Regulation of globin gene expression: boxes and borders

Regulatie van globine genexpressie:
boxen en grenzen

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam

op gezag van de Rector Magnificus
Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaats vinden op
donderdag 6 december 2001 om 15.45 uur

door

Gaetano Zafarana

geboren te Siracusa (Italië)

Promotiecommissie

Promotor:	Prof. dr. F.G. Grosveld
Overige leden:	Prof. dr. B.A. Oostra Dr. ir. D.N. Meijer Prof. dr. I.P. Touw
Copromoter:	Dr. J.N.J. Philipsen

Dit proefschift werd bewerkt binnen de vakgroep Gene Structure and Expression, National Institute for Medical Research, Engeland en de vakgroep Celbiologie en Genetica, Faculteit der Geneeskunde en Gezondheidswetenschappen, Erasmus Universiteit Rotterdam.

Contents

List of abbreviations

A	Adenine
ATP	Adenosine triphosphate
BAC	Bacterial Artificial Chromosome
bp	base pair(s)
bZip	basic-region-leucine zipper
C	Cytosine
cDNA	complementary DNA
C/EBP	CAAT enhancer binding protein
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dIdC	Deoxyinosinic-doxycytidylic acid
DDT	1,4-Dithio-DL-threitol
EDTA	Ethylenediaminotetraacetic acid
EGTA	Ethylene glycol-bis(2-aminoethyl-)N, N, N', N'-tetraacetic acid
G	Guanine
GTF	General transcription factor
HbA	Adult haemoglobin
HbS	Sickle haemoglobin
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanolsulphonic acid
HMG	High mobility group
HS	Hypersensitive site
hsp	Heat shock protein
IgH	Immunoglobulin heavy chain
kb	Kilo base(s)
kD	Kilo Dalton
LCR	Locus control region
MEL	Mouse erythroleukemia
mRNA	messenger RNA
NF-E6	Nuclear factor-erythroid 6
PAC	P1 artificial chromosome
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PEV	Position effect variegation
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
T	Thymidine
TAF	TBP associated factor
TBP	TATA binding protein
TCA	Trichloroacetic acid
TF	Transcription factor
TrisHcl	Tris(hydroxymethyl)aminomethane hydrochloride
YAC	Yeast artificial chromosome

Greek letters		Symbol		Factor
α	alpha	M	mega	10^{6}
β	beta	k	kilo	10^{3}
γ	gamma	m	milli	10^{-3}
δ	delta	μ	micro	10^{-6}
ε	epsilon	n	nano	10^{-9}
ζ	zeta	p	pico	10^{-12}
θ	theta	f	femto	10^{-15}

Aim of the Thesis

The experimental work presented in this thesis pertains to the regulation of gene expression and focuses on the regulation of the β globin genes. The expression of the globin genes is restricted to the erythropoietic tissue and is regulated during development. Therefore the study of globin gene expression can yield valuable insights in how organisms regulate their genes in a coordinated tissue and developmental manner.

Two main issues will be addressed in the chapters of this thesis. The first concerns the mechanism by which a transcriptionally competent domain is formed in the chromosomal environment. This topic is the subject of Chapter 2. Once transcriptionally competent domains are formed, gene transcription can take place and it is orchestrated by factors that interact specifically with the regulatory elements of the gene itself. This interaction determines tissue specificity and developmental regulation of a gene. The second issue addressed concerns the transcriptional regulation of the β globin gene via their CCAAT boxes studied through NF-E6, a factor that binds to these elements (Chapters 3 and 4). NF-E6 interaction with its target sequence participates in the change in globin gene expression concomitant with the progressive changes that occur during the development of an embryo to a fetus.

Chapter 1

Regulation of globin gene expression

Gaetano Zafarana, Joost Gribnau, Frank Grosveld and Sjaak Philipsen

Regulation of globin gene expression

Haematopoiesis: transcriptional regulation determines the developmental fate of cells.

Blood cells are produced continuously throughout the life span of the individual. Millions of cells are produced every minute to cater for the body's cellular demand for gaseous exchange, to ensure a defence against pathogens and to stop bleeding. This process, called haematopoiesis, requires an extraordinary plasticity of the system involved. Haematopoietic cells originate from rare, self-renewing, pluripotent haematopoietic cells (HSCs). Proliferative expansion of progenitor cells and progressive commitment of progenitors to single lineage differentiation is achieved through the combinatory effect of signalling molecules. Via an ill-defined cascade of secondary messengers the transcription of lineage restricted genes is activated and their expression results in an irreversible differentiation of these precursors to red cells, neutrophils, monocytes/macrophages, megakaryocytes, mast cells and lymphocytes.

During vertebrate embryogenesis, haematopoiesis occurs in successive waves. The first cells of haematopoietic origin (embryonic cells) are produced in the blood islands that arise in the extraembryonic mesoderm of the yolk sac. Intraembryonic precursor cells located in the paraaortic splanchnopleura, which later develops into the aortic-gonad-mesonephros (AGM) region, initiate the definitive haematopoietic system. In the mouse, multipotent haematopoietic progenitors are found in the AGM region before the appearance of fetal liver stem cell activity. At midgestation HSCs migrate to the fetal liver, which becomes the site of definitive haematopoiesis. Shortly after birth and thereafter, the bone marrow becomes the major site of haematopoiesis (Dzierzak et al., 1998).

Mice are widely used in the laboratory to study haematopoiesis. For the studies of erythroid development and globin gene expression, transgenic mice carrying the human globin loci have provided extremely useful tools for the study of human globin gene regulation. Mice have been instrumental in the analysis of transcription factors involved in erythropoiesis and the developmental regulation of the globin genes, because loss of function mutants of these transcription factors can be created in the laboratory ("knockout" mice). In addition, gain of function experiments have also been performed, in which particular transcription factors, and mutant factors, have been expressed in the erythroid lineage via regular transgenesis.

The erythrocytes.

Erythrocytes or red blood cells represent the majority of the cells in the blood. In man there are approximately five million erythrocytes per millilitre of blood, versus 7000 leukocytes per millilitre. However these figure vary according to the age, sex and state of health of each

individual. In an adult, each erythrocyte has a life span of about three months after which time it is destroyed in the spleen or liver. Between 2-10 million erythrocytes are destroyed and replaced each minute in the human body. The rate of destruction and replacement is determined largely by the amount of oxygen in the atmosphere that is available for carriage by the blood. If the quantity carried is low, then the marrow is stimulated to produce more erythrocytes than the liver destroys. This is one of the ways in which mammals adapt to the reduced oxygen content at high altitudes. When the oxygen content is high, the situation is reversed.

The mechanism of production of erythrocytes is known as erythropoiesis. It is a differentiation process that takes place via a series of intermediate precursor cells that arise from a pluripotent stem cell. Ultimately the mature, fully haemoglobinised erythrocyte, leaves the haematopoietic site and enters circulation. 'Early' and 'late' progenitor cells with significant but declining proliferative potential are classified as Burst Forming Units-Erythroid (BFU-E) and Colony Forming Units-Erythroid (CFU-E) respectively, based on the size and morphology of the erythroblast colonies formed in semi-solid medium *in vitro*. Haemoglobin accumulation begins at approximately the pro-erythroblast stage (Nienhuis and Benz, 1977). Small enucleated cells (7-8μm) are produced by definitive erythropoiesis whilst larger nucleated cells originate from primitive erythropoiesis.

Haemoglobin.

The main function of red blood cells is to transport atmospheric oxygen through the body and exchange it for CO_2 produced by cellular respiration. CO_2 is exchanged again for oxygen in the lungs. In erythrocytes, the molecule responsible for the transport of oxygen is haemoglobin. Not surprisingly, mature erythrocytes have haemoglobin as their main and most abundant soluble protein. Human haemoglobin is a globular protein with a diameter of approximately 5.5 nm and a molecular weight of 64.4 kD (Perutz, 1960). The molecule consists of two pairs of unlike globin polypeptide chains ($\alpha_2\beta_2$). A haeme prosthetic group, ferroprotoporphirin IX, responsible for the characteristic red colour of blood, is linked covalently at a specific site to each chain. A ferrous iron atom is located within each haeme group, and each one of these can combine loosely with one molecule of oxygen. Combination of oxygen with haemoglobin, to form oxyhaemoglobin, occurs under conditions where the partial pressure of oxygen is high, such as in the lung alveolar capillaries. When the partial pressure of oxygen is low, as in the capillaries that supply metabolically active tissues, the bonds holding oxygen become unstable and oxygen is released. This diffuses in solution into the surrounding cells.

In the developing human erythroid precursors, eight genes direct the synthesis of six structurally different globin peptides, designated α,β,γ,δ,ε and ζ (Fig. 1). The α-chain gene is

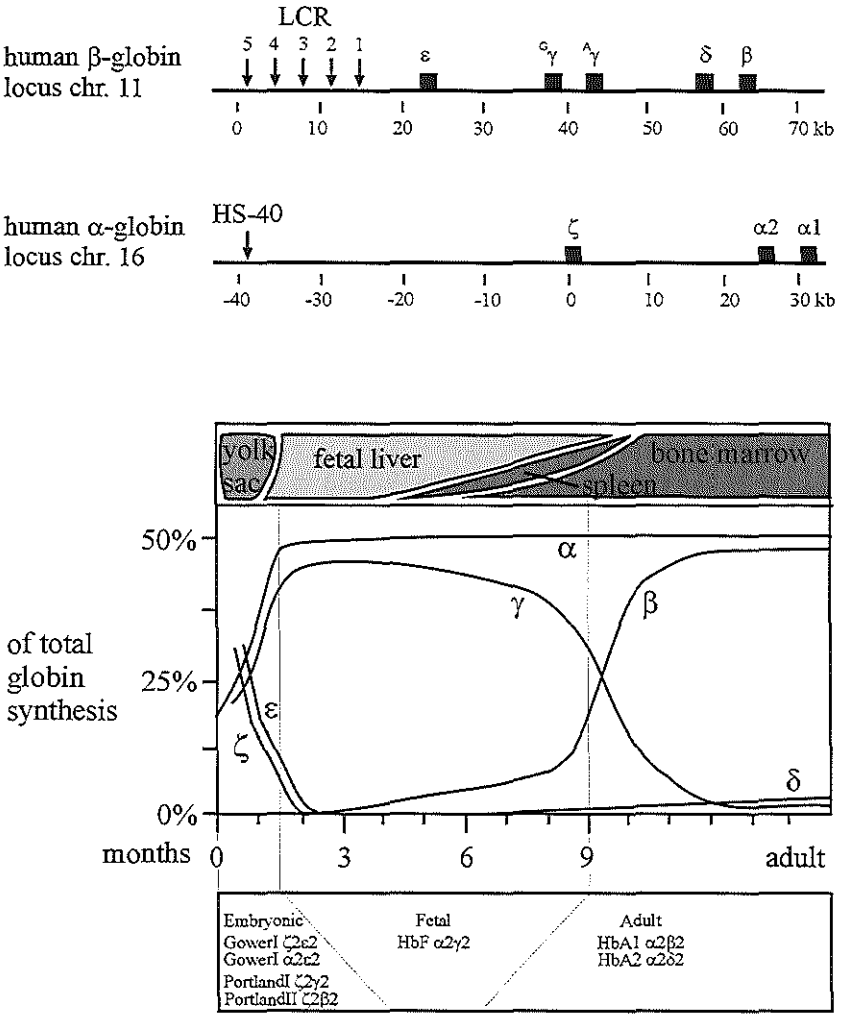

Fig. 1. Organization of the human α- and β-globin loci, the expression of the globin genes and the site of erythropoiesis during development. The α-like and β-like proteins that are expressed at the same stage of development can form different types of hemoglobin with each hetero-tetramer having specific oxygen affinity characteristics. DNaseI hypersensitive sites of important distal regulatory elements are indicated by arrows.

duplicated in humans and is located on chromosome 16, downstream from the ζ gene. The ε, $^{G}\gamma$, $^{A}\gamma$, δ and β genes are arranged in sequential order on chromosome 11. Strong structural homology between the α- and β-chains indicates that they arose following duplication of a single gene at an early point in evolution. Their separation onto two different chromosomes reflects the long time that has elapsed in the development of oxygen-binding proteins (Hardison, 1998).

Haemoglobinopathies.

Co-ordinate output from the α and β loci during all phases of ontogeny ensure a stoichiometric balance between the α-like and β-like polypeptide chains in the erythrocytes. Haemoglobinopathies are hereditary blood disorders that influence globin production and function. Most of these syndromes are the result of deletions or mutations of the structural gene sequences or of important gene regulatory elements. The study of these genetic diseases has provided valuable insights in some general mechanisms that control gene expression within the context of a specific tissue type and also during development. Thalassemia, Sickle cell anaemia and HPFH are the most common traits and will be discussed.

Thalassemia.

In 1925 Thomas Cooley and Pearl Lee, first described a severe form of anaemia associated with splenomegaly and characteristic bone changes. George Whipple later named this disorder Thalassemia from θαλσσα, the sea, by because the first cases on which he carried out autopsies were all of Mediterranean background. Thalassemias refer to a group of inherited haemoglobin disorders all characterised by reduced synthesis of either α or β-like globin chains. In the case of β-thalassemia this leads to an imbalanced synthesis of globin chains which is the determinant of the severity of the disease (Orkin, 1986; Weatherall and Clegg, 1981). The mutation causing β-thalassemia result in a deficit of β-globin production which ranges from minimal (mild β^{+}-thalassemia) to a complete absence (β^{0}-thalassemia). Heterozygotes for β-thalassemia are clinically asymptotic with minor haematological abnormalities, whereas homozygotes or compound heterozygotes for β-thalassemia have severe disease. The affected individuals are transfusion dependent which is required to prevent the severe skeletal abnormalities caused by bone marrow expansion. The anaemia results not only from the lack of β-chains but also from the surplus of α-chains. The latter forms insoluble membrane-damaging precipitates that cause premature red cell destruction. Chelation therapies are necessary in combination to blood transfusions to avoid liver and

kidney damage caused by the systemic iron overload released from the haemoglobin prosthetic group.

Detailed molecular analysis of a number of deletion-type β-thalassemias has helped in the identification of important regulatory elements (Grosveld et al., 1987; Kioussis et al., 1983; Kulozik et al., 1991; Taramelli et al., 1986). However, most β-thalassemias are caused by a wide variety of point mutations that affect the production of β-chains. The most interesting once are mutations in the β gene promoter region. These attenuate transcriptional initiation (see section on the β-globin promoter). Other point mutations alter the sequence at an exon/intron junction and the AAUAAA cleavage signal at the mRNA 3' end. Consideration of the effects of these mutations, especially those involving gene spicing and transcription, has extended our understanding of how eukaryotic genes are constructed and expressed.

Sickle cell anaemia.

Sickle cell anaemia is probably the best known of the haemoglobinopathies. It is caused by a single amino acid change (glutamine to valine) in the human β-globin polypeptide (Ingram, 1956). Sickle cell anaemia is of historic importance because it marks the development of a new technique, protein fingerprinting. More importantly this was the first report that a disease is caused by a single amino acid change in a protein.

Sickle haemoglobin (HbS) differs physiologically from adult haemoglobin (HbA) primarily in that it polymerises when deoxygenated. HbS precipitates cause the erythrocytes to deform and assume the classical sickled shape. Anaemia results because of the lysis of the red blood cells that have lost their flexibility and get trapped in small capillaries. It is likely that this polymorphism has reached high frequencies in certain populations because of the beneficial effect conveyed to heterozygotes under the selective pressure of malaria, which is caused by the parasite *Plasmodium falciparum*. It has also been suggested that the invasion of sickle cells traits erythrocytes by *P. falciparum* merozoites may be retarded under conditions of low oxygen (Pasvol et al., 1978).

HPFH.

Hereditary persistence of foetal haemoglobin (HPFH) is a disorder characterised by the production of foetal haemoglobin (HbF) in adult erythrocytes. HPFH is found in combination with sickle cell anaemia or β-thalassemia. The presence of HbF in the adult erythrocyte has no harmful effects. Rather, the severity of the thalassemia is often alleviated in these patients because HbF can replace HbA and also reduce HbS polymerisation. A number of non-deletion HPFH is caused by point mutations in the promoters of the γ-globin promoters. These

mutations disrupt important elements within the γ promoters, suggesting that factors binding sites for repressors of foetal globin synthesis in the adult have been removed by the mutations or that new sites have been created. This argument will be expanded later in the context of the regulation of the γ-globin gene.

One class of deletions located in the 3' half of the locus are also involved in HPFH (Forget, 1998) for review). Among these, an interesting set of deletions appear to delineate a small region situated between the γ- and δ-globin gene that may be involved in silencing γ-globin. However, if this area is deleted from a YAC containing the complete β-globin locus, no effect is observed on the silencing of the γ-globin genes in transgenic mouse livers. In contrast, the analysis shows that this area is directly involved in the activation of β-globin (Calzolari et al., 1999). Therefore the HPFH observed in patients is not caused by a simple 'expression by lack of repression' mechanism. It is still unclear whether the persistence of foetal haemoglobin in the adult is due the juxtaposition of uncharacterised sequences brought close by the deletions.

Since the developmental fate of a cell is determined by its gene expression program, it is extremely important to understand how gene expression is regulated in eukaryotic cells. In the next paragraphs, we will therefore discuss eukaryotic transcriptional regulation in general, and globin gene regulation in particular.

Control of gene expression.

In eukaryotic cells, the ability to express biologically active proteins is regulated at several steps in the pathway leading from DNA to protein. The first level of control is chromatin structure. In the nucleus, DNA is compacted into chromatin with the aid of histone proteins. This physical structure affects the ability of transcriptional regulatory proteins (termed transcription factors) and RNA polymerases to gain access to specific genes and to activate transcription from them. For many genes control at the first step of expression, transcriptional regulation, is the most important mode of regulation. Regulation of transcription is accomplished through the combined action of transcription factors binding to cis-regulatory sequences other than the structural gene itself. These cis-regulatory elements will be discussed below.

