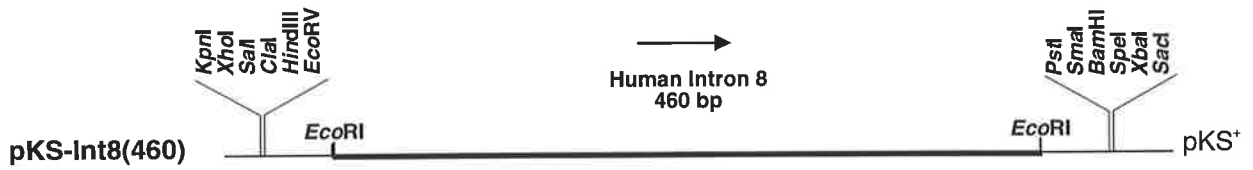
Ligate a *Kpn*I-*Bam*HI fragment into

*Kpn*I/*Bgl*II digested pALASp-LUC

*Kpn*I
*Xho*I
*Sal*I
*Cla*I
*Hin*dIII
*Eco*RV

*Pst*I
*Sma*I

*Hin*dIII

Human Intron 8
460 bp

pALASp-
Int8(460)-LUC   *Sma*I   *Eco*RI   *Eco*RI   *Bam*HI/
                                                                              *Bgl*II   LUC

                                                                    -293 bp   +1
                                                                                        +28 bp

D

Intron 8

*Hind*III   *Bgl*I   3'-flanking   *Bam*HI   *Hind*III   pTC-EA1

Isolate a *Bgl*I-*Bam*HI fragment

*Bgl*I   3'-flanking 2.935 kb   *Bam*HI

Ligate into the *Bam*HI linearised pALASp-LUC

pALASp-3'-LUC

*Bgl*I   *Hind*III   *Bam*HI/ *Bgl*I   3'-flanking 2.935 kb   *Bam*HI

LUC

-293 bp   +1   +28 bp

pALASp-3'R-LUC

*Bgl*I   *Hind*III   *Bam*HI   3'-flanking 2.935 kb   *Bam*HI/ *Bgl*I

LUC

-293 bp   +1   +28 bp

E

pKS-Int8(460)

*Kpn*I *Xho*I *Sal*I *Cla*I *Hind*III *Eco*RV

*Eco*RI

Human Intron 8
460 bp

*Eco*RI

*Pst*I *Sma*I *Bam*HI *Spe*I *Xba*I *Sac*I

pKS⁺

Ligate an *Eco*RV-*Sma*I fragment
into *Sma*I linearised pALASp-Int1-LUC

Ligate an *Eco*RV-*Sma*I fragment
into blunted *Sal*I linearised pALASp-Int1-LUC

pALASp-Int1-LUC

*Sma*I *Kpn*I *Sac*I *Mlu*I *Nhe*I *Xho*I

*Bgl*II

−293 bp

+1

Human Intron 1
4.976 kb

*Hind*III

LUC

+4976 bp

*Bam*HI *Sal*I

pALASp-Int1-5'I8-LUC

Human Intron 8

*Sma*I/ *Eco*RV

*Sma*I *Bgl*II

−293 bp

+1

Human Intron 1

*Hind*III

LUC

+4976 bp

pALASp-Int1-3'I8-LUC

Human Intron 8

*Bgl*II

−293 bp

+1

Human Intron 1

*Hin*dIII

LUC

+4976 bp

*Sal*I/ *Eco*RV

*Sma*I/ *Sal*I

pALASp-Int1-3'I8R-LUC

*Bgl*II

−293 bp

+1

*Hind*III

LUC

+4976 bp

*Sal*I/ *Sma*I

Human Intron 8

*Eco*RV/ *Sal*I

F

pALASp-Int1-LUC

*Sma*I *Kpn*I *Sac*I *Mlu*I *Nhe*I *Xho*I

*Bgl*I

*Sac*I

Human Intron 1
4.976 kb

*Sac*I

*Hind*III

LUC

-293 bp

+1

+4976 bp

Intron 1
3.984 kb

*Sac*I

*Sac*I

Ligate a *Sac*I fragment from pALASp-Int1-LUC

into *Sac*I linearised ptk-LUC

ptk-Int1-LUC

*Sma*I *Kpn*I

*Sac*I

Human Intron 1

*Sac*I

*Bam*HI

tk promoter

*Bgl*I

LUC

+1

G

pKS-Int8(460)

KpnI XhoI SaII ClaI HindIII EcoRV

EcoRI

Human Intron 8
460 bp

PstI SmaI BamHI SpeI XbaI SacI

EcoRI

pKS⁺
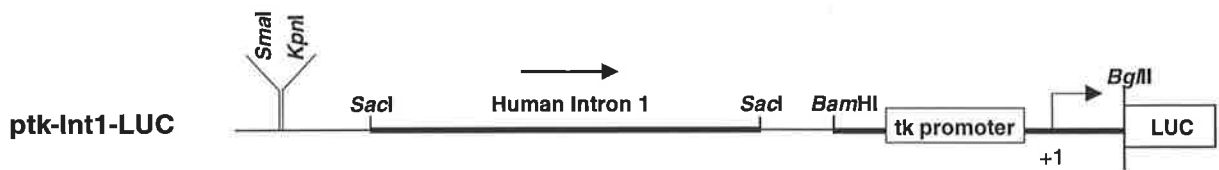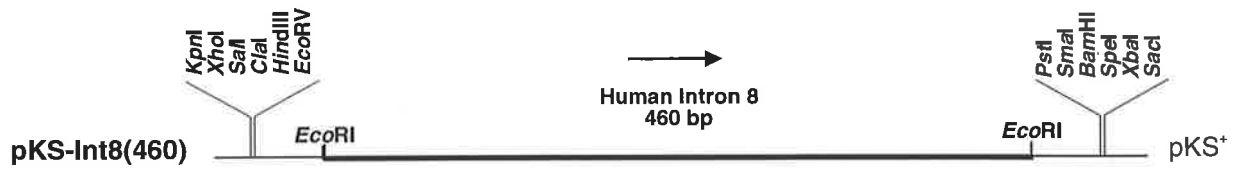
Individually ligate a *Kpn*I-*Sac*I fragment
and a *Kpn*I-*Sma*I fragment into ptk-LUC

ptk-Int8(460)-LUC

KpnI

Human Intron 8

SacI BamHI

tk promoter

BgllI

+1

LUC

ptk-Int8(460)R-LUC

SmaI

Human Intron 8

KpnI BamHI

tk promoter

BgllI

+1

LUC

H

Murine Intron 8
330 bp
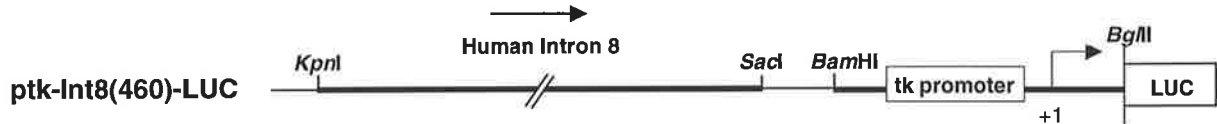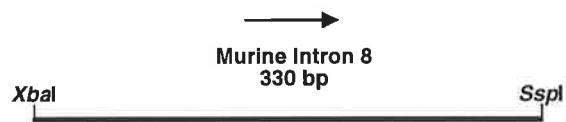
XbaI

SspI

Ligate an *Xba*I-*Ssp*I fragment into
*Nhe*I/*Sma*I linearised ptk-LUC

ptk-mInt8(330)R-LUC

SmaI/
SspI

Murine Intron 8

XbaI/
NheI BamHI

tk promoter

BgllI

+1

LUC
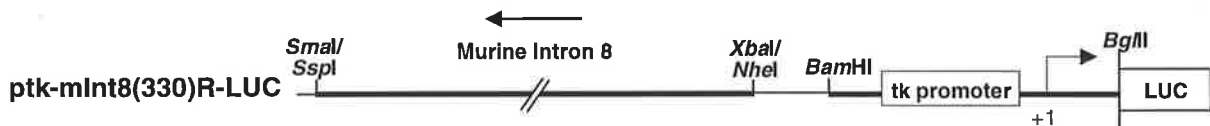
kinase promoter (Fig. 2.3H). A 330 bp *Xba*I-*Ssp*I fragment containing the murine ALAS2 intron 8 sequence was isolated from a genomic clone (kindly provided by Helen Chandler) and ligated in the reverse orientation into the *Sma*I/*Nhe*I linearised ptk-LUC. The resulting construct is designated ptk-mInt8(330)R-LUC. However, repeated attempts to synthesise a plasmid construct containing the murine ALAS2 intron 8 sequence in the native orientation were unsuccessful.

2.4.3   Construction of Human ALAS2 Intron 8 Deletion/Reporter Gene Plasmids

A series of ALAS2 intron 8 deletion constructs ranging in length from 400 bp to 115 bp were synthesised.

To synthesise a plasmid construct containing 400 bp of human ALAS2 intron 8 sequence, a 400 bp *Pst*I fragment was isolated from pKS-Int8(460), blunted with T4 DNA polymerase enzyme and ligated into the *Sma*I linearised pALASp-LUC to create pALASp-Int8(400)-LUC (Fig. 2.4A). To generate a plasmid containing 279 bp of intron 8 sequence extending from the native *Hin*dIII site to the *Eco*RI site introduced at the end of intron 8, the plasmid pALASp-Int8(460)-LUC was digested with *Hin*dIII and the 600 bp *Hin*dIII fragment containing intron 8 and the ALAS2 promoter was religated with the vector fragment and the resulting plasmid designated pALASp-Int8(279)-LUC (Fig. 2.4B). To further analyse this region, the 279 bp intron 8 fragment was subdivided into two smaller fragments of 115 bp and 177 bp which were amplified by PCR using the template, pALASp-Int8(400)-LUC   and   I8 primer 3   with   I8 primer 4,   and   I8 primer 5   with I8 primer 6, respectively (Fig. 2.4C). This resulted in the amplification of a 115 bp and a 177 bp fragment which were each digested *Xho*I and *Sal*I site and individually ligated into the *Xho*I linearised pALASp-LUC plasmid in both orientations. The resulting plasmids were designated as pALASp-Int8(115)-LUC and pALASp-Int8(177)-LUC, respectively. To synthesise a plasmid construct containing a 235 bp fragment of intron 8 sequence from the native *Pst*I site extending to the introduced *Sal*I site, the polymerase chain reaction was employed using pTC-EA1 as the template and I8 primer 1 and I8 primer 4 (Fig. 2.4D). The amplified fragment of approximately 400 bp was digested with *Pst*I and *Sal*I and the 235 bp fragment ligated into the similarly digested pBluescript KS+ vector to generate pKS-Int8(235). This plasmid was digested with *Sma*I and *Sal*I and ligated into the

Figure 2.4   Construction of human ALAS2 intron 8 deletion constructs

Human ALAS2 intron 8 deletion constructs were generated, ranging in length from 400 bp to 115 bp. The details of the cloning procedures are outlined in the text. The mutated sites are represented by a *cross*.

A

pKS-Int8(460)

*Kpn*I *Xho*I *Sal*I *Cla*I *Hind*III *Eco*RV

*Eco*RI   *Pst*I

Human Intron 8
460 bp

*Eco*RI

*Pst*I *Sma*I *Bam*HI *Spe*I *Xba*I *Sac*I

pKS+

*Pst*I   400 bp   *Pst*I
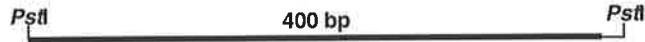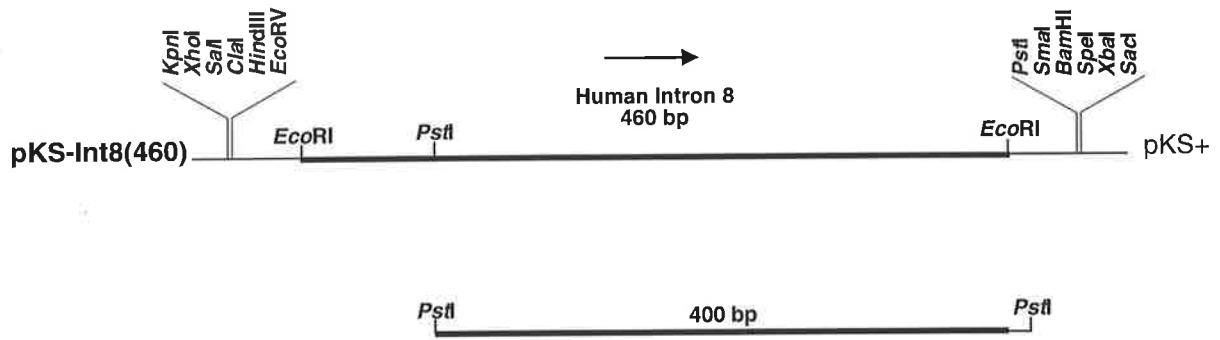
Ligate a blunted *Pst*I fragment

into *Sma*I linearised pALASp-LUC

pALASp-
Int8(400)-LUC

*Sma*I/
*Pst*I

Intron 8
400 bp

*Eco*RI

*Kpn*I *Sac*I *Mlu*I *Nhe*I *Xho*I

*Pst*I *Sma*I

*Bgl*II

-293 bp

+1

*Hind*III

LUC

+28 bp

B

pALASp-
Int8(460)-LUC

*Sma*I *Kpn*I *Xho*I *Sal*I *Cla*I *Hind*III *Eco*RV

*Eco*RI   *Hind*III

Intron 8
460 bp

*Eco*RI

*Pst*I *Sma*I

*Bam*HI/
*Bgl*II

-293 bp

+1

*Hind*III

LUC

+28 bp

*Hind*III   *Hind*III
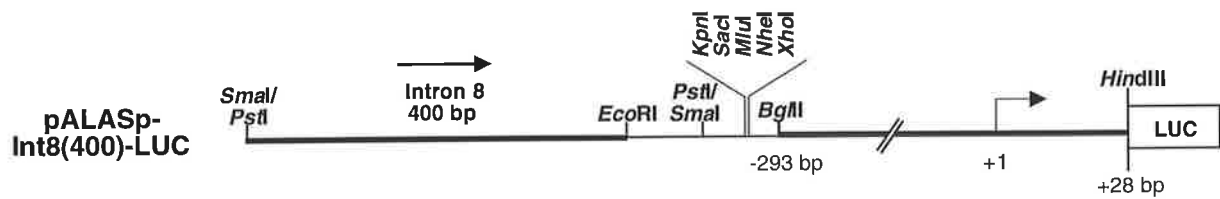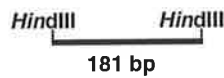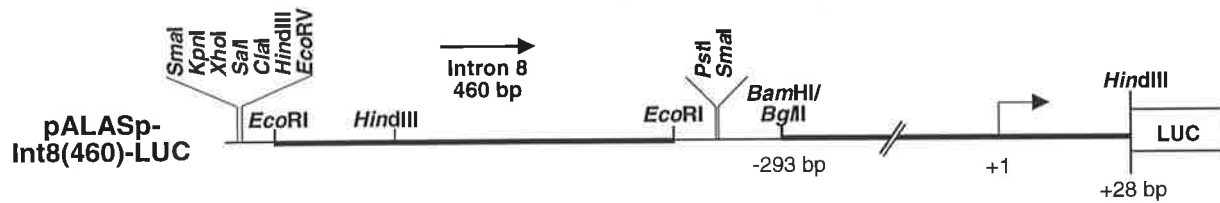
181 bp

Digest with *Hind*III, discard 181 bp

fragment and ligate the 600 bp

*Hind*III fragment with the vector

pALASp-
Int8(279)-LUC

*Sma*I *Kpn*I *Xho*I *Sal*I *Cla*I

*Hind*III

Intron 8
279 bp

*Eco*RI

*Pst*I *Sma*I

*Bam*HI/
*Bgl*II

-293 bp

+1

*Hind*III

LUC

+28 bp

C

pALASp-Int8(400)-LUC

*Sma*I/*Pst*I — Intron 8 400 bp — *Eco*RI — *Pst*I/*Sma*I — *Kpn*I *Sac*I *Mlu*I *Nhe*I *Xho*I — *Bgl*II (-293 bp) — +1 — *Hind*III (+28 bp) — LUC

*Sma*I/*Pst*I — I8 Primer 3 — I8 Primer 4 — I8 Primer 5 — I8 Primer 6

*Sma*I/*Pst*I — *Hind*III — *Xho*I site — *Xho*I site — *Sal*I site — *Sal*I site
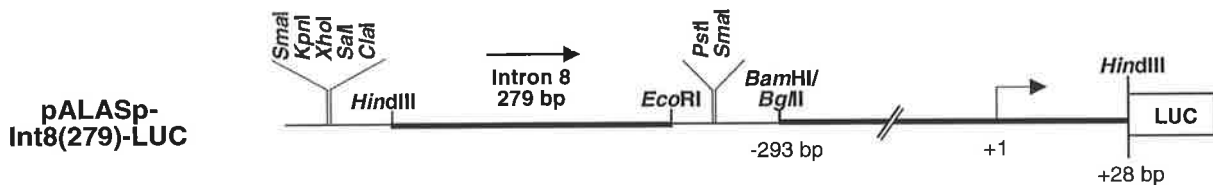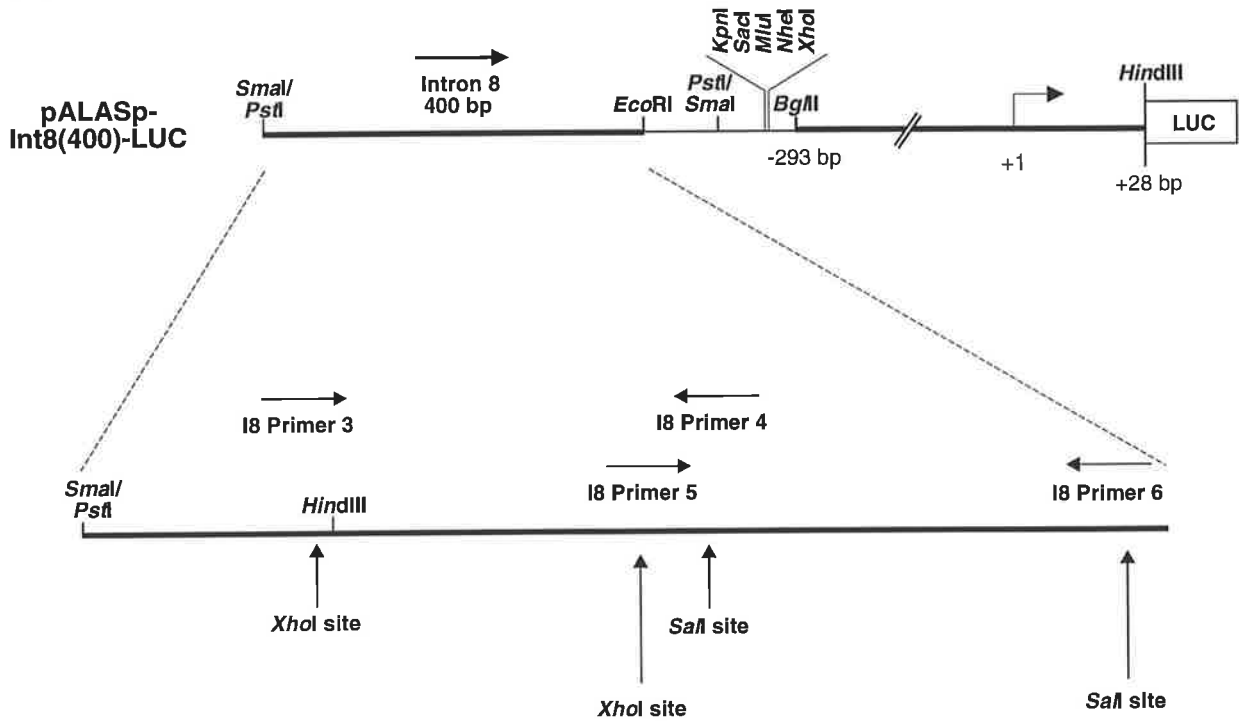
Digest amplified products with *Xho*I and *Sal*I
and ligate into *Xho*I linearised pALASp-LUC

*Xho*I — 115 bp — *Sal*I

*Xho*I — 177 bp — *Sal*I

pALASp-Int8(115)-LUC

*Sma*I *Kpn*I *Sac*I *Mlu*I *Nhe*I — *Xho*I — Intron 8 115 bp — *Sal*I/*Xho*I — *Bgl*II (-293 bp) — +1 — *Hind*III (+28 bp) — LUC

pALASp-Int8(177)-LUC

*Sma*I *Kpn*I *Sac*I *Mlu*I *Nhe*I — *Xho*I — Intron 8 177 bp — *Sal*I/*Xho*I — *Bgl*II (-293 bp) — +1 — *Hind*III (+28 bp) — LUC

D

Exon 8 / Intron 8 / Exon 9 map (pTC-EA1) showing I8 Primer 1, I8 Primer 4, *Eco*RI, *Pst*I, *Hind*III, and *Sal* I site.
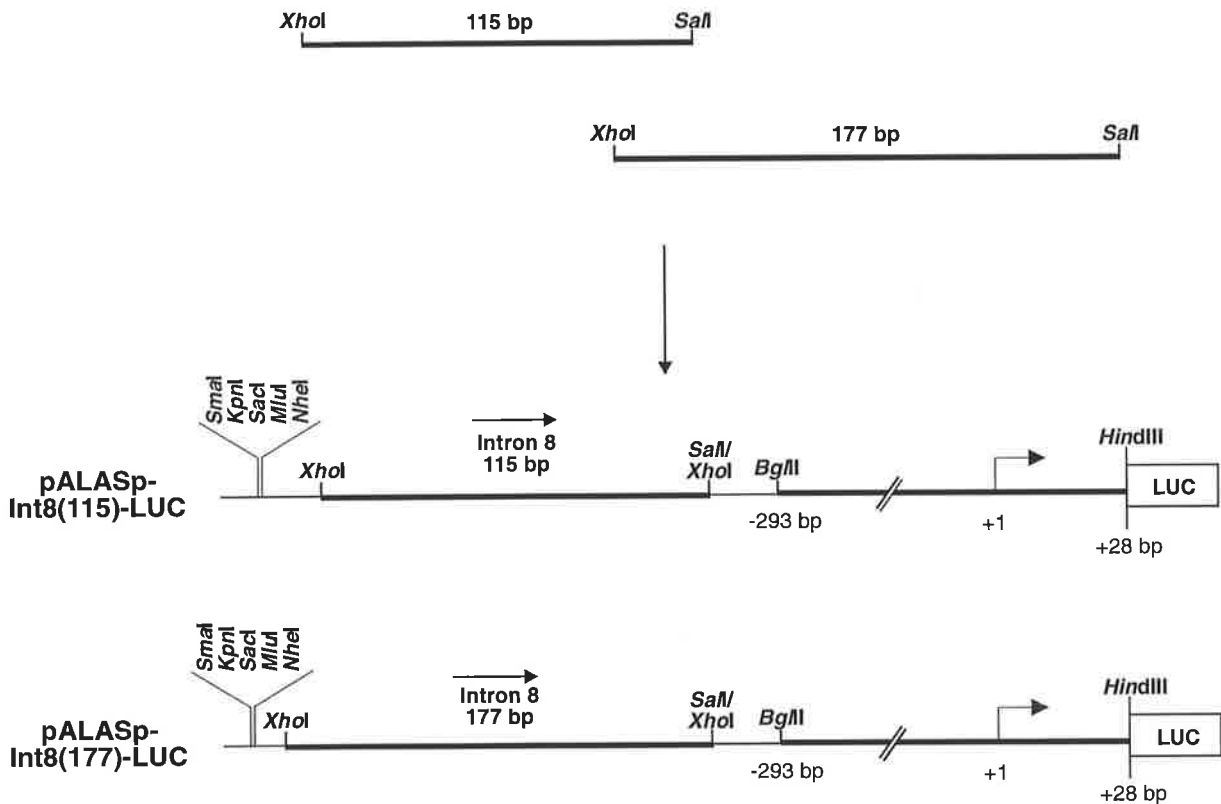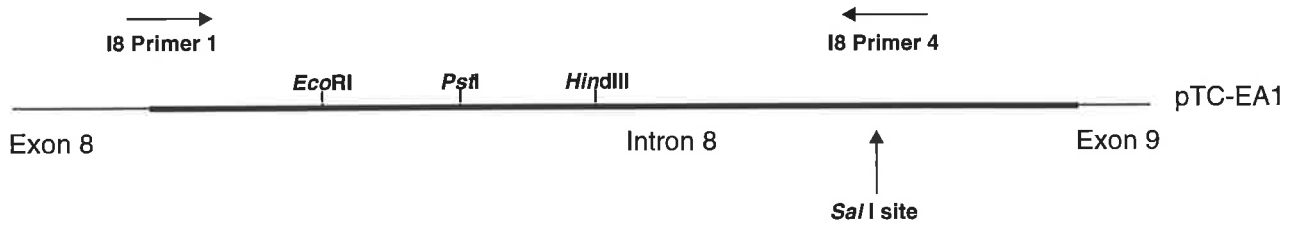
Digest amplified product with *Pst*I and *Sal*I

and ligate into pBluescript KS⁺

pKS-Int8(235) — Intron 8 235 bp, *Sal*I, *Hind*III, *Pst*I, pKS⁺ (*Kpn*I, *Xho*I / *Sma*I, *Bam*HI, *Spe*I, *Xba*I, *Sac*I)
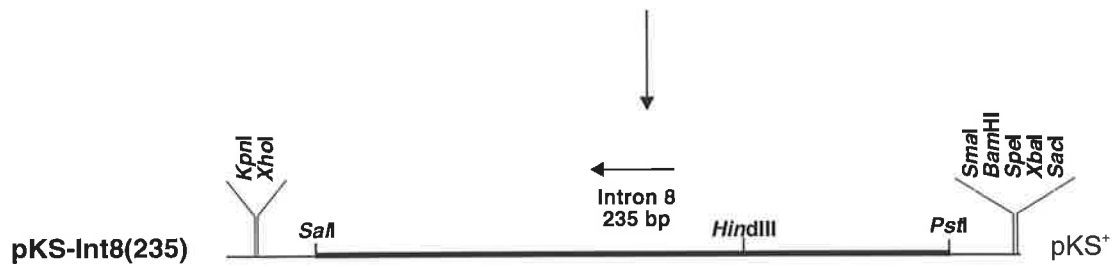
Isolate a *Sma*I-*Sal*I fragment and ligate

into *Sma*I/*Xho*I linearised pALASp-LUC

pALASp-Int8(235)-LUC — *Sma*I, *Pst*I, *Hind*III, Intron 8 235 bp, *Sal*I/*Xho*I, *Bgl*II, -293 bp, +1, *Hind*III, +28 bp, LUC
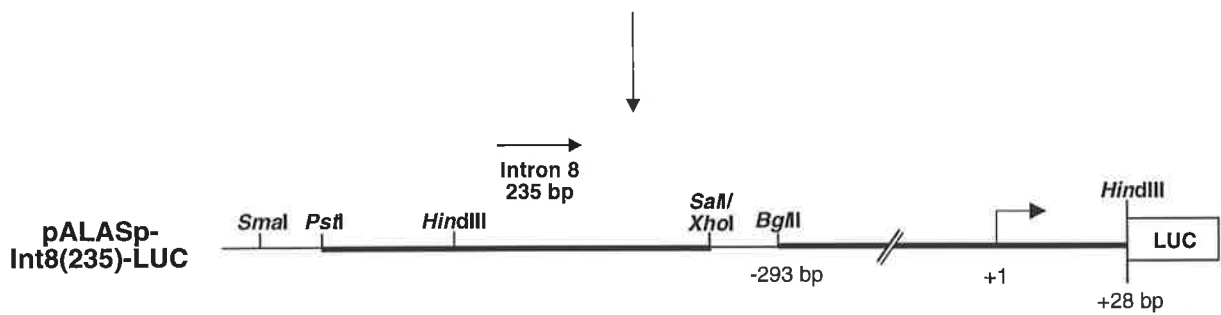
Isolate a *Sma*I-*Hind*III fragment, blunt with Klenow

and ligate into *Sma*I linearised pALASp-LUC

pALASp-Int8(120)-LUC — *Sma*I, *Pst*I, *Hind*III/*Sma*I, Intron 8 120 bp, *Kpn*I, *Sac*I, *Mlu*I, *Nhe*I, *Xho*I, *Bgl*II, -293 bp, +1, *Hind*III, +28 bp, LUC
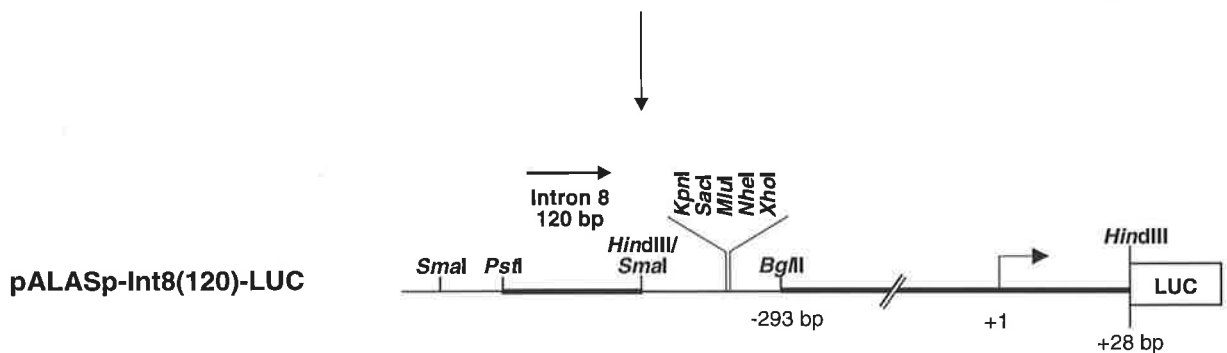
SmaI/XhoI linearised pALASp-LUC to generate pALASp-Int8(235)-LUC (Fig. 2.4D). A 120 bp *SmaI-Hind*III fragment was then isolated from this plasmid, blunted with Klenow enzyme and ligated into the *Sma*I linearised pALASp-LUC to create pALASp-Int8(120)-LUC (Fig. 2.4D).

2.4.4 Site-directed Mutagenesis of the Human ALAS2 Intron 8 Sequence

Site-directed mutagenesis was performed with the plasmid pKS-Int8(235) containing 235 bp of human ALAS2 intron 8 sequence. Following mutagenesis, a *SmaI-Xho*I fragment harbouring a single mutation was excised from pKS-Int8(235) and ligated into the similarly digested pALASp-LUC (Fig. 2.5A).

To generate constructs containing the mutated CACCC site A in combination with either the mutated CACCC site B, GATA site A or GATA site B, a two-step cloning procedure was performed (see Fig. 2.5B). The plasmid pALASp-mut1-LUC containing the mutated CACCC site A was digested with *Hind*III and *Sal*I, the 115 bp fragment removed and replaced with a *Hind*III-*Sal*I fragment isolated from pALASp-mut2, mut3 and mut4-LUC containing a mutation in CACCC site B, GATA site A and GATA site B, respectively (Fig. 2.5B). A *SmaI-Xho*I fragment harbouring the double mutations introduced into the human intron 8 sequence was then isolated from pALASp-mut5-LUC and pALASp-mut7-LUC (Fig. 2.5B) and ligated into the similarly digested ptk-LUC vector.

2.4.5 Synthesis of Plasmid Constructs for use in Transactivation Assays

The plasmid reporter constructs, pC1G3tk-CAT, containing four copies of the EKLF responsive CACCC (-94/-87) derived from the adult β-globin promoter ligated upstream of the thymidine kinase promoter fused to the CAT reporter gene, and pβ-globin-CAT containing the murine adult β-globin promoter, (kindly provided by Dr. J. Bieker), were employed as positive controls in co-transfection assays. Plasmid constructs were then synthesised which contained the luciferase reporter gene. An *NdeI-Xho*I fragment from pC1G3tk-CAT (Bieker & Southwood, 1995) was blunted with Klenow enzyme and ligated into the *Sma*I linearised pGL2-Basic to generate p(CAC)$_4$tk-LUC. The plasmid, pβ-globin-LUC was constructed by the removal of the

Figure 2.5   Strategy used to introduce mutations in the human ALAS2 intron 8 sequence

A   A *Pvu*II restriction site was introduced at each CACCC and GATA site (*boxed*) in the plasmid pKS-Int8(235). A *Sma*I-*Xho*I fragment harbouring a single mutation at either CACCC site A (CAC-A), CACCC site B (CAC-B), GATA site A (GATA-A) or GATA site B (GATA-B) was excised and ligated into the similarly digested pALASp-LUC.

B   Plasmid constructs containing a mutation in CAC-A in combination with a mutation in CAC-B, GATA-A or GATA-B in human ALAS2 intron 8 were generated as described in the text.

A

pKS-Int8(235)

Human Intron 8
235 bp

XhoI   SalI   GATA-B   GATA-A   CAC-B   HindIII   CAC-A   PstI   SmaI   pKS⁺

Introduce *Pvu*II site

by Site-directed Mutagenesis
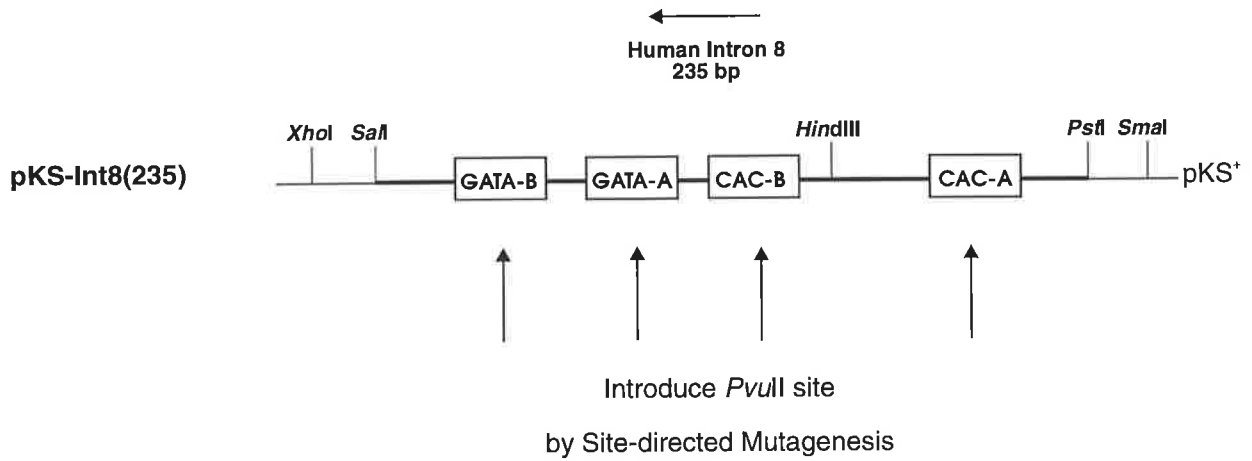
Isolate a *Sma*I-*Xho*I fragment from
pKS-Int(235) harbouring a single mutation
and ligate into pALASp-LUC

Human Intron 8
235 bp

pALASp-mut1-LUC   SmaI   PstI   X   HindIII   CAC-B   GATA-A   GATA-B   SalI   XhoI   ALAS   +1   LUC

pALASp-mut2-LUC   SmaI   PstI   CAC-A   HindIII   X   GATA-A   GATA-B   SalI   XhoI   ALAS   +1   LUC

pALASp-mut3-LUC   SmaI   PstI   CAC-A   HindIII   CAC-B   X   GATA-B   SalI   XhoI   ALAS   +1   LUC

pALASp-mut4-LUC   SmaI   PstI   CAC-A   HindIII   CAC-B   GATA-A   X   SalI   XhoI   ALAS   +1   LUC
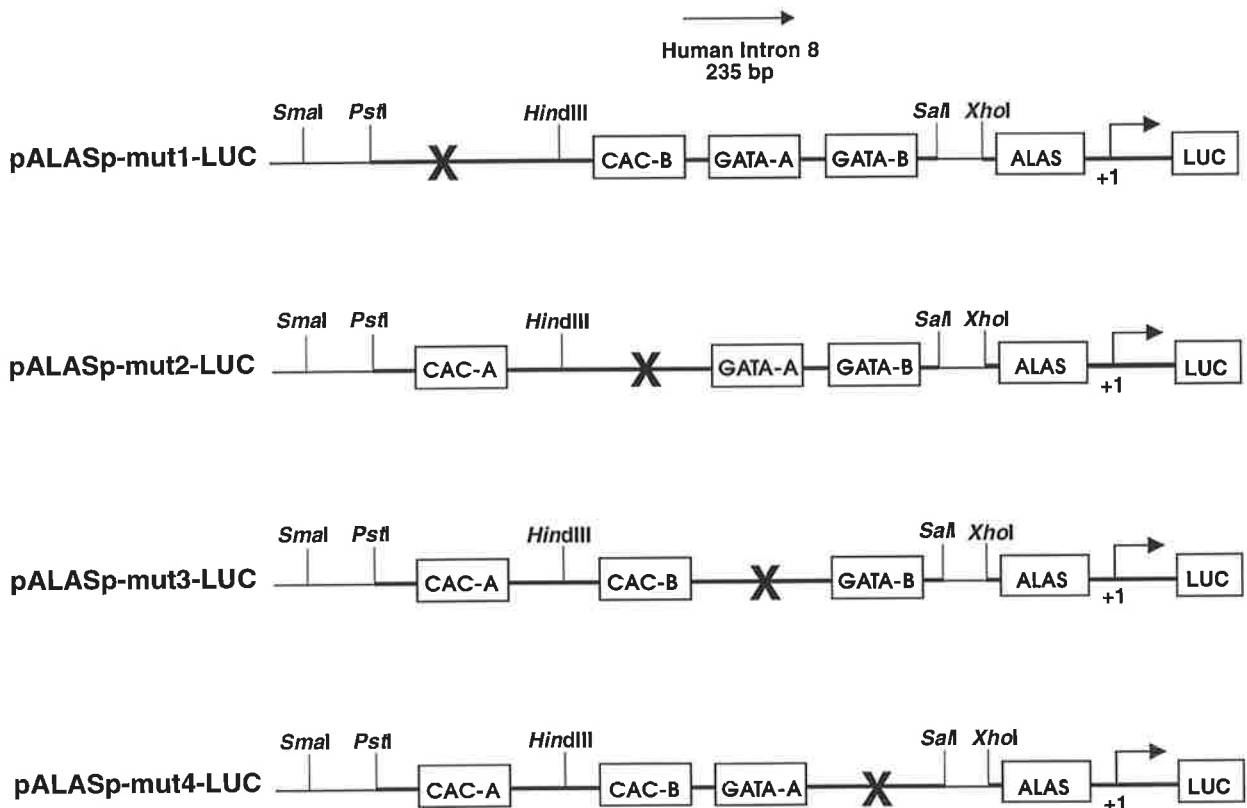
B
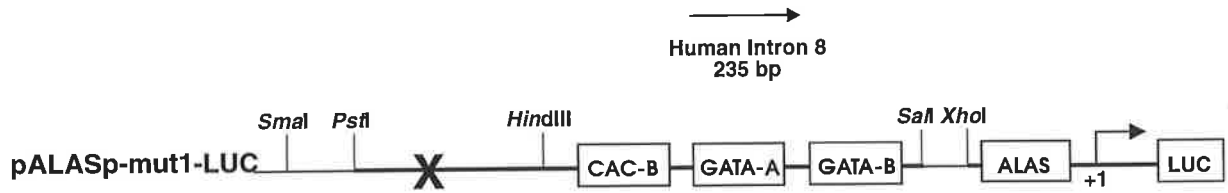
Excise a *Hind*III-*Sal*I fragment from pALASp-mut1-LUC and

replace with a *Hind*III-*Sal*I fragment from pALASp-mut2 or mut3 or mut4 -LUC
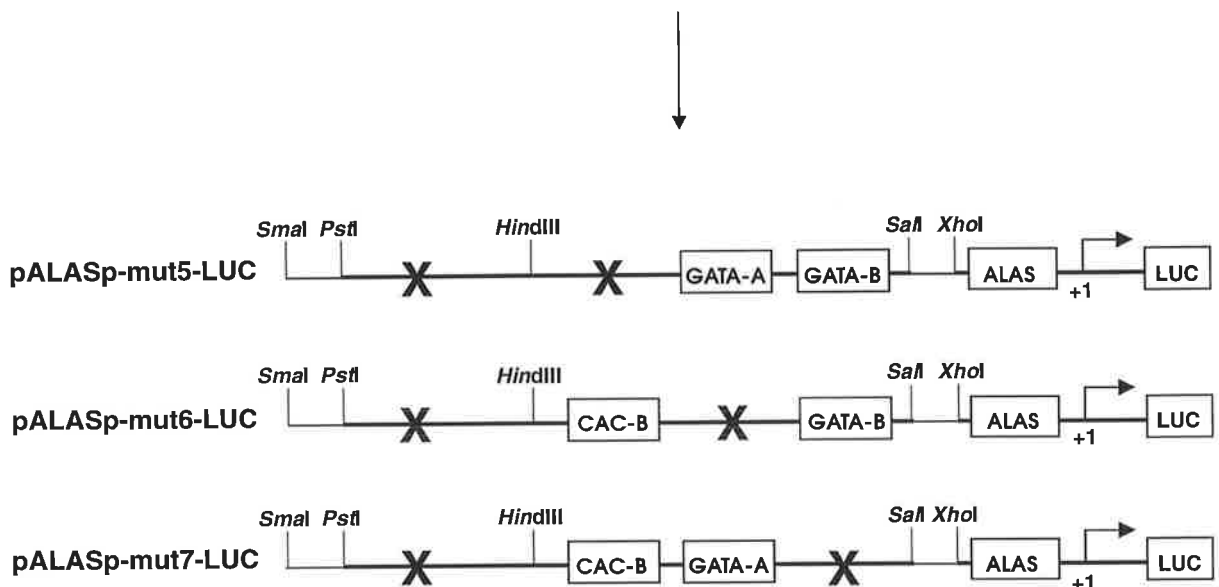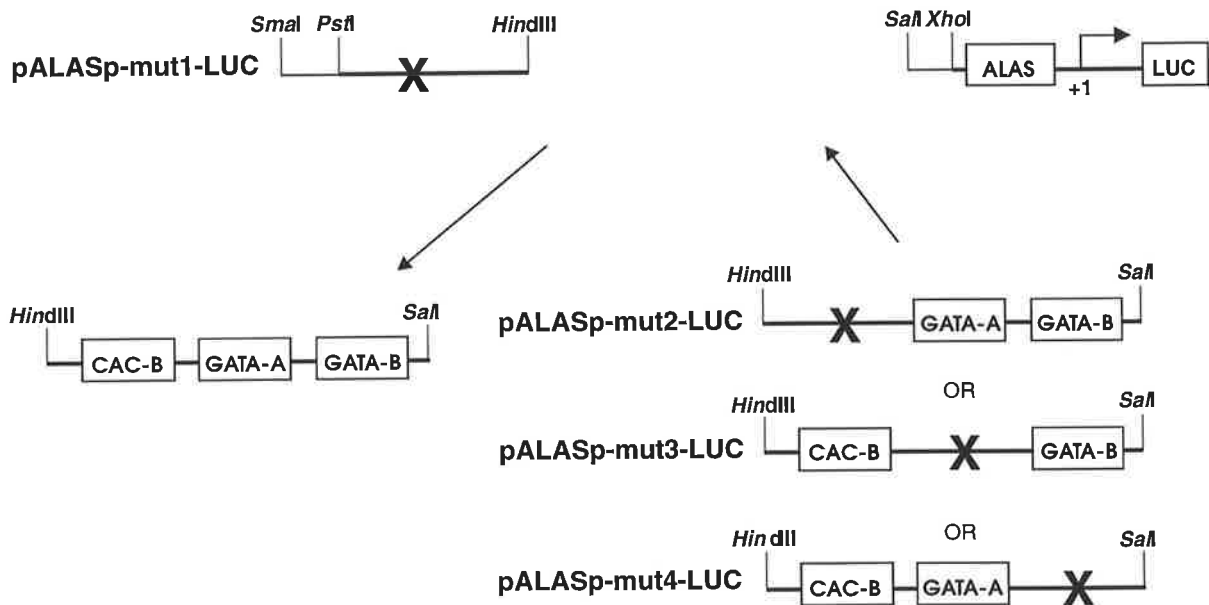
*Xba*I-*Bgl*II fragment from pβ-globin-CAT (Bieker & Southwood, 1995) and the subsequent ligation into the *Nhe*I/*Bgl*II linearised pGL2-Basic.

A series of ALAS2 promoter constructs were synthesised for use in transactivation assays. Plasmid constructs containing 152 bp (-124 bp to +28 bp) of ALAS2 promoter sequence were generated by PCR using as templates the plasmids pALAS-293-LUC and pALAS-293mut7-LUC (containing a mutation introduced at -54), and primer 1 with primer 3. Following digestion of the amplified products with *Bgl*II and *Hin*dIII, the resulting fragments were individually cloned into the similarly digested pGL2-Basic vector. The resulting constructs were designated pALAS-124-LUC and pALAS-124mut-LUC, respectively (Fig. 2.6A).

ALAS2 promoter constructs which terminated at -4 and did not contain the transcription initiation site were created by PCR using as templates the plasmids, pALAS-124-LUC and pALAS-124mut-LUC and primer 3 and RW-142 (Fig. 2.6B). These amplified products were digested with *Bgl*II and *Hin*dIII, ligated into the similarly digested pGL2-Basic and designated pALAS-124-CAP-LUC and pALAS-124mut-CAP-LUC, respectively (Fig. 2.6B).

ALAS2 promoter constructs (-124 bp to +28 bp) containing a mutation in the putative CACCC site located at -3/+5, (5'-ACCCACCT-3' to 5'-ACCC<u>GTT</u>-3') were generated by PCR using as templates the plasmids, pALAS-124-LUC and pALAS-124mut-LUC, and primer 3 with primer 4 (Fig. 2.6C). The amplified products were digested with *Bgl*II and *Hin*dIII and ligated into the similarly digested pGL2-Basic vector. The resulting plasmids were designated pALAS-124CAPmut-LUC and pALAS-124mutCAPmut-LUC and contained an introduced mutation at +2/+4 and the wild-type or the mutated -54 CACCC sequence, respectively (Fig. 2.6C).

Two additional ALAS2 promoter constructs (-27 bp to +28 bp) were synthesised for use in transactivation assays, pALAS-27-LUC and pALAS-27mut-LUC containing an introduced mutation at +2/+4 (Fig. 2.6D). These two fragments were amplified by PCR using as templates the plasmids, pALAS-124-LUC and pALAS-124CAPmut-LUC and primer GATA-27 with primer 1, then digested with *Pvu*II and *Hin*dIII and ligated into the *Sma*I/*Hin*dIII digested pGL2-Basic vector (Fig. 2.6D).

Plasmid constructs containing 152 bp of ALAS2 promoter sequence (-124 bp to +28 bp) were also generated in the vector, pCATOO containing two synthetic poly A sites

Figure 2.6   Synthesis of chimeric human ALAS2 promoter-luciferase reporter gene plasmids for use in transactivation assays

A series of human ALAS2 promoter-luciferase constructs were generated by PCR (details are described in the text). The primers used in the amplification of ALAS2 promoter sequence are indicated. The location of the mutated sites are shown and are represented by a *cross*.

A

B

pALAS-124-LUC

Primer 3   RW-142

*Bgl*I   GGGTGGGTGGGG   *Hind*III   LUC

-124 bp   -54   +1   +28 bp

*Hind*III site at -4

pALAS-124 mut-LUC

Primer 3   RW-142

*Bgl*I   X   *Hind*III   LUC

-124 bp   -54   +1   +28 bp

*Hind*III site at -4
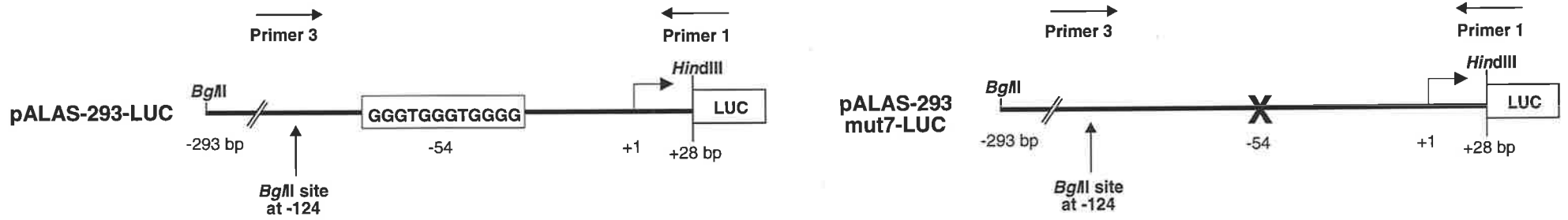
Digest amplified products with *Bgl*I and *Hind*III

and ligate into pGL2-Basic

pALAS-124 -CAP-LUC

*Bgl*I   GGGTGGGTGGGG   *Hind*III   LUC

-124 bp   -54   -4 bp

pALAS-124mut -CAP-LUC

*Bgl*I   X   *Hind*III   LUC

-124 bp   -54   -4 bp

C

pALAS-124-LUC

*Bgl*II   GGGTGGGTGGGG   *Hind*III LUC

-124 bp   -54   +1 +28 bp

mutation at +2/+4

Primer 3   Primer 4

pALAS-124 mut-LUC

*Bgl*II   X   *Hind*III LUC

-124 bp   -54   +1 +28 bp

mutation at +2/+4

Digest amplified products with *Bgl*II and *Hind*III

and ligate into pGL2-Basic

pALAS-124 CAPmut-LUC

*Bgl*II   GGGTGGGTGGGG   *Hind*III X LUC

-124 bp   -54   +1 +28 bp

pALAS-124mut CAPmut-LUC

*Bgl*II   X   *Hind*III X LUC

-124 bp   -54   +1 +28 bp

D

pALAS-124-LUC

GATA-27   Primer 1

*Bgl*I   *Hind*III   LUC

-124 bp   +1   +28 bp

introduce *Pvu*II site at -27

pALAS-124 CAPmut-LUC

GATA-27   Primer 1

*Bgl*I   *Hind*III   LUC

-124 bp   +1   +28 bp

introduce *Pvu*II site at -27

Digest amplified products with *Pvu*II and *Hind*III

and ligate into *Sma*I-*Hind*III digested pGL2-Basic

pALAS-27-LUC

*Sma*I/ *Pvu*II   *Hind*III   LUC

-27 bp   +1   +28 bp

pALAS-27mut-LUC

*Sma*I/ *Pvu*II   *Hind*III   LUC

-27 bp   +1   +28 bp

inserted upstream of the CAT reporter gene (Fig. 2.7). A *Bgl*II-*Hin*dIII fragment was isolated from the plasmids, pALAS-124mut-LUC containing a mutation at the -54 CACCC sequence and pALAS-124CAPmut-LUC containing a mutation at +2/+4, blunted with Klenow enzyme and individually ligated into the *Sma*I linearised pCATOO vector (Fig. 2.7).

2.5 METHODS FOR THE *IN VITRO* DIFFERENTIATION OF ERYTHROID CELL LINES

2.5.1 Cell Maintenance

The human erythroleukemia cell line, K562 and the murine erythroleukemia MEL DS-19 cell line were both maintained in RPMI 1640 medium containing 10% fetal calf serum. The murine erythroleukemia cell line, MELC 707T was maintained in Ham's F12 medium supplemented with 10% FCS. The murine erythroleukemia MEL F4-12B2 (kindly provided by Dr. G. Bergholz, Hamburg, Germany), and J2E cell lines, COS-1 and CV-1 cells were all maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum.

All cells were routinely maintained in 175cm^2 flasks (Falcon) at 37°C in an atmosphere of 5% CO_2 and subcultured every 2-3 days. The COS-1 cell line was grown in 1700cm^2 roller bottles (Corning) at 37°C in the absence of CO_2, to obtain large quantities of cells required for transient transfection experiments. To subculture the suspension K562, MEL DS-19, MELC 707T or J2E cell lines, cells in approximately 25ml of culture media were pelleted by centrifugation at 1200 rpm for 5 minutes, then resuspended into fresh media. To subculture the adherent MEL F4-12B2 cell line, the culture media was removed and cells washed once in PBS before the addition of 2ml of trypsin solution (0.1% trypsin, 0.02% EDTA). Cells were left to detach at room temperature for approximately one minute, resuspended in 10ml of culture media and then seeded into fresh media. The parental CV-1 and the derived COS-1 cell lines were passaged as described for the MEL F4-12B2 cell line except that these cells were washed twice with PBS and then trypsinised for 4-5 minutes. All cells were routinely subcultured by splitting approximately 1:4 into fresh media and incubated at 37°C.

Figure 2.7   Map of the human ALAS2 promoter-pCATOO plasmids

Two ALAS2-CAT expression plasmids were generated, containing -124 bp to +28 bp of ALAS2 promoter sequence in the plasmid, pCATOO. The location of the introduced mutations (*cross*) and the synthetic polyA site are shown. The details of the cloning procedures employed in the generation of these two constructs are outlined in the text.

Smal/
Bgll

hALAS2 promoter

X
-54

-124 bp

poly A

HindIIII/
Smal

+28bp

pALAS-124mut-CAT

CAT

SV40 small T Intron
poly A site

Smal/
Bgll

hALAS2 promoter

HindIIII/
Smal

X
+2

-124 bp

+28bp

poly A

pALAS-124CAP
mut-CAT

CAT

SV40 small T Intron
poly A site
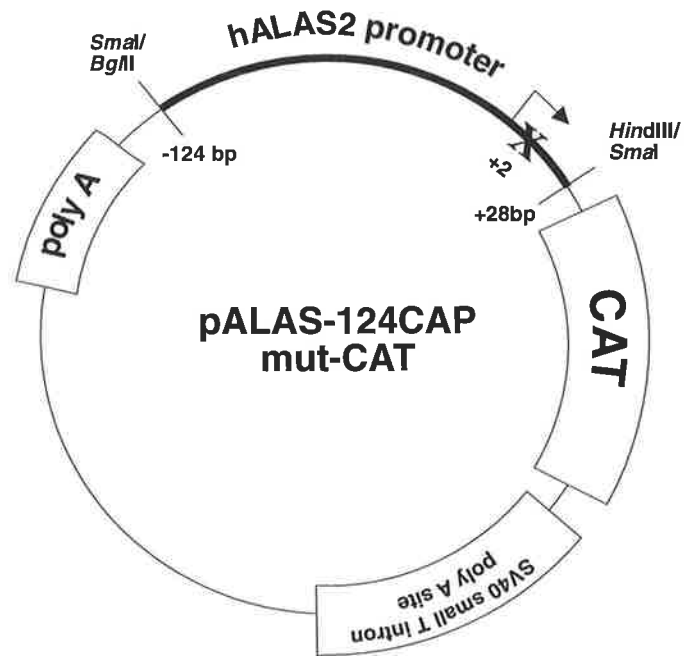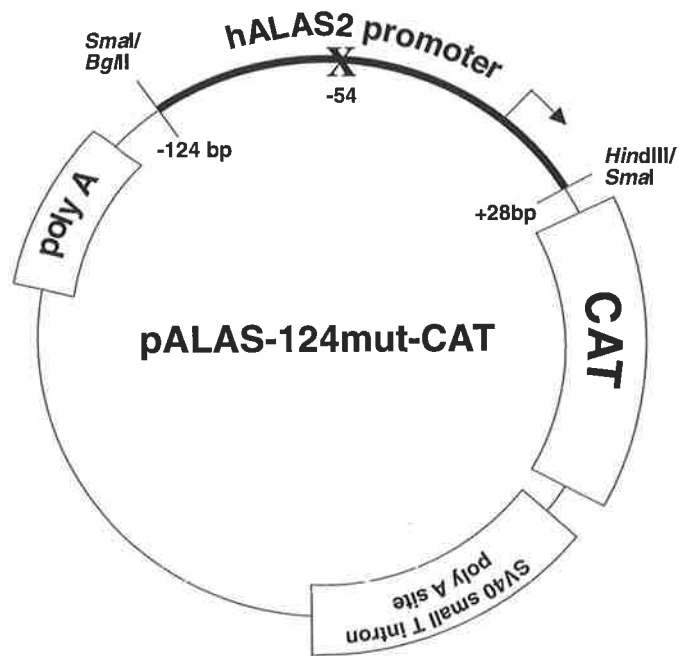
2.5.2 *In vitro* Differentiation of Erythroid Cell Lines

i) *Dimethyl sulfoxide (Me2SO)-induced differentiation of MEL cell lines*

MEL cell lines were chemically induced to differentiate with Me_2SO (Sigma). The DS-19 cell line was seeded at a density of 2.5×10^5 cells/ml in RPMI medium supplemented with 10% FCS containing 2% Me_2SO, and grown in culture for 3 days (Sassa, 1983). MELC 707T cells (5×10^5/ml) in Ham's F12 medium supplemented with 10% FCS and 1.9% Me_2SO was induced to differentiate over a period of 6 days (Friend *et al.*, 1971). The adherent MEL F4-12B2 cell line (5×10^5 cells/ml) in DMEM medium supplemented with 10% FCS was induced to differentiate by culturing with 1.86% Me_2SO for 6 days (Ostertag *et al.*, 1972).

ii) *Hemin-induced differentiation of the K562 cell line*

The K562 cell line was induced to differentiate with hemin using a 1mM stock of freshly prepared hemin. 1mM hemin was prepared by dissolving hemin in 0.25% calcium carbonate and adjusting the pH to 7.4, by the addition of 1M HCl. This solution was then sterilised by filtration through a 0.2μm filter. The suspension K562 cell line was seeded at a density of approximately 1×10^6 cells/ml in RPMI medium supplemented with 10% FCS containing 20μM hemin and maintained in this culture for 3 days (Weber-Benarous *et al.*, 1988).

iii) *Staining of cells with 2,7 diaminofluorene (DAF), an indicator of hemoglobin synthesis*

To stain for hemoglobin positive cells, the diaminofluorene staining procedure was employed (Kaiho & Mizuno, 1985; Worthington *et al.*, 1987). A stock solution of diaminofluorene was prepared by dissolving 50mg of diaminofluorene in 5ml of 90% acetic acid and stored at 4°C. To 5ml of 200mM Tris-HCl, pH 7.6, 100μl of DAF solution and 100μl of hydrogen peroxide were mixed together until the solution turned pale blue and then 100μl of this solution was mixed with 100μl of cell suspension, incubated at room temperature for approximately 5 minutes and the number of blue stained cells counted.

2.6   METHODS FOR EXPRESSION OF PROMOTER CONSTRUCTS IN TISSUE CULTURE CELL LINES

A number of parameters identified in previous studies performed by Dr. G. Braidotti (1992) and shown to influence the transfection efficiency of mammalian cells included the passage number and cell density. Consequently, cell lines were maintained in culture for a maximum of five to seven passages. Adherent cell lines including MEL F4-12B2 and COS-1, were electroporated when the cell monolayers were approximately 70-80% confluent. Exponentially growing cultures of the suspension K562 cell line were achieved by seeding the culture at a density of approximately 5×10^5 cells/ml into fresh medium, 18-24 hours prior to transfection. Cells used for electroporation, were harvested as described in Section 2.5.1, washed twice in 10ml of PBS, pelleted by centrifugation and resuspended in the appropriate electroporation buffer.

All plasmid DNA utilised in these transfection studies was subjected to one round of purification by the cesium-chloride/ethidium bromide equilibrium density gradient procedure and quantified by spectrophotometry as described in Section 2.2.2. The plasmid DNA was analysed by agarose gel electrophoresis to determine concentration, supercoiling and RNA contamination. One round of cesium-chloride density gradient purification was sufficient to remove essentially all of the RNA.

As an internal control for tranfection efficiency, a β-galactosidase expression plasmid, pRSV-β-gal, containing the β-galactosidase gene under the control of the Rous sarcoma viral promoter was co-transfected with each of the plasmid constructs. Optimal β-galactosidase activity was obtained in K562 and COS-1 cell lines co-transfected with 5 μg of pRSV-βgal and the MEL F4-12B2 cell line with 10μg of this vector (data not shown).

2.6.1 Transfection of Erythroid MEL Cell Lines by Electroporation

i)   *Transient transfection of the erythroid MEL F4-12B2 cell line*

For electroporation, the MEL F4-12B2 cell line was grown to 80% confluency, harvested by trypsinisation, resuspended in media and washed twice in PBS as described in Section 2.5.1. MEL cells (10^7) resuspended in 500μl of cold PBS containing 10mM Hepes, pH 7.5, 2 pmol of reporter construct, 10μg of pRSV-βgal and 250μg of sonicated salmon sperm DNA were added to a cuvette, gently mixed and incubated on ice for 10 minutes. These cells were then exposed to a single 300-volt pulse at a capacitance of 960 microfarads using the Bio-Rad Gene Pulser and incubated on ice for a further 10 minutes before seeding into petri dishes (60 x 15 mm) containing 5ml of DMEM supplemented with 10% FCS.

Cells were routinely harvested at either 24 or 48 hours following transfection and assayed for luciferase or chloramphenicol acetyltransferase (CAT) activities, respectively. The transfected cells were washed once in PBS before resuspending in the appropriate buffer. As an internal control for transfection efficiency, β-galactosidase activity was determined at the same time as the luciferase or CAT activity.

ii)   *Stable transfection of the erythroid MELC 707T cell line*

For the stable transfection of MELC 707T cells, cells were harvested as described in Section 2.5.1. MEL cells (1×10^8) were resuspended in 500μl of cold Hepes buffered saline (HBS) (20mM Hepes, pH 7.05, 137mM NaCl, 5mM KCl, 700μM Na_2HPO_4 and 60mM dextrose), 10μg of linearised construct, 1μg of the linearised selectable plasmid DNA and 250μg of sonicated salmon sperm DNA were added to a cuvette, gently mixed and incubated on ice for 10 minutes. These cells were then exposed to a single 400-volt pulse at a capacitance of 250 microfarads, incubated on ice for a further 10 minutes and resuspended in Ham's F12 supplemented with 10% FCS at ~ 1×10^6 cells/ml. These cells were then seeded at ~1×10^5 cells/well in a 96 well tray in a volume of 50μl. 48 hours following transfection, 50μl of Ham's F12 10% FCS supplemented with 400μg/ml of G418 was added to a obtain a final concentration of 200μg/ml. Selection with G418 (200μg/ml) was

maintained for 10 days in which all of the mock-transfected cells were killed. Pools of G418 resistant colonies were seeded into 96 well microtitre trays containing a feeder layer of adherent STO cells and serially diluted 1/3 to obtain clonal cell lines. Feeder cells were pre-treated with the chemical mitomycin C and seeded at a density of 1×10^4 cells/well (J. Beale, personal communication). The G418 resistant colonies were grown on these feeder cells for 7-10 days, clonal cell lines obtained and aliquots frozen at -80°C.

2.6.2 Transient Transfection of the K562 Cell Line by Electroporation

Electroporation of the K562 cell line was performed as described by Max-Audit *et al.* (1993). Exponentially growing cells, were harvested and washed in PBS as described in Section 2.5.1. 10^7 cells in 200µl of PBS containing 10mM Hepes, pH 7.5 were mixed with 2 pmol of reporter construct, 5µg of pRSV-βgal and 250µg of sonicated salmon sperm DNA and incubated at room temperature for 10 minutes then exposed to a single 200-volt pulse at 960 microfarads. Following electroporation, cells were left at room temperature for an additional 10 minutes before seeding into petri dishes (60x15mm) containing 5ml of RPMI medium supplemented with 10% FCS.

Cells were routinely harvested at 24 hours following transfection, washed once in PBS before resuspending cells in 100µl of the appropriate buffer.

2.6.3 Transient Transfection of the COS-1 or CV-1 Cell Lines by Electroporation

The electroporation of the African green monkey kidney fibroblast COS-1 or CV-1 cells were performed by a modification of the method described by Chu *et al.* (1987) and was optimised by J. Beale (personal communication). The cells were harvested by trypsinisation, and washed twice in PBS as described in Section 2.5.1. COS-1 or CV-1 cells (5×10^6) were resuspended in 500µl of cold Hepes buffered saline (HBS) (20mM Hepes, pH 7.05, 137mM NaCl, 5mM KCl, 700µM Na_2HPO_4 and 60mM dextrose) were incubated on ice for 10 minutes with 2 pmol of the reporter construct, 5µg of pRSV-βgal, 250µg of sonicated salmon sperm DNA and electroporated at a single 300-volt pulse at 960 microfarads. Following transfection, these cells were incubated on ice for a further 10 minutes and then seeded into petri dishes (60x15mm) containing 5ml of DMEM

supplemented with 10% FCS. Cells were harvested 24 hours following electroporation using a rubber policeman and washed once in PBS before resuspending cells in the appropriate buffer.

2.6.4 Reporter Gene Assays

Following electroporation of cells with luciferase reporter gene constructs, transfected cells were harvested, washed once in PBS and treated with 100µl of 1x Cell Culture Lysis Reagent (Promega) on ice for 10 minutes. Cells were then snap frozen, thawed on ice, centrifuged for 5 minutes to remove cellular debris, and the supernatants assayed to determine total protein concentration, luciferase activity and β-galactosidase activity.