Further levels of control are RNA processing, transport, translation and degradation. The primary RNA transcript is capped, polyadenylated and the introns are removed. The resulting mRNA is then transported to the cytoplasm where it is translated in a polypeptide chain and eventually degraded. Finally, the protein synthesised can be selectively activated or inactivated by covalent modifications such as phosphorylation, acetylation, enzymatic

cleavage or compartmentalisation. For this overview, we will now turn our attention to transcription initiation.

Transcription initiation.

Transcription in eukaryotes is carried out by three different polymerases. RNA polymerase I synthesises the ribosomal RNAs except for the 5S species. RNA polymerase II synthesises the mRNA precursors coding for proteins and small nuclear RNAs (snRNAs) involved in RNA splicing. RNA polymerase III synthesises the 5S and the tRNAs. Here, we will only consider transcription by RNA polymerase II (RNA-polII).

Initiation of transcription is thought of as a stepwise process mediated by RNA-polII and a complex array of general transcription factors (GTFs). This process requires the assembly at the promoter of a large multi-protein complex named Pre Initiation Complex (PIC). The PIC contains RNA-polII and the GTFs TFII A,-B,-D,-E,-F and –H. Virtually every promoter employs these components of the PIC, and all the subunits are conserved between man and yeast. The RNA-polII/GTFs multiprotein complex has an intrinsic ability for low level transcription from core promoters *in vitro*. The first step in the assembly in the initiation complex is the binding of TFIID to the TATA element. TFIID contains the TATA binding protein (TBP) and at least eight TBP-associated factors (TAFIIs). The presence of TAFIIs enables a response to activators. Activators bind to specific TAFIIs and this may facilitate recruitment of TFIID to promoters thus accelerating the first step in initiation complex formation. This initial complex acts as a binding site for TFIIB, which can recruit RNA polymerase II and TFIIF in the complex. Subsequently, TFIIE and TFIIH associate with the initiation complex. Once the complete complex is assembled, an ATP-dependent activation step is necessary for transcription elongation to occur. The reader is referred to several excellent review articles for more details (Green, 2000; Roeder, 1996a; Roeder, 1996b; Tjian, 1995; Tjian and Maniatis, 1994). In order to obtain high levels of transcription such as those required for the globin genes, this process needs to be activated by specific transcription factors. We will now turn our attention to the elements bound by these transcription factors.

Elements involved in transcriptional regulation.

Promoters.

Promoters are DNA sequences located immediately upstream to the RNA start site. Promoters are typically about 100 bp in length and are the minimal requirement for accurate and efficient initiation of transcription of a gene. A typical eukaryotic promoter consists of a series of conserved DNA motives to which RNA polymerases and accessory factors required for

transcription can bind. A large number of protein-encoding genes transcribed by RNA-polII have a short DNA sequence consisting of A and T nucleotides located 30 bp upstream of the transcriptional start site. This motif is designated the TATA box and it is often referred to as the core promoter. Two other core promoter elements have been recognized: the Initiator element (Inr) coinciding with the transcription start site, and the Downstream Promoter Element (DPE) located some 30 bp downstream of the transcription initiation site. These three elements can be found alone but also in any combination in eukaryotic promoters (Conaway and Conaway, 1991; Kadonaga, 1990; Smale, 1997; Tjian, 1996; Weis and Reinberg, 1992). It is estimated that some hundred proteins are present in the multi-protein complexes that interact with these elements. Genes without a TATA box often contain one or more copies of the GC box, that is a linear sequence rich in G and C nucleotides. Typically these are housekeeping genes, which are thought to be constitutively transcribed, and their transcripts have multiple 5' ends. The globin genes all contain canonical TATA boxes. The β-globin promoter also contains an Inr element; this has not been studied in detail for the other globin genes (Antoniou et al., 1995).

The TATA box functions to ensure that transcripts are initiated efficiently and accurately. In addition to the TATA box, one or more elements of 8 to 10 base pairs called upstream promoter elements (UPEs) have been identified in many promoters (Wasylyk, 1988). Examples of these elements are the CCAAT box located between –70 and –80 and the CACC boxes of the globin genes (Hardison et al., 1998; Riemer et al., 1998).

Mutagenesis studies have shown that the number and type of UPEs determine the strength of a promoter. An important observation is that UPEs can function regardless of the orientation with respect to the TATA box (Mantovani, 1998). However, insertions of odd multiples of half a DNA helix turn between the TATA box and the UPE seems to be detrimental to transcription more than the insertion of even multiples. This observation suggests that proteins bound to the UPE require a stereospecific alignment on the DNA helix to interact with the transcription initiation complex (Schule et al., 1988). UPEs have important functions in regulating the rate of transcription of a gene. UPEs contain binding sites for factors required for positive and negative regulation of gene transcription. These elements will be discussed in more detail in the context of the globin promoters

Enhancers.

Enhancers are elements that can strongly activate transcription of a linked gene at a relatively long distance from promoters. These elements activate transcription independently of their orientation and are able to function from a wide range of distances either upstream or

downstream of the RNA start site (Blackwood and Kadonaga, 1998). For example the enhancer of the Drosophila *cut* locus is 85 kb upstream of its promoter (Jack and DeLotto, 1995), whereas the murine immunoglobulin Hμ core enhancer lies within the second intron of its transcriptional unit (Banerji et al., 1983). On the other hand, the T cell receptor α-chain gene enhancer resides up to 96 kb downstream of the promoter (Winoto and Baltimore, 1989). The first enhancer element to be found was a 200 bp long segment of the simian virus 40 (SV40) (Banerji et al., 1981). Since then many cellular enhancers have been identified in several inducible and tissue specific genes. The immunoglobulin heavy chain gene enhancer was the first cellular enhancer to be identified (Banerji et al., 1983; Gillies et al., 1983). Enhancers are composed of a number of transcription factors binding motifs. These sites are bound by tissue-specific and ubiquitous factors. It is striking that several protein-binding sites that are known to be important for the function of enhancers are also found in promoters. IgH genes have provided some suggestion that interactions between particular enhancer and promoter factors may be important for the regulated expression of genes. In the human globin loci, enhancers have been described 3' of the Aγ and the β-globin genes (Antoniou et al., 1988; Behringer et al., 1987; Bodine and Ley, 1987; Kollias et al., 1987; Trudel and Costantini, 1987).

The functioning of enhancers is still a matter of debate. Several factors may be involved in the activation of transcription by enhancers. The current data suggests that first sequence-specific DNA-binding proteins interact directly with sequences in the enhancer. Then, numerous co-activators interact with DNA bound factors. One of the most captivating questions is how do enhancer–binding proteins and their associate co-activators establish a productive interaction with their cognate promoter. Direct contact between the enhancer and the promoter, with the intervening DNA looping out, is the most attractive hypothesis (Ptashne and Gann, 1997). Support for the looping hypothesis comes from examination of the regulation of transcription of *E. coli* (Gralla, 1991). An important limitation to the hypothesis that similar events occur in eukaryotes is that all of the *E. coli* loops are quite small (< 500 bp). In eukaryotes, enhancers can act over kilobases of DNA. It is possible that the repositioning of eukaryotic enhancers to greater distances is because the intervening DNA can be folded into nucleosomes and the chromatin fibre. This model has not been rigorously tested, but it was shown *in vitro* by electron microscopy that two Sp1 binding sites placed 1.8 kb apart directly interact in the presence of the eukaryotic transcriptional regulator Sp1 (Mastrangelo et al., 1991; Su et al., 1991). Such a model would be consistent with the observation that enhancers can stimulate transcription of a promoter when linked via a biotin-steptavidin bridge (Mueller-Storm et al., 1989).

Locus Control Regions.

Early experiments indicated that the inclusion of additional elements were required besides enhancer and promoter sequences to ensure correct levels or tissue specific expression in mammals *in vivo* (Chada et al., 1986; Chada et al., 1985; Magram et al., 1985). The first indication that such elements exist was obtained from the analysis of naturally occurring deletions in a number of thalassemia patients. These deletions removed the upstream part of the β-globin locus, leaving the gene intact but silent (Kioussis et al., 1983). A number of erythroid-specific DNaseI hypersensitive sites (HSs) were mapped within the deleted region, indicating putative regulatory elements (Tuan et al., 1985). A DNA fragment containing these sites used to generate transgenic mice demonstrated that a linked human β-globin gene showed high level expression, tissue-specific and transgene copy-number dependent expression (Grosveld et al., 1987). Since the β-globin LCR several other LCRs have been identified (Fraser and Grosveld, 1998).

Locus Control Regions share many of the properties of a typical enhancer such as augmenting transcription of a linked gene regardless of their distance and orientation relative to the transcription initiation site. However, the operational difference between these two elements is that LCRs are identified by *in vivo* expression analysis in transgenic animals. Thus, enhancers are characterised in transient transfection experiments and LCRs are defined as stable integrated elements able to drive high-level expression of a linked transgene in a tissue-specific manner irrespective of the site of integration in the host genome. Like enhancers, LCRs are modular in structure: multiple binding sites for sequence-specific transcriptional activators are functionally important for the correct operation of the element (Philipsen et al., 1993).

Detailed analysis of the β-globin LCR and the human CD2 LCR have shown that LCRs have the capability to establish an open chromatin configuration at the site of integration even when this is normally in the repressed state (Festenstein and Kioussis, 2000). However the mode of action of the LCRs in chromatin modification is still poorly understood. Several experiments have been performed to dissect the properties of LCRs that will be discussed later in more detail in the context of the β-globin LCR. We will now give an overview of the transcriptional regulators of globin gene expression.

Transcription factors involved in globin gene expression.

Temporal and tissue specific expression of the globin genes is influenced by the various regulatory elements located proximal to the genes and at different positions in the locus. These elements are composed of discrete DNA sequence motifs, which constitute binding

sites for sequence specific DNA binding proteins. These transcription factors in turn interact with each other and the RNA-polII/GTF multi-protein complex described above, in order to modulate transcription.

Analogous to what has been observed in other tissues, the factors that regulate globin genes are themselves both tissue-specific and ubiquitously expressed. The identification of such transcription factors has been difficult however, principally because they are expressed in very small amounts. Nonetheless, a limited number of these factors have been isolated from erythroid cells following the refinement of sequence-specific DNA affinity chromatographic techniques and molecular cloning approaches.

At present, only a small number of transcription factors including GATA-1 and EKLF have been studied in detail for their role in the transcription of the globin genes. In addition, a number of factors have been shown to affect erythropoiesis through their importance for hematopoietic differentiation in general, such as scl/tal-1, GATA-2 and PU.1. In the next paragraphs we will discuss some of the best studied examples of factors that are directly involved in erythropoiesis.

GATA-1.

GATA-1 was the first cloned member of a family of six DNA binding proteins that recognise a central GATA consensus motif (Weiss and Orkin, 1995a). GATA binding sites are present in all the regulatory elements in the globin genes. However, GATA sites are also found in many other genes. Originally GATA-1 was thought to be erythroid specific, but it was later shown to be present in uncommitted haematopoietic precursor cells, megakaryocytes, eosinophils, mast cells and the Sertoli cells in the testis (Ito et al., 1993; Martin et al., 1990; Romeo et al., 1990). A GATA motif is often accompanied by a G-rich motif in erythroid specific regulatory sequences, suggesting that GATA-1 binds together with members of the Sp/XKLF family of zinc finger proteins, such as EKLF, FKLF and Sp1, to specify an erythroid combination (Philipsen et al., 1990). The family of GATA proteins is characterised by the presence two "zinc finger" domains that interact with the major groove of the DNA helix. The carboxy-terminal finger is required for the binding to the GATA motif. The N-terminal finger can also bind DNA (Omichinski et al., 1993; Trainor et al., 1996). Besides stabilising the binding of GATA-1 to DNA, it functions to interact with zinc fingers present in its cofactor Friend of GATA-1 (FOG-1) and the Sp/XKLF factors (Merika and Orkin, 1995; Tsang et al., 1997). GATA-1 has also been reported to interact with itself (Crossley *et al.*, 1995; Yang and Evans, 1995), and is found in a large complex with the transcription factors E47, scl/tal-1, Lmo-2 and Ldb1. This complex binds to a GATA motif juxtaposed to an E-box

(GANNTC), and although this observation is interesting in light of the strong hematopoietic phenotypes of the scl/tal1 and Lmo-2 knockouts, the significance of these observations for erythropoiesis requires further study (Wadman et al., 1997; Warren et al., 1994; Yamada et al., 1998). Finally, GATA-1 is associated with the general transcriptional co-activator CBP that has acetyltransferase activity (Blobel et al., 1998). Acetylation of GATA-1 correlates with increased DNA binding affinity, presenting another mode of regulating GATA-1 activity (Boyes et al., 1998; Hung et al., 1999). It appears that once bound via the C-terminal finger to the DNA, GATA-1 provides a platform for interacting proteins to bind and exert their action on the transcriptional machinery. Of the other five GATA factors, only GATA-2 and GATA-3 are expressed in the haematopoietic system in addition to many other tissues (Weiss and Orkin, 1995a).

Inactivation of GATA-1 leads to a fatal defect in erythropoiesis, and GATA-1 deficient ES cells do not contribute to the erythroid lineage in chimeric animals (Pevny et al., 1995; Pevny et al., 1991). The GATA-1 deficient cells are blocked at a stage precedent to the proerythroblast and go into programmed cell death (apoptosis) (Weiss and Orkin, 1995b). Thus, it can be concluded that GATA-1 is not required for the specification of the erythroid lineage, but that it is essential for terminal erythroid differentiation. Inactivation of GATA-2 also leads to early embryonic death. Mice lacking GATA-2 survive to E10-11 but succumb to anemia due to a severe decrease of early embryonic red blood cells (Tsai et al., 1994; Tsai and Orkin, 1997). Experiments *in vitro* indicate that GATA-2 serves specific functions in the proliferation and survival of progenitor cells but not in the terminal differentiation of erythroid cells (Heyworth et al., 1999). This indicates that GATA-2 acts earlier in the hematopoietic hierarchy and is not directly involved in the expression of the globin genes.

Studies on GATA-3 have demonstrated that it is clearly important for the development of T-cells (Ting et al., 1996). GATA-3 is unlikely to be directly involved in erythropoiesis, since its expression is undetectable in this lineage and the GATA-3 knockout mice do not display overt anemia (Pandolfi et al., 1995).

FOG-1.

FOG-1 (Friend of GATA-1) was identified in a two-yeast hybrid screen as a protein that interacts with the N-terminal finger of GATA-1. FOG-1 contains nine zinc fingers of which finger 6 specifically interact with the N-terminal finger of GATA-1 both *in vitro* and *in vivo* (Crispino et al., 1999; Tsang et al., 1997). It is co-expressed with GATA-1 during development and is present in erythroid and megakaryocytic cells. Inactivation of FOG-1 in mice leads to a fatal anemia in the fetal liver. FOG-1 deficient red blood cells show a block in

erythroid maturation that is very similar to the phenotype of GATA-1 deficient erythroid cells. Unlike in GATA-1 *null* cells, megakaryocytes also fail to develop, indicating that FOG-1 is required earlier in megakaryocytic differentiation than GATA-1 (Tsang et al., 1998). Recent evidence suggests that FOG-1 acts as a repressor of GATA-1 mediated transcription (Deconinck et al., 2000)

Nuclear factor-erythroid 2 (NF-E2).

NF-E2 was first identified as a DNA binding activity associated to an AP-1 motif in the promoter of the porphobilinogen deaminase gene (Mignotte et al., 1989). It was later shown to bind the AP-1 consensus sequence in 5'HS2 of the β-globin LCR, which is largely responsible for the activity of this element (Talbot et al., 1990). The gene encoding NF-E2 was cloned after microsequencing of the affinity-purified protein and shown to be expressed in several haematopoietic lineages including erythroid, mast cells and megakaryocytes (Andrews et al., 1993). NF-E2 is a heterodimer of a 45-kD hematopoietic-specific protein (p45) and a widely expressed p18 subunit, both of which belong to the basic-region- leucine zipper (bZip) family of nuclear proteins. The p18 subunit, also called MafK, belongs to the family of Maf proteins, which do not have a canonical transcriptional activation domain but are essential for binding site recognition in conjunction with a heterodimeric partner (Igarashi et al., 1994). Stimulation requires the p45 transactivation domain and is mediated by interaction with the co-activator CBP (Cheng et al., 1997).

Mice lacking p45 NF-E2 develop apparently normal *in utero*, but the vast majority of pups die within the first week of *post partum*. Erythroid cell development is only subtly affected. The mice only show a mild hypochromic anemia, the developmental switch of the genes in the β-globin locus is normal and globin gene synthesis is normally balanced. Death results from haemorrhage caused by the absence of circulating platelets (Shivdasani and Orkin, 1995; Shivdasani et al., 1995).

It is unclear whether the absence of a more dramatic erythroid phenotype reflects redundancy at the level of transcription factors. A number of p45 NF-E2-related factors are known to be widely expressed and will be discussed later. Mice lacking the p18 subunit show no erythroid phenotype and again this may be due to compensation by one of the other Maf homologues (Kotkow and Orkin, 1996). The subtle erythroid defect observed in Maf-G/MafK compound homozygotes is in agreement with this notion (Onodera et al., 2000).

Nrf-1, Nrf-2, Nrf-3, Bach 1, Bach 2.

Five other factors that bind to the NF-E2/AP-1 of HS2 have been identified. These are Nrf-1, Nrf-2, Nrf-3, Bach-1 and Bach-2. These proteins together with NF-E2 belong to the family of Cap'n'Collar (CNC)-type of bZip family of proteins. Some of these proteins are limited in expression to a number of tissue types but none are erythroid-specific. These proteins are also capable of forming heterodimers with Maf family members (Itoh et al., 1995; Kobayashi et al., 1999; Oyake et al., 1996; Toki et al., 1997).

Inactivation of Nrf-1 results in an early embryonic lethal phenotype. Development of the embryo is arrested to a stage antecedent to the formation of the primitive streak and there is no detectable mesoderm. However Nrf-1 deficient ES cells contribute to mesodermally derived tissues including blood. Erythroid cells have normal levels of hemoglobin, suggesting that this factor is not directly involved in globin gene expression (Farmer et al., 1997). The phenotype of the Nrf-1 knockout is still under debate since Nrf-1 knockout mice generated by another laboratory succumb to fetal anemia (Chan et al., 1998). However, the erythroid defect is non-cell-autonomous in both cases, negating a direct role of Nrf-1 in erythroid cells.

Inactivation of Nrf-2 does not result in any clear phenotype, suggesting also that this factor is not essential for globin gene expression (Chan et al., 1996). Compound Nrf-2/NF-E2 knockouts do not show a phenotype beyond that of the NF-E2 knockout alone (Kuroha et al., 1998; Martin et al., 1998).

Bach1 and Bach2 were identified in a two-hybrid screen as proteins interacting with MafK. Besides the bZip domain, these two proteins contain a BTB domain. In *Drosophila melanogaster*, BTB proteins are involved in a variety of processes including chromatin remodelling (Oyake et al., 1996). In this regard the Bach proteins could substitute for NF-E2 at the 5'HS2 of the globin LCR in the NF-E2 knockout. Interestingly, it has recently been shown that the Bach proteins are capable of forming DNA loops between binding sites *in vitro*, suggesting they might be involved in loop formation between the LCR and the globin promoters (Yoshida et al., 1999).

EKLF.

EKLF (Erythroid Krüppel like Factor) is a member of the Sp/XKLF family of zinc finger proteins (Miller and Bieker, 1993; Philipsen and Suske, 1999). Its three zinc fingers bind specifically the sequence CCACACCT which is found in the promoter of the β-globin gene, and related sequences in the promoters of the other globin genes and 5'HS3 of the LCR (Feng et al., 1994; Gillemans et al., 1998). Like other members of the Sp/XKLF family, EKLF acts as a transcriptional activator in reporter assays through a proline-rich domain that can be

subdivided into stimulatory and inhibitory subdomains (Chen and Bieker, 1996). EKLF is already expressed in the primitive erythroid cells, although the adult β-globin gene is its only known target at present (Southwood et al., 1996). Ablation of EKLF in the mouse germ line results in death of the fetuses due to severe β-thalassemia. The adult murine β^{maj} and β^{min} globin expression is markedly reduced, whilst α-globin and embryonic β-like globin gene expression is not affected, showing that EKLF is essential for the activation of the β-globin promoter (Nuez et al., 1995; Perkins et al., 1995). Direct activation of the adult β-globin gene *in vitro* is achieved by interaction with ERC-1, a SWI/SNF-related chromatin remodelling complex. In the presence of both ERC-1 and EKLF a DNaseI hypersensitive site is formed on the β-globin promoter *in vitro* and transcription ensues (Armstrong et al., 1998). It has also been shown that DNaseI hypersensitive site formation of the 5'HS3 of the LCR is contributed by the binding of EKLF (Gillemans et al., 1998). The dynamics of this process is still obscure but the current data demonstrate that EKLF is a substrate of p300 and CBP and that this association leads to protein acetylation *in vitro* and increased activity *in vivo* (Zhang and Bieker, 1998). Thus, the decompacted status of chromatin at the β-globin promoter and at 5'HS3 could result from EKLF binding followed by recruitment of a chromatin remodelling complex such as ERC-1, and acetyltransferases such as CBP and p300. This in turn decondenses the area and renders it transcriptionally competent. EKLF is also phosphorylated at serine and threonine residues within its transactivation region. Phosphorylation at these sites is associated with transactivation activity (Ouyang et al., 1998). It has been postulated that the phosphorylation status of EKLF may be different in cells that contain EKLF but do not express β-globin, such as the primitive erythroblasts. Phosphorylation would allow extracellular effectors to modify the activity of EKLF at appropriate times during erythroid differentiation. Interestingly, overexpression of EKLF in transgenic mice results in a reduction of the circulating platelets, suggesting that EKLF may have a role in determining the balance between megakaryocytic and erythroid lineages, perhaps by activating erythroid specific genes (Tewari et al., 1998).

Structure and regulation of the human globin genes.

All the human globin genes have a similar structural organisation. They have three coding exons separated by two introns, probably reflecting their common origin from a single ancestral gene. The exons code for 141 and 146 amino acids in the α- and β-like globin chains respectively (Hardison, 1998). The length of the introns varies between the α- and β-like globin genes. At least in the case of the β-globin gene, intron 2 appears to be important for

polyadenylation , release of the transcript and its transport to the cytoplasm (Custodio et al., 1999). Conserved sequences important for tissue and developmental expression of the individual globin genes are both in close proximity to and more distant from the globin genes (Fig. 1 and 2). These include sequences in the promoter and both 5' and 3' of the mRNA coding sequences. These regulatory sequences are bound by a combination of both tissue restricted and ubiquitous transcription factors (Fig. 3). The transcriptional regulation of the individual globin genes will be discussed below.

The ε-globin gene.

The ε-globin gene has a TATA box situated around 30 bp upstream of the transcription initiation site, a CCAAT box at –82 and a CCAC box at approximately – 100. These elements are required for expression in conjunction with the GATA sites in the promoter. It is not known which of these putative binding sites are the most important sites *in vivo*. For example, inactivation of the transcription factor EKLF, which is able to bind the CCAC box motif, does not result in defective expression of the ε-globin gene (Wijgerde et al., 1996). The ε-globin gene requires the LCR for expression, but is silenced autonomously i.e. it is still silenced in transgenic experiments if there are no other globin genes present in the construct. A silencer element has been identified between the – 177 and – 329 bp upstream of the start site (Raich et al., 1992). In transgenic animals, silencing appears to be mediated by YY1 (a ubiquitously expressed zinc finger protein) and GATA-1 binding sites (Raich et al., 1995). Deletion of the silencer results in continuous expression of the embryonic gene into adulthood, although at very modest levels (Raich et al., 1992). However, when the silencer is deleted from a YAC containing the entire β-globin locus, the ε-gene expression is suppressed rather than enhanced, indicating that this region is also important for the activation of the gene (Liu et al., 1997). Interestingly, deletion of this ε-activator/suppresser region also results in a down regulation of γ-globin transcription, perhaps by interfering with the LCR/γ-globin interaction.

The γ-globin genes.

The γ-globin genes are principally expressed during the fetal period of development. Normally, their expression diminishes to low levels in adulthood. However, in the condition known as Hereditary Persistence of Fetal Hemoglobin (HPFH), γ-globin chains are synthesised at elevated levels in adult erythrocytes. This observation is clinically relevant because reactivation of the fetal genes would be therapeutically important in the alleviation of

β-thalassemia and sickle cell anemia. Thus, the regulation of γ-globin expression has been subjected to intense scrutiny.

In the presence of the LCR, a human γ-globin transgene is expressed in the embryonic yolk sac and the fetal liver of mice and is switched off autonomously in the late fetal liver and during adulthood (Dillon and Grosveld, 1991). This indicates that expression of the γ-globin gene, like the ε-globin gene, is blocked by the action of stage-specific negative regulators. The factors responsible for γ-globin silencing have not yet been identified, but mutations associated with non-deletion HPFHs suggest that sequences around the promoters are likely to be involved. These mutations occur in transcription factors binding sites, either creating new factor binding sites or destroying existing ones.

Both γ-globin promoters contain a canonical TATA box, two tandemly duplicated CCAAT boxes 27 bp apart and one CACC box. The proximal promoter region also contains a G-rich sequence, which has been shown to bind two factors, Sp1 and a binding activity termed stage selector protein (SSP) (Jane et al., 1992). Binding of Sp1 is probably not functionally significant. The SSP complex does appear to be important during the switch from γ- to β-globin expression (Jane et al., 1995). Recent experiments in transgenic mice show that mutations in the stage selector element (SSE) which prevent *in vitro* binding of SSP result in the down-regulation of the γ-globin gene during the switching period (Ristaldi et al., 2001). SSP has been partially purified and appears to be a heterodimer of the ubiquitously expressed CP2 protein and an unknown erythroid factor (Jane et al., 1995).

Point mutations in the distal CCAAT box at positions -114 and -117 result in persistent expression of the γ-globin gene in the adult stage (Berry et al., 1992; Wood, 1993). The -117 G to A mutation, associated with Greek HPFH, has been reported to cause decreased *in vitro* binding of the erythroid-specific factor NFE-3 and of GATA-1 (Berry et al., 1992; Mantovani et al., 1989). However, mutations that specifically abolish binding of both factors to the CCAAT box *in vitro*, fail to give a HPFH phenotype in transgenic mice (Ronchi et al., 1996). One clear conclusion from this work was that the high level of γ-globin expression consequent to the -117 mutation is dependent on the presence of a functional proximal CCAAT box.

Other factors bind the CCAAT boxes. CP1 binds both CCAAT boxes and is thought to act as a positive transcriptional activator. CP1 is an heteromeric protein formed by three subunits : NFY-A, -B and –C (Mantovani, 1998). The CCAAT displacement protein (CDP) binds over a broad region of the promoter extending at least from -202 to -50. CDP acts as a transcriptional repressor (Superti-Furga et al., 1988). Another factor that was shown to interact with the CCAAT boxes of the β-like globin genes was originally designated NFE-6 (Berry et al.,

1992). NFE-6 contains the γ-member of the C/EBP family of proteins (Wall et al., 1996). Overexpression studies in transgenic mice show that this factor acts as a positive regulator of the γ-globin gene during the switching period, but overexpression does not alleviate γ-globin silencing in adult mice (Zafarana et al., 2000). Finally, recent data suggest that members of the nuclear hormone receptor family are involved in the regulation of the stage-specific activity of the γ-globin gene promoters through binding sites that overlap with the CCAAT box motifs (Filipe et al., 1999). Members of this family are known to function both as repressors and activators of transcription, and are therefore very interesting candidate switching factors.

The γ-globin CCAC box binds at least five transcription factors present in erythroid cells: Sp1, Sp3, BKLF, EKLF and the recently described FKLF (Philipsen and Suske, 1999). Knockout studies in mice suggest that Sp1, Sp3, BKLF and EKLF do not have a major role in γ-globin activation *in vivo* (Bouwman et al., 2000; Marin et al., 1997; Perkins et al., 1997; Wijgerde et al., 1996). Thus, the function of the γ-globin CACC box remains elusive. *In vitro* transactivation experiments indicate that FKLF is a potent activator of γ-globin expression (Asano et al., 1999). It will be of great interest to determine the effect of the FKLF knockout on γ-globin expression in mice.

Several other binding sites in the upstream part of the promoter are further characterised by other HPFH mutations. The most intriguing mutations are centred on the -200 region. Besides creating novel binding sites, mutations in this region could lead to structural alterations of the DNA. The -200 region is capable of forming a triple stranded structure, leaving an exposed single stranded region in the DNA (Ulrich et al., 1992). A suppresser binding to this structure would be displaced by the transcription factor that binds to the novel sequence created by the HPFH mutation. Interestingly, insertion of an additional C at -200 does not change the profile of factor binding *in vitro* but destabilizes triple strand formation (Bacolla et al., 1995). The effect of this mutation on γ-globin expression is being analysed in transgenic mice (Imam A. pers. comm.). Lastly, an enhancer element is present 750 bp downstream of the Aγ gene. The role of this element is not clear. Recent experiments in which this element was deleted from the locus indicate that it has no role in either silencing or enhancing γ-globin expression in transgenic mice (Liu et al., 1998).

The δ-globin gene.

The δ-globin gene contributes to only 3% of the adult β-like globin chains. This low-level expression is attributed to a mutated CCAAT box and an imperfect CACC box in the promoter of the gene. Replacement of the -90 region with the proximal CACC box of the β-globin gene increases δ-globin expression approximately 10-fold in cultured cells (Donze et al., 1996; Ristaldi et al., 1999). This increase correlates with binding of EKLF to this CACC box. A similar increase is observed with the insertion of a functional CCAAT box (Tang et al., 1997; Tang and Rodgers, 1998).

The β-globin gene.

The β-globin promoter contains a TATA box at – 30, a G-rich sequence at –50, an imperfect palindromic CCAAT box at – 75 and two CACC boxes at – 75 and – 110. In addition to the promoter, two enhancer elements have been mapped. The first is located in the second intron and the second a few hundred bases downstream from the polyadenylation site (Antoniou et al., 1988). The β-globin promoter is sufficient to drive expression of a linked gene in transgenic mice, but the expression levels are generally low at 1% of the endogenous mouse β-globins (Chada et al., 1986; Chada et al., 1985; Magram et al., 1985). High level expression is only achieved when these sequences are linked to the LCR (Grosveld et al., 1987). Combined deletion of the CCAAT and the CACC box reduces LCR-mediated expression dramatically in MEL cells (Antoniou and Grosveld, 1990). The TATA box region binds the general transcription factor complex TFIID, and appears to act in concert with an Inr element overlying the transcription initiation site. Using the MEL cell system, it has been shown that mutations drastically affecting the binding of TBP to the TATA box result in decreased transcription levels (Antoniou et al., 1995). Mutations found in the CACC box and TATA motifs of some β-thalassemic patients have also indicated an important function for these elements in proper regulation of β-globin gene expression (Hardison et al., 1998). There are no reports of patients with CCAAT box mutations yet, and hence the functional significance of this element remains elusive. The effects of such mutations in the context of the complete human β-globin locus are currently being explored in transgenic mice.

The G-rich sequence called the β-DRE repeats consists of an imperfect repeat of a 10 bp motif (Stuve and Myers, 1990). Its position is analogous to the SSE element in the γ-globin promoter. The β-DRE is bound by βDREf, a factor reported to be able to bend DNA (Dyer et al., 1998). The functional significance of the β-DRE repeats remains unknown.

The CCAAT box region binds several factors: CP1, GATA-1 and NFE-6. As for the γ-globin gene, the functional relevance of CP1 has not been established yet. GATA-1 binding to the β-globin CCAAT box is very unstable (half life < 1 min) and is probably not very important at this position (Delvoye et al., 1993). NFE-6 has been identified as C/EBPγ (Wall et al., 1996). Despite the observation that in C88 mouse erythroleukemia (MEL) cells the NFE-6 binding activity on the CCAAT box diminishes on induction of globin synthesis, NFE-6 does not appear to be a classical repressor. Raising the levels of C/EBPγ in these cells does not result in repression of the mouse globin genes. However, C/EBPγ may have a role *in vivo* at the time of the switch in expression from the fetal to the adult genes. Overexpression of this factor in transgenic mice results in a shift in the ratio of γ- to β-globin expression, suggesting that C/EBPγ downregulates the adult β-globin indirectly by favouring expression of the γ-globin gene at the time of switching, when both genes are co-expressed in the mouse fetal liver erythrocytes (Zafarana et al., 2000).

The functional factor *in vivo* of the CACC box is EKLF. EKLF is essential for the expression of the β-globin gene. During development, EKLF levels play a role in the dynamic switch from γ- to β-globin expression: higher EKLF levels favour β-globin expression (Tewari et al., 1998; Wijgerde et al., 1996). This identifies EKLF as the first factor that is directly involved in γ- to β-globin switching. Ablation of EKLF in mice leads to fetal anemia and absence of expression of the adult β-globin genes (Nuez et al., 1995; Perkins et al., 1995). Even though EKLF is expressed during the embryonic stages of development (Southwood et al., 1996), the expression and timing of embryonic and fetal β-like globin genes is not affected in EKLF knockouts (Wijgerde et al., 1996).

The function of the two enhancers has been tested. The activity of the intragenic enhancer is associated with polyadenylation and release of the nascent transcript from the template (Custodio et al., 1999). This interesting observation adds a level of complexity to the regulation of the globin genes. Transcription is a tightly regulated and highly interactive process. These intragenic sequences may enhance transcription by increasing the rate of processing of the primary RNA transcript still closely associated with the DNA template. It will be of interest to determine the role of these sequences in the context of the entire locus.

The second enhancer contains four GATA-1 binding sites and has been shown to stimulate transcription of a linked promoter in transfection experiments (Antoniou et al., 1988). The 3' enhancer functions as an adult stage specific activator in transgenic mice when the transgene is not linked to the LCR (Behringer et al., 1987). When the enhancer is deleted from a YAC containing the entire β-globin locus, the expression of the β-globin gene is severely affected

in transgenic mice (Liu et al., 1997). Thus, the 3' enhancer has an important function in the expression of the β-globin gene.

The aforementioned individual regulatory sequences have important functions, and it has become apparent that the correct developmental expression of the β-like globin genes is only achieved in the context of the whole locus. Competition between the genes for the activating function of the LCR, the relative distance of the genes from the LCR and the order and orientation of the genes with respect to the LCR are important parameters in globin gene regulation (Dillon et al., 1997; Hanscombe et al., 1991). Recent progress made in this area of research is discussed in the next section.