Cells transfected with chloramphenicol acetyltransferase (CAT) reporter gene constructs were harvested, washed once in PBS and resuspended in 100µl of 250mM Tris-HCl, pH 7.6. The cells were lysed by three cycles of freeze-thawing and the lysate was centrifuged for 5 minutes to remove cellular debris. The supernatant was then assayed to determine total protein concentration and chloramphenicol acetyltransferase (CAT) activity.

i) *Bradford Protein Assay*

The protein content of the cell lysates was determined using 1µl of cell extract in the Bio-Rad protein microassay procedure according to the manufacturer's instructions. Bovine serum albumin was used as the standard. Subsequent luciferase and CAT assays were performed with 100µg of cell lysate.

ii) *Assay for Luciferase Activity*

Luciferase activity was measured using a luciferase assay system (Promega) and measurements determined in a Berthold model LB9502 luminometer. A 100µg sample of cell lysate was added into a small test tube and placed in the luminometer. 300µl of luciferase reaction buffer (39mM glycylglycine, 23.3mM $MgSO_4$, 7.8mM rATP) and 300µl of 111µM Beetle luciferin were then injected, and the luciferase activity measured.

iii) *Assay for β-galactosidase Activity*

β-Galactosidase activity was assayed as described by the method of Herbomel *et al.* (1984). A constant amount of protein (25μg) was added to 500μl of reaction buffer (100mM $NaPO_4$, pH 7.3, 10mM KCl, 1mM $MgCl_2$, 50mM β-mercaptoethanol) and the reaction initiated with the addition of 100μl of 2mg/ml ONPG. The reaction was incubated at 37°C until a yellow shade was obvious (20-60 minutes) and the reaction terminated with the addition of 1ml 1M Na_2CO_3. The absorbance at 420nm was determined and β-galactosidase activity expressed as (A_{420}nm/μg of protein) x 100 extrapolated to 1 hour of reaction.

iv) *Assay for Chloramphenicol Acetyltransferase (CAT) Activity*

Prior to the CAT assay, the cell lysate was incubated at 65°C for 5 minutes to remove deacetylase activity (Adrisani *et al.*, 1987). CAT activity was then assayed as described by Gorman *et al.* (1982). To 65μl of 250mM Tris-HCl, pH 7.6 containing 100μg of protein was added 5μl of 10mM acetyl-coenzyme A, 2.5μl of [14C] chloramphenicol and 17.5μl of water. The reaction mixture was incubated at 37°C for 2 hours. The reaction was terminated following the addition of 1ml of ethyl acetate, the solution vortexed to extract the chloramphenicol, the upper, organic phase was transferred to a clean Eppendorf tube and the ethyl acetate was evaporated. The residue was dissolved in 15μl of ethyl acetate and spotted onto silica plates (Merck). Acetylated [14C] chloramphenicol was resolved by thin layer chromatography in a solvent of chloroform:methanol (9.5:0.5). The silica plates were air dried and autoradiographed at -80°C for 24-48 hours. After autoradiography, the spots corresponding to acetylated chloramphenicol were cut out and the amount of radioactivity quantitated by liquid scintillation counting or by densitometry.

2.7   METHODS FOR ISOLATION AND ANALYSIS OF RNA

2.7.1   Isolation of RNA from Tissue Culture Cells

Total RNA was extracted from tissue culture cells using the guanidinium isothiocyanate procedure described by Chomczynski & Sacchi, (1987). Following RNA extraction, the absorbance values at 260nm and 280nm of each RNA sample were determined on a Shimadzu UV-160A spectrophotometer. The A_{260}/A_{280} ratio of the RNA samples were consistently in the range 1.8-2.0. The relationship of one A_{260} unit is equal to 40µg/ml RNA was used to determine the concentration of RNA.

2.7.2   Northern Hybridisation Analysis of RNA

i)   *Northern Blots*

RNA for Northern blots (20µg) was denatured in 10mM sodium phosphate, pH 7.4, 50% formamide, 2.2M formaldehyde, 0.5mM EDTA at 65°C for 5 minutes, prior to electrophoresis on 1% agarose gels containing 1.1M formaldehyde and 10mM sodium phosphate, pH 7.4. The RNA was transferred onto Nytran[a] filters (Schleicher & Schuell) using capillary transfer with 20xSSPE. Following transfer, the RNAs were covalently cross-linked to the filters by exposure to 120mJoules of UV radiation in a Stratagene UV Stratalinker[a] 1800 (see manufacturer's instruction manual). Filters were pre-hybridised for 4-16 hours at 42°C in 50% formamide, 5x SSPE, 5x Denhardt's solution, 0.1% SDS, 0.05% sodium pyrophosphate and 200µg/ml of sonicated salmon sperm DNA in roller bottles. Hybridisations were carried out for 18-24 hours under the same conditions, except for the addition of radiolabelled probe (1-5 x 10^8 cpm/µg). Filters were washed twice in 1xSSPE, 0.1% SDS and 0.05% sodium pyrophosphate at room temperature for 10 minutes each, followed by one wash in 0.5xSSPE, 0.1% SDS and 0.05% sodium pyrophosphate at 65°C for 20-30 minutes.

ii)   *Removal of Probe from Filter*

Northern Blots were stripped of probe for re-use by washing blots in 5mM Tris-HCl, pH 8.0, 0.2mM EDTA, pH 8.0, 0.05% sodium pyrophosphate, 0.1% SDS and

0.1x Denhardt's solution for 2 hours at 65-67°C. Filters were then rinsed in 2xSSPE and either re-used immediately or stored at 4°C between clean pieces of 3MM Whatman paper.

2.8   SOUTHERN BLOT ANALYSIS AND HYBRIDISATION CONDITIONS

2.8.1   Isolation of Genomic DNA from Tissue Culture Cells

4×10^7 cells were harvested, washed once in PBS and resuspended in 700μl of 25mM Tris-HCl, pH 8.0, 75mM NaCl and 24mM EDTA, pH 8.0. 0.1 volume of 10% SDS was added and gently mixed by inversion to lyse the cells and nuclei. 200μg of proteinase K was added and incubated at 37°C for 2.5 hours with occasional inversions. The digestion was then extracted by inversion for 5 minutes, three times with an equal volume of TE-buffered phenol and then twice with an equal volume of chloroform:isoamyl alcohol (24:1). Genomic DNA was precipitated with 0.1 volume of cold 2M KCl and 2 volumes of cold 100% ethanol, carefully washed with 70% ethanol and resuspended in 500μl of TE at 4°C overnight.

2.8.2   Southern Blot Analysis

20μg of genomic DNA was digested with the appropriate restriction enzymes and electrophoresed on 0.8% agarose gels in 1xTBE. Following staining with ethidium bromide, the gels were visualised under UV light and photographed. Gels were treated with 0.25M HCl for 10 minutes to partially depurinate the genomic DNA, prior to the denaturation and neutralisation steps. Denaturation, neutralisation and capillary transfer of the DNA to Nytran^a filters (Schleicher & Schuell) was performed as described in the manufacturer's protocol, except that the transfer was carried out in 10xSSPE. Following transfer, the DNAs were covalently cross-linked to the filters (as described in Section 2.7.2). The filters were pre-hybridised for 4-16 hours at 42°C in 50% formamide, 5xSSPE, 5xDenhardt's solution, 0.1% SDS, 0.05% sodium pyrophosphate and 200μg/ml of sonicated salmon sperm DNA in roller bottles. Hybridisations were performed for 18-24 hours under the same conditions, except for the addition of radiolabelled probe (1-5 x 10^8

cpm/µg). Filters were washed once in 2xSSC and 0.1% SDS at room temperature for 10 minutes and then once at 65°C for 10-20 minutes.

2.9   METHODS FOR GEL SHIFT ASSAYS

2.9.1   Preparation of Nuclear Protein Extracts

The procedure of Partington *et al.* (in press) was used to prepare nuclear extracts from K562, MEL F4-12B2 and COS-1 cells. Nuclear extracts were also prepared from COS-1 cells were transfected with 10µg of the murine GATA-1 cDNA expression clone, pXM/GF-1 or pMT2/RINFE and pMT2/p18w-1, encoding the erythroid-specific subunit of NF-E2, p45 (Andrews *et al.*, 1993a) and its ubiquitously expressed partner, p18 (Andrews *et al.*, 1993b), respectively.

Cells (4-8x10^7) were harvested and washed twice in PBS and then washed three times in PBS containing protease inhibitors (0.5mM PMSF, 0.5mM benzamidine, 10mM β -glycerophosphate, 2mM levamisole). Cells were resuspended in 10ml of lysis buffer (2mM MgCl$_2$, 20mM Hepes, pH 7.5, 0.5mM PMSF, 0.5mM benzamidine, 10mM β -glycerophosphate, 2mM levamisole, 1µg/ml leupeptin, 1µg/ml aprotinin, 1µg/ml pepstatin, 5µg/ml bestatin) containing 0.05% Triton X-100, gently mixed and incubated on ice for 30 minutes. Nuclei were harvested by centrifugation at 2000 rpm at 4°C for 5 minutes, washed in lysis buffer without Triton X-100 and then resuspended in 2 volumes of this buffer. 4M NaCl was then added dropwise to a final concentration of 350mM and gently mixed. Chromatin was pelleted by centrifugation at 12000 rpm at 4°C for 15 minutes. The supernatant was transferred to a clean Eppendorf and glycerol added to a final concentration of 20% and stored at -80°C in 50µl aliquots.

For the detection of CACCC binding proteins, a rapid procedure described by Andrews and Faller, (1991), and used by Crossley *et al.* (1996) was employed to prepare nuclear extracts from K562, MEL and CV-1 cells and CV-1 cells transfected with 20µg of the murine EKLF cDNA expression clone, pSG5/EKLF.

2.9.2 Preparation of Radiolabelled Annealed Oligonucleotide Probes

Single-stranded oligonucleotides were gel purified and 100ng labelled with [γ ^{32}P]-ATP using T4 polynucleotide kinase as described in Section 2.2.10. A three molar excess of the unlabelled complementary oligonucleotide was annealed to the [^{32}P]-labelled oligonucleotide in 100mM NaCl by incubation at 100°C for 2 minutes, 70°C for 10 minutes and then allowing the samples to slowly cool to room temperature. The [^{32}P]-labelled annealed oligonucleotides were precipitated and washed to remove unincorporated radioactivity and resuspended in 100μl of TE.

Unlabelled oligonucleotides were also annealed for use in competition assays. 1μg of each sense and anti-sense oligonucleotide in 100mM NaCl in a volume of 100μl, were annealed as described above.

2.9.3 Gel Shift Assay

Binding reactions used in the detection of GATA and NF-E2 binding proteins contained 5μg of nuclear protein, 2μg poly(dI-dC) in 15μl of 25mM Hepes, pH 7.9 containing 60mM KCl, 7.5% glycerol, 0.1mM EDTA, 5mM MgCl$_2$, 0.75mM dithiothreitol and 2mM spermidine and incubated on ice for 10 minutes. Radiolabelled probe (1ng) was added to the reaction and incubated on ice for a further 30 minutes. In supershift assays, 2μl of the GATA-1 specific monoclonal antibody, N-6 (Ito *et al.*, 1993), (provided by Dr. G. Partington) was incubated in the binding reaction prior to the addition of probe. Retarded nuclear protein complexes were resolved on a pre-electrophoresed 6% non-denaturing polyacrylamide gel in 0.25xTBE at 180 V for 2.5 hours at 4°C.

Binding reactions used for the detection of GATA binding proteins was also employed for the detection of the thyroid hormone receptor. For the optimal binding of retarded complexes, the binding reaction was incubated at 25°C rather than on ice.

For the detection of protein binding to the -27 GATA site, the binding reaction protocol described by Fong and Emerson (1992) was used. 1μl (20ng) of purified recombinant human TBP (Promega) was used in the binding reaction and poly(dI-dC) was omitted. The reaction was incubated at 25°C for 30 minutes and retarded nuclear protein complexes were resolved on a pre-electrophoresed 5% non-denaturing polyacrylamide gel

(containing 0.05% NP40, 5% glycerol, 0.5mM dithiothreitol and 1mM EDTA, pH 7.0) in 0.5xTBE at 250 V for 2 hours at 4°C.

In experiments designed to determine the binding affinity constants (K_d) of GATA-1 and TBP, binding reactions and electrophoresis conditions were performed as described above with a constant amount of radiolabelled oligonucleotide probes and serial dilutions of TBP or a purified GST-GATA-1 zinc finger fusion protein (GST-GATA-1(f)) prepared as described by Smith and Johnson (1988).

Binding reactions used in the detection of CACCC binding proteins and the supershift assays using polyclonal antibodies to BKLF, EKLF and Sp1 were performed as described by Crossley et al. (1996). The polyclonal antibodies to BKLF and EKLF were generously provided by Dr. M. Crossley. The Sp1 polyclonal antibody; Sp1 (PEP 2) (Santa Cruz Biotechnology, Inc.), was a gift from Dr. M. Shannon.

For the detection of ets binding in gel shift assays, the binding protocol reported by Higashino et al. (1993) was employed. Briefly, the binding reaction consisted of 5µg of nuclear protein, 2µg of poly(dI-dC) in 20mM Hepes, pH 7.9 containing 50mM NaCl, 0.5mM EDTA, 5% glycerol, 1mM dithiothreitol and 1ng of probe. The binding reaction was incubated on ice for 15 minutes and following the addition of probe incubated at 25°C for 30 minutes. The retarded nuclear protein complexes were resolved on a pre-electrophoresed 4% non-denaturing polyacrylamide gel in 0.25xTBE buffer.

Gel shift competition assays were performed with unlabelled competitor oligonucleotides included in the binding reactions. Following electrophoresis, the gels were transferred to Whatman 3MM paper, dried and exposed to Kodak X-Omat AR film.

2.10 MISCELLANEOUS METHODS

2.10.1 Densitometric Quantitation of Bands on Autoradiograms

Quantitation of bands on autoradiograms was performed on a Molecular Dynamics computing laser densitometer, using their ImageQuant data analysis software. Exposure times were adjusted so that the signals were within the linear range of the film used. In addition to autoradiography, radioactivity was also detected by using storage Phosphor

2.10.2 Computer Analysis

The computer software package "GCG Sequence Analysis Software Package" (Devereux *et al.*, 1984), and in particular the program, "FIND" was used to identify putative transcription factor binding sites in the human ALAS2 gene sequence.

CHAPTER THREE

EVALUATION OF ERYTHROID CELL LINES SUITABLE

FOR TRANSIENT TRANSFECTION STUDIES AND

SYNTHESIS OF A STABLY TRANSFECTED MEL CELL LINE

CHAPTER THREE: EVALUATION OF ERYTHROID CELL LINES SUITABLE FOR TRANSIENT TRANSFECTION STUDIES AND SYNTHESIS OF A STABLY TRANSFECTED MEL CELL LINE

3.1   INTRODUCTION

As described in Section 1.13, the development of transformed erythroid cell lines such as MEL, J2E and K562 cells has provided the opportunity to investigate the regulatory processes which occur during the latter stages of erythroid differentiation. Transcriptional activation of the genes encoding the enzymes of the heme biosynthetic pathway must be coordinated with that of the α- and β- globins for the efficient synthesis of hemoglobin. Following exposure of MEL cells to Me_2SO or J2E cells to Epo (Klinken *et al.*, 1988), the levels of ALAS2 mRNA (Elferink *et al.*, 1988; Fujita *et al.*, 1991; Busfield *et al.*, 1993) and transferrin receptor (TfR) mRNA (Chan *et al.*, 1994; Busfield *et al.*, 1993) are markedly increased, in parallel to the dramatic increase in globin gene transcription (Ross *et al.*, 1972; Nudel *et al.*, 1977; Orkin *et al.*, 1977; Busfield *et al.*, 1993). These erythroid cell lines thus provide an abundant source of material that can be readily manipulated for molecular expression studies.

The primary objective of this thesis is to identify regions of the human ALAS2 gene required for transcriptional regulation in erythroid cells. To this end, several MEL cell lines have been investigated for their suitability in transient transfection experiments and these studies are described in this chapter. Initial studies with a luciferase reporter gene construct demonstrated the lack of suitability of two erythroid cell lines; the suspension cell line, DS-19, and the semi-adherent cell line, MELC 707T. Therefore a Me_2SO-responsive stably transfected cell line was generated in the MELC 707T cell line, in which 1.9 kb of the 5'-flanking region of the ALAS2 gene was fused to the chloramphenicol acetyltransferase reporter gene. Such an approach would permit expression of this reporter gene to be investigated during erythroid differentiation of this stable cell line and these studies have been included in this chapter. However, while these studies were in progress, an adherent cell line, MEL F4-12B2, was obtained and luciferase activity was readily detected in transient expression studies performed in this cell line. In addition to the successful transient

expression of this cell line, expression of a luciferase reporter gene construct was also examined in the erythroid J2E and K562 cell lines.

3.2   RESULTS

3.2.1   Northern Blot Analysis of MEL Cell Lines Chemically Induced to Differentiate *in vitro*

To ensure that the MEL cell lines, suspension DS-19 (Ohta *et al.*, 1976) and the semi-adherent MELC 707T (Harrison *et al.*, 1978), were indeed suitable for expression studies, the levels of endogenous ALAS2, α- and β- globin mRNAs during erythroid differentiation were examined. The conditions for the induction of erythroid differentiation in these MEL cell lines with the chemical inducer, Me_2SO have been previously reported (Sassa, 1983; Friend *et al.*, 1971). Me_2SO-induced erythroid cells were stained with diaminofluorene (as described in Section 2.5.2) to determine the number of cells synthesising hemoglobin and total RNA was isolated for Northern blot analysis. Total RNA ($20\mu g$) was processed for Northern blotting as described in Section 2.7 and the same filter was subsequently stripped and reprobed for the different mRNAs.

As seen in Fig. 3.1A, the proportion of diaminofluorene positive cells prior to treatment of the DS-19 cells with Me_2SO was approximately 8%, and as expected this rapidly increased to 96% of the cells after 3 days in the presence of the chemical inducer, indicating that an extremely high proportion of the cells were synthesising hemoglobin. In contrast, the proportion of diaminofluorene positive cells in the MELC 707T cell line increased more slowly than the DS-19 cells from approximately 1.3% in uninduced cells to 89% at day 7 (Fig. 3.1A).

The expression of transcripts encoding ALAS2, the α- and β- globins and GAPDH was determined by Northern blot analysis. As seen in Fig. 3.1B, endogenous ALAS2, α- and β-globin mRNA levels in the DS-19 cell line, exhibited a significant increase (approximately 30-50 fold) in Me_2SO-induced cells at day 3 compared with uninduced cells (day 0). The level of ALAS2 mRNA in the Me_2SO-induced MELC 707T cell line was significantly increased by day 1 post-induction, whereas transcripts for the α- and β-globins were detected at approximately day 3 and day 2 post-induction, respectively (Fig. 3.1B).

Figure 3.1   Me$_2$SO-induced differentiation of the MEL DS-19 and MELC 707T cell lines

The DS-19 cell line (2.5x10^5 cells/ml) was induced to differentiate with 2% Me$_2$SO for 3 days and the MELC 707T cell line (5x10^5 cells/ml) was induced to differentiate with 1.9% Me$_2$SO for 7 days. During the time course experiments, aliquots of Me$_2$SO-induced MEL cells were isolated daily, stained for hemoglobin synthesis and total RNA isolated for Northern blot analysis. Untreated DS-19 and MELC 707T cells were included as negative controls.

A   Me$_2$SO-induced DS-19 and MELC 707T cells were stained daily with diaminofluorene and positive cells are an indicator of hemoglobin synthesis. The number of hemoglobin positive cells are expressed relative to the total number of cells. The data is represented as the percentage (%) of differentiated cells during the time course experiments.

B   Total RNA was then isolated at the indicated times following the addition of Me$_2$SO to the DS-19 and MELC 707T cell lines and 20μg of total RNA was subjected to Northern blot analysis. The filter was sequentially probed for ALAS2, the α- and β- globins and GAPDH mRNAs.

The highest levels of all three transcripts were observed at approximately day 4. However, the level of GAPDH mRNA declined during erythroid differentiation and was reduced approximately 4-fold (at day 7 post-induction) compared with the level detected in uninduced MEL cells (day 0). The downregulation of the constitutively expressed GAPDH transcript during Me₂SO-induced differentiation which served as a loading control has been previously reported (Fraser & Curtis, 1987). This downregulation of non-erythroid specific mRNAs appears to be a characteristic feature of terminally differentiating erythroid cells (Sherton & Kabat, 1976; Fraser & Curtis, 1987) and therefore GAPDH could not be used as a loading control.

In conclusion, Northern blot analysis of these two MEL cell lines chemically induced to differentiate with Me₂SO demonstrated that both cell lines contain all of the factors necessary for the induction of ALAS2 gene transcription.

3.2.2 Optimisation of Transient Transfection of Erythroid MEL Cell Lines

To identify and investigate the regulatory regions of the human ALAS2 gene responsible for basal and erythroid-specific expression of the gene, it was decided to employ transient expression studies. A number of methods are available to introduce recombinant DNA molecules into mammalian cells *in vitro* (reviewed by Spandidos & Wilkie, 1984) including the co-precipitation of DNA with either calcium phosphate or DEAE-dextran sulphate, and electroporation. The introduction of DNA molecules by electroporation has been successfully used to introduce DNA into both plant and mammalian cells (Chu *et al.*, 1987; Maxwell & Maxwell, 1988). This technique involves the exposure of cultured cells to a pulsed electric field, which presumably creates transient pore formation in the cell membrane and facilitates the uptake of DNA molecules (Chu *et al.*, 1987). A number of parameters can influence the transfection efficiency of plasmid DNA into cells by electroporation and include the quantity of DNA, the salt concentration and temperature of the electroporation buffer, the volume of electroporation buffer, voltage and the addition of carrier DNA such as sonicated salmon sperm DNA (Potter *et al.*, 1984; Chu *et al.*, 1987).

In the first experiments, it was important to optimise the transfection of each of the two MEL cell lines, DS-19 and MELC 707T. The expression of two reporter genes, the firefly luciferase gene and the bacterial chloramphenicol acetyltransferase (CAT) gene were

examined in both cell lines. The plasmids, pRSV-LUC and pRSV-CAT containing either the luciferase or the chloramphenicol acetyltransferase reporter genes, respectively, driven by the Rous sarcoma viral promoter were electroporated into both cell lines and expression compared.

The suspension MEL DS-19 cell line was transiently transfected following a transfection protocol described for a semi-adherent *aprt⁻* MEL cell line (Deisseroth & Hendrick, 1978). The DS-19 cells (2×10^7) were electroporated with either 20µg of pRSV-LUC or pRSV-CAT plasmid DNA and 250µg of sonicated salmon sperm DNA in 500µl of cold Hepes-buffered saline (HBS) containing 20mM Hepes, pH 7.05, 137mM NaCl, 5mM KCl, 0.7mM Na_2HPO_4 and 6mM dextrose at either 380, 400 or 420 volts at 250 microfarads. In these experiments, 20µg of the plasmids, pGL2-Basic (containing the promoterless luciferase gene) or pIBI-CAT H1 (containing the promoterless CAT gene) were included as negative controls. Cells transfected with the luciferase or CAT reporter constructs were harvested at 24 and 48 hours following electroporation, respectively. Luciferase and CAT activities were determined with 100µg of total protein as described in Section 2.6.4. Optimal CAT activity was detected in cells which were electroporated at a single 400-volt pulse at 250 microfarads, but no luciferase activity was detected at any of the three voltages examined (data not shown). The addition of salmon sperm DNA was found to increase the expression of the CAT reporter construct approximately 5-fold (data not shown).

To investigate whether the inability to detect luciferase activity was restricted to the DS-19 cell line, similar transfections were performed with the semi-adherent MELC 707T cell line. Exponentially growing cells were transfected using the same protocol and plasmid constructs (20µg each) as that described for the successful electroporation of the MEL DS-19 cell line. As seen in Fig. 3.2, the percentage (%) of acetylated chloramphenicol was approximately 2.5-fold higher in MELC 707T cells compared with DS-19 cells electroporated with the same amount of plasmid DNA. However, luciferase activity was again not detectable in lysates of cells transfected with the plasmid, pRSV-LUC (data not shown).

In conclusion, transient transfection studies using either the MEL DS-19 or MELC 707T cell lines were found to be not suitable for three reasons: the inability to detect

MEL cell line	DS-19		MELC 707T	
Plasmid	pRSV-CAT	pIBICAT	pRSV-CAT	pIBICAT
% Conversion	3.6	0.6	8.8	0.7
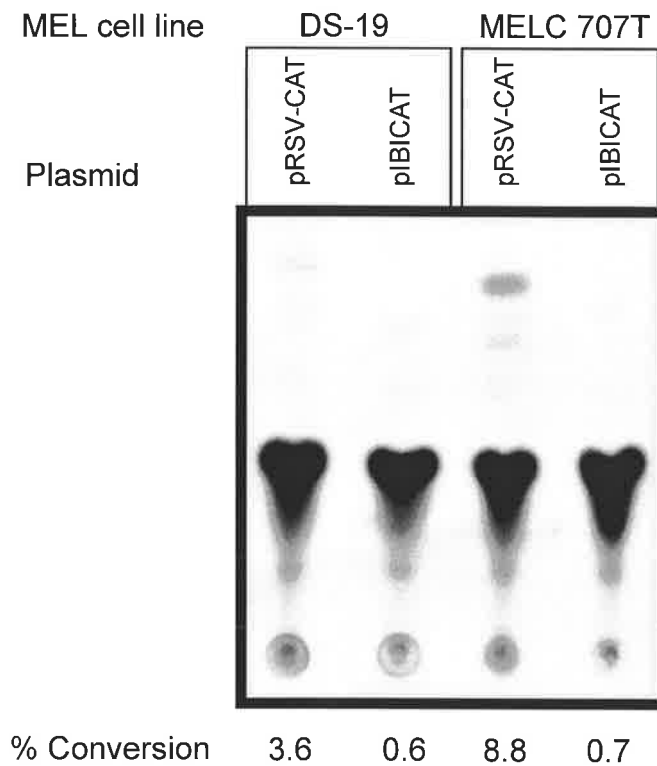
Figure 3.2   Expression of the plasmid pRSV-CAT in transiently transfected MEL DS-19 and MELC 707T cell lines

Exponentially growing MEL DS-19 and MELC 707T cells (2×10^7) were electroporated with 20µg of the plasmid pRSV-CAT and 250µg of sonicated salmon sperm DNA in 500µl of cold buffer containing 20mM Hepes, pH 7.05, 137mM NaCl, 5mM KCl, 0.7mM Na_2HPO_4 and 6mM dextrose at a single 400-volt pulse at 250 microfarads (µF). The promoterless plasmid, pIBI-CAT was included as a negative control. Cells were harvested 48 hours following transfection, the protein concentration of the cell lysate determined and 100µg of total protein assayed for CAT activity.

expression of the luciferase reporter gene, the difficulty in transfecting these cells which would require large quantities of the CAT reporter gene plasmid DNA and time consuming CAT assays performed by thin layer chromatography. An alternative approach, involving the generation of cell lines stably transfected with ALAS2 promoter/CAT reporter gene constructs may overcome some of these problems. In addition, the generation stably transfected cell lines would permit expression of the reporter gene to be examined over a period of days as the cells were induced to differentiate.

3.2.3   Generation of a Stably Transfected MEL Cell Line: Determination of Selection Conditions

To distinguish between transfected cells and non-transfected cells, the bacterial gene encoding aminoglycoside phosphotransferase was employed as a selectable marker. The aminoglycoside antibiotic 3'-deoxystreptamine (G418) normally blocks protein synthesis in mammalian cells by interfering with ribosomal function. However, expression of the aminoglycoside phosphotransferase gene in mammalian cells results in the detoxification of G418. The plasmid, pPGKneoβ containing the aminoglycoside phosphotransferase gene driven by the murine *pgk-1* promoter was generously provided by Dr. L. Whyatt (Fig. 3.3).

Initial experiments were performed to determine the dose of the drug G418 required for the effective selection against the parental MELC 707T cell line over a 7-10 day period. Cells were grown in culture medium containing G418 (ranging from 50μg/ml to 800μg/ml) for a period of 10 days. The optimal concentration of G418 was determined to be 200μg/ml and was used in the selection of stably transfected cells.

3.2.4   Generation of MEL Cell Line Stably Transfected with a 1.9kb ALAS2 Promoter/CAT Reporter Construct

As described in Section 3.2.2, expression of CAT activity in transiently transfected MEL (DS-19 or MELC 707T) cells was extremely low (approximately 3-9%) probably due to the inefficiency of transfecting the cells and the sensitivity of the CAT reporter gene. The creation of a MEL cell line permanently expressing an ALAS2 promoter/CAT reporter gene construct could be expected to overcome these problems and thus permit time course

experiments to be performed with large cell numbers. Since higher CAT activity was observed in transiently transfected MELC 707T cells in comparison to DS-19 cells, the MELC 707T cell line was chosen for the generation of stably transfected cell lines. A 1.9 kb ALAS2 promoter fragment (-1.9 kb to +28 bp) thought to contain the necessary elements required for promoter activity (Fig. 3.3) was ligated upstream of the CAT reporter gene (pALAS-1.9kb-CAT) and stably integrated into the genome of the MELC 707T cell line. The plasmid, pRSV-CAT, which efficiently expresses in transiently transfected MELC 707T cells (Section 3.2.2) was included as a positive control.

To increase the efficiency of plasmid DNA integration into the genome, linearised plasmid DNA was used (Chu *et al.*, 1987). The plasmids pALAS-1.9kb-CAT, pRSV-CAT and pPGKneoβ were linearised with the restriction enzymes *Hind*III, *Bam*HI and *Hind*III, respectively (Fig. 3.3). To obtain a high proportion of selected colonies containing the reporter plasmid, cells were co-transfected with 10μg of the CAT reporter plasmid and 1μg of the selectable gene plasmid, pPGKneoβ.

The MELC 707T cell line was electroporated with 10μg of linearised pALAS-1.9kb-CAT or pRSV-CAT and 1μg of linearised pPGKneoβ as described in Section 2.6.1. A control was included in which the selectable plasmid DNA, pPGKneoβ was omitted. Cells were allowed to recover for 48 hours following transfection and the media supplemented with G418 to a final concentration of 200μg/ml. The non-transfected cells were selected against during the next 10 day period by which time all cells were dead in the control plate. In order to obtain clonal cell lines, G418 resistant cells were serially diluted 1/3 and seeded into 96 well microtitre trays containing a feeder layer of the adherent STO cell line which has been successfully used in the generation of permanent cell lines (J. Beale, personal communication). This feeder layer of cells would provide the nutrients required to support growth of the transfected MELC 707T cells for a period of approximately one week. Clonal cell lines transfected with the pALAS-1.9kb-CAT or pRSV-CAT constructs were initially screened for CAT activity and approximately 80% of the clones expressed acetyl transferase activity (data not shown).

Two independently derived clones, A and B, expressing CAT activity and stably transfected with the pALAS-1.9kb-CAT construct, were chosen for further characterisation. Genomic Southern blot analysis was performed to confirm the integration of the pALAS-1.9kb-CAT construct into the MELC 707T cell line. Genomic DNA was

Figure 3.3   **Restriction map of the plasmids, pALAS-1.9kb-CAT, pRSV-CAT and pPGKneoβ**

The plasmids, pALAS-1.9kb-CAT and pPGKneoβ were linearised with the restriction enzyme, *Hin*dIII and pRSV-CAT with *Bam*HI. The location of restriction sites used to linearise the three plasmids are indicated.

pPGKneoβ

*Hind*III

m *pgk-1* promoter

*Bgl*II

Neo R

β-globin intron

*Sma*I/
*Xba*I

hALAS2 promoter

*Hind*III/
*Sma*I

-1.9 kb

+28bp

pALAS-1.9kb-CAT

7.83 kb

CAT

*Hind*III

*Nru*I

RSV promoter

*Hind*III

pRSV-CAT

4.4 kb

CAT

*Bam*HI
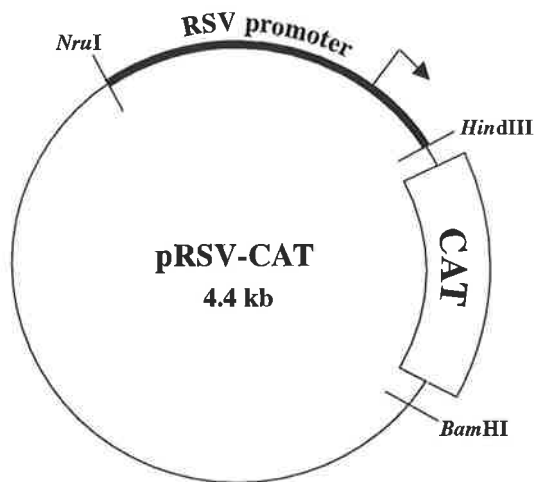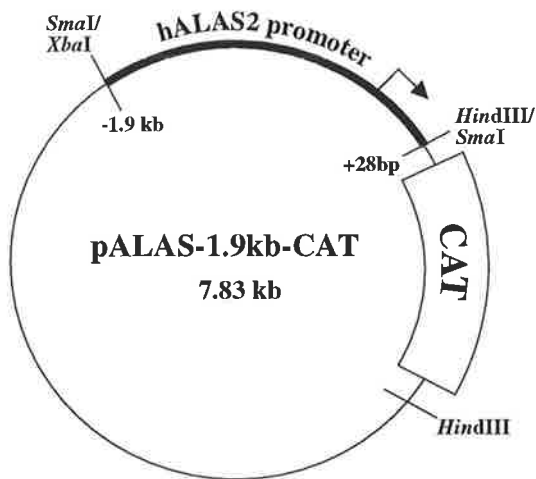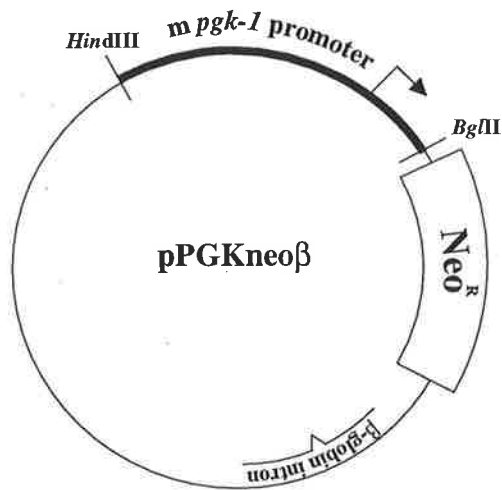
isolated from the parental line and the two stable cell lines, A and B (as described in Section 2.8.1) and approximately 30μg of genomic DNA digested with the restriction enzymes *Eco*R1 or *Sac*I (as described in Section 2.2.3). As a positive control, genomic DNA isolated from the parental cell line was mixed with 10pg of pALAS-1.9kb-CAT plasmid DNA and also digested with these restriction enzymes. The digested genomic DNA was resolved on a 0.8% agarose gel in 1xTBE buffer and transferred to Nytran[a] as described in Section 2.8.2. A *Kpn*I-*Sca*I fragment of approximately 2.7 kb was isolated from the plasmid, pALAS-1.9kb-CAT, and contained the 1.9 kb ALAS2 promoter and approximately 800 bp of the CAT gene (Fig. 3.4A). This fragment was radiolabelled as described in Section 2.10 and used to probe the genomic Southern blot. The results are shown in Fig. 3.4B. Bands of approximately 2.2 kb and 1.5 kb were detected in DNA from both clonal lines digested with *Eco*RI and *Sac*I, respectively. The 2.2 kb *Eco*RI fragment corresponded to the 1.9 kb ALAS2 promoter and 284 bp of the CAT gene. The 1.5 kb *Sac*I hybridising fragment corresponded to the ALAS2 promoter (Fig. 3.4B). In conclusion, clones A and B both contained 1.9 kb of the human ALAS2 promoter ligated upstream of the CAT reporter gene stably integrated into the MELC 707T cell line.

To show that clones A and B, could still be induced to differentiate with Me_2SO, a time course experiment was performed with these two clones and the parental cell line as described in Section 3.2.2. As seen in Fig. 3.5A, the proportion of diaminofluorene positive cells in clone A progressively increased during the 6 day period in a pattern extremely similar in timing and the number of hemoglobin synthesising cells to that observed with the parental cell line. In contrast, clone B cells showed a lag in hemoglobin synthesis with greater than background staining only observed at day 3 post-induction and thereafter. Even after 6 days, only 62% of the clone B cells stained positive for hemoglobin synthesis compared with 95% observed for clone A and parental. (Fig. 3.5A).

The expression of endogenous ALAS2 and β-globin mRNAs in clones A and B were then examined by Northern blot (as described in Section 2.7). Northern blots were initially probed with the radiolabelled human ALAS2 cDNA, stripped and reprobed with the radiolabelled murine β-globin cDNA. As seen in Fig. 3.5B, the timing and level of expression of both ALAS2 and β-globin mRNAs in clone A cells was extremely similar to the parental cell line. Exposure of the clone B cells to Me_2SO resulted in increased levels of β-globin mRNA but this accumulation was delayed compared to the parental cell line

Figure 3.4   Southern blot analysis of genomic DNA isolated from two MELC 707T cell lines stably transfected with the pALAS-1.9kb-CAT construct

Genomic DNA was isolated from the parental MELC 707T cell line and two independently derived clones A and B, stably transfected with pALAS-1.9kb-CAT. Genomic DNA was digested with either *Eco*RI or *Sac*I, electrophoresed on a 0.8 % agarose/1xTBE gel and the DNA was transferred to Nytran[a] as described in Section 2.8.2. A ^{32}P-labelled *Kpn*I-*Sca*I fragment isolated from pALAS-1.9kb-CAT was used as the probe. Hybridisation was performed as described in Section 2.8.2. The filter was washed once in 2xSSC and 0.1% SDS at room temperature for 10 minutes, then once in 2xSSC and 0.1% SDS at 65°C for 10 minutes.

A   Restriction map of the *Hin*dIII linearised pALAS-1.9kb-CAT construct. Above the map is shown the probe used in the Southern blot analysis of clone A and B. The expected size of the fragments detected with the *Kpn*I-*Sca*I probe (approximately 2.7 kb), following digestion of the plasmid pALAS-1.9kb-CAT with *Eco*RI or *Sac*I are indicated.

B   Autoradiograph of the filter hybridised with the 2.7 kb ^{32}P-labelled *Kpn*I-*Sca*I fragment. Genomic DNA isolated from the parental MEL cell line (lanes 1 and 5), and clones A (lanes 3 and 7) and B (lanes 4 and 8) were digested with *Eco*RI and *Sac*I. As a positive control, genomic DNA isolated from the parental cell line was mixed with 10pg of pALAS-1.9kb-CAT (lanes 2 and 6) and digested with *Eco*RI and *Sac*I. The approximate size of the bands detected with the probe are indicated.

A
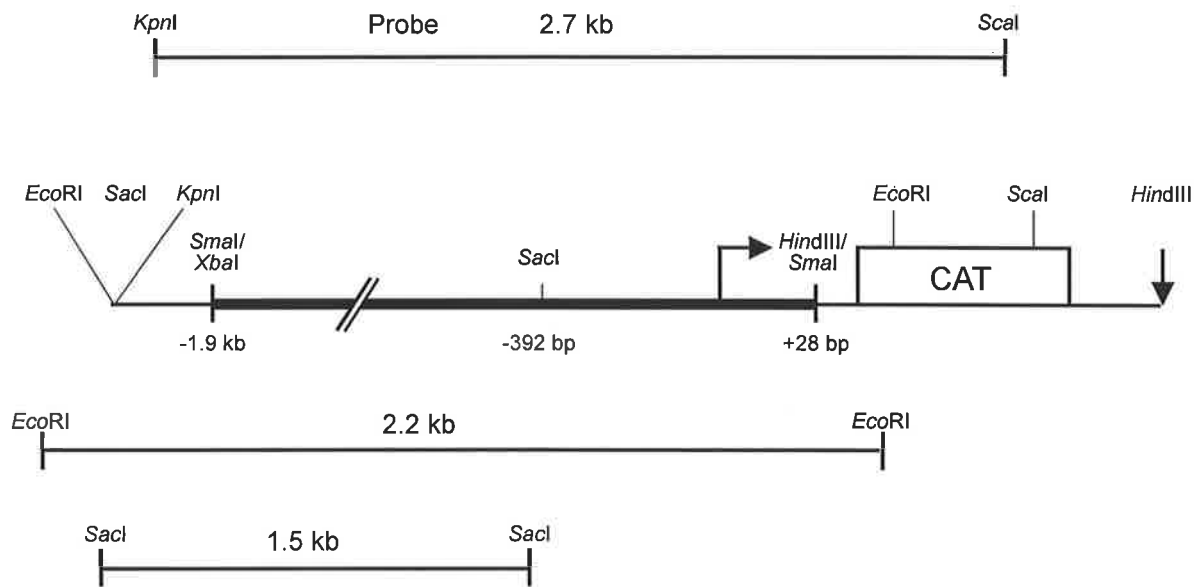
B
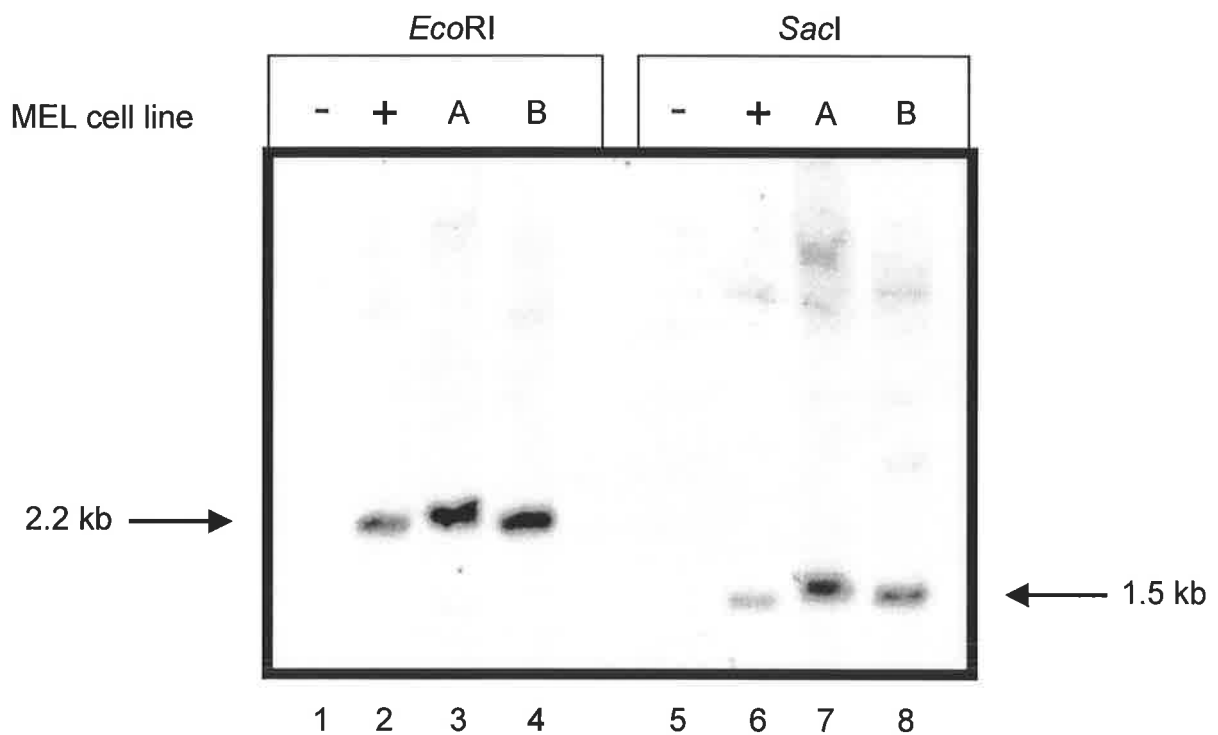
(Fig. 3.5B). Similarly, the upregulation of the level of ALAS2 mRNA from these cells was delayed compared with the parental cell line and this may account for the slower rate of hemoglobin synthesis.

The expression of CAT activity in clones A and B was also examined following induction with Me₂SO. As seen in Fig. 3.5C, high levels of CAT activity were detected in clone A, at days 1 and 2 post-induction when compared with the uninduced state (day 0), and expression then gradually decreased during the next 4 day period. Interestingly, induction of CAT activity occurred at the same time with the clone B cell line although the expression of the CAT reporter gene was reduced (Fig. 3.5C). This is in contrast to the endogenous erythroid messages that showed a lag in their induction compared with clone A and the parental cell line. The addition of Me₂SO was essential to activate expression of the CAT reporter gene driven by the ALAS2 promoter, since clone A or B cells cultured during a 6 day period in the absence of the chemical inducer showed only minimal CAT activity. No CAT activity was detected in the parental cell line (Fig. 3.5C).

In conclusion, a 1.9 kb fragment of the ALAS2 promoter fused to the CAT reporter gene was stably integrated into the genome of the MELC 707T cell line. These cells were treated with Me₂SO to induce erythroid differentiation and cells stained positive for hemoglobin synthesis. An apparent difference in the Me₂SO-induced differentiation of the two stable cell lines was observed, with a lag in both the activation of the ALAS2 and β-globin mRNAs and the synthesis of hemoglobin with clone B. However, examination of the CAT reporter gene driven by the 1.9 kb of ALAS2 promoter demonstrated that it was activated and expressed over a similar time period to that observed with the endogenous ALAS2 gene. Although the reason for this is unclear, these differences may depend on the site of integration of the pALAS-1.9kb-CAT construct into the genome, with the possible integration of this construct into a region required for the expression and differentiation of a number of erythroid-specific genes in clone B.

Figure 3.5   Me₂SO-induced differentiation of the MELC 707T cell line stably transfected with the pALAS-1.9kb-CAT construct

The parental MELC 707T cell line and the two clones A and B stably transfected with pALAS-1.9kb-CAT (5×10^5 cells/ml) were induced to differentiate with 1.9% Me₂SO for a period of 6 days. Uninduced cells from all three cell lines were included as a negative control. During the time course experiments, aliquots of the Me₂SO-induced parental, clones A and B were isolated daily, stained for hemoglobin synthesis, total RNA isolated for Northern blot analysis and cells isolated to determine CAT activity.

A   Me₂SO-induced parental MELC 707T cells, clone A and B cells, were stained daily with diaminofluorene and positive cells are an indicator of hemoglobin synthesis. The data is represented as the % of differentiated cells during the time course experiment.

B   Total RNA was then isolated at the indicated times following the addition of Me₂SO to the parental MELC 707T cell line and from clones A and B, and 20µg of total RNA was subjected to Northern blot analysis. The filter was sequentially probed for ALAS2 and β-globin mRNAs.

C   Cells were harvested at the indicated times from the Me₂SO-induced and the uninduced parental MELC 707T cell line and clones A and B. The protein concentration of each cell lysate was determined and 100µg of total protein assayed for CAT activity. Following autoradiography, the radioactivity in each spot was determined and CAT activities are expressed as the % of acetylated chloramphenicol.

A
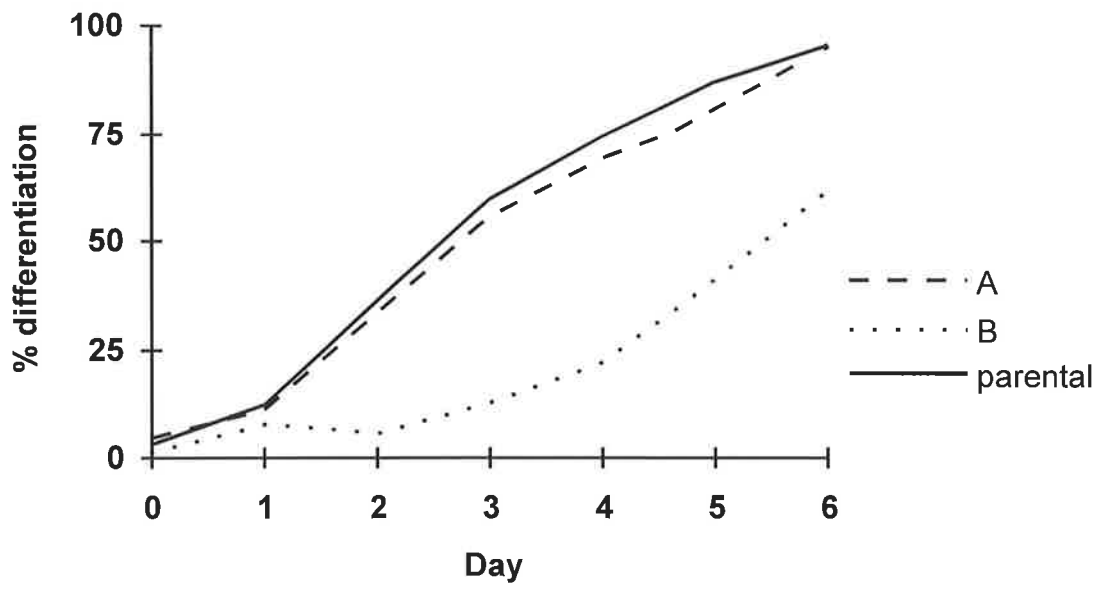
B
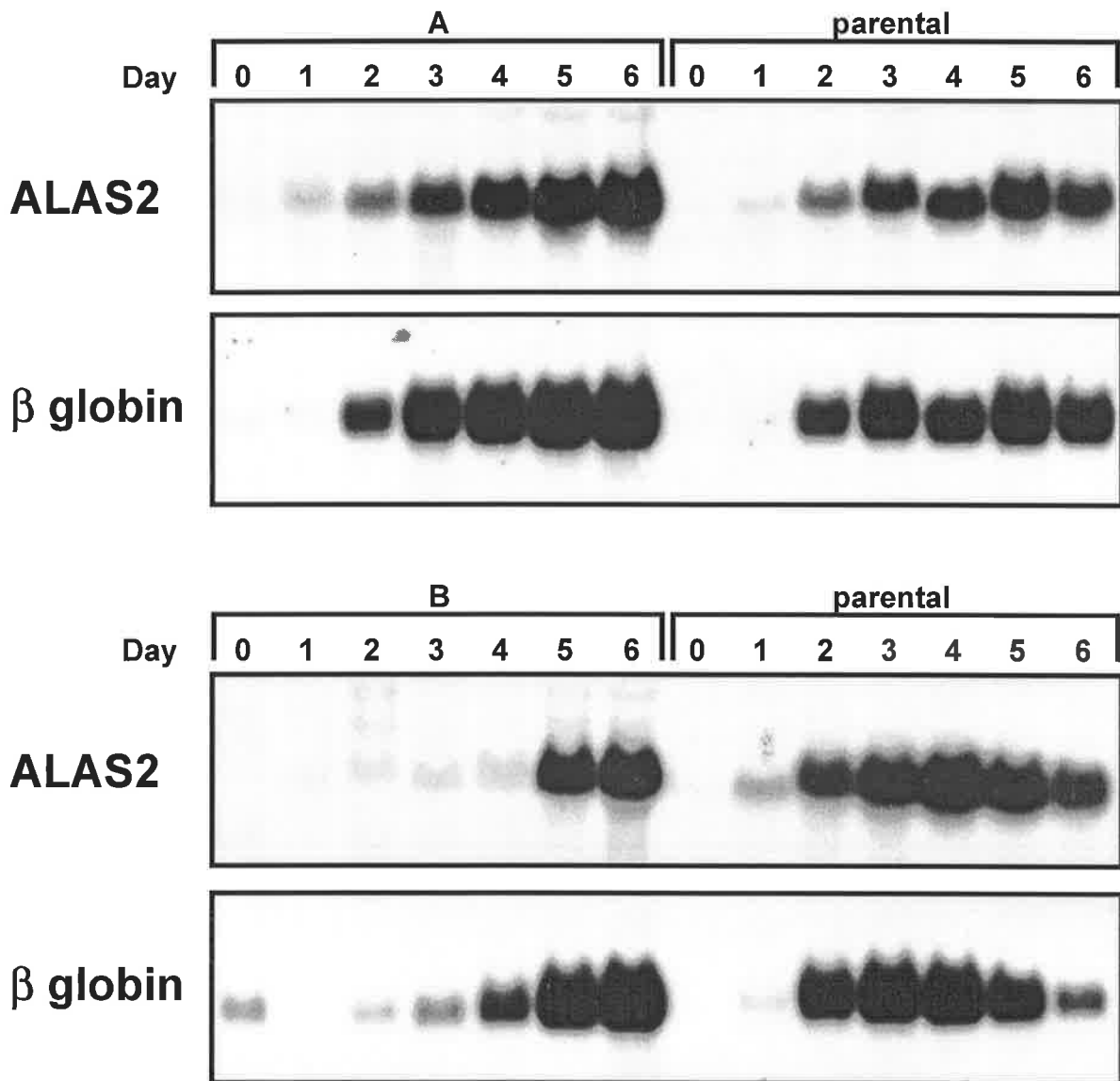
C

Induced with Me$_2$SO
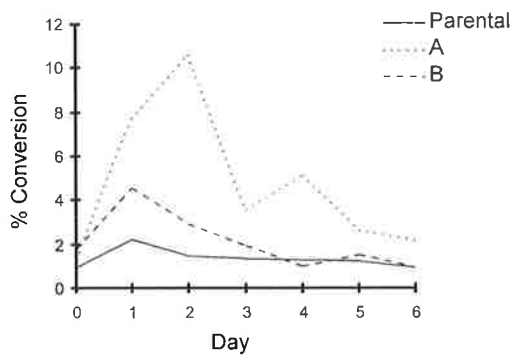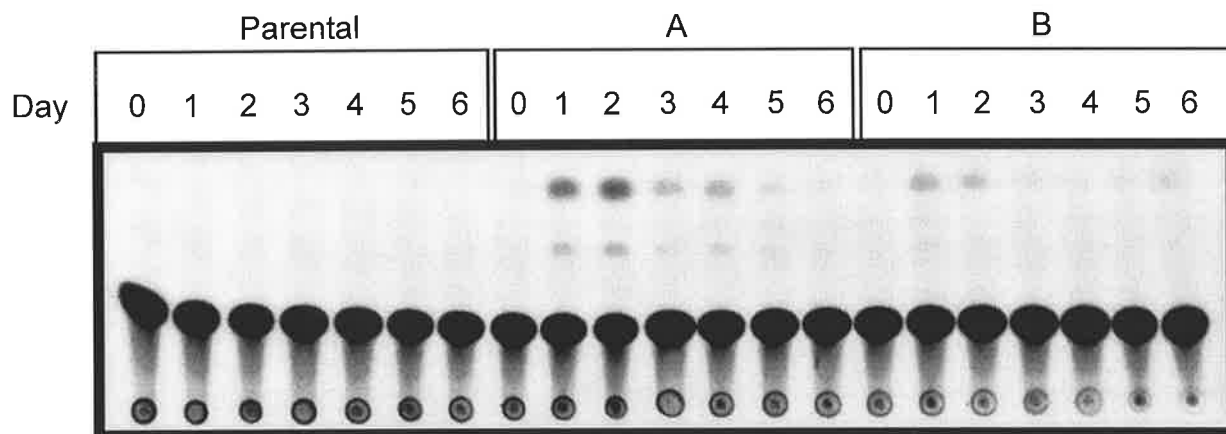
Uninduced
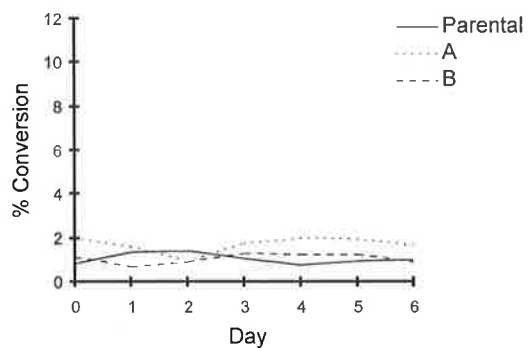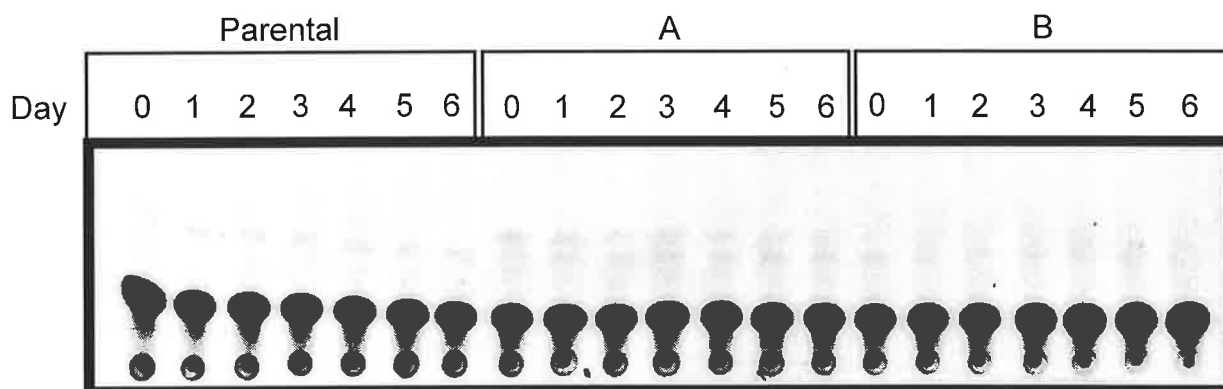
3.2.5 Detection of the Luciferase Reporter Gene in an Adherent MEL Cell Line and Northern Blot Analysis of this Cell Line Induced to Differentiate with Me₂SO

During the course of this project, an adherent MEL cell line, MEL F4-12B2 was acquired. This cell line was investigated to determine whether the inability to detect luciferase activity was a general feature of MEL cell lines.

The transfection of the MEL F4-12B2 cell line by electroporation was optimised with the pRSV-CAT plasmid. Electroporation buffer consisting of phosphate buffered saline (PBS) containing 10 mM Hepes, pH 7.5, successfully used in the transfection of the human K562 cell line by electroporation (Max-Audit *et al.*, 1993) was compared with the Hepes-buffered saline used for the successful transfection of the DS-19 and MELC 707T cell lines as described in Section 3.2.2. In initial experiments, PBS containing 10 mM Hepes, pH 7.05 produced higher activity of the CAT reporter construct (data not shown) and therefore was used in the remainder of these studies. In addition to this, CAT activity was readily detected with both 8μg (2 pmol) or 20μg of plasmid DNA and therefore 2 pmol of plasmid DNA was used in subsequent transient transfection assays (data not shown). To optimise the buffer volume, approximately 10^7 MEL F4-12B2 cells were transfected with 2 pmol of reporter plasmid and 250μg of salmon sperm DNA in a volume of either 200μl or 500μl of cold PBS containing 10mM Hepes, pH 7.5. These cells were exposed to a single volt pulse ranging from 100 to 300 volts in increments of 50 volts, at a capacitance of 960 microfarads. The promoterless CAT plasmid, pIBI-CAT was again included as a negative control (at one voltage 300V, 960μF). Cells were harvested and assayed 48 hours after electroporation (Section 2.6.4). As seen in Fig. 3.6, maximum CAT activity was observed at a 300-volt pulse in a final volume of 500μl of electroporation buffer. Therefore transfection of the MEL F4-12B2 cell line by electroporation consisted of 2 pmol of plasmid DNA, 250 μg of sonicated salmon sperm DNA in 500μl of cold PBS containing 10mM Hepes, pH 7.5, at a single 300-volt pulse at 960 microfarads.

To determine whether luciferase activity could be detected in this MEL cell line, cells were electroporated with 2 pmol of the plasmids pRSV-LUC (8μg) or pGL2-Basic using the optimised conditions described above. Cells were harvested 24 hours post-electroporation and assayed for luciferase activity. Expression of pRSV-LUC was readily detected in this cell line and at approximately 100-fold higher than pGL2-Basic, the

Figure 3.6 **Optimisation of electroporation conditions for the transient transfection of the MEL F4-12B2 cell line**

Exponentially growing MEL F4-12B2 cells (10^7) in 200μl or 500μl of cold PBS containing 10mM Hepes, pH 7.5, 2 pmol of pRSV-CAT (approximately 8μg) and 250μg of sonicated salmon sperm DNA were exposed to a single volt pulse ranging from 100V to 300V at a capacitance of 960 microfarads. The promoterless plasmid, pIBI-CAT was included as a negative control and cells electroporated at 300V, 960 microfarads. Cells were harvested at 48 hours following transfection, the protein concentration of the cell lysates determined and 100μg of total protein assayed for CAT activity.

Volume of EP Buffer	200μl						500μl					
Plasmid	pRSV-CAT					pIBICAT	pRSV-CAT					pIBICAT
Voltage	100	150	200	250	300	300	100	150	200	250	300	300
% Conversion	1.8	3.8	40	63	64	1.3	0.9	1.8	4.2	19	70	0.5
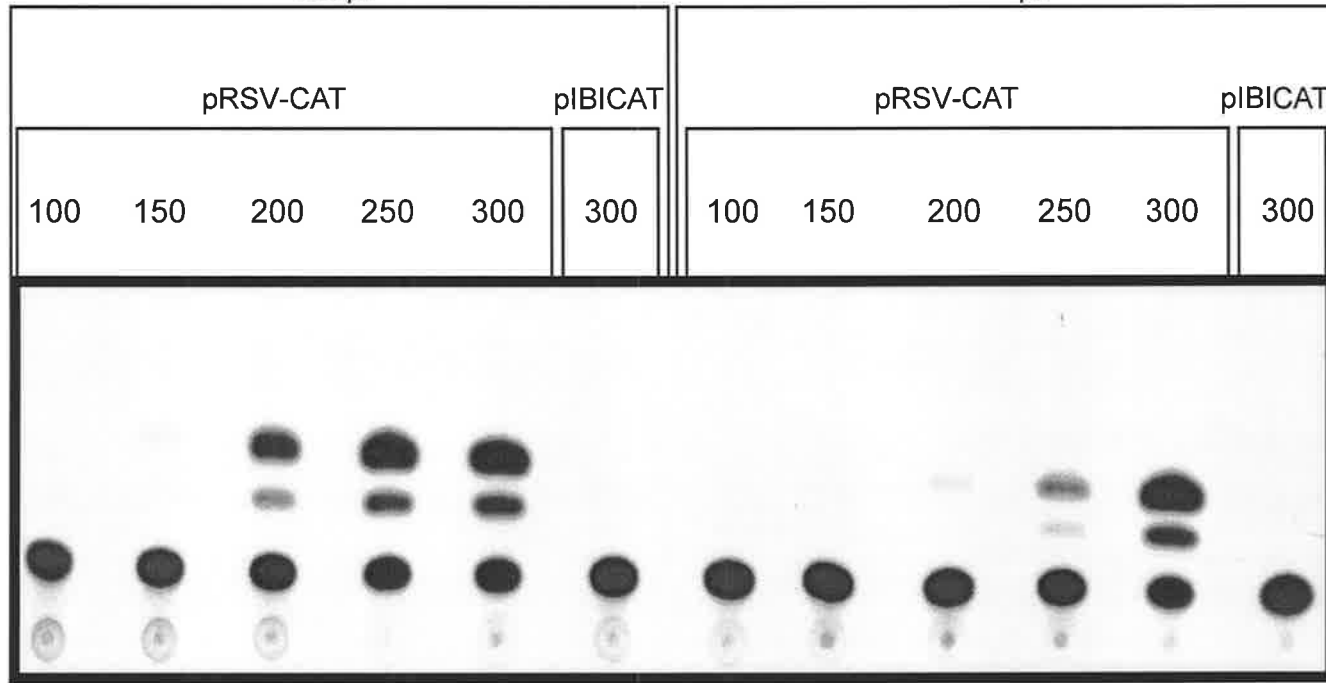
negative control. For comparison, luciferase activity was also assayed at 48 hours following electroporation. These results indicated a similar level of expression to that observed 24 hours post-transfection (data not shown). Consequently, all cell lysates were routinely assayed 24 hours following transfection.

In conclusion, expression of the luciferase reporter gene was readily detected in the transiently transfected erythroid MEL F4-12B2 cell line. To ensure that the MEL F4-12B2 cell line was suitable for expression studies, these cells were induced to terminally differentiate with Me$_2$SO (Ostertag et al., 1972). Me$_2$SO-induced cells were stained with diaminofluorene to determine the number of cells synthesising hemoglobin and total RNA isolated for Northern blot analysis as described previously. The levels of endogenous ALAS2, α- and β- globin mRNAs were examined during erythroid differentiation.

As seen in Fig. 3.7A, the proportion of diaminofluorene positive cells was approximately 2% in the uninduced culture (day 0) and this was significantly increased to approximately 75% in cultures grown for a period of 6 days in the presence of Me$_2$SO, indicating that a high proportion of the cells in the culture were synthesising hemoglobin. ALAS2 mRNA was initially detected at day 2 post-induction with approximately a 15-fold increase in this transcript detected by day 3 post-induction (Fig. 3.7B). A similar increase was also observed for transcripts encoding PBGD (data not shown). Transcripts for α- and β- globin were detected in uninduced MEL F4-12B2 cells (day 0), but following treatment with Me$_2$SO the levels of α- and β- globin transcripts were markedly increased (approximately 50-fold) at day 2 post-induction. Maximum levels of all four transcripts were observed at approximately day 4 post-induction whereas the level of GAPDH mRNA remained relatively unchanged throughout the Me$_2$SO-induced differentiation of the MEL F4-12B2 cell line (Fig. 3.7B).

The expression of transcripts encoding the transcription factors, GATA-1, EKLF and BKLF were also examined in the differentiating MEL F4-12B2 cells, using radiolabelled cDNAs to probe the Northern blot as previously described. EKLF and BKLF transcripts were distinguished by probing with the following cDNA fragments; a 733 bp NcoI-PvuII of the murine EKLF cDNA (Donze et al., 1995) and a 513 bp NcoI-NdeI of the murine BKLF cDNA (Crossley et al., 1996), which did not contain the highly conserved zinc finger coding region. As seen in Fig. 3.7B, the levels of GATA-1 and BKLF mRNAs were marginally

Figure 3.7   Me$_2$SO-induced differentiation of the MEL F4-12 B2 cell line

The MEL F4-12B2 cell line (5x10^5 cells/ml) was induced to differentiate with 1.86% Me$_2$SO and grown in culture for a period of 6 days. During the time course experiment, aliquots of Me$_2$SO-induced MEL cells were isolated daily, stained for hemoglobin synthesis and total RNA isolated for Northern blot analysis. Untreated MEL F4-12B2 cells was included as a negative control.

A   Me$_2$SO-induced MEL F4-12B2 cells were stained daily with diaminofluorene and positive cells are an indicator of hemoglobin synthesis. The number of hemoglobin positive cells are expressed relative to the total number of cells. The data is represented as the % of differentiated cells during the 6 day time course experiment.

B   Total RNA was then isolated at the indicated times following the addition of Me$_2$SO to the MEL F4-12B2 cell line and 20μg of total RNA was subjected to Northern blot analysis. The filter was sequentially probed for ALAS2, the α- and β- globins, GATA-1, EKLF, BKLF and GAPDH mRNAs.

increased (approximately 2-3 fold), whereas expression of the EKLF transcript remained similar throughout the Me₂SO-induced differentiation of the MEL F4-12B2 cells.