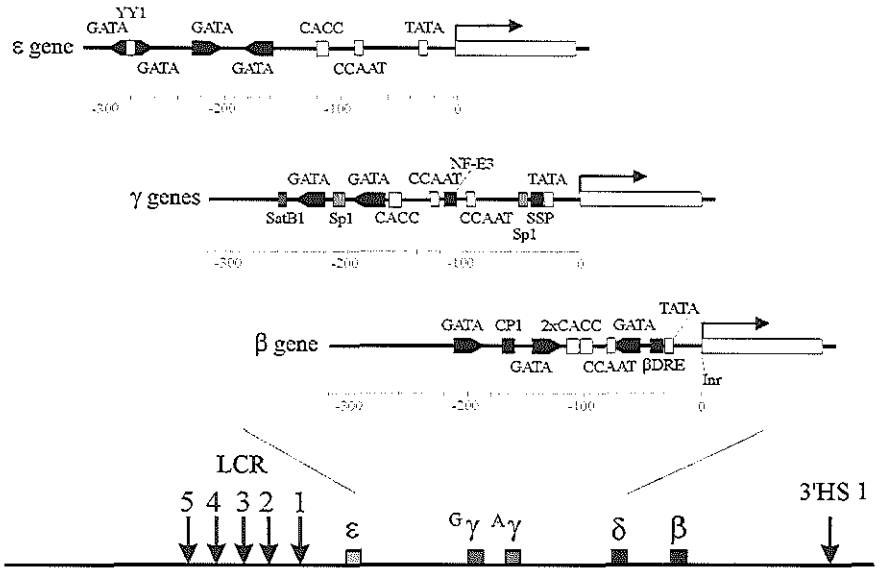

Fig. 2. The promoter areas of the human β-like globin genes. The motifs reported in the promoters of the different β-like globin genes are shown. See text for details on the proteins binding to these elements. The scale is in base pairs with base 0 representing the transcription start sites of the genes.

The β-globin Locus Control Region.

The discovery of the β-globin Locus Control Region (LCR) followed the observation that in a human γδβ-thalassemia one of the patient's chromosome 11 presented a 100 kb deletion that had eliminated the entire upstream region of the locus, leaving the β-globin gene intact but silent. Cloning and expression of the β-globin gene from the mutant allele showed that it was completely normal (Kioussis et al., 1983). The other allele was expressed in the patient,

indicating that the silencing of the mutant allele was not due to a lack of transactivating factors, but rather that a region important in the regulation of the entire β-globin locus had been deleted. Within the deleted area, 5 to 25 kb 5' of the ε-globin gene, a series of five developmentally stable, erythroid specific DNaseI hypersensitive sites (HSs) were found (Tuan et al., 1985). These are now termed 5'HS1-5. Linkage of this region to a cloned β-globin gene resulted in erythroid-specific, high level expression of the gene in transgenic mice. This expression was independent of the integration site in the host genome and was dependent on the copy number of the transgene. Thus, all the transgenic animals expressed the transgene at predictable levels, and the expression levels per copy of the transgene were similar to those of the endogenous mouse β-globin genes (Grosveld et al., 1987). The area was therefore termed the Locus Control Region (LCR).

Subsequent work showed that the activity was localised within approximately 200-300 bp sequences marked by the HSs (Fig. 3). Each of these 'core' sequences bind a limited number of erythroid-specific and ubiquitously expressed factors, including GATA factors, Sp/XKLF factors, NF-E2, AP-1 and YY1 (Philipsen et al., 1990; Pruzina et al., 1991; Talbot et al., 1990).

Only 5'HS2 appears to act as a strong classical enhancer in transient transfection assays (Ney et al., 1990; Tuan et al., 1989). The other HSs have very low classical enhancer activity when tested in transient expression experiments (Tuan et al., 1989). The enhancer activity of 5'HS2 is entirely due to the tandem NF-E2/AP1 binding site present in this LCR fragment. Mutations that eliminate NFE-2 but preserve AP-1 binding markedly reduce HS2 activity but do not abolish it (Ney et al., 1990). Similarly, when globin transcription is analysed in erythroid cells of *null* mutant mice for NFE-2, no dramatic reduction in globin synthesis is observed (Shivdasani and Orkin, 1995). This suggests that there is redundancy at these sites or the factors that bind these sites *in vivo* have still to be identified.

5'HS3 is the most powerful individual element conferring an active chromatin structure to a linked transgene when integrated in the genome. For instance, 5'HS3 can activate single copy transgenes in mice, while 5'HS2 requires multiple copies to activate gene expression (Ellis et al., 1996). Since gene therapy approaches that rely on the use of viral vectors usually result in single copy integrations, it is important that gene therapy constructs are functionally analyzed in single copy transgenic assays. It must be emphasized here that the presence of the complete LCR is required for full level β-globin gene expression in all the erythroid cells (Milot et al., 1996). Thus, gene therapy vectors aimed at complete restoration of β-globin expression should not rely on the use of partial LCRs.

Two GATA-1 motifs and flanking CACC motifs are required for LCR activity of a minimal 5'HS3 fragment in transgenic mice (Philipsen et al., 1993). Further experiments with 5'HS3 in transgenic mice have shown that EKLF is the active factor at the CACC motifs *in vivo*, and that binding of EKLF induces a change in chromatin structure (Gillemans et al., 1998). This correlates well with the fact that EKLF has been reported to associate with E-RC1, a member of the SWI/SNF family of proteins, which modify chromatin structure (Armstrong et al., 1998). The classical transactivator protein Sp1 can also bind to these CACC motifs *in vitro*, but it does not function as an activator of the minimal 5'HS3 (Gillemans et al., 1998). This underscores the importance of demonstrating specificity *in vivo*. As yet, such specificity can not be achieved with biochemical experiments, and this is an important goal for the next decade because we need to understand globin gene activation at the molecular level if we want to have clear targets for therapeutic intervention.

Other widely expressed proteins, such as USF and YY1, have been shown to interact with the LCR, but the functional relevance of such interactions has not been defined (Ellis et al., 1993).

In transgenic experiments, globin genes without an LCR are expressed only in a proportion of mice and at low levels that do not correlate with copy number (Chada et al., 1986; Chada et al., 1985; Kollias et al., 1987; Magram et al., 1985; Trudel and Costantini, 1987). This type of expression reflects the influence of the chromatin environment at the site of integration of the transgene in the host genome. This effect is generally known as position effect. Except when the globin LCR is integrated in the X chromosome where it is subject to X-inactivation (Whyatt et al., 2000), such position effects are overcome by the full LCR in a dominant fashion (Grosveld et al., 1987). However, deletion of single HSs of the LCR in transgenic mice carrying the complete β-globin locus results in loss of position independence and lowered expression. Loss of position independence upon deletion of the HS sites of the LCR has been shown using transgenes of different sizes based on cosmid and YAC technology (Bungert et al., 1995; Milot et al., 1996; Peterson et al., 1996). This makes it very difficult to compare the data in the literature, but the data collectively suggest that all the HSs are required for proper LCR function and that the individual HSs act in concert as a holocomplex in order to activate the globin genes (Ellis et al., 1996; Hanscombe et al., 1991).

The function of the LCR has also been studied within the intact murine and human β-globin loci by transferring the manipulated chromosomes to the appropriate cell types. These experiments offer the advantage over the use of transgenic animals that the effects observed are studied on the locus in its native chromosomal context. Deletion of the murine LCR results in a marked reduction of transcription of the murine β-globin genes, whereas deletion

of the human LCR results in the complete abrogation of β-globin gene expression (Epner et al., 1998; Reik et al., 1998). This is in agreement with the data obtained from patients with LCR deletions.

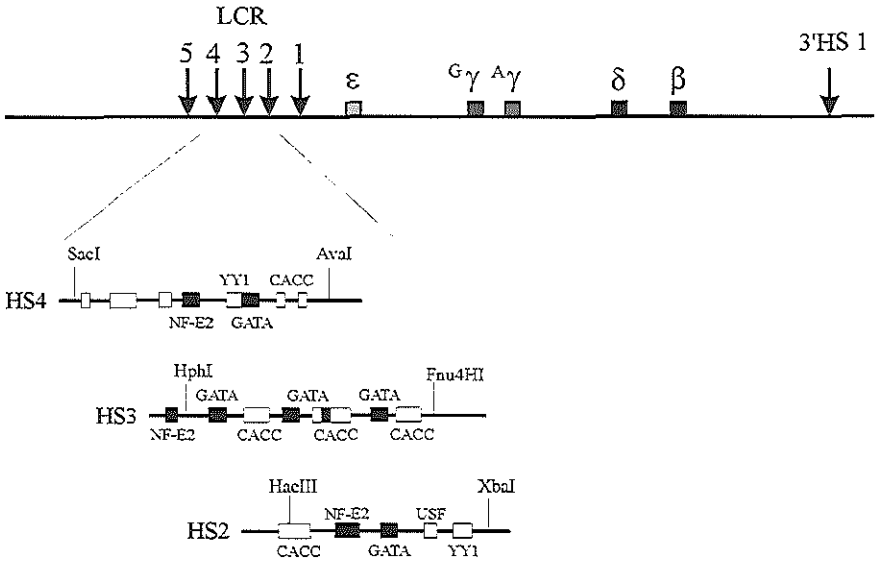

Fig. 3. Trans-acting factor binding sites in the LCR of the human β-like globin locus. The motifs reported in DNaseI hypersensitive sites 2,3 and 4 of the LCR are shown; the factors binding to these boxes are discussed in the text. Restriction sites used to define the functional elements of the LCR in transgenic mice are indicated.

Domain boundaries.

The development of very complex phenotypes observed amongst vertebrates has been generally accomplished by the expansion of their genome. Intuitively, one advantage of a large genome resides on its greater degree of informational capacity. However such an expanded genome presents a vertebrate with dilemmas not faced by bacteria or lower eukaryotes. Besides the structural puzzle of packaging this large amount of DNA into the nucleus, the presence of such a quantity of genetic information confronts vertebrates with serious strategical problems of information recognition and processing. With increasing genome size, there can be an increased probability that illicit transcription may accidentally activate areas of the genome that should not be expressed in a particular cell type with possible catastrophic results.

Attempts to develop models to explain the way in which vertebrate cells overcome this problem has led to the concept of gene domain (Benyajati and Worcel, 1976). The concept of

gene domain embodies the idea that chromatin structure and chromosomal organisation influence gene expression in a direct and non-random manner (Dillon and Grosveld, 1994). According to this view, domains form units of independent gene activity. In other words, genes within a domain would be subjected to its regulatory environment, but would be insulated from the regulatory environment of surrounding domains. This supposition implies the presence of boundaries, DNA elements that mark the borders between adjacent domains, allowing them to maintain different functional states. The presence of domain boundaries might explain why position effects are often observed when genes are relocated to a new chromosomal environment either by rearrangement or transformation. When genes are transposed to new sites without their associated boundary DNA segments, they would be unprotected from the regulatory influences of the new surroundings. In contrast, if a gene is flanked by boundaries, these DNA sequences should insulate it by delimiting a domain of independent gene activity.

Several different assays have been used to identify and describe boundaries. The concept of boundary is functionally defined to embrace two activities (Bell and Felsenfeld, 1999). The first relates to the ability of these elements to protect a test gene against position effects and act as a barrier between active and inactive chromatin. The second concerns their ability to block the influence of an enhancer on a promoter. This last class of elements is commonly referred to as insulators. In the last few years several elements from different loci and different species have been described to posses boundary activity. For completion a summary of these elements is presented in Table 1. A few of those examples will be discussed in more detail. Insulators are exemplified by the scs and scs' elements which flank the *hsp70* genes at the 87A7 locus in Drosophila. These specialised chromatin structures (scs) define the junctions between the decondensed chromatin of the actively transcribed 87A7 heat-shock locus and adjacent condensed chromatin. Each element consists of a pair of nuclease-hypersensitive sites bordering a 250-300 bp segment of DNA (Udvardy et al., 1985; Vazquez and Schedl, 1994). The properties of the scs elements are consistent with their function as insulators: upregulation of transcription is lost when scs sequences are interposed between an enhancer and a promoter (Kellum and Schedl, 1992). One can also observe that when scs elements are placed on either side of a gene for eye colour and introduced in Drosophila, the resulting flies have all similar eye colour independent of the transgene's site of integration, indicating that these sequences are able to insulate the reporter from 'position effects' (Kellum and Schedl, 1991; Kellum and Schedl, 1992). However, the scs elements do not protect a white mini-gene from exhibiting a variegating phenotype when resident in the

Element	Locus/gene	Organism	Enhancer blocking	Transgene Protection	Deletion assay	Insulate Heterochromatin	Associated Protein(s)
HMR-R	HMR	Yeast	ND	ND	+	+	SMC, SMC3
HMR-L	HMR	Yeast	ND	ND	ND	+	ND
scs	Hsp 70	Drosophila	+	+	ND	-	Zw5
scs'	hsp 70	Drosophila	+	+	ND	-	BEAF-32
Gypsy	Retrotrasposon	Drosophila	+	+	ND	+	SU(HW), Mod(mdg4)
Fab 7	Abdominal-B	Drosophila	+	ND	+	ND	ND
Fab 8	Abdominal-B	Drosophila	+	ND	ND	ND	ND
Mcp	Abdominal-B	Drosophila	ND	ND	+	ND	ND
eve promoter	even-skipped	Drosophila	+	ND	ND	ND	GAGA
sns	Early histone	Sea Urchin	+	ND	ND	ND	ND
Repeat organiser	rRNA	Xenopus	+	ND	ND	ND	XUBF, CTCF
HS2-6	TCRα/δ-Dad1	Human	+	ND	ND	ND	ND
BEAD-1	TCRα/δ	Human	+	ND	ND	ND	CTCF
HS4	β-globin	Chicken	+	ND	ND	+	CTCF
HS5	β-globin	Human	-	+	+	ND	ND

Table 1. Eukaryotic regulatory elements with putative boundary activity.

pericentric heterochromatin (Kellum and Schedl, 1991) and thus cannot be defined as true barrier elements. On the other hand, the *gypsy* insulator element, made up of multiple copies of the *su(Hw)*-binding site, can clearly function as a barrier and protects a transgene from silencing on insertion into heterochromatin (Roseman et al., 1993; Sigrist and Pirrotta, 1997). Therefore, when defining boundary elements operationally, more than one experimental parameter should be taken into consideration. Another attribute should be regarded as a 'rule of thumb' when distinguishing between an insulator and a barrier: barriers serve as a boundary between heterochromatin and euchromatin, whilst insulators operate within euchromatic domains. Indeed, previous observation showing the spreading of gene silencing over hundreds of kilobases and multiple genes (as seen in PEV) imply that many insulators capable of establishing domains of enhancer function are not able to serve as barriers to heterochromatin spreading.

The best-characterised example of a border element in vertebrates is the 1.2-kb DNA fragment containing the 5' HS4 of the chicken β-globin locus. This element displays both attributes of boundary and insulator: in cell lines it provides a directional block to enhancer action on a promoter and it is capable of protecting a gene from position effect in transgenic flies and mice (Chung et al., 1993); (Wang et al., 1997)). The DNaseI hypersensitive site

5'HS4 is also reminiscent of the scs/scs' elements of Drosophila. Its position within the chicken β-globin locus coincides with the region of transition between an active chromatin conformation, marked both by histone hyperacetylation and heightened sensitivity to DNaseI, and an inactive domain extending further 5' that is insensitive to nuclease and less acetylated (Hebbes et al., 1994). The activity of this element seems to be contained within a 250 bp CG-rich core fragment (Chung et al., 1997) that contains binding sites for CTCF (CCCTC-binding factor), a highly conserved and ubiquitous eleven-zinc finger DNA-binding protein implicated in both transcriptional silencing and activation (Burcin et al., 1997; Klenova et al., 1993). Interestingly, functional CTCF-binding sites are also present in other insulators from diverse vertebrate species, suggesting that the enhancer blocking activity of CTCF is a functional component of vertebrate domain boundaries (Bell and Felsenfeld, 1999).

Recently CTCF sites have been demonstrated to be required for the enhancer blocking function of the *Igf2*/H19 imprinted-control region (ICR) (Bell and Felsenfeld, 2000; Hark et al., 2000). The data shows that CTCF binding is abolished on fully methylated DNA, but more interestingly that hemi-methylated DNA (Hark et al., 2000) does not inhibit binding. This indicates that CTCF may bind to the ICR when DNA is transiently hemi-methylated during replication. This conditional binding may be a mechanism for the establishment of heritably stabilized higher order chromatin structures or a mechanism for demethylating the paternal chromosome in the female germ line (Brandeis et al., 1994; Macleod et al., 1994). Overall these results establish a strong correlation between methylation and an open (unblocked) boundary, suggesting that at least some insulators may act not only as fixed boundaries, but also as switches that provide or negate enhancer access to a gene promoter.

Despite the fact that CTCF-binding sites appear to be necessary and sufficient for the enhancer blocking activity, mutation analysis of the HS4 core fragment of the chicken β-globin locus shows that in addition to the CTCF-binding sites other sequences are necessary to prevent heterochromatisation of a test gene (Bell et al., 2001). Therefore, the insulator activity of HS4 is probably distinct from its border activity. It is also likely that the insulator is only part of a larger complex that functions as a barrier to limit the spread of heterochromatin silencing. The notion that multiprotein complexes operate on boundary elements comes from studies conducted in Drosophila. It has been shown that the insulating activity of the gypsy retrotransposon requires not only the DNA-binding protein suppresser of Hairy wings but also the protein encoded by *mod (mdg4)*. Significantly, *mod (mdg4)* has properties like those of the trithorax-Group (trx-G) gene family (Gerasimova and Corces, 1998) in that it inhibits repression from a Polycomb group proteins (Pc-G). When a Polycomb response element (PRE) is placed between two genes, one of which is flanked by gypsy sites,

the protected gene is not repressed but the unprotected gene displays variegated expression (Mallin et al., 1998). Unfortunately, the identity of other proteins complexing *in vivo* with CTCF at the 5' end of the chicken β-globin locus is still to be determined (Bell et al., 1999). Of equal importance, it has not yet been demonstrated that HS4 is able to insulate a test gene from a full LCR.