In conclusion, Northern blot analysis of the MEL F4-12B2 cell line induced to differentiate following treatment with Me₂SO, demonstrated that the endogenous transcripts encoding ALAS2, α- and β- globins were all significantly upregulated and this correlated with an increase in the number of cells synthesising hemoglobin. These findings indicate that this cell line contains all of the factors required for the induction of ALAS2 gene transcription. Therefore the MEL F4-12B2 cell line is suitable for use as an *in vitro* model to investigate the transcriptional regulation of the human ALAS2 gene. Transient expression studies performed in this cell line with the human ALAS2/luciferase reporter gene plasmids are described in Chapters 4, 5 and 6.

3.2.6 An Investigation into the Suitability of the Erythroid J2E and K562 Cell Lines for Transient Transfection Studies

Preliminary transient transfection studies were also performed with the erythroid J2E and K562 cell lines. However, initial attempts to detect expression of the plasmid pRSV-LUC transfected in the erythroid J2E cell line by electroporation was unsuccessful. It was assumed the lack of detectable luciferase activity in J2E cells was due to the same reasons that activity could not be found in the suspension and semi-adherent MEL cell lines, and was therefore not pursued further. An investigation into transfecting the J2E cell line by Dr. S. P. Klinken who generated this cell line, has revealed that these cells are extremely difficult to transfect (P. Klinken, personal communication). However, the transfection protocol, "receptor-mediated transferrinfection" using a novel transferrin/poly-L-lysine conjugate has been successfully employed to introduce plasmid constructs into this cell line (Taxman *et al.*, 1993).

The K562 cell line has been successfully used in transient transfection studies to identify critical regulatory regions responsible for the expression of numerous erythroid-specific genes and was also included in this study (Yu *et al.*, 1991; Rahuel *et al.*, 1992; Fischer *et al.*, 1993; Max-Audit *et al.*, 1993; Fraizer *et al.*, 1994). Traditionally, these studies were performed with either the CAT (Rahuel *et al.*, 1992) or the growth hormone reporter gene systems (Yu *et al.*, 1991) and recently the luciferase reporter gene (Fischer *et*

al., 1993; Max-Audit *et al.*, 1993; Fraizer *et al.*, 1994) has been successfully used in this cell line. Consequently, preliminary transient transfection experiments were performed by electroporation following the standardised transfection protocol reported by Max-Audit *et al.* (1993) used in the successful detection of luciferase activity, with 2 pmol of the plasmid, pRSV-LUC. Expression of the luciferase reporter gene was detected in these transiently transfected K562 cells (data not shown).

The suitability of the K562 cell line, which is known to express the embryonic and fetal globins, was investigated further by inducing these cells to differentiate with hemin for a period of 3 days (Weber-Benarous *et al.*, 1988). These cells were stained with diaminofluorene to determine the number of cells synthesising hemoglobin, and total RNA isolated for Northern blot analysis to examine the level of endogenous ALAS1 and ALAS2 mRNAs.

As seen in Fig. 3.8A, the proportion of diaminofluorene positive cells was approximately 14.5% in the uninduced culture (day 0) and this was significantly increased to approximately 67% in cultures grown for a period of 3 days in the presence of hemin. However, only transcripts encoding ALAS1 but not ALAS2 were detected in these cells chemically induced to differentiate with hemin (Fig. 3.8B).

In conclusion, expression of the luciferase reporter gene was detected in transiently transfected K562 cells and these cells can also be induced to differentiate with hemin and synthesise hemoglobin. However, Northern blot analysis failed to detect ALAS2 mRNA in either the uninduced or the hemin-induced K562 cells.

3.3   DISCUSSION

In this chapter, several MEL cell lines, the J2E and the K562 cell lines were analysed to evaluate their suitability for transient expression studies of the human ALAS2 promoter. Transient transfection studies performed in two uninduced MEL cell lines; suspension DS-19 and semi-adherent MELC 707T, failed to detect expression of a luciferase reporter gene construct but the reason for this is not known. However, while expression of the bacterial CAT reporter gene was detected in both these MEL cell lines, this reporter gene was not considered suitable for transient expression studies. This was due to the low level of CAT activity and the consequent requirement for large quantities of the CAT reporter

Figure 3.8 **Hemin-induced differentiation of the K562 cell line**

K562 cells (1×10^6 cells/ml) were induced to differentiate with 20µM hemin for 3 days. During the time course experiment, aliquots of hemin-induced K562 cells were isolated daily, stained for hemoglobin and total RNA isolated for Northern blot analysis.

A Hemin-induced K562 cells were stained daily for hemoglobin synthesis. The number of hemoglobin positive cells are expressed relative to the total number of cells. The data is represented as the % of differentiated cells during the 3 day time course experiment.

B Total RNA was then isolated at the indicated times following the addition of hemin to the K562 cell line and 20µg of total RNA was subjected to Northern blot analysis. Total RNA isolated from uninduced DS-19 cells (day 0) and DS-19 cells cultured with Me$_2$SO for 3 days, and fetal liver were included as positive controls for the detection of the ALAS2 transcript. The filter was sequentially probed for ALAS2 and then ALAS1 mRNAs.

A
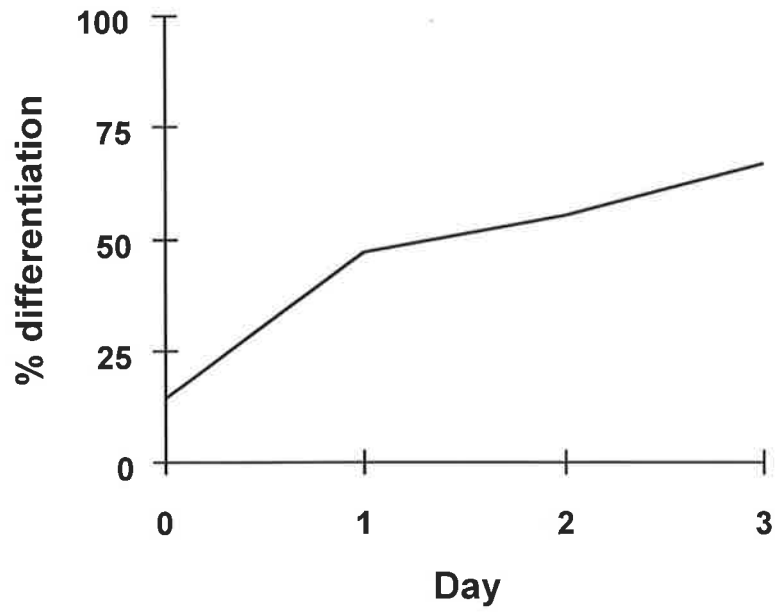
B
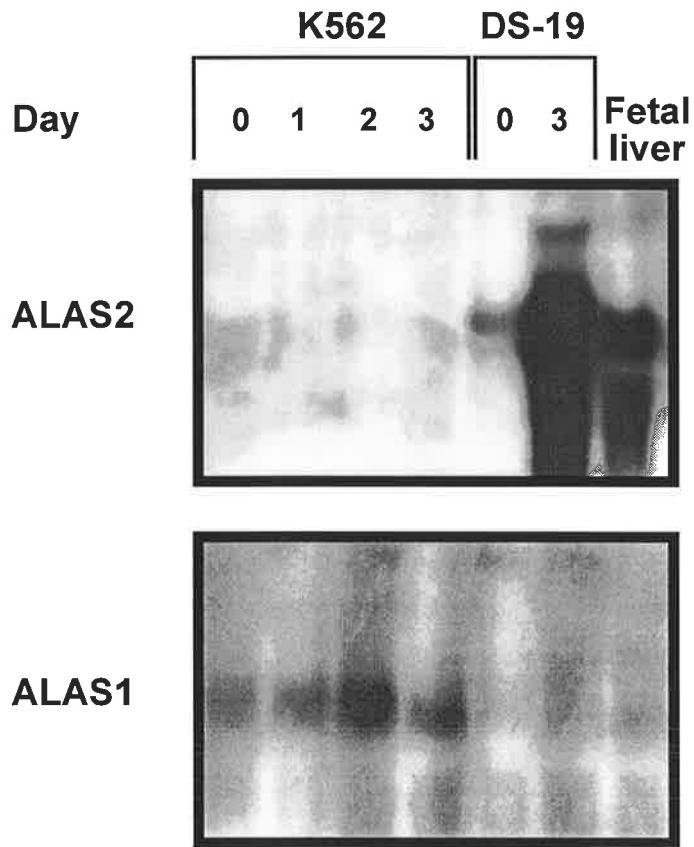
gene plasmid DNA. In contrast, luciferase activity was readily detected in the transiently transfected adherent MEL F4-12B2 cell line, indicating that the inability to detect luciferase activity in the DS-19 and MELC 707T cells is not a common feature associated with MEL cell lines.

Northern blot analysis of Me_2SO-induced MEL cell lines; DS-19, MELC 707T and MEL F4-12B2 demonstrated that the levels of ALAS2 mRNA, α- and β- globin mRNA were all significantly upregulated during erythroid differentiation and this was associated with an increase in the number of cells synthesising hemoglobin. The coordinated activation of the ALAS2 and globin genes during Me_2SO-induced differentiation is in agreement with previous studies (Fraser & Curtis, 1987; Fujita *et al.*, 1991) demonstrating that MEL F4-12B2 cells contain all of the factors necessary for the transcriptional activation of the ALAS2 gene during erythroid differentiation.

The level of several transcription factors implicated to play a role in the transcription of other erythroid-specific genes were examined to determine whether a correlation existed between their accumulation and the transcriptional activation of ALAS2. The level of GATA-1 mRNA was marginally increased following exposure to Me_2SO and remained similar throughout erythroid differentiation of the MEL F4-12B2 cell line, although Busfield *et al.* (1995) reported a marginal decrease in the level of GATA-1 mRNA following exposure to Me_2SO. This discrepancy may be attributable to differences between the strains of the various MEL cell lines. The findings observed with the expression of GATA-1 mRNA in Me_2SO-induced MEL cells are in direct contrast to the rapid increase in GATA-1 mRNA which has been observed in both the J2E cell line (Busfield *et al.*, 1995) and erythroid progenitor cells (Dalyot *et al.*, 1993) following exposure to Epo. Dalyot *et al.* (1993) demonstrated that following the addition of Epo to erythroid progenitor cells, the burst in GATA-1 activity occurs immediately prior to globin mRNA expression (Dalyot *et al.*, 1993). Similarly, the level of EKLF mRNA encoding the erythroid-enriched CACCC box binding protein, has been previously reported to remain unchanged in MEL DS-19 cells chemically induced with HMBA to differentiate (Miller & Bieker, 1993). In both Me_2SO-induced and uninduced MEL F4-12B2 cells, EKLF also showed little change, suggesting that the level of EKLF mRNA is not influenced by the onset of terminal differentiation (Miller & Bieker, 1993).

Therefore the question still remains: "How is the transcription of the genes encoding the enzymes of the heme biosynthetic pathway and the globins upregulated during the Me$_2$SO-induced differentiation of MEL cells?" There exists at least two possible explanations: the involvement of another erythroid-specific factor, or protein phosphorylation. Examination of the phosphorylation status of GATA-1 during Me$_2$SO-induced MEL cell differentiation revealed that phosphorylation at serine 310 is increased during erythroid differentiation (Crossley *et al.*, 1994b). However, changes in the phosphorylation status of GATA-1 during Me$_2$SO-induced differentiation does not significantly alter the ability of the protein to bind DNA as determined in gel shift assays (Crossley *et al.*, 1994b; Busfield *et al.*, 1995), bend DNA or to transactivate a reporter construct containing a functional GATA-1 binding site (Crossley *et al.*, 1994b). Chen & Bieker (1996) demonstrated that the transactivation domain of EKLF is composed of activation and inhibitory subdomains. The inhibitory domain of EKLF has been proposed to function by interfering with efficient DNA binding. Therefore during the Me$_2$SO-induced differentiation of MEL cells, EKLF may be activated by modifying this negative subdomain. Activation of EKLF my be vital, since EKLF has been proposed to function in co-operation with other transcription factors, such as GATA-1 (Merika & Orkin, 1995; Gregory *et al.*, 1996). In addition, a number of putative phosphorylation sites within the EKLF protein have been identified and phosphorylation may play a role in the activation of EKLF (Chen & Bieker, 1996).

Another recently identified member of the Krüppel-like family of transcription factors, BKLF, increases during erythroid differentiation (Crossley *et al.*, 1996). Northern blot analysis of BKLF mRNA in Me$_2$SO-induced MEL F4-12B2 cells revealed a marginal increase in the level of the BKLF transcript in differentiating cells. However, target genes that are regulated by BKLF remain to be identified.

Finally, stable MELC 707T cell lines carrying 1.9 kb of ALAS2 promoter driving the expression of the CAT reporter gene were created to determine whether this region of the ALAS2 promoter was sufficient for the activation of transcription during Me$_2$SO-induced differentiation. Me$_2$SO-treatment resulted in a substantial increase in CAT activity indicating that the elements necessary for the expression of the ALAS2 gene following Me$_2$SO-induction are located within this 1.9 kb region. However, in contrast to the increase in the level of the endogenous ALAS2 mRNA transcript during

Me$_2$SO-induced erythroid differentiation, CAT activity peaked at day 1-2. This observation may indicate that the stability of the CAT reporter transcript is affected during Me$_2$SO-induced differentiation. Alternatively, the transcriptional activation of the reporter construct may be subject to silencing with the onset of differentiation, significantly earlier than could be expected based on endogenous mRNA levels in the parental line. Indeed, a characteristic feature of terminally differentiating erythroid cells is an increase in chromatin condensation and the silencing of genes not required for erythroid development (Sherton & Kabat, 1976; Fraser & Curtis, 1987). The role of the erythroid-specific enhancers, HS-40 and LCR of the α- and β- globin gene clusters, respectively, has been proposed to delay the silencing of these erythroid-specific genes and maintain their transcription in an otherwise "repressive" environment (Walters et al., 1995; Walters et al., 1996; Sutherland et al., 1997). Two lines of evidence support this proposal. Firstly, the formation of the DNase I hypersensitivity sites of the LCR occurs prior to detectable transcription of the β-globin genes (Groudine et al., 1983; Forrester et al., 1986; 1987). Secondly, deletion of the DNase I HS2 of the LCR from a reporter construct stably integrated into K562 cells, significantly increases the rate at which expression of the β-geo reporter construct is silenced (Walters et al., 1996). It is possible that other erythroid-specific genes may utilise a similar mechanism to maintain transcription of their genes in a "repressive" environment. In this regard, the finding that 1.9 kb of ALAS2 promoter is sufficient to activate gene transcription in Me$_2$SO-induced MEL cells but not maintain expression suggests the possible involvement of other regions of the ALAS2 gene, located either further upstream in the 5'-flanking region, within the ALAS2 gene or in the 3'-flanking region. Therefore analogous to the role of the erythroid-specific enhancers of the α- and β-globin gene clusters, these regions of the ALAS2 gene may be responsible for the delay in the silencing of the ALAS2 gene during erythroid differentiation.

To define the minimal region of the ALAS2 promoter that is activated during the Me$_2$SO-induction of MEL cells, a series of ALAS2 deletion constructs stably transfected in the MELC 707T cell line could be generated. However, making stable cell lines is a very time consuming approach and since the erythroid MEL F4-12B2 and K562 cell lines can be both induced to differentiate *in vitro*, synthesise hemoglobin and are readily transfected, transient analysis was chosen rather than the generation of stable cell lines.

In contrast to the correlation observed between the upregulation of ALAS2 mRNA and hemoglobin synthesis in Me$_2$SO-induced MEL cell lines, endogenous ALAS2 mRNA was not detected during the hemin-induced differentiation of K562 cells by Northern blot. Instead endogenous ALAS1 mRNA was detected. Although, Minegishi *et al.* (1994) reported the detection of ALAS2 mRNA in K562 cells, repeated Northern blot analysis of ALAS2 mRNA from hemin-induced K562 cells which have been performed in our laboratory on a number of occasions (Dr. T. Cox, personal communication), are consistent with the findings presented in this chapter. It is also interesting to note that the β-globin gene is not transcribed in these cells even following induction with hemin (Charnay & Maniatis, 1983). As described in Section 1.14, this absence of adult globins is apparently not attributable any major insertions, deletions or rearrangements in or adjacent to the β-globin gene (Dean *et al.*, 1983; Mueller *et al.*, 1983; Weber-Benarous *et al.*, 1988). The absence of endogenous β-globin expression in K562 cells is also not due to the reduced expression of EKLF (Donze *et al.*, 1995). Nevertheless, these cells are readily transfected and have been employed widely for similar studies, suggesting that these cells contain the necessary transcription factors required for erythroid-specific gene expression. Therefore we decided to pursue the use of K562 cells, in addition to the MEL F4-12B2 cell line in our transient transfection assays.

The African green monkey kidney fibroblast COS-1 cell line previously shown capable of expressing transiently transfected luciferase reporter gene constructs was included as a non-erythroid control cell line (C. Hahn, personal communication). Transfection of this cell line by electroporation has been previously optimised by J. Beale.

In the ensuing chapters, preliminary transient transfection studies are described which were aimed at identifying regions of the ALAS2 gene responsible for erythroid-specific expression of the gene. These studies were all performed in the uninduced erythroid MEL F4-12B2 and K562 cell lines, and the non-erythroid COS-1 cell line, using the luciferase reporter gene constructs following the protocols and optimised conditions described in this chapter.

CHAPTER FOUR

TRANSIENT TRANSFECTION STUDIES OF THE ALAS2 5'-FLANKING AND INTRONIC REGIONS ON THE TISSUE-SPECIFIC EXPRESSION OF THE ALAS2 GENE

CHAPTER FOUR : TRANSIENT TRANSFECTION STUDIES OF THE ALAS2 5'-FLANKING AND INTRONIC REGIONS ON THE TISSUE-SPECIFIC EXPRESSION OF THE ALAS2 GENE

4.1   INTRODUCTION

As described in Section 1.7.3, the structural organisation of the human, murine and avian ALAS2 genes are remarkably similar (see Fig. 1.5). DNase I hypersensitivity studies performed on the murine ALAS2 gene, identified five hypersensitive sites which were present in both uninduced and Me_2SO-induced MEL cells but not in non-erythroid cells (Schoenhaut & Curtis, 1989). These hypersensitivity sites were located in the immediate promoter region, at the 5' end of intron 1 (two sites), within intron 3 (one site) and at the 3' end of intron 8 (one site) (Schoenhaut & Curtis, 1989). Such DNase I hypersensitive sites are indicative of nucleosome-free regions of DNA and usually correspond to regions that interact with factors important for transcriptional regulation (Elgin, 1988; Gross & Garrard, 1988). Therefore to identify the sequences in the human ALAS2 gene responsible for tissue-specific expression, regions of the sequenced human ALAS2 gene (22 kb) corresponding to hypersensitive sites in the murine ALAS2 gene and 5'-flanking (10.3 kb) and 3'-flanking (3 kb) sequences were examined for putative *cis*-acting elements (performed by Dr. T. Cox). Numerous putative binding sites for the erythroid-enriched transcription factor GATA-1, CACCC and CCAAT box binding proteins were identified throughout the 5'-flanking sequence.

The binding sites identified within the immediate human ALAS2 promoter region included two GATA-1 binding sites, a GATA motif at the TATA box location, an NF-E2 binding site, an Ets-like motif, CACCC and CCAAT box binding sites, all clustered within 140 bp of sequence located immediately upstream of the transcription initiation site. Similarly, within intron 1, sixteen consensus GATA-1 binding sites and six CACCC boxes were identified in the 4.9 kb intronic sequence. Analysis of the 850 bp intron 3 identified five putative GATA-1 binding sites and two CACCC box binding sites. Interestingly, intron 8 (562 bp) contained four putative consensus GATA-1 binding sites and two CACCC box binding sites located within 270 bp of sequence. The 3'-flanking region of the human ALAS2 gene contained two putative GATA-1 binding sites, two CACCC box

binding sites and a putative NF-E2-like sequence with a single nucleotide (5'-GGCTGACTCAT-3') mismatch from the consensus NF-E2 motif (5'-(T/C)GCTGAG/CTCA(C/T)-3') (Andrews *et al.*, 1993a).

In this chapter, a series of 5'-flanking ALAS2 deletion/luciferase constructs were synthesised and examined in transient expression studies in the erythroid MEL F4-12B2 and K562 cell lines and the non-erythroid, COS-1 cell line. Deletion analysis of the ALAS2 5'-flanking sequence demonstrated that the first 300 bp of ALAS2 promoter resulted in maximal expression of the luciferase reporter gene. However, these ALAS2 5'-flanking constructs efficiently expressed in both erythroid and non-erythroid cells. Therefore to identify region(s) of the ALAS2 gene which may confer tissue-specific expression, constructs were generated containing the 293 bp of ALAS2 promoter together with either intron 1, intron 3, intron 8 or the 3'-flanking sequence and the effect of these sequences on the tissue-specific expression of the promoter examined in transient transfection studies.

4.2 RESULTS

4.2.1 Analysis of the 5'-Flanking Region of the Human ALAS2 Gene

In order to investigate the regions of the human ALAS2 gene that contribute to expression, a series of 5'-flanking ALAS2 deletion constructs were generated in which different 5'-lengths (-10.3 kb to -27 bp), and with a common 3' end (+28) were fused to the firefly luciferase reporter gene as described in Section 2.4.1. These constructs were transiently transfected into either K562, MEL F4-12B2, or COS-1 cells as described in Section 2.6 and luciferase activity determined in cell lysates. As an internal control for transfection efficiency, the β-galactosidase expression plasmid, pRSV-β-gal, was co-transfected with each of the constructs. To standardise for the different lengths of the human ALAS2 promoter/luciferase constructs, 2 pmol of plasmid DNA was routinely used for all constructs, which is equivalent to approximately 7.5μg and 21μg of pALAS-27-LUC and pALAS-10.3kb-LUC, respectively. The plasmid pGL2-Basic was also included as a negative control. The luciferase activity of each plasmid construct was corrected for expression following the subtraction of luciferase activity obtained with pGL2-Basic. The luciferase and β-galactosidase activities were determined in all three cell lysates, at 24 hours

following transfection as described in Section 2.6.4, and the data expressed as "Relative luciferase activity".

The activity of the longest construct (pALAS-10.3kb-LUC) in each cell line was assigned a value of 100 (Fig. 4.1B). The promoter expressed strongly in both erythroid cell lines and maximal activity was seen with 293 bp of promoter (pALAS-293-LUC). A low level of activity was obtained with the -27 bp promoter construct (pALAS-27-LUC) compared with 293 bp of promoter. Expression of these constructs was also observed in COS-1 cells and followed a similar pattern to that in erythroid cells except that 1.9 kb of promoter (pALAS-1.9 kb-LUC) gave maximal expression (Fig. 4.1B).

In conclusion, transient transfection analysis of the human ALAS2 5'-flanking sequence demonstrated that the first 300 bp of promoter resulted in maximal expression of the luciferase reporter gene in erythroid cells. The identification of regulatory elements located within this region and the corresponding transcription factors that drive expression of the human ALAS2 promoter are described in Chapter 5. Since promoter constructs with 5'-flanking sequence from 293 bp to 10.3 kb expressed in transiently transfected non-erythroid cells as well as in erythroid cells, an enhancer sequence located elsewhere or native chromatin structure may be required for the tissue-restricted expression of the gene *in vivo*.

4.2.2 Analysis of Other Potential Regulatory Regions of the Human ALAS2 Gene

To investigate whether sequences located elsewhere in the ALAS2 gene were responsible for tissue-specific expression, regions of the human ALAS2 gene corresponding to DNase I hypersensitivity sites in the murine ALAS2 gene as well as the human ALAS2 3'-flanking (Fig. 4.2A), were examined in transient expression studies. A series of constructs containing the human ALAS2 intron 1 (4.973 kb), intron 3 (850 bp) and intron 8 (460 bp) fragments and the 3'-flanking region (2.935 kb) were generated from the human genomic clone, pTC-EA1 (Cox *et al.*, 1991) and ligated into the vector, pALAS-293-LUC (Fig. 4.1) to generate pALASp-Int1-LUC, pALASp-Int3-LUC, pALASp-Int8(460)-LUC and pALASp-3'-LUC, respectively (Fig. 4.2). The 293 bp of ALAS2 promoter was chosen since maximal expression of the luciferase reporter gene was obtained in erythroid cells. The plasmid pALAS-293-LUC described in Section 4.2.1 is referred to as pALASp-LUC in this

Figure 4.1   Transient expression analysis of the 5'-flanking region of the human ALAS2 gene

A   A partial restriction map of the 5'-flanking region of the human ALAS2 gene. Restriction enzyme sites used to generate deletion fragments are indicated. The transcription initiation site at +1 of the ALAS2 gene is depicted by the *arrow*.

B   Expression of the 5'-flanking ALAS2 deletion constructs in transiently transfected K562, MEL and COS-1 cells. Luciferase activities were standardised relative to β-galactosidase activity (pRSV-βgal) as an internal control for variation in transfection efficiencies. The normalised luciferase activities are expressed relative to pALAS-10.3kb-LUC which was set at 100%. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation.

A

ClaI −10.3 kb | KpnI −5.7 kb | XbaI −1.9 kb | BglI −293 bp | +1 | ALAS2

B
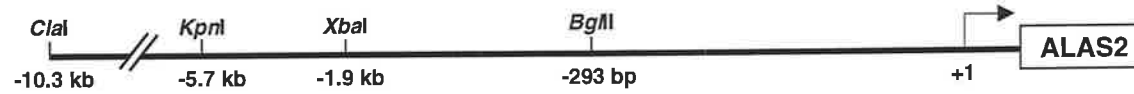
Relative luciferase activity

	K562	MEL	COS-1
pALAS-10.3kb-LUC	100±15	100±31	100±12
pALAS-5.7kb-LUC	84±5	229±25	76±15
pALAS-1.9kb-LUC	184±23	311±38	187±8
pALAS-293-LUC	234±6	608±115	78±17
pALAS-124-LUC	151±25	216±40	73±13
pALAS-27-LUC	56±19	75±18	6±1

section so as to distinguish between constructs containing 5'-flanking sequence or intronic sequences.

As described in Section 2.4.2, the complete ALAS2 intron 1 fragment was ligated downstream of the ALAS2 promoter in the native orientation and the resulting construct contained -293 bp to +4.973 kb of contiguous ALAS2 sequence. Plasmid constructs containing intron 8 ligated upstream of the ALAS2 promoter were synthesised, since repeated attempts to clone the intron 8 fragment downstream of the luciferase reporter gene were unsuccessful. However, only clones containing the intron 8 fragment in the native orientation were isolated. Initial expression studies with plasmid constructs containing intron 8 indicated that a fragment ligated immediately upstream of the ALAS2 promoter was capable of increasing promoter activity. Therefore plasmid constructs containing either intron 3 or the ALAS2 3'-flanking region in both orientations were also ligated upstream of the ALAS2 promoter.

These human ALAS2 promoter intronic/luciferase reporter gene plasmids were co-transfected with the pRSV-βgal plasmid and transiently expressed in K562, MEL F4-12B2, or COS-1 cells as previously described. The plasmid construct pALASp-LUC containing the 293 bp human ALAS2 promoter fused to the luciferase reporter gene was included as a positive control and assigned a value of 1.0 in each cell line.

As seen in Fig. 4.2B, the inclusion of intron 1 (pALASp-Int1-LUC) increased transcription of the ALAS2 promoter 3.2 and 2.4 -fold in K562 and MEL cell lines, respectively, and significantly reduced expression approximately 4-fold in COS-1 cells and in another non-erythroid cell line JTC-12 (data not shown). The addition of human intron 3 in the native orientation, upstream of the 293 bp human ALAS2 promoter (pALASp-Int3-LUC), reduced expression approximately 1.6-fold in K562 cells and approximately 2-fold in COS-1 cells (Fig. 4.2B). A similar result was obtained when a plasmid construct containing the intron 3 fragment in the reverse orientation was examined (data not shown). Interestingly, the addition of the 460 bp intron 8 fragment in the native orientation, upstream of the 293 bp human ALAS2 promoter (pALASp-Int8(460)-LUC) significantly increased luciferase expression approximately 12 and 4 -fold in K562 and MEL cells, respectively, but did not alter expression in COS-1 cells (Fig. 4.2B). Transient expression studies performed in K562 cells, demonstrated that the addition of the 3'-flanking sequence (pALASp-3'-LUC) reduced the level of expression approximately

Figure 4.2   Expression analysis of human ALAS2 intronic sequences corresponding to DNase I hypersensitivity sites in the murine ALAS2 gene and the 3'-flanking region

A   The human ALAS2 gene is composed of 11 exons spanning 22 kb. Regions of the human ALAS2 gene corresponding to DNase I hypersensitivity sites identified in the murine ALAS2 gene, and the 3'-flanking region are indicated.

B   A series of human ALAS2 intronic constructs were synthesised containing 293 bp of ALAS2 promoter and either intron 1, intron 3, intron 8 or the 3'-flanking sequence. The 293 bp ALAS2 promoter construct is referred to as pALASp-LUC. The size, location and orientation (*arrowed*) of these intronic sequences and the 3'-flanking sequence relative to the ALAS2 promoter are indicated. These constructs and pALASp-LUC were co-transfected with pRSV-βgal and transiently expressed in K562, MEL and COS-1 cells. Luciferase activities were standardised relative to β-galactosidase activity as an internal control for variation in transfection efficiencies. The normalised luciferase activities are expressed relative to pALASp-LUC assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not tested in a particular cell line.

A

EXONS: 1, 2 (AUG), 3, 4, 5, 6, 7, 8, 9, 10, 11 (TAA) — 22 kb

Immediate promoter, Intron 1, Intron 3, Intron 8, 3'-flanking

B
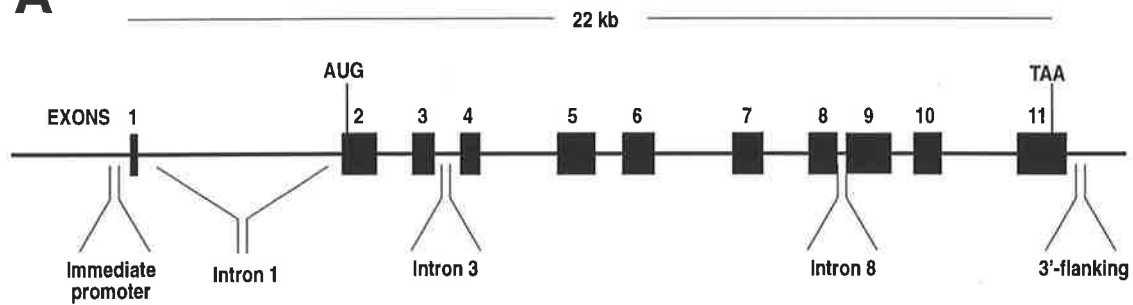
	Fold Induction		
	K562	MEL	COS-1
pALASp-LUC	1.0	1.0	1.0
pALASp-Int1-LUC	3.2±0.11	2.4±0.1	0.25±0.1
pALASp-Int3-LUC	0.6±0.02	ND	0.48±0.05
pALASp-Int8(460)-LUC	12±0.4	4±0.8	0.9±0.2
pALASp-3'-LUC	0.66±0.03	ND	ND
pALASp-Int1 5'Int8-LUC	7.8±0.3	ND	0.29±0.06
pALASp-Int1 3'Int8-LUC	8.6±0.2	ND	0.25±0.05
pALASp-Int1 3'Int8R-LUC	8.9±0.1	ND	0.27±0.01

1.5-fold irrespective of fragment orientation (Fig. 4.2B). However, since the primary aim was to identify regions of the ALAS2 gene capable of increasing the basal transcription of the promoter in erythroid cells, no further studies were performed with either intron 3 or the 3'-flanking sequence.

In conclusion, transient expression studies demonstrated that sequences located within intron 1 and intron 8 were capable of transactivating the ALAS2 promoter in an erythroid-specific manner. To investigate whether sequences located within these introns functioned cooperatively, additional plasmid constructs containing both intron 1 and intron 8 were generated in which intron 8 was ligated upstream of the ALAS2 promoter and intron 1 in the native orientation (pALASp-Int1-5'Int8-LUC) or downstream of the luciferase reporter gene in both orientations (pALASp-Int1-3'Int8-LUC) and (pALASp-Int1-3'Int8R-LUC) as described in Section 2.4.2. Similarly, to investigate whether these ALAS2 intronic sequences were also capable of increasing expression of the heterologous thymidine kinase promoter, plasmid constructs containing intron 1 or intron 8 sequence were also synthesised (refer to Section 2.4.2). A 3.984 kb *SacI* intron 1 fragment containing the majority of putative transcription factor binding sites (thirteen GATA-1 binding sites, five CACCC boxes and numerous CCAAT boxes) was ligated upstream of the thymidine kinase promoter (ptk-Int1-LUC). Plasmid clones containing intron 1 in the native orientation were isolated and tested in transient transfection assays. The plasmid constructs, ptk-Int8(460)-LUC and ptk-Int8(460)R-LUC containing 460 bp of intron 8 sequence ligated upstream of the thymidine kinase promoter in the native and reverse orientations, respectively, (Section 2.4.2) were also examined.

These plasmid constructs were co-transfected with pRSV-βgal and transiently expressed in K562, MEL and COS-1 cells as described previously. As seen in Fig. 4.2B, the plasmid construct pALASp-Int1-5'Int8-LUC, transactivated the ALAS2 promoter approximately 7.8-fold in K562 cells. Similarly, the plasmid constructs, containing intron 8 ligated downstream of the luciferase reporter gene in the native (pALASp-Int1-3'Int8-LUC) or reverse orientation (pALASp-Int1-3'Int8R-LUC) transactivated the ALAS2 promoter 8.6 and 8.9 -fold, respectively, in K562 cells (Fig. 4.2B). The fold induction obtained with these constructs fall between the 3.2 and 12 -fold obtained when intron 1 and intron 8 were tested individually in K562 cells, while the addition of intron 1 consistently reduced

expression in COS-1 cells approximately 4-fold, irrespective of the presence, location or orientation of intron 8 (see Fig. 4.2B).

As seen in Fig. 4.3, the addition of intron 1 also increased the transcription from the thymidine kinase promoter (ptk-Int1-LUC) approximately 6-fold in K562 cells and only marginally (1.27-fold) in COS-1 cells. In the natural orientation (ptk-Int8(460)-LUC), intron 8 transactivated the thymidine kinase promoter approximately 25 and 12 -fold in K562 and MEL cells, respectively, but only 1.8-fold in COS-1 cells (Fig. 4.3). In contrast, intron 8 in the reverse orientation (ptk-Int8(460)R-LUC) displayed enhancer activity but this was considerably reduced to 4 and 2 -fold in K562 and MEL cells, respectively, (Fig. 4.3) indicating a possible orientation dependence.

These results demonstrated that sequences located within ALAS2 intron 1 and intron 8 can increase expression of either the native ALAS2 promoter or a heterologous promoter in an erythroid cell-specific fashion. However, these two intronic sequences did not function cooperatively and the relevance of these transient expression studies to the *in vivo* situation is unknown. Due to the small size of intron 8 compared with intron 1, and the large increase in reporter gene expression obtained with the addition of intron 8 alone, the sequences located within intron 8 that were responsible for high level erythroid-specific expression were investigated further. Studies described in Chapter 6, involve the localisation of the erythroid-specific enhancer activity to a 235 bp fragment within intron 8 and the subsequent identification of the elements that constitute this enhancer.

4.2.3 Transactivation of Human ALAS2/Luciferase Reporter Gene Constructs by GATA-1

As described in Section 4.1, numerous putative GATA-1 binding sites were identified in the ALAS2 promoter, intron 1 and intron 8. Therefore to investigate the putative role of GATA-1 in the erythroid-specific regulation of the human ALAS2 gene, the ability of exogenously expressed murine GATA-1 to transactivate the ALAS2 promoter, intron 1 and intron 8 was investigated in co-transfection assays.

Co-transfection experiments were performed initially in K562 cells, with 2 pmol of the control plasmids, pRSV-LUC or ptk-LUC, and pALASp-LUC containing the ALAS2 promoter and 5μg of the GATA-1 cDNA expression clone, pXM/GF-1. GATA-1

Figure 4.3   Transactivation of the thymidine kinase promoter by ALAS2 intron 1 and intron 8

Plasmid constructs containing human ALAS2 intron 1 or intron 8 ligated upstream of the thymidine kinase (tk) promoter were co-transfected with the plasmid, pRSV-βgal in K562, MEL and COS-1 cells. The size and orientation (*arrowed*) of these intronic sequences are indicated. Luciferase activities were standardised relative to β-galactosidase activity as an internal control for transfection efficiency. The normalised luciferase activities are expressed relative to ptk-LUC, assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not tested in a particular cell line.

	Fold induction		
	K562	**MEL**	**COS-1**
ptk-LUC	1.0	1.0	1.0
ptk-Int1-LUC	6±0.2	ND	1.27±0.1
ptk-Int8(460)-LUC	25±4	12±1.5	1.8±0.2
ptk-Int8(460)R-LUC	4±1.3	2±0.5	ND

transactivated both the Rous sarcoma viral and thymidine kinase promoters 6 and 20 -fold, respectively, but the ALAS2 promoter only 2.5-fold (data not shown). The minor transactivation of the ALAS2 promoter by GATA-1 observed in K562 cells is possibly attributable to interference by endogenous GATA-1, although the Rous sarcoma viral and thymidine kinase promoters were transactivated by GATA-1 and the reason for this is not known. Therefore to eliminate the possible interference by endogenous GATA-1 in K562 cells, co-transfection experiments with exogenously expressed GATA-1 were subsequently performed in COS-1 cells and consisted of 2 pmol of the reporter plasmid and 5μg of the GATA-1 cDNA expression clone, pXM/GF-1 as described in Section 2.6.3. The plasmid pRSV-βgal was not included as an internal control for transfection efficiency in these co-transfection experiments, since initial experiments demonstrated that exogenously expressed GATA-1 also transactivated the Rous sarcoma viral promoter. The plasmid pGL2-Basic was therefore included as a control. Cells were harvested 24 hours following transfection and 100μg of total protein assayed for luciferase activity as described in Section 2.6.4. The fold transactivation obtained with exogenously expressed GATA-1 was calculated following subtraction of the activity obtained with pGL2-Basic (approximately 1.2-fold).

As shown in Fig. 4.4, GATA-1 transactivated the plasmids, pALASp-LUC and pALASp-Int1-LUC, 4 and 9.5 -fold, respectively, in COS-1 cells. Interestingly, the plasmid pALASp-Int8(460)-LUC containing intron 8 was also only transactivated 3.6-fold by GATA-1 (Fig. 4.4). Test plasmids containing intron 1 and intron 8 cloned either upstream (pALASp-Int1-5'Int8-LUC) or downstream (pALASp-Int1-3'Int8-LUC and pALASp-Int1-3'Int8R-LUC) were transactivated 8.5, 9.4, and 10 -fold, respectively, by GATA-1 irrespective of the orientation of the intron 8 fragment (Fig. 4.4). Therefore the increase in expression observed with exogenously expressed GATA-1 in COS-1 cells is due to sequences located in the promoter and intron 1 but not within intron 8.

In conclusion, exogenously expressed murine GATA-1 is capable of transactivating both the ALAS2 promoter and sequences located within intron 1. Studies aimed at identifying the binding sites through which GATA-1 transactivated the ALAS2 promoter are described in the following chapter. However, despite the identification of four putative GATA-1 binding sites within intron 8, this region cannot be transactivated by GATA-1 in non-erythroid COS-1 cells. Consequently, the erythroid-specific enhancer function of intron

Figure 4.4   **Transactivation of human ALAS2 promoter/intronic constructs by GATA-1**

Plasmid constructs containing either the human ALAS2 promoter, intron 1 and intron 8 were co-transfected with 5μg of the murine GATA-1 cDNA expression clone, pXM/GF-1 in COS-1 cells and luciferase activities determined. The location and orientation (*arrowed*) of intron 8 are indicated. The data are averages obtained from constructs tested in triplicate in at least three experiments and are represented as the mean ± standard deviation. pGL2-Basic was included as a control for transactivation by GATA-1 and assigned a value of 1.0.

Construct	Fold induction with GATA-1 in COS-1 cells
pALASp-LUC	4.0±0.4
pALASp-Int1-LUC	9.5±0.3
pALASp-Int8(460)-LUC	3.6±0.2
pALASp-Int1 5'Int8-LUC	8.5±0.6
pALASp-Int1 3'Int8-LUC	9.4±0.1
pALASp-Int1 3'Int8R-LUC	10.1±0.05
pGL2-Basic	1.0

8 may be independent of GATA-1 or require the interaction of GATA-1 with additional erythroid-specific transcription factors. An investigation into the critical elements responsible for intron 8 enhancer activity are described in Chapter 6.

4.3   DISCUSSION

In this chapter, two aspects of ALAS2 gene regulation were examined, promoter expression and the identification of enhancer regions within the ALAS2 gene responsible for tissue-specific expression. The results obtained from the transient expression studies with human ALAS2 intronic sequences are discussed in relation to the LCR of the β-globin gene cluster which serves as a model in understanding promoter-enhancer interactions.

Deletion analysis of the human ALAS2 gene 5'-flanking region demonstrated that the first 300 bp of promoter sequence contains the information required to direct strong transient expression in erythroid cells. A detailed analysis of the elements in this region are described in Chapter 5. However, ALAS2 promoter deletion constructs from -10.3 kb to -293 bp expressed efficiently in COS-1 cells suggesting that the erythroid-restricted expression of the ALAS2 gene must rely on additional regulatory elements located elsewhere or native chromatin structure. Similar observations have been made for the promoters of other erythroid cell-specific genes including the rabbit α-globin (James-Pederson *et al.*, 1995), murine cytosolic glutathione peroxidase (O'Prey *et al.*, 1993), human Wilm's tumour gene (WT1) (Fraizer *et al.*, 1994) and the proximal promoter of the human glycophorin B gene (Rahuel *et al.*, 1992) when examined in non-erythroid cells. The question therefore arises as to how the ALAS2 gene is silenced in non-erythroid cells, *in vivo*. Deletion analysis of the ALAS2 5'-flanking region did not indicate that the ALAS2 promoter was negatively controlled, such as that reported for the human glycophorin B promoter where a repressor localised at position -75 is displaced by GATA-1 in erythroid cells (Rahuel *et al.*, 1992).

The human ALAS2 promoter lacks a canonical TATA box but located at -30/-23 is the sequence (5'-GGATAAAT-3') which resembles a binding site for GATA-1 and a detailed analysis of this site is described in Chapter 5. Interestingly, other erythroid-specific genes including chicken β-globin (Fong & Emerson, 1992), human glycophorin B (Rahuel *et al.*, 1992) and rat pyruvate kinase (Max-Audit *et al.*, 1993) all lack a consensus TATA

box but instead contain a GATA motif, suggesting that GATA-1 may play a role in erythroid-specific expression possibly by preventing the assembly of nucleosomes at this site, thereby allowing access of transcription factors required for transcription initiation (Barton *et al.*, 1993). Therefore expression of the ALAS2 promoter in non-erythroid cells may reflect an inadequate assembly of repressive nucleosomes on transiently transfected constructs. However, other regulatory regions of the ALAS2 gene may be involved in conferring tissue-specific expression of the gene *in vivo*. Erythroid cell-specific enhancers have been identified in the flanking regions of several erythroid-specific genes including the chicken β-globin (Fong & Emerson, 1992), human WT1 (Fraizer *et al.*, 1994) and the murine cytosolic glutathione peroxidase genes (O'Prey *et al.*, 1993).

The correlation between DNase I hypersensitive sites and important transcriptional regulatory elements has been well documented. The classical example is the LCR of the β-globin gene cluster consisting of five DNase I hypersensitive sites, four of which are erythroid-specific. However, the precise role of the LCR in β-globin gene expression is surrounded by controversy. Currently, there are two models; the "binary" and "DNA looping". The "binary" model (Walters *et al.*, 1995) proposes that the primary function of the LCR is to establish and/or maintain domains permissive for transcription rather than functioning to increase the rate of transcription (Walters *et al.*, 1995; Martin *et al.*, 1996; Milot *et al.*, 1996; Sutherland *et al.*, 1997). In contrast, the "DNA looping" model proposes that the LCR directly interacts with elements located within individual globin gene promoters of the β-globin gene cluster and increases the rate of gene transcription (as reviewed by Orkin, 1995a). Evidence supporting both models will be discussed.

Several lines of evidence obtained from transfected erythroid cell lines and studies performed in transgenic mice support the "binary" model (Walters *et al.*, 1995; 1996; Martin *et al.*, 1996; Sutherland *et al.*, 1997). Studies performed by Walters *et al.* (1995) demonstrated that the presence of a linked enhancer element (HS2 of the LCR or the SV40 enhancer) in both transient and stably transfected cells increases the number of cells actively expressing a reporter gene but not the level of expression in transcriptionally active cells. In addition to this, deletion of the HS2 site from a reporter construct stably integrated into K562 cells, significantly increases the rate at which expression of the β-geo reporter gene is silenced (Walters *et al.*, 1996). The erythroid-specific hypersensitivity site, HS-40 of the α-globin gene cluster (Higgs *et al.*, 1990) is also capable of functioning as an enhancer

element in both stable erythroid cell lines (Sharpe *et al.*, 1993a) and transiently transfected cells (Pondel *et al.*, 1992; Ren *et al.*, 1993) during erythroid differentiation. Sutherland *et al.* (1997) demonstrated that the addition of the HS-40 enhancer of the α-globin locus increased transcription of the *lacZ* reporter gene in transgenic mice by increasing the number of transcriptionally active cells rather than increasing the expression of the reporter gene in transcriptionally active cells.

The second model, "DNA looping", proposes that the role of the LCR is to directly interact with elements located within individual promoters of the β-globin gene cluster. The individual sites of the LCR of the β-globin locus have been assayed in a number of expression systems including transient expression assays, stably transformed MEL cell lines, transgenic mice and homologous recombination with varying results suggesting that these may have different functional capabilities (as reviewed in Orkin, 1995a; Martin *et al.*, 1996). However, it is difficult to imagine how these individual HS sites can interact with the promoters of the β-globin gene cluster and an alternative model suggests that these sites may function cooperatively or form a holocomplex which then interacts with promoters of the β-globin gene cluster (Fraser *et al.*, 1993). Using *in situ* hybridisation, it was possible to detect within a single cell, transcription of the γ-globin gene and the accumulation of β -globin mRNA in the cytoplasm indicating the previous transcriptional activity of the β -globin gene (Wijgerde *et al.*, 1995). These findings supported the "flip-flop" model where the individual globin genes compete for LCR function; the LCR activates globin genes one at a time, and these interactions are dynamic that is, constantly changing from one gene to another. Milot *et al.* (1996) demonstrated that deletion of a single HS site from the LCR in transgenic mice significantly reduced expression of a human globin transgene and resulted in the loss of position independence and copy number dependence. Therefore these results suggest that the individual HS sites of the LCR function cooperatively and the stability of the LCR-gene interaction may influence the level of globin gene expression by determining the frequency and duration of transcription periods (Milot *et al.*, 1996).

The LCR-globin gene interaction is thought to occur via DNA looping and evidence exists which implicates a possible role of NF-E2 in mediating such an interaction. Tandem binding sites for NF-E2 located within HS2 are essential for enhancer activity of the HS2 in the LCR (Ney *et al.*, 1990; Talbot *et al.*, 1991) and for the enhancer-dependent

transcription of the epsilon promoter which is mediated via a GATA-1 binding site in the promoter (Gong & Dean, 1993). A functional NF-E2 binding site has also been identified within HS-40 (Zhang et al., 1995), implicating a possible role for NF-E2 in enhancer-promoter interactions and supporting the "DNA looping" model.

Recent studies have implicated a direct role for NF-E2 in the formation of the DNase I hypersensitivity sites. The identification of binding sites for NF-E2 in each of the four hypersensitivity sites of the human LCR and sequence conservation in both the mouse and goat LCRs (Stamatoyannopoulos et al., 1995) alluded to a possible functional role. Armstrong and Emerson (1996) using an in vitro Drosophila chromatin assembly system demonstrated that purified NF-E2 protein was capable of forming a DNase I hypersensitive site at HS2 in vitro as observed in vivo. NF-E2 is also capable of disrupting local nucleosome organisation in a process that is ATP dependent and although other chromatin remodelling factors have been identified (Kwon et al., 1994; Cote et al., 1994; Tsukiyama & Wu, 1995) it is currently unknown whether NF-E2 can also functionally interact with these proteins. Interestingly, the disruption of nucleosomes by NF-E2 also facilitated the binding of GATA-1 to inverted GATA motifs located approximately 60 bp downstream of the NF-E2 binding site (Armstrong & Emerson, 1992). The association of NF-E2 and GATA motifs have been noted in all four hypersensitivity sites of the human LCR and both sites are required for the formation of HS4 in the human globin LCR (Stamatoyannopoulos et al., 1995). Similarly positioned sites have also been identified in the murine HS2, 3 and 4, goat HS2 and 3 and chicken β-globin enhancer (Stamatoyannopoulos et al., 1995), suggesting that a conserved mechanism may be utilised to create these hypersensitive sites among species. However, whether the role of NF-E2 is simply to disrupt nucleosomes at the LCR and permit expression of downstream globin genes or has a dual function in both disrupting nucleosomes and allowing enhancer-promoter interactions, permitting the access of other transcription factors and stimulating globin gene expression remains to be elucidated. Evidence supporting both models does exist and information obtained from understanding the role of the LCR in globin gene regulation may serve as a general model applicable to other erythroid cell-specific genes.

DNase I hypersensitive sites have also been identified in the promoters or enhancer regions of numerous erythroid cell expressed genes including the promoter of the erythroid carbonic anhydrase I gene (Sowden et al., 1992), the erythroid-specific enhancer in the

L-type pyruvate kinase gene (Lacronique *et al.*, 1995), the 3' enhancers of the human (Groudine *et al.*, 1983) and chicken (Choi & Engel, 1986; Hesse *et al.*, 1986) β-globin genes and the 3' enhancer of the murine cytosolic glutathione peroxidase gene (O'Prey *et al.*, 1993). These DNase I hypersensitive sites correspond to regions of the gene responsible for tissue-specific expression. Interestingly, putative binding sites for NF-E2 have been identified in the promoter of the erythroid carbonic anhydrase I gene (Sowden *et al.*, 1992), the 3' enhancers of the human and chicken β-globin genes (Groudine *et al.*, 1983; Choi & Engel, 1986; Hesse *et al.*, 1986), and the murine cytosolic glutathione peroxidase gene (O'Prey *et al.*, 1993) suggesting the possible involvement of NF-E2 in the formation of these hypersensitivity sites. Therefore based on the correlation between the occurrence of DNase I hypersensitivity sites and their role in transcriptional regulation, the identification of several DNase I hypersensitivity sites in the murine ALAS2 gene and the structural similarities between the human and murine ALAS2 genes, the corresponding regions of the human ALAS2 gene were examined for tissue-specific expression.

Studies described in this chapter demonstrated that ALAS2 intron 1 moderately transactivated the ALAS2 and thymidine kinase promoters in erythroid cells while it reduced expression of the ALAS2 promoter in non-erythroid COS-1 cells. The reason is unknown but may suggest the presence of regulatory sequences located within intron 1 capable of conferring tissue-specific expression. Sequence analysis identified sixteen putative binding sites for GATA-1 that were dispersed throughout the 4.973 kb fragment with thirteen of these GATA sites found within the 3.9 kb intron 1 fragment which was capable of transactivating the thymidine kinase promoter in erythroid cells. In addition, five putative consensus CACCC boxes and numerous putative CCAAT boxes were also identified within the 3.9 kb intron 1 fragment. However, the functional roles of these putative transcription factor binding sites within intron 1 were not investigated further.

The identification of positive regulatory elements located within intron 1 that are involved in transcriptional regulation have also been identified in the rabbit (James-Pederson *et al.*, 1995) and human (Brickner *et al.*, 1991) α-globin genes, the human IFN-γ (Hardy *et al.*, 1985) and HLA-DRα genes (Peterlin *et al.*, 1987), the quail muscle-specific troponin I gene (Konieczny *et al.*, 1987) and the human collagen α1(I) gene (Bornstein *et al.*, 1988). The addition of ALAS2 intron 1 reduced expression of the ALAS2 promoter in non-erythroid cells, suggesting the presence of negative regulatory elements located

within this intron. Tissue-specific silencers have also been identified within intron 1 of the human CD4 gene (Donda *et al.*, 1996) and the rat renin gene (Voigtlander *et al.*, 1995).

Preliminary transient expression studies demonstrated that intron 3 reduced the level of expression in both erythroid and non-erythroid cells suggesting the binding of negative regulatory elements to this region. In contrast, intron 3 did not appear to affect expression of a reporter construct driven by the thymidine kinase promoter (data not shown) indicating a possible interaction between sequences located within intron 3 and the ALAS2 promoter but this was not investigated further. Examination of the intron 3 sequence revealed a clustering of five putative binding sites for GATA-1, but only one of these sites conformed to the consensus GATA-1 motif (Merika & Orkin, 1993; Ko & Engel, 1993) as well as two putative CACCC boxes.

The most significant and interesting finding obtained from transient expression studies described in this chapter, was the identification of an erythroid-specific enhancer element capable of transactivating both the human ALAS2 and the heterologous thymidine kinase promoters. This enhancer element is located within intron 8 of the gene and as such would be located approximately 16 kb downstream of the ALAS2 promoter. Within the reporter constructs, intron 8 was positioned immediately adjacent to the ALAS2 or thymidine kinase promoters and whether intron 8 functions irrespective of the distance between the promoter and intron 8 needs to be investigated further. However, it would be extremely difficult to recapitulate the *in vivo* situation *in vitro* and determine the importance of distance between the ALAS2 promoter and intron 8. Interestingly, the erythroid-specific enhancer element transactivated the thymidine kinase promoter more efficiently in the native genomic orientation in contrast to a plasmid construct containing intron 8 in the reverse orientation. The possible orientation dependence of this enhancer suggests that transcription factors binding to this enhancer may only functionally interact with factors bound to the promoter in one orientation. However, we were unable to confirm this result with the ALAS2 promoter constructs due to the inability to isolate constructs containing intron 8 fragment in the reverse orientation.

As described in Chapter 6, there is extensive sequence conservation between the human, canine and the 3' end of the murine intron 8 sequence (Cox, 1993). This sequence conservation again suggests a possible functional role for the intron 8 sequence in contributing to ALAS2 gene expression. The elements located within intron 8 which are

responsible for erythroid-specific enhancer activity were characterised and these functional elements were found to be conserved between all three species examined.

Numerous tissue-specific enhancers have been identified in the 3'-flanking sequence of genes including the chicken β-globin gene (Fong & Emerson, 1992) and the murine cytosolic glutathione peroxidase gene (O'Prey *et al.*, 1993). Therefore the ALAS2 3'-flanking sequence which contained two putative consensus GATA-1 binding sites, two consensus CACCC boxes and an NF-E2-like sequence was also examined in transient expression studies. However, the addition of the ALAS2 3'-flanking region to the ALAS2 promoter did not increase luciferase expression in erythroid cells but rather resulted in a decrease in luciferase expression. This finding suggested that the 3'-flanking region did not contain any positive regulatory elements and was not investigated further.

Co-transfection assays performed with exogenously expressed GATA-1 and reporter constructs in non-erythroid cells, is one approach successfully used to investigate the role of GATA-1 in transactivating erythroid-specific promoters (Martin & Orkin, 1990; Evans & Felsenfeld, 1991; Zon *et al.*, 1991; Rahuel *et al.*, 1992; Crossley *et al.*, 1994a) and enhancers (Wu *et al.*, 1995). This method has also been used in the identification of critical DNA binding sites for GATA-1 within a regulatory region (Rahuel *et al.*, 1992). Exogenously expressed murine GATA-1 was found to be capable of transactivating both the ALAS2 promoter and sequences located within intron 1 and this correlated with numerous potential GATA-1 binding sites in both of these regions. However, similar transactivation assays were not performed with the thymidine kinase-ALAS2 intronic constructs since GATA-1 significantly increased the expression of the thymidine kinase promoter. Sequence analysis of the thymidine kinase promoter revealed numerous putative GATA sites through which exogenously expressed GATA-1 may transactivate this promoter. As described in Chapter 5, three putative GATA-1 binding sites were identified in the ALAS2 promoter and co-transfection assays were performed with exogenously expressed GATA-1 in COS-1 cells, to identify the specific sites through which GATA-1 transactivated the ALAS2 promoter. Sequence analysis identified approximately sixteen sites located within intron 1 which conform to the consensus binding site for GATA-1 (Merika & Orkin, 1993; Ko & Engel, 1993) with thirteen of these located within the 3.984 kb fragment of intron 1 which transactivated the thymidine kinase promoter. In the future, GATA-1 transactivation studies will need to be performed to determine which of

these sites are functional. In contrast, sequences within intron 8 were not transactivated by GATA-1 in non-erythroid cells, despite the identification of putative GATA-1 binding sites and possible explanations are discussed in Chapter 6. Similarly, an erythroid-specific enhancer element located within the 5'-flanking region of the EpoR gene and corresponding to a DNase I hypersensitivity site has been identified but not yet sequenced (Youssoufian, 1994). This element is inactive in non-erythroid COS-1 cells and is not transactivated by GATA-1 (Youssoufian, 1994).

In conclusion, examination of regions of the human ALAS2 gene which corresponded to DNase I hypersensitive sites in the murine ALAS2 gene was an extremely useful and successful approach in the identification of transcriptional regulatory sequences responsible for the tissue-specific expression of the human ALAS2 gene. Studies described in the following two chapters, involve the characterisation of *cis*-acting elements and the *trans*-acting factors that bind to these sites, that are required for the maximal expression of the ALAS2 promoter and the erythroid-specific enhancer element located within intron 8.

CHAPTER FIVE

IDENTIFICATION OF REGULATORY ELEMENTS AND CORRESPONDING TRANSCRIPTION FACTORS THAT DRIVE EXPRESSION OF THE ALAS2 PROMOTER

CHAPTER FIVE : IDENTIFICATION OF REGULATORY ELEMENTS AND CORRESPONDING TRANSCRIPTION FACTORS THAT DRIVE EXPRESSION OF THE ALAS2 PROMOTER

5.1   INTRODUCTION

As mentioned in Section 4.2.1, the first 300 bp of human ALAS2 promoter resulted in maximal expression of the luciferase reporter gene in transient expression studies. Sequence analysis of this region revealed potential binding sites for the erythroid-enriched transcription factors, GATA-1 (Evans & Felsenfeld, 1989; Tsai *et al.*, 1989) and NF-E2 (Mignotte *et al.*, 1989b, Andrews *et al.*, 1993a), CACCC (Kadonaga *et al.*, 1987; Kingsley & Winoto, 1992; Hartzog & Myers, 1993; Miller & Bieker, 1993; Crossley *et al.*, 1996) and CCAAT (Chodosh *et al.*, 1988) box binding proteins and the Ets family of proteins (reviewed by Wasylyk *et al.*, 1993). Interestingly, all of these sites are clustered within the first 140 bp from the transcription initiation site (Cox, 1993) (see Fig. 5.1).

Three putative GATA-1 binding sites were identified at -126/-121 (on the non-coding strand), -102/-97 and -30/-23 (see Fig. 5.1). The -124 GATA site (5'-AGATAA-3') conforms to the consensus for GATA-1 (Merika & Orkin, 1993; Ko & Engel, 1993), while the -100 site (5'-AGATA<u>C</u>-3') deviates by a single nucleotide. The sequence (5'-GGATAAAT-3') centred at -27 in the ALAS2 promoter represents a non-canonical TATA box that exhibits some similarity with a GATA motif. Similar sequences in this location have been identified in the promoters of other erythroid cell-specific genes such as the chicken β-globin (Fong & Emerson, 1992), rat pyruvate kinase (Max-Audit *et al.*, 1993) and human glycophorin B (Rahuel *et al.*, 1992). There is evidence that GATA-1 binding sites are linked functionally to nearby CACCC boxes in erythroid promoters (Merika & Orkin, 1995; Gregory *et al.*, 1996) and the sequence (5'-GGGTGGGTGGGG-3') located at -59/-48 in the ALAS2 promoter contains two putative overlapping CACCC boxes on the non-coding strand (see Fig. 5.1). Several transcription factors that bind *in vitro* to CACCC boxes have been identified and include Sp1 (Kadonaga *et al.*, 1987), Sp1-related proteins (Kingsley & Winoto, 1992), CAC C and CAC D (Harztog & Myers, 1993), BKLF (Crossley *et al.*, 1996), corresponding to the previously identified protein TEF-2/GT-IC (Xiao *et al.*, 1987) and the erythroid-enriched

protein, EKLF (Miller & Bieker, 1993). Two additional transcription factors demonstrated to play a critical role in erythroid-specific gene transcription during differentiation are those of the AP-1 leucine zipper family, NF-E2 (Mignotte *et al.*, 1989b) and the thyroid hormone receptor (Zenke *et al.*, 1990). Partially overlapping the CACCC sequence, is an NF-E2-like sequence at -49/-39 with a 9/11 match to the consensus NF-E2 binding motif (Andrews *et al.*, 1993a). The sequence from -21 to -6 exhibits similarity to a thyroid hormone responsive element with a 11/16 match to the to the ideal thyroid hormone responsive element (TRE) (Glass *et al.*, 1988). A putative binding site for the ubiquitous CCAAT box transcription factor, CP1 (Chodosh *et al.*, 1988) was identified at -90/-84 (5'-GGCCAAT-3') in the ALAS2 promoter and is identical to the functional CCAAT box located in the human β-globin promoter (deBoer *et al.*, 1988). The sequence located at -113 5'-GCAAGGAAGGGAC-3' (see Fig. 5.1) was initially identified on the basis of strong sequence homology to a functional enhancer element located in the adenovirus type 5 (Herbomel *et al.*, 1984) and polyoma virus (Hearing & Schenk, 1986) which was later found to bind members of the Ets family of transcription factors (Higashino *et al.*, 1993).

Numerous putative transcription factor binding sites for GATA-1 and CACCC box proteins are clustered within the first 140 bp of ALAS2 promoter, but no putative transcription factor binding sites have been identified in the sequence between -140 bp to -293 bp (Fig. 5.1). Analysis of 5'-flanking sequence from -293 bp to -10.3 kb revealed numerous putative binding sites for GATA-1, CACCC box binding proteins and Ets family of transcription factors. These sites were not further characterised since the studies described in this chapter were directed at investigating the elements responsible for the strong transient expression of the first 300 bp of the human ALAS2 promoter. Gel shift assays have been employed to investigate the specificity of protein-DNA interactions in the ALAS2 promoter and the functional contribution of such binding sites evaluated by site-directed mutagenesis and transient expression analysis of ALAS2 promoter/reporter gene constructs.

Figure 5.1   **Putative *cis*-acting elements in the promoter of the human ALAS2 gene**

The promoter sequence of the first 300 bp is represented in plain text whereas the sequence of exon 1 is in bold text. The transcription initiation site at +1 is indicated by the *arrow*. Sequences in the promoter with homology to the consensus GATA, Ets, NF-E2, CCAAT, CACCC and TATA box binding sites and the putative TRE and Inr are *boxed* and are clustered within 140 bp of promoter. The CACCC sequence contains two overlapping CACCC boxes. The GATA and CACCC boxes on the non-coding strand are *underlined* by *arrows*.

-300
TTGAGGAGATCTATAGTCAGAGAGGTGGTTTTGCTCAAGGTATCCCAG

-252
CTCTATGTGGATGGTCTGATTCCAAAGCCCAAATGAGCTAATTTTACT

-204
GTCCTATAGAGAGAAGAAAATAGGGAAGAGCCAGAGCTGGGGGAGGGG

-156
GTACACTAGAGGGAGGGGCTACTTTGGGTTTTATCTCTAGCAAGGAAG
GATA-1   Ets

-108
GGACTGAGATACCTTTGGGGCCAATGCAGGGCCAGGCCGGCAGAAGGC
     GATA-1   CCAAT

-60
AGGGTGGGTGGGGCTGAGTCAGAGGAGAAGGGATAAATGCCAGGTCCT
   CACCC   NF-E2   TATA-like   TRE-like

-12
AACCCAAGTACCCACCTGTCATTCGTTCGTCCTCAGTGCAGGGCAACA
              +1   Inr

5.2   RESULTS

5.2.1   Analysis of Putative GATA-1 Binding Sites in the Human ALAS2 Promoter

i)   *GATA-1 Protein Binds at -124 and -100 GATA Sites in the Promoter*

To investigate the putative GATA-1 binding sites identified in the human ALAS2 promoter, a series of complementary oligonucleotides were synthesised (see Section 2.1.7), radiolabelled, annealed and the binding of nuclear proteins to these probes was determined using gel shift assays (as described in Section 2.9). GATA-124 and GATA-100 probes corresponded to the putative GATA-1 binding sites centred at -124 and -100, respectively. A β-globin GATA-1 consensus (GATA-cons) probe was included as a control probe (Wall *et al.*, 1988).

To detect GATA-1 binding, the procedure of Partington *et al.* (in press) was employed to prepare nuclear extracts from K562, MEL and COS-1 cells, and COS-1 cells transfected with 10μg of the murine GATA-1 cDNA expression clone, pXM/GF-1 as described in Section 2.9.1. Complex formation was optimised by testing various binding buffers and concentrations of the synthetic competitor DNA, poly(dI-dC) (data not shown). Consequently, the binding reaction used in the detection of GATA binding proteins contained 5μg of nuclear protein, 2μg of poly(dI-dC) in 15μl of 25mM Hepes, pH 7.9 containing 60mM KCl, 7.5% glycerol, 0.1mM EDTA, 5mM $MgCl_2$, 0.75mM dithiothreitol and 2mM spermidine and incubated on ice for 10 minutes. Radiolabelled probe (1ng) was added to the reaction and incubated on ice for a further 30 minutes. The retarded nuclear protein complexes were then resolved on a 6% non-denaturing polyacrylamide gel as described in Section 2.9.3.