The DNaseI hypersensitive site 5 (HS5) of the human β-globin locus control region is considered to have the same properties as the HS4 of the chicken β-globin locus. However, contrary to a report where HS5 was shown to display characteristics of an insulating element (Li and Stamatoyannopoulos, 1994), we showed (chapter 2 of this thesis) that in transgenic mice HS5 alone was not sufficient to insulate a β-globin gene from the full human β-globin LCR. The disagreement between our results and those of Li and Stamatoyannopolous may be due to the difference in the experimental conditions used for the assay. In the aforementioned assay the HS3 site of the human β-globin LCR was used to drive expression of a β-globin gene. Because HS3 retains about 30% of the transcriptional activity of the full LCR (Collis et al., 1990; Philipsen et al., 1993), one cannot exclude that HS5 retains some residual insulating activity that is capable of blocking HS3 in small transgenes. Furthermore, our results suggest that genomic sequences 5' of HS5 may be needed for the full insulating activity of the globin insulator. Indeed, the presence of other elements 5' of HS5 has been shown recently (Bulger et al., 1999).

Finally, in a very elegant experiment, the order of the HS sites of the β-globin LCR was reversed, thus placing HS5 proximal to the ε-globin gene. The effect of this inversion resulted in a downregulation of all the β-genes in the locus (Tanimoto et al., 1999). The most straightforward conclusion may be that the LCR is a directional element but it may also mean that HS5 is insulating the genes from the LCR. Elucidation of this hypothesis awaits experiments where for instance the inverted LCR is recombined back into the locus second to the deletion of HS5.

In conclusion the identity of the insulator at the 5' border of the human β-globin locus and its precise location of such element/s is still a matter of debate.

Developmental control of the β-globin genes.

The sequences required for correct developmental expression of the globin genes have been extensively studied in transgenic mice. Introduction of individual globin genes is hampered by lack of expression or low non copy-dependent expression, suggesting that expression is dependent on the position of integration in the mouse genome (Chada et al., 1986; Chada et

al., 1985; Kollias et al., 1987; Magram et al., 1985; Trudel and Costantini, 1987). Although the analysis is made difficult by these position effects, the results suggest that the individual globin genes contain the necessary information for developmental regulation (Starck et al., 1994). Linkage of the individual genes to the LCR achieves high level of expression of the linked gene independent of the site of integration, thus allowing furthering the analysis (Grosveld et al., 1987). The results of such experiments show that the ε- and γ-globin genes are regulated autonomously (Dillon and Grosveld, 1991; Raich et al., 1992). The ε-globin gene is restricted to embryonic red cells and the γ-globin gene is expressed in both embryonic and foetal derived red cells until it is silenced autonomously around day 16 of mouse development. If the β-globin gene is linked directly to the LCR, it is not silent at the embryonic stage of development, but immediately activated in the earliest erythroid cells (Hanscombe et al., 1991). Thus, the human β-globin gene *per se* is not silenced in embryonic and fetal erythroid cells. Proper expression is only achieved when the other genes are placed *in cis* between the β-globin gene and the LCR. This indicates that the β-globin gene is regulated competitively in embryonic and fetal erythroid cells (Strouboulis et al., 1992). The ε- and γ-globin genes are located closer to the LCR and therefore have a competitive advantage over the more distantly located β-globin gene for direct interactions with the LCR. It has indeed been demonstrated that proximity to the LCR is an important factor in the ability for a gene to compete for the activating activity of the LCR. Introduction of a second marked β-globin gene in the ε-globin gene position in the context of a full β-globin locus shows the marked β-globin gene is expressed in embryonic and fetal erythroid cells, and competes efficiently with the γ-globin genes for activation by the LCR (Dillon et al., 1997). Such a competitive switching mechanism provides a simple solution for the fact that the output of the α- and β-globin loci has to be kept in balance through development, even though the α-globin locus does not contain a fetal stage-specific gene.

Thus, the LCR provides an activating function at all stages of development, but its action on individual genes at particular developmental stages is blocked by the binding of stage-specific repressors to individual gene promoters. This autonomous silencing appears to be the major mechanism by which the expression of the early genes is regulated. Activation at later stages of development is achieved by competition between the genes for direct interaction with the LCR, and hence the relative distance of each gene from the LCR is an important parameter. During periods of switching, the genes on the same allele are interdependent, with only one gene being transcribed at any time (Wijgerde et al., 1995). The looping model best explains this mode of regulation, but other models have been put forward (see Fraser and Grosveld,

1998, for a recent review). The looping model proposes that the initiation of transcription requires an interaction between the LCR and the gene, looping out the intervening DNA. The HSs of the LCR would form one holocomplex that interacts as a unit with different regulatory sequences in or surrounding the gene (Dillon et al., 1997; Ellis et al., 1996; Hanscombe et al., 1991). Transcription initiation can only take place while this interaction is maintained and this implies that the level of transcription of the globin genes is the product of how frequently the LCR interacts with a given gene and the stability of this interaction (Wijgerde et al., 1995). This model predicts that if any factor necessary for activating a particular gene would be absent, this gene would lose its competitive advantage over another gene capable of being transcribed during the same stage of development. The validity of this assumption is probably best shown in experiments performed in mice lacking EKLF. The absence of EKLF leads to a complete lack of β-globin expression with a concomitant increase in γ-gene expression in

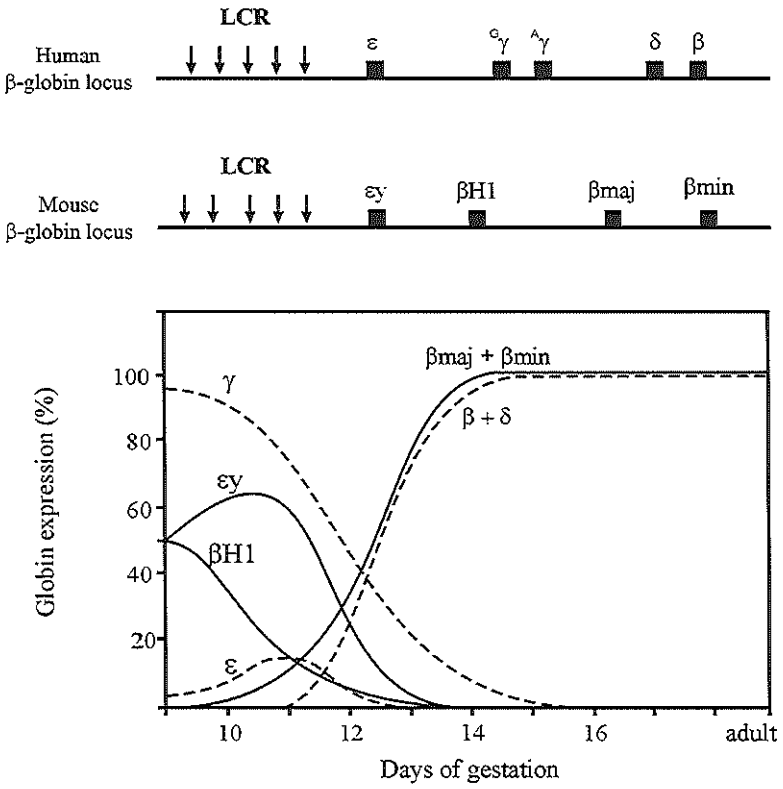

Fig. 4. Expression of the genes of the human β-like globin locus in transgenic mice compared with the endogenous mouse globin genes. The mouse locus does not contain a fetal stage-specific gene, but the human fetal γ-globin genes are expressed appropriately in fetal liver erythropoiesis in transgenic mice.

fetal liver-derived erythroid cells. Furthermore, a decrease of EKLF levels in EKLF+/- mice results in a reduction in the number of transcriptionally active β-globin genes. This is accompanied by an increase in the number of actively transcribed γ-globin genes. Finally, in the adult blood of EKLF +/- EKLF mice where the γ-globin genes are silenced, the levels of β-globin globin return to near normal levels (Wijgerde et al., 1996).

In conclusion the looping model is an interesting candidate for the overall control of the β-globin locus. Direct evidence for this model awaits the application of very sensitive techniques that can demonstrate physical interactions between the LCR and the globin gene promoters *in vivo*. It will then be of interest to determine whether a similar mechanism operates of gene activation operates in the α-globin locus.

Transcriptional regulation of the α-globin locus.

The human α-globin locus is found in a GC-rich isochore in the sub-telomeric region of the short arm of human chromosome 16. It contains one embryonic gene (ζ) and two fetal/adult genes (α1 and α2), in addition to a globin gene of unknown function (θ). The genes are present in the order (telomere) ζ-α2-α1-θ (centromere). Four housekeeping genes are found between the α-globins and the telomeres, and at least five genes have been identified in close proximity at the centromeric side. Most of these genes are ubiquitously expressed (Flint et al., 1997; Higgs et al., 1998). The most important distant regulatory element of the α-globin genes, called HS-40 because it is found 40 kb upstream of the ζ-globin gene (Higgs et al., 1990), is located in the intron of a ubiquitously expressed gene (Vyas et al., 1995). It would therefore appear that the genes in this area of the genome are not organized in distinct chromosomal domains. Thus, the α-like globin cluster is positioned in a very different chromosomal environment than the β-globin locus.

Unlike the promoters of the β-like globin genes, the human α-like globin promoters have features characteristic of methylation-free CpG islands (Bird et al., 1987). Typically, a CpG island is an approximately 1 kb, GC-rich DNA sequence containing the promoter and the first part of the associated gene. CpG islands remain free of DNA methylation, the only stable modification of DNA that is known to occur in higher eukaryotes. The C residue in the dinucleotide sequence 5'-CG-3' can be methylated at the N5 position of the pyrimidine ring by DNA methyltransferases. DNA methylation is generally associated with gene inactivation and repressive chromatin structures, and most 5'-CG-3' dinucleotides undergo this stable modification (Antequera and Bird, 1993). The α-globin locus has an open, DNaseI sensitive chromatin conformation in erythroid as well as non-erythroid cells and tissues, again in

marked contrast to the β-globin cluster that only assumes an open chromatin structure in erythroid cells (Craddock et al., 1995). Possibly related to this open chromatin structure, the β locus replicates early in the cell cycle in erythroid cells only, while the α locus is early replicating in both erythroid and non-erythroid cell lines (Epner et al., 1988; Smith and Higgs, 1999). Thus, there are many striking differences between the α- and β-globin chromosomal domains. Yet, the genes from both clusters are expressed co-ordinately at high levels in erythroid cells only. In the next sections, we will discuss transcriptional regulation of the α-like globin genes.

The ζ-globin gene.

The ζ-globin gene is expressed in erythroid cells of yolk sac origin during the first 6-7 weeks of gestation. The gene is silenced in fetal liver and bone marrow erythropoiesis via a gene-intrinsic mechanism. Thus, downregulation of ζ-globin expression is independent of the presence of the adult α1 and α2-globin genes and of its location in the α-globin gene cluster (Albitar et al., 1991; Gourdon et al., 1994; Sabath et al., 1993; Spangler et al., 1990; Tang et al., 1992). It has been demonstrated that a human ζ-globin gene with 0.5 kb of 5' and 2.2 kb of 3' flanking sequences is appropriately silenced in transgenic mice. Further mapping has identified an 108 bp sequence in the 3' flanking region that functions as a developmental silencer of ζ-globin expression (Liebhaber et al., 1996; Spangler et al., 1990). This silencing activity was attributed to a binding site for the transcription factor NF-κB in this element (Wang and Liebhaber, 1999). However, the significance of this element in ζ-globin silencing in the natural context of the α-locus remains unknown, since the ζ gene is silenced in patients with a deletion of this region (Tang et al., 1992).

Finally, ζ-globin downregulation is further enhanced by the instability of ζ-globin mRNA in erythroid cells of fetal and adult origin (Russell et al., 1998).

The α-globin genes.

The two α-globin genes, α1 and α2, are expressed in fetal liver and adult bone marrow and spleen derived erythroid cells, and probably also in the embryonic yolk sac. The two α-globin genes are almost identical: the first 600 bp of the promoter are completely conserved, as is most of the transcribed region. Yet, the ratio of α1 to α2 mRNAs is 1:2.6 in definitive erythroid cells (Albitar et al., 1992). The molecular mechanism underlying the higher expression level of α2 globin remains unknown, but mutations in the α2-globin gene, such as

the Constant Spring mutation that destabilizes α-globin mRNA, have a particularly severe effect on α-globin protein synthesis (Hunt et al., 1982).

The α-globin promoters contain canonical TATA-box and CCAAT-box motifs, as well as binding sites for Sp/XKLF and GATA factors (Rombel et al., 1995). The CCAAT-box belongs to a different class then the β-gene CCAAT boxes and has been shown to interact with the heteromeric CP2 protein (Lim et al., 1993; Lim et al., 1992). It has been suggested that the CP2 complex is also involved in γ- to β-globin switching (Cunningham et al., 1995; Jane et al., 1995), but its significance for globin gene regulation will remain elusive until knockout mice of the CP2 components have been generated.

The importance of the Sp/XKLF binding sites in the α-globin promoters has been deduced from *in vivo* footprinting experiments, and site-directed mutagenesis (Rombel et al., 1995). Surprisingly, the α-globin genes are still expressed at normal levels in EKLF knockout mice, the only erythroid-specific member of the Sp/XKLF family, as well as in knockout mice of other more widely expressed Sp/XKLF factors such as Sp1, Sp3, LKLF and BKLF (Philipsen and Suske, 1999). These data either negate a role for Sp/XKLF factors in α-globin expression, or suggest the involvement of another family member. The recently described FKLF factor is a good candidate (Asano et al., 1999), but we will have to wait for the generation of FKLF knockout mice to test this hypothesis.

Patients suffering from the ATR-X (for X-linked α-thalassemia and mental retardation) syndrome have provided interesting clues about transcriptional regulation of the α-globin locus (Gibbons et al., 1991). The ATRX gene is a member of a subgroup of the helicase superfamily that includes proteins involved in a wide range of cellular functions, including DNA recombination and repair and regulation of transcription. The complex ATR-X phenotype suggests that the ATRX protein, when mutated, down-regulates expression of several genes, including the alpha-globin genes, indicating that it could be a global transcriptional regulator. As a working model, ATRX would have an important function in establishing and/or maintaining a transcriptionally competent chromatin structure of the α-globin locus (Gibbons et al., 1997; Gibbons et al., 1995). Interestingly, ATRX does not appear to be required for expression of the genes in the β-globin locus (Gibbons et al., 1995). Knockout and transgenic ATRX mice will directly address these issues in the hopefully near future.

Regulatory elements of the α-globin locus: HS-40.

Transgenic mice carrying a 70 kb fragment of the human β-globin locus encompassing the LCR and the five structural β-like globin genes faithfully reproduce the quantitative and spatio-temporal expression pattern of the human β-like globin genes (Strouboulis et al., 1992). For the human α-globin locus, α-thalassemia patients with deletions upstream from the ζ-globin gene first drew attention to a potential role for this region in α-globin expression (Hatton et al., 1990). A strong, erythroid DNaseI hypersensitive site is present in the deleted region. This HS is located in the intron of a ubiquitously expressed gene, some 40 kb upstream from the ζ-globin gene and was termed HS-40 (Higgs et al., 1990). A small deletion encompassing HS-40 had catastrophic consequences for α-globin expression in cultured erythroid cells, in agreement with the patient data (Bernet et al., 1995). Biochemical analysis revealed the HS-40 contains GATA, Sp/XKLF and NFE-E2/AP1 motifs, i.e. very similar transcription factor binding sites to those found in the β-globin LCR (Gourdon et al., 1995; Zhang et al., 1995).

Despite all these observations, it has been exceedingly difficult to obtain both quantitative and qualitative expression in transgenic mice of α-globin transgenes linked to their own regulatory elements. Early experiments had shown that the human α-globin genes can be expressed at high levels when linked *in cis* to the β-globin LCR (Hanscombe et al., 1989). Indeed, HS-40 could drive embryonic and fetal expression of human α-globin, but expression was gradually extinguished at later developmental time points (Higgs et al., 1998; Higgs et al., 1990). Many different constructs were tried, and the best results, i.e. sustained expression during adulthood, were obtained with a large 150 kb construct with the α-globin genes located approximately in the middle. However, even with this construct large variations in expression levels were observed (Higgs et al., 1998). Thus, the extent of the functional α-globin domain remains elusive. It is possible that the telomeric position of the human α-globin locus is important for its regulation. Such a telomeric position is not easily reproduced in transgenic mice, but novel constructs might be designed to include fragments that reconstruct a chromosomal telomere in transfected cells and transgenic mice (Higgs et al., 1998).

Perspective.

The integrated application of biochemistry, molecular biology, transgenic and knockout mice has made major contributions to our understanding of erythropoiesis and globin gene regulation over the past fifteen years or so. Sadly, despite the impressive progress that has

been made in these research areas, we are still not able to apply all this knowledge at the practical level to provide a reliable, safe and cost-effective cure for thalassemic patients. For instance, we now know which DNA elements could be used to direct high-level globin expression in erythroid cells, but it is still a major problem to deliver such constructs to the appropriate cells, i.e. the hematopoietic stem cells. A large scientific effort is geared towards overcoming this hurdle, but it is impossible to predict when this will result in a major breakthrough.

At the same time, there are many new and exciting developments in the areas of gene expression and functional analysis of the genome. Early embryonic lethality is a common complication in knockout mice, precluding the analysis of the function of the factor at later stages. The development of "conditional" knockout strategies allows researchers to inactivate the gene of interest in specific tissues and cells at any point in time, by fine-tuning the expression of a bacterial recombinase called Cre. The gene is flanked by recognition sites for Cre recombinase, and will only be inactivated after Cre-mediated excision of the DNA between the recognition sites (Kilby et al., 1993). This approach will be very useful for the study of the role of transcription factors in adult erythropoiesis, in particular since most of these knockouts die at the embryonic or fetal stage of development, before the onset of bone marrow erythropoiesis.

Genome projects represent another area were spectacular progress has been made. The first draft of the complete nucleotide sequence of the human genome has been published (2001), and will soon be followed by the genome sequence of the mouse. This will have a huge impact on biomedical science. For instance, genes that modify γ-globin expression in adults could be identified by searching the database. This approach would save a huge amount of time in comparison to classical approaches such as positional cloning (Craig et al., 1996). We might then be able to predict the function of the protein involved, and test this hypothesis in transgenic and knockout mice. Since reactivation of γ-globin expression would be beneficial to most β-thalassemia and sickle cell anemia patients this is clearly a very important area of research. Transcription factor knockouts have been instrumental in the elucidation of developmental cascades, such as those underlying hematopoiesis. However, the identity of transcription factor target genes has remained largely elusive, and these need to be identified in order to understand the function of, for instance, GATA-1 in erythropoiesis. With the complete genome sequence in hand, we will have access to all the 40,000 or so human and mouse genes. The rapidly developing DNA microchip technology will enable researchers to obtain genome-wide profiles of gene expression in experimental versus control samples, thus allowing the description of target genes in unprecedented detail. It is expected that DNA

microchips will soon find applications in the clinic, not only for gene expression analysis but also for mutation detection (1999; Iyer et al., 1999). Electronic access to research and patient data will be of vital importance to basic scientists and clinicians. A very useful web site with comprehensive information on globin gene regulation and the haemoglobinopathies is provided by the Laboratories of Computer Science & Engineering and Biochemistry & Molecular Biology at the Pennsylvania State University (USA): http://globin.cse.psu.edu/ .

In conclusion, we can expect many more exciting results that will increase our understanding of erythropoiesis and globin gene regulation in ever more detail. The application of this knowledge should enable biomedical research to deliver one of its major promises: an effective cure for the anemias.

References

(1999) The chip challenge. *Nature Genetics*, **21**, 61-62.

(2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Albitar, M., Cash, F.E., Peschle, C. and Liebhaber, S.A. (1992) Developmental switch in the relative expression of the alpha 1- and alpha 2-globin genes in humans and in transgenic mice. *Blood*, **79**, 2471-4.

Albitar, M., Katsumata, M. and Liebhaber, S.A. (1991) Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol*, **11**, 3786-94.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. and Orkin, S.H. (1993) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic- leucine zipper protein. *Nature*, **362**, 722-8.

Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A*, **90**, 11995-9.

Antoniou, M., de Boer, E., Spanopoulou, E., Imam, A. and Grosveld, F. (1995) TBP binding and the rate of transcription initiation from the human beta-globin gene. *Nucleic Acids Res*, **23**, 3473-80.

Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988) The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *Embo J*, **7**, 377-84.

Antoniou, M. and Grosveld, F. (1990) beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev*, **4**, 1007-13.

Armstrong, J.A., Bieker, J.J. and Emerson, B.M. (1998) A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell*, **95**, 93-104.

Asano, H., Li, X.S. and Stamatoyannopoulos, G. (1999) FKLF, a novel Kruppel-like factor that activates human embryonic and fetal beta-like globin genes. *Mol Cell Biol*, **19**, 3571-9.

Bacolla, A., Ulrich, M.J., Larson, J.E., Ley, T.J. and Wells, R.D. (1995) An intramolecular triplex in the human gamma-globin 5'-flanking region is altered by point mutations associated with hereditary persistence of fetal hemoglobin. *J Biol Chem*, **270**, 24556-63.

Banerji, J., Olson, L. and Schaffner, W. (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, **33**, 729-40.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, **27**, 299-308.

Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. and Townes, T.M. (1987) Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. *Proc Natl Acad Sci U S A*, **84**, 7056-60.

Bell, A.C. and Felsenfeld, G. (1999) Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev*, **9**, 191-8.

Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482-5.

Bell, A.C., West, A.G. and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**, 387-96.

Bell, A.C., West, A.G. and Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science*, **291**, 447-50.

Benyajati, C. and Worcel, A. (1976) Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. *Cell*, **9**, 393-407.

Bernet, A., Sabatier, S., Picketts, D.J., Ouazana, R., Morle, F., Higgs, D.R. and Godet, J. (1995) Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood*, **86**, 1202-11.

Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, **358**, 499-502.

Bird, A.P., Taggart, M.H., Nicholls, R.D. and Higgs, D.R. (1987) Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *Embo J*, **6**, 999-1004.

Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. *Science*, **281**, 61-3.

Blobel, G.A., Nakajima, T., Eckner, R., Montminy, M. and Orkin, S.H. (1998) CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A*, **95**, 2061-6.

Bodine, D.M. and Ley, T.J. (1987) An enhancer element lies 3' to the human A gamma globin gene. *Embo J*, **6**, 2997-3004.

Bouwman, P., Gollner, H., Elsasser, H.P., Eckhoff, G., Karis, A., Grosveld, F., Philipsen, S. and Suske, G. (2000) Transcription factor Sp3 is essential for post-natal survival and late tooth development. *Embo J*, **19**, 655-661.

Boyes, J., Byfield, P., Nakatani, Y. and Ogryzko, V. (1998) Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature*, **396**, 594-8.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature*, **371**, 435-8.

Bulger, M., van Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R., Groudine, M. and von Doorninck, J.H. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A*, **96**, 5129-34.

Bungert, J., Dave, U., Lim, K.C., Lieuw, K.H., Shavit, J.A., Liu, Q. and Engel, J.D. (1995) Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev*, **9**, 3083-96.

Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenkov, V.V. and Renkawitz, R. (1997) Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol Cell Biol*, **17**, 1281-8.

Calzolari, R., McMorrow, T., Yannoutsos, N., Langeveld, A. and Grosveld, F. (1999) Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *Embo J*, **18**, 949-58.

Chada, K., Magram, J. and Costantini, F. (1986) An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature*, **319**, 685-9.

Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Costantini, F. (1985) Specific expression of a foreign beta-globin gene in erythroid cells of transgenic mice. *Nature*, **314**, 377-80.

Chan, J.Y., Kwong, M., Lu, R., Chang, J., Wang, B., Yen, T.S. and Kan, Y.W. (1998) Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. *Embo J*, **17**, 1779-87.

Chan, K., Lu, R., Chang, J.C. and Kan, Y.W. (1996) NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proc Natl Acad Sci U S A*, **93**, 13943-8.

Chen, X. and Bieker, J.J. (1996) Erythroid Kruppel-like factor (EKLF) contains a multifunctional transcriptional activation domain important for inter- and intramolecular interactions. *Embo J*, **15**, 5888-96.

Cheng, X., Reginato, M.J., Andrews, N.C. and Lazar, M.A. (1997) The transcriptional integrator CREB-binding protein mediates positive cross talk between nuclear hormone receptors and the hematopoietic bZip protein p45/NF-E2. *Mol Cell Biol*, **17**, 1407-16.

Chung, J.H., Bell, A.C. and Felsenfeld, G. (1997) Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A*, **94**, 575-80.

Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*, **74**, 505-14.

Collis, P., Antoniou, M. and Grosveld, F. (1990) Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *Embo J*, **9**, 233-40.

Conaway, J.W. and Conaway, R.C. (1991) Initiation of eukaryotic messenger RNA synthesis. *J Biol Chem*, **266**, 17721-4.

Craddock, C.F., Vyas, P., Sharpe, J.A., Ayyub, H., Wood, W.G. and Higgs, D.R. (1995) Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *Embo J*, **14**, 1718-26.

Craig, J.E., Rochette, J., Fisher, C.A., Weatherall, D.J., Marc, S., Lathrop, G.M., Demenais, F. and Thein, S. (1996) Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat Genet*, **12**, 58-64.

Crispino, J.D., Lodish, M.B., MacKay, J.P. and Orkin, S.H. (1999) Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell*, **3**, 219-28.

Cunningham, J.M., Vanin, E.F., Tran, N., Valentine, M. and Jane, S.M. (1995) The human transcription factor CP2 (TFCP2), a component of the human gamma-globin stage selector protein, maps to chromosome region 12q13 and is within 250 kb of the NF-E2 gene. *Genomics*, **30**, 398-9.

Custodio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F. and Antoniou, M. (1999) Inefficient processing impairs release of RNA from the site of transcription. *Embo J*, **18**, 2855-66.

Deconinck, A.E., Mead, P.E., Tevosian, S.G., Crispino, J.D., Katz, S.G., Zon, L.I. and Orkin, S.H. (2000) FOG acts as a repressor of red blood cell development in Xenopus. *Development*, **127**, 2031-2040.

Delvoye, N.L., Destroismaisons, N.M. and Wall, L.A. (1993) Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol Cell Biol*, **13**, 6969-83.

Dillon, N. and Grosveld, F. (1991) Human gamma-globin genes silenced independently of other genes in the beta-globin locus. *Nature*, **350**, 252-4.

Dillon, N. and Grosveld, F. (1994) Chromatin domains as potential units of eukaryotic gene function. *Curr Opin Genet Dev*, **4**, 260-4.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997) The effect of distance on long-range chromatin interactions. *Mol Cell*, **1**, 131-9.

Donze, D., Jeancake, P.H. and Townes, T.M. (1996) Activation of delta-globin gene expression by erythroid Krupple-like factor: a potential approach for gene therapy of sickle cell disease. *Blood*, **88**, 4051-7.

Dyer, M.A., Hayes, P.J. and Baron, M.H. (1998) The HMG domain protein SSRP1/PREIIBF is involved in activation of the human embryonic beta-like globin gene. *Mol Cell Biol*, **18**, 2617-28.

Dzierzak, E., Medvinsky, A. and de Bruijn, M. (1998) Qualitative and quantitative aspects of haematopoietic cell development in the mammalian embryo. *Immunol Today*, **19**, 228-36.

Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993) Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *Embo J*, **12**, 127-34.

Ellis, J., Tan-Un, K.C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S. and Grosveld, F. (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *Embo J*, **15**, 562-8.

Epner, E., Forrester, W.C. and Groudine, M. (1988) Asynchronous DNA replication within the human beta-globin gene locus. *Proc Natl Acad Sci U S A*, **85**, 8081-5.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G. and Groudine, M. (1998) The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*, **2**, 447-55.

Farmer, S.C., Sun, C.W., Winnier, G.E., Hogan, B.L. and Townes, T.M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Genes Dev*, **11**, 786-98.

Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem*, **269**, 1493-500.

Festenstein, R. and Kioussis, D. (2000) Locus control regions and epigenetic chromatin modifiers [In Process Citation]. *Curr Opin Genet Dev*, **10**, 199-203.

Filipe, A., Li, Q., Deveaux, S., Godin, I., Romeo, P.H., Stamatoyannopoulos, G. and Mignotte, V. (1999) Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *Embo J*, **18**, 687-97.

Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A. and Higgs, D.R. (1997) The relationship between chromosome structure and function at a human telomeric region. *Nat Genet*, **15**, 252-7.

Forget, B.G. (1998) Molecular basis of hereditary persistence of fetal hemoglobin. *Ann N Y Acad Sci*, **850**, 38-44.

Fraser, P. and Grosveld, F. (1998) Locus control regions, chromatin activation and transcription. *Curr Opin Cell Biol*, **10**, 361-5.

Gerasimova, T.I. and Corces, V.G. (1998) Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*, **92**, 511-21.

Gibbons, R.J., Bachoo, S., Picketts, D.J., Aftimos, S., Asenbauer, B., Bergoffen, J., Berry, S.A., Dahl, N., Fryer, A., Keppler, K., Kurosawa, K., Levin, M.L., Masuno, M., Neri, G., Pierpont, M.E., Slaney, S.F. and Higgs, D.R. (1997) Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain [letter]. *Nat Genet*, **17**, 146-8.

Gibbons, R.J., Picketts, D.J., Villard, L. and Higgs, D.R. (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell*, **80**, 837-45.

Gibbons, R.J., Wilkie, A.O., Weatherall, D.J. and Higgs, D.R. (1991) A newly defined X linked mental retardation syndrome associated with alpha thalassaemia. *J Med Genet*, **28**, 729-33.

Gillemans, N., Tewari, R., Lindeboom, F., Rottier, R., de Wit, T., Wijgerde, M., Grosveld, F. and Philipsen, S. (1998) Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region in vivo. *Genes Dev*, **12**, 2863-73.

Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, **33**, 717-28.

Gourdon, G., Sharpe, J.A., Higgs, D.R. and Wood, W.G. (1995) The mouse alpha-globin locus regulatory element. *Blood*, **86**, 766-75.

Gourdon, G., Sharpe, J.A., Wells, D., Wood, W.G. and Higgs, D.R. (1994) Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucleic Acids Res*, **22**, 4139-47.

Gralla, J.D. (1991) Transcriptional control--lessons from an E. coli promoter data base. *Cell*, **66**, 415-8.

Green, M.R. (2000) TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci*, **25**, 59-63.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-85.

Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D.R. and Grosveld, F. (1989) High-level, erythroid-specific expression of the human alpha-globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes. *Genes Dev*, **3**, 1572-81.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) Importance of globin gene order for correct developmental expression. *Genes Dev*, **5**, 1387-94.

Hardison, R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol*, **201**, 1099-117.

Hardison, R., Riemer, C., Chui, D.H., Huisman, T.H. and Miller, W. (1998) Electronic access to sequence alignments, experimental results, and human mutations as an aid to studying globin gene regulation. *Genomics*, **47**, 429-37.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486-9.

Hatton, C.S., Wilkie, A.O., Drysdale, H.C., Wood, W.G., Vickers, M.A., Sharpe, J., Ayyub, H., Pretorius, I.M., Buckle, V.J. and Higgs, D.R. (1990) Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood*, **76**, 221-7.

Hebbes, T.R., Clayton, A.L., Thorne, A.W. and Crane-Robinson, C. (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *Embo J*, **13**, 1823-30.

Heyworth, C., Gale, K., Dexter, M., May, G. and Enver, T. (1999) A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal. *Genes Dev*, **13**, 1847-60.

Higgs, D.R., Sharpe, J.A. and Wood, W.G. (1998) Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin Hematol*, **35**, 93-104.

Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.M. and Ayyub, H. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev*, **4**, 1588-601.

Hung, H.L., Lau, J., Kim, A.Y., Weiss, M.J. and Blobel, G.A. (1999) CREB-Binding protein acetylates hematopoietic transcription factor GATA- 1 at functionally important sites. *Mol Cell Biol*, **19**, 3496-505.

Hunt, D.M., Higgs, D.R., Winichagoon, P., Clegg, J.B. and Weatherall, D.J. (1982) Haemoglobin Constant Spring has an unstable alpha chain messenger RNA. *Br J Haematol*, **51**, 405-13.

Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. and Yamamoto, M. (1994) Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins [see comments]. *Nature*, **367**, 568-72.

Ingram, V.M. (1956) *Nature*, **252**, 792-4.

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J.D. and Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*, **362**, 466-8.

Itoh, K., Igarashi, K., Hayashi, N., Nishizawa, M. and Yamamoto, M. (1995) Cloning and characterization of a novel erythroid cell-derived CNC family transcription factor heterodimerizing with the small Maf family proteins. *Mol Cell Biol*, **15**, 4184-93.

Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson, J., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. and Brown, P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83-87.

Jack, J. and DeLotto, Y. (1995) Structure and regulation of a complex locus: the cut gene of Drosophila. *Genetics*, **139**, 1689-700.

Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L. and Nienhuis, A.W. (1992) Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *Embo J*, **11**, 2961-9.

Jane, S.M., Nienhuis, A.W. and Cunningham, J.M. (1995) Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein [published erratum appears in EMBO J 1995 Feb 15;14(4):854]. *Embo J*, **14**, 97-105.

Kadonaga, J.T. (1990) Gene transcription: basal and regulated transcription by RNA polymerase II. *Curr Opin Cell Biol*, **2**, 496-501.

Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, **64**, 941-50.

Kellum, R. and Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol*, **12**, 2424-31.

Kilby, N.J., Snaith, M.R. and Murray, J.A. (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet*, **9**, 413-21.

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983) Beta-globin gene inactivation by DNA translocation in gamma beta- thalassaemia. *Nature*, **306**, 662-6.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E. and Lobanenkov, V.V. (1993) CTCF, a conserved nuclear factor required for optimal transcriptional activity

of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol*, **13**, 7612-24.

Kobayashi, A., Ito, E., Toki, T., Kogame, K., Takahashi, S., Igarashi, K., Hayashi, N. and Yamamoto, M. (1999) Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3. *J Biol Chem*, **274**, 6443-52.

Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) The human beta-globin gene contains a downstream developmental specific enhancer. *Nucleic Acids Res*, **15**, 5739-47.

Kotkow, K.J. and Orkin, S.H. (1996) Complexity of the erythroid transcription factor NF-E2 as revealed by gene targeting of the mouse p18 NF-E2 locus. *Proc Natl Acad Sci U S A*, **93**, 3514-8.

Kulozik, A.E., Bellan-Koch, A., Bail, S., Kohne, E. and Kleihauer, E. (1991) Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood*, **77**, 2054-8.

Kuroha, T., Takahashi, S., Komeno, T., Itoh, K., Nagasawa, T. and Yamamoto, M. (1998) Ablation of Nrf2 function does not increase the erythroid or megakaryocytic cell lineage dysfunction caused by p45 NF-E2 gene disruption. *J Biochem (Tokyo)*, **123**, 376-9.

Li, Q. and Stamatoyannopoulos, G. (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood*, **84**, 1399-401.