As seen in Fig. 5.2, a major protein complex was obtained with the GATA-124 probe (lanes 2 and 3) and GATA-cons probe (lanes 12 and 13) using nuclear extracts from K562 and MEL cells. A complex with the same mobility was detected with the GATA-100 probe although the intensity was reduced (lanes 7 and 8). Similar results were also observed with all three probes using nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 5, 10 and 15) but were not detected with nuclear extracts from mock-transfected COS-1 cells (lanes 4 and 9), although a minor band was observed with the GATA-cons probe (lane 14).

Figure 5.2   Gel shift analysis of the -124 GATA and -100 GATA sites

Radiolabelled double-stranded oligonucleotides containing the -124 GATA site (GATA-124 probe), the -100 GATA site (GATA-100 probe) and a consensus GATA-1 binding site (GATA-cons probe) were incubated with nuclear extracts from K562 (lanes 2, 7 and 12), MEL (lanes 3, 8 and 13) and COS-1 cells (lanes 4, 9 and 14) and COS-1 cells expressing recombinant GATA-1 (lanes 5, 10 and 15). Nuclear extract was omitted from lanes 1, 6 and 11. The retarded complex corresponding to GATA binding is indicated by the *arrow*.
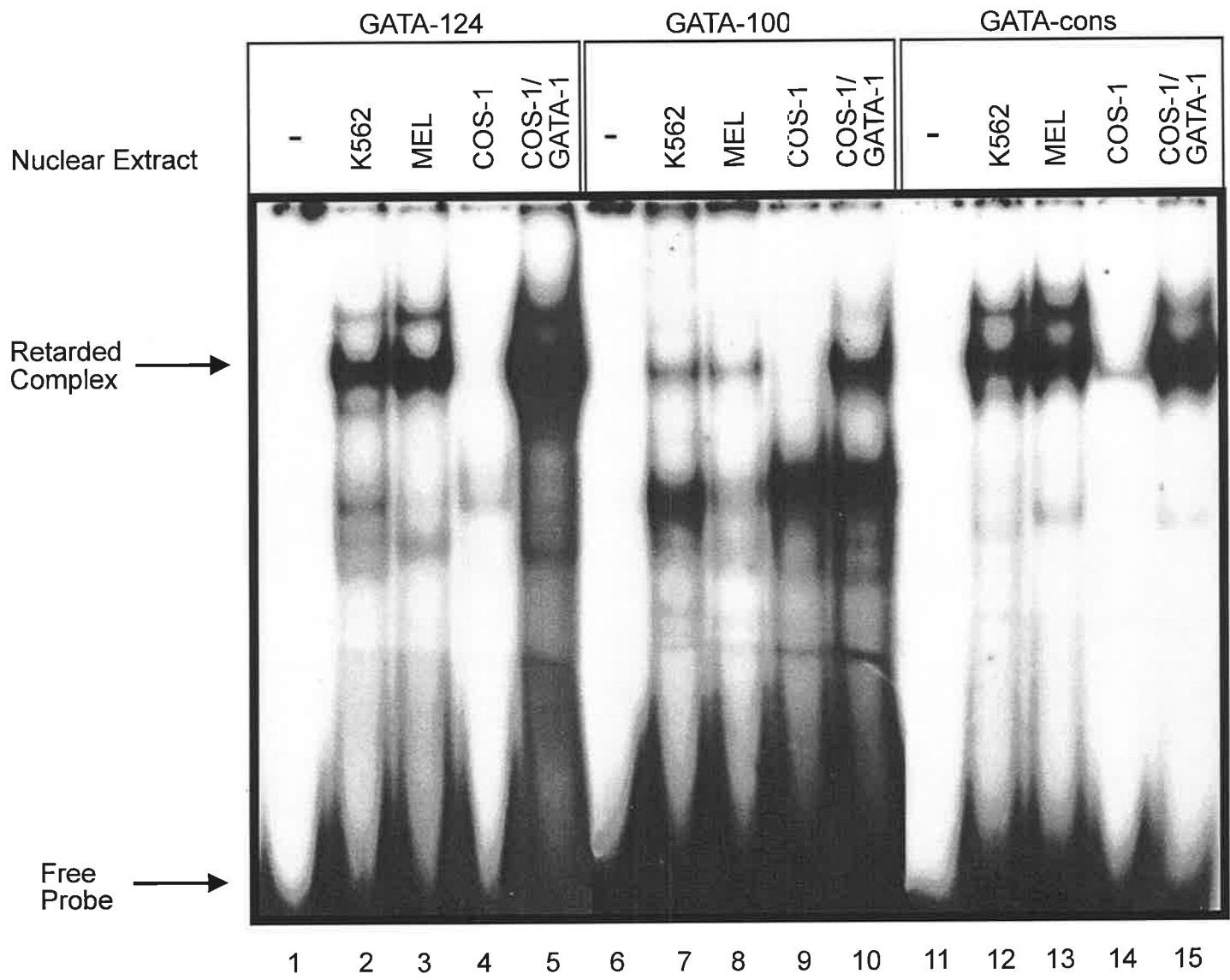
To confirm whether the protein complex that bound to the -124 and -100 sites in erythroid extracts was indeed GATA-1, gel supershift assays were undertaken with the GATA-1 monoclonal antibody, N-6 (Ito *et al.*, 1993) and nuclear extracts from either K562, MEL or COS-1 cells, or COS-1 cells expressing recombinant GATA-1. As seen in Fig. 5.3, the antibody substantially supershifted the major band obtained with the GATA-124 probe (compare lanes 2, 4 and 8 with lanes 1, 3 and 7), GATA-100 probe (compare lane 12 with lane 11) and the GATA-cons probe (compare lane 10 with lane 9). However, the GATA-1 antibody appeared to supershift the major retarded protein complex obtained in the MEL cell derived and COS-1 cells expressing recombinant murine GATA-1 nuclear extracts more efficiently compared with K562 nuclear extracts. This finding is probably due to the fact that the monoclonal antibody was generated against murine GATA-1 rather than human GATA-1.

Competition experiments were performed with the GATA-cons probe, nuclear extracts from COS-1 cells expressing recombinant GATA-1 and an increasing amount of unlabelled competitor oligonucleotide ranging from 10ng (10-fold excess) to 100ng (100-fold excess). As seen in Fig. 5.4, the binding of GATA-1 (lane 2) was effectively and specifically inhibited by a 10-fold molar excess of either GATA-cons in self competition (lane 3) or GATA-124 (lane 6) but a 50-fold molar excess of GATA-100 was required for a similar level of inhibition (lane 10). These experiments indicated that the affinity of the -124 GATA site for protein binding is comparable with the β-globin GATA-1 consensus sequence and greater than that of the -100 GATA site. The bands of higher and lower mobility relative to the major retarded complex that were competed in these experiments most likely represent a dimer of GATA-1 (Crossley *et al.*, 1995; Calligaris *et al.*, 1995) and degraded GATA-1 protein, respectively.

ii)   GATA-1 protein and TBP Bind to the -27 Site

The sequence (5'-GGATAAAT-3') centred at -27 in the human ALAS2 promoter (see Fig. 5.1) represents a non-canonical TATA box that exhibits some similarity with a GATA motif. Gel shift experiments were used to determine whether the sequence in the ALAS2 promoter binds GATA-1 and TATA-binding protein (TBP). Complementary synthetic oligonucleotides were synthesised containing the sequence 5'-GGATAAAT-3'

Figure 5.3   GATA-1 supershift assays

For supershift assays, the GATA-1 monoclonal antibody, N-6 was added to nuclear extracts from K562 (lane 2), MEL (lane 4) and COS-1 (lane 6) cells and COS-1 expressing recombinant GATA-1 (lane 8) prior to the addition of the radiolabelled GATA-124 probe. The GATA-cons probe was incubated with nuclear extracts from MEL cells (lanes 9 and 10) while the GATA-100 probe was incubated with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 11 and 12). The GATA-1 monoclonal antibody was added to nuclear extracts from MEL (lane 10) and COS-1 cells expressing recombinant GATA-1 (lane 12) prior to the addition of the radiolabelled GATA-cons and GATA-100 probes, respectively. The retarded complex in the absence of antibody and the supershifted complex are indicated.
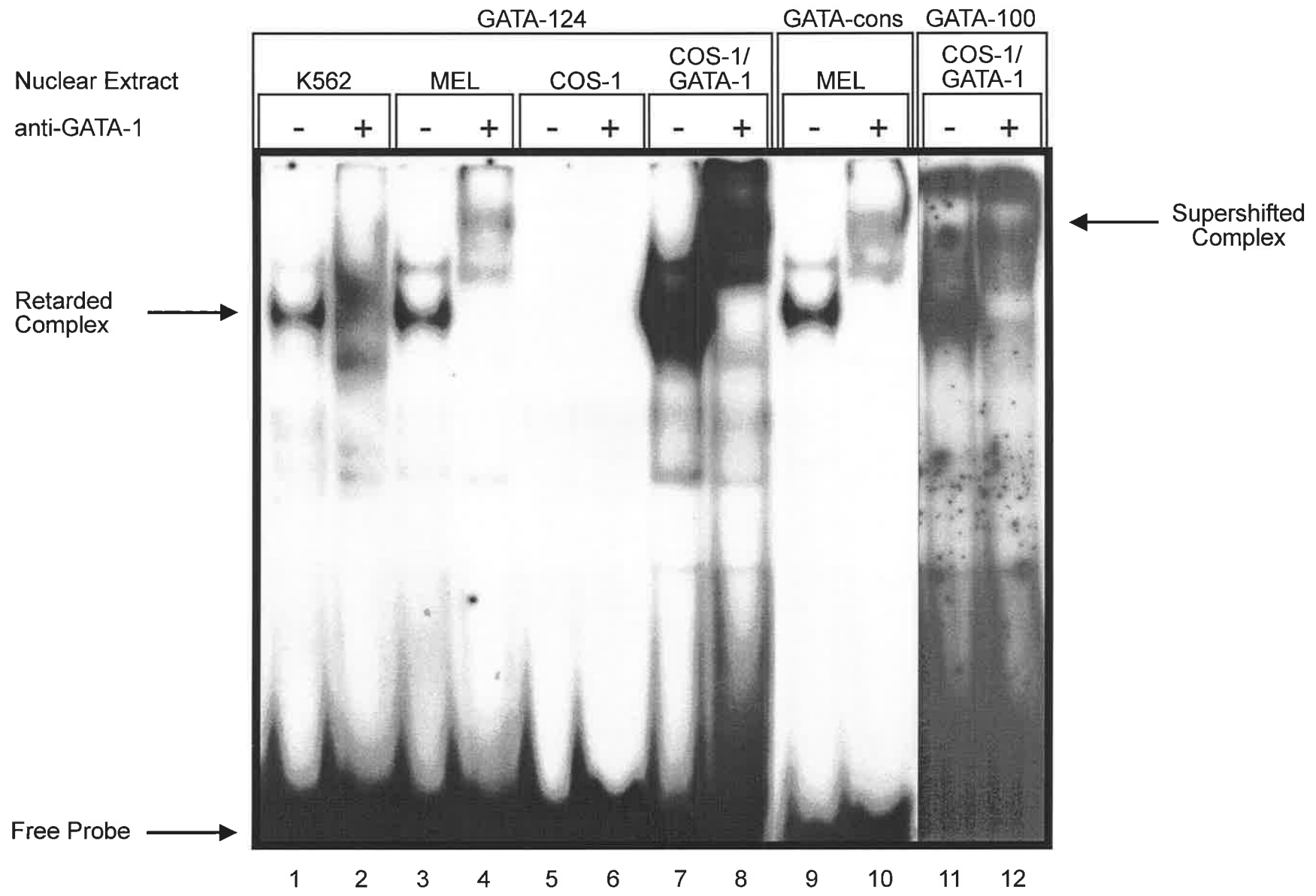
Figure 5.4   Oligonucleotide competition for GATA-1 binding in a gel shift assay

The radiolabelled GATA-cons probe was incubated with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 2-12). Nuclear extract was omitted from lane 1. The retarded complex (*arrowed*) was competed with a 10, 50 and 100 -fold molar excess of the GATA-cons oligonucleotide in self-competition (lanes 3, 4 and 5), the GATA-124 oligonucleotide (lanes 6, 7 and 8), the GATA-100 oligonucleotide (lanes 9, 10 and 11) and a 100-fold molar excess of a non-specific (NS) competitor (lane 12).

		Self			GATA-124			GATA-100			NS
Competitor	−	−	10x	50x	100x	10x	50x	100x	10x	50x	100x	100x

Retarded Complex →

Free Probe →

1   2   3   4   5   6   7   8   9   10   11   12
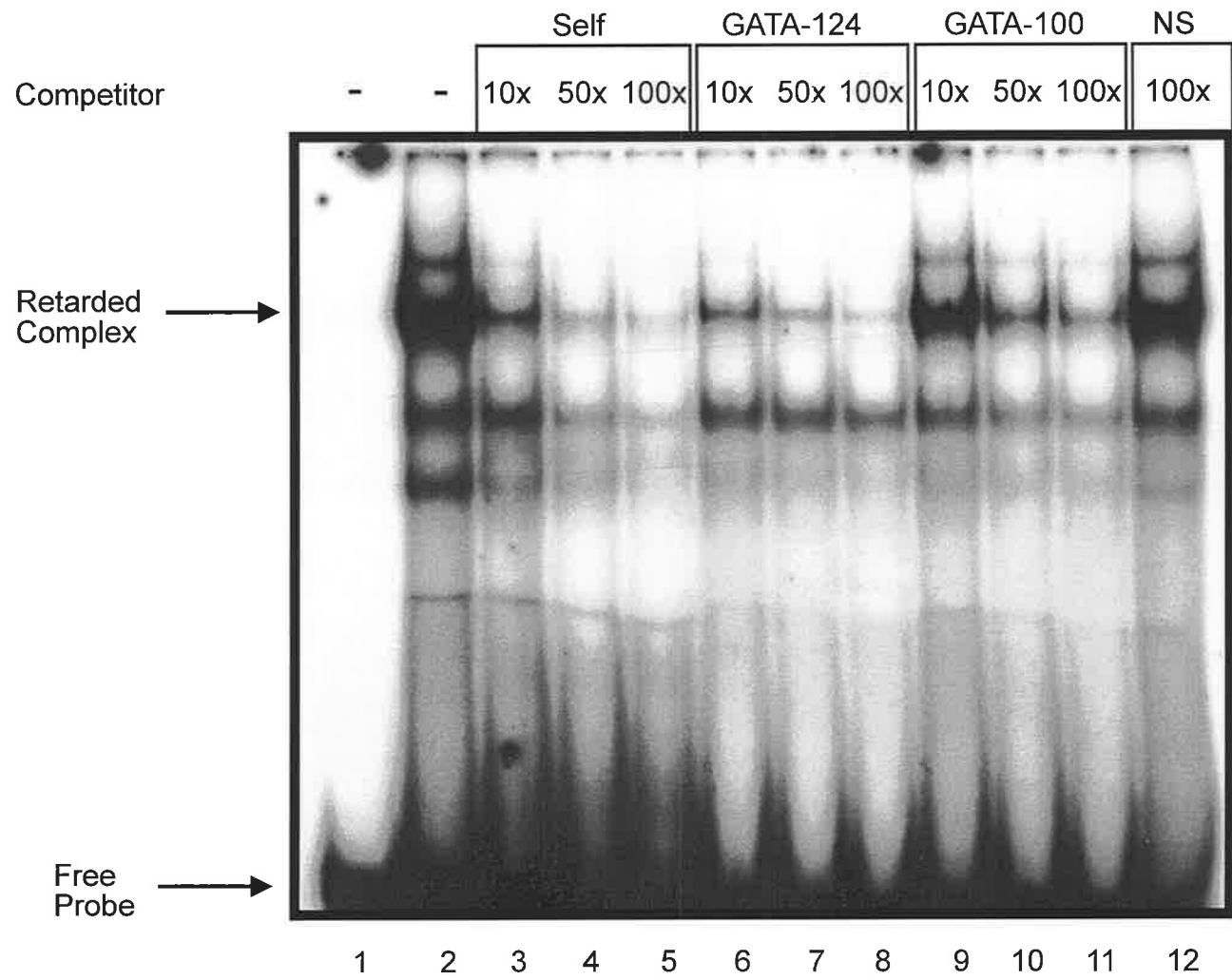
and 5'-GTATAAAT-3', corresponding to the native -27 site (GATA-27) and a canonical TATA box, respectively (see Section 2.1.7). These oligonucleotides were radiolabelled, annealed and then used in gel shift experiments (as described in Section 2.9).

As seen in Fig. 5.5A, a major retarded protein complex was observed following the incubation of the GATA-27 probe with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (Fig. 5.5A, lane 3) but not with nuclear extracts from mock-transfected COS-1 cells (lane 2). In competition experiments, the complex was abolished using a 50-fold molar excess of either the GATA-27 oligonucleotide as a self-competitor (lane 4) or the GATA-cons oligonucleotide (lane 6) but not with the canonical TATA box oligonucleotide (lane 5). The complex was identified as GATA-1 since it was supershifted with the GATA-1 monoclonal antibody and had the same mobility as the complex detected with the GATA-cons probe (data not shown). A retarded complex of similar mobility to GATA-1 was observed with the TATA probe and nuclear extracts from mock-transfected COS-1 cells (lane 8) or COS-1 cells expressing recombinant GATA-1 (lane 9) but the mobility of the complex was not affected with the GATA-1 monoclonal antibody (data not shown) and its identity is unknown.

In other experiments, the DNA binding affinity of GATA-1 for the -27 GATA sequence was compared with that of the β-globin GATA consensus site (GATA -27G) in gel shift assays using a purified GST-GATA-1(f) fusion protein (see Section 2.9.3). An increasing concentration of GST-GATA-1(f) was incubated with a constant amount of each probe and the extent of DNA binding determined. An approximately 20-40 fold difference in the concentration of protein required to give 50% DNA binding was observed, with GATA-1 exhibiting a higher binding affinity for the GATA-27G probe compared with the GATA-27 probe (data not shown).

To investigate whether the GATA-27 probe could also bind TBP, gel shift experiments were performed according to the binding reaction protocol described by Fong and Emerson (1992) (see Section 2.9.3). A specific protein complex was detected following incubation of the GATA-27 probe with recombinant human TBP (Fig. 5.5B, lane 2) and a corresponding complex was seen with the canonical TATA box probe (lane 4) but not with the GATA-27G probe (data not shown). These results demonstrate that in addition to GATA-1, the -27 non-canonical TATA box can bind TBP *in vitro*. The DNA binding affinity of TBP for the -27 GATA site was compared with that of the canonical TATA

Figure 5.5   Identification of proteins binding to the -27 GATA site

A   A radiolabelled double-stranded oligonucleotide containing the -27 GATA site (GATA-27 probe) was incubated with nuclear extracts isolated from COS-1 cells (lane 2) and COS-1 cells expressing recombinant GATA-1 (lanes 3-6). The major retarded complex (*arrowed*) was specifically inhibited with a 50-fold molar excess of the GATA-27 oligonucleotide in self-competition (lane 4) and the GATA-cons oligonucleotide (lane 6) but not with the canonical TATA box oligonucleotide (lane 5). The canonical TATA probe was incubated with nuclear extracts from COS-1 cells (lane 8) or COS-1 cells expressing recombinant GATA-1 (lane 9). Nuclear extracts were omitted from lanes 1 and 7.

B   The radiolabelled GATA-27 probe (lane 2) and the canonical TATA probe (lane 4) were incubated with recombinant TATA-Binding Protein (TBP). The major retarded complex is *arrowed*. TBP was omitted from lanes 1 and 3.

A

	GATA-27						TATA	
Nuclear Extract	–	COS-1	COS-1/ GATA-1				COS-1	COS-1/ GATA-1
Competitor	–	–	–	Self	TATA	GATA	–	–	–

Retarded Complex →

Free Probe →

1   2   3   4   5   6   7   8   9

B

	GATA-27		TATA	
TBP	–	+	–	+

Retarded Complex →

Free Probe →

1   2   3   4
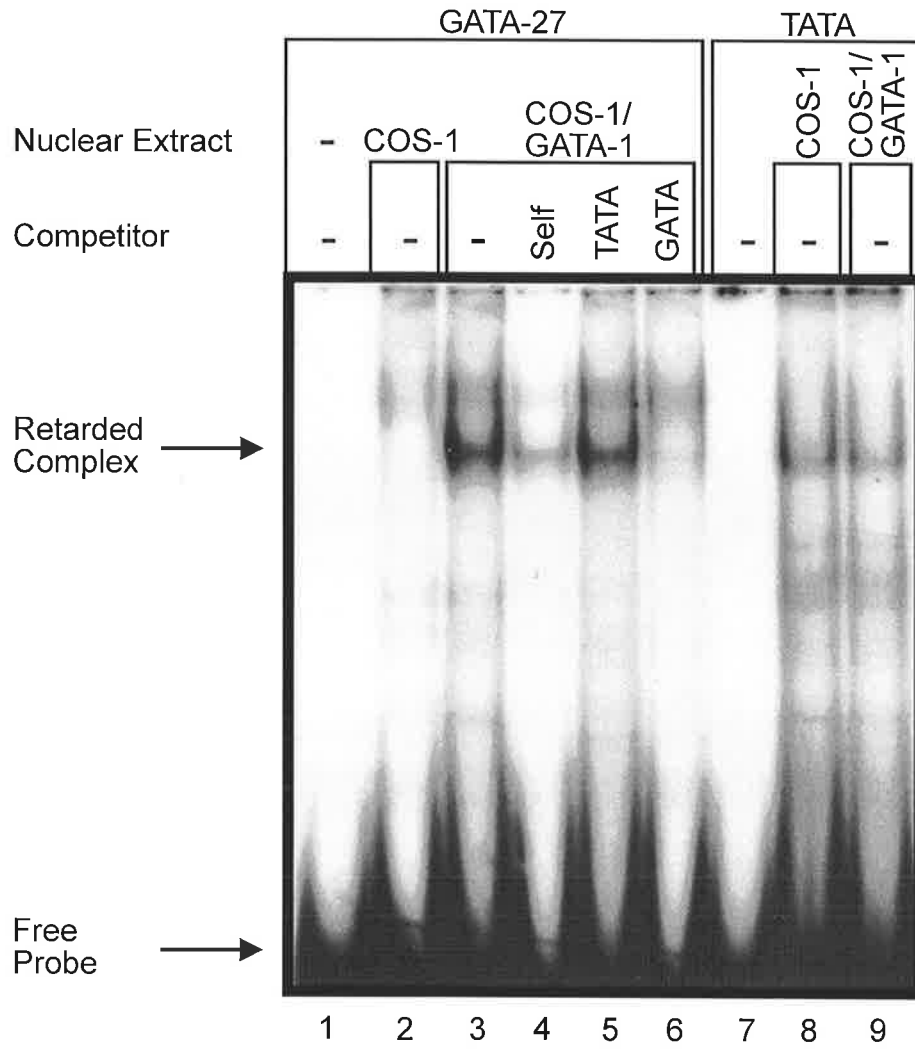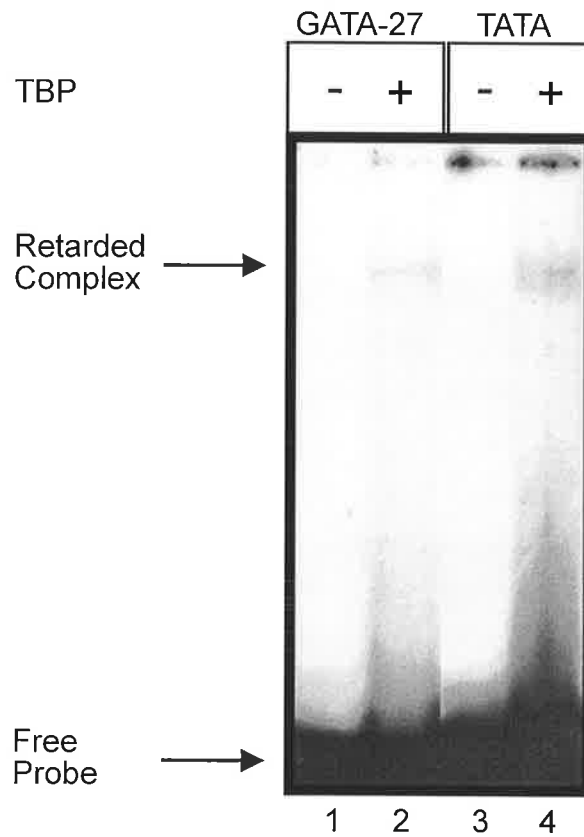
probe in experiments where an increasing concentration of purified TBP (ranging from 0.1nM to 100nM) was incubated with a constant amount of each probe. A specific retarded protein complex was detected with the TATA probe with 1nM of TBP but a corresponding complex was not observed with the GATA-27 probe over this range of TBP concentrations (data not shown). These data demonstrate that TBP has a weak affinity for the -27 GATA sequence in the ALAS2 promoter compared with a consensus TATA box.

In conclusion, these gel shift experiments demonstrated that the -27 non-canonical TATA box identified in the human ALAS2 promoter can bind either GATA-1 or TBP *in vitro*. However, this sequence binds both GATA-1 and TBP with a weak affinity compared with consensus binding sequences for GATA-1 and TBP.

iii)   *Mutational Analysis of the GATA-1 Binding Sites*

To investigate the functional contribution of the GATA-1 binding motifs identified in the ALAS2 promoter, these sites were inactivated by mutagenesis in the plasmid construct, pALAS-293-LUC, and expression analysed in K562, MEL and COS-1 cells (Fig. 5.6). Mutagenesis of either the -124 GATA site (pALAS-293mut1-LUC) or the -100 GATA site (pALAS-293mut2-LUC) reduced promoter expression relative to wild-type in K562 cells to 64% and 73%, respectively, and this was further reduced to 57% when both sites were mutated (pALAS-293mut3-LUC) (Fig. 5.6). In MEL cells, mutagenesis of these sites lowered expression to 36% and 78%, respectively, and a value of 34% was obtained when both sites were mutated. The reason for the greater contribution of the -124 GATA site in MEL cells compared with K562 cells is not known. These mutations had no effect when tested in COS-1 cells (Fig. 5.6), demonstrating the inactivity of these GATA-1 binding sites in non-erythroid cells.

To investigate the requirement for the -27 GATA site in transcription initiation, the sequence was mutated (5'-GGATAAAT-3' to 5'-GCAGCTGT-3') so that binding of both GATA-1 and TBP was abolished in gel shift assays (data not shown). Expression of the mutated promoter construct, pALAS-293mut4-LUC, was reduced to 36% of wild-type in both K562 and MEL cells (Fig. 5.6) and to 33% in COS-1 cells. The -27 GATA site was also converted to a sequence (5'-AGGGATAAAT-3' to 5'-CATGATAAG-3') which bound GATA-1 but not TBP in gel shift assays. This mutation (pALAS-293mut5-LUC)

Figure 5.6   Effect of mutating the GATA motifs on ALAS2 promoter expression

GATA sites located at -124, -100 and -27 in pALAS-293-LUC, were each mutated to a *Pvu*II site represented by a *cross*. The -27 GATA site was converted to a canonical TATA box and to a consensus GATA-1 binding site (*boxed*). These constructs were co-transfected with a β-galactosidase expression construct (pRSV-βgal) and transiently expressed in the K562, MEL and COS-1 cells. The normalised luciferase activities of the mutant constructs are expressed relative to pALAS-293-LUC which was set at 100%. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not tested in a particular cell line.

	Relative luciferase activity		
	K562	MEL	COS-1
pALAS-293-LUC	100	100	100
pALAS-293 mut1-LUC	64±3	36±6	96±4
pALAS-293 mut2-LUC	73±7	78±12	98±9
pALAS-293 mut3-LUC	57±12	34±3	102±6
pALAS-293 mut4-LUC	36±5	36±2	33±9
pALAS-293 mut5-LUC	30±3	ND	40±5
pALAS-293 mut6-LUC	81±6	67±5	132±6
pALAS-293 mut7-LUC	41±7	ND	137±4

reduced expression in K562 and COS-1 cells to 30% and 40%, respectively, compared to wild-type (Fig. 5.6).

The -27 binding site was also mutated to a canonical TATA box (5'-GGATAAAT-3' to 5'-GTATAAAT-3') which, in gel shift assays, bound TBP (Fig. 5.5B, lane 4) but not GATA-1 (Fig. 5.5A, lane 9). This mutation (pALAS-293mut6-LUC) consistently reduced promoter activity in K562 and MEL cells to 81% and 67%, respectively, compared with wild-type but increased expression in COS-1 cells to 132% (Fig. 5.6). Inactivation of the -124 GATA and -100 GATA sites and conversion of the -27 GATA site to a TATA box (pALAS-293mut7-LUC) reduced expression in K562 cells to 41% relative to wild-type. Hence, for maximal expression in transiently transfected erythroid cells, a non-canonical TATA box is required at the -27 position that can bind both GATA-1 and TBP *in vitro*.

iv)   ***GATA-1 Transactivates the ALAS2 Promoter in Non-erythroid Cells***

Co-transfection assays performed in non-erythroid COS-1 cells (as described in Section 4.2.3), demonstrated that exogenously expressed murine GATA-1 was capable of transactivating the ALAS2 promoter approximately 4-fold. To identify the sites through which GATA-1 transactivates the ALAS2 promoter, plasmid constructs containing mutated GATA motifs were examined in similar co-transfection assays in COS-1 cells (Fig. 5.7). Mutagenesis of either the -124 GATA site (pALAS-293mut1-LUC) or the -100 GATA site (pALAS-293mut2-LUC), reduced GATA-1 transactivation in COS-1 cells to 2.2 and 2.8 -fold, respectively, and this was further reduced to 1.4-fold when both sites were mutated in combination (pALAS-293mut3-LUC). Conversion of the -27 site to the canonical TATA box sequence (pALAS-293mut6-LUC) marginally reduced the level of transactivation to 3.2-fold (Fig. 5.7) and is consistent with the reduced activity of the same construct in K562 and MEL cells. Transactivation by GATA-1 was virtually abolished following the inactivation of the -124 GATA and -100 GATA sites and conversion of the -27 site to a canonical TATA box (pALAS-293mut7-LUC) (Fig. 5.7).

Figure 5.7   Transactivation of the ALAS2 promoter by GATA-1

ALAS2 promoter constructs with mutated GATA motifs were co-transfected with the murine GATA-1 cDNA expression clone, pXM/GF-1 in COS-1 cells and luciferase activities determined. The plasmid, pGL2-Basic was included as a control for transactivation by GATA-1 and assigned a value of 1.0. The data are averages obtained from constructs tested in triplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not examined.

Fold Induction with GATA-1 in COS-1 cells

Construct	Fold Induction
pALAS-293-LUC	4.0±0.4
pALAS-293 mut1-LUC	2.2±0.7
pALAS-293 mut2-LUC	2.8±0.8
pALAS-293 mut3-LUC	1.4±0.2
pALAS-293 mut4-LUC	ND
pALAS-293 mut5-LUC	ND
pALAS-293 mut6-LUC	3.2±0.1
pALAS-293 mut7-LUC	1.2±0.1
pGL2-Basic	1.0
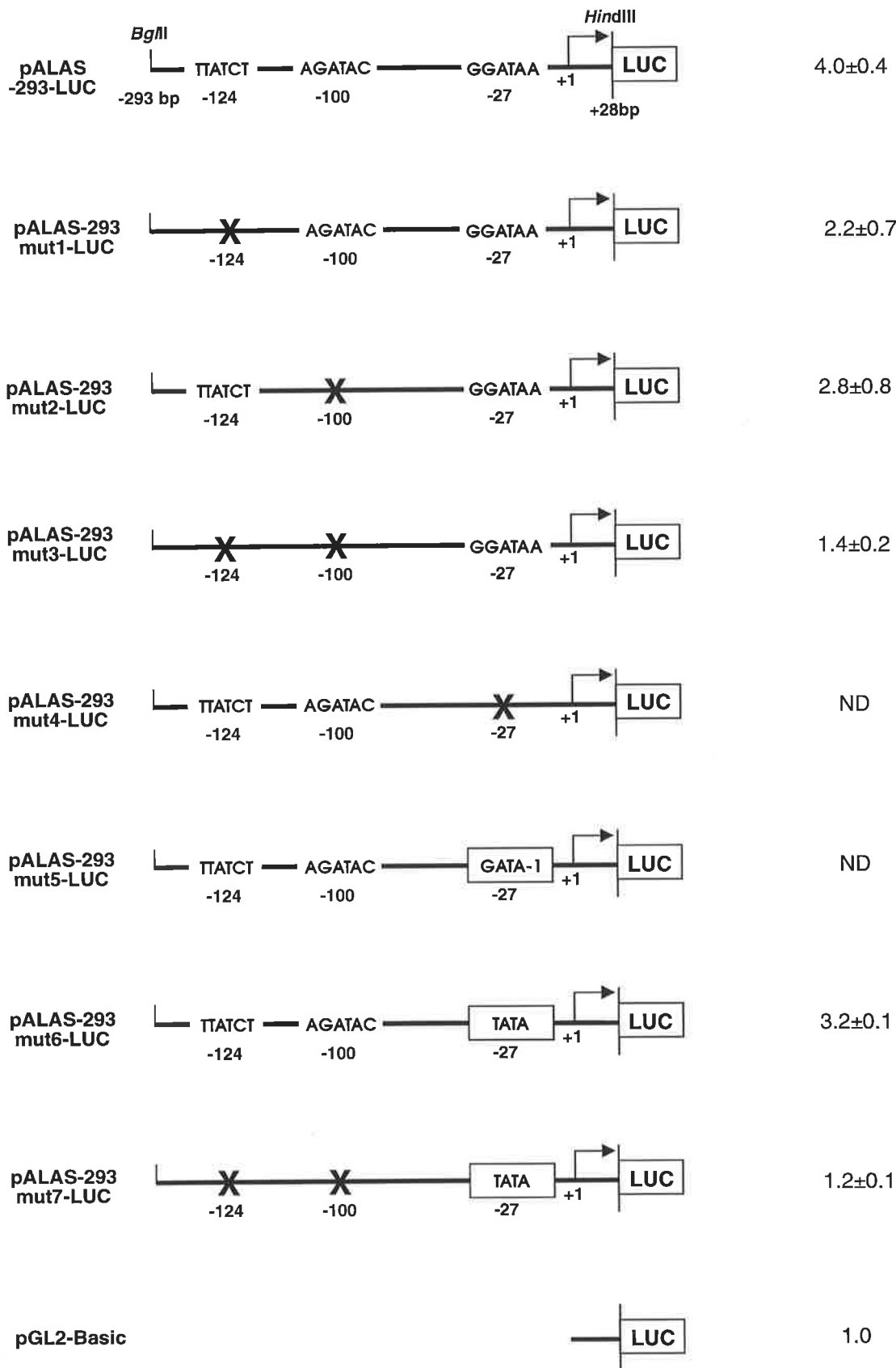
5.2.2 Analysis of a Putative CACCC Sequence in the Human ALAS2 Promoter

i) Sp1, EKLF and BKLF bind to the CACCC Sequence

The sequence (5'-GGGTGGGTGGGG-3') located at -59/-48 in the human ALAS2 promoter contains two putative overlapping CACCC boxes on the non-coding strand (Fig. 5.1). Currently, several transcription factors have been identified which bind *in vitro* to CACCC boxes and include ubiquitous Sp1 (Kadonaga *et al.*, 1987) and Sp1-related proteins (Hagen *et al.*, 1992; Kingsley & Winoto, 1992), CAC C and CAC D (Hartzog & Myers, 1993), BKLF (Crossley *et al.*, 1996) and the erythroid-enriched protein, EKLF (Miller & Bieker, 1993). To investigate the identity of the protein(s) capable of binding to the putative CACCC site identified in the human ALAS2 promoter, gel shift assays were performed. Two complementary oligonucleotide pairs were synthesised, the first pair contained the CACCC sequence from the ALAS2 promoter, referred to as the CACCC-54 probe and the second pair contained the previously characterised CACCC site from the murine adult β-globin promoter, capable of binding Sp1, EKLF and BKLF (Crossley *et al.*, 1996) and referred to as the β-globin CACCC probe (Section 2.1.7). These synthetic oligonucleotides were radiolabelled, annealed, and tested for protein binding ability with nuclear extracts derived from MEL and CV-1 cells or CV-1 cells transfected with 10μg of the murine EKLF cDNA expression clone, pSG5/EKLF (Miller & Bieker, 1993) in gel shift assays (as described in Section 2.9). Preliminary gel shift experiments with the CACCC-54 probe were performed using the same procedure as described for detection of GATA binding proteins. However, only two major retarded protein complexes were detected (data not shown) and a procedure described by Crossley *et al.* (1996) was adopted which permitted the identification of at least three CACCC binding proteins.

The CACCC-54 probe bound three major protein complexes from MEL cell nuclear extracts in gel shift assays (Fig. 5.8A, lane 2). Of these, the most rapidly migrating complex was identified as BKLF since an antibody to BKLF (lane 4) partially but specifically inhibited binding whereas antibodies to EKLF (lane 5) or preimmune serum (lane 3) had no effect. The slowest major migrating complex contained Sp1 and probably Sp1-related proteins since it was supershifted with an antibody to Sp1 (lane 6). However, the Sp1 antibody also partially inhibited binding to the second and third (BKLF) protein complexes.

The remaining major retarded complex was unaffected by the antibodies employed and its identity is unknown.

The CACCC-54 probe was also incubated with nuclear extracts from mock-transfected CV-1 cells (lane 7) and CV-1 cells expressing recombinant murine EKLF (lane 8). A complex of high mobility was observed only with nuclear extracts from cells expressing recombinant EKLF, and this complex was confirmed as EKLF using an antibody to EKLF (lane 11). The slowest migrating complex observed with mock-transfected CV-1 nuclear extracts was confirmed immunologically as Sp1 (lane 12). Together, the data demonstrate that the CACCC-54 probe can bind Sp1, BKLF and EKLF but the EKLF complex cannot be detected in the MEL cell nuclear extracts employed.

In similar gel shift experiments, Crossley et al. (1996) have shown using nuclear extracts from a different MEL cell line, that an EKLF-responsive CACCC box at -94/-87 in the promoter of the murine adult β-globin gene (Miller & Bieker, 1993) strongly binds BKLF and Sp1 but only weakly binds EKLF. For comparison, protein binding by this β -globin CACCC box (Miller & Bieker, 1993) using MEL F4-12B2 cell nuclear extracts was investigated (Fig. 5.8B). The results were almost identical to those observed with the CACCC-54 probe with major complexes detected for BKLF (lane 4) and Sp1 (lane 5) but no complex corresponding to EKLF (Fig. 5.8B). A complex of high mobility which was observed only with nuclear extracts prepared from cells expressing recombinant EKLF (lane 7) but not with mock-transfected CV-1 cell nuclear extracts (lane 6) was identified as EKLF using the EKLF antibody (lane 8). Apparently, there is insufficient EKLF in the MEL F4-12B2 cell nuclear extracts for detection by gel shift analysis using either CACCC probe. The slowest migrating complex detected with the β-globin CACCC and the Sp1 consensus probes in both MEL (lanes 2 and 11) and CV-1 nuclear extracts expressing recombinant EKLF (lane 7) was identified as Sp1 since it was supershifted with an antibody to Sp1 (lanes 5, 9 and 12).

Competition experiments were performed using the β-globin CACCC probe, nuclear extracts from CV-1 cells expressing recombinant EKLF and an increasing amount of unlabelled competitor oligonucleotide ranging from 5ng (5-fold excess) to 100ng (100-fold excess). As seen in Fig. 5.9, the CACCC-54 sequence and the β-globin CACCC box bind EKLF with similar affinities. The binding of EKLF to the β-globin CACCC probe (lane 3) was substantially reduced by competition with a 25-fold molar excess of either the β-globin

Figure 5.8   Gel shift analysis of the -54 CACCC sequence

A   Radiolabelled double-stranded oligonucleotides containing the -54 CACCC sequence (CACCC-54 probe) was incubated with nuclear extracts isolated from MEL (lanes 2-6) and CV-1 (lane 7) cells, and CV-1 cells expressing recombinant EKLF (lanes 8-12). Nuclear extract was omitted from lane 1. For supershift assays, preimmune (PI) serum (lanes 3 and 9), anti-BKLF (lanes 4 and 10), anti-EKLF (lanes 5 and 11) and anti-Sp1 (lanes 6 and 12) antibodies were added prior to the addition of the probe. The retarded complexes corresponding to Sp1, EKLF, BKLF and an unknown protein (?) are *arrowed*.

B   Radiolabelled double-stranded oligonucleotides containing the CACCC site from the murine adult β-globin promoter (β-globin CACCC probe) was incubated with nuclear extracts from MEL (lanes 2-5) and CV-1 (lane 6) cells, and CV-1 cells expressing recombinant EKLF (lanes 7-9). The Sp1-cons probe containing a consensus Sp1 binding site was incubated with nuclear extracts from MEL cells (lanes 11-12). Nuclear extracts were omitted from lanes 1 and 10. For supershift assays, preimmune (PI) serum (lane 3), anti-BKLF (lane 4), anti-EKLF (lane 8) and anti-Sp1 (lanes 5, 9 and 12) antibodies were added prior to the addition of probe.

A
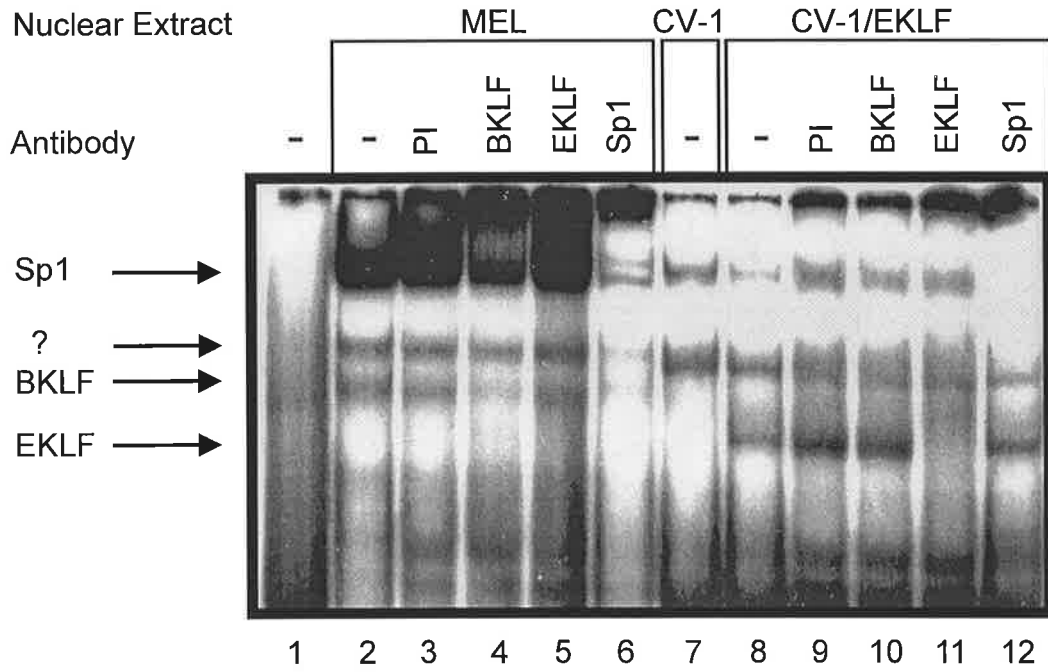
B
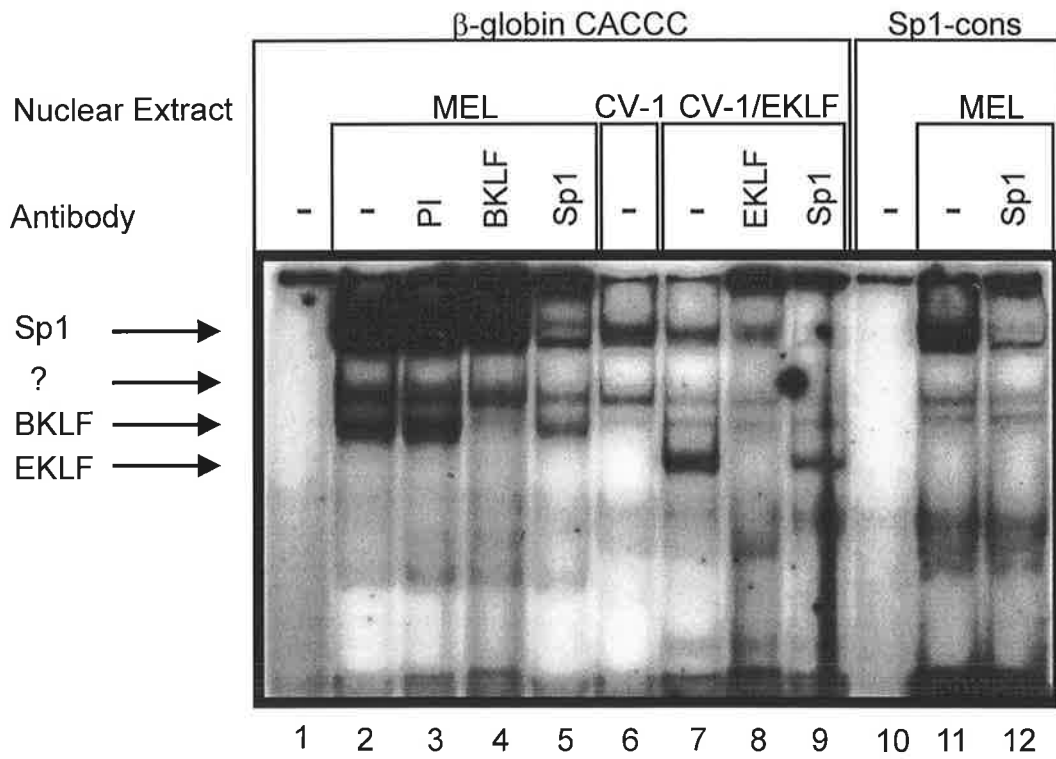
CACCC oligonucleotide in self competition (lane 6) or the CACCC-54 oligonucleotide (lane 9) but was not affected with a 100-fold molar excess of a non-specific competitor oligonucleotide (lane 10).

ii) *Mutational Analysis of the CACCC Sequence*

The -54 CACCC sequence was mutated (5'-GGGTGGGTGGGG-3' to 5'-GGCAGCTGGGGG-3') so that both of the constituent overlapping CACCC boxes were destroyed. As seen in Fig. 5.10, expression of this mutant promoter construct (pALAS-293mut8-LUC) in K562 and MEL cells was reduced to 59% and 46%, respectively, relative to pALAS-293-LUC, demonstrating the functional importance of this CACCC sequence. To investigate a possible cooperative interaction of CACCC-binding proteins with GATA motifs in the ALAS2 promoter (Frampton *et al.*, 1990; Merika & Orkin, 1995; Gregory *et al.*, 1996), two additional mutant promoter constructs were created. Mutagenesis of the CACCC sequence and the -124 GATA site (pALAS-293mut9-LUC), or a triple mutation of the CACCC sequence together with the -124 GATA and -100 GATA sites (pALAS-293mut10-LUC), reduced expression in K562 cells to 44% and 38%, respectively (Fig. 5.10). In COS-1 cells, expression of the promoter construct with only the CACCC sequence mutated (pALAS-293mut8-LUC) was markedly reduced to 33% relative to wild-type and mutations in the GATA sites did not further lower expression, establishing that promoter activity in these cells is driven predominantly by a CACCC binding protein, perhaps Sp1 or a Sp1-related protein.

iii) *Attempts to Identify the Transcriptional Activator at the -54 CACCC Site*

a) **EKLF transactivates the ALAS2 promoter**

To investigate the identity of the transcriptional activator at the -54 CACCC site in the human ALAS2 promoter, co-transfection assays were performed. Transactivation experiments have demonstrated that the functional CACCC site located in the murine adult β-globin promoter can respond transcriptionally to exogenously expressed EKLF in the non-erythroid CV-1 cell line (Miller & Bieker, 1993; Bieker & Southwood, 1995) and in the K562 cell line, which synthesises reduced levels of EKLF in comparison to MEL cells

Figure 5.9   Oligonucleotide competition for proteins binding to the β-globin CACCC site in a gel shift assay

The radiolabelled β-globin CACCC probe was incubated with nuclear extracts from CV-1 cells (lane 2) and CV-1 cells expressing recombinant EKLF (lanes 3-10). Nuclear extract was omitted from lane 1. The retarded complexes were competed with a 5, 10 and 25 -fold molar excess of the β-globin CACCC oligonucleotide in self competition (lanes 4, 5 and 6), the CACCC-54 oligonucleotide (lanes 7, 8 and 9) and a 100-fold molar excess of a non-specific (NS) competitor (lane 10). The retarded complexes corresponding to Sp1 and EKLF are *arrowed*.
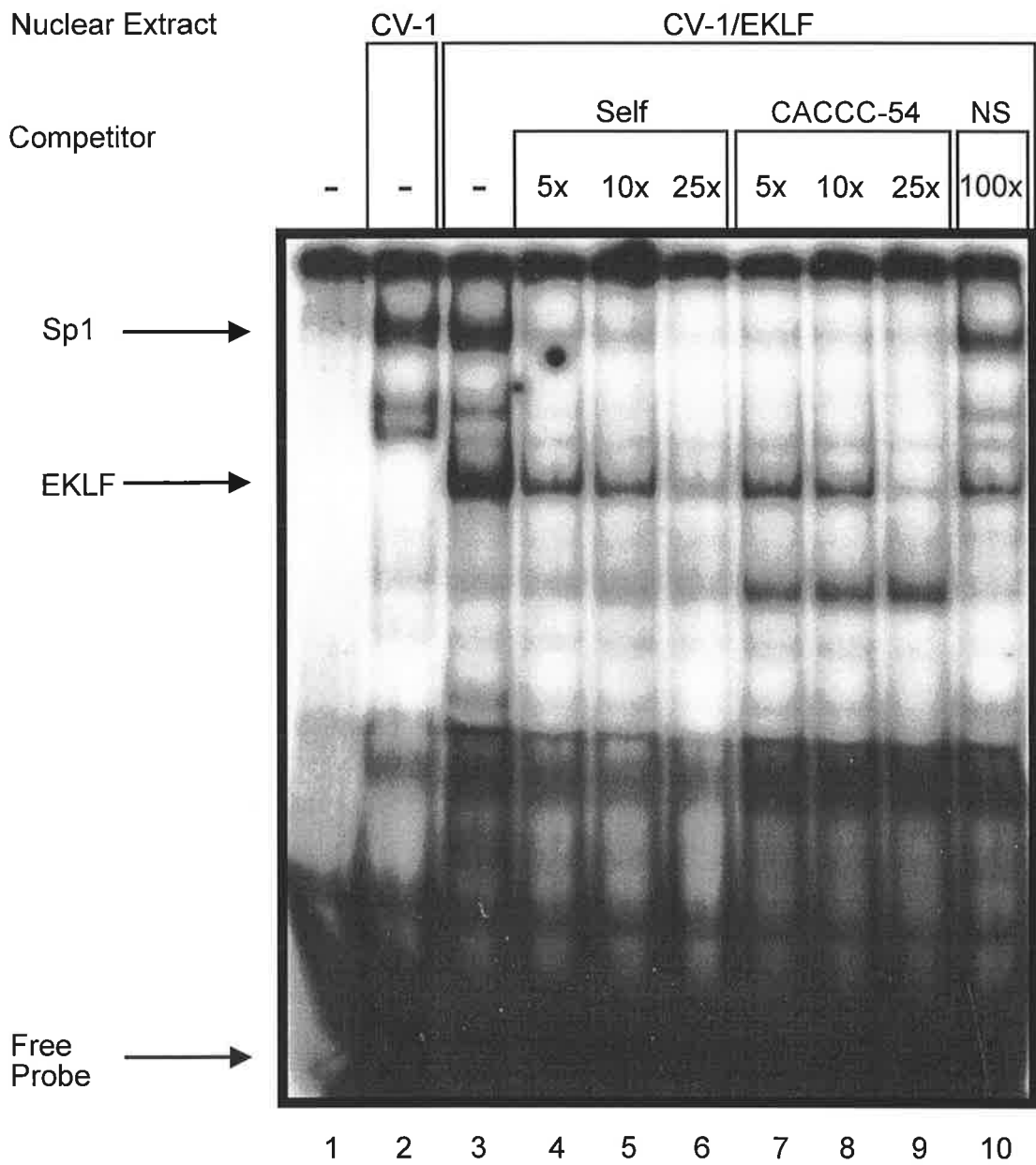
Figure 5.10   Effect of mutating the -54 CACCC sequence on ALAS2 promoter expression

The -54 CACCC sequence in pALAS-293-LUC was mutated to a *Pvu*II site. To examine the possible interaction of the CACCC sequence with GATA sites, the -124 GATA and -100 GATA sites were mutated in combination with the -54 CACCC sequence. Mutated sites are represented by a *cross*. Constructs were co-transfected with a β-galactosidase expression construct (pRSV-βgal) and transiently expressed in the K562, MEL and COS-1 cells. The luciferase activities were standardised relative to β-galactosidase activity as an internal control for transfection efficiency and expressed relative to pALAS-293-LUC (set at 100%). The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not tested in a particular cell line.

Relative luciferase activity

	K562	MEL	COS-1
pALAS-293-LUC	100	100	100
pALAS-293-mut8-LUC	59±12	46±11	33±8
pALAS-293-mut9-LUC	44±6	ND	36±2
pALAS-293-mut10-LUC	38±5	ND	34±3

(Donze *et al.*, 1995). Consequently, the level of transactivation of the ALAS2 promoter by exogenously expressed EKLF was compared to that obtained with the adult β-globin promoter.

The plasmid reporter constructs, pC1G3tk-CAT containing four copies of the functionally EKLF-responsive CACCC (-94/-87) from the adult β-globin promoter ligated upstream of the thymidine kinase promoter and pβ-glob-CAT containing the murine adult β-globin promoter (-168 bp to +37 bp), both ligated upstream of the CAT reporter gene were obtained from Dr. J. Bieker. However, due to the increased sensitivity of the luciferase reporter gene in comparison to the CAT reporter gene, plasmid constructs containing the luciferase reporter gene were synthesised as described in Section 2.4.5. These plasmid constructs, $p(CAC)_4$tk-LUC and pβ-glob-LUC were employed as positive controls in co-transfection assays with exogenously expressed EKLF. To eliminate the possibility of CACCC-like sequences being located upstream in the ALAS2 promoter, plasmid constructs with -124 bp of wild-type ALAS2 promoter (pALAS-124A-LUC) or a mutation in the -54 CACCC sequence (pALAS-124mut-LUC) were synthesised for use in transactivation experiments (refer to Section 2.4.5).

Co-transfection assays performed in CV-1 cells by Bieker and Southwood (1995), demonstrated that 2μg of the plasmid pC1G3tk-CAT and 10μg of the EKLF cDNA expression clone, pSG5/EKLF (Miller & Bieker, 1993) was required to transactivate the reporter construct approximately 25-fold. Preliminary co-transfection assays were performed with the plasmids, $p(CAC)_4$tk-LUC and pβ-globin-LUC (2μg) and pSG5/EKLF (10μg). The plasmid, $p(CAC)_4$tk-LUC was not transactivated but repressed approximately 10-fold by exogenously expressed EKLF in CV-1 cells. Similarly, pβ-glob-LUC was also repressed by exogenously expressed EKLF in CV-1 cells. In contrast, examination of these constructs containing the CAT reporter gene demonstrated that pC1G3tk-CAT was transactivated approximately 5-fold by EKLF in CV-1 cells but expression of plasmid pβ-glob-CAT was unaffected with exogenously expressed EKLF (data not shown).

Donze *et al.* (1995) have performed transactivation experiments with exogenously expressed EKLF in K562 cells and similar studies were undertaken in this cell line with the ALAS2 promoter. Co-transfection assays were optimised and consisted of 2 pmol (approximately 7.5μg) of the luciferase reporter constructs and 7.5μg of pSG5/EKLF. The

pRSV-βgal plasmid that served as an internal control for transfection efficiency in earlier experiments was not was included in these transactivation experiments since numerous promoters contain CACCC binding sites. Therefore the possibility existed that EKLF may also transactivate the Rous sarcoma viral (RSV) promoter and increase expression of the β-galactosidase reporter gene and this was observed in preliminary transactivation experiments (data not shown). The plasmids, pGL2-Basic and ptk-LUC, containing a 164 bp fragment of the thymidine kinase promoter were included as controls. The fold transactivation was determined following subtraction of the background activity obtained with pGL2-Basic or ptk-LUC and assigned a value of 1.0.

As seen in Fig. 5.11, the plasmid, pALAS-124A-LUC containing 124 bp of ALAS2 promoter was consistently induced 3.1-fold by exogenously expressed EKLF in K562 cells. Mutagenesis of the -54 CACCC sequence (pALAS-124mut-LUC) reduced this to 1.8-fold indicating that EKLF can function through this site. Transactivation of the ALAS2 promoter by EKLF was compared with pβ-glob-LUC and p(CAC)$_4$tk-LUC which were transactivated 4.0-fold and 9.5-fold, respectively, by EKLF in K562 cells (Fig. 5.11).

In conclusion, the -54 CACCC sequence located in the human ALAS2 promoter can respond transcriptionally to exogenously expressed EKLF but gel shift analysis demonstrated that this site was capable of binding not only EKLF but at least two other CACCC box binding proteins, Sp1 and BKLF.

b)   **Can BKLF transactivate the ALAS2 promoter?**

Co-transfection studies performed in the *Drosophila* SL2 cell line, devoid of endogenous Sp1-like activity have demonstrated that BKLF can function as a transcriptional activator (Crossley *et al.*, 1996). To investigate whether the -54 CACCC sequence in the ALAS2 promoter can respond transcriptionally to exogenously expressed murine BKLF, co-transfection assays were performed in CV-1 cells which contain reduced levels of BKLF in comparison to MEL cells (Crossley *et al.*, 1996).

Preliminary co-transfection studies in CV-1 cells were undertaken with 2 pmol of the reporter plasmid and 7.5μg of the murine BKLF cDNA expression clone, pMT2/BKLF (provided by Dr. M. Crossley). The plasmid construct, p(CAC)$_4$tk-LUC was not transactivated by BKLF, but activity was reduced, approximately 4-fold. The plasmid, pβ-glob-LUC was only marginally transactivated (approximately 1.3-fold), by exogenously

Figure 5.11   Transactivation of the ALAS2 promoter by EKLF

The constructs p(CAC)$_4$tk-LUC, ptk-LUC, pβ-glob-LUC, pALAS-124A-LUC, pALAS-124mut-LUC and pGL2-Basic were co-transfected with the EKLF cDNA expression clone, pSG5/EKLF in K562 cells and luciferase activities determined. The mutated -54 CACCC sequence is represented by a *cross*. The data are averages obtained from constructs tested in quadruplicate in at least six experiments and are represented as the mean ± standard deviation. The plasmids ptk-LUC and pGL2-Basic were included as controls for transactivation by EKLF, assigned a value of 1.0 and transactivation of the plasmid constructs by EKLF corrected for background.

p(CAC)₄tk-LUC

*Nde*I

(CCACACCCT)₄

*Xho*I

tk

+1

LUC

9.5 ± 1

ptk-LUC

tk

+1

LUC

1.0

pβ-glob-LUC

*Xba*I

-168 bp

CCACACCCT

-90

*Bgl*II

+1

LUC

+37 bp

4.0 ± 0.8

pALAS-124A-LUC

*Pvu*II

-124 bp

AGATAC

-100

GGGTGGGTGGGG

-54

*Hind*III

+1

LUC

+28 bp

3.1 ± 0.2

pALAS-124mut-LUC

AGATAC

-100

X

-54

+1

LUC

1.8 ± 0.4

pGL2-Basic

LUC

1.0
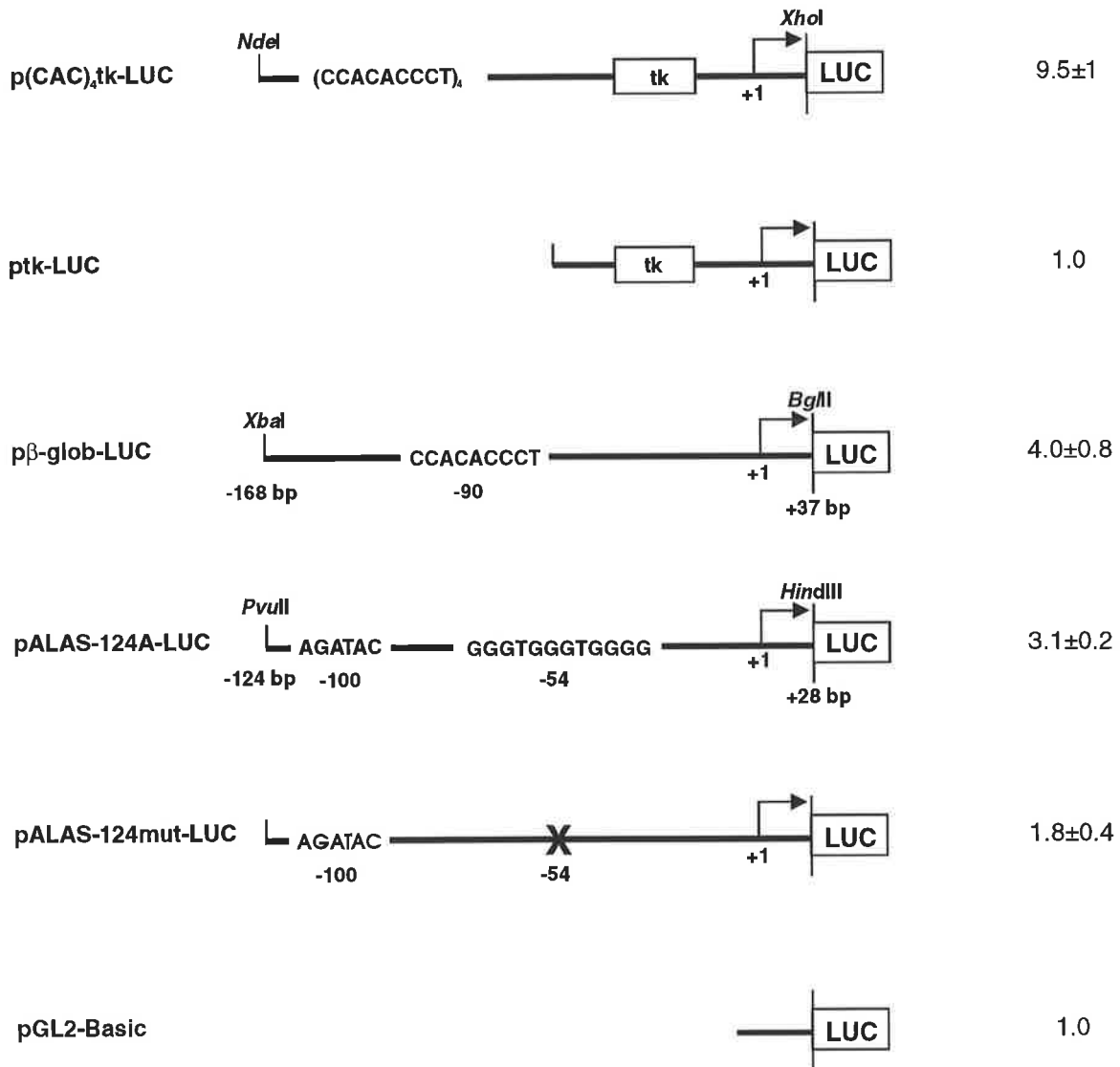
expressed BKLF in CV-1 cells. These results suggested that BKLF co-transfection assays would not be suitable in this cell line.

To investigate whether K562 cells are an appropriate cell line for co-transfection studies with exogenously expressed BKLF, gel shift assays were performed with the β-globin CACCC probe to examine the level of BKLF protein in K562 cells in comparison to MEL cells. Although, Crossley *et al.* (1996) had reported a similar level of BKLF in both MEL and K562 cell lines, variation in the level of nuclear proteins between strains of MEL cell lines had been observed. As seen in Fig. 5.12, the binding profile of nuclear proteins to the β-globin CACCC probe with nuclear extracts from K562 (lane 2) and MEL (lane 6) cells was extremely similar. An antibody to BKLF (lanes 4 and 8) specifically inhibited binding of the most rapidly migrating complex observed in both K562 and MEL nuclear extracts, whereas antibodies to EKLF or preimmune serum had no effect. Consequently, gel shift analysis indicated that a major retarded protein complex corresponding to BKLF was present in both K562 and MEL cells. However, in spite of this finding, preliminary co-transfection experiments were undertaken in this cell line.

As seen in Fig. 5.13, p(CAC)$_4$tk-LUC was not transactivated by exogenously expressed BKLF but was reduced (0.48-fold), whereas expression of pβ-globin-LUC was only marginally transactivated (1.7-fold) by BKLF. In contrast, pALAS-124A-LUC was transactivated approximately 12-fold by BKLF in K562 cells, but mutagenesis of the -54 CACCC sequence (pALAS-124mut-LUC) only marginally reduced this to 10-fold indicating that BKLF may function through an alternative site.

Sequence analysis of the 152 bp ALAS2 promoter revealed a putative CACCC site (5'-ACCCACCT-3') located at -3/+5 (at the transcription initiation site). To investigate whether BKLF could function through this site, a series of ALAS2 plasmid constructs were generated by PCR, in which this region was deleted or mutated (refer to Section 2.4.5).

As seen in Fig. 5.13, deletion of the putative CACCC site at -3/+5 (pALAS-124-CAP-LUC) abolished transactivation of the ALAS2 promoter by BKLF (0.85-fold) in K562 cells. Mutagenesis of the -3/+5 putative CACCC site at +2/+4 did not significantly affect the ability of exogenously expressed BKLF to transactivate the ALAS2 promoter, since pALAS-124CAPmut-LUC was transactivated approximately 8.5-fold. Similarly, pALAS-124mutCAPmut-LUC containing mutations at the -54 CACCC sequence and at +2/+4 was transactivated approximately 9.5-fold with exogenously expressed BKLF

Figure 5.12   Expression of BKLF in erythroid cell lines by gel shift analysis

The radiolabelled β-globin CACCC probe was incubated with nuclear extracts from K562 (lanes 2-5) and MEL (lanes 6-9) cells. Nuclear extract was omitted from lane 1. For supershift assays, preimmune (PI) serum (lanes 3 and 7), anti-BKLF (lanes 4 and 8) and anti-EKLF (lanes 5 and 9) antibodies were added prior to the addition of probe. The retarded complex corresponding to BKLF is *arrowed*.