Liebhaber, S.A., Wang, Z., Cash, F.E., Monks, B. and Russell, J.E. (1996) Developmental silencing of the embryonic zeta-globin gene: concerted action of the promoter and the 3'-flanking region combined with stage-specific silencing by the transcribed segment. *Mol Cell Biol*, **16**, 2637-46.

Lim, L.C., Fang, L., Swendeman, S.L. and Sheffery, M. (1993) Characterization of the molecularly cloned murine alpha-globin transcription factor CP2. *J Biol Chem*, **268**, 18008-17.

Lim, L.C., Swendeman, S.L. and Sheffery, M. (1992) Molecular cloning of the alpha-globin transcription factor CP2. *Mol Cell Biol*, **12**, 828-35.

Liu, Q., Bungert, J. and Engel, J.D. (1997) Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A*, **94**, 169-74.

Liu, Q., Tanimoto, K., Bungert, J. and Engel, J.D. (1998) The A gamma-globin 3' element provides no unique function(s) for human beta-globin locus gene regulation. *Proc Natl Acad Sci U S A*, **95**, 9944-9.

Luzzatto, L., Nwachuku-Jarrett, E.S. and Reddy, S. (1970) Increased sickling of parasitised erythrocytes as mechanism of resistance against malaria in the sickle-cell trait. *Lancet*, **1**, 319-21.

Macleod, D., Charlton, J., Mullins, J. and Bird, A.P. (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev*, **8**, 2282-92.

Magram, J., Chada, K. and Costantini, F. (1985) Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature*, **315**, 338-40.

Mallin, D.R., Myung, J.S., Patton, J.S. and Geyer, P.K. (1998) Polycomb group repression is blocked by the Drosophila suppressor of Hairy-wing [su(Hw)] insulator. *Genetics*, **148**, 331-9.

Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res*, **26**, 1135-43.

Mantovani, R., Superti-Furga, G., Gilman, J. and Ottolenghi, S. (1989) The deletion of the distal CCAAT box region of the A gamma-globin gene in black HPFH abolishes the binding of the erythroid specific protein NFE3 and of the CCAAT displacement protein. *Nucleic Acids Res*, **17**, 6681-91.

Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell*, **89**, 619-28.

Martin, D.I., Zon, L.I., Mutter, G. and Orkin, S.H. (1990) Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature*, **344**, 444-7.

Martin, F., van Deursen, J.M., Shivdasani, R.A., Jackson, C.W., Troutman, A.G. and Ney, P.A. (1998) Erythroid maturation and globin gene expression in mice with combined deficiency of NF-E2 and nrf-2. *Blood*, **91**, 3459-66.

Mastrangelo, I.A., Courey, A.J., Wall, J.S., Jackson, S.P. and Hough, P.V. (1991) DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc Natl Acad Sci U S A*, **88**, 5670-4.

Merika, M. and Orkin, S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol*, **15**, 2437-47.

Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P.H. (1989) Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res*, **17**, 37-54.

Miller, I.J. and Bieker, J.J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol*, **13**, 2776-86.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P. (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**, 105-14.

Mueller-Storm, H.P., Sogo, J.M. and Schaffner, W. (1989) An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge [published erratum appears in Cell 1989 Oct 20;59(2):405]. *Cell*, **58**, 767-77.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H. and Nienhuis, A.W. (1990) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res*, **18**, 6011-7.

Nienhuis, A.W. and Benz, E.J., Jr. (1977) Regulation of hemoglobin synthesis during the development of the red cell. (Second of three parts). *N Engl J Med*, **297**, 1371-81.

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316-8.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438-46.

Onodera, K., Shavit, J.A., Motohashi, H., Yamamoto, M. and Engel, J.D. (2000) Perinatal synthetic lethality and hematopoietic defects in compound mafG::mafK mutant mice [In Process Citation]. *Embo J*, **19**, 1335-45.

Orkin, S.H. (1986) *Disorder of Haemoglobin synthesis: The Thalassemias*.

Ouyang, L., Chen, X. and Bieker, J.J. (1998) Regulation of erythroid Kruppel-like factor (EKLF) transcriptional activity by phosphorylation of a protein kinase casein kinase II site within its interaction domain. *J Biol Chem*, **273**, 23019-25.

Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M. and Igarashi, K. (1996) Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol Cell Biol*, **16**, 6083-95.

Pandolfi, P.P., Roth, M.E., Karis, A., Leonard, M.W., Dzierzak, E., Grosveld, F.G., Engel, J.D. and Lindenbaum, M.H. (1995) Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis [see comments]. *Nat Genet*, **11**, 40-4.

Pasvol, G., Weatherall, D.J. and Wilson, R.J. (1978) Cellular mechanism for the protective effect of haemoglobin S against P. falciparum malaria. *Nature*, **274**, 701-3.

Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC- transcription factor EKLF. *Nature*, **375**, 318-22.

Perkins, A.C., Yang, H., Crossley, P.M., Fujiwara, Y. and Orkin, S.H. (1997) Deficiency of the CACC-element binding protein, BKLF, leads to a progressive myeloproliferative diease and impaired expression of SHP-1. *Blood, Supplement 1*, **90**, 575a.

Perutz, M.F. (1960) *Nature*, **185**, 416-22.

Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G. and Stamatoyannopoulos, G. (1996) Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A*, **93**, 6605-9.

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. and Costantini, F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development*, **121**, 163-72.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257-60.

Philipsen, S., Pruzina, S. and Grosveld, F. (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *Embo J*, **12**, 1077-85.

Philipsen, S. and Suske, G. (1999) A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res*, **27**, 2991-3000.

Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) The beta-globin dominant control region: hypersensitive site 2. *Embo J*, **9**, 2159-67.

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. and Philipsen, S. (1991) Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Res*, **19**, 1413-9.

Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature*, **386**, 569-77.

Raich, N., Clegg, C.H., Grofti, J., Romeo, P.H. and Stamatoyannopoulos, G. (1995) GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. *Embo J*, **14**, 801-9.

Raich, N., Papayannopoulou, T., Stamatoyannopoulos, G. and Enver, T. (1992) Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood*, **79**, 861-4.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol*, **18**, 5992-6000.

Riemer, C., ElSherbini, A., Stojanovic, N., Schwartz, S., Kwitkin, P.B., Miller, W. and Hardison, R. (1998) A database of experimental results on globin gene expression. *Genomics*, **53**, 325-37.

Ristaldi, M.S., Casula, S., Porcu, S., Marongiu, M.F., Pirastu, M. and Cao, A. (1999) Activation of the delta-globin gene by the beta-globin gene CACCC motif. *Blood Cells Mol Dis*, **25**, 193-209.

Ristaldi, M.S., Drabek, D., Gribnau, J., Poddie, D., Yannoutsous, N., Cao, A., Grosveld, F. and Imam, A.M. (2001) The role of the -50 region of the human gamma-globin gene in switching. *Embo J*, **20**, 5242-5249.

Roeder, R.G. (1996a) Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. *Methods Enzymol*, **273**, 165-71.

Roeder, R.G. (1996b) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, **21**, 327-35.

Rombel, I., Hu, K.Y., Zhang, Q., Papayannopoulou, T., Stamatoyannopoulos, G. and Shen, C.K. (1995) Transcriptional activation of human adult alpha-globin genes by hypersensitive site-40 enhancer: function of nuclear factor-binding motifs occupied in erythroid cells. *Proc Natl Acad Sci U S A*, **92**, 6454-8.

Romeo, P.H., Prandini, M.H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. and Uzan, G. (1990) Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature*, **344**, 447-9.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996) Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *Embo J*, **15**, 143-9.

Roseman, R.R., Pirrotta, V. and Geyer, P.K. (1993) The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *Embo J*, **12**, 435-42.

Russell, J.E., Morales, J., Makeyev, A.V. and Liebhaber, S.A. (1998) Sequence divergence in the 3' untranslated regions of human zeta- and alpha-globin mRNAs mediates a difference in their stabilities and contributes to efficient alpha-to-zeta gene development switching. *Mol Cell Biol*, **18**, 2173-83.

Sabath, D.E., Spangler, E.A., Rubin, E.M. and Stamatoyannopoulos, G. (1993) Analysis of the human zeta-globin gene promoter in transgenic mice. *Blood*, **82**, 2899-905.

Schule, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988) Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. *Nature*, **332**, 87-90.

Shivdasani, R.A. and Orkin, S.H. (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci U S A*, **92**, 8690-4.

Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J. and Orkin, S.H. (1995) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, **81**, 695-704.

Sigrist, C.J. and Pirrotta, V. (1997) Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes. *Genetics*, **147**, 209-21.

Smale, S.T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta*, **1351**, 73-88.

Smith, Z.E. and Higgs, D.R. (1999) The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression. *Hum Mol Genet*, **8**, 1373-86.

Southwood, C.M., Downs, K.M. and Bieker, J.J. (1996) Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn*, **206**, 248-59.

Spangler, E.A., Andrews, K.A. and Rubin, E.M. (1990) Developmental regulation of the human zeta globin gene in transgenic mice. *Nucleic Acids Res*, **18**, 7093-7.

Starck, J., Sarkar, R., Romana, M., Bhargava, A., Scarpa, A.L., Tanaka, M., Chamberlain, J.W., Weissman, S.M. and Forget, B.G. (1994) Developmental regulation of human gamma- and beta-globin genes in the absence of the locus control region. *Blood*, **84**, 1656-65.

Strouboulis, J., Dillon, N. and Grosveld, F. (1992) Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev*, **6**, 1857-64.

Stuve, L.L. and Myers, R.M. (1990) A directly repeated sequence in the beta-globin promoter regulates transcription in murine erythroleukemia cells. *Mol Cell Biol*, **10**, 972-81.

Su, W., Jackson, S., Tjian, R. and Echols, H. (1991) DNA looping between sites for transcriptional activation: self- association of DNA-bound Sp1. *Genes Dev*, **5**, 820-6.

Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988) The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. *Embo J*, **7**, 3099-107.

Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) Detailed analysis of the site 3 region of the human beta-globin dominant control region. *Embo J*, **9**, 2169-77.

Tang, D.C., Ebb, D., Hardison, R.C. and Rodgers, G.P. (1997) Restoration of the CCAAT box or insertion of the CACCC motif activates [corrected] delta-globin gene expression [published erratum appears in Blood 1997 Sep 1;90(5):2120]. *Blood*, **90**, 421-7.

Tang, D.C. and Rodgers, G.P. (1998) Activation of the human delta-globin gene promoter in primary adult erythroid cells. *Br J Haematol*, **103**, 835-8.

Tang, W., Luo, H.Y., Albitar, M., Patterson, M., Eng, B., Waye, J.S., Liebhaber, S.A., Higgs, D.R. and Chui, D.H. (1992) Human embryonic zeta-globin chain expression in deletional alpha-thalassemias. *Blood*, **80**, 517-22.

Tanimoto, K., Liu, Q., Bungert, J. and Engel, J.D. (1999) Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature*, **398**, 344-8.

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. and Grosveld, F.G. (1986) Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res*, **14**, 7017-29.

Tewari, R., Gillemans, N., Wijgerde, M., Nuez, B., von Lindern, M., Grosveld, F. and Philipsen, S. (1998) Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *Embo J*, **17**, 2334-41.

Ting, C.N., Olson, M.C., Barton, K.P. and Leiden, J.M. (1996) Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature*, **384**, 474-8.

Tjian, R. (1995) Molecular machines that control genes. *Sci Am*, **272**, 54-61.

Tjian, R. (1996) The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes. *Philos Trans R Soc Lond B Biol Sci*, **351**, 491-9.

Tjian, R. and Maniatis, T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, **77**, 5-8.

Toki, T., Itoh, J., Kitazawa, J., Arai, K., Hatakeyama, K., Akasaka, J., Igarashi, K., Nomura, N., Yokoyama, M., Yamamoto, M. and Ito, E. (1997) Human small Maf proteins form heterodimers with CNC family transcription factors and recognize the NF-E2 motif. *Oncogene*, **14**, 1901-10.

Trainor, C.D., Omichinski, J.G., Vandergon, T.L., Gronenborn, A.M., Clore, G.M. and Felsenfeld, G. (1996) A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high- affinity interaction. *Mol Cell Biol*, **16**, 2238-47.

Trudel, M. and Costantini, F. (1987) A 3' enhancer contributes to the stage-specific expression of the human beta-globin gene. *Genes Dev*, **1**, 954-61.

Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W. and Orkin, S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221-6.

Tsai, F.Y. and Orkin, S.H. (1997) Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood*, **89**, 3636-43.

Tsang, A.P., Fujiwara, Y., Hom, D.B. and Orkin, S.H. (1998) Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev*, **12**, 1176-88.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M. and Orkin, S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, **90**, 109-19.

Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A*, **82**, 6384-8.

Tuan, D.Y., Solomon, W.B., London, I.M. and Lee, D.P. (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A*, **86**, 2554-8.

Udvardy, A., Maine, E. and Schedl, P. (1985) The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol*, **185**, 341-58.

Ulrich, M.J., Gray, W.J. and Ley, T.J. (1992) An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. *J Biol Chem*, **267**, 18649-58.

Vazquez, J. and Schedl, P. (1994) Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. *Embo J*, **13**, 5984-93.

Vyas, P., Vickers, M.A., Picketts, D.J. and Higgs, D.R. (1995) Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics*, **29**, 679-89.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A. and Rabbitts, T.H. (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J*, **16**, 3145-57.

Wall, L., Destroismaisons, N., Delvoye, N. and Guy, L.G. (1996) CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells. *J Biol Chem*, **271**, 16477-84.

Wang, Y., DeMayo, F.J., Tsai, S.Y. and O'Malley, B.W. (1997) Ligand-inducible and liver-specific target gene expression in transgenic mice. *Nat Biotechnol*, **15**, 239-43.

Wang, Z. and Liebhaber, S.A. (1999) A 3'-flanking NF-kappaB site mediates developmental silencing of the human zeta-globin gene. *Embo J*, **18**, 2218-28.

Warren, A.J., Colledge, W.H., Carlton, M.B., Evans, M.J., Smith, A.J. and Rabbitts, T.H. (1994) The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell*, **78**, 45-57.

Wasylyk, B. (1988) Transcription elements and factors of RNA polymerase B promoters of higher eukaryotes. *CRC Crit Rev Biochem*, **23**, 77-120.

Weatherall, D.J. and Clegg, J.B. (1981) *The thalassemia syndrome*. Blackwell Scientific, Oxford.

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *Faseb J*, **6**, 3300-9.

Weiss, M.J. and Orkin, S.H. (1995a) GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol*, **23**, 99-107.

Weiss, M.J. and Orkin, S.H. (1995b) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A*, **92**, 9623-7.

Whyatt, D., Lindeboom, F., Karis, A., Ferreira, R., Milot, E., Hendriks, R., de Bruijn, M., Langeveld, A., Gribnau, J., Grosveld, F. and Philipsen, S. (2000) An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature*, **406**, 519-524.

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition. *Genes Dev*, **10**, 2894-902.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, **377**, 209-13.

Winoto, A. and Baltimore, D. (1989) A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor alpha locus. *Embo J*, **8**, 729-33.

Wood, W.G. (1993) Increased HbF in adult life. *Baillieres Clin Haematol*, **6**, 177-213.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R. and Rabbitts, T.H. (1998) The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci U S A*, **95**, 3890-5.

Yoshida, C., Tokumasu, F., Hohmura, K.I., Bungert, J., Hayashi, N., Nagasawa, T., Engel, J.D., Yamamoto, M., Takeyasu, K. and Igarashi, K. (1999) Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells*, **4**, 643-55.

Zafarana, G., Rottier, R., Grosveld, F. and Philipsen, S. (2000) Erythroid overexpression of C/EBPgamma in transgenic mice affects gamma- globin expression and fetal liver erythropoiesis. *Embo J*, **19**, 5856-63.

Zhang, Q., Rombel, I., Reddy, G.N., Gang, J.B. and Shen, C.K. (1995) Functional roles of in vivo footprinted DNA motifs within an alpha- globin enhancer. Erythroid lineage and developmental stage specificities. *J Biol Chem*, **270**, 8501-5.

Zhang, W. and Bieker, J.J. (1998) Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A*, **95**, 9855-60.

Chapter 2

Characterization of hypersensitive site 5 of the human β-globin Locus Control Region.

Gaetano Zafarana, Selina Raguz, Sara Pruzina, Alex Harper Sjaak Philipsen, Frank Grosveld and Dies Meijer

Characterization of hypersensitive site 5 of the human β-globin Locus Control Region.

Introduction

DNaseI hypersensitive sites (HS) flank the human β-globin multigene cluster (Forrester et al., 1987; Forrester et al., 1986; Tuan et al., 1985). The set of erythroid-specific HS sites spread over a 20 kb region upstream of the embryonic epsilon globin gene has been named Locus Control region (LCR). Addition of the LCR to a β gene results in high levels of expression, in an integration site independent and copy number dependent fashion of the transgene in cultured cells and transgenic animals (Grosveld et al., 1987). The LCR comprises five hypersensitive sites (HS 1-5). Previous work from several laboratories has shown that the individual HS 2-4 have LCR activity although to a lower level than the full LCR (Ellis et al., 1993; Fraser et al., 1990; Philipsen et al., 1993; Philipsen et al., 1990; Pruzina et al., 1991; Talbot et al., 1990). These hypersensitive sites can only be detected in the chromatin of erythroid cells in vivo. HS 5 was described as a non-tissue specific site (Tuan et al., 1985) and has not been studied in the same detail. One group reported on its ability to cooperate with the polyoma enhancer to give position independent expression in transfection experiments (Yu et al., 1994), whereas Li and Stamatoyannopolous reported that HS5 may act as an insulator with the same function as described earlier for the equivalent HS site from the chicken β-globin locus (Chung et al., 1993; Li and Stamatoyannopoulos, 1994).