Nuclear Extract / K562 / MEL

Antibody: −, −, PI, BKLF, EKLF (K562); −, PI, BKLF, EKLF (MEL)
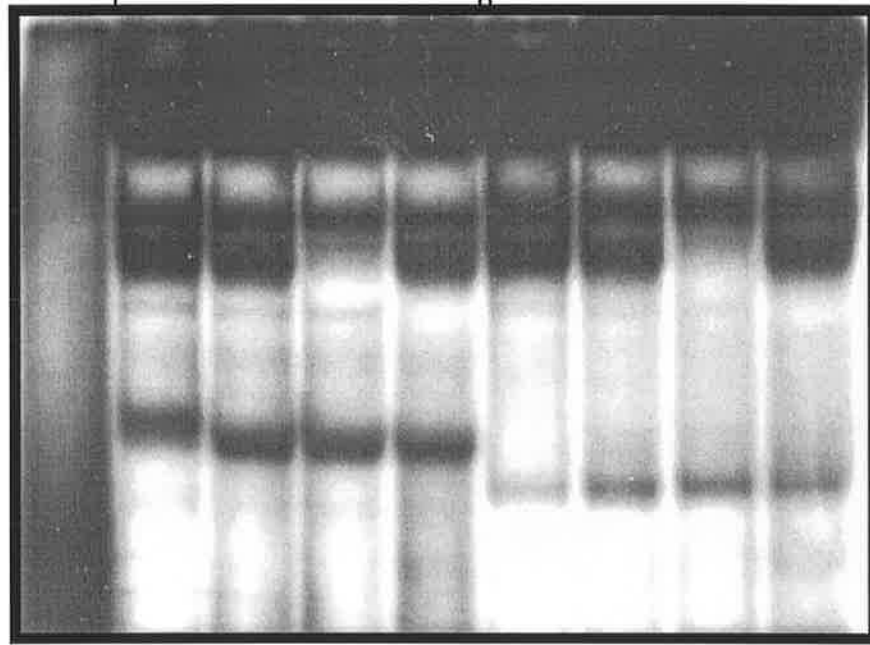
BKLF →

1 2 3 4 5 6 7 8 9

Figure 5.13   Transactivation of the ALAS2 promoter by BKLF

A series of ALAS2 promoter plasmid constructs were co-transfected with the BKLF cDNA expression clone, pMT2/BKLF in K562 cells and luciferase activities determined. The mutated CACCC sites located at -54 and -3/+5 are represented by a *cross*. The plasmid pGL2-Basic was included as a control for transactivation by BKLF, assigned a value of 1.0 and transactivation of the plasmid constructs by BKLF corrected for background. The data are averages obtained from constructs tested in quadruplicate in at least four experiments and are represented as the mean ± standard deviation.

Fold Induction with BKLF in K562 cells

Construct	Diagram details	Fold Induction
p(CAC)₄tk-LUC	NdeI — (CCACACCCT)₄ — tk — +1 — XhoI — LUC	0.48±0.02
pβ-glob-LUC	XbaI (-168 bp) — CCACACCCT (-90) — +1 — BglII (+37 bp) — LUC	1.7±0.05
pALAS-124A-LUC	BglII (-124 bp) — GGGTGGGTGGGG (-54) — GATA (-27) — +1 — HindIII (+28 bp) — LUC	11.9±3.9
pALAS-124mut-LUC	X (-54) — GATA (-27) — +1 — LUC	10±2.3
pALAS-124 -CAP-LUC	BglII (-124 bp) — GGGTGGGTGGGG (-54) — GATA (-27) — -2 — HindIII — LUC	0.85±0.13
pALAS-124mut -CAP-LUC	X (-54) — GATA (-27) — -2 — HindIII — LUC	2±0.7
pALAS-124 CAPmut-LUC	GGGTGGGTGGGG (-54) — GATA (-27) — X +1 — HindIII (+28 bp) — LUC	8.55±2
pALAS-124mut CAPmut-LUC	X (-54) — GATA (-27) — X +1 — LUC	9.5±2.6
pALAS -27-LUC	PvuII (-27) — +1 — LUC	12±1.1
pALAS-27 CAPmut-LUC	(-27) — X +1 — LUC	11±0.7
pGL2-Basic	— LUC	1.0

in K562 cells. Plasmid constructs containing ALAS2 promoter sequence from -27 bp to +28 (pALAS-27-LUC), were also transactivated approximately 11-12-fold by BKLF in K562 cells, irrespective of the mutation introduced at +2/+4 (Fig. 5.13). In conclusion, transactivation experiments performed with exogenously expressed BKLF in K562 cells demonstrated that BKLF was capable of increasing expression of the luciferase reporter gene through a region in the ALAS2 promoter. In contrast, the β-globin promoter was only marginally transactivated (1.7-fold) with exogenously expressed BKLF.

In an attempt to localise the site through which BKLF transactivated the ALAS2 promoter, gel shift assays were performed. The CAP-probe containing ALAS2 promoter sequence from (-10 bp to +12 bp) (see Section 2.1.7) and the β-globin CACCC probe from the murine adult β-globin promoter which binds BKLF were incubated with nuclear extracts isolated from K562 cells. As seen in Fig. 5.14A, three retarded protein complexes were observed with the β-globin CACCC probe (lane 3) and the most rapidly migrating complex was identified as BKLF since binding was inhibited with an antibody to BKLF (lane 4). In comparison, binding of nuclear proteins to the CAP-probe was extremely weak (lane 1) and there was no evidence that BKLF was capable of binding to this sequence, since addition of an antibody to BKLF (lane 2) did not effect the binding profile observed with this probe.

Transactivation studies described in K562 cells with exogenously expressed BKLF indicated that the region of ALAS2 promoter located from -27 bp to +28 bp was transactivated approximately 12-fold. A 55 bp fragment containing this region was isolated and radiolabelled and designated the CAP-55 probe. To investigate whether sequences located within this region were capable of binding BKLF directly, or if in the presence of exogenously expressed BKLF, proteins were inhibited from binding to this region, nuclear extracts were prepared from K562 cells transfected with an increasing amount of the BKLF cDNA expression clone, pMT2/BKLF. As seen in Fig. 5.14B, an increase in the retarded complex binding to the radiolabelled β-globin CACCC probe and corresponding to BKLF was observed in K562 nuclear extracts transfected with 30μg of pMT2/BKLF (lane 7) in comparison to mock-transfected K562 nuclear extracts (lane 5). Although in this gel shift assay the intensity of the complex corresponding to BKLF was extremely weak in comparison to what has been observed previously, there was no significant increase in the retarded protein complexes observed with the K562 nuclear extracts and the CAP-55 probe (compare lane 4 with lane 2). However, an additional although extremely weak protein

Figure 5.14   Identification of binding sites for BKLF in the ALAS2 promoter (-27 bp to +28 bp) by gel shift analysis

A   Radiolabelled double-stranded oligonucleotides containing the putative CACCC site at -3/+5 (CAP probe) and the β-globin CACCC probe were incubated with nuclear extracts from K562 (lanes 1-4). An anti-BKLF antibody was added to the K562 nuclear extract prior to the addition of the CAP probe (lane 2) and the β-globin CACCC probe (lane 4). The retarded complex corresponding to BKLF is *arrowed.*

B   A 55 bp fragment of the ALAS2 promoter (-27 bp to +28 bp) was radiolabelled and designated the CAP-55 probe. The CAP-55 probe and the β-globin CACCC probe were incubated with nuclear extracts isolated from K562 cells (lanes 2 and 5), K562 cells transfected with 7.5µg of the BKLF cDNA expression vector, pMT2/BKLF (lanes 3 and 6) and K562 cells transfected with 30µg of this vector (lanes 4 and 7). Nuclear extract was omitted from lane 1 and contains the CAP-55 probe. The retarded complex corresponding to BKLF, and a faster migrating complex detected with the CAP-55 probe (?) and nuclear extracts from K562 cells are *arrowed.*

A
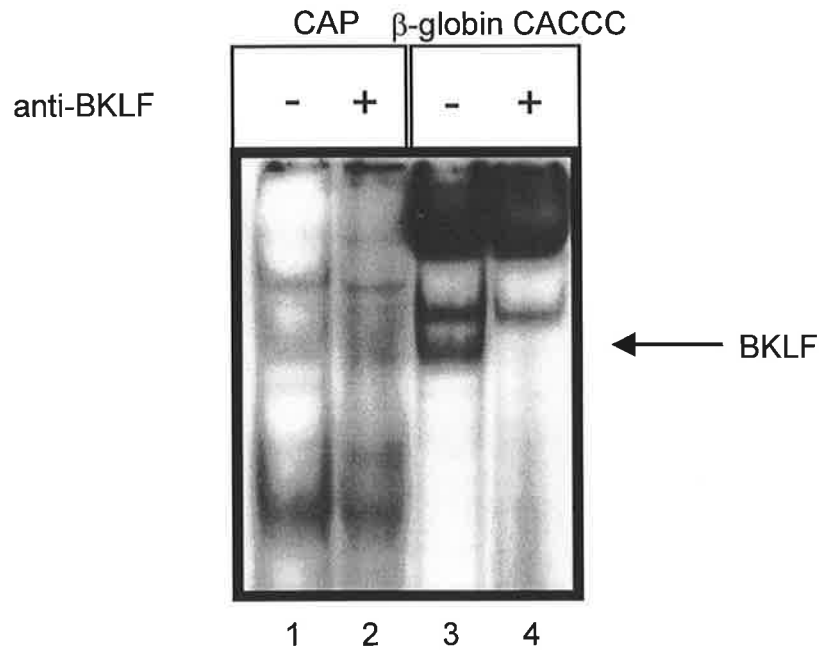
B
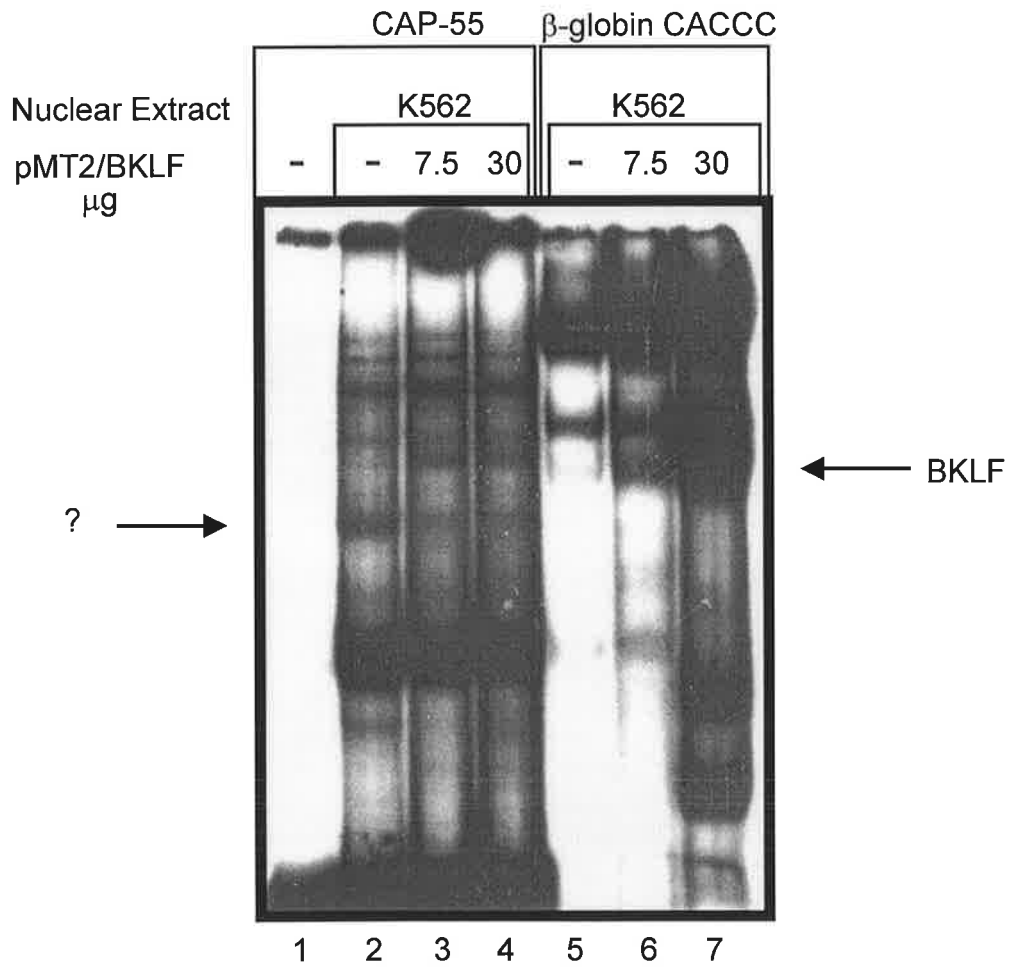
complex was observed in mock-transfected K562 nuclear extracts (lane 2) but not with K562 nuclear extracts transfected with either 7.5µg (lane 3) or 30µg (lane 4) of pMT2/BKLF. Hence, gel shift analysis of ALAS2 promoter sequence from -27 bp to +28 bp did not detect the binding of BKLF to any sequences located within this region.

To eliminate the possibility of BKLF transactivating sequences located within the luciferase vector, a 152 bp ALAS2 promoter (-124 bp to +28 bp) fragment was inserted into the vector, pCATOO containing two synthetic polyA sites inserted upstream of the CAT reporter gene as described in Section 2.4.5. Two plasmid constructs were examined, one containing a mutation at the -54 CACCC sequence (pALAS-124mut-CAT) and the second, a mutation introduced at +2/+4 (pALAS-124CAPmut-CAT). Transactivation experiments consisted of 15µg of the CAT reporter construct due to the low level CAT expression and 15µg of pMT2/BKLF in K562 cells. To determine whether BKLF was also capable of transactivating other promoters, 2 pmol of the plasmids, pRSV-βgal and ptk-LUC were co-transfected with 7.5µg of pMT/BKLF.

As seen in Fig. 5.15A, BKLF consistently transactivated the ALAS2 promoter-CAT constructs approximately 2.7-fold irrespective of the mutations introduced at either the -54 CACCC site or at +2/+4. Similarly, the Rous sarcoma viral and the thymidine kinase promoters were transactivated 6.5 and 10 -fold, respectively, by BKLF in K562 cells (Fig. 5.15B).

In conclusion, the -54 CACCC sequence identified in the human ALAS2 promoter is capable of binding BKLF but co-transfection studies performed with exogenously expressed BKLF in K562 cells, demonstrated that this site does not respond transcriptionally to exogenously expressed BKLF. However, the ALAS2 promoter can be transactivated by BKLF approximately 10-fold in K562 cells. Attempts to localise and identify the BKLF-responsive site in the ALAS2 promoter focused on the region located between -27 bp to +28 bp which included the transcription initiation site. Mutational analysis of a putative CACCC site located at -3/+5, failed to abolish the transactivation of the ALAS2 promoter by BKLF. In addition, gel shift analysis of sequences located within this region and the putative CACCC site at -3/+5 were unable to detect the binding of BKLF.

Figure 5.15   Transactivation of the ALAS2, Rous sarcoma viral (RSV) and thymidine kinase promoters by BKLF in K562 cells

A   K562 cells were electroporated with 15μg of ALAS2/pCATOO plasmids containing a mutation at -54 (pALAS-124mut-CAT) or at +2/+4 (pALAS-124CAPmut-CAT) and 15μg of pMT2/BKLF. Cells were harvested 48 hours following transfection, the protein concentration of the cell lysate determined and 100μg of total protein assayed for CAT activity. The % of acetylated chloramphenicol and the fold induction obtained with exogenously expressed BKLF were determined. The autoradiograph shown is representative of a typical experiment and the data obtained from constructs tested in duplicate in three experiments and are represented as the mean ± standard deviation.

B   The plasmids, pRSV-βgal and ptk-LUC (2 pmol) were electroporated with 7.5μg of pMT2/BKLF in K562 cells. Cells were harvested 24 hours following transfection, and 100μg of total protein assayed for β-galactosidase or luciferase activity. The data are averages obtained from constructs tested in triplicate in three experiments and are represented as the mean ± standard deviation.

A

Plasmid	pALAS-124mut-CAT				pALAS-124CAPmut-CAT			
pMT2/BKLF	−	−	+	+	−	−	+	+

| Fold induction with BKLF | 1 | | 2.62±0.16 | | 1 | | 2.7±0.38 | |

B

Fold induction with BKLF

pRSV-βgal   RSV promoter   +1   βgal   6.5±0.95

ptk-LUC   tk promoter   +1   LUC   10±0.4
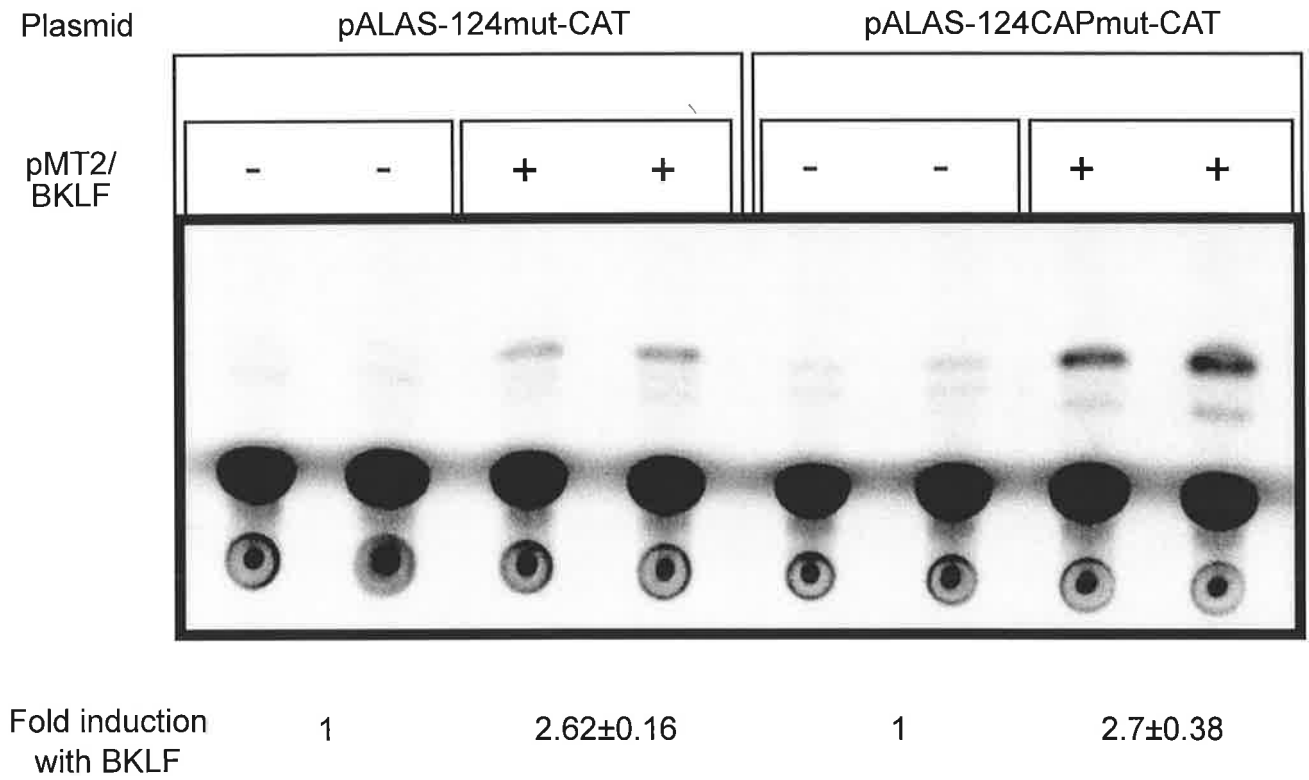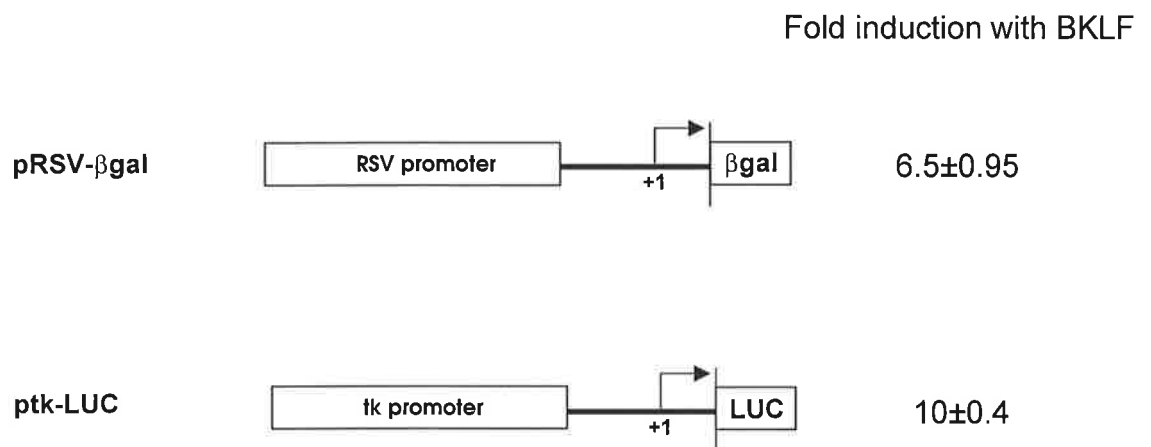
5.2.3 Analysis of a Putative NF-E2-like Sequence in the Human ALAS2 Promoter

i)   *Gel Shift Analysis of an NF-E2-like Sequence*

Partially overlapping the -54 CACCC sequence, is an NF-E2-like sequence (5'-GGCTGAGTCAG-3') located at -49/-39 (Fig. 5.1) with a 9/11 match to the consensus DNA binding motif for NF-E2 (5'-(T/C)GCTGAG/CTCA(C/T)-3') (Andrews *et al.*, 1993a). To investigate the proteins capable of binding to the putative NF-E2-like binding site identified in the human ALAS2 promoter, two complementary oligonucleotides (-32 bp to -52 bp) were synthesised that encompassed both the putative NF-E2 sequence and one of the CACCC sites (-48/-56) (see Section 2.1.7). These oligonucleotides were radiolabelled and annealed and designated the NF-E2-44 probe. The binding of nuclear proteins to this site was investigated with nuclear extracts from K562, MEL or COS-1 cells using gel shift assays as previously described. The preparation of nuclear extracts and the binding protocol described for the detection of GATA-1 binding was also employed in these studies.

As seen in Fig. 5.16, two major retarded protein complexes were obtained with the NF-E2-44 probe in K562 and MEL cell nuclear extracts (lanes 2 and 6) and minor bands of similar mobility with COS-1 nuclear extracts (lane 9). An additional minor band located between the two major retarded complexes was obtained with nuclear extracts from MEL cells and was also detected in COS-1 cells, although the intensity was increased (lanes 6 and 9).

To investigate the identity of these retarded protein complexes, competition experiments were performed with a 50-fold molar excess of unlabelled oligonucleotides containing the NF-E2-44 probe in self competition, or an AP-1 or Sp1 consensus binding sites (designated AP-1 and Sp1-cons oligonucleotides). As seen in Fig. 5.16A, the two major retarded protein complexes obtained with K562 (lane 2) and MEL (lane 6) cell nuclear extracts, were substantially reduced with a 50-fold molar excess of the NF-E2-44 oligonucleotide in self competition (lane 3) or the Sp1-cons oligonucleotide (lanes 4 and 7). These two protein complexes were not affected with a 50-fold molar excess of the AP-1 cons oligonucleotide (lanes 5 and 8) although the minor band observed with MEL cell nuclear extracts was abolished (compare lane 8 with lane 6). The major protein complex observed with COS-1 nuclear extracts (lane 9) was abolished with a 50-fold excess of the NF-E2-44 oligonucleotide in self competition (lane 10) and the AP-1 cons oligonucleotide

(lane 12). The two minor bands observed in the COS-1 cell nuclear extracts were substantially reduced with a 50-fold excess of the Sp1-cons oligonucleotide (lane 11).

These competition experiments demonstrated that the NF-E2-44 probe substantially bound Sp1 protein (or a related protein) in erythroid cell nuclear extracts (as expected from the inclusion of the CACCC site in this probe and were also observed with the -54 CACCC probe under the same binding conditions) and AP-1 in non-erythroid cell extracts. However, the detection of an additional protein complex in erythroid nuclear extracts corresponding to NF-E2 was not observed. A consensus NF-E2 probe was not included in these experiments which would have unequivocally established that these nuclear extracts contained NF-E2.

As a source of recombinant murine NF-E2, nuclear extracts were also prepared from COS-1 cells co-transfected with 10μg of each cDNA expression clone, pMT2/RINFE and pMT2/p18w-1, encoding the erythroid-specific subunit of NF-E2, p45 (Andrews *et al.*, 1993a) and its ubiquitously expressed partner, p18 (Andrews *et al.*, 1993b), respectively. However, no additional retarded complexes were observed following the incubation of the NF-E2-44 probe with these nuclear extracts in comparison to mock-transfected COS-1 nuclear extracts (data not shown). This finding is in agreement with gel shift studies performed by Andrews *et al.* (1993a) where the ALAS2 NF-E2-like sequence failed to inhibit binding of purified NF-E2 protein to a consensus NF-E2 binding site.

ii)   *Mutational Analysis of the Putative NF-E2 Binding Site*

Gel shift analysis of the putative NF-E2-like sequence at -49/-39 failed to detect binding of an erythroid-specific complex. To unequivocally eliminate a functional role of this site in the expression of the ALAS2 promoter, the sequence was mutated (5'-GGCTGAGTCAG-3' to 5'-GGCAGCTGCAG-3') and the resulting construct pALAS-293mut11-LUC, transiently transfected into K562, MEL and COS-1 cells. As seen in Fig. 5.16B, this mutation did not affect expression of the ALAS2 promoter construct in erythroid cells. However, expression of this mutant promoter construct in COS-1 cells, was reduced to 78% of the wild-type, presumably reflecting a contribution from AP-1.

In transactivation experiments in COS-1 cells, over-expression of recombinant murine NF-E2 protein, using the cDNA expression clones encoding p45 and p18, failed to

Figure 5.16   Gel shift and mutational analysis of the -44 NF-E2-like sequence in the ALAS2 promoter

A   Gel shift analysis of the -44 NF-E2-like sequence

A radiolabelled double-stranded oligonucleotide containing the -44 NF-E2-like sequence and one of the CACCC sites (NF-E2-44 probe) was incubated with nuclear extracts isolated from K562 (lanes 2-5), MEL (lanes 6-8) and COS-1 (lanes 9-12) cells. Nuclear extract was omitted from lane 1. The major and minor retarded complexes are *arrowed*. These retarded complexes were competed with a 50-fold molar excess of the NF-E2-44 oligonucleotide in self competition (lanes 3 and 10), the Sp1-cons oligonucleotide (lanes 4, 7 and 11) and an AP-1 cons oligonucleotide (lanes 5, 8 and 12).

B   Effect of mutating the -44 NF-E2-like sequence on ALAS2 promoter expression

The -44 NF-E2-like sequence in pALAS-293-LUC was mutated to a *Pvu*II site, represented by a *cross*. This construct and pALAS-293-LUC were co-transfected with a β-galactosidase expression construct (pRSV-βgal) and transiently expressed in K562, MEL and COS-1 cells. The luciferase activities were standardised relative to β-galactosidase activity as an internal control for transfection efficiency and expressed relative to pALAS-293-LUC (set at 100%). The data are averages obtained from constructs tested in triplicate in two experiments and are represented as the mean ± standard deviation.

A

Nuclear Extract

Competitor

B

Relative luciferase activity

		K562	MEL	COS-1
pALAS-293-LUC		100	100	100
pALAS-293 mut11-LUC		109±3	96±8	78±5
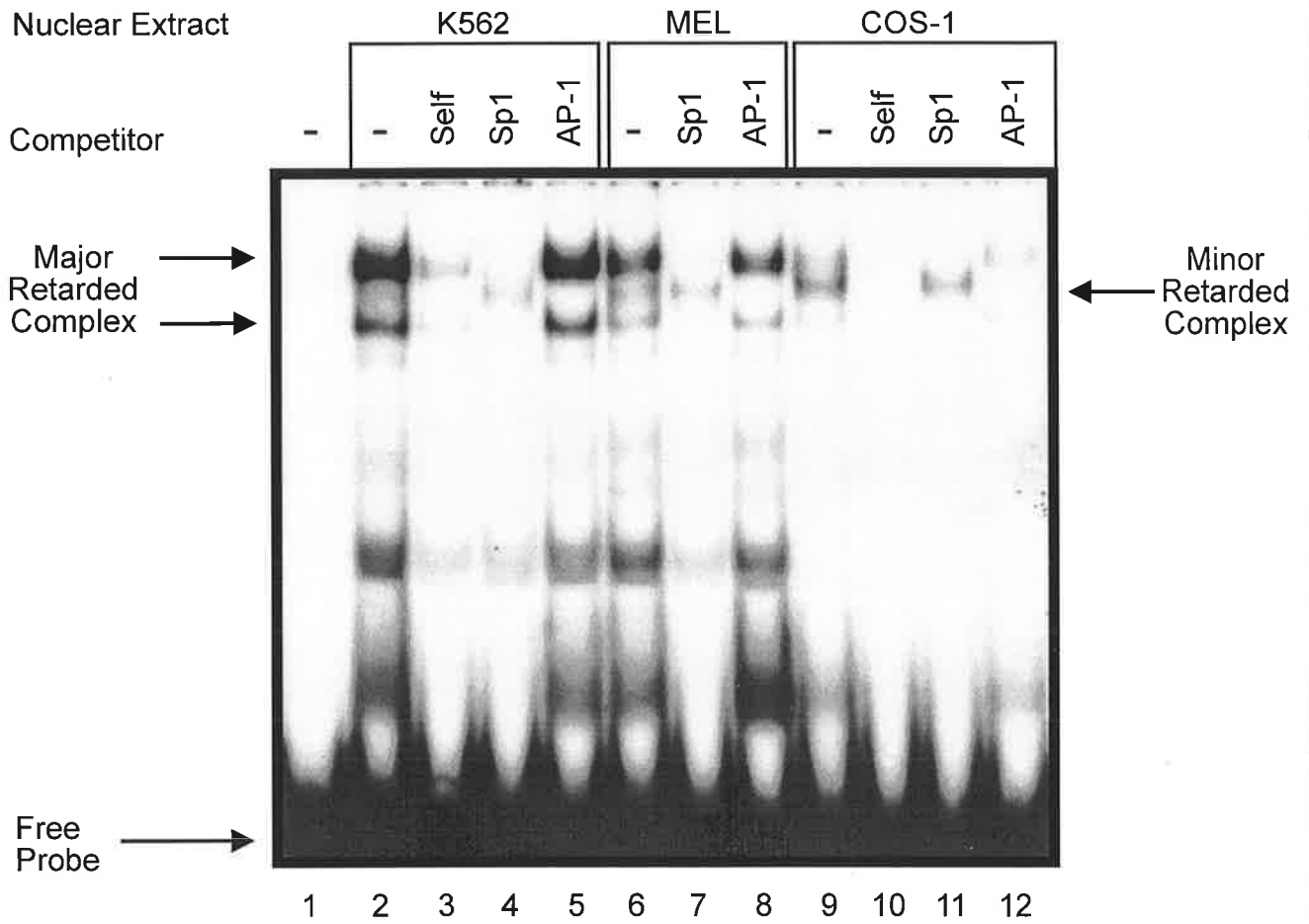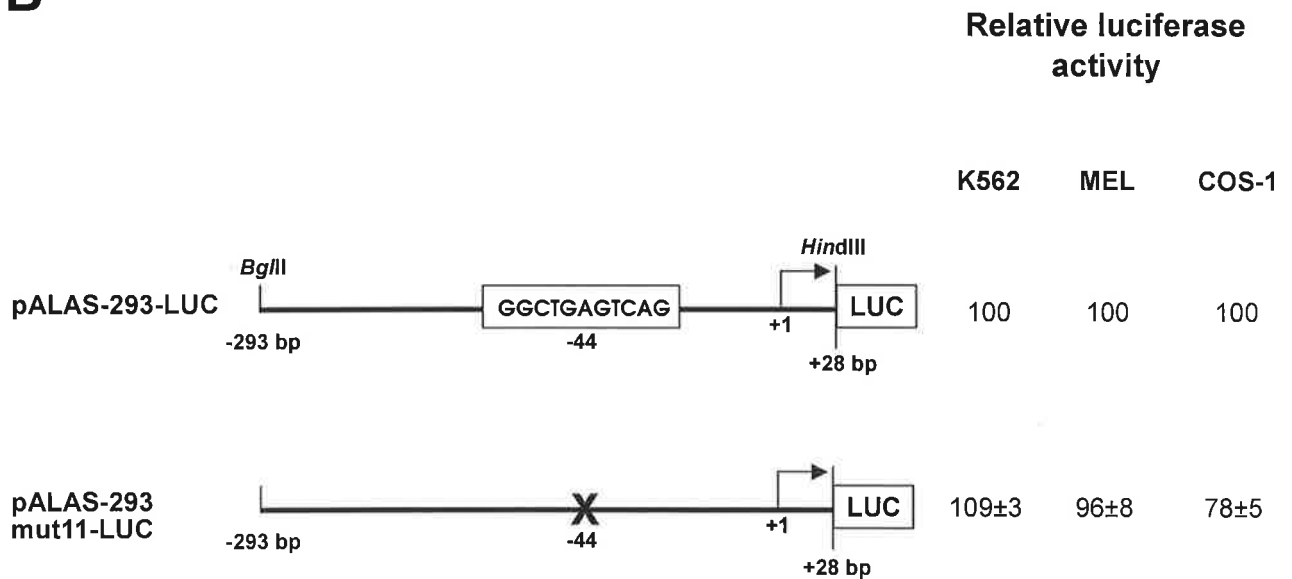
increase expression of pALAS-293-LUC (data not shown). These experiments established that the NF-E2-like sequence is inactive in erythroid cells.

5.2.4 Analysis of a Putative Thyroid Hormone Receptor Binding Site (TRE) in the Human ALAS2 Promoter

i)   Gel Shift Analysis of a Putative Thyroid Hormone Receptor Binding Site

A putative thyroid hormone receptor binding site (TRE) was identified at -21/-6 (5'-CCAGGTCCTAACCCAA-3') in the human ALAS2 promoter with a 11/16 match to the ideal TRE consensus (5'-TCAGGTCATGACCTGA-3') (Glass *et al.*, 1988). To investigate whether the TRE-like sequence could bind nuclear proteins, gel shift assays were performed. Two complementary oligonucleotide pairs were synthesised, the first pair contained the putative TRE sequence from the ALAS2 promoter, and the second pair contained the consensus TRE binding site, and are referred to as the TRE-13 and TRE-cons probes, respectively (see Section 2.1.7). These oligonucleotides were radiolabelled, annealed and examined for protein binding activity in gel shift assays (as described in Section 2.9). Nuclear extracts were prepared from K562 and MEL cells, and COS-1 cells which does not express the thyroid hormone receptor. The nuclear extracts and binding protocol described for the detection of GATA-1 binding was also employed in these experiments, with the exception that the binding reaction was incubated at 25°C rather than on ice for optimal binding of retarded complexes.

As seen in Fig. 5.17, two major protein complexes were observed with the TRE-cons probe and K562 (lane 2) and MEL (lane 5) cell nuclear extracts. The faster mobility complex was also observed with COS-1 (lane 8) nuclear extracts. Competition experiments performed with the TRE-cons probe and nuclear extracts from K562 and MEL cells, demonstrated that a 50-fold molar excess of the TRE-cons oligonucleotide as a self competitor (lanes 3 and 6) but not the TRE-13 oligonucleotide (lanes 4 and 7) was required to abolish binding of the two retarded protein complexes (although this is difficult to see on the reproduction of the autoradiogram). The intensity of the faster mobility complex observed with COS-1 cell nuclear extracts (lane 8) was only marginally reduced with a 50-fold molar excess of the TRE-cons oligonucleotide in self competition (lane 9) and unaffected with a 50-fold molar excess of the TRE-13 probe (lane 10). In contrast, protein

binding to the TRE-13 probe was not detected with nuclear extracts isolated from either K562, MEL or COS-1 cells (data not shown). In conclusion, gel shift analysis of the putative TRE identified in the ALAS2 promoter, failed to detect binding of nuclear proteins to this site.

ii)   *Mutational Analysis of a Putative TRE*

To eliminate a functional role of this site in the expression of the ALAS2 promoter, the putative TRE was inactivated by mutagenesis in the plasmid pALAS-293-LUC and expression analysed in K562, MEL and COS-1 cells. The thyroid hormone receptor is known to bind as a homodimer or a heterodimer to TREs and a *Hind*III site was introduced at the first half site to abolish the possible binding of this dimer to the putative TRE in the ALAS2 promoter. However, mutagenesis of the putative TRE (5'-CCAGGTCCTAACCCAA-3' to 5'-CAAGCTTCTAACCCAA-3') did not affect expression of pALAS-293mut12-LUC relative to wild-type in either K562, MEL or COS-1 cells (Fig. 5.17B).

5.2.5   Analysis of a Putative Ets Binding Site in the Human ALAS2 Promoter

i)   *Gel Shift Analysis of a Putative Ets Site*

A putative Ets binding site (5'-TAGCAAGGAAGGGAC-3') located at -119/-105 in the ALAS2 promoter, was initially identified by sequence homology to a functional enhancer element (5'-TAAGCAGGAAGTGAC-3') found in both the adenovirus type 5 (Herbomel *et al.*, 1984) and polyoma virus (Hearing & Schenk, 1986) and known to bind members of the Ets family of transcription factors. The Ets-like sequence 5'-AGGAAGGGAC-3' (-114 to -105), containing the core ets binding motif, GGAA is located between the -124 and -100 GATA sites in the ALAS2 promoter. Two complementary oligonucleotides (-132 bp to -90 bp) were synthesised that encompassed the two GATA-1 binding sites at -124 and -100 and the putative ets site (see Section 2.1.7). These oligonucleotides were radiolabelled, annealed and referred to as the ets-110 probe. A probe containing the Ets binding site located in the CYP_{24} promoter (ets-CYP_{24} probe) (Dr.

Figure 5.17   Gel shift and mutational analysis of the putative thyroid hormone receptor binding site (TRE) in the ALAS2 promoter

A   Gel shift analysis of the TRE-like sequence

A radiolabelled double-stranded oligonucleotide containing a consensus TRE (TRE-cons probe) was incubated with nuclear extracts isolated from K562 (lanes 2-4), MEL (lanes 5-7) and COS-1 (lanes 8-10) cells. Nuclear extract was omitted from lane 1. The two major retarded complexes are *arrowed*. These retarded complexes were competed with a 50-fold molar excess of the TRE-cons oligonucleotide in self competition (lanes 3, 6 and 9) but not with an oligonucleotide containing the ALAS2 TRE-like sequence (TRE-13) (lanes 4, 7 and 10).

B   Effect of mutating the TRE-like sequence on ALAS2 promoter expression

The -13 TRE sequence in pALAS-293-LUC was mutated to a *Pvu*II site (*cross*). This construct and pALAS-293-LUC were co-transfected with a β-galactosidase expression construct (pRSV-βgal) and transiently expressed in K562, MEL and COS-1 cells. The luciferase activities were standardised relative to β-galactosidase activity and expressed relative to pALAS-293-LUC (set at 100%). The data are averages obtained from constructs tested in triplicate in three experiments and are represented as the mean ± standard deviation.

A

Nuclear Extract: K562, MEL, COS-1
Competitor: −, −, Self, TRE-13, −, Self, TRE-13, −, Self, TRE-13

Retarded Complexes

Free Probe

Lanes: 1 2 3 4 5 6 7 8 9 10

B

Relative luciferase activity

	K562	MEL	COS-1
pALAS-293-LUC	100	100	100
pALAS-293 mut12-LUC	95±2	91±6	104±3

pALAS-293-LUC: *Bgl*II ... −293 bp ... CCAGGTCCTAACCCAA (−13) ... *Hind*III ... +1 ... LUC ... +28 bp

pALAS-293 mut12-LUC: −293 bp ... X (−13) ... +1 ... LUC ... +28 bp
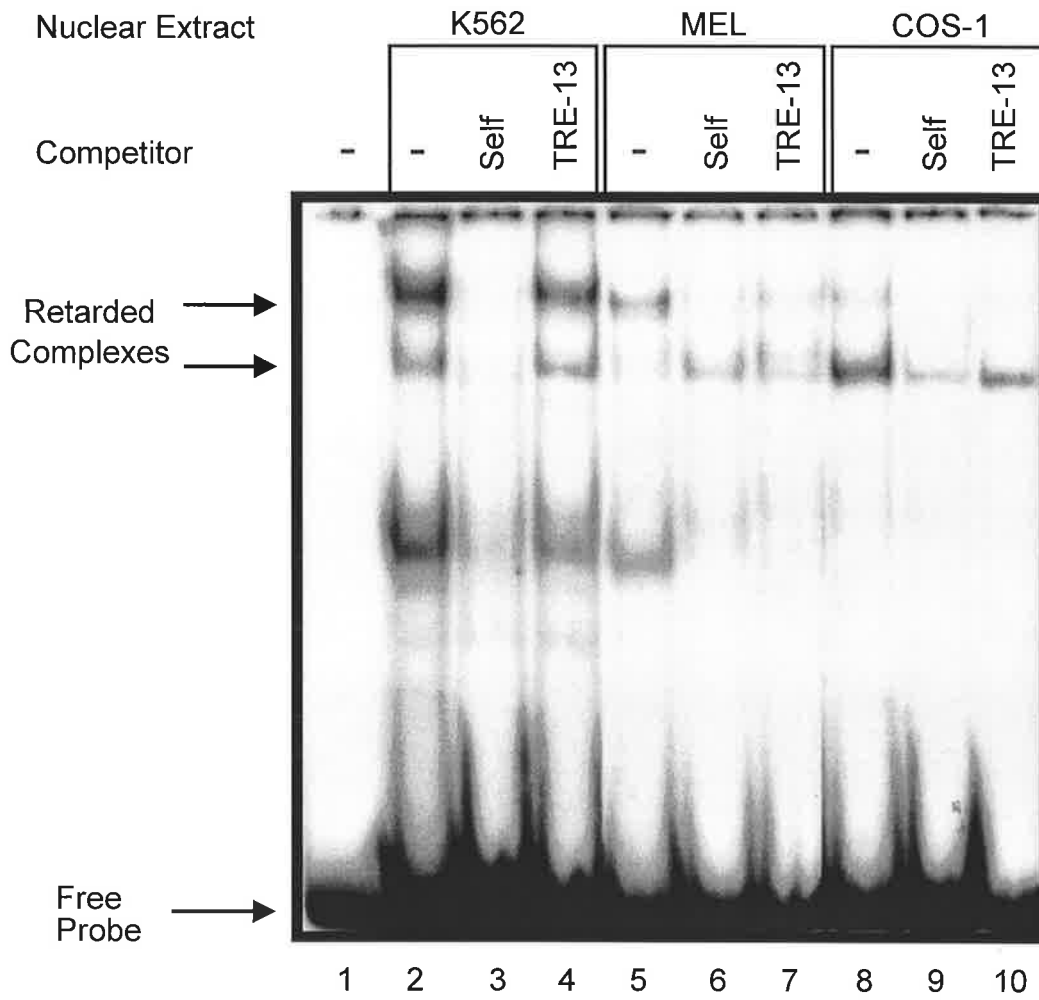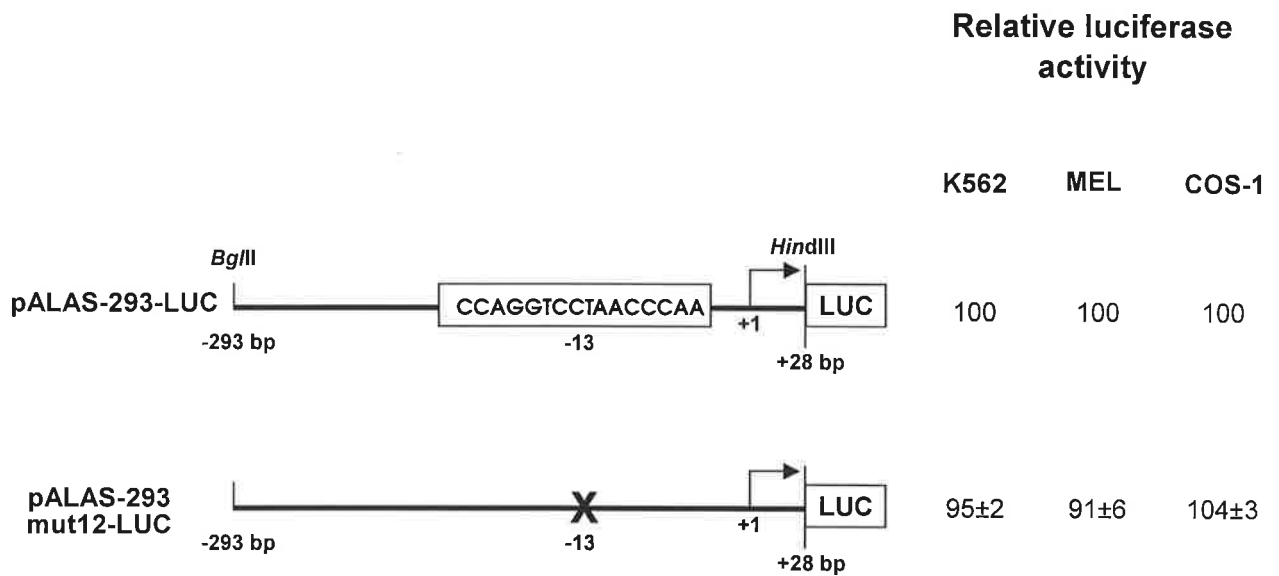
Prem Dwivedi, personal communication) and the GATA-cons probe were included as positive controls.

The binding protocol reported by Higashino *et al.* (1993) used to detect binding of the ets factor, E1A-F to the adenovirus type 5 E1A enhancer was employed (Section 2.9). As seen in Fig. 5.18A, a major retarded protein complex was detected with the ets-110 probe and nuclear extracts isolated from MEL cells (lane 2). In competition experiments, the protein complex was abolished with a 50-fold molar excess of the ets-110 oligonucleotide in self competition (lane 3) or the GATA-cons oligonucleotide (lane 4) but was not affected with a 50-fold molar excess of the ets-CYP_{24} oligonucleotide (lane 5). Consequently, only binding of GATA-1 to the ets-110 probe was detected.

Control gel shift experiments were performed using the radiolabelled ets-CYP_{24} probe and MEL nuclear extracts. The addition of a 50-fold molar excess of the ets-CYP_{24} oligonucleotide in self competition abolished the retarded protein complexes, whereas addition of a 50-fold molar excess of either the ets-110 oligonucleotide or a mutant homologue of the ets-CYP_{24} oligonucleotide containing a mutation in the core ets DNA binding site did not affect the retarded protein complexes (data not shown). In conclusion, gel shift assays were unable to detect binding of nuclear protein(s) to the putative Ets site (-114/-105) in the ALAS2 promoter.

ii) *Mutational Analysis of a Putative Ets Binding Site*

The putative Ets site (-114/-105) was inactivated by mutagenesis (5'-A GGAA GGGAC-3' to 5'-A CCAA GGGAC-3'), to eliminate a functional role of this site in the expression of the ALAS2 promoter, and analysed in K562, MEL and COS-1 cells. Expression of the plasmid, pALAS-293mut13-LUC was similar in comparison to wild-type when tested in K562, MEL and COS-1 cells (Fig. 5.18B).

Figure 5.18   Gel shift and mutational analysis of a putative Ets binding site (-114/-105) in the ALAS2 promoter

A   Gel shift analysis of the -110 Ets site

A radiolabelled double-stranded oligonucleotide containing the Ets-like sequence (-114/-105) and the -124 and -100 GATA sites (ets-110 probe) were incubated with nuclear extracts isolated from MEL cells (lanes 2-5). Nuclear extract was omitted from lane 1. The major retarded complex (*arrowed*) was competed with a 50-fold molar excess of the ets-110 oligonucleotide in self competition (lane 3), the GATA-cons oligonucleotide (lane 4) and a Ets binding sequence from the CYP_{24} promoter (lane 5).

B   Effect of mutating the -110 Ets site on ALAS2 promoter expression

The -110 Ets sequence in pALAS-293-LUC was mutated to a *Pvu*II site (*cross*). This construct and pALAS-293-LUC were co-transfected with a β-galactosidase expression construct (pRSV-βgal) and transiently expressed in K562, MEL and COS-1 cells. The luciferase activities were standardised relative to β-galactosidase activity and expressed relative to pALAS-293-LUC (set at 100%). The data are averages obtained from constructs tested in triplicate in three experiments and are represented as the mean ± standard deviation.

A

Nuclear Extract
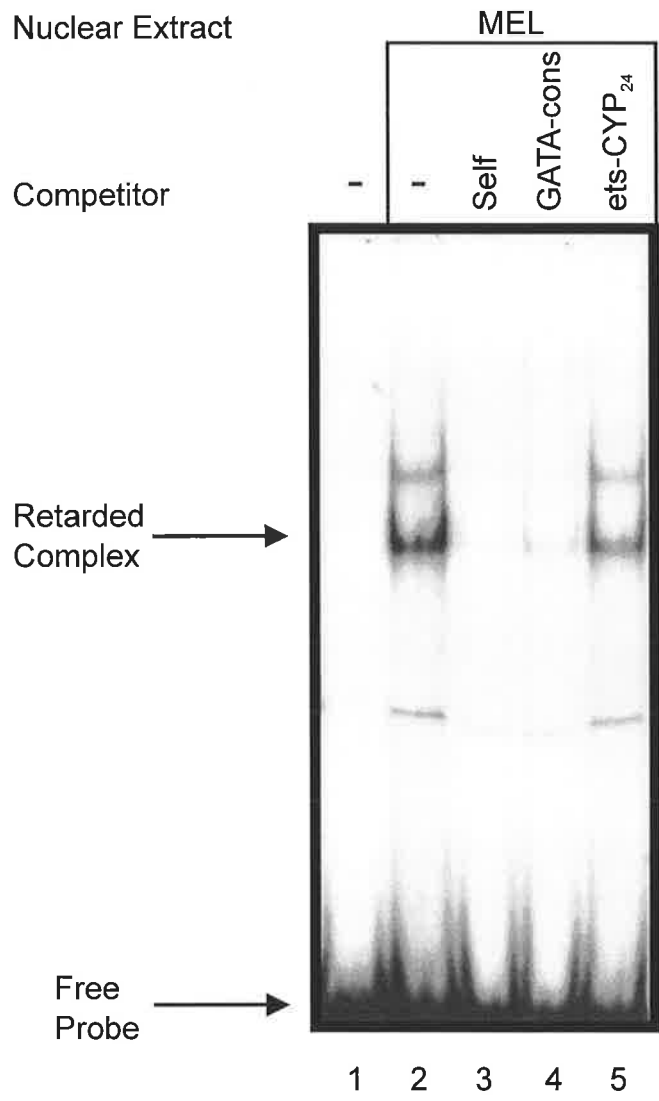
B
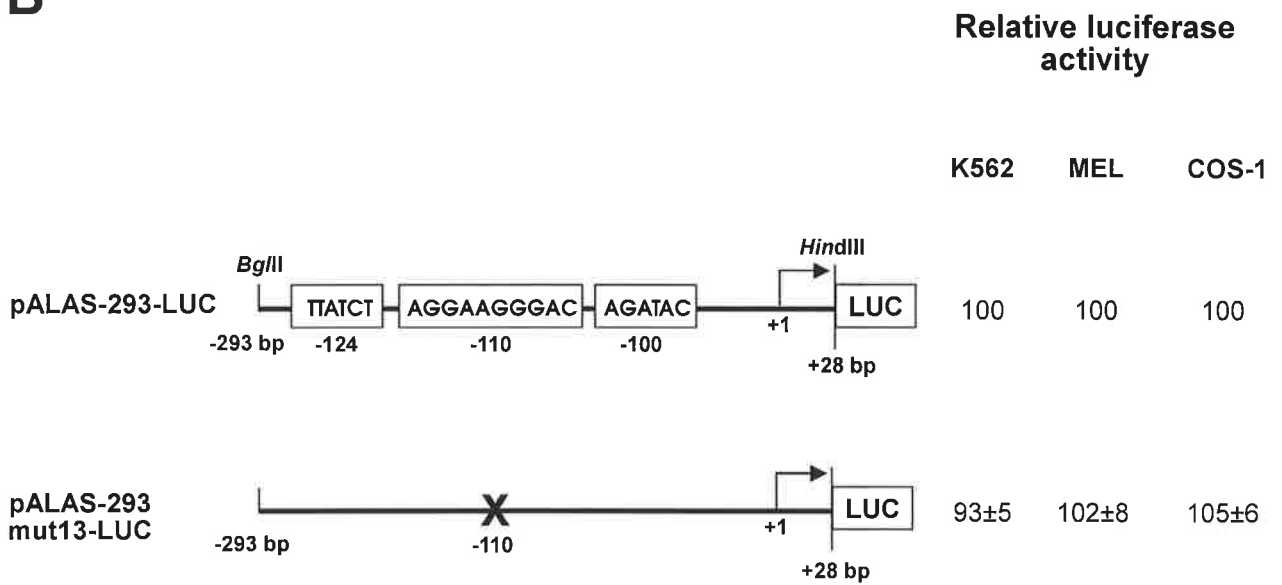
5.3 DISCUSSION

As described in Section 4.2.1, deletion analysis of the human ALAS2 gene 5'-flanking region demonstrated that the first 300 bp of promoter sequence can direct strong transient expression in erythroid cells. Sequence analysis of this region revealed several putative transcription factor binding sites (see Fig. 5.1) clustered within the first 140 bp. These included GATA and CACCC box motifs which are a feature of the regulatory regions of many other erythroid-specific genes including human PBGD (Mignotte *et al.*, 1989b), β-globin (deBoer *et al.*, 1988) and glycophorin B (Rahuel *et al.*, 1992), murine GATA-1 (Tsai *et al.*, 1991) and the EpoR (Zon *et al.*, 1991) and the rat pyruvate kinase (Max-Audit *et al.*, 1993).

Two potential GATA-1 binding sites are centred at -124 and -100 with an inverted palindromic arrangement. *In vitro* binding studies showed that each site bound GATA-1 protein present in erythroid cell nuclear extracts and the functional activity of these sites was demonstrated by transient expression analysis of mutated promoter constructs in erythroid cells. Transactivation assays with exogenously expressed GATA-1 in non-erythroid COS-1 cells confirmed the response of each of these sites to GATA-1. The contribution of the -124 site to ALAS2 expression was moderately greater than that of the -100 site, consistent with the deviation of the -100 GATA site by a single nucleotide from the consensus sequence (Merika & Orkin, 1993; Ko & Engel, 1993).

The human ALAS2 and the murine ALAS2 (Schoenhaut & Curtis, 1989) promoters both lack a canonical TATA box. However, located at -30/-23 in the human ALAS2 promoter, there is the sequence 5'-GGATAAAT-3' which binds TBP or GATA-1 *in vitro*. Protein binding reactions performed with purified GST-GATA-1(f) and TBP indicated that the affinities of these proteins for this site were considerably reduced compared with the consensus sites for these proteins. Conversion of the -30/-23 sequence to a consensus GATA-1 binding site, which binds GATA-1 *in vitro* but not TBP, significantly reduced transient expression in erythroid cells to 30% of the wild-type. This finding demonstrated the importance of a functional TATA box and presumably the requirement of the general transcription factor, TFIID, in the transcriptional initiation of this gene. Conversion of the -30/-23 sequence to a canonical TATA box, which binds TBP *in vitro* but not GATA-1, consistently reduced transient expression in erythroid cells to 70-80% of the wild-type, also supporting a role for GATA-1 in transcriptional initiation. A similar role for GATA-1 has

been proposed for the erythroid-specific human glycophorin B (Rahuel *et al.*, 1992) and chicken β-globin (Fong & Emerson, 1992) gene promoters which also possess non-canonical TATA boxes. For the chicken β–globin promoter, there is evidence that GATA-1 bound at the -30 position prevents the assembly of a repressive nucleosome (Barton *et al.*, 1993) and imparts erythroid cell-specificity through the interaction with another GATA-1 molecule bound to the 3' enhancer (Fong & Emerson, 1992). Perhaps, by analogy, GATA-1 bound to the -27 site of the ALAS2 promoter may facilitate transcriptional initiation *in vivo* by inhibiting nucleosome formation. The minor inhibition that was observed following conversion of the -27 site to a TATA box in transient transfection assays is probably attributable to poor nucleosome assembly. The mechanism by which such bound GATA-1 at the -27 site in the ALAS2 promoter, could be replaced by TFIID *in vivo* is not known but an "initiator-like" element (Smale & Baltimore, 1989) located at +7/+12 (5'-TCATTC-3') (see Fig. 5.1) may play a role in this process. Interestingly, initiator elements have been identified in the erythroid promoter of the human PBGD gene (Beaupain *et al.*, 1990) and the promoter of the human β-globin gene (Lewis & Orkin, 1995).

In the ALAS2 promoter, a CACCC sequence at -59/-48 was identified which consists of two overlapping CACCC boxes on the non-coding strand. This sequence is functionally important for expression since mutagenesis reduced promoter activity in erythroid cells to 45-60% of the wild-type, although the contributions of the two overlapping CACCC sites remain to be elucidated.

Gel supershift assays demonstrated that the ALAS2 CACCC sequence mimics the β-globin CACCC box (Crossley *et al.*, 1996) and is able to bind not only EKLF but also Sp1 and BKLF. While the *in vivo* function of these proteins has been difficult to define, a specific role for EKLF in adult β-globin gene transcription has been established (Bieker & Southwood, 1995; Donze *et al.*, 1995; Nuez *et al.*, 1995; Perkins *et al.*, 1995; Wijgerde *et al.*, 1996) and an EKLF-responsive CACCC box identified at -94/-87 in the murine adult β-globin gene promoter (Bieker & Southwood, 1995). Naturally occurring mutations in the CACCC site of the human β-globin promoter result in β-thalassemia (Orkin *et al.*, 1984; Kulozik *et al.*, 1991; Feng *et al.*, 1994), whereas disruption of the EKLF gene in mice dramatically reduces β-globin expression (Nuez *et al.*, 1995; Perkins *et al.*, 1995; Wijgerde *et al.*, 1996). Interestingly, in these embryos a number of erythroid genes with functional

CACCC sites located within their promoters, including PBGD, are expressed at normal levels (Nuez et al., 1995; Perkins et al., 1995) suggesting that CACCC proteins other than EKLF can regulate these genes. However, expression of the ALAS2 gene was not examined in these mice and this is currently under investigation.

Since competition experiments indicated that the two CACCC sites bound EKLF with a similar affinity, this raised the possibility that EKLF may also regulate expression of the ALAS2 gene. Transactivation experiments performed in K562 cells, provided support for this, with the ALAS2 promoter being consistently transactivated approximately 3-fold by exogenously expressed EKLF (comparable to the 4-fold level observed with the β-globin promoter) and mutagenesis of the -54 CACCC sequence significantly inhibited this transactivation.

Although mutagenesis of the CACCC sequence in the ALAS2 promoter reduced expression (45-60%), in other promoters including human PBGD (Frampton et al., 1990) and glycophorin B (Rahuel et al., 1992), murine GATA-1 (Tsai et al., 1991) and EpoR (Zon et al., 1991) and the rat pyruvate kinase (Max-Audit et al., 1993), mutagenesis of a CACCC element almost abolished expression in erythroid cells. Similarly, mutagenesis of a nearby GATA site in these promoters also markedly reduce expression (Frampton et al., 1990; Tsai et al., 1991; Zon et al., 1991; Rahuel et al., 1992; Max-Audit et al., 1993) supporting the proposal that proteins binding to the GATA and CACCC sites function in a co-operative manner. There is also evidence for transcriptional synergism between CACCC and nearby GATA-1 binding sites based on co-transfection studies performed in Drosophila SL2 cells which are devoid of endogenous Sp factors (Merika & Orkin, 1995; Gregory et al., 1996). Indeed, the orientation of GATA and CACCC sites can influence whether Sp1 or EKLF bind to the CACCC site and functionally interact with GATA-1 (Gregory et al., 1996). In the present studies, mutagenesis of the GATA and CACCC sites in the ALAS2 promoter reduced but did not abolish expression or provide evidence for a co-operative interaction between these sites in the promoter. However, future experiments which are planned in the Drosophila SL2 cell line may shed light on whether there is a synergistic interaction between the GATA and CACCC sites in the ALAS2 promoter and also the role of EKLF or Sp1 in the transactivation of the ALAS2 promoter.

To investigate a possible role for BKLF, similar co-transfection assays were performed. Exogenously expressed BKLF transactivated the ALAS2 promoter, but not the

β-globin promoter, in K562 cells containing endogenous BKLF protein. However, BKLF did not transactivate the ALAS2 promoter through the -54 CACCC site, and attempts to identify the region through which BKLF functioned were unsuccessful. Exogenously expressed BKLF also transactivated the Rous sarcoma viral and thymidine kinase promoters, suggesting that BKLF may affect expression of a variety of promoters. In contrast, exogenously expressed BKLF did not transactivate the plasmid construct, p(CAC)$_4$tk-LUC or the ALAS2 promoter in non-erythroid COS-1 cells. These findings are in agreement with co-transfection studies performed by Dr. M. Crossley (personal communication) which have indicated that BKLF may function in co-operation with a general corepressor protein, where the binding of BKLF to this protein alleviates repression of a promoter construct. In conclusion, BKLF appears to possess a general function of transactivating a subset of promoters and is unlikely to be directly involved in the transcriptional activation of the human ALAS2 promoter *in vivo*.

In addition to GATA and CACCC box sequences, other possible binding sites including an NF-E2 site, a thyroid hormone response element (TRE) and an Ets site were identified in the ALAS2 promoter (see Fig. 5.1). These sites were also investigated to determine their role in promoter expression.

The putative NF-E2 site located at -49/-39, contained a mismatch at both extremities of the 11 bp consensus sequence for NF-E2 (Andrews *et al.*, 1993a) and partially overlapped the CACCC sequence by two nucleotides. Gel shift assays were performed with a probe containing the NF-E2 site and one of the CACCC sites but binding of either AP-1 or NF-E2 was not detected with erythroid cell nuclear extracts. However, the retarded protein complexes that were detected corresponded to the binding of nuclear proteins to the CACCC site, since the same binding profile was observed with the -54 CACCC probe using the same binding conditions. In addition, nuclear extracts isolated from COS-1 cells transfected with the cDNA expression clones encoding the subunits of NF-E2, p45 and p18, failed to detect binding of any additional protein complexes compared with mock-transfected COS-1 cell nuclear extracts. Mutational analysis of the NF-E2 site in the ALAS2 promoter demonstrated that this site was inactive in erythroid cells and expression was reduced to approximately 78% in COS-1 and is probably attributable to binding of AP-1 to this site as determined by gel shift competition assays.

In preliminary transactivation experiments performed in COS-1 cells, overexpression of recombinant NF-E2 protein (p45 and p18) failed to transactivate the ALAS2 promoter. However, two essential controls were required in these transactivation experiments. These included the analysis of COS-1 nuclear extracts co-transfected with the two cDNA clones encoding the subunits of NF-E2 (p45 and p18) with a consensus NF-E2 oligonucleotide for the detection of recombinant NF-E2 binding in gel shift assays. In addition, a promoter construct containing a functional NF-E2 binding site would also be required to demonstrate the transactivation of this promoter by exogenously expressed NF-E2.

In conclusion, the putative NF-E2 site identified in the ALAS2 promoter is probably non-functional in erythroid cells, since mutagenesis of the NF-E2 site did not alter expression in erythroid cells. This finding is in agreement with gel shift competition studies performed by Andrews et al. (1993a) where this sequence failed to inhibit binding of purified NF-E2 protein to a consensus NF-E2 binding site.

Functional binding sites for NF-E2 have been identified in the α- (Strauss et al., 1992; Zhang et al., 1995) and β- globin (Talbot et al., 1990; 1991; Ney et al., 1990) enhancers but are not a common occurrence in erythroid cell-specific promoter sequences. A functional NF-E2 site has only been identified in the promoter of the human PBGD gene (Mignotte et al., 1989b) but is not conserved in the murine PBGD gene (Porcher et al., 1991). Putative binding sites for NF-E2 have also been identified in the human ferrochelatase promoter (Tugores et al., 1994) and in the erythroid-specific enhancer element located in the promoter of the murine ferritin H subunit gene (Beaumont et al., 1994b). Therefore it appears that functional NF-E2 sites appear to be restricted to a subset of erythroid-specific expressed genes.

As described in Section 1.5.4, expression of the v-erbA oncoprotein in chick erythroblasts can arrest erythroid differentiation and suppress erythroid-specific gene transcription including the erythrocyte anion transporter band 3, erythroid-specific carbonic anhydrase II and ALAS2 which are normally upregulated during erythroid development (Zenke et al., 1988). Interestingly, overexpression of the thyroid hormone receptor as a gag-c-erbA fusion protein in erythroid cells can also inhibit erythroid differentiation and erythroid-specific gene transcription in the absence of thyroid hormone. However, this repression is relieved following the addition of thyroid hormone suggesting a possible role for thyroid hormone in the regulation in erythroid-specific genes. A thyroid hormone

response element that also resembles an NF-E2 binding site has been identified in the erythroid-specific carbonic anhydrase II promoter located at -660 bp from the transcription initiation site (Disela *et al.*, 1991). This element can confer thyroid hormone dependent transcriptional regulation to the heterologous thymidine kinase promoter in transient transfection assays and the v-erbA oncoprotein can also function through this site (Disela *et al.*, 1991).

A putative thyroid hormone response element (TRE) was identified in the 293 bp ALAS2 promoter located at -21/-6, adjacent to the -27 GATA site. This element 5'-CCAGGTCCTAACCCAA-3' shared extensive sequence homology to the palindromic TRE 5'-TCAGGTCATGACCTGA-3' described by Glass *et al.* (1988) with only a 5 bp mismatch. Preliminary gel shift analysis of the putative ALAS2 TRE (-21/-6) incubated with erythroid nuclear extracts failed to detect the binding of a protein complex to this sequence compared with the consensus TRE probe. In competition studies, the putative ALAS2 TRE did not affect the binding of retarded protein complexes obtained with the consensus TRE probe.

Mutagenesis of one of the putative ALAS2 TRE half sites did not alter promoter expression in erythroid cells suggesting that this sequence probably does not function as a thyroid hormone response element *in vivo*. These transient transfection studies performed in erythroid cells assumed that there was sufficient endogenous thyroid hormone receptor, and thyroid hormone present in the foetal calf serum, to detect an observable effect on expression following mutagenesis of the TRE. Future experiments would include co-transfecting the reporter construct with cDNA expression plasmids, encoding either the α or β thyroid hormone receptors and supplementing thyroid hormone to medium containing fetal calf serum stripped of steroid hormones. Consequently, in the absence of thyroid hormone, expression of the reporter construct would be repressed and then repression alleviated with the addition of thyroid hormone.

Other regions of the ALAS2 gene which may respond to thyroid hormone through a thyroid hormone response element may exist and can be investigated in transient transfection assays. The 5'-flanking ALAS2 promoter/intronic luciferase reporter constructs described in Section 4.2.2, could be used in the identification of possible thyroid hormone response elements located in the ALAS2 gene which may be subject to transcriptional regulation by thyroid hormone.

An Ets-like sequence located at -114 to -105, between the -124 and -100 GATA sites (see Fig. 5.1) was identified by its sequence similarity to the enhancer of the adenovirus type 5 E1A enhancer sequence which binds the ets factor, E1A-F (Higashino *et al.*, 1993). As described in Section 1.5.5, several lines of evidence have implicated a possible role for members of the Ets family of transcription factors in erythropoiesis. Studies performed by Metz and Graf (1991a; 1991b; 1992) demonstrated the requirement for a functional interaction between v-Ets and v-Myb or v-erbA in the transformation of erythroblast-like cells, suggesting that they may regulate specific target genes. The Myb-Ets fusion protein of the ME26 virus is capable of directly transactivating the GATA-1 promoter indicating a possible role in the activation of GATA-1 (Aurigemma *et al.*, 1992). An interaction between the myeloid expressed MafB protein and c-ets-1 via their DNA binding domains can inhibit erythroid differentiation and repress expression of the transferrin receptor gene, α and β- globin and PBGD, suggesting a possible mechanism for repressing erythroid-specific gene expression in myeloid cells (Sieweke *et al.*, 1996). However, expression of ALAS2 was not examined. Although members of the Maf family of proteins can heterodimerise with the NF-E2 p45 protein, the MafB protein does not heterodimerise with p45 and inhibit erythroid-specific gene expression (Igarashi *et al.*, 1994).

The association of Ets binding sites with GATA and CACCC box binding sites have been observed in the tissue-specific enhancer of the erythroid cytosolic glutathione peroxidase gene (O'Prey *et al.*, 1993). Functional GATA-1 and Ets binding sites have also been identified in the promoter of the megakaryocyte-specific human glycoprotein IIB gene and exogenously expressed c-ets-1 and human GATA-1 can transactivate this promoter in HeLa cells (Lemarchandel *et al.*, 1993).

Mutational analysis of the core ets motif GGAA to CCAA in the ALAS2 promoter did not affect promoter expression in erythroid cells. Examination of the proteins binding to an oligonucleotide probe encompassing both GATA-1 binding sites and the putative -110 Ets site, only detected a single retarded protein complex in gel shift assays. This complex was effectively inhibited by a β-globin GATA consensus oligonucleotide and corresponded to GATA-1. Consequently, the putative -110 Ets site is probably not functional. However, this finding does not eliminate the possibility that the ALAS2 gene may be subjected to

transcriptional regulation by the Ets family of transcription factors, since numerous putative Ets binding sites have been identified throughout the ALAS2 gene.

Northern blot analysis of MEL and K562 cell lines performed by O'Prey *et al.* (1993) demonstrated that these cell lines express very low levels of ets-1 and ets-2 whereas MEL cells express significant levels of Fli-1 and PU.1. Co-transfection assays could therefore be performed with ALAS2 5'-flanking/intronic constructs and exogenously expressed Ets factors in erythroid cell lines which synthesise reduced levels of endogenous Ets proteins. Such an approach may assist in the identification of "Ets" responsive regions of the ALAS2 gene. Similar co-transfection assays could be performed with exogenously expressed MafB and c-ets-1 in erythroid cells to examine whether endogenous ALAS2 mRNA is repressed in the presence of these factors, and possibly identify regions of the ALAS2 gene subjected to transcriptional repression through the use of these ALAS2 reporter constructs. The transfection of erythroid cell lines with ALAS2 reporter constructs and the availability of cDNA expression clones encoding Ets proteins such as ets-1, ets-2, Fli-1 or PU.1 provides an excellent model system to investigate further a possible role for Ets protein(s) in ALAS2 expression.

A putative CCAAT box located at -90/-84 in the ALAS2 promoter was also identified (see Fig. 5.1) and is identical to the consensus sequence for the transcription factor CP1 (Chodosh *et al.*, 1988). This sequence is also identical (5'-GGCCAAT-3') to the functional CCAAT box located at approximately -70 in the human (deBoer *et al.*, 1988) and murine (Cowie & Myers, 1988) β-globin promoters. Examination of the corresponding region in the murine ALAS2 promoter revealed a CCAAT-like box. In the present study, time did not permit a mutational analysis of the putative CCAAT box in the human ALAS2 promoter, but based on its homology with the globin sequence it seems highly likely that it is functional. This sequence may be responsible for the remaining activity of the transfected promoter construct in erythroid cells following mutagenesis of both the -124 GATA and -100 GATA binding sites and the CACCC sequence.

A number of CCAAT-box binding proteins have been identified, which are either tissue-specific or ubiquitously expressed (Johnson & McKnight, 1989). These include the ubiquitous factor CP1 (also known as CBF or NF-Y) composed of a heterodimeric complex of at least three subunits, A, B and C (all necessary for DNA binding) (Kim *et al.*, 1996b), and C/EBPγ, which is highly expressed in MEL cells (Wall *et al.*, 1996).

Several lines of evidence have demonstrated the functional importance of CCAAT boxes in the regulation of other erythroid-specific genes. The promoters of both the human embryonic ζ globin (Watt *et al.*, 1990) and β-globin (deBoer *et al.*, 1988) genes and the murine EKLF gene (Crossley *et al.*, 1994a) all contain functional GATA sites and CCAAT boxes that are essential for erythroid-specific transcriptional activation. However, a physical association of the proteins binding to these sites has not been reported. Interestingly, the CCAAT box binding protein BB1, is essential for repression of murine β-globin gene transcription in undifferentiated MEL cells (Macleod & Plumb, 1991). A number of naturally occurring mutations in the distal CCAAT box of the γ-globin gene promoter have been identified in patients with Hereditary persistence of fetal hemoglobin (HPFH) resulting in continued expression of the γ-globin gene in the adult (Katsube & Fukumaki, 1995), highlighting the functional significance of CCAAT box binding proteins in globin gene transcription. Marziali *et al.* (1997) demonstrated that the ferritin H-gene promoter contains a functional CCAAT binding site for CP1/NF-Y which is required to sustain ferritin H-gene transcription by heme.

The studies described in this chapter examined the role of a number of putative transcription binding sites in ALAS2 promoter expression by transient transfection studies performed in uninduced erythroid cells. However, expression of the ALAS2 gene is significantly transcriptionally upregulated in erythroid cells induced to differentiate. Therefore to determine whether the functional GATA-1 and CACCC box binding sites are also major contributors to ALAS2 promoter expression, ALAS2 promoter constructs with these mutated sites could be stably transfected into erythroid cells, induced to differentiate and expression examined. This system would also permit an examination of gene regulation under more physiological conditions with proper chromatin assembly. These studies are planned for the future.

All of the eight enzymes of the heme biosynthetic pathway have now been cloned (May *et al.*, 1995; Taketani *et al.*, 1995). The large requirement for heme during erythropoiesis in contrast to non-erythroid cells, may have necessitated the evolution of distinct transcriptional regulatory processes for the expression of these genes in erythroid cells. To highlight this, there are two genes encoding the rate-limiting ALAS enzyme, the housekeeping gene and the erythroid gene and these are located on different chromosomes (Sutherland *et al.*, 1988; Cox *et al.*, 1990; Bishop *et al.*, 1990). As expected, the promoter

architecture of the housekeeping gene is different from that of the erythroid gene and contains multiple binding sites for the ubiquitous transcription factors, Sp1 and NRF-1 (Braidotti *et al.*, 1993). As described in this chapter, the ALAS2 promoter contains functional binding sites for GATA-1 and CACCC box binding proteins which are a common feature of many erythroid cell expressed genes (deBoer *et al.*, 1988; Mignotte *et al.*, 1989b; Tsai *et al.*, 1991; Zon *et al.*, 1991; Rahuel *et al.*, 1992; Max-Audit *et al.*, 1993). In contrast to ALAS, there is only one structural gene for the other enzymes of the heme pathway and these have either a composite promoter which contains binding sites for both ubiquitous and erythroid-specific transcription factors (May *et al.*, 1995; Taketani *et al.*, 1995) or, alternatively, two separate promoters, one with a housekeeping function and the other that is erythroid cell-specific, including the human ALAD and PBGD genes (Kaya *et al.*, 1994; Chretien *et al.*, 1988). Functional sites for GATA-1, NF-E2 and CACCC box binding proteins have been characterised in the erythroid promoter for human PBGD (Mignotte *et al.*, 1989b), although the NF-E2 site is absent from the corresponding murine PBGD promoter (Porcher *et al.*, 1991). Binding sites for GATA-1, NF-E2 and CACCC box binding proteins have also been identified in the promoter of the human ferrochelatase gene, which is transcribed from a single promoter in both erythroid and non-erythroid cells (Tugores *et al.*, 1994). These studies, together with the information on globin gene expression (reviewed by Orkin, 1995a), confirm that there is likely to be only a small number of erythroid cell-specific factors that act in a combinatorial fashion to ensure the coordinated regulation of heme and globin synthesis during erythropoiesis.

CHAPTER SIX

CHARACTERISATION OF AN ERYTHROID-SPECIFIC

ENHANCER LOCATED WITHIN INTRON 8

OF THE HUMAN ALAS2 GENE

CHAPTER SIX : CHARACTERISATION OF AN ERYTHROID-SPECIFIC ENHANCER LOCATED WITHIN INTRON 8 OF THE HUMAN ALAS2 GENE

6.1 INTRODUCTION

Transient expression studies described in Section 4.2.2, demonstrated that intron 8 of the human ALAS2 gene was capable of increasing expression of both the human ALAS2 promoter and the thymidine kinase promoter in erythroid cells and this activity corresponded with a DNase I hypersensitivity site located in this intron in the murine ALAS2 gene (Schoenhaut & Curtis, 1989). Studies described in this chapter, further investigate the erythroid-specific enhancer activity identified within the human ALAS2 intron 8 sequence. As shown in this chapter, sequence comparison of intron 8 of the human, murine and canine ALAS2 genes identified extensive sequence conservation. Transient transfection studies described in this chapter demonstrate that the corresponding fragment from the murine intron 8 region is also transactivated the thymidine kinase promoter in erythroid cells only.

To localise the enhancer activity within the human intron 8 sequence, a series of ALAS2 intron 8 deletion constructs were synthesised and examined in transient expression studies in the erythroid K562 cell line. The elements responsible for the erythroid-specific enhancer activity of intron 8 in erythroid K562 and MEL cell lines, were identified by mutational analysis and the identity of the proteins binding to these sites determined by gel shift assays.

6.2 RESULTS

6.2.1 Conservation of ALAS2 Intron 8 Sequence

A detailed comparison of the intron 8 sequence from the human and murine ALAS2 genes was performed based on the correlation between a DNase I hypersensitive site in intron 8 of the murine ALAS2 gene (Schoenhaut & Curtis, 1989) and the enhancer function of intron 8 of the human ALAS2 gene as described in Section 4.2.2. Intron 8 sequence from the canine ALAS2 gene was also available through the canine genome mapping project. A comparison of sequence across all three species was performed and is shown in Fig. 6.1

(Cox, 1993). This comparison revealed extensive sequence conservation between all three at the 5' end, and between the human and canine sequences with the 3' end of the murine intron 8 sequence (see Fig. 6.1, nucleotides 545 to 1150 in the ALAS2 intron 8 sequence). The murine intron 8 contains additional 5' sequence (Fig. 6.1, nucleotides 21 to 541) which is not present in the human or canine intron 8 sequences and exhibited sequence similarity to a murine-specific repetitive element (Cox, 1993). Therefore the insertion of this element is thought to have post-dated the divergence of the murine, canine and human lineages. Of particular note in the intronic sequence similarity (see Fig. 6.2A), is the complete conservation of two putative GATA-1 binding sites (902-907, 936-941) and one CACCC box binding site (884-893), and a less conserved CACCC sequence (817-829) present in all three species.

It was of interest to investigate whether intron 8 in the murine ALAS2 gene could also confer erythroid-specific enhancer activity, as seen with the human gene; a 330 bp intron 8 fragment containing these conserved binding sites, was ligated in the reverse orientation upstream of the thymidine kinase promoter (ptk-mI8(330)R-LUC) as described in Section 2.4.2. This construct was co-transfected with pRSV-βgal and transiently expressed in K562, MEL and COS-1 cells as described in Section 2.6.

As described in Section 4.2.2, a 460 bp human ALAS2 intron 8 fragment in the reverse orientation transactivated the thymidine kinase promoter 4 and 2 -fold in K562 and MEL cells, respectively. Similarly, the murine ALAS2 intron 8 fragment oriented in the same direction, transactivated the thymidine kinase promoter 7 and 2 -fold in K562 and MEL cells, respectively, and only 1.2-fold in COS-1 cells (Fig. 6.2B). Therefore the corresponding region from the murine ALAS2 gene also contains comparable *trans*-acting activity.

6.2.2 Localisation of Erythroid-Specific Enhancer Activity to a 235 bp Fragment Within ALAS2 Intron 8

To identify the functional control elements within human ALAS2 intron 8 responsible for enhancer activity in erythroid cells, a series of intron 8 deletion constructs were synthesised. Different lengths of the intron ranging from 460 bp to 115 bp, were generated by the polymerase chain reaction and cloned in the natural orientation in the

Figure 6.1   Comparison of the intron 8 sequences from the human, murine and canine ALAS2 genes

Alignment of the ALAS2 intron 8 nucleotide sequence from the human, murine and canine genes. Regions of sequence homology are shaded. Conserved putative GATA and CACCC box binding sites and non-conserved GATA sites are *boxed*. Two non-conserved putative GATA motifs in the human ALAS2 intron 8 sequence are also *boxed*.

```
human ALAS2 Intron 8  1   A     AAT   T  GCC .................................................  50
dog ALAS2 Intron 8    1      ---C  G  ATG .................................................  50
mouse ALAS2 Intron 8  1   A  ---   A  ATAAGGTTAATATTACTATGATGAAATATAGTCA                    50

                     51   ..................................................................  100
                     51   ..................................................................  100
                     51   GAAGCAACTTGGGGAGGAAAGGGCTTACGTGGCTTACATATCCTGTATCA                  100

                    101   ..................................................................  150
                    101   ..................................................................  150
                    101   CAAGCCCACTGAGGGAAAACCAAGGCATGAACTCAAACCAGGTAAGAACCT                 150

                    151   ..................................................................  200
                    151   ..................................................................  200
                    151   GGAGGCAGGAGCTGATTCAGAGATGTTGGTTACTGGCTTACTCCTCTTGG                  200

                    201   ..................................................................  250
                    201   ..................................................................  250
                    201   TTTGTTCAGCCTGCTCGCTGATAGAACACGGTAACATGAGCCCACAGTGG                  250

                    251   ..................................................................  300
                    251   ..................................................................  300
                    251   TATTACCCCCAAAGGTCTGGGCCATCCCACATCAATCACTAATTATAAAA                  300

                    301   ------------------------------------------------------------------  350
                    301   ------------------------------------------------------------------  350
                    301   ATGCTCTGCAGCCTAGTCTTACACATTTTCCCTCCTCTCAGATGACTCTT                  350

                    351   ..................................................................  400
                    351   ..................................................................  400
                    351   GCTTGCATCAAGTTGACATAAAGCTAGCCAGCAGTGAGCAAACACTTGTA                  400

                    401   ..................................................................  450
                    401   ..................................................................  450
                    401   TATAAAGGCTTCCCTTATGTACAATGTGGAGAAAGCACCCCCAGCAAACA                  450

                    451   ------------------------------------------------------------------  500
                    451   ------------------------------------------------------------------  500
                    451   GTTGTTGGGAAAGAATTTATAACCCCCGGCTTTTGTTTTTTGTTTTTGTT                  500

                    501   -------------------------------------------TT                       550
                    501   -------------------------------------------CC                       550
                    501   TTTTAACTTGACATGAACCATTTTTATTTTTATTTATTTTTGAGA            A           550

                    551   CC     AGAGGAGGC C  A -----------C       T      T                   600
                    551   AA    -----------------                   GAC          T            600
                    551   T T AT  TTTCCTCCTT T  CCCCTCTTTCCT C       T  CC AG A               600

                    601   CTCCTCCCC----------------  -TATAGGGGTTGTTTGGAC                       650
                    601   CTCCTCCCC----------------  -TATAGAAGTTATTTGGGT                       650
                    601   CCCCACCTTGTCATTAATATCTGTT  G ATTATAGAAATGACATCAAAT                  650

                    651        C      --   C   CC                      TCT                    700
                    651     C        --    T       A G AAC T       T        C                  700
                    651   TCT CAT  AAA                  ---    -- CAAA    A    T TT             700

                    701   CGCATCCCACTGCAGAGTTGATGAGAAAGTCTGGTAGAGTTTTTTGAAAA                  750
                    701   CCTACCCCACTGCAGAGTTGCTGAGAAAGTCTGGGAGGGGTTTTTGAGAA                  750
                    701   CCCATCTCAATGCAGAGTCACCAAGAAAGTCTAGAAACGTTTTTCTAAAA                  750

                    751    -T G   --CCAAATAGTTAGATTCAACTT      TTGA                           800
                    751   T A- CT G   GCCAAATAGCCAGGTTCAAGTC-     C  --G                      800
                    751   G T  -CC  ACAAAACAGCTGGGTTCAACCA        -----  A                    800

                    801   GCTGTGTTTCTAAACCCCTGCCCCACCCTAGCCCCAAGCTTCATCTTAGC                  850
                    801   GTTGTAGTTCTAAATTCCTCCCCCACTCCAGCCCCAACCTTCATTTTAGT                  850
                    801   ACTGTT-------TTCCT-CCTCACCG-AGCTCCAAGCCTCATCTTAGT                   850

                    851   -----GACCCTATCCAGCTAAAGGTCCCCACCCCAGCTCCTG                          900
                    851      TT -----GGGACTAACCAGCTAAAGGTCCCCACCCAGCTACCA                      900
                    851   ATT    AACCAGGCCCTAACCAAGTAACAGTCCCCACCCTGCTCCTG                    900

                    901   CCTATCTAGTCATTGCATATGGCAAGACTTGAAAGTCCTATCTCAAAGCA                  950
                    901   CCTATCTCATCACTGCATGTGGCAGGCCTTGGAAGTCCTATCTCAAAGCA                  950
                    901   CCTATCTAGTCATCACACATGGCAGTATGTGACAGTCCTATCTCAGAACC                  950

                    951   GCAGAATTATCAGCTACGACTGCCTTGTCATGGACA-----GATGAGCAG                  1000
                    951   GCAGAACTATCAGCT--GTCTGTCCTGTCATGGACAAGAGAGAGATAAGCAG                1000
                    951   ACAGAA------GCAAG----GCTCTAACATGGACAGGAGAGAGATGTGCAG                1000

                   1001   AGGCCTGGGAAGACAGCCTGGAGGCCCAACTTCTGGTGCACCCGCCTTGTG                 1050
                   1001   AGGTCTGGGAAAGGAATCTGGAACTCCAGCTTCTGGTGCACATCTTCATG                  1050
                   1001   AGGACTGGGAAGCGAGTCTGAAAGCCCCAGCTTCTAGTGTCCCGGCCTGTG                 1050

                   1051   -TTATCTGGCACATGATCCTGTTG            -          A                    1100
                   1051   TTTATCTGG-ACATAATCCTGTC     T  A      A          C                  1100
                   1051   TTCCCAGGGGGATGAACCAG---     G         A    T CT      T              1100

                   1101   T       T ATT  TT C G                                               1150
                   1101   C       C CCC  CA -A                                                1150
                   1101   -----------------------                                             1150
```

Figure 6.2   Sequence comparison of highly conserved CACCC boxes and GATA sites within intron 8 and transient expression analysis of a murine intron 8 fragment containing these sites

A   Comparison of two CACCC box sequences and two GATA-1 binding sites identified within intron 8 that are highly conserved between the human, murine and canine ALAS2 genes.

B   The construct, ptk-mInt8(330)R-LUC containing a 330 bp fragment of murine ALAS2 intron 8 sequence ligated upstream of the thymidine kinase promoter and ptk-LUC were co-transfected with pRSV-βgal and transiently expressed in K562, MEL and COS-1 cells. The normalised luciferase activities are expressed relative to ptk-LUC assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. The transactivation of the thymidine kinase promoter by the human ALAS2 intron 8 fragment is also shown. The orientation of the ALAS2 intron 8 fragment is indicated by the *arrow*.
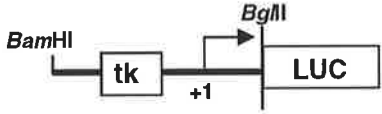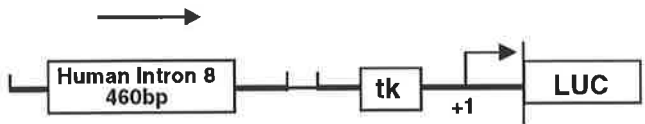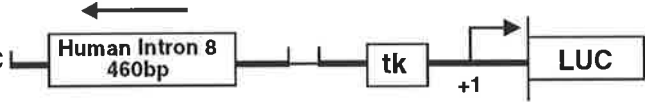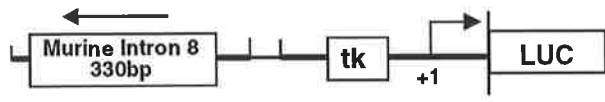
A

CACCC box

	817-829	884-893
	5' CCTCCCCCACCCT 3'	5' TCCCCACCCA 3'
Human	5' CCTCCCCCACTCC 3'	5' TCCCCACCCA 3'
Canine	5' CCT-CCTCACCC- 3'	5' TCCCCACCCT 3'
Murine		

CACCC box

	817-829	884-893
Human	5' CCTCCCCCACCCT 3'	5' TCCCCACCCA 3'
Canine	5' CCTCCCCCACTCC 3'	5' TCCCCACCCA 3'
Murine	5' CCT-CCTCACCC- 3'	5' TCCCCACCCT 3'

GATA-1 binding motif

	902-907	936-941
Human	5' CTATCT 3'	5' CTATCT 3'
Canine	5' CTATCT 3'	5' CTATCT 3'
Murine	5' CTATCT 3'	5' CTATCT 3'

B

	Fold Induction		
	K562	MEL	COS-1
ptk-LUC	1.0	1.0	1.0
ptk-Int8(460)-LUC	25±4	12±1.5	1.8±0.2
ptk-Int8(460)R-LUC	4±1.3	2±0.5	ND
ptk-mInt8 (330)R-LUC	7±1.5	2±0.1	1.2±0.1

plasmid pALASp-LUC as described in Section 2.4.3. The restriction sites used to generate the deletion fragments of human intron 8 and the putative GATA and CACCC box sites are indicated in Fig. 6.3A. The erythroid K562 cell line was chosen for the testing of these intron 8 deletion constructs since the intron 8 enhancer activity was greater in these transiently transfected cells compared with MEL cells. These ALAS2 intron 8 deletion constructs, the plasmid pALASp-Int8(460)-LUC containing 460 bp of intron 8 sequence and the 293 bp ALAS2 promoter plasmid pALASp-LUC were co-transfected with the pRSV-β-gal in K562 cells and assayed as previously described. Expression of these intron 8 deletion constructs were then compared to the plasmid, pALASp-Int8(460)-LUC.

As seen in Fig. 6.3B, constructs containing 400 bp (pALASp-Int8(400)-LUC) and 235 bp (pALASp-Int8(235)-LUC) of ALAS2 intron 8 sequence, transactivated the human ALAS2 promoter approximately 11.5 and 12.9 -fold, respectively, in K562 cells compared with 12-fold obtained with 460 bp of intron 8 sequence (pALASp-Int8(460)-LUC). These findings demonstrate that the full enhancer activity is located within 235 bp of ALAS2 intron 8 sequence. However, deletion of a 120 bp fragment (containing a CACCC-like sequence) from the 5' end of the 400 bp intron 8 sequence (pALASp-Int8(279)-LUC), reduced the level of transactivation to 3.2-fold in K562 cells (Fig. 6.3B). In contrast, the 3' end of intron 8 sequence containing two putative GATA-1 binding sites (pALASp-Int8(177)-LUC) did not affect expression (Fig. 6.3B).

The 235 bp intron 8 fragment that contained full enhancer activity was divided into two smaller fragments of 120 bp and 115 bp. However, 120 bp of intron 8 sequence (pALASp-Int8(120)-LUC) did not transactivate the ALAS2 promoter (Fig. 6.3B). The plasmid construct pALASp-Int8(115)-LUC containing two putative GATA-1 binding sites in intron 8 that are evolutionary conserved and a conserved CACCC box, transactivated the ALAS2 promoter approximately 2.3-fold in K562 cells (Fig. 6.3B). These results established that the erythroid-specific enhancer activity of the 235 bp fragment is significantly reduced when this fragment is divided into two regions of 120 bp and 115 bp, suggesting that there may be a cooperative interaction between proteins binding to these sites within these two regions.
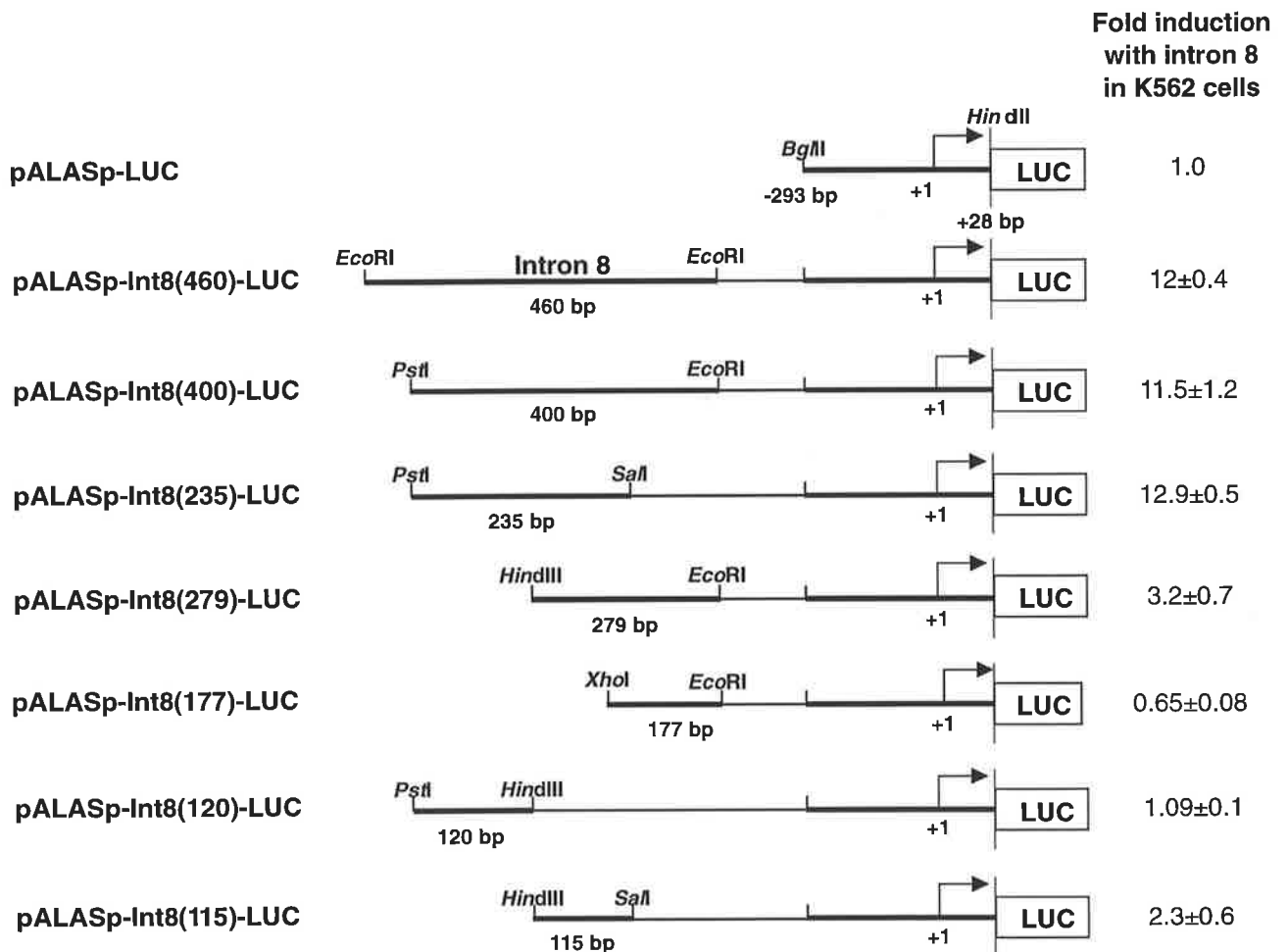
Figure 6.3   Deletion analysis of the human ALAS2 intron 8 enhancer

A   Nucleotide sequence and restriction map of the human ALAS2 intron 8 sequence. The polymerase chain reaction was employed to generate deletion fragments of intron 8 sequence (from 460 bp to 115 bp). The native and introduced (*asterixed*) restriction sites are indicated. The orientation of the primers used in the polymerase chain reaction are depicted by the direction of the *arrow*. The putative GATA-1 binding sites and the CACCC sites in intron 8 are *boxed*.

B   A series of human ALAS2 intron 8 deletion constructs and the plasmid pALASp-LUC, containing 293 bp of ALAS2 promoter were co-transfected with pRSV-βgal and transiently expressed in K562 cells. Luciferase activities were standardised relative to β -galactosidase activity as an internal control. The normalised luciferase activities are expressed relative to pALASp-LUC assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation.

A

```
5'-TTCTTATATACCCTCCAGAGAGGAGGCCCTTACAAAATTCTTTTCTG
```

```
                                                    EcoRI
CCTCCTCCCCAAAGCTATAGGGGTTGTTTGGACAGAATTCACAGCCCCAG
```

```
                                                    PstI
GCGTCTGCCATCCTGGACTCCCTCTCTCCACTCGCATCCCACTGCAGAGT
```

```
TGATGAGAAAGTCTGGTAGAGTTTTTTGAAAAGACCTTGAACTAGGCCAA
```

```
ATAGTTAGATTCAACTTGAGTATGTGAAGAGCTGTGTTTCTAAACCCCTC
```

```
             XhoI*       HindIII
CCCCACCCTAGCCCCAAGCTTCATCTTAGCTCCACTCCTGACCCTATCCA
```

```
GCTAAAGGTCCCCACCCAGCTACTGCCTATCTAGTCATTGCATATTGGCA
```

```
                 XhoI*           SaII*
AGACTTGAAAGTCCTATCTCAAAGCAGCAGAATTATCAGCTACGACTGCC
```

```
TTGTCATGGACAGATGAGCAGAGGCCTGGGAAGACTGGAGCCCCAACTTC
```

```
                         EcoRI*
TGGTGCACCCCCTTGTGTTATCTGGCACCATGATCCTGTTGCTCTGGGAC
```

```
TGATTATGGGATCTGTGTATATCTTATTCCTTTCTGTCTCCAG-3'
```

B

	Fold induction with intron 8 in K562 cells
pALASp-LUC	1.0
pALASp-Int8(460)-LUC	12±0.4
pALASp-Int8(400)-LUC	11.5±1.2
pALASp-Int8(235)-LUC	12.9±0.5
pALASp-Int8(279)-LUC	3.2±0.7
pALASp-Int8(177)-LUC	0.65±0.08
pALASp-Int8(120)-LUC	1.09±0.1
pALASp-Int8(115)-LUC	2.3±0.6

6.2.3 Identification of Functional Control Elements in Human ALAS2 Intron 8

Sequence analysis of the 235 bp enhancer of intron 8 (see Fig. 6.3A, *PstI-SalI* fragment), revealed two putative CACCC box binding sites and two putative GATA sites located on the non-coding strand (see Fig. 6.3A). Interestingly, two corresponding GATA sites, designated GATA site A and GATA site B, were identified in the murine and canine intron 8 sequences (see Fig. 6.2A and Fig. 6.4A). Also, the two putative CACCC boxes identified in the human sequence were present in the corresponding regions of the murine and canine intron 8 sequences (see Fig. 6.2A and Fig. 6.4A). The first CACCC site designated CACCC site A was partially conserved whereas the second CACCC site, CACCC site B was identical in sequence in all three species (Fig. 6.4A).

The functional contributions of the conserved GATA-1 and CACCC box binding sites identified in the human intron 8 sequence (Fig. 6.4) were investigated by mutating these sites individually or in various combinations in the plasmid pALASp-Int8(235)-LUC as described in Section 2.4.4. These mutant constructs and the plasmid pALASp-Int8(235)-LUC containing the wild-type intron 8 sequence, were co-transfected with pRSV-βgal and transiently expressed in K562 and MEL cells as described in Section 2.6. The relative luciferase activities of the constructs were expressed relative to pALASp-Int8(235)-LUC, assigned a value of 100%.

As seen in Fig. 6.4B, mutagenesis of CACCC site A in intron 8 (pALASp-mut1-LUC), substantially reduced expression of the enhancer in K562 and MEL cells to 37% and 47%, respectively, relative to wild-type. Similarly, mutagenesis of CACCC site B in intron 8 (pALASp-mut2-LUC) reduced enhancer activity in K562 and MEL cells to 43% and 58%, respectively (Fig. 6.4B). Inactivation of both these CACCC sites (pALASp-mut5-LUC) further reduced expression to 21% and 35%, respectively. Interestingly, the conserved GATA site A did not contribute to enhancer activity as indicated by mutagenesis of this site (pALASp-mut3-LUC), and in fact activity was marginally increased to 119% and 122% in K562 and MEL cells, respectively (Fig. 6.4B). The non-functional contribution of this site was confirmed by a double mutation of GATA site A with CACCC site A which expressed at 49% of the wild-type level in K562 cells (pALASp-mut6-LUC) and a marginal increase in expression of the enhancer was again observed following mutagenesis of the GATA site A. In comparison, GATA site B was important for enhancer activity since mutagenesis of this site alone (pALASp-mut4-LUC)

Figure 6.4   Effect of mutating conserved CACCC boxes and GATA sites on human ALAS2 intron 8 enhancer activity

A   Sequence comparison of putative CACCC box and GATA-1 binding sites identified within the 235 bp human ALAS2 intron 8 enhancer with the murine and canine ALAS2 intron 8 sequences. These sites are designated as CACCC site A and CACCC site B, and GATA site A and GATA site B, respectively. The orientation of these sites (*arrowed*) and the approximate spacing between each of these sites are indicated.

B   CACCC site A, CACCC site B, GATA site A and GATA site B were each mutated to a *Pvu*II site represented by a *cross* in the 235 bp fragment of the human ALAS2 intron 8 enhancer and ligated upstream of the 293 bp ALAS2 promoter. Constructs containing the mutated CACCC site A and CACCC site B (ptk-mut1-LUC) or mutated CACCC site A and GATA site B (ptk-mut2-LUC) in the 235 bp fragment of the human ALAS2 intron 8 enhancer were also ligated upstream of the thymidine kinase promoter. These constructs were co-transfected with the β-galactosidase expression construct, pRSV-βgal and transiently expressed in K562 and MEL cells. The normalised luciferase activities of the mutant constructs are expressed relative to pALASp-Int8(235)-LUC containing 235 bp of wild-type intron 8 sequence ligated upstream of the 293 bp ALAS2 promoter or ptk-Int8(460)-LUC containing 460 bp of wild-type human ALAS2 intron 8 sequence, respectively, (set at 100%). The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± standard deviation. ND (not determined) corresponds to those constructs not tested in a particular cell line.

A

Human	5'-CCCCACCC-3'	59 bp	5'-CCCCACCC-3'	15 bp	5'-CTATCT-3'	36 bp	5'-CTATCT-3'
Murine	5'-CCTCACCC-3'	63 bp	5'-CCCCACCC-3'	15 bp	5'-CTATCT-3'	36bp	5'-CTATCT-3'
Canine	5'-CCCCACTC-3'	59 bp	5'-CCCCACCC-3'	15 bp	5'-CTATCT-3'	36 bp	5'-CTATCT-3'

CACCC site A   CACCC site B   GATA site A   GATA site B

B

	Relative luciferase activity	
	K562	MEL
pALASp-Int8(235)-LUC	100	100
pALASp-mut1-LUC	37±5	47±4
pALASp-mut2-LUC	43±7	58±3
pALASp-mut3-LUC	119±12	122±6
pALASp-mut4-LUC	36±3	45±7
pALASp-mut5-LUC	21±1.5	35±3
pALASp-mut6-LUC	49±7	ND
pALASp-mut7-LUC	7±1.2	21±4.3
ptk-Int8(460)F-LUC	100	100
ptk-mut1-LUC	62±6	34±3.5
ptk-mut2-LUC	17±1.5	11.5±4

reduced expression in K562 and MEL cells to 36% and 45%, respectively. Mutagenesis of both CACCC site A and GATA site B (pALASp-mut7-LUC) significantly reduced expression in K562 and MEL cells to 7% and 21%, respectively, thus severely inhibiting the ability of intron 8 to transactivate the ALAS2 promoter.

The effect of these mutations introduced into intron 8, on conferring erythroid-specific enhancer activity to the heterologous thymidine kinase promoter was also examined. A 235 bp intron 8 fragment containing mutations in both the CACCC site A and CACCC site B or CACCC site A and GATA site B was ligated upstream of the thymidine kinase promoter. Expression of these mutant constructs were determined and expressed relative to the plasmid ptk-Int8(460)-LUC containing 460 bp wild-type intron 8 sequence, since deletion analysis had demonstrated that the 235 bp intron 8 fragment functioned as efficiently as the 460 bp fragment. As seen in Fig. 6.4B, mutagenesis of CACCC site A in combination with CACCC site B (ptk-mut1-LUC) reduced enhancer activity in K562 and MEL cells to 62% and 34%, respectively, relative to wild-type. Inactivation of both CACCC site A and GATA site B (ptk-mut2-LUC) almost abolished the enhancer activity in these cells (Fig. 6.4B). These studies therefore establish that proteins binding to the CACCC sites A and B and the GATA site B contribute substantially to the erythroid-specific enhancer activity located within intron 8, with a possible cooperative interaction between the proteins binding to these sites.

6.2.4 GATA-1 Protein Binds to the Erythroid-Specific Enhancer in Human ALAS2 intron 8

As discussed above, there are two putative GATA sites, A and B which are conserved between the human, canine and murine ALAS2 intron 8 sequences (Fig. 6.4A) and located in the 235 bp enhancer region of human ALAS2 intron 8. Both of these sites are identical in the core sequence 5'-AGATAG-3', located on the non-coding strand, and conform to the consensus GATA-1 binding site (Merika & Orkin, 1993; Ko & Engel, 1993). However, mutational analysis of these GATA sites inferred that only GATA site B was important for enhancer activity in transiently transfected erythroid cells. The binding of nuclear proteins to both of these sites was investigated using GATA site A and GATA site B probes (Section 2.1.7) in gel shift assays with nuclear extracts from K562,

MEL and COS-1 cells and COS-1 cells expressing recombinant murine GATA-1 (Section 2.9). A β-globin GATA-1 consensus sequence (GATA-cons) was included as a control probe (Wall *et al.*, 1988).

As seen in Fig. 6.5, a major retarded complex was obtained with the GATA site B probe (lanes 6 and 7) using nuclear extracts from K562 and MEL cells. A complex of the same mobility was detected with the GATA site A probe (lanes 1 and 2) although the intensity was reduced (Fig. 6.5). The major retarded protein complex was also observed with all three probes using nuclear extracts from COS-1 cells expressing recombinant murine GATA-1 (lanes 4, 9 and 11) but was not detected with nuclear extracts from mock-transfected COS-1 cells (lanes 3 and 8). This major retarded protein complex was confirmed immunologically as GATA-1, since it was substantially supershifted with the GATA-1 monoclonal antibody, N-6 (Ito *et al.*, 1993) (lanes 5, 10 and 12). These results demonstrate that both GATA sites A and B within intron 8, can bind GATA-1 *in vitro* but only GATA site B is important in enhancer function as determined by mutational analysis.

6.2.5 Gel Shift Analysis of CACCC Sites Located Within Human ALAS2 Intron 8

As described in Section 5.2.2, several transcription factors bind *in vitro* to CACCC boxes, including Sp1 (Kadonaga *et al.*, 1987), Sp1-related proteins (Kingsley & Winoto, 1992), CAC C and CAC D (Hartzog & Myers, 1993), BKLF (Crossley *et al.*, 1996) and EKLF (Miller & Bieker, 1993). Gel shift assays were performed with radiolabelled probes containing either CACCC site A and CACCC site B (Section 2.1.7) to investigate nuclear protein binding to these sites. The CACCC site from the murine adult β-globin promoter which binds EKLF, BKLF and Sp1 *in vitro* (Crossley *et al.*, 1996) (β-globin CACCC) and an Sp1 consensus sequence (Sp1-cons) were included as control probes.

The binding of nuclear proteins to the CACCC site A and CACCC site B probes with nuclear extracts from K562 and MEL cells was extremely weak. Therefore in order to visualise the retarded complexes binding to these probes, these gel shift assays were exposed for a longer period compared with gel shift assays performed with the control probes. As a result, it is very difficult to see the retarded complexes obtained with the CACCC site A and CACCC site B probes due to the high background. A major slow mobility complex was observed with both CACCC site A and CACCC site B probes using

Figure 6.5   **Gel shift analysis of the GATA sites in the human ALAS2 intron 8**

Radiolabelled double-stranded oligonucleotides containing the GATA site A (GATA-A probe) and GATA site B (GATA-B probe) and a consensus GATA-1 binding site (GATA-cons) were incubated with nuclear extracts from K562 (lanes 1 and 6), MEL (lanes 2 and 7) and COS-1 (lanes 3 and 8) cells and COS-1 cells expressing recombinant GATA-1 (lanes 4 and 5, 9-12). For supershift assays, the GATA-1 monoclonal antibody, N-6, was added to nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 5, 10 and 12) prior to the addition of probe. The retarded complex corresponding to GATA-1 in the absence of antibody and the supershifted complex are indicated by *arrows*.
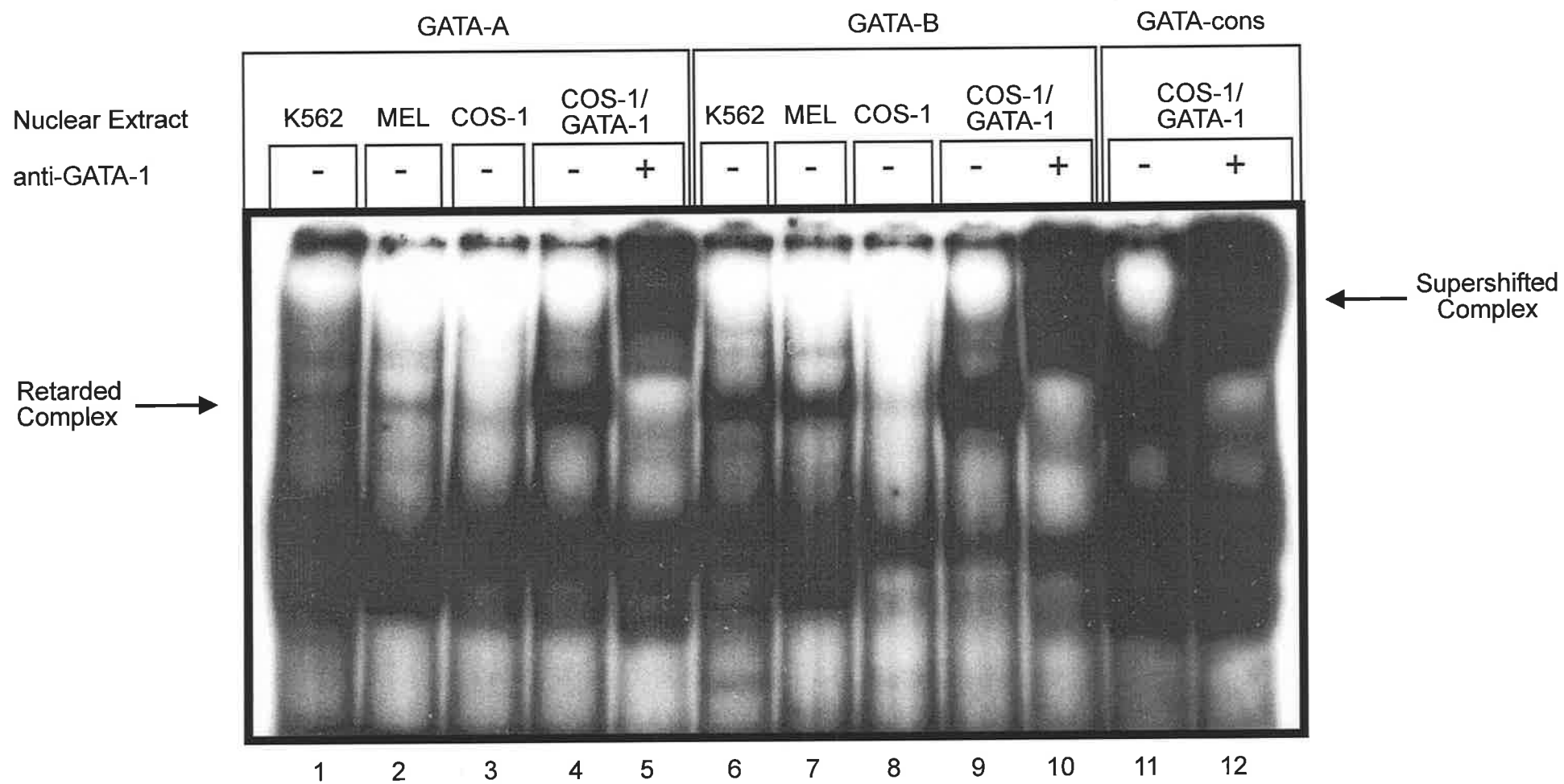
nuclear extracts from K562 (Fig. 6.6A, lanes 1 and 9) and MEL (lanes 4 and 12) cells. This complex contained Sp1 and probably Sp1-related proteins since it was completely supershifted with the Sp1 antibody (lanes 3, 6 and 11). In contrast, this slow migrating complex was not observed with the CACCC site A probe (lane 7) and CV-1 nuclear extracts expressing recombinant EKLF but was detected with the CACCC site B probe, although the intensity was markedly reduced (and difficult to see in Fig. 6.6A, lane 14).

These findings and gel shift experiments described in Section 5.2.2 (see Fig. 5.8) suggested that the level of Sp1 was reduced in nuclear extracts isolated from CV-1 cells compared with erythroid cells. To investigate this further, gel shift assays were performed with the Sp1-cons probe and K562 and CV-1 nuclear extracts. As seen in Fig. 6.6B, the level of the slow migrating complex (confirmed immunologically as Sp1), detected with the Sp1-cons probe was more abundant in K562 cells (lane 6) compared with both CV-1 cells (lane 8) and CV-1 cells expressing recombinant EKLF (lane 10).

In addition to the binding of Sp1 to the CACCC site A and CACCC site B probes, major more rapidly migrating complexes were also detected with both the CACCC site A and CACCC site B probes and MEL cell nuclear extracts (Fig. 6.6A, lanes 4 and 12). One of these complexes is thought to correspond to the binding of BKLF, since it was partially inhibited with an antibody to BKLF (but unfortunately this is not visible in Fig. 6.6A, lanes 5 and 13). However, the identity of the protein complex which migrated faster than Sp1 is unknown since it was unaffected by antibodies to Sp1 or BKLF.

To investigate whether the CACCC site A and CACCC site B probes were capable of binding EKLF protein, the β-globin CACCC probe that binds EKLF was included as a control, and all of these probes incubated with nuclear extracts from CV-1 cells expressing recombinant EKLF. As seen in Fig. 6.6B, a complex of high mobility was observed only with nuclear extracts from cells expressing recombinant EKLF and the β-globin CACCC probe (lane 4) and was confirmed as EKLF using an EKLF antibody (lane 5). However, the corresponding complex was not detected with either the CACCC site A (lane 7) or CACCC site B (lane 14) probes.
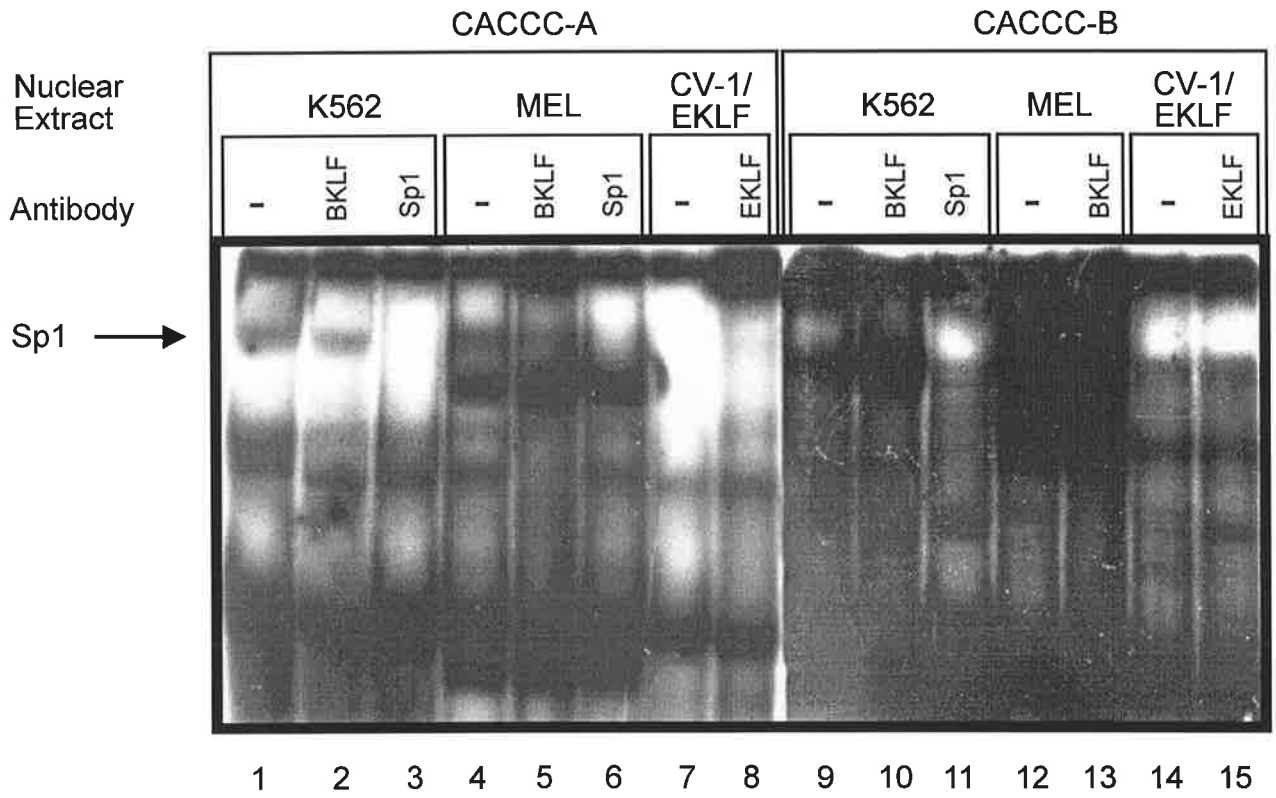
In conclusion, the data demonstrate that the CACCC site A and CACCC site B probes displayed a weak binding affinity for Sp1 in comparison to the β-globin CACCC site or a Sp1 consensus sequence. Although these sequences exhibit an extremely weak binding

Figure 6.6   Gel shift analysis of CACCC sites A and B in the human ALAS2 intron 8
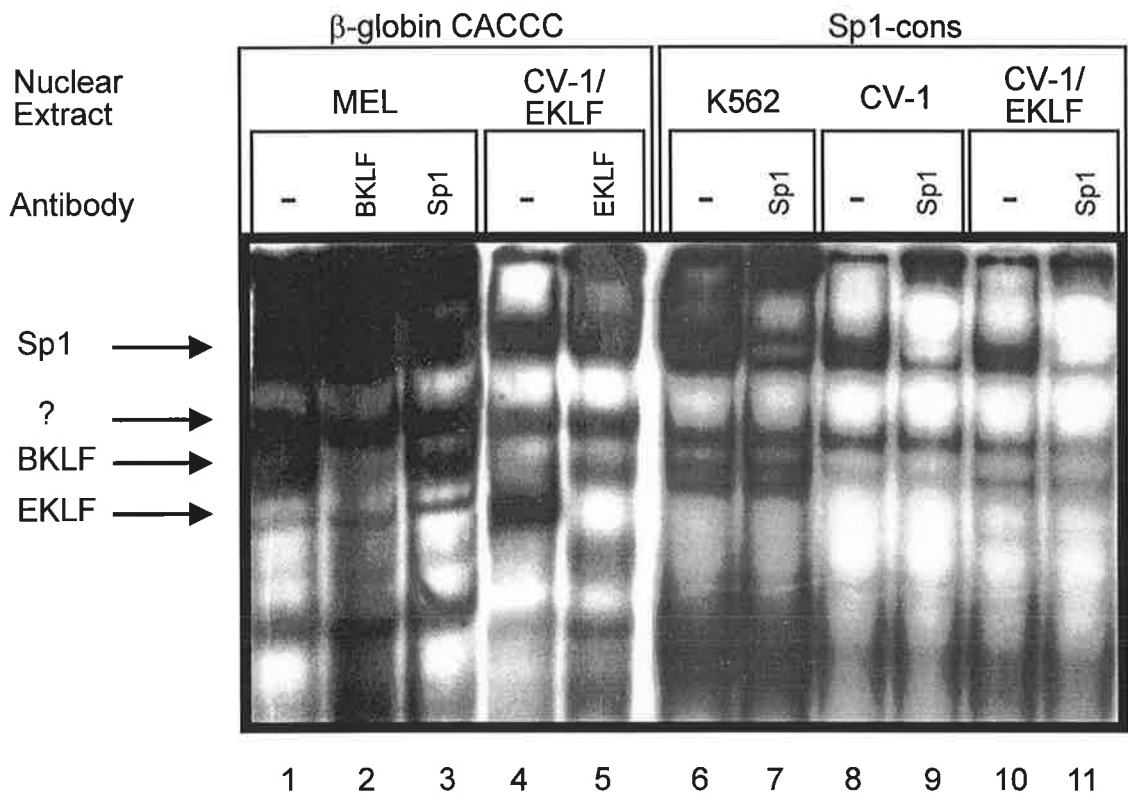
A   Radiolabelled double-stranded oligonucleotides containing the CACCC site A (CACCC-A probe) and CACCC site B (CACCC-B probe) were incubated with nuclear extracts from K562 (lanes 1-3, 9-11) and MEL (lanes 4-6, 12 and 13) cells, and CV-1 cells expressing recombinant EKLF (lanes 7 and 8, 14 and 15). For supershift assays, anti-BKLF (lanes 2, 5, 10 and 13), anti-Sp1 (lanes 3, 6 and 11) and anti-EKLF (lanes 8 and 15) antibodies were added prior to the addition of probe. The complex corresponding to Sp1 is indicated by the *arrow*.

B   The radiolabelled β-globin CACCC probe was incubated with nuclear extracts prepared from MEL cells (lanes 1-3) and CV-1 cells expressing recombinant EKLF (lanes 4 and 5). Antibodies specific to BKLF (lane 2), Sp1 (lane 3) and EKLF (lane 5) were added prior to the addition of the radiolabelled β-globin CACCC probe. The radiolabelled Sp1-cons probe was incubated with nuclear extracts from K562 (lanes 6 and 7) and CV-1 (lanes 8 and 9) cells and CV-1 cells expressing recombinant EKLF (lanes 10 and 11). For supershift assays, an anti-Sp1 antibody (lanes 7, 9 and 11) was added prior to the addition of probe. The retarded complexes corresponding to Sp1, BKLF, EKLF and an unknown protein (?) are *arrowed*.

A

	CACCC-A								CACCC-B						
Nuclear Extract	K562			MEL			CV-1/ EKLF		K562			MEL		CV-1/ EKLF	
Antibody	-	BKLF	Sp1	-	BKLF	Sp1	-	EKLF	-	BKLF	Sp1	-	BKLF	-	EKLF

Sp1 →

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

B

	β-globin CACCC					Sp1-cons					
Nuclear Extract	MEL			CV-1/ EKLF		K562		CV-1		CV-1/ EKLF	
Antibody	-	BKLF	Sp1	-	EKLF	-	Sp1	-	Sp1	-	Sp1

Sp1 →
? →
BKLF →
EKLF →

1 2 3 4 5 6 7 8 9 10 11

affinity for BKLF (difficult to see in Fig. 6.6A), the binding of EKLF to either intron 8 CACCC site was not detected.

A series of competition experiments were performed to investigate whether the human intron 8 CACCC site A or CACCC site B sequences were capable of inhibiting the binding of protein complexes to either the β-globin CACCC box or a Sp1 consensus sequence. Competition experiments performed with the β-globin CACCC probe and nuclear extracts from K562 cells (Fig. 6.7A) showed that the binding of nuclear proteins was effectively inhibited with a 10-fold molar excess of the β-globin CACCC oligonucleotide in self competition (lane 2) but a 200-fold molar excess of the CACCC site B oligonucleotide (lane 11) was required to inhibit binding. However, a 200-fold molar excess of the CACCC site A oligonucleotide (lane 7) very weakly competed for binding of protein complexes to the β-globin CACCC box, in comparison to a 200-fold molar excess of a non-specific competitor oligonucleotide (lane 12).

Similar competition experiments were performed using the Sp1-cons probe and nuclear extracts isolated from K562 cells. As seen in Fig. 6.7B, the binding of the major slowly migrating complex (lane 1) was moderately inhibited with a 10-fold molar excess of the Sp1-cons oligonucleotide in self competition (lane 2) but a 200-fold molar excess of either the CACCC site A oligonucleotide (lane 6) or the CACCC site B oligonucleotide (lane 11) was required for a similar level of inhibition. There was no effect with a non-specific competitor oligonucleotide at a 200-fold molar excess (lane 7).

Competition experiments were also performed with the CACCC site B probe and nuclear extracts from MEL cells. The binding of nuclear proteins to the CACCC site B probe was inhibited with a 25-fold molar excess of the CACCC site B oligonucleotide in self competition and in comparison, a 50-fold molar excess of the CACCC site A oligonucleotide was required for a similar level of inhibition. However, a 25-fold molar excess of the β-globin CACCC oligonucleotide completely abolished the retarded protein complexes and there was no effect with a 200-fold molar excess of a non-specific competitor oligonucleotide (data not shown).

In conclusion, gel shift analysis of the two CACCC sites identified in the human ALAS2 intron 8 sequence, demonstrated the binding of Sp1 or an Sp1-related protein and the possible binding of BKLF. However, the binding affinity of Sp1 to either site was

Figure 6.7   Oligonucleotide competition for CACCC binding proteins in a gel shift assay

A   The radiolabelled β-globin CACCC probe was incubated with nuclear extracts from K562 cells (lanes 1-12). The retarded complexes were competed with a 10 and 50 -fold molar excess of the β-globin CACCC oligonucleotide in self competition (lanes 2 and 3), a 25, 50, 100 and 200 -fold molar excess of the CACCC-A (lanes 4-7) and CACCC-B (lanes 8-11) oligonucleotides and a 200-fold molar excess of a non-specific (NS) competitor (lane 12). The retarded complexes corresponding to Sp1, BKLF and an unknown protein (?) are *arrowed*. An additional retarded complex (*asterixed*) was observed in these competition experiments and co-migrated with BKLF but its identity is unknown.

B   The radiolabelled Sp1-cons probe was incubated with nuclear extracts from K562 cells (lanes 1-11). The retarded complexes were competed with a 10-fold molar excess of the Sp1-cons oligonucleotide in self competition (lane 2), a 25, 50, 100 and 200 -fold molar excess of the CACCC-A (lanes 3-6) and the CACCC-B (lanes 8-11) oligonucleotides and a 200-fold molar excess of a non-specific (NS) competitor (lane 7).

A
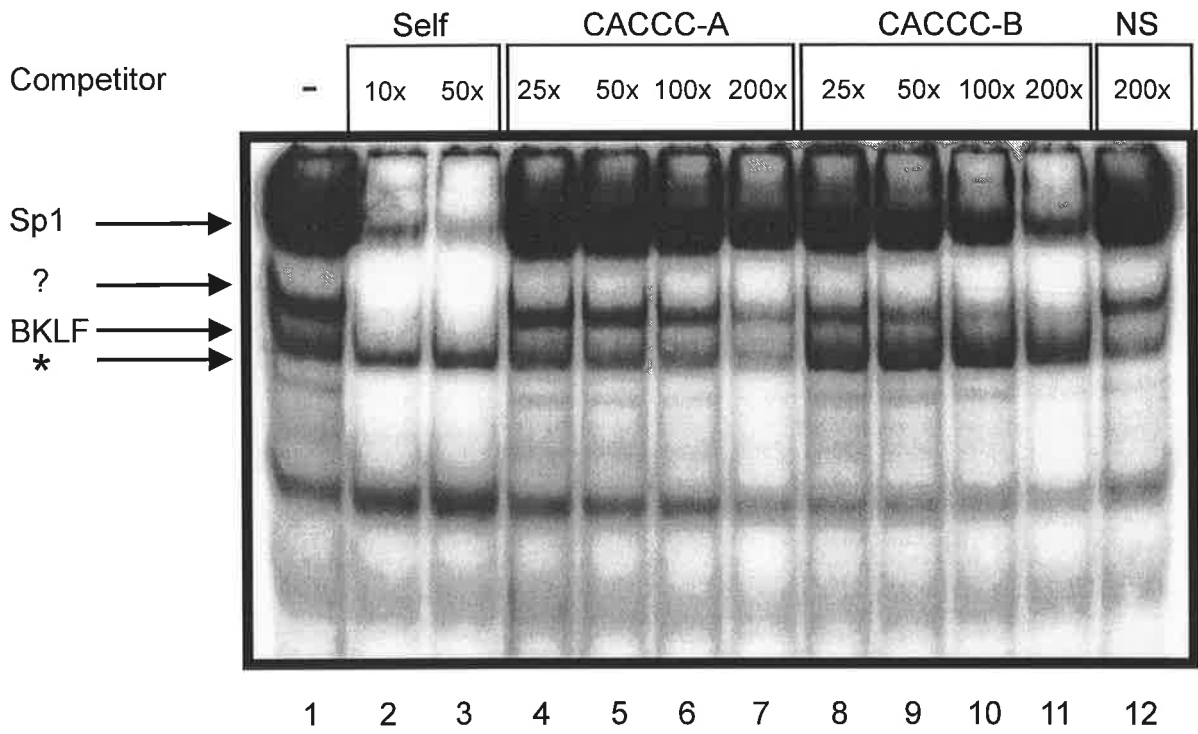
B
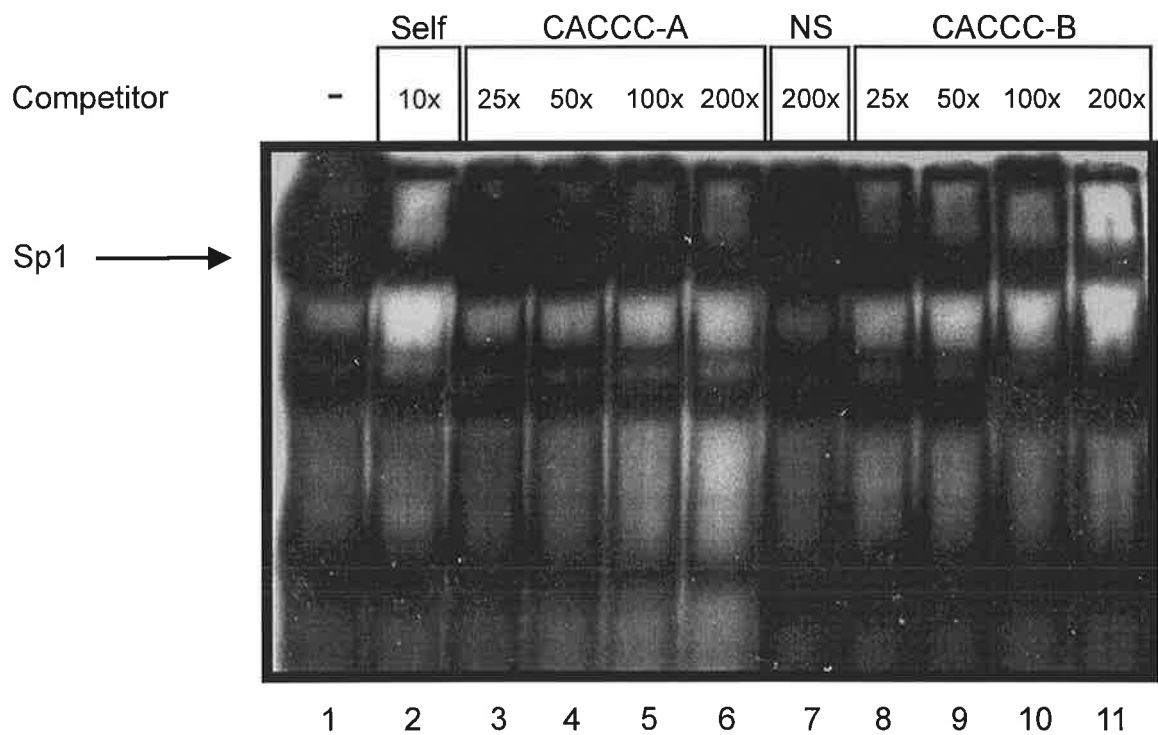
considerably reduced in comparison to that observed with the murine β-globin CACCC box or the human ALAS2 -54 CACCC sequence (as shown in Fig. 5.8)

6.2.6 Attempts to Transactivate the Erythroid-Specific Intron 8 Enhancer in Non-Erythroid Cells

Mutational analysis demonstrated that only one of the two GATA-1 binding sites and both CACCC box binding sites were important for the erythroid-specific enhancer activity of intron 8. However, gel shift analysis of the two CACCC boxes only detected binding of Sp1 or Sp1-related proteins with the possible binding of BKLF to both sites but not EKLF. As described in Section 4.2.3, co-transfection studies were performed in COS-1 cells with exogenously expressed murine GATA-1 to investigate whether the inactivity of the intron 8 enhancer in non-erythroid COS-1 cells was solely attributable to the absence of endogenous GATA-1 protein. Although the human ALAS2 promoter was transactivated by GATA-1 in COS-1 cells, no additional expression was observed with the inclusion of the human intron 8 sequence (pALASp-Int8(460)-LUC) (as described in Section 4.2.3). A similar result was obtained from GATA-1 transactivation assays performed in the parental CV-1 cell line from which COS-1 cells are derived (data not shown). These findings suggested the possible involvement of other transcription factors, either binding specifically to the two functional CACCC sites within intron 8 or to additional but as yet unidentified sites located elsewhere within the 235 bp intron 8. It is possible that these additional factors are required to achieve full enhancer activity in non-erythroid cells. The inability to activate intron 8 in COS-1 or CV-1 cells may be attributable to the relatively low levels of endogenous Sp1 protein in comparison to erythroid K562 cells (see Fig. 6.6B, compare lane 6 with lane 8) and this may be insufficient to cooperate with exogenously expressed GATA-1. Alternatively, an erythroid-specific CACCC box binding protein or an Sp1-related protein may be required that is not expressed in the kidney cell line examined and these possibilities are discussed further in Section 6.3.

To investigate whether the erythroid-specific CACCC box binding protein, EKLF, may be involved in the creation of the erythroid-specific intron 8 enhancer, EKLF transactivation experiments were performed. Preliminary co-transfection assays were performed with the plasmid pALASp-Int8(235)-LUC and exogenously expressed murine

171

EKLF and GATA-1 in COS-1 and CV-1 cells. However, exogenously expressed EKLF failed to transactivate the positive control plasmid, p(CAC)$_4$tk-LUC. Therefore transactivation assays were performed in K562 cells as described in Section 5.2.2, to determine whether EKLF could transactivate sequences located within intron 8.

As seen in Fig. 6.8, exogenously expressed EKLF transactivated the ALAS2 promoter (pALASp-LUC) 4.4-fold in K562 cells, while the construct pALASp-Int8(235)-LUC containing an additional 235 bp of intron 8 sequence was also transactivated to the same extent, demonstrating the inability of EKLF to transactivate the intron 8 enhancer. This finding is in agreement with gel shift studies performed on the two functional CACCC sites located within intron 8 which failed to detect the binding of recombinant EKLF protein to either site.

6.2.7 Intron 8 Enhancer Activity is not Modulated by a Non-Canonical TATA Box

As described in Section 5.2.1, the ALAS2 promoter contains a non-canonical TATA box that binds either GATA-1 or TBP *in vitro*. As described earlier, there is evidence that GATA-1 bound to the 3' enhancer of the chicken β-globin gene interacts with GATA-1 bound to a non-canonical TATA box in the promoter, imparting an erythroid-specific enhancer activity to transcription initiation (Fong & Emerson, 1992). A similar interaction between ALAS2 intron 8 and the non-canonical TATA box in the ALAS2 promoter may also occur. To investigate this, the plasmid construct pALASt-Int8(460)-LUC, was synthesised in which 460 bp human ALAS2 intron 8 fragment was ligated upstream of the ALAS2 promoter containing a consensus TATA box as described in Section 2.4.2. This plasmid and pALASp-Int8(460)-LUC were transiently expressed in K562, MEL and COS-1 cells.
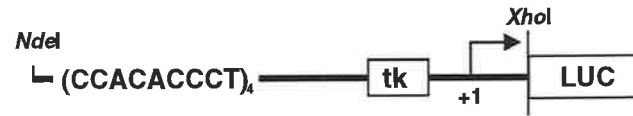
As previously described in Section 4.2.2, intron 8 transactivated the native ALAS2 promoter (pALASp-Int8(460)-LUC) 12.5 and 4.2 -fold in K562 and MEL cells, respectively. In comparison, intron 8 transactivated the promoter containing a consensus TATA box (pALASt-Int8(460)-LUC) 15 and 3 -fold in K562 and MEL cells, respectively, but not in COS-1 cells (Fig. 6.9). These findings indicate that a functional GATA-1 site at the TATA box location is not required for the enhancer activity of intron 8 *in vitro*. This

Figure 6.8   EKLF does not transactivate the human ALAS2 intron 8

The constructs p(CAC)$_4$tk-LUC, ptk-LUC, pALASp-LUC, pALASp-Int8(460)-LUC and pGL2-Basic were co-transfected with the EKLF cDNA expression clone, pSG5/EKLF in K562 cells and luciferase activities determined. The orientation of the intron 8 fragment is indicated by the *arrow*. The data are averages obtained from constructs tested in quadruplicate in three experiments and are represented as the mean ± standard deviation. The plasmids ptk-LUC and pGL2-Basic were included as controls for transactivation by EKLF, assigned a value of 1.0 and transactivation of the plasmid constructs by EKLF corrected for background.
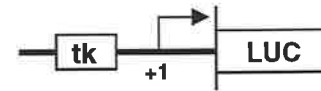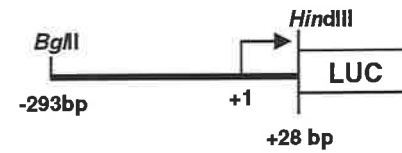
Fold Induction with EKLF in K562 cells

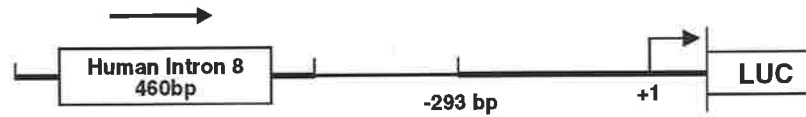p(CAC)₄tk-LUC — *Nde*I ⌐ (CCACACCCT)₄ — tk — +1 — *Xho*I — LUC — 10.6±0.5

ptk-LUC — tk — +1 — LUC — 1.0

pALASp-LUC — *Bgl*II — -293bp — +1 — *Hind*III — LUC — +28 bp — 4.4±0.6

pALASp-Int8(460)-LUC — Human Intron 8 460bp — -293 bp — +1 — LUC — 3.96±0.4
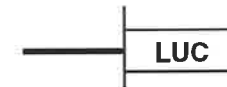
pGL2-Basic — LUC — 1.0

Figure 6.9 **Intron 8 enhancer activity is not modulated by a non-canonical TATA box**

The human ALAS2 intron 8 was ligated upstream of the 293 bp ALAS2 promoter containing a canonical TATA box at -27 GATA site (*boxed*) (pALASt-Int8(460)-LUC). This plasmid and pALASp-Int8(460)-LUC containing the native ALAS2 promoter together with pALASt-LUC and pALASp-LUC, were co-transfected with pRSV-βgal and transiently expressed in K562, MEL and COS-1 cells. The normalised luciferase activities of the intron 8 constructs are expressed relative to their appropriate promoter construct and assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in three experiments and are represented as the mean ± standard deviation.

Fold induction with Intron 8

	K562	MEL	COS-1
pALASp-LUC	1.0	1.0	1.0
pALASp-Int8(460)-LUC	12.5±0.3	4.2±0.4	0.95±0.2
pALASt-LUC	1.0	1.0	1.0
pALASt-Int8(460)-LUC	15±1.2	3±1.2	1.4±0.1

finding is also in agreement with the observation that intron 8 can increase expression of the heterologous thymidine kinase promoter as described in Section 4.2.2.

6.3 DISCUSSION

Studies described in this chapter focused on the characterisation of the elements responsible for the erythroid-specific enhancer activity located within intron 8 of the human ALAS2 gene. In contrast to the chicken β-globin gene which requires the interaction of a GATA-1 molecule bound to the 3' enhancer with another GATA-1 molecule bound at the non-canonical TATA box (Fong & Emerson, 1992), the human ALAS2 erythroid-specific enhancer efficiently transactivated the human ALAS2 promoter construct containing a TATA sequence rather than the native non-canonical TATA box (GATA motif) located at -27 . Comparison of the human, murine and canine intron 8 sequences identified extensive sequence conservation between the human, canine and 3' end of the murine intron 8 sequence. Deletion analysis of the human ALAS2 intron 8 sequence localised the erythroid-specific enhancer element to a 235 bp fragment. Sequence analysis of this fragment identified two putative binding sites for GATA-1 which were conserved between the human, canine and murine ALAS2 genes, and two putative CACCC boxes, both sufficiently conserved between the human, canine and murine ALAS2 genes as to predict a maintained function. Although the murine ALAS2 intron 8 contains additional 5' sequence which is not present in the human or canine intron 8 sequences, transient expression analysis demonstrated that the corresponding fragment from the murine intron 8 region fragment also transactivated the thymidine kinase promoter in erythroid cells only. These findings implicate a possible functional role of the sequences located within intron 8 of the ALAS2 gene *in vivo*. These conserved binding sites were investigated further but the involvement of other transcription factors binding to sequences located within the 235 bp intron 8 enhancer cannot be eliminated and *in vitro* DNA footprint analysis could be used to investigate this possibility. However, in addition to the GATA-1 and CACCC box binding sites identified by sequence analysis, putative binding sites for other transcription factors were not identified within the human ALAS2 intron 8 sequence.

The two conserved GATA sites identified within the ALAS2 intron 8 sequence were identical in sequence and both conformed to the consensus GATA-1 binding site (Merika &

Orkin, 1993; Ko & Engel, 1993). Mutational analysis demonstrated that only one of the two conserved GATA sites, GATA site B was functionally important in erythroid cells. Although *in vitro* binding studies showed that each GATA site bound GATA-1 protein in erythroid nuclear extracts, the DNA-binding affinity of GATA-1 for these sites was not investigated. The reason why only one of these GATA-1 binding sites is functional may be attributable to its location and the possible requirement to interact with other proteins. To investigate this further, plasmid constructs could be synthesised in which these two GATA sites are exchanged within intron 8, and then examined in transient expression studies to determine whether GATA site A can functionally replace GATA site B.

One interesting feature observed from mutational analysis of GATA site A was the consistent moderate increase in expression following destruction of this site. This finding indicated that the binding of GATA-1 to GATA site A may exhibit a repressor-like activity either directly through the action of GATA-1 or via a repressor protein binding to a sequence located adjacent to or overlapping the GATA site. Raich *et al.* (1995) demonstrated that GATA-1 is capable of exerting a negative effect on the expression of the human ε-globin gene, where a mutation of a GATA motif increased expression of a transgene. Two putative GATA-1 binding sites were also identified at the 3' end of the human ALAS2 intron 8 sequence but transient expression analysis of a fragment containing these sites suggested that these did not contribute to enhancer activity. Therefore although a number of putative GATA-1 binding sites were identified within the human ALAS2 intron 8 sequence, only one site was functional as determined by mutational analysis. Putative GATA-1 binding sites have been identified in the proximal promoters of erythroid-cell expressed genes including the human ε- (Gong *et al.*, 1991) and β- globin promoters (deBoer *et al.*, 1989), the human PBGD (Mignotte *et al.*, 1989b) and the murine EKLF promoter (Crossley *et al.*, 1994a), but not all of these sites are functional. Clusters of GATA sites have been identified which bind a single GATA-1 molecule (Tsai *et al.*, 1991; Trainor *et al.*, 1996) but we have no evidence suggesting that a single GATA-1 molecule is capable of binding to the GATA sites within the human ALAS2 intron 8 sequence.
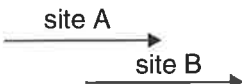
Two putative CACCC boxes were identified within this erythroid-specific enhancer and mutational analysis established that both sites were required for enhancer activity. The contribution of CACCC site A appeared moderately greater than that of CACCC site B as revealed by transient expression studies in erythroid cells. Although, mutational analysis

indicated a possible cooperative interaction between proteins binding to CACCC site A, CACCC site B and GATA site B, evidence for such an interaction cannot be concluded from these studies.

In vitro binding studies were performed to investigate the identity of the protein(s) binding to the two functional CACCC sites identified in the human ALAS2 enhancer. Currently, several proteins have been reported to bind *in vitro* to CACCC boxes and in particular members of the Krüppel family of transcription factors Sp1 (Kadonaga *et al.*, 1987), EKLF (Miller & Bieker, 1993) and BKLF (Crossley *et al.*, 1996). Gel shift analysis of the two functional CACCC sites in the intron 8 enhancer demonstrated that the CACCC site A and CACCC site B each bound Sp1 or an Sp1-related protein in erythroid cell nuclear extracts (possibly BKLF) but each site displayed a weak binding affinity for Sp1 compared to that detected with the murine β-globin CACCC box or the -54 CACCC sequence in the human ALAS2 promoter. In addition, the profile of nuclear proteins binding to these two CACCC sites did not mimic the previously characterised proteins capable of binding to the murine β-globin CACCC box including EKLF, Sp1 and BKLF (Crossley *et al.*, 1996) and the -54 CACCC sequence in the human ALAS2 promoter (as described in Section 5.2.2). A sequence comparison of the two CACCC sites from the human ALAS2 intron 8 sequence with the murine β-globin CACCC box and the single CACCC sites comprising the -54 CACCC sequence in the human ALAS2 promoter is shown in Table 6.1. Interestingly, CACCC site A (5'-CCCCACCCT-3') is identical in sequence to the functional distal CACCC site identified in the erythroid promoter of the human PBGD gene (Mignotte *et al.*, 1989b) which binds BKLF in MEL cell nuclear extracts (Crossley *et al.*, 1996) and purified EKLF protein (approximately 57% relative to the murine β-globin CACCC box) (Dr. A. Perkins, personal communication). However, one of the CACCC sites from the murine HS2 which displayed a weak binding affinity for BKLF in competition gel shift assays (Crossley *et al.*, 1996) is also identical to the CACCC site A. This finding suggests that the flanking sequence may also be critical in determining the binding affinity of a protein for a specific sequence. CACCC site B is identical in sequence to the proximal CACCC site in the murine GATA-1 promoter (Tsai *et al.*, 1991) which binds BKLF but whether this sequence also binds EKLF is not known. The CACCC site B is also identical to one of the CACCC sites that constitute the overlapping -54 CACCC sequence in the human ALAS2 promoter but specifically which site binds Sp1, EKLF and BKLF has not been

Table 6.1   Comparison of CACCC sequences that bind EKLF and BKLF

A summary of CACCC sequences from erythroid-specific genes that bind either purified EKLF or BKLF from erythroid nuclear extracts, determined from competition gel shift assays. The data is compiled from Crossley *et al.* (1996) and information obtained from Dr. Andrew Perkins (personal communication). The sequences of the CACCC site A and CACCC site B in the human ALAS2 intron 8 are compared to CACCC sequences identified in other erythroid-specific genes and the -54 CACCC sequence of the human ALAS2 promoter, comprised of two overlapping CACCC boxes.

	CACCC Sequence	Sequence binds BKLF	Sequence binds EKLF
Human ALAS2 intron 8 CACCC site A	5'-CCCCACCCT-3'	weakly	NO
Human PBGD erythroid promoter (distal site)	5'-CCCCACCCT-3'	strongly	weakly
Murine HS2 (site 2)	5'-CCCCACCCT-3'	weakly	ND
Human ALAS2 intron 8 CACCC site B	5'-CCCCACCCA-3'	weakly	NO
Murine GATA-1 (proximal site)	5'-CCCCACCCA-3'	strongly	ND
Human ALAS2 -54 CACCC sequence	site A → site B → 5'-CCCCACCCACCCA-3'	strongly	strongly
Human ALAS2 -48/-56 CACCC site A	5'-CCCCACCCA-3'	ND	ND
Human ALAS2 -51/-60 CACCC site B	5'-ACCCACCCA-3'	ND	ND
Murine β-globin promoter	5'-CCACACCCT-3'	strongly	strongly

determined. Therefore from examination of the CACCC sites from other erythroid-specific promoters capable of binding BKLF or EKLF it was extremely surprising that binding of these two proteins to either CACCC site in the human ALAS2 intron 8 was not detected. Although binding of Sp1 to both CACCC sites within intron 8 was observed, it was markedly reduced in comparison to the binding detected with either the murine β-globin CACCC box or the human ALAS2 promoter -54 CACCC sequence. To investigate these CACCC sites further, gel shift assays could be performed with purified Sp1, EKLF and BKLF proteins.

To investigate whether the inactivity of the ALAS2 enhancer in non-erythroid cells was due to the absence of endogenous GATA-1 protein, co-transfection assays were performed with exogenously expressed GATA-1 in COS-1 cells and the parental CV-1 cell line. However, exogenously expressed GATA-1 did not transactivate the ALAS2 promoter through intron 8. This finding is in direct contrast to the Wilm's Tumour gene erythroid-specific enhancer located 3' of the gene which can be transactivated in non-erythroid cells by exogenously expressed GATA-1 protein (Wu *et al.*, 1995). Therefore the inability to transactivate the human intron 8 enhancer in non-erythroid cells by GATA-1, suggested that either the levels of endogenous Sp1 protein were not appropriate in this cell line to cooperate with GATA-1 and activate the intron 8 enhancer, or another CACCC box binding protein was required. Alternatively, another erythroid factor binding elsewhere in the 235 bp intron 8 enhancer fragment may also be necessary to attain erythroid-specific enhancer activity.

Comparison of the relative concentrations of Sp1 protein in erythroid and non-erythroid CV-1 cells by gel shift analysis indicated that the Sp1 protein was more abundant in erythroid cell lines. Consequently, the inability to activate the enhancer in intron 8 may be attributable to the differences in the levels of exogenously expressed GATA-1 and endogenous Sp1 in COS-1 cells. Therefore future experiments would involve the over-expression of both Sp1 and GATA-1 in COS-1 cells to determine whether these two proteins can now transactivate the ALAS2 intron 8 enhancer in non-erythroid cells.

Another CACCC binding protein that may be involved is EKLF and EKLF may functionally interact with GATA-1 to activate the erythroid-specific enhancer. However, the inability to transactivate the intron 8 enhancer by exogenously expressed murine EKLF in K562 cells is in agreement with gel shift studies performed on the two functional CACCC

sites located within intron 8 which failed to detect the binding of recombinant EKLF protein to either site. This result is in contrast to that obtained with EKLF transactivation assays performed with the human ALAS2 promoter (as described in Section 5.2.2), where exogenously expressed EKLF transactivated the ALAS2 promoter through the functional CACCC sequence at -59/-48 (5'-GGGTGGGTGGGG-3').

Gregory *et al.* (1996) investigated the interaction of GATA-1 with proteins binding to the CACCC site, namely Sp1 or EKLF within erythroid promoters in transactivation studies performed in the *Drosophila* SL2 cell line which is devoid of endogenous Sp factors. Their findings suggested that in addition to the sequence of the CACCC element, the orientation of the CACCC element may also be an important factor in determining whether Sp1 or EKLF is the *in vivo* transcriptional activator which cooperates with GATA-1. The two functional CACCC sites and the two GATA sites within intron 8 are each oriented as direct repeats, with both GATA sites located on the non-coding strand. In contrast, the functional GATA sites (-124 and -100) in the ALAS2 promoter are arranged as an inverted palindrome. Similar transactivation studies could be performed in *Drosophila* SL2 cells, with both the human ALAS2 promoter and intron 8 in an attempt to investigate whether EKLF or Sp1 can functionally interact with GATA-1 to activate the enhancer.

As mentioned, gel supershift analysis identified the binding of Sp1 or Sp1-related proteins to both functional CACCC sites in the human ALAS2 intron 8 enhancer element. However, numerous transcription factors including Sp2, Sp3 and Sp4 (Hagen *et al.*, 1992; Kingsley & Winoto, 1992), BTEB (Imataka *et al.*, 1992) and BTEB2 (Sogawa *et al.*, 1993) recognise a similar DNA binding motif to Sp1 and a putative role for these factors in the transcriptional activation of the human ALAS2 gene through the erythroid-specific enhancer located within intron 8 may be eliminated by their tissue distribution.

The Sp1-related proteins, Sp2, Sp3 and Sp4 were cloned based on their homology with the Sp1 DNA-binding domain (Hagen *et al.*, 1992; Kingsley & Winoto, 1992). The DNA-binding specificities of Sp3 and Sp4 proteins are highly conserved and recognise GC/GT boxes with specificity and affinity closely similar to Sp1, while Sp2 has different DNA binding specificities (Kingsley & Winoto, 1992). Northern blot analysis of the mRNA tissue distribution of Sp2 and Sp3 in comparison to Sp1, demonstrated that these two proteins are also widely expressed in tissues and cell lines (Hagen *et al.*, 1992; Kingsley & Winoto, 1992). Interestingly, Sp2 mRNA is expressed in erythroid K562 cell lines whereas

Sp3 mRNA is detected in both erythroid K562 and MEL cell lines (Kingsley & Winoto, 1992). In contrast, Sp4 transcripts are abundantly expressed in the brain but barely detected in other organs (Hagen *et al.*, 1992).

The frequency of Sp binding sites in the regulatory regions of genes and their tissue distribution suggests that these transcription factors are involved in the regulation of a wide variety of viral and cellular genes in different tissues. Co-transfection studies performed in mammalian cells and *Drosophila* SL2 cells, demonstrated that Sp4 functions as a transcriptional activator while Sp3 represses both Sp1- and Sp4- mediated transcription (Hagen *et al.*, 1994; Hagen *et al.*, 1995, Majello *et al.*, 1994; Majello *et al.*, 1995, De luca *et al.*, 1996). However, Dennig *et al.* (1996) demonstrated that Sp3 was also capable of functioning as a transcriptional activator but was silenced by an inhibitory domain located between the second glutamine-rich region and the zinc fingers. Gene targeting by homologous recombination was used to investigate the role of Sp4 during murine development (Supp *et al.*, 1996) and suggested an important role for Sp4 in normal male reproductive behaviour.

There are two other GC-box binding proteins that recognise a similar DNA binding motif to Sp1, designated BTEB (Imataka *et al.*, 1992) and BTEB2 (Sogawa *et al.*, 1993). BTEB can function both as a transcriptional activator or repress transcriptional activation of a promoter by Sp1 (Imataka *et al.*, 1992). Northern blot analysis of BTEB mRNA detected transcripts in all tissues examined and was abundantly expressed in the kidney, lung, brain and testis (Imataka *et al.*, 1992). BTEB2 can also function as a transcriptional activator but transcripts were only detected in the testis and placenta. Interestingly, intracellular levels of Sp1 do vary between tissues and expression is relatively low in the testis and brain (Imataka *et al.*, 1992) suggesting that BTEB2 may functionally replace Sp1 in these tissues or that some testis or placenta-expressed genes specifically require BTEB2.

Therefore the identification of the CACCC binding protein involved in the transcriptional activation of the erythroid-specific enhancer within intron 8 of the human ALAS2 gene is extremely difficult due to the number of transcription factors capable of binding to CACCC or GC boxes with a similar DNA binding affinity to that of Sp1. Although, the Sp2 and Sp3 transcription factors are widely expressed in tissues and cell lines (Hagen *et al.*, 1992; Kingsley & Winoto, 1992), Sp2 is expressed in the erythroid K562 cell line and Sp3 is expressed in both K562 and MEL cells (Kingsley & Winoto,

1992). Therefore to investigate the possible role of these proteins in the transcriptional activation of the ALAS2 intron 8 enhancer, transactivation assays could be performed in the *Drosophila* SL2 cell line with cDNA expression clones encoding these proteins. To examine whether these proteins are capable of binding to the two CACCC sites in intron 8, gel shift assays could be performed with either nuclear extracts prepared from SL2 cells transfected with either a Sp2 or Sp3 cDNA expression clone and the binding of retarded protein complexes compared with nuclear extracts isolated from K562 and MEL cells. The generation of purified recombinant Sp2 and Sp3 proteins from GST-fusion proteins and the use of antibodies specific to Sp2 or Sp3 in gel supershift assays would assist in determining the identity of the nuclear proteins binding to the CACCC sites within intron 8. In contrast, the restricted expression of Sp4 transcripts in the brain and evidence generated from gene targeting studies, suggest that Sp4 is probably not the transcriptional activator of the erythroid-specific ALAS2 intron 8 enhancer. Similarly, examination of the tissue-specific expression of BTEB and BTEB2 suggested that these factors may also not be involved in the transcriptional activation of the ALAS2 enhancer but these cannot be entirely eliminated.

In conclusion, the identity of the CACCC binding protein involved in the transcriptional activation of the erythroid-specific enhancer located in intron 8 of the human ALAS2 remains to be resolved. Gel supershift analysis identified the binding of Sp1 or Sp1-related protein to both functional CACCC sites in the enhancer, but there also exists the possibility that the protein responsible for activating the erythroid-specific enhancer may not be detected with the current gel shift assay conditions. However, numerous highly related transcription factors are capable of binding to the Sp1 sequence and their role may be eliminated by their tissue distribution.

In addition to the identification of protein(s) binding to the CACCC boxes within intron 8 of the ALAS2 gene, another erythroid factor may be required for the activation of the erythroid-specific enhancer in non-erythroid cells. Tsang *et al.* (1997) has recently reported the isolation of a cofactor for GATA-1, designated Friend of GATA-1 (FOG) which is coexpressed with GATA-1 in both erythroid and megakaryocytic lineages. FOG can functionally cooperate with GATA-1 to synergistically activate transcription of the *p45 NF-E2* gene and also demonstrated to play a critical role in the terminal differentiation of both erythroid and megakaryocytic lineages (Tsang *et al.*, 1997). Therefore it would be of

great interest to investigate whether coexpression of both GATA-1 and FOG in COS-1 cells is capable of activating the intron 8 enhancer in these non-erythroid cells.

GATA-1 binding sites are often found in close association with CACCC boxes in the regulatory regions of many erythroid-specific genes (deBoer *et al.*, 1988, Mignotte *et al.*, 1989b; Frampton *et al.*, 1990; Zon *et al.*, 1991; Rahuel *et al.*, 1992; Max-Audit *et al.*, 1993). Interestingly, the distance between these two motifs are generally observed in an approximate multiple of 10 bp indicating that stereospecific alignment of these binding sites may be required. There is evidence for a synergistic interaction between the proteins binding to the GATA and CACCC binding sites, namely GATA-1 with either EKLF or Sp1 (Merika & Orkin, 1995; Gregory *et al.*, 1996) and mutational analysis of the two conserved CACCC binding sites and one GATA-1 binding site located within human ALAS2 intron 8, indicated that proteins binding to all of these sites was required to attain full erythroid-specific enhancer activity. Interestingly, the distance between the CACCC and GATA-1 binding sites identified within intron 8 are highly conserved between all three species, with the exception of a 5 bp insertion between the CACCC site A and CACCC site B in the murine intron 8 sequence. The distance between the functional CACCC site A and GATA site B, essential for the erythroid-specific enhancer activity of intron 8, consists of a spacing of approximately 110 bp in both the human and canine ALAS2 genes and approximately 114 bp in the murine ALAS2 gene. The distance between the CACCC site B and GATA site B is approximately 51 bp in all three ALAS2 intron 8 sequences. Therefore based on these distances, both CACCC sites are probably located on the same side of the helix. The two GATA sites are predicted to reside on opposite sides of the helix with GATA site B on the same side of the helix as both CACCC sites within human ALAS2 intron 8. This arrangement may explain why only GATA site B is functionally important for enhancer activity. In contrast, an additional 5 bp of sequence is located between the two CACCC sites in the murine intron 8 sequence and this indicates that these two sites may be located on opposite sides of the helix and the functional significance of this is currently being examined. In addition, the CACCC-like sequences 5'-CCTCACCC-3' and 5'-CCCCACTC-3' present in the murine and canine intron 8 sequences, respectively, are predicted to correspond to the functional CACCC site A in the human ALAS2 intron 8. However, whether these sites are indeed functional remains to be determined.

Although enhancer elements have been considered to function at a distance and increase the rate of transcription of a linked promoter (Treisman & Maniatis, 1985; Weber & Shaffner, 1985), recent studies (Walters *et al.*, 1995; Walters *et al.*, 1996; Sutherland *et al.*, 1997) as described in Chapter 4, have suggested that the primary role of enhancer elements is to establish and maintain an active domain permissive for transcription rather than functioning to increase the rate of gene transcription. Sequences located within human ALAS2 intron 8 were shown to contribute to promoter expression, either by contributing to promoter activity or to chromatin structure. In the present study, it cannot be distinguished whether the reduced expression observed when the GATA or CACCC box binding sites within intron 8 were mutated, represent a decrease in the proportion of transfected cells expressing the reporter gene or from a decrease in promoter activity. However, in this study the primary aim was to identify regions of the human ALAS2 gene responsible for tissue-specific expression, resulting in the identification of elements located within intron 8, demonstrated to be functionally important for expression. These elements are highly conserved between species alluding to a possible functional role *in vivo* and this is currently under investigation.

During erythroid differentiation, large quantities of heme are required for hemoglobin synthesis. The transcription of the ALAS2 gene is markedly upregulated together with the genes for the other heme pathway enzymes to accommodate the increased demand for heme. As described in Section 5.1, the promoter of the erythroid ALAS gene (ALAS2) contains functional GATA and CACCC binding sites which are a feature of the regulatory regions of many erythroid cell specific genes, but importantly it was shown that this promoter is active in both erythroid and non-erythroid cells in transient expression studies (as described in Section 4.2.1). Consequently, an examination of other regions of the ALAS2 gene was undertaken and resulted in the identification of an erythroid-specific enhancer element within intron 8 located approximately 16 kb downstream from the transcription initiation site. Studies described in this chapter, demonstrated that functional GATA and CACCC binding sites constitute this enhancer element and are responsible for the high level expression of the gene *in vitro*. These elements were also found to be highly conserved in the corresponding region of the murine and canine ALAS2 genes alluding to a possible functional role *in vivo*. In contrast to ALAS, there is only one structural gene

encoding the other enzymes of the heme pathway (May *et al.*, 1995; Taketani *et al.*, 1995) but erythroid-specific enhancer elements are yet to be identified in any of these genes.

CHAPTER SEVEN

FINAL DISCUSSION AND SUMMARY

CHAPTER SEVEN : FINAL DISCUSSION AND SUMMARY

ALAS, has been of particular interest to our research laboratory since it is the rate-controlling enzyme of the heme biosynthetic pathway, and is likely to represent a target of gene regulation. There are two separate ALAS genes located on different chromosomes, encoding for either housekeeping (ALAS1) or erythroid-specific (ALAS2) forms of the enzymes (Sutherland *et al.*, 1988; Cox *et al.*, 1990; Bishop *et al.*, 1990). Expression of ALAS2 is developmentally regulated and is markedly increased during erythropoiesis to accommodate the increased demand for heme during hemoglobin production (May *et al.*, 1995). Therefore unlike the housekeeping enzyme, ALAS1 which is expressed in the liver and possibly all non-erythroid tissues, distinct regulatory mechanisms are thought to be involved in regulating ALAS2 gene expression in erythroid cells. At the onset of erythroid differentiation, transcription of the ALAS2 gene (Elferink *et al.*, 1988) is markedly up-regulated together with an increase in the transcription of genes for the other heme pathway enzymes (Beaumont *et al.*, 1984; Romeo *et al.*, 1986; Raich *et al.*, 1989) and for globin (Karlsson & Nienhuis, 1985) most probably in a coordinated fashion. In addition to the ALAS2 gene being regulated at the level of gene transcription, studies have also demonstrated that expression of the ALAS2 gene is regulated at the post-transcriptional level by an iron-responsive element (Cox *et al.*, 1991; Bhasker *et al.*, 1993)

The work presented in this thesis was aimed at investigating the transcriptional regulation of the human ALAS2 gene during erythropoiesis. The primary aim was to identify regions of the ALAS2 gene subject to transcriptional regulation and investigate the transcription factors responsible for the regulation of this gene during erythropoiesis. Studies performed in transformed erythroid cell lines, which can be induced to differentiate *in vitro*, have previously demonstrated that the increase in ALAS, together with other enzymes of the heme biosynthetic pathway and globins occurs primarily at the level of transcription (Beaumont *et al.*, 1984; Karlsson & Nienhuis, 1985; Romeo *et al.*, 1986; Elferink *et al.*, 1988; Raich *et al.*, 1989). Therefore transient expression studies were performed in transformed erythroid cell lines and a non-erythroid cell line, with a series of human ALAS2/reporter gene constructs in an attempt to identify regions of the ALAS2 gene which confer erythroid-specific expression. The MEL cell line which can be chemically induced to differentiate with Me$_2$SO (Friend *et al.*, 1971) was chosen for transient

expression studies. The erythroid K562 cell line has been used successfully in transient expression studies to identify critical regulatory regions responsible for the expression of numerous erythroid-specific genes (Rahuel *et al.*, 1992; Fischer *et al.*, 1993; Max-Audit *et al.*, 1993) and was also included in this study.

The structural organisation of the human, murine and chicken ALAS2 genes are remarkably similar, consisting of 11 exons and 10 introns with an IRE located in the 5'-untranslated region of the mRNA (Schoenhaut & Curtis, 1989; Conboy *et al.*, 1992; Lim *et al.*, 1994). DNase I hypersensitivity studies performed on the murine ALAS2 gene resulted in the identification of a number of hypersensitivity sites located in the immediate promoter region, intron 1, intron 3 and intron 8 (Schoenhaut & Curtis, 1989). These hypersensitivity sites are indicative of nucleosome-free or nucleosome remodelled regions of DNA where transcription factors essential for transcriptional regulation have bound (Elgin, 1988; Gross & Garrard, 1988). Therefore based on the highly conserved structural organisation of the ALAS2 gene and the availability of the entire nucleotide sequence of the human ALAS2 gene together with extensive 5'-flanking (10.3 kb) and 3'-flanking (3 kb) sequence, the corresponding regions of the human ALAS2 gene were examined for putative *cis*-acting elements. Numerous binding sites for the erythroid-enriched transcription factor, GATA-1, CACCC and CCAAT box binding proteins were identified throughout the 5'- and 3'- flanking sequences, while a single putative binding site for the erythroid-enriched factor, NF-E2, was identified in both the 5'- and 3'-flanking sequences. Although putative transcription factor binding sites were identified within intronic sequences of the human ALAS2 gene, only regions of the human ALAS2 gene corresponding to DNase I hypersensitivity sites in the murine ALAS2 gene were investigated. Multiple GATA-1 and CACCC box binding sites were identified in introns 1, 3 and 8 and these are a common occurrence in the regulatory regions of many erythroid-specific genes (deBoer *et al.*, 1988; Mignotte *et al.*, 1989b; Zon *et al.*, 1991; Max-Audit *et al.*, 1993).

As described in Chapter 4, a series of 5'-flanking ALAS2 deletion/luciferase constructs ranging in length from -10.3 kb to -27 bp upstream of the transcription initiation site were synthesised and transiently transfected into the erythroid MEL F4-12B2 and K562 cell lines and the non-erythroid COS-1 cell line. Deletion analysis of the 5'-flanking region of the human ALAS2 gene demonstrated that the first 300 bp upstream of the transcription initiation site could function as a promoter in erythroid cells. Sequence analysis of this

300 bp region revealed potential binding sites for the erythroid-enriched transcription factors, GATA-1 and NF-E2, as well as CACCC and CCAAT box binding proteins, the Ets family of proteins and the thyroid hormone receptor. Interestingly, all these putative transcription factor binding sites were clustered within the first 140 bp of promoter sequence. Transient expression studies of 5'-flanking sequence extending upstream of the 300 bp promoter, also indicated the presence of possible negative regulatory regions but these were not investigated further during the course of this project. However, ALAS2 promoter constructs containing -300 bp to -10.3 kb of 5'-flanking sequence also expressed efficiently in non-erythroid COS-1 cells indicating that additional sequence or native chromatin structure may be required for the erythroid-specific expression of the human ALAS2 gene *in vivo*.

Therefore in an attempt to identify regions of the ALAS2 gene responsible for erythroid-specific expression, the regions of the human ALAS2 gene corresponding to DNase I hypersensitivity sites in the murine ALAS2 gene and the 3'-flanking sequence were also examined in transient expression studies. As described in Chapter 4, ALAS2 intron 1 moderately transactivated the ALAS2 and heterologous thymidine kinase promoters in erythroid cells only, indicating the presence of regulatory sequences located within intron 1 capable of conferring tissue-specific expression. In contrast, sequences within ALAS2 intron 3 reduced the level of ALAS2 promoter expression in both erythroid and non-erythroid cell lines suggesting the presence of negative regulatory elements located within this region. However, in contrast to intron 1, intron 3 did not appear to affect expression of a reporter construct driven by the thymidine kinase promoter indicating a possible interaction between the ALAS2 promoter and intron 3 but this was not investigated further. The most significant and interesting finding obtained from transient expression studies described in Chapter 4, was the identification of an erythroid-specific enhancer element located within intron 8 of the human ALAS2 gene, capable of transactivating both the human ALAS2 promoter and the heterologous thymidine kinase promoter.

Sequence analysis of the ALAS2 3'-flanking sequence revealed putative consensus binding sites for GATA-1, CACCC and CCAAT box binding proteins and an NF-E2-like sequence. Erythroid-specific enhancers have been identified in the 3'-flanking sequence of the human (Trudel & Costantini, 1987) and chicken (Fong & Emerson, 1992) β-globin genes, the murine cytosolic glutathione peroxidase gene (O'Prey *et al.*, 1993) and the

Wilm's Tumour (WT1) gene (Wu *et al.*, 1995). However, the addition of the 3'-flanking sequence upstream of the ALAS2 promoter did not affect expression in transient transfection studies.

The final two chapters of this thesis examined in detail the transcription factors responsible for the strong transient expression of the first 300 bp of the human ALAS2 promoter and the high level erythroid-specific expression of an enhancer element identified within intron 8 of the ALAS2 gene. Mutational analysis and gel shift assays were performed to determine the transcription factors responsible for expression of the ALAS2 promoter and the erythroid-specific enhancer element located within intron 8.

As described in Chapter 5, sequence analysis identified putative binding sites for both erythroid-enriched and ubiquitous transcription factors, located within the first 140 bp of ALAS2 promoter sequence. The transcription factors that bound to these putative binding sites were examined in gel shift assays while the functional contribution of these sites in ALAS2 expression were investigated by site-directed mutagenesis and transient expression analysis of ALAS2 promoter/luciferase reporter gene constructs.

Two potential GATA-1 binding sites with an inverted palindromic arrangement, located at -124 and -100, respectively, both bound GATA-1 in *in vitro* DNA binding studies. Mutational analysis demonstrated that both sites were functional and transactivation assays performed with exogenously expressed GATA-1 demonstrated that both sites respond to GATA-1. A GATA-1 binding site was also identified at the -27 TATA box location which can also bind TBP in gel shift assays. Conversion of the GATA-1 binding motif to a canonical TATA box consistently reduced transient expression in erythroid cells only to 70-80% of wild-type, suggesting a role for GATA-1 in transcriptional initiation. However, conversion of the native -27 GATA site to a sequence which bound only GATA-1 and not TBP, significantly reduced expression indicating that the sequence at the TATA box location, must be capable of binding both GATA-1 and TBP in attaining maximal promoter expression.

Two overlapping CACCC boxes on the non-coding strand were identified at -59/-48 in the ALAS2 promoter. Mutational analysis demonstrated the functional role of this sequence in driving ALAS2 promoter expression in both erythroid and non-erythroid cells. Gel shift analysis demonstrated the binding of at least three CACCC binding proteins, including Sp1, EKLF and BKLF. Interestingly, the ALAS2 promoter was transactivated by

exogenously expressed EKLF in transactivation assays, suggesting the possible involvement of EKLF in the transcriptional activation of the ALAS2 promoter. Studies are currently in progress to examine the expression of ALAS2 mRNA in mice containing a targeted disruption of the EKLF gene (Perkins *et al.*, 1995).

A putative NF-E2 binding site (-49/-39) identified in the ALAS2 promoter was determined to be non-functional in erythroid cells by mutational analysis and binding of NF-E2 protein to this sequence was not detected by gel shift analysis. In contrast, to the general distribution of GATA and CACCC box motifs in the promoters and enhancers of erythroid cell expressed genes, functional binding sites for NF-E2 are not a common occurrence (Mignotte *et al.*, 1989b; Talbot *et al.*, 1990; Strauss *et al.*, 1992) suggesting that NF-E2 may be involved in the transcriptional regulation of only a subset of erythroid-specific genes.

A putative CCAAT box located at -90/-84 in the ALAS2 promoter was not investigated in the present study by gel shift assays or mutational analysis. However, based on the conserved sequence homology with the functional CCAAT box present in the human (deBoer *et al.*, 1988) and murine (Cowie & Myers, 1988) β-globin promoters, it is postulated to be functional in the human ALAS2 promoter. In addition, mutagenesis of both the -124 and -100 GATA sites and the -54 CACCC sequence in the ALAS2 promoter, reduced expression to 38% in erythroid cells, and the putative CCAAT box (-90/-84) may be responsible for this remaining activity.

As described in Chapter 1, several lines of evidence have implicated a possible role for the thyroid hormone receptor in the transcriptional regulation of the ALAS2 gene. A putative thyroid hormone response element was identified in the ALAS2 promoter (-21/-6). Gel shift assays did not detect binding of nuclear proteins to this site and preliminary transient expression studies suggested that this site was non-functional. However, other regions of the ALAS2 gene may respond to thyroid hormone and could be investigated further in transient transfection studies. Although several lines of evidence have also implicated a role for members of the Ets family of transcription factors in erythropoiesis, an Ets-like sequence located between the two functional GATA-1 binding sites at -124 and -100 was also shown to be non-functional in transient expression studies.

In conclusion, the ALAS2 promoter contains functional binding sites for GATA-1 and CACCC box binding proteins which are critical regulatory sequences of many erythroid

cell expressed genes (deBoer *et al.*, 1988; Mignotte *et al.*, 1989b; Zon *et al.*, 1991; Max-Audit *et al.*, 1993).

As shown in Chapter 6, sequence comparison of intron 8 of the human, murine and canine ALAS2 genes revealed extensive conservation between the human, canine and 3' end of the murine intron 8 sequence. An interesting feature was the complete conservation in sequence and location of two putative GATA-1 binding sites and one CACCC box binding site and a second CACCC-like sequence present in all three species. Since the erythroid-enriched factor, GATA-1, and CACCC box binding proteins have been demonstrated to functionally interact (Merika & Orkin, 1995; Gregory *et al.*, 1996), the clustering of these binding sites within intron 8 implicates a possible role for these sites in ALAS2 gene transcription. Transient expression studies performed with either the human or murine intron 8 sequences, demonstrated that both sequences conferred erythroid-specific enhancer activity to a similar degree. Deletion analysis of the human ALAS2 intron 8 resulted in the subsequent localisation of the erythroid-specific enhancer activity to a 235 bp fragment containing the conserved GATA sites and CACCC boxes. Although gel shift assays demonstrated that both GATA sites could bind GATA-1 *in vitro*, mutational analysis established that only one GATA-1 binding site in combination with both CACCC boxes was required to attain full enhancer activity. Gel shift assays of the two functional CACCC boxes only detected the binding of Sp1 or an Sp1-related protein but not EKLF. In agreement with this, exogenously expressed EKLF did not transactivate the ALAS2 promoter through sequences located within intron 8 suggesting that this factor is not involved in enhancer activity.

Studies are currently in progress in our laboratory to confirm the predicted functional role of the GATA-1 and CACCC box binding sites in intron 8 of the murine ALAS2 gene using similar transient expression studies as those performed with the human ALAS2 intron 8. Although, transient expression studies performed in the erythroid cell lines was an extremely successful approach in the identification of an erythroid-specific enhancer element located within intron 8 of the human ALAS2 gene, the most significant experiments would involve determining whether sequences located within intron 8 function *in vivo*. The approach undertaken will be to delete the murine ALAS2 intron 8 using gene targeting technology in embryonic stem cells. The mice generated (chimeric and hemizygous males) are predicted to be anemic and the extent of anemia will be dependent on the strength of the

enhancer. A number of parameters could be examined in these mice including the level of ALAS2 gene transcription in comparison to a wild-type mouse. In addition, DNase I hypersensitivity mapping studies as those described by Schoenhaut and Curtis (1989) could be performed to examine the formation of DNase I hypersensitivity sites in the intron 8 enhancer deficient mouse.

A characteristic feature of terminally differentiating erythroid cells is an increase in chromatin condensation and the silencing of genes not required for erythroid differentiation (Sherton & Kabat, 1976; Fraser & Curtis, 1987). Therefore enhancer elements of the globin gene clusters have been proposed to function to delay the silencing of globin genes (Walters et al., 1996; Sutherland et al., 1997). It is possible that other erythroid-specific genes may utilise a similar mechanism to maintain transcription of their genes in a "repressive" environment. The erythroid-specific enhancer identified within intron 8 of the human ALAS2 gene may represent one such example. Therefore it would be interesting to examine whether deletion of the ALAS2 intron 8 enhancer in erythroid progenitor cells alters the rate of ALAS2 gene transcription or reduces the number of cells expressing ALAS.

A recurrent theme observed throughout the course of this study was the critical role of GATA-1 and CACCC box binding proteins in the transcriptional regulation of the human ALAS2 gene. As described in Section 1.4, previous studies directed at investigating the transcription factors responsible for expression of other erythroid cell expressed genes including globin, enzymes of the heme biosynthetic pathway and the EpoR have also demonstrated the functional importance of GATA-1 and CACCC box binding proteins (deBoer et al., 1988; Mignotte et al., 1989b; Zon et al., 1991). There is evidence for transcriptional synergism between GATA-1 binding sites and nearby CACCC box binding sites based on the mutagenesis of these sites (Frampton et al., 1990; Zon et al., 1991; Tsai et al., 1991) and co-transfection studies performed in D. melanogaster SL2 cells which are devoid of endogenous Sp factors (Merika & Orkin, 1995; Gregory et al., 1996). However, mutational analysis of the GATA and CACCC sites in the ALAS2 promoter did not provide any evidence for a cooperative interaction between these sites. In contrast, mutational analysis of the two conserved CACCC box binding sites and one GATA-1 binding site identified in the ALAS2 intron 8 sequence indicated that the proteins binding to all of these sites was required to attain full erythroid-specific enhancer activity. Interestingly, unlike intron 1, intron 8 was not transactivated by exogenously expressed GATA-1 in

co-transfection assays indicating the requirement of additional proteins in the transcriptional activation of the ALAS2 intron 8 enhancer. Studies are currently in progress to investigate the identity of these factors.

Transient expression studies performed in the erythroid MEL and K562 cell lines was an extremely successful approach in providing an insight into the regions of the human ALAS2 gene subjected to transcriptional regulation in erythroid cells. However, this approach only allowed the identification of transcription factors involved in driving expression of the ALAS2 gene in undifferentiated erythroid cells. As previously described by Elferink *et al.* (1988) and as shown in Chapter 3, the levels of endogenous ALAS2 mRNA are markedly increased in MEL cells chemically induced to differentiate following exposure to Me_2SO. A similar result has also been observed in the J2E cell line induced to differentiate with the natural stimulator of erythropoiesis, Epo (Busfield *et al.*, 1993). Therefore it would be of great interest to determine whether the regions of the ALAS2 gene demonstrated to increase expression of the ALAS2 promoter in transient transfection studies, namely intron 1 and intron 8 are also activated during erythroid differentiation. These studies would require the generation of erythroid cell lines stably transfected with human ALAS2 gene constructs to permit expression of a reporter gene construct to be examined throughout erythroid differentiation.

A preliminary investigation into the regions of the ALAS2 gene which are activated during the Me_2SO-induced differentiation of MEL cells, involved the generation of a MEL cell line permanently expressing 1.9 kb of ALAS2 5'-flanking sequence fused to the CAT reporter gene. Interestingly, induction of this permanent cell line to differentiate revealed that the CAT reporter gene was transcriptionally activated at a similar time as the endogenous ALAS2 gene during Me_2SO-induced differentiation. This result indicated that all of the elements necessary for the temporal expression of the ALAS2 gene following Me_2SO-induction are located within this 1.9 kb of 5'-flanking sequence. Therefore the minimal region of the ALAS2 promoter that is activated during the Me_2SO-induction of MEL cells, could be defined with the generation of a series of deletion ALAS2 5'-flanking/CAT reporter constructs stably transfected into MEL cells. Although, expression of the CAT reporter gene was activated in response to Me_2SO at a similar time to endogenous ALAS2, expression was not maintained during erythroid differentiation, suggesting that other regions of the ALAS2 gene may be required for continued expression

during erythroid differentiation. Therefore analogous to the proposed role of the LCR of the β-globin gene cluster, involved in delaying the silencing of globin gene expression during erythroid differentiation (Walters *et al.*, 1995, Walters *et al.*, 1996; Sutherland *et al.*, 1997), intron 8 of the human ALAS2 gene may play a similar role. A similar strategy could also be employed to examine the expression of human ALAS2 intronic sequences including intron 1 and intron 8 in stably transfected MEL cells. Such an approach would permit an examination of the expression of the ALAS2 constructs during erythroid differentiation and also determine the importance of correct nucleosome assembly in ALAS2 gene expression.

As described in Chapter 4, the ALAS2 promoter contains a non-canonical TATA box (at -27) that exhibits some similarity to a GATA motif. However, conversion of the -27 GATA motif to a canonical TATA box only marginally reduced promoter activity. Interestingly, there is evidence to suggest that GATA-1 bound at the -30 position prevents the assembly of a repressive nucleosome at the chicken β-globin promoter (Barton *et al.*, 1993) and imparts erythroid cell specificity through the interaction with another GATA-1 molecule bound to the 3' enhancer (Fong & Emerson, 1992). Therefore expression of a reporter gene from MEL cell lines stably transfected with an ALAS2 promoter construct containing a canonical TATA box could be compared with the native ALAS2 promoter construct where correct nucleosome assembly would be expected. Transient transfection studies suggested that the human ALAS2 intron 8 enhancer transactivated the ALAS2 promoter containing a canonical TATA box to a similar degree as the native ALAS2 promoter. However, such a strategy could also be employed to investigate whether the model proposed for the chicken β-globin promoter-3' enhancer interaction (Fong & Emerson, 1992) may also be applicable to the ALAS2 gene.

In contrast to MEL cell lines, the J2E cell line which differentiates following exposure of these cells to Epo is a more appealing model system to investigate the regulatory regions of the ALAS2 gene which are transcriptionally activated during Epo-induced erythroid differentiation. Taxman and Wojchowski (1995) examined the transcriptional activation of the murine β-globin promoter in J2E cells induced to terminally differentiate with Epo. Minimal promoter constructs containing GATA-1 binding sites were highly induced by Epo in J2E cells suggesting the involvement of GATA-1. Treatment of J2E cells with okadaic acid and cAMP-elevating agents also induced erythroid differentiation indicating the possible involvement of serine/threonine phosphorylation

pathways and possibly including GATA-1 (Taxman & Wojchowski, 1995). However, transient expression studies performed with the murine β-globin promoter has revealed distinct differences between J2E cells and MEL cells (Taxman & Wojchowski, 1995). These include the identification of a positive regulatory domain (-346 bp to -107 bp) in the murine β-globin promoter in J2E cells but not present in MEL cells, and differences in the contribution of the CACCC and CCAAT elements to globin gene expression (Taxman & Wojchowski, 1995). These findings suggest that there may be distinct mechanisms operating in these two erythroid cell lines. To investigate this further, similar transient expression studies with ALAS2/reporter gene constructs should also be performed in the J2E cell line and then compared with the data obtained in MEL and K562 cell lines. Although preliminary attempts to detect activity of reporter gene constructs introduced in J2E cells by electroporation were unsuccessful, transfection of reporter constructs has been achieved in this cell line by "receptor-mediated transferrinfection" and this method could be employed in future expression studies. J2E cell lines that are stably transfected with ALAS2 gene/reporter constructs could be generated and these cell lines subsequently induced to differentiate with Epo. Expression of the ALAS2/CAT reporter constructs during differentiation could then be analysed and subsequently compared to those results obtained in stably transfected MEL cells chemically induced to differentiate with Me_2SO. However, time did not permit this aspect to be pursued during the course of the project.

REFERENCES

REFERENCES

Adams, M. L., Ostapiuk, I. & Grasso, J. A. (1989) *Biochim. Biophys. Acta* **1012,** 243-253

Adrisani, O. M., Hayes, T. E. & Dixon, J. E. (1987) *Nucleic Acids Res.* **15,** 5715-5728

Alland, L., Muhle, R., Hou Jr., H., Potes, J., Chin, L., Schreiber-Agus, N. & Depinho, R. A. (1997) *Nature* **387,** 49-55

Andersson, L. C., Jokinen, M. & Gahmberg, C. G. (1979) *Nature* **278,** 364-365

Andrews, N. C. & Faller, D. V. (1991) *Nucleic Acids Res.* **19,** 2499

Andrews, N. C., Erjument-Bromage, H., Davidson, M. B., Tempst, P. & Orkin, S. H. (1993a) *Nature* **362,** 722-728

Andrews, N. C., Kotkow, K. J., Ney, P. A., Erdjument-Bromage, H., Tempst, P., & Orkin, S. H. (1993b) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 11488-11492

Antoniou, M. & Grosveld, F. (1990) *Genes & Dev.* **4,** 1007-1013

Aplan, P. D., Nakahara, K., Orkin, S. H. & Kirsch, I. R. (1992) *EMBO J.* **11,** 4073-4081

Arceci, R. J., King, A. A. J., Simon, M. C., Orkin, S. H., & Wilson, D. B. (1993) *Mol. Cell. Biol.* **13,** 2235-2246

Armstrong, J. A. & Emerson, B. M. (1996) *Mol. Cell. Biol.* **16,** 5634-5644

Arosio, P., Adelman, T. G. & Drysdale, J. W. (1978) *J. Biol. Chem.* **253,** 4451-4458

Aurigemma, R. E., Blair, D. G. & Ruscetti, S. K. (1992) *J. Virol.* **66,** 3056-3061

Aziz, N. & Munro, H. N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84,** 8478-8482

Barber, D. L. & D'Andrea, A. D. (1994) *Mol. Cell. Biol.* **14,** 6506-6514

Barton, M. C., Madani, N. & Emerson, B. M. (1993) *Genes & Dev.* **7,** 1796-1809

Bawden, M. J., Borthwick, I. A., Healy, H. M., Morris, C. P., May, B. K. & Elliott, W. H. (1987) *Nucleic Acids Res.* **15,** 8563

Beaumont, C., Deybach, J. C., Grandchamp, B., da Silva, V., de Verneuil, H. & Nordmann, Y. (1984) *Exp. Cell Res.* **154,** 474-484

Beaumont, C., Dugast, I., Renaudie, F., Souroujon, M. & Grandchamp, B. (1989) *J. Biol. Chem.* **264,** 7498-7504

Beaumont, C., Jones, R., Seyhan, A. & Grandchamp, B. (1994a) *Adv. Exp. Med. Biol.* **356,** 211-218

Beaumont, C., Seyhan, A., Yachou, A.-K., Grandchamp, B. & Jones, R. (1994b) *J. Biol. Chem.* **269,** 20281-20288

Beaupain, D., Eleouet, J. F. & Romeo, P. H. (1990) *Nucleic Acids Res.* **18,** 6509-6515

Begley, C. G., Aplan, P. D., Denning, S. M., Haynes, B. F., Waldmann, T. A. & Kirsch, I. R. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86,** 10128-10132

Ben-David, Y., Giddens, E. B., Letwin, K. & Bernstein, A. (1991) *Genes & Dev.* **5,** 908-919

Benz, E. J. Jr., Murnane, M. J., Tonkonow, B. L., Berman, B. W., Mazur, E. M., Cavallesco, C., Jenko, T., Snyder, E. L., Forget, B. G. & Hoffman, R. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77,** 3509-3513

Bernet, A., Sabatier, S., Picketts, D. J., Ouazana, R., Morle, F., Higgs, D. R. & Godet, J. (1995) *Blood* **86,** 1202-1211

Berrodin, T. J., Marks, M. S., Ozato, K., Linney, E. & Lazar, M. A. (1992) *Mol. Endocrinol.* **6,** 1468-1478

Bessis, M., Lessin, L. S. & Beutler, E. (1983) in: *Hematology 3rd Ed*, W. J. Williams, E. Beutler, A. J. Erslev & M. A. Lichtman, eds. (McGraw-Hill) pp. 257-279

Beug, H., Leutz, A., Kahn, P. & Graf, T. (1984) *Cell* **39,** 579-588

Bhasker, C. R., Burgiel, G., Neupert, B., Emery-Goodman, A., Kuhn, L. C. & May, B. K. (1993) *J. Biol. Chem.* **268,** 12699-12705

Bieker, J. J. & Southwood, C. M. (1995) *Mol. Cell. Biol.* **15,** 852-860

Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513-1523

Bishop, D. F., Kitchen, H. & Wood, W. A. (1981) *Arch. Biochem. Biophys.* **206,** 380-391

Bishop, D. F., Henderson, A. S. & Astrin, K. H. (1990) *Genomics* **7,** 207-214

Bishop, T. R., Miller, M. W., Frelin, L. P., Dierks, P. & Boyer, S. H. (1991) *Am. J. Hum. Genet. Suppl.* **49,** 425

Bishop, T. R., Miller, M. W., Beall, J., Zon, L. I. & Dierks, P. (1996) *Nucleic Acids Res.* **24,** 2511-2518

Blanchard, K. L., Acquaviva, A. M., Galson, D. L. & Bunn, H. F. (1992) *Mol. Cell. Biol.* **12,** 5573-5585

Blenis, J. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 5889-5892

Blobel, G. A., Simon, M. C. & Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 626-633

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

Boehm, T., Buluwela, L., Williams, D., White, L. & Rabbitts, T. H. (1988) *EMBO J.* **7,** 2011-2017

Boehm, T., Foroni, L., Kennedy, M. & Rabbitts, T. H. (1990) *Oncogene* **5,** 1103-1105

Bondurant, M. C., Lind, R. N., Koury, M. J. & Ferguson, M. E. (1985) *Mol. Cell. Biol.* **5,** 675-683

Bornstein, P., McKay, J., Liska, D. J., Apone, S. & Devarayalu, S. (1988) *Mol. Cell. Biol.* **8,** 4851-4857

Borthwick, I. A., Srivastava, G., Day, A. R., Pirola, B. A., Snoswell, M. A., May, B. K. & Elliott, W. H. (1985) *Eur. J. Biochem.* **150,** 481-484

Borthwick, I. A., Srivastava, G., Pirola, B. A., May, B. K. & Elliott, W. H. (1986) *Methods Enzymol.* **123,** 395-401

Bottomley, S. S. (1993) in *Wintrobe's Clinical Hematology, 9th Ed.,* pp. 852-871 (G. R. Lee, T. C. Bithell, J. Foerster, J. W. Athens and J. N. Lukens, eds.), Lea & Febiger, Philadelphia

Bottomley, S. S., May, B. K., Cox, T. C., Cotter, P. D. & Bishop, D. F. (1995) *J. Bioenergetics & Biomembranes* **27**, 161-168

Bowerman, B., Eaton, B. A. & Priess, J. R. (1992) *Cell* **68**, 1061-1075

Boyd, D., Vecoli, C., Belcher, D. M., Jain, S. K. & Drysdale, J. W. (1985) *J. Biol. Chem.* **260**, 11755-11761

Braidotti, G. (1992) *PhD thesis,* University of Adelaide.

Braidotti, G., Borthwick, I. A. & May, B. K. (1993) *J. Biol. Chem.* **268**, 1109-1117

Brenner, D. A. & Frasier, F. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 849-853

Brenner, D. A., Didier, J. M., Frasier, F., Christensen, S. R., Evans, G. A. & Dailey, H. A. (1992) *Am. J. Hum. Genet.* **50**, 1203-1210

Brent, G. A., Harney, J. W., Chen, Y., Warne, R. L., Moore, D. D. & Larsen, P. R. (1989) *Mol. Endocrinol.* **3**, 1996-2004

Brickner, H. E., Zhu, X.-X. & Atweh, G. F. (1991) *J. Biol. Chem.* **266**, 15363-15368

Bridges, K. R. & Cudkowicz, A. (1984) *J. Biol. Chem.* **259**, 12970-12977

Briegel, K., Lim, K.-C., Plank, C., Beug, H., Engel, J. D., & Zenke, M. (1993) *Genes & Dev.* **7**, 1097-1109

Bruns, G. P. & London, I. M. (1965) *Biochem. Biophys. Res. Commun.* **18**, 236-242

Bull, B. S., Breton-Gorius, J. & Beutler, E. (1983) in: *Hematology 3rd Ed,* W. J. Williams, E. Beutler, A. J. Erslev & M. A. Lichtman, eds. (McGraw-Hill) pp. 297-316

Buluwela, L., Forster, A., Buehm, T. & Rabbitts, T. H. (1989) *Nucleic Acids Res.* **17**, 452

Busfield, S. J. & Klinken, S. P. (1992) *Blood* **80**, 412-419

Busfield, S. J., Riches, K. J., Sainsbury, A. J., Rossi, E., Garcia-Webb, P. & Klinken, S. P. (1993) *Growth Factors* **9**, 87-97

Busfield, S. J., Spadaccini, A., Riches, K. J., Tilbrook, P. A. & Klinken, S. P. (1995) *Eur. J. Biochem.* **230**, 475-480

Cairns, L. A., Crotta, S., Minuzzo, M., Moroni, E., Granucci, F., Nicolis, S., Schiro, R., Pozzi, L., Giglioni, B., Ricciardi-Castagnoli, R. & Ottolenghi, S. (1994) *EMBO J.* **13**, 4577-4586

Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I. & Santoro, C. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11598-11602

Casey, J. L., Hentze, M. W., Koeller, D. M., Caughman, S. W., Rouault, T. A., Klausner, R. D. & Harford, J. B. (1988) *Science* **240**, 924-928

Caterina, J. J., Ciavatta, D. J., Donze, D., Behringer, R. R. & Townes, T. M. (1994a) *Nucleic Acids Res.* **22**, 1006-1011

Caterina, J. J., Donze, D., Sun, C.-W., Ciavatta, D. J. & Townes, T. M. (1994b) *Nucleic Acids Res.* **22**, 2383-2391

Chambon, P. (1994) *Semin. Cell Biol.* **5**, 115-125

Chan, L.-N. L. & Gerhardt, E. M. (1992) *J. Biol. Chem.* **267**, 8254-8259

Chan, J. Y., Han, X.-L. & Kan, Y. W. (1993a) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11371-11375

Chan, L.-N. L. & Shao, J. Y. (1993) *Blood* **82 (Supl.),** 458a

Chan, R. Y. Y., Schulman, H. M. & Ponka, P. (1993b) *Biochem. J.* **292,** 343-349

Chan, R. Y., Seiser, C., Schulman, H. M., Kuhn, L. C. & Ponka, P. (1994).
Eur. J. Biochem. **220,** 683-692

Chandler, H. (1996) *Honours thesis,* University of Adelaide

Charnay, P. & Maniatis, T. (1983) *Science* **220,** 1281-1283

Chen, Q., Cheng, J.-T., Tsai, L.-H., Schneider, N., Buchanan, G., Carroll, A., Crist, W.,
Ozanne, B., Siciliano, M. J. & Baer, R. (1990) *EMBO J.* **9,** 415-424

Chen, J.-J. & London, I. M., (1995) *TIBS* **20,** 105-108

Chen, X. & Bieker, J. J. (1996) *EMBO J.* **15,** 5888-5896

Chodosh, L., Baldwin, A. S., Carthew, R. W. & Sharp, P. A. (1988) *Cell* **53,** 11-24

Choi, O.-R. & Engel, J. D. (1986) *Nature* **323,** 731-734

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

Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M.
& Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85,** 6-10

Chu, G., Hayakawa, H. & Berg, P. (1987) *Nucleic Acids Res.* **15,** 1311-1325

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

Conboy, J. G., Cox, T. C., Bottomley, S. S., Bawden, M. J. & May, B. K. (1992) *J. Biol. Chem.* **267,** 18753-18758

Conklin, K. F. & Groudine, M. (1986) *Mol. Cell. Biol.* **6,** 3999-4007

Cote, J., Quinn, J., Workman, J. L. & Peterson, C. L. (1994) *Science* **265,** 53-60

Coustry, F., Maity, S. N. & de Crombrugghe, B. (1995) *J. Biol. Chem.* **270,** 468-475

Cowie, A. & Myers, R. M. (1988) *Mol. Cell. Biol.* **8,** 3122-3128

Cox, T. C., Bawden, M. J., Abraham, N. G., Bottomley, S. S., May, B. K., Baker, E., Chen, L. Z. & Sutherland, G. R. (1990) *Am. J. Hum. Genet.* **46,** 107-111

Cox, T. C., Bawden, M. J., Martin, A. & May, B. K. (1991) *EMBO J.* **10,** 1891-1902

Cox, T. C. (1993) *PhD thesis,* University of Adelaide.

Cox, T. C., Bottomley, S. S., Wiley, J. S., Bawden, M. J., Matthews, C. S. & May, B. K. (1994) *N. Engl. J. Med.* **330,** 675-659

Crews, C. M. & Erikson, R. L. (1993) *Cell* **74,** 215-217

Crosby, J. S., Lee, K., London, I. M. & Chen, J.-J. (1994) *Mol. Cell. Biol.* **14,** 3906-3914

Crossley, M., Tsang, A. P., Bieker, J. J. & Orkin, S. H. (1994a) *J. Biol. Chem.* **269,** 15440-15444

Crossley, M. & Orkin, S. H. (1994b) *J. Biol. Chem.* **269,** 16589-16596

Crossley, M., Merika, M., & Orkin, S. H. (1995) *Mol. Cell. Biol.* **15**, 2448-2456

Crossley, M., Whitelaw, E., Perkins, A., Williams, G., Fujiwara, Y., & Orkin, S. H. (1996) *Mol. Cell. Biol.* **16**, 1695-1705

Cunningham, T. S. & Cooper, T. G. (1991) *Mol. Cell. Biol.* **11**, 6205-6215

Curtin, P. T., Liu, D., Liu, W., Chang, J. C. & Kan, Y. W. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7082-7086

Dalyot, N., Fibach, E., Ronchi, A., Rachmilewitz, E. A., Ottolenghi, S. & Oppenheim, A. (1993) *Nucleic Acids Res.* **21**, 4031-4037

Damm, K., Thomson, C. C. & Evans, R. M. (1989) *Nature* **339**, 593-597

Darr, D. & Fridovich, I. (1994) *J. Invest. Dermatol.* **102**, 671-675

Dean A., Ley, T. J., Humphries, R. K., Fordis, M. & Schechter, A. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5515-5519

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. & Grosveld, F. (1988) *EMBO J.* **7**, 4203-4212

Debuire, B., Henry, C., Bernissa, M., Biserte, G., Claverie, J. M., Saule, S., Martin, P. & Stehelin, D. (1984) *Science* **224**, 1456-1459

Deisseroth, A. & Hendrick, D. (1978) *Cell* **15**, 55-63

Delfau-Larue, M.-H., Martasek, P. & Grandchamp, B. (1994) *Hum. Mol. Genet.* **3**, 1325-1330

De luca, P., Majello, B., & Lania, L. (1996) *J. Biol. Chem.* **271**, 8533-8536

Dennig, J., Beato, M. & Suske, G. (1996) *EMBO J.* **15**, 5659-5667

Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395

Dierks, P. (1990) in *Biosynthesis of Heme and Chlorophylls,* pp. 201-233
(Dailey, H. A., ed.) McGraw-Hill, New York

Disela, C., Glineur, C., Bugge, T., Sap, J., Stengl, G., Dodgson, J., Stunnenburg, H.,
Beug, H. & Zenke, M. (1991) *Genes & Dev.* **5**, 2033-2047

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985)
J. Embryol. Exp. Morph. **87**, 27-45

Dogra, S. C., Hahn, C. N. & May, B. K. (1993) *Arch. Biochem. Biophys.* **300**, 531-534

Dogra, S. C. & May, B. K. (1996) *Arch. Biochem. Biophys.* **327**, 271-278

Donda, A., Schulz, M., Burki, K., De Libero, G. & Uematsu, Y. (1996) *Eur. J. Immunol.*
26, 493-500

Donze, D., Townes, T. M. & Bieker, J. J. (1995) *J. Biol. Chem.* **270**, 1955-1959

Dorfman, D. M., Wilson, D. B., Bruns, G. A. P. & Orkin, S. H. (1992) *J. Biol. Chem.* **267**,
1279-1285

Drew, P. D. & Ades, I. Z. (1989) *Biochem. Biophys. Res. Commun.* **162**, 102-107

Driscoll, C., Dobkin, C. S. & Alter, B. P. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**,
7470-7474

Dusanter-Fourt, I., Casadevall, N., Lacombe, C., Muller, O., Billat, C., Fischer, S. & Mayeux, P. (1992) *J. Biol. Chem.* **267,** 10670-10675

Dynan, W. (1986) *Trends Genet.* **2,** 196-197

Eckardt, K.-U. & Bauer, C. (1989) *Eur. J. Clin. Invest.* **19,** 117-127

Edwards, C. Q., Griffen, L. M. & Kushner, J. P. (1991) *Hosp. Pract.* **26,** 30-36

Elferink, C. J., Sassa, S. & May, B. K. (1988) *J. Biol. Chem.* **263,** 13012-13016

Elgin, S. C. R. (1988) *J. Biol. Chem.* **263,** 19259-19262

Elliott, W. H., May, B. K., Bawden, M. J. & Hansen, A. J. (1989) *Biochem. Soc. Symp.* **55,** 13-27

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

Evans, M. J. & Scarpulla, R. C. (1990) *Genes & Dev.* **4,** 1023-1034

Evans, T. & Felsenfeld, G. (1989) *Cell* **58,** 877-885

Evans, T. & Felsenfeld, G. (1991) *Mol. Cell. Biol.* **11,** 843-853

Fagard, R. & London, I. M. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78,** 866-870

Feng, W. C., Southwood, C. M. & Bieker, J. J. (1994) *J. Biol. Chem.* **269,** 1493-1500

Ferreira, G. C., Neame, P. J. & Dailey, H. A. (1993) *Protein Sci.* **2,** 1959-1965

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

Fischer, K. D., Haese, A. & Nowock, J. (1993) *J. Biol. Chem.* **268**, 23915-23923

Fong, T. C. & Emerson, B. M. (1992) *Genes & Dev.* **6**, 521-532

Forrester, W. C., Thompson, C., Elder, J. T. & Groudine, M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1359-1363

Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. (1987) *Nucleic Acids Res.* **15**, 10159-10177

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

Fraizer, G. C., Wu, Y.-J., Hewitt, S. M., Maity, T., Ton, C. C., Huff, V. & Saunders, G. F. (1994) *J. Biol. Chem.* **269**, 8892-8900

Frampton, J., Walker, M., Plumb, M. & Harrison, P. R. (1990) *Mol. Cell. Biol.* **10**, 3838-3842

Fraser, P. J. & Curtis, P. J. (1987) *Genes & Dev.* **1**, 855-861

Fraser, P., Hurst, J., Collis, P. & Grosveld, F. (1990) *Nucleic Acids Res.* **18**, 3503-3508

Fraser, P., Pruzina, S., Antoniou, M., & Grosveld, F. (1993) *Genes & Dev.* **7**, 106-113

Friend, C., Scher, W., Holland, J. G. & Sato, T. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 378-382

Fu, Y.-H. & Marzluf, G. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5331-5335

Fujita, H., Yamamoto, M., Yamagami, T., Hayashi, N. & Sassa, S. (1991) *J. Biol. Chem.* **266**, 17494-17502

Fukuda, Y., Fujita, H., Taketani, S. & Sassa, S. (1993) *Br. J. Hematol.* **85**, 670-675

Fukuda, Y., Fujita, H., Garbaczewski, L. & Sassa, S. (1994) *Blood* **83**, 1662-1667

Galson, D. L., Hensold, J. O., Bishop, T. R., Schalling, M., D'Andrea, A. D., Jones, C., Auron, P. E. & Housman, D. E. (1993) *Mol. Cell. Biol.* **13**, 2929-2941

Galson, D. L., Tsuchiya, T., Tendler, D. S., Huang, L. E., Ren, Y., Ogura, T. & Bunn, H. F. (1995) *Mol. Cell. Biol.* **15**, 2135-2144

Ganguly, S. & Skoultchi, A. I. (1985) *J. Biol. Chem.* **260**, 12167-12173

Gibson, K. D., Laver, W. G. & Neuberger, A. (1958) *Biochem. J.* **70**, 71-81

Gilles-Gonzalez, M. A., Ditta, G. S. & Helinski, D. R. (1991) *Nature* **350**, 170-172

Glass, C. K., Holloway, J. M., Devary, O. V. & Rosenfeld, M. G. (1988) *Cell* **54**, 313-323

Glass, C. K., Lipkin, S. M., Devary, O. V. & Rosenfeld, M. G. (1989) *Cell* **59**, 697-708

Godin, I. E., Garcia-Porrero, J. A., Coutinho, A., Dieterlen-Lievre, F. & Marcos, M. A. (1993) *Nature* **364**, 67-70

Golay, J., Introna, M. & Graf, T. (1988) *Cell* **55**, 1147-1158

Goldberg, M. A., Glass, G. A., Cunningham, J. M. & Bunn, H. F. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7972-7976

Goldberg, M. A., Dunning, S. P. & Bunn, H. F. (1988) *Science* **242,** 1412-1415

Goldberg, M. A., Gaut, C. C. & Bunn, H. P. (1991) *Blood* **77,** 271-277

Goldwasser, E., Jacobson, L. O., Fried, W. & Plzak, L. F. (1958) *Blood* **13,** 55-60

Gong, Q. H., Stern, J. & Dean, A. (1991) *Mol. Cell. Biol.* **11,** 2558-2566

Gong, Q. & Dean, A. (1993) *Mol. Cell. Biol.* **13,** 911-917

Gonzalez, F. J. (1989) *Pharmacol. Rev.* **40,** 243-288

Goossen, B., Caughman, S. W., Harford, J. B., Klausner, R. D. & Hentze, M. W. (1990) *EMBO J.* **9,** 4127-4133

Goossen, B. & Hentze, M. W. (1992) *Mol. Cell. Biol.* **12,** 1959-1966

Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044-1051

Gourdon, G., Sharpe, J. A., Wells, D., Wood, W. G. & Higgs, D. R. (1994) *Nucleic Acids Res.* **22,** 4139-4147

Graber, S. E. & Krantz, S. B. (1989) *Hematol. Oncol. Clin. North Am.* **3,** 369-400

Grandchamp, B., Romeo, P-H., Dubart, A., Raich, N., Rosa, J., Nordmann, Y. & Goossens, M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5036-5040

Grandchamp, B., de Verneuil, H., Beaumont, C., Chretien, S., Walter, O. & Nordmann, Y. (1987) *Eur. J. Biochem.* **162,** 105-110

Granick, S. (1966) *J. Biol. Chem.* **241**, 1359-1375

Granick, J. L. & Sassa, S. (1978) *J. Biol. Chem.* **253**, 5402-5406

Grayzel, A. I., Horchner, P. & London, I. M. (1966) *Proc. Natl. Acad. Sci. U.S.A.* **55**, 650-655

Green, A. R., Salvaris, E. & Begley, C. G. (1991) *Oncogene* **6**, 475-479

Green, A. R., Rockman, S., DeLuca, E. & Begley, C. G. (1993) *Exp. Hematology* **21**, 525-531

Gregory, R. C., Taxman, D. J., Seshasayee, D., Kensinger, M. H., Bieker, J. J. & Wojchowski, D. M. (1996) *Blood* **87**, 1793-1801

Gross, D. S. & Garrard, W. T. (1988) *Ann. Rev. Biochem.* **57**, 159-197

Gross, M. & Rabinovitz, M. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1565-1568

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

Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G. & Papayannopoulou, T. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7551-7555

Guo, G. G., Gu, M. & Etlinger, J. D. (1994) *J. Biol. Chem.* **269**, 12399-12402

Hagen, G., Muller, S., Beato, M. & Suske, G. (1992) *Nucleic Acids Res.* **20**, 5519-5525

Hagen, G., Muller, S., Beato, M. & Suske, G. (1994) *EMBO J.* **13**, 3843-3851

Hagen, G., Dennig, J., Preiss, A., Beato, M. & Suske, G. (1995) *J. Biol. Chem.* **270**, 24989-24994

Hamilton, J. W., Bement, W. J., Sinclair, P. R., Sinclair, J. F., Alcedo, J. A. & Wetterhahn, K. E. (1991) *Arch. Biochem. Biophys.* **289,** 387-392

Hamilton, J. W., Bement, W. J., Sinclair, P. R., Sinclair, J. F., Alcedo, J. A. & Wetterhahn, K. E. (1992) *Arch. Biochem. Biophys.* **298,** 96-104

Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D. R. & Grosveld, F. (1989) *Genes & Dev.* **3,** 1572-1581

Hansen, A. J., May, B. K. & Elferink, L. A. (1989) *DNA* **3,** 179-191

Hardy, K. J., Peterlin, B. M., Atchison, R. E. & Stobo, J. D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82,** 8173-8177

Harford, J. B., Casey, J. L., Koeller, E. M. & Klausner, R. D. (1991) *Intracellular Trafficking of Proteins,* pp. 302-334 (Steer, C. J. and Hanover, J. eds.) Cambridge University Press, Cambridge, UK

Harrison, P. R., Rutherford, T., Conkie, D., Affara, N., Sommerville, J., Hissey, P. & Paul, J. (1978) *Cell* **14,** 61-70

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

Hatton, C. S., R., Wilkie, A. O. M., Drysdale, H. C., Wood, W. G., Vickers, M. A., Sharpe, J., Ayyub, H., Pretonus, I. M., Buckle, V. J. & Higgs, D. G. (1990) *Blood* **76,** 221-227

Hayashi, N., Kurashima, Y. & Kikuchi, G. (1972) *Arch. Biochem. Biophys.* **148,** 10-21

Healy, H. M. (1990) *PhD Thesis*, University of Adelaide

Hearing, P. & Schenk, T. (1986) *Cell* **45,** 229-236

Heinzel, T., Lavinsky, R. M., Mullen, T.-M., Soderstrom, M., Laherty, C. D., Torchia, J., Yang, W.-M., Brard, G., Ngo, S. D., Davie, J. R., Seto, E., Eisenman, R. N., Rose, D. W., Glass, C. K. & Rosenfeld, M. G. (1997) *Nature* **387,** 43-48

Hentze, M. W., Rouault, T. A., Caughman, S. W., Dancis, A., Harford, J. B. & Klausner, R. D. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84,** 6730-6734

Hentze, M. W. & Argos, P. (1991) *Nucleic Acids Res.* **19,** 1739-1740

Herberlein, C., Fischer, K. D., Stoffel, M., Nowock, J., Ford, A., Tessmer, U. & Stocking, C. (1992) *Mol. Cell. Biol.* **12,** 1815-1826

Herbomel, P., Bourachot, B. & Yaniv, M. (1984) *Cell* **39,** 653-662

Hesse, J., Nichol, J., Lieber, M. & Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83,** 4312-4316

Higashino, F., Yoshida, K., Fujinaga, Y., Kamio, K. & Fujinaga, K. (1993) *Nucleic Acids Res.* **21,** 547-553

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I. M. & Ayyub, H. (1990) *Genes & Dev.* **4,** 1588-1601

Hodge, M. R., Kim, G., Singh, K. & Cumsky, M. G. (1989) *Mol. Cell. Biol.* **9,** 1958-1964

Hodin, R. A., Lazar, M. A., Wintman, B. I., Darling, D. S., Koenig, R. J., Larsen, P. R., Moore, D. D. & Chin, W. W. (1989) *Science* **244,** 76-79

Hofer, E., Hofer-Warbinek, R. & Darnell Jr., J. E. (1982) *Cell* **29,** 887-893

Horlein, A. J., Naar, A. M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C. K. & Rosenfeld, M. G. (1995) *Nature* **377,** 397-404

Horton, M. A. (1983) *Exp. Cell Res.* **144,** 361-366

Hu, B., Wright, E., Campbell, L. & Blanchard, K. L. (1997) *Mol. Cell. Biol.* **17,** 851-856

Hug, B. A., Wesselschmidt, R. L., Fiering, S., Bender, M. A., Epner, E., Groudine, M., & Ley, T. J. (1996) *Mol. Cell. Biol.* **16,** 2906-2912

Iacopetta, B. J., Morgan, E. H. & Yeoh, G. C. T. (1982) *Biochim. Biophys. Acta* **687,** 204-210

Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. & Yamamoto, M. (1994) *Nature* **367,** 568-572

Ihle, J. N., Witthuhn, B. A., Quelle, F. W., Yamamoto, K., Thierfelder, W. E., Kreider, B. & Silvennoinen, O. (1994) *TIBS* **19,** 222-227

Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M. & Fijii-Kuriyama, Y. (1992) *EMBO. J.* **11,** 3663-3671

Ishida, N., Fujita, H., Fukuda, Y., Noguchi, T., Doss, M., Kappas, A. & Sassa, S. (1992) *J. Clin. Invest.* **89,** 1431-1437

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J. D. & Yamamoto, M. (1993) *Nature* **362,** 466-468

Izumo, S. & Mahdavi, V. (1988) *Nature* **334,** 539-542

Jaffe, E. K., Bagla, S. & Michini, P. A. (1991) *Biol. Trace Element Res.* **28**, 223-231

James-Pederson, M., Yost, S., Shewchuk, B., Zeigler, T., Miller, R. & Hardison, R. (1995) *J. Biol. Chem.* **270**, 3965-3973

Jelkmann, W. (1992) *Physiol. Rev.* **72**, 449-489

Johnson, P. F. & McKnight, S. (1989) *Annu. Rev. Biochem.* **58**, 799-839

Joulin, V., Bories, D., Eleouet, J.-F., Labastie, M. C., Chretien, S., Mattei, M.-G. & Romeo, P.-H. (1991) *EMBO J.* **10**, 1809-1816

Jover, R., Lindberg, R. L. & Meyer, U. A. (1996) *Mol. Pharmacol.* **50**, 474-481

Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R (1987) *Cell* **51**, 1079-1090

Kaiho, S-I. & Mizuno, K. (1985) *Anal. Biochem.* **149**, 117-120

Kallianpur, A. R., Jordan, J. E. & Brandt, S. J. (1994) *Blood* **83**, 1200-1208

Kappas, A., Sassa, S., Galbraith, R. A. & Nordmann, Y. (1989) in *Metabolic and Molecular Basis of Inherited Disease,* pp.1305-1365 (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D. eds.) McGraw-Hill, New York

Karlsson, S. & Nienhuis, A. W. (1985) *Ann. Rev. Biochem.* **54**, 1071-1108

Katsube, T. & Fukumaki, Y. (1995) *J. Biochem. (Toyko)* **117**, 68-76

Kaya, A. H., Plewinska, M., Wong, D. M., Desnick, R. J., & Wetmur, J. G. (1994) *Genomics* **19**, 242-248

Kelley, C., Blumberg, H., Zon, L. I. & Evans, T. (1993) *Development* **118**, 817-827

Kiebler, M., Keil, P., Schneider, H., van der Klei, I. J., Pfanner, N. & Neupert, W. (1993) *Cell* **74**, 483-492

Kikuchi, G., Kumar, A., Talmage, P. & Shemin, D. (1958) *J. Biol. Chem.* **233**, 1214-1219

Kikuchi, G. & Hayashi, N. (1981) *Mol. Cell. Biochem.* **37**, 27-41

Kim, C. G., Swendeman, S. L., Barnhart, K. M. & Sheffery, M. (1990) *Mol. Cell. Biol.* **10**, 5958-5966

Kim, H. Y., LaVaute, T., Iwai, K., Klausner, R. D. & Rouault, T. A. (1996a) *J. Biol. Chem.* **271**, 24226-24230

Kim, I. S., Sinha, S., deCrombrugghe, B. & Maity, S. N. (1996b) *Mol. Cell. Biol.* **16**, 4003-4013

Kingsley, C. & Winoto, A. (1992) *Mol. Cell. Biol.* **12**, 4251-4261

Kioussis, D., Vanin, E. DeLange, T., Flavell, R. A. & Grosveld, F. G. (1983) *Nature* **306**, 662-666

Klausner, R. D., Rouault, T. A. & Harford, J. B. (1993) *Cell* **72**, 19-28

Klinken, S. P., Nicola, N. A. & Johnson, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8506-8510

Klinken, P., Busfield, S., Keil, U., Farr, T., Colley, S., Callus, B., Chappel, D. & Papadimitriou, J. (1993) *J. Comput.-Assist. Microsc.* **5**, 81-84

Klinken, S. P. (1996) *personal communication*

Ko, L. J. & Engel, J. D. (1993) *Mol. Cell. Biol.* **13,** 4011-4022

Koenig, R. J., Lazar, M. A., Hodin, R. A., Brent, G. A., Larsen, P. R., Chin, W. W. & Moore, D. D. (1988) *Nature* **337,** 659-661

Kohno, H., Furukawa, T., Yoshinaga, T., Tokunaga, R. & Taketani, S. (1993) *J. Biol. Chem.* **268,** 21350-21363

Konieczny, S. F. & Emerson Jr., C. P. (1987) *Mol. Cell. Biol.* **7,** 3065-3075

Kotkow, K. J. & Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 4640-4647

Kotkow, K. J. & Orkin, S. H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 3514-3518

Kraut, N., Frampton, J., McNagny, K. M. & Graf, T. (1994) *Genes & Dev.* **8,** 33-44

Kreidberg, J. A. & Kelley, J. T. (1986) *Mol. Cell. Biol.* **6,** 2903-2909

Kruh, J. & Borsook, H. (1956) *J. Biol. Chem.* **220,** 905-915

Kudla, B., Caddick, M. X., Langdon, T., Martinez-Rossi, N. M., Bennett, C. F., Sibley, S., Davies, R. W. & Arst Jr., H. N. (1990) *EMBO J.* **9,** 1355-1364

Kulozik, A. E., Bellan-Koch, A., Bail, S., Kohne, E. & Kleihauer, E. (1991) *Blood* **77,** 2054-2058

Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods in Enzymol.* **154,** 367-382

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. (1994) *Nature* **370,** 477-481

Lacronique, V., Lopez, S., Miquerol, L., Porteu, A., Kahn, A. & Raymondjean, M. (1995) *J. Biol. Chem.* **270,** 14989-14997

Landschulz, K. T., Noyes, A. N., Rogers, O. & Boyer, S. H. (1989) *Blood* **73,** 1476-1486

Lathrop, J. T. & Timko, M. P. (1993) *Science* **259,** 522-525

Lauble, H., Kennedy, M. C., Beinert, H. & Stout, C. D. (1992) *Biochem.* **31,** 2735-2748

Laverriere, A. C., MacNeill, C., Mueller, C., Poelman, R. E., Burch, J. B. E. & Evans, T. (1994) *J. Biol. Chem.* **269,** 23177-23184

Lazar, M. A., Hodin, R. A., Darling, D. S. & Chin, W. W. (1988) *Mol. Endocrinol.* **2,** 893-901

Lazar, M. A., Hodin, R. A., Darling, D. S. & Chin, W. W. (1989a) *Mol. Cell. Biol.* **9,** 1128-1136

Lazar, M. A., Hodin, R. A. & Chin, W. W. (1989b) *Proc. Natl. Acad. Sci. U.S.A.* **86,** 7771-7774

Leibold, E. A. & Munro, H. N. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85,** 2171-2175

Lemarchandel, V., Ghysdael, J., Mignotte, V., Rahuel, C. & Romeo, P. H. (1993) *Mol. Cell. Biol.* **13,** 668-676

Leong, S. A., Williams, P. H. & Ditta, G. S. (1985) *Nucleic Acids Res.* **13,** 5965-5976

Leonard, M., Brice, M., Engel, J. D. & Papayannopoulo, T. (1993) *Blood* **82,** 1071-1079

LePrince, D., Gegonne, A., Coll, J., de Taisne, C., Schneeberger, A., Legrou, C. & Stehelin, D. (1983) *Nature* **306**, 395-397

Lewis, S. M. (1990) in: *Postgraduate haematology* pp. 1-25 (A. V. Hoffbrand and S. M. Lewis, eds.) Heinemann Medical Books

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

Liebhaber, S. A., Griese, E. U., Weiss, I., Cash, F. E., Ayyub, H., Higgs, D. R. & Horst, J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9431-9435

Lim, G. B. & Wintour, E. M. (1993) *Today's Life Sci.* **5**, 16-22

Lim, K. C., Ishihara, H., Riddle, R. D., Yang, Z., Andrews, N., Yamamoto, M. & Engel, J. D. (1994) *Nucleic Acids Res.* **22**, 1226-1233

Lloyd, J. A., Krakowsky, J. M., Crable, S. C. & Lingrel, J. A. (1992) *Mol. Cell. Biol.* **12**, 1561-1567

Lopez, A. F., To, L. B. & Yang, Y. C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2761-2765

Lozzio, C. B. & Lozzio, B. B (1975) *Blood* **45**, 321-334

Lu, S.-J., Rowan, S., Bani, M. R. & Ben-David, Y. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8398-8402

Luckow, B & Schutze, G. (1987) *Nucleic Acids Res.* **15**, 5490

Macleod, K. & Plumb, M. (1991) *Mol. Cell. Biol.* **11**, 4324-4332

Mager, D. & Bernstein, A. (1979) *J. Cell Physiol.* **100**, 467-479

Magram, J., Chada, K. & Costantini, F. (1985) *Nature* **315**, 338-340

Maguire, D. J., Day, A. R., Borthwick, I. A., Srivastava, G., Wigley, P. L., May, B. K. & Elliott, W. H. (1986) *Nucleic Acids Res.* **14**, 1379-1391

Majello, B., De Luca, P., Hagen, G., Suske, G. & Lania, L. (1994) *Nucleic Acids Res.* **22**, 4914-4921

Majello, B., De Luca, P., Suske, G. & Lania, L. (1995) *Oncogene* **10**, 1841-1848

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. & Evans, R. M. (1995) *Cell* **83**, 835-839

Marceau, M., Lewis, S. D., Kojiro, C. L., Mountjoy, K. & Shafer, J. A.(1990) *J. Biol. Chem.* **265**, 20421-20429

Martasek, P., Camadro, J. M., Delfau-Larue, M.-H., Dumas, J. B., Montagne, J. J., de Verneuil, H., Labbe, P. & Grandchamp, B. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3020-3028

Martin, D. I. K. & Orkin, S. H. (1990) *Genes & Dev.* **4**, 1886-1898

Martin, D. I. K., Zon, L. I., Mutter, G. & Orkin, S. H. (1990) *Nature* **344**, 444-446

Martin, D. I. K., Fiering, S. & Groudine, M. (1996) *Current Opinion in Genetics & Development* **6**, 488-495

Marziali, G., Perrotti, E., Ilari, R., Testa, U., Coccia, E. M. & Battistini, A. (1997) *Mol. Cell. Biol.* **17**, 1387-1395

Matthews, C. S. (1997) *personal communication*

Mattia, E., Rao, K. K., Shapiro, D. S., Sussman, H. H. & Klausner, R. D. (1984) *J. Biol. Chem.* **259**, 2689-2692

Max-Audit, I., Eleouet, J.-F. & Romeo, P.-H. (1993) *J. Biol. Chem.* **268**, 5431-5437

Maxwell, I. H. & Maxwell, F. (1988) *DNA* **7**, 557-562

May, B. K., Dogra, S. C., Sadlon, T. J., Bhasker, C. R., Cox, T. C. & Bottomley, S. S. (1995) *Prog. Nucleic Acid Res. Mol. Biol.* **51**, 1-51

McClung, C. R., Somerville, J. E., Guerinot, M. L. & Chelm, B. K. (1987) *Gene* **54**, 133-139

Medvinsky, A. L., Samoylina, N. L., Muller, A. M. & Dzierzak, E. A. (1993) *Nature* **364**, 64-67

Meguro, K., Igarashi, K., Yamamoto, M., Fujita, H. & Sassa, S. (1995) *Blood* **86**, 940-948

Melefors, O., Goossen, B., Johansson, H. E., Stripecke, R., Gray, N. K. & Hentze, M. W. (1993) *J. Biol. Chem.* **268**, 5974-5978

Merika, M., & Orkin, S. H. (1993) *Mol. Cell. Biol.* **13**, 3999-4010

Merika, M. & Orkin, S. H. (1995) *Mol. Cell. Biol.* **15**, 2437-2447

Metcalf, D., Begley, C. G. & Johnson, G. R. (1986) *Blood* **67**, 37-45

Metz, T. & Graf, T. (1991a) *Genes & Dev.* **5**, 369-380

Metz, T. & Graf, T. (1991b) *Cell* **66**, 95-105

Metz, T. & Graf, T. (1992) *Oncogene* **7**, 597-605

Mignotte, V., Eleouet, J.-F., Raich, N. & Romeo, P.-H. (1989)
Proc. Natl. Acad. Sci. U.S.A. **86**, 6548-6552

Mignotte, V., Wall, L., deBoer, E., Grosveld, F. & Romeo, P.-H. (1989)
Nucleic Acids Res. **17**, 37-54

Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609-1614

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

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., deBoer, E., Langeveld, A.,
Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. & Fraser, P. (1996) *Cell* **87**,
105-114

Minegishi, N., Minegishi, M., Tsuchiya, S., Fujie, H., Nagai, T., Hayashi, N., Yamamoto,
M. & Konno, T. (1994) *J. Biol. Chem.* **269**, 27700-27704

Mitsuhashi, T., Tennyson, G. E. & Nikodem, V. M. (1988) *Proc. Natl. Acad. Sci. U.S.A.*
85, 5804-5808

Miura, O., Nakamura, N., Quelle, F. W., Witthuhn, B. A., Ihle, J. N. & Aoki, N. (1994)
Blood **84**, 1501-1507

Mohler, J., Vani, K., Leung, S. & Epstein, A. (1991) *Mech. Dev.* **34**, 3-10

Moi, P., Chan, K., Asunis, I., Cao, A. & Kan, Y. W. (1994) *Proc. Natl. Acad. Sci. U.S.A.*
91, 9926-9930

Moreau-Gachelin, F., Tavitian, A. & Tambourin, P. (1988) *Nature* **331**, 277-280

Morley, B. J., Abbott, C. A., Sharpe, J. A., Lida, J., Chan-Thomas, P. S. & Wood, W. G. (1992) *Mol. Cell. Biol.* **12**, 2057-2066

Mouthon, M.-A., Bernard, O., Mitjavila, M.-T., Romeo, P.-H., Vainchenker, W. & Mathieu-Mahul, D. (1993) *Blood* **81**, 647-655

Mueller, R. F., Murray, J. C., Gelinas, R., Farquhar, M. & Papayannopoulou, T. H. (1983) *Hemoglobin* **7**, 245-256

Müller-Eberhard, U. & Fraig, M. (1993) *Am. J. Hematol.* **42**, 59-62

Mullner, E. W., Neupert, B. & Kuhn, L. C. (1989) *Cell* **58**, 373-382

Munoz, A., Zenke, M., Gehring, U., Sap, J., Beug, H. & Vennstrom, B. (1988) *EMBO J.* **7**, 155-159

Murphy, G. & Kavanagh, T. (1989) *Nucleic Acids Res.* **16**, 5198

Nakahashi, Y., Taketani, S., Sameshima, Y. & Tokunaga, R. (1990) *Biochim. Biophys. Acta.* **1037**, 321-327

Nandi, D. L. (1978) *J. Biol. Chem.* **253**, 8872-8877

Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. (1990) *Genes & Dev.* **4**, 993-1006

Nishimura, K., Taketani, S. & Inokuchi, H. (1995) *J. Biol. Chem.* **270**, 8076-8080

Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K. T. & Kawai, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86,** 7711-7715

Novak, T. J. & Rothenberg, E. V. (1988) *DNA* **5,** 439-451

Nudel, U., Salmon, J., Fibach, E., Terada, M., Rifkind, R., Marks, P. A. & Bank, A. (1977) *Cell* **12,** 463-469

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

Nunez, M. T., Glass, J., Fischer, S., Lavidor, L. M., Lenk, E. M. & Robinson, S. H. (1977) *Br. J. Haemotol.* **36,** 519-526

Nunn, M. F., Seeburg, P. H., Moscovici, C. & Duesberg, P. H. (1983) *Nature* **306,** 391-395

Nunn, M. F. & Hunter, T. (1989) *J. Virol.* **63,** 398-402

Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D. & Graves, B. J. (1992) *Genes & Dev.* **6,** 975-990

O' Prey, J., Ramsey, S., Chambers, I., & Harrison, P. R. (1993) *Mol. Cell. Biol.* **13,** 6290-6303

Ohta, Y., Tanaka, M., Terada, M., Miller, O. J., Bank, A., Marks, P. & Rifkind, R. A. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 1232-1236

Orkin, S. H. & Swerdlow, P. S. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74,** 2475-2479

Orkin, S. H., Antonarakis, S. E. & Kazazian Jr, H. H. (1984) *J. Biol. Chem.* **259,** 8679-8681

Orkin, S. H. (1992) *Blood* **80,** 575-581

Orkin, S. H. (1995a) *Eur. J. Biochem.* **231,** 271-281

Orkin, S. H., (1995b) *J. Biol. Chem.* **270,** 4955-4958

Orkin, S. H. (1996) *Current Opinion in Genetics & Dev.* **6,** 597-602

Osada, H., Grutz, G., Axelson, H., Forster, A. & Rabbitts, T. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 9585-9589

Ostertag, W., Melderis, H., Steinheider, G., Kluge, N. & Dube, S. (1972) *Nature* **239,** 231-234

Palma, J. F., Gao, X., Lin, C.-H., Wu, S. & Solomon, W. B. (1994) *Blood* **84,** 1288-1297

Partington, G. A., Bertwistle, D., Nicolas, R. H., Kee, W.-J., Pizzey, J. A. & Patient, R. K. *Developmental Biol.*, in press

Pelicci, P. G., Tabilio, A., Thomopoulos, P., Titeux, M., Vainchenker, W., Rochant, H. & Testa, U. (1982) *FEBS Lett.* **145,** 350-354

Penny, L. A. & Forget, B. G. (1991) *Genomics* **11,** 974-980

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

Peterlin, B. M., Hardy, K. J. & Larsen, A. S. (1987) *Mol. Cell. Biol.* **7,** 1967-1972

Pevney, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S.-F., D'Agati, V., Orkin, S. H. & Costantini, F. (1991) *Nature* **349,** 257-260

Pevney, L., Lin, C.-S., D'Agati, V., Simon, M. C., Orkin, S. H. & Costantini, F. (1995) *Development* **121,** 163-172

Pfeifer, K., Kim, K. S., Kogan, S. & Guarente, L. (1989) *Cell* **56,** 291-301

Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9,** 2159-2167

Philipsen, S., Pruzina, S. & Grosveld, F. (1993) *EMBO J.* **12,** 1077-1085

Philpott, C. C., Haile, D., Rouault, T. A. & Klausner, R. D. (1993) *J. Biol. Chem.* **268,** 17655-17658

Piechaczyk, M., Blanchard, J. M., Marty, L., Dani, C., Panabieres, F., El Sabouty, S, Fort, P. & Jeanteur, P. (1984) *Nucleic Acids Res.* **12,** 6951-6963

Pirola, B. A., Mayer, F., Borthwick, I. A., Srivastava, G., May, B. K. & Elliott, W. H. (1984) *Eur. J. Biochem.* **144,** 577-579

Pirola, B. A. (1986) *PhD thesis,* University of Adelaide

Pondel, M. D., George, M. & Proudfoot, N. J. (1992) *Nucleic Acids Res.* **20,** 237-243

Ponka, P. & Schulman, H. M. (1993) *Stem Cells* **11 (suppl 1),** 24-35

Porcher, C., Pitiot, G., Plumb, M., Lowe, S., de Verneuil, H. & Grandchamp, B. (1991) *J. Biol. Chem.* **266,** 10562-10569

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F. W. & Orkin, S. H. (1996) *Cell* **86,** 47-57

Porter, D. L. & Goldberg, M. A. (1993) *Exp. Hematol.* **21**, 399-404

Potter, H., Weir, L. & Leder, P. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7161-7165

Prasad, K. S., Jordan, J. E., Koury, M. J., Bondurant, M. C. & Brandt, S. J. (1995)
J. Biol. Chem. **270**, 11603-11611

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. & Philipsen, S. (1991)
Nucleic Acids Res. **19**, 1413-1419

Pugh, C. W., Tan, C. C., Jones, R. W. & Ratcliffe, P. J. (1991)
Proc. Natl. Acad. Sci. U.S.A. **88**, 10553-10557

Rahuel, C., Vinit, M.-A., Lemarchandel, V., Cartron, J.-P., & Romeo, P.-H. (1992)
EMBO J. **11**, 4095-4102

Raich, N., Mignotte, V., Dubart, A., Beaupain, D., Leboulch, P., Romana, M., Chabret, C.,
Charnay, P., Papayannopoulou, T., Goossens, M. & Romeo, P.-H. (1989) *J. Biol. Chem.*
264, 10186-10192

Raich, N. & Romeo, P.-H. (1993) *Stem Cells* **11**, 95-104

Raich, N., Clegg, C. H., Grofti, J., Romeo, P.-H. & Stamatoyannopoulos, G. (1995)
EMBO J. **14**, 801-809

Rao, K. K., Shapiro, D. S., Mattia, E., Bridges, K. & Klausner, R. D. (1985)
Mol. Cell. Biol. **5**, 595-600

Ren, S., Luo, X. N. & Atweh, G. F. (1993) *Blood* **81**, 1058-1066

Reuben, R. C., Wife, R. L., Breslow, R., Rifkind, R. A. & Marks, P. A. (1976)
Proc. Natl. Acad. Sci. U.S.A. **73**, 862-866

Riddle, R. D., Yamamoto, M. & Engel, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86,** 792-796

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R. P., Metcalf, D. & Begley, C. G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 7075-7079

Rohde, M., Srivastava, G., Rylatt, D. B., Bundesen, P., Zamattia, J., Crane, D. I. & May, B. K. (1990) *Arch. Biochem. Biophys.* **280,** 331-335

Romana, M., Dubart, A., Beaupain, D., Chabret, C., Goossens, M. & Romeo, P.-H. (1987) *Nucleic Acids Res.* **15,** 7343-7356

Romeo, P.-H., Raich, N., Dubart, A., Beaupain, D., Pryor, M., Kushner, J., Cohen-Solal, M. & Goossens, M. (1986) *J. Biol. Chem.* **261,** 9825-9831

Romeo, P.-H., Prandini, M.-H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. & Uzan, G. (1990) *Nature* **344,** 447-449

Ross, J., Itawa, Y. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69,** 3620-3623

Rougeon, F. & Mach, B (1977) *Gene* **1,** 229-239

Rouault, T. A., Hentze, M. W., Caughman, S. W., Harford, J. B. & Klausner, R. D. (1988) *Science* **241,** 1207-1210

Rouault, T. A., Stout, C. D., Kaptain, S., Harford, J. B. & Klausner, R. D. (1991) *Cell* **64,** 881-883

Rutherford, T. R., Clegg, J. B. & Weatherall, D. J. (1979) *Nature* **280,** 164-165

Ryan, T. M., Behringer, R. R., Martin, N. C., Townes, T. M., Palmiter, R. D. & Brinster, R. L. (1989) *Genes & Dev.* **3,** 314-323

Sadlon, T. J. (1995) *PhD thesis,* University of Adelaide

Sadlon, T. J. (1997) *personal communication*

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed.,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sanchez-Garcia, I. & Rabbitts, T. H. (1993) *Seminars in Cancer Biol.* **4,** 349-358

Sanger, F., Niucklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463-5467

Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H. & Vennstrom, B. (1986) *Nature* **324,** 635-640.

Sap, J., Munoz, A., Schmitt, J., Stunnenberg, H. & Vennstrom, B. (1989) *Nature* **340,** 242-244

Sassa, S. (1975) *J. Exp. Med.* **143,** 305-315

Sassa, S (1983) in *Regulation of Hemoglobin Biosynthesis*, pp.359 (Goldwasser, E. ed.) Elsevier, North Holland

Schuetze, S., Stenberg, P. E. & Kabat, D. (1993) *Mol. Cell. Biol.* **13,** 5670-5678

Schoenhaut, D. S. & Curtis, P. J. (1986) *Gene* **48,** 55-63

Schoenhaut, D. S. & Curtis, P. J. (1989) *Nucleic Acids Res.* **17,** 7013-7028

Schuster, S. J., Badiavas, E. V., Costa-Giomi, P., Weinmann, R., Erslev, A. J. & Caro, J. (1989) *Blood* **73,** 13-16

Semenza, G. L. & Wang, G. L. (1992) *Mol. Cell. Biol.* **12,** 5447-5454

Semenza, G. L., Roth, P. H., Fang, H.-M. & Wang, G. L. (1994) *J. Biol. Chem.* **269,** 23757-23763

Sharma, R. C. & Schimke, R. T. (1996) *BioTechniques* **20,** 42-44

Sharpe, J. A., Ayyub, H., Pretonus, I. M., Buckle, V. J. & Higgs, D. G. (1990) *Blood* **76,** 221-227

Sharpe, J. A., Summerhill, R. J., Vyas, P., Gourdon, G., Higgs, D. R. & Wood, W. G. (1993a) *Blood* **82,** 1666-1671

Sharpe, J. A., Wells, D. J., Whitelaw, E., Vyas, P., Higgs, D. R. & Wood, W. G. (1993b) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 11262-11266

Sherton, C. C. & Kabat, D. (1976) *Dev. Biol.* **48,** 118-131

Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. (1995a) *Nature* **373,** 432-437

Shivdasani, R. A. & Orkin, S. H. (1995b) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 8690-8694

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

Sieweke, M. H., Tekotte, H., Frampton, J. & Graf, T. (1996) *Cell* **85,** 49-60

Simon, M. C., Pevney, L., Wiles, M. V., Keller, G.,Costantini, F. & Orkin, S. H. (1992) *Nature Genet.* **1,** 92-98

Smale, S. T. & Baltimore, D. (1989) *Cell* **57**, 103-113

Smith, D. B. & Johnson, K. S. (1988) *Gene* **67**, 31-40

Smith, A. (1990) in *Biosynthesis of Heme and Chlorophylls* pp.435 (Dailey, H. A. ed.) McGraw-Hill, New York

Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H. & Fujii-Kuriyama, Y. (1993) *Nucleic Acids Res.* **21**, 1527-1532

Southwood, C. M., Downs, K. M. & Bieker, J. J. (1996) *Dev. Dyn.* **206**, 248-259

Sowden, J., Edwards, M., Morrison, K., Butterworth, P. H. W. & Edwards, Y. H. (1992) *Biochem. J.* **288**, 545-551

Spandidos, D. A. & Wilkie, N. M. (1984) in *Transcription and Translation*, pp1-48 (Hames, B. D. & Higgins, S. J., eds.)

Sposi, N. M., Zon, L. I., Care, A., Valtieri, M., Testa, U., Gabbianelli, M., Mariani, G., Bottero, L., Mather, C., Orkin, S. H. & Peschle, C. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6353-6357

Srivastava, G., Borthwick, I. A., Brooker, J. D., May, B. K. & Elliott, W. H. (1982) *Biochem. Biophys. Res. Commun.* **109**, 305-312

Srivastava, G., Borthwick, I. A., Brooker, J. D., Wallace, J. C., May, B. K. & Elliott, W. H. (1983) *Biochem. Biophys. Res. Commun.* **117**, 344-349

Srivastava, G., Borthwick, I. A., Maguire, D. J., Elferink, C. J., Bawden, M. J., Mercer, J. F. B. & May, B. K. (1988) *J. Biol. Chem.* **263**, 5202-5209

Srivastava, G., Hansen, A. J., Bawden, M. J. & May, B. K. (1990) *Mol. Pharmacol.* **38,** 486-493

Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. & Varmus, H. (1994) in *The Molecular Basis of Blood Diseases,* pp. 108-109, Saunders, W. B. Company, Philadelphia

Stamatoyannopoulos, J. A., Goodwin, A., Joyce, T. & Lowrey, C. (1995) *EMBO J.* **14,** 106-116

Strauss, E. C., Andrews, N. C., Higgs, D. R. & Orkin, S. H. (1992) *Mol. Cell. Biol.* **12,** 2135-2142

Sugawara, A., Yen, P. M., Darling, D. S. & Chin, W. W. (1993) *Endocrinology* **133,** 965-971

Supp, D. M., Witte, D. P., Branford, W. W., Smith, E. P. & Potter, S. S. (1996) *Developmental Biology* **176,** 284-299

Sutherland, G. R., Baker, E., Callen, D. F., Hyland, V. J., May, B. K., Bawden, M. J., Healy, H. M. & Borthwick, I. A. (1988) *Am. J. Hum. Genet.* **43,** 331-335

Sutherland, H. G. E., Martin, D. I. K. & Whitelaw, E. (1997) *Mol. Cell. Biol.* **17,** 1607-1614

Swaroop, A., Xu, J., Pawar, H., Jackson, A., Skolnick, C. & Agarwal, N. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89,** 266-270

Taketani, S., Nakahashi, Y., Osumi, T. & Tokunaga, R. (1990) *J. Biol. Chem.* **256,** 19377-19380

Taketani, S., Inazawa, J., Nakahashi, Y., Abe, T. & Tokunaga, R. (1992) *Eur. J. Biochem.* **205,** 217-222

Taketani, S., Kohno, H., Furukawa, T., Yoshinaga, T. & Tokinaga, R. (1994) *Biochim. Biophys. Acta.* **1183,** 547-549

Taketani, S., Inazawa, J., Abe, T., Furukawa, T., Kohno, H., Tokunaga, R., Nishimura, K. & Inokuchi, H. (1995) *Genomics* **29,** 698-703

Talbot, D. S., Philipsen, S., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9,** 2169-2178

Talbot, D. S. & Grosveld, F. (1991) *EMBO J.* **10,** 1391-1398

Taxman, D. J., Lee, E. S. & Wojchowski, D. M. (1993) *Anal. Biochem.* **213,** 97-103

Taxman, D. J. & Wojchowski, D. M. (1995) *J. Biol. Chem.* **270,** 6619-6627

Theil, E. C. (1990) *J. Biol. Chem.* **265,** 4771-4774

Thompson, C. C., Weinberger, C., Lebo, R. & Evans, R. M. (1987) *Science* **237,** 1610-1614

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. & Palmiter, R. D. (1985) *EMBO J.* **4,** 1715-1723

Trainor, C. D., Evans, T., Felsenfeld, G. & Boguski, M. S. (1990) *Nature* **343,** 92-96

Trainor, C. D., Omichinski, J. G., Vandergon, T. L., Gronenborn, A. M., Clore, G. M. & Felsenfeld, G. (1996) *Mol. Cell. Biol.* **16,** 2238-2247

Treisman, R. & Maniatis, T. (1985) *Nature* **315,** 73-75

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

Tsai, S. F., Martin, D. I. K., Zon, L. I., D'Andrea, A. D., Wong, G. G., & Orkin, S. H. (1989) *Nature* **339,** 446-451

Tsai, S. F., Strauss, E. & Orkin, S. H. (1991) *Genes & Dev.* **5,** 919-931

Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M. & Orkin, S. H. (1997) *Cell* **90,** 109-119

Tsukiyama, T. & Wu, C. (1995) *Cell* **83,** 1011-1020

Tuan, D., Solomon, W., Li, Q. & London, I. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82,** 6384-6388

Tugores, A., Magness, S. T. & Brenner, D. A (1994) *J. Biol. Chem.* **269,** 30789-30797

Tur-Kaspa, R., Teicher, L., Levine, B. J., Skoultchi, A. I. & Shafritz, D. A. (1986) *Mol. Cell. Biol.* **6,** 716-718

Urban-Grimal, D., Volland, C., Garnier, T., Dehoux, P. & Labbe-Bois, R. (1986) *Eur. J. Biochem.* **156,** 511-519

Valge-Archer, V. E., Osada, H., Warren, A. J., Forster, A., Li, J., Baer, R. & Rabbitts, T. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 8617-8621

Virbasius, C. A., Virbasius, J. V. & Scarpulla, R. C. (1993) *Genes & Dev.* **7,** 2431-2445

Visvader, J. E., Begley, C. G. & Adams, J. M. (1991) *Oncogene* **6,** 187-194

Visvader, J. E., Elefanty, A. G., Strasser, A. & Adams, J. M. (1992) *EMBO J.* **11,** 4557-4564

Visvader, J. E. & Adams, J. M. (1993) *Blood* **82**, 1493-1501

Voigtlander, T., Ripperger, A., Ganten, D. & Bader, M. (1995) *Adv. Exp. Med. Biol.* **377**, 285-292

Volland, C. & Felix, F. (1984) *Eur. J. Biochem.* **142**, 551-557

Vyas, P., Vickers, M. A., Simmons, D. L., Ayyub, H., Craddock, C. F. & Higgs, D. R. (1992) *Cell* **69**, 781-793

Wada, O., Sassa, S., Takaku, F., Yano, Y., Uratta, G. & Nakao, K. (1967) *Biochim. Biophys. Acta* **148**, 585-587

Wadman, I., Li, J., Bash, R. O., Forster, A., Osada, H., Rabbitts, T. H. & Baer, R. (1994) *EMBO J.* **13**, 4831-4839

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

Wall, L., Destroismaisons, N., Delvoye, N. & Guy, L. G. (1996) *J. Biol. Chem.* **271**, 16477-16484

Walters, M. & Martin, D. I. K. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10444-10448

Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. & Martin, D. I. K. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7125-7129

Walters, M. C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. & Martin, D. I. K. (1996) *Genes & Dev.* **10**, 185-195

Wang, G. L. & Semenza, G. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4304-4308

Ward, J. H., Kushner, J. P. & Kaplan, J. (1982) *J. Biol. Chem.* **257**, 10317-10323

Ward, J. H., Jordan, I., Kushner, J. P. & Kaplan, J. (1984) *J. Biol. Chem.* **259**, 13235-13240

Warner, C. A., Yoo, H. W., Tsai, S.-F., Roberts, A. G. & Desnick, R. J. (1990) *Am. J. Hum. Genet.* **47 (Abstr.)**, 83

Warnick, G. R. & Burnham, B. F. (1971) *J. Biol. Chem.* **246**, 6880-6885

Warren, A. J., Colledge, W. H., Carlton, M. B. L., Evans, M. J., Smith, A. J. H. & Rabbitts, T. H. (1994) *Cell* **78**, 45-57

Wasylyk, B., Hahn, S. L. & Giovane, A. (1993) *Eur. J. Biochem.* **211**, 7-18

Watowich, S. S., Yoshimura, A., Longmore, G. D., Hilton, D. J., Yoshimura, Y. & Lodish, H. F. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2140-2144

Watowich, S. S., Hilton, D. J. & Lodish, H. F. (1994) *Mol. Cell. Biol.* **14**, 3535-3549

Watson, D. K., McWilliams, M. J., Lapis, P., Lautenberger, J. A., Schweinfest, C. W. & Papas, T. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7862-7866

Watt, P., Lamb, P., Squire, L. & Proudfoot, N. (1990) *Nucleic Acids Res.* **18**, 1339-1350

Waxman, H. S. & Rabinovitz, M. (1965) *Biochem. Biophys. Res. Commun.* **19**, 538-545

Waxman, H. S. & Rabinovitz, M. (1966) *Biochim. Biophys. Acta* **129**, 369-379

Waxman, H. S., Freedman, M. L. & Rabinovitz, M. (1967) *Biochim. Biophys. Acta* **145**, 353-360

Weatherall, D. (1994) *Blood* **80,** 1379-1381

Weber-Benarous, A., Cone, R. D., London, I. M. & Mulligan, R. C. (1988) *J. Biol. Chem.* **263,** 6142-6145

Weber, F. & Shaffner, W. (1985) *Nature* **315,** 75-77

Weinberger, C., Thompson, C. C., Ong, E. S., Lebo, R., Gruol, D. J. & Evans, R. M. (1986) *Nature* **324,** 641-646

Weiss, M. J., Keller, G. & Orkin, S. H. (1994) *Genes & Dev.* **8,** 1184-1197

Weiss, M. J. & Orkin, S. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 9623-9627

Wickrema, A., Krantz, S. B., Winkelmann, J. C. & Bondurant, M. C. (1992) *Blood* **80,** 1940-1949

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

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. & Fraser, P. (1996) *Genes & Dev.* **10,** 2894-2902

Witthuhn, B. A., Quelle, F. W., Silvennoinen, O., Yi, T., Tang, B., Miura, O. & Ihle, J. N. (1993) *Cell* **74,** 227-236

Wolfe, A. P. (1997) *Nature* **387,** 16-17

Wood, J. S. & Dixon, R. L. (1972) *Biochem. Pharmacol.* **21,** 1735-1744

Wood, J. S. (1973) *Mol. Pharmacol.* **10,** 389-397

Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1585-1588

Worthington, R. E., Bossie-Codreanu, J. & Van Zant, G. (1987) *Exp. Hematol.* **15,** 85-92

Wu, Y., Fraizer, G. C. & Saunders, G. F. (1995) *J. Biol. Chem.* **270,** 5944-5949

Xiao, J. H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. & Chambon, P. (1987) *Genes & Dev.* **1,** 794-807

Yamamoto, M., Kure, S., Engel, J. D. & Hiraga, K. (1988) *J. Biol. Chem.* **263,** 15973-15979

Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H. & Engel, J. D. (1990) *Genes & Dev.* **4,** 1650-1662

Yamauchi, K., Hayashi, N. & Kikuchi, G. (1980) *J. Biol. Chem.* **255,** 1746-1751

Yang, H.-Y. & Evans, T. (1992) *Mol. Cell. Biol.* **12,** 4562-4570

Yang, J. M., London, I. M. & Chen, J.-J. (1992) *J. Biol. Chem.* **267,** 20519-20524

Yen, P. M., Ikeda, M., Wilcox, E. C., Brubaker, J. H., Spanjaard, R. A., Sugawara, A. & Chin, W. W. (1994) *J. Biol. Chem.* **269,** 12704-12709

Yomogida, K., Yamamoto, M., Yamagami, T., Fujita, H. & Hayashi, N. (1993) *J. Biochem. (Tokyo)* **113,** 364-371

Youssoufian, H. (1994) *Blood* **83,** 1428-1435

Yu, C.-H., Motamed, K., Chen, J., Bailey, A. D. & Shen, C.-K. J. (1991) *J. Biol. Chem.* **266,** 8907-8915

Zenke, M., Kahn, P., Disela, C., Vennstrom, B., Leutz, A., Keegan, K., Hayman, M. J., Choi, H. R., Yew, N., Engel, J. D & Beug, H. (1988) *Cell* **52,** 107-119

Zenke, M., Munoz, A., Sap, J., Vennstrom, B. & Beug, H. (1990) *Cell* **61,** 1035-1049

Zhang, Q., Reddy, P. M., Yu, C.-Y., Bastiani, C., Higgs, D. R., Stamatoyannopoulos, G., Papayannopoulou, T. & Shen, C.-K. J. (1993) *Mol. Cell. Biol.* **13,** 2298-2308

Zhang, Q., Rombel, I., Reddy, G. N., Gang, J. B. & Shen, C. K. (1995) *J. Biol. Chem.* **270,** 8501-8505

Zon, L. I., Tsai, S. F., Burgess, S., Matsudaira, P., Bruns, G. A. & Orkin, S. H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87,** 668-672

Zon, L. I., Youssoufian, H., Mather, C., Lodish, H. F. & Orkin, S. H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88,** 10638-10641

Zon, L. I., Yamaguchi, Y., Yee, K., Albee, E. A., Kimura, A., Bennett, J. C., Orkin, S. H. & Ackerman, S. J. (1993) *Blood* **81,** 3234-3241

PUBLICATIONS

The manuscript entitled "Transcriptional Regulation of the Human Erythroid 5-Aminolevulinate Synthase Gene: Identification of Promoter Elements and Role of Regulatory Proteins" has been recently accepted for publication in The Journal of Biological Chemistry. A copy of this manuscript has been included.

Surinya, K.H., Cox, T.C., and May, B.K., (1997) Transcriptional regulation of the human erythroid 5-aminolevulinate synthase gene: identification of promoter elements and role of regulatory proteins.
Journal of Biological Chemistry, v. 272 (42), pp. 26585-26594.

REFERENCES

1. May, B. K., Dogra, S. C., Sadlon, T. J., Bhasker, C. R., Cox, T. C., and Bottomley, S. S. (1995) *Prog. Nucleic Acid Res. Mol. Biol.* **51,** 1-51

2. Bottomley, S. S., May, B. K., Cox, T. C., Cotter, P. D., and Bishop, D. F. (1995) *J. Bioenergetics & Biomembranes* **27,** 161-168

3. Dierks, P.(1990) in *Biosynthesis of Heme and Chlorophylls*, pp. 201-233 (Dailey, H. A., ed.), McGraw-Hill, New York

4. Sutherland, G. R., Baker, E., Callen, D. F., Hyland, V. J., May, B. K., Bawden, M. J., Healy, H. M., and Borthwick, I. A. (1988) *Am. J. Hum. Genet.* **43,** 331-335

5. Cox, T. C., Bawden, M. J., Abraham, N. G., Bottomley, S. S., May, B. K., Baker, E., Chen, L. Z., and Sutherland, G. R. (1990) *Am. J. Hum. Genet.* **46,** 107-111

6. Bishop, D. F., Henderson, A. S., and Astrin, K. H. (1990) *Genomics* **7,** 207-214

7. Kappas, A., Sassa, S., Galbraith, R. A., and Nordmann, Y. (1995) in *Metabolic and Molecular Basis of Inherited Disease*, 7th Ed., pp.2103-2159 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds.), McGraw-Hill, New York

8. Maguire, D. J., Day, A. R., Borthwick, I. A., Srivastava, G., Wigley, P. L., May, B. K., and Elliot, W. H. (1986) *Nucleic Acids Res.* **14,** 1379-1391

9. Yomogida, K., Yamamoto, M., Yamagami, T., Fujita, H., and Hayashi, N. (1993) *J. Biochem. (Tokyo)* **113,** 364-371

10. Schoenhaut, D. S., and Curtis, P. J. (1989) *Nucleic Acids Res.* **17,** 7013-7028

11. Cox, T. C., Bawden, M. J., Martin, A., and May, B. K. (1991) *EMBO J.* **10,** 1891-1902

12. Lim, K. C., Ishihara, H., Riddle, R. D., Yang, Z., Andrews, N., Yamamoto, M., and Engel, J. D. (1994) *Nucleic Acids Res.* **22,** 1226-1233

13. Conboy, J. G., Cox, T. C., Bottomley, S. S., Bawden, M. J., and May, B. K. (1992) *J. Biol. Chem.* **267,** 18753-18758

14. Cox, T. C., Bottomley, S. S., Wiley, J. S., Bawden, M. J., Matthews, C. S., and May, B. K. (1994) *N. Engl. J. Med.* **330,** 675-679

15. Bhasker, C. R., Burgiel, G., Neupert, B., Emery-Goodman, A., Kuhn, L. C., and May, B. K. (1993) *J. Biol. Chem.* **268,** 12699-12705

16. Lathrop, J. T., and Timko, M. P. (1993) *Science* **259,** 522-525

17. Beaumont, C., Deybach, J. C., Grandchamp, B., Silva, V. D., de Verneuil, H., and Nordmann, Y. (1984) *Exp. Cell Res.* **154,** 474-484

18. Karlsson, S., and Nienhuis, A. W. (1985) *Annu. Rev. Biochem.* **54,** 1071-1108

19. Orkin, S. H. (1995) *J. Biol. Chem.* **270,** 4955-4958

20. Bieker, J. J., and Southwood, C. M. (1995) *Mol. Cell. Biol.* **15**, 852-860

21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

22. Herbomel, P., Bourachot, B., and Yaniv, M. (1984) *Cell* **39**, 653-662

23. Partington, G. A., Bertwistle, D., Nicolas, R. H., Kee, W. -J., Pizzey, J. A., and Patient, R. K. *Developmental Biol.*, in press

24. Andrews, N. C., and Faller, D. V. (1991) *Nucleic Acids Res.* **19**, 2499

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

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

27. Evans, M. J., and Scarpulla, R. C. (1990) *Genes & Dev.* **4**, 1023-1034

28. Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J. D., and Yamamoto, M. (1993) *Nature* **362**, 466-468

29. Fong, T. C., and Emerson, B. M. (1992) *Genes & Dev.* **6**, 521-532

30. Crossley, M., Merika, M., and Orkin, S. H. (1995) *Mol. Cell. Biol.* **15**, 2448-2456

31. Smith, D. B. and Johnson, K. S. (1988) *Gene* **67**, 31-40

32. Crossley, M., Whitelaw, E., Perkins, A., Williams, G., Fujiwara, Y., and Orkin, S. H. (1996) *Mol. Cell. Biol.* **16**, 1695-1705

33. Chodosh, L., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988) *Cell* **53**, 11-24

34. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211**, 7-18

35. Merika, M., and Orkin, S. H. (1993) *Mol. Cell. Biol.* **13**, 3999-4010

36. Ko, L. J., and Engel, J. D. (1993) *Mol. Cell. Biol.* **13**, 4011-4022

37. Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I., and Santoro, C. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11598-11602

38. Max-Audit, I., Eleouet, J.-F., and Romeo, P.-H. (1993) *J. Biol. Chem.* **268**, 5431-5437

39. Rahuel, C., Vinit, M.-A., Lemarchandel, V., Cartron, J.-P., and Romeo, P.-H. (1992) *EMBO J.* **11**, 4095-4102

40. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51**, 1079-1090

41. Kingsley, C., and Winoto, A. (1992) *Mol. Cell. Biol.* **12**, 4251-4261

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

43. Donze, D., Townes, T. M. and Bieker, J. J. (1995) *J. Biol. Chem.* **270**, 1955-1959

44. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1993) *Nature* **362**, 722-728

45. Andrews, N. C., Kotkow, K. J., Ney, P. A., Erdjument-Bromage, H., Tempst, P., and Orkin, S. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11488-11492

46. deBoer, E., Antoniou, M., Mignotte, V., Wall, L., and Grosveld, F. (1988) *EMBO J.* **7**, 4203-4212

47.   Mignotte, V., Eleouet, J.-F., Raich, N., and Romeo, P.-H. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6548-6552

48.   Tsai, S. F., Strauss, E., and Orkin, S. H. (1991) *Genes & Dev.* **5,** 919-931

49.   Zon, L. I., Youssoufian, H., Mather, C., Lodish, H. F., and Orkin, S. H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 10638-10641

50.   Barton, M. C., Madani, N., and Emerson, B. M. (1993) *Genes & Dev.* **7,** 1796-1809

51.   Smale, S. T., and Baltimore, D. (1989) *Cell* **57,** 103-113

52.   Frampton, J., Walker, M., Plumb, M., and Harrison, P. R. (1990) *Mol. Cell. Biol.* **10,** 3838-3842

53.   Merika, M., and Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 2437-2447

54.   Gregory, R. C., Taxman, D. J., Seshasayee, D., Kensinger, M. H., Bieker, J. J., and Wojchowski, D. M. (1996) *Blood* **87,** 1793-1801

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

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

57.   Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F., and Fraser, P. (1996) *Genes & Dev.* **10,** 2894-2902

58.   Orkin, S. H., Antonarakis, S. E., and Kazazian Jr, H. H. (1984) *J. Biol.Chem.* **259,** 8679-8681

59.   Kulozik, A. E., Bellan-Koch, A., Bail, S., Kohne, E., and Kleihauer, E. (1991) *Blood* **77,** 2054-2058

60.   Feng, W. C., Southwood, C. M., and Bieker, J. J. (1994) *J. Biol. Chem.* **269,** 1493-1500

61.   Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., and Martin, D. I. K. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7125-7129

62.   Walters, M. C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M., and Martin, D. I. K. (1996) *Genes & Dev.* **10,** 185-195

63.   Sutherland, H. G. E., Martin, D. I. K., and Whitelaw, E. (1997) *Mol. Cell. Biol.* **17,** 1607-1614

64.   O' Prey, J., Ramsey, S., Chambers, I., and Harrison, P. R. (1993) *Mol. Cell. Biol.* **13,** 6290-6303

65.   James-Pederson, M., Yost, S., Shewchuk, B., Zeigler, T., Miller, R., and Hardison, R. (1995) *J. Biol. Chem.* **270,** 3965-3973

66.   Fraizer, G. C., Wu, Y.-J., Hewitt, S. M., Maity, T., Ton, C. C., Huff, V., and Saunders, G. F. (1994) *J. Biol. Chem.* **269,** 8892-8900

67.   Taketani, S., Inazawa, J., Abe, T., Furukawa, T., Kohno, H., Tokunaga, R., Nishimura, K., and Inokuchi, H. (1995) *Genomics* **29,** 698-703

68.   Braidotti, G., Borthwick, I. A., and May, B. K. (1993) *J. Biol. Chem.* **268,** 1109-1117

69.   Kaya, A. H., Plewinska, M., Wong, D. M., Desnick, R. J., and Wetmur, J. G. (1994) *Genomics* **19,** 242-248

70.   Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., and Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 6-10

71.   Porcher, C., Pitiot, G., Plumb, M., Lowe, S., de Verneuil, H., and Grandchamp, B. (1991) *J. Biol. Chem.* **266,** 10562-10569

72.   Tugores, A., Magness, S. T., and Brenner, D. A. (1994) *J. Biol. Chem.* **269,** 30789-30797

FOOTNOTES

[1]   The abbreviations used are: ALAS, 5-aminolevulinate synthase; EKLF, erythroid Krüppel-like factor; BKLF, basic Krüppel-like factor; bp, base pair(s); kb, kilobase(s); MEL, murine erythroleukemia; PBS, phosphate-buffered saline; Hepes, N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid; RSV, Rous sarcoma virus; V, volt; TBP, TATA binding protein; PBGD, Porphobilinogen deaminase; GST, glutathione S-transferase.

FORMAL PRESENTATIONS

I have presented my work as a poster presentation at the:

18th Annual Conference on the Organisation and Expression of the Genome (1996)
Lorne, Victoria, Australia

17th Annual Conference on the Organisation and Expression of the Genome (1995)
Lorne, Victoria, Australia

ERRATUM

page 3, line 19: "these cells become increasing specialised" should read "these cells become increasingly specialised.

page 4, line 5: "The activation of the EpoR results in its rapid auto-phosphorylation" should read "The EpoR is activated by JAK2.

page 22, line 16: "yeast factors" should read "mammalian factors"

page 31, line 21: "Epo-induced J2E cells" should read "FVA cells"

page 61, line 6: "Erythroid transformed cell lines" should read "Transformed erythroid cells"

page 69, line 18: "appropriate antibody" should read "appropriate antibiotic"

page 80, line 15: "To eliminate the possible introduction of random errors" should read "To minimise the possible introduction of random errors".

Figure 3.1: "the percentage of hemoglobin cells relative to the total number of viable cells"

page 109, line 10: "erythroid messages" should read "erythroid mRNA"

page 123, line 23: "to investigate the putative role of GATA-1" should read "to investigate the role of GATA-1"

Figure 6.5: the EMSA is overexposed. Later experiments have demonstrated that GATA-1 binds to GATA-A and GATA-B with a similar affinity.

Figure 6.7: The band marked * is not EKLF.