To clarify the role of HS5 in the regulation of the β-globin domain, we have studied the role of HS5 using transgenic mice as a non-selective in vivo expression system. Interestingly and in contrast to the published results, we show that HS5 is erythroid specific but that HS5 alone does not have any β-globin gene transcriptional activator properties. In addition we show that HS5 does not function as an insulator as it does not prevent integration site dependent expression of a reporter gene in transgenic mice.

Results

Hypersensitive site analysis in transgenic mice and human cells.

Fig. 1 gives an overview of the human β-globin locus and the strategy used for hypersensitive site analysis. We studied the chromatin structure in tissues derived from line 72 transgenic mice that contain a single copy of the 70 kb human β-globin locus (Strouboulis et al., 1992). Thymocytes and fetal liver were used as a source of non-erythroid cells (T-cells) and erythroid precursors, respectively. HS5 as well 2 and 3 were absent from thymic lymphoid cells but present in erythroid cells (Fig. 2 A,B). Thymic nuclei showed the presence of the murine vav gene HS (Fig. 2 C; Ogilvy et al., 1998) and demonstrated that the absence of detectable LCR hypersensitive sites was not due to technical problems. The erythroid specificity of HS5 was further confirmed by analysis of the chromatin from human non-immortalized cells.

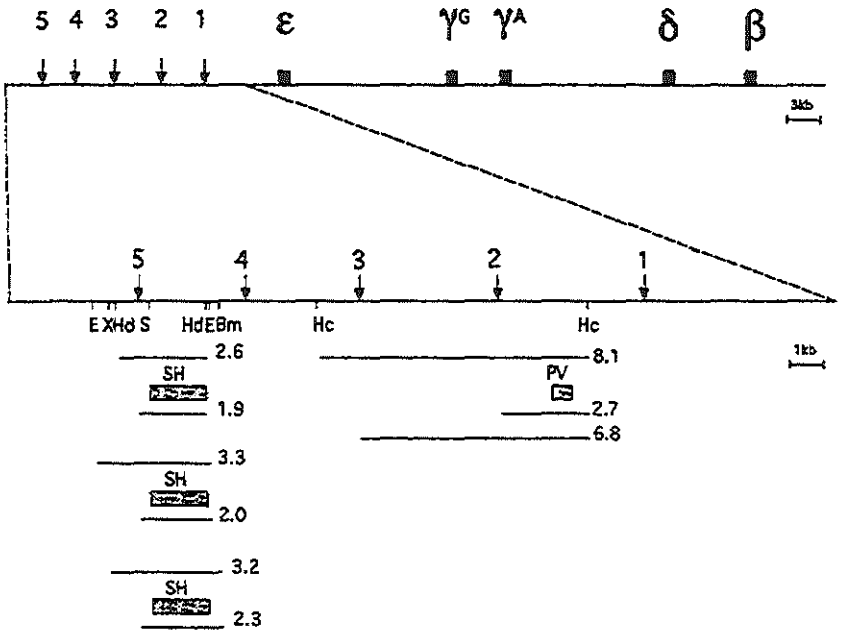

Fig. 1. The human β-globin locus.
The top line shows the human β-globin locus; the bottom part shows the complete LCR. Restriction enzyme sites and fragment sizes of digests used in this study, including predicted HS fragments are shown in kb below. Hybridization probes are shown in shaded boxes. Arrows show hypersensitive sites. E=EcoRI, X=XbaI, Hd=HindIII, Bm=BamHI, HC=HincII, S=SacI. Probes used: PV=450 bp PvuII-EcoRI fragment and SH= 1.3 kb SacI-HindIII fragment.

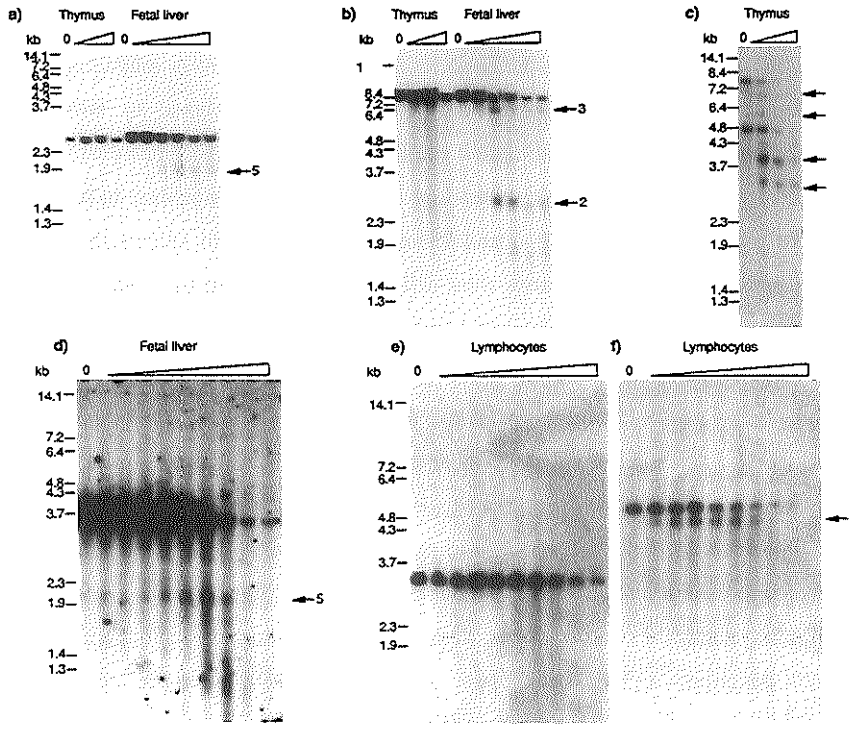

Fig. 2. In vivo DNAseI hypersensitive site mapping.
Nuclei were prepared from 13.5 dpc fetal livers and young (thymus) heterozygous animals of the β-locus line 72 (a, b, c); human fetal liver at 16 weeks of gestation (d) and adult peripheral blood (e, f) and digested with increasing amounts of DNAseI. DNA was digested with HindIII (a), HincII (b), BglII (c), BamHI-EcoRI (d) and BamHI-XbaI (e, f) Southern blotted and probed with SH (a, d, e) and PV (b) (Fig. 1). The wedge above each panel indicates amounts of DNAseI. The lane marked "0" indicates that no DNAseI was added. Arrows indicate DNAseI hypersensitive sites. As control for hypersensitivity, the same thymus DNAseI series in (a, b) were used in (c) and probed a 950 NcoI fragment of the murine vav gene. A duplicate filter of (e) was used in (f) and probed with a 600-bp SacI-HindIII fragment of the human CD2 gene to detect the 3' HS.

The hypersensitive end fragment for HS5 was detected in human fetal liver (Fig. 2 D) but was absent from peripheral blood lymphocytes (Fig. 2 E). The quality of the lymphocyte fade-out was controlled by probing the same filter for the presence of the human CD2 3'HS (Greaves et al., 1989; Fig. 2 F).

Fine mapping of HS5 and *in vitro* DNAseI footprinting.
In order to determine the DNA-protein interactions that gave rise to the erythroid specificity of HS5 found by the *in vivo* DNAseI mapping, the position of HS5 was mapped in more detail

using nuclei prepared from a MEL cell clone stably transfected with 3 copies of minilocus ε (Lindenbaum and Grosveld, 1990; Fig.3). HS5 mapped to a 200-bp core, 2.0 kb 5' of the second EcoRI site marked in Fig. 1. To determine the protein-DNA interactions at HS5, a 270-bp fragment containing the HS5 core was cloned and analyzed by DNAseI *in vitro* footprinting. Several footprints throughout the fragment, as well as hypersensitive sites were found when fetal liver, MEL, adult spleen and liver nuclear extracts were used (Fig.4). FP 1 and 2 were erythroid specific. FP 6 and 9 were also detected in spleen and liver nuclear extracts, although the protected bands were different from those in the erythroid extracts. FP 3, 4, 5 and 7 were detected in all extracts but the intensity and extent of protection differed in erythroid and non-erythroid extracts. Hypersensitive sites were only detected on the sense strand (Fig.4 A). The HS between FP 2 and 3 was only present in fetal liver extracts. The HS between FP 3 and 4 was MEL cell specific and another on FP 3 was present in both erythroid extracts. Based on the DNaseI footprinting data, overlapping oligonucleotides covering the footprinted areas were designed and used in gel retardation experiments. All oligonucleotides tested (except for oligonucleotide 920 encompassing FP 5 and 6; Fig. 4 C) failed to detect protein binding under the conditions that detected most transcription factors (data not shown). Oligo 920 contained a YY1binding site and the gel shift could be competed with a knownYY1 binding oligonucleotide derived from HS2 (Philipsen et al., 1990). We note the presence of a 10 base, purine rich direct repeat with 3 mismatches on FP 1 and 2 that could be a weak GATA binding site (Whyatt et al., 1993). The lack of erythroid factor binding activity was not due to poor extract quality since control experiments with oligonucleotides that contained erythroid and ubiquitous factor recognition sites produced the correct binding patterns (data not shown). It is interesting to note that the position of the 3' HS cleavage determined by in vivo DNAseI digestion co-mapped with the YY1 binding site *in vitro* (Figs.3 and 4 C).

The appearance of HS5 was weaker than the other LCR HS but also appeared to comprise 200-300 bp of sensitivity as we had observed in HS 2-4 (Collis et al., 1990; Philipsen et al., 1993; Pruzina et al., 1991). We therefore tested the activation properties of HS5 in transgenic mice.

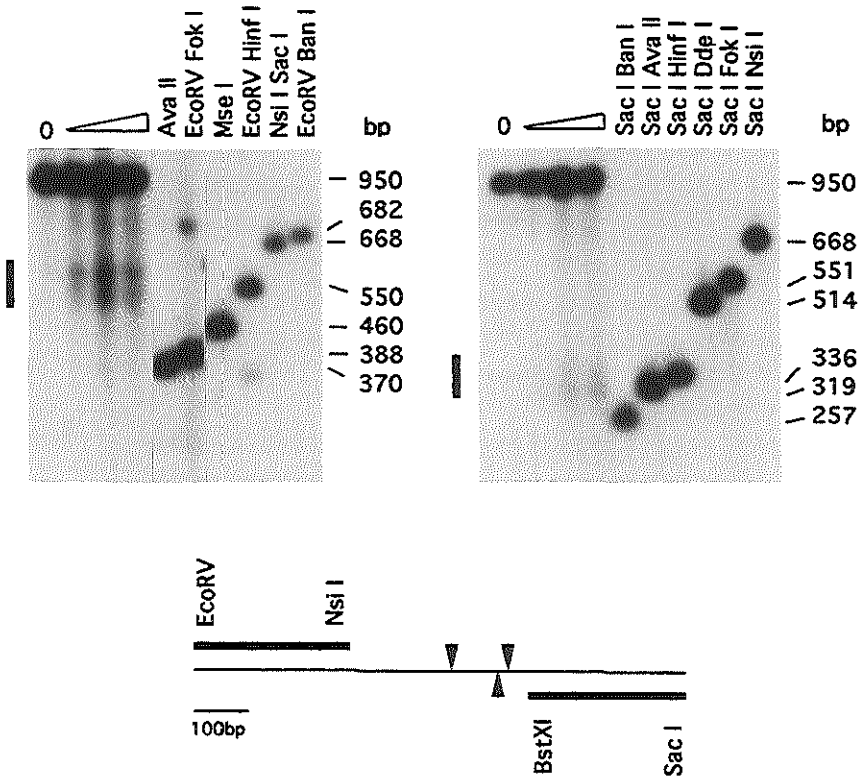

Fig. 3. DNaseI fine mapping of HS5 *in vivo*.
Nuclei were isolated from MEL cells with three copies of a cosmid containing the human β-globin LCR and ε-globin gene (Lindenbaum and Grosveld, 1990), and digested with increasing amounts of DNAseI. DNA was purified, cut with SacI and EcoRV, and Soutern blotted. Internal molecular weight markers were obtained by cutting DNA from the no-DNAseI sample (0) with the restriction enzymes indicated. Probes used are a 271 bp EcoRV-NsiI fragment (5' probe; left panel) and a 300 bp BstXI-SacI fragment (3' probe, right panel). A schematic representation of the location of major DNAseI cleavage sites (arrows) and the positions of the probes is shown below.

Generation of transgenic mice containing HS5 β-globin constructs

The erythroid-specific HS of the β-globin LCR confer high level, position independent, copy number dependent expression to a linked β-globin gene in transgenic mice (Forrester et al., 1989; Grosveld et al., 1990). To determine whether the human HS5 was capable of LCR function, a 2.5 kb HS5 fragment cloned 5' of a β-globin gene was obtained as a 7.5 kb EcoRV fragment and used to generate transgenic mice. Expression levels of β-globin mRNA were analyzed by quantitative S1 nuclease protection with probes for the 5' ends of the human and mouse β-globin globin mRNAs. RNA was obtained from 4 founders and 4 bred lines.

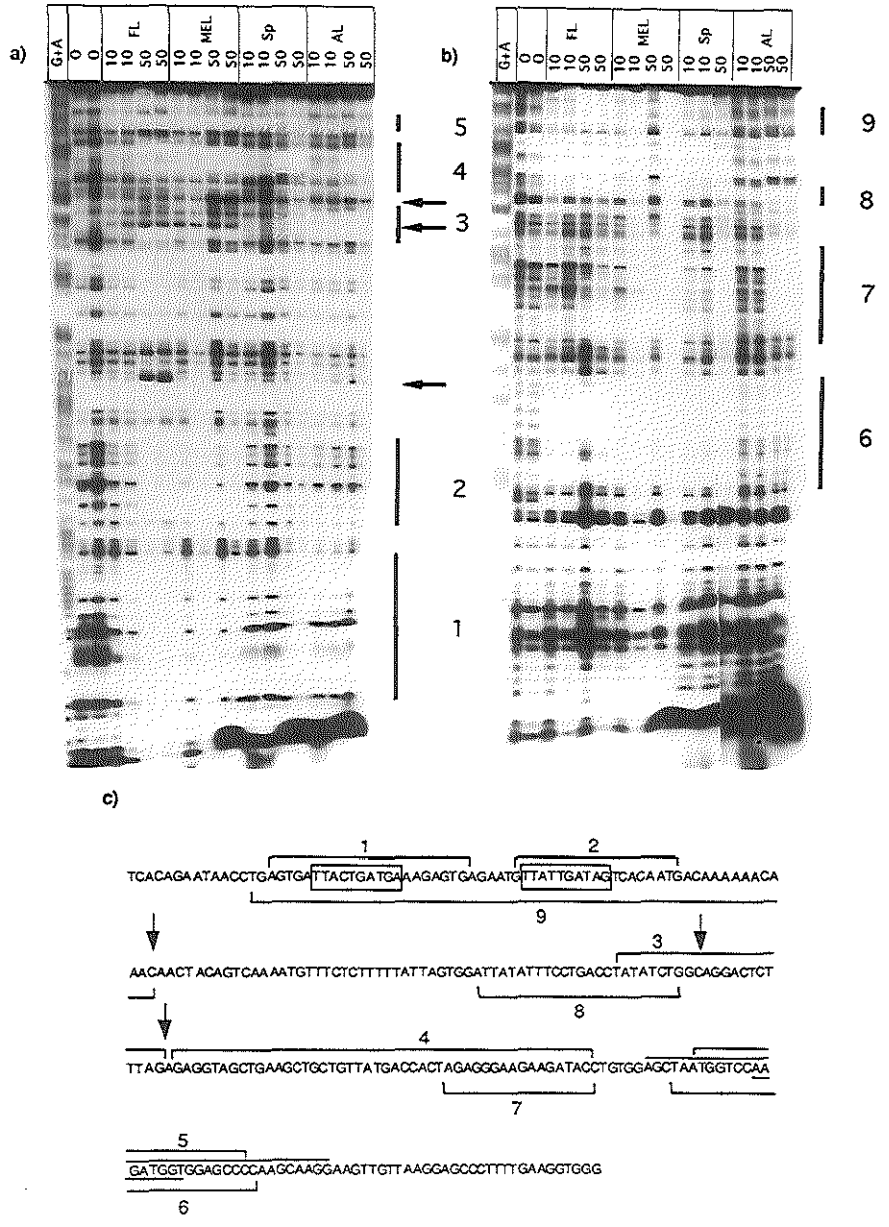

Fig. 4. *In vitro* DNAseI footprinting of the 270 bp BamHI-EcoRI HS5 core fragment.
DNA was preincubated with 10 μg or 50 μg of protein and treated with 0.50 μg and 1.0 μg or 1.0 μg and 2.0 μg of DNAseI respectively. Extracts were derived from 14.5 dpc fetal liver (FL), MEL cells (MEL), adult spleen (Sp) and liver (AL). Sense (a) and antisense strand (b) footprints are marked by bars and numbered. Arrows indicate hypersensitive sites. G+A are Maxam and Gilbert sequencing reactions. (c) Footprints on the sense and antisense strand are indicated by brackets and hypersensitive sites by arrows. The YY1 consensus binding site is underlined and the oligonucleotide 920 used in gel retardation experiments is overlined. The purine-rich 10 bp GATA-like motifs are boxed.

Heterozygous 13.5 dpc fetal liver RNA from line 72 was used as control (Strouboulis et al., 1992). The protected bands were quantified in every test sample using a Phosphor Imager. In all the mice analyzed, the expression per copy of the human β-globin mRNA was very low (0.6 – 2.2 % of the murine β-globin; data not shown). Interestingly these levels are comparable to those obtained from transgenic mice containing the β-globin gene without a linked LCR (Philipsen et al., 1993).

Assays for insulator function

Transgenic experiments were set up to test the possibility that HS5 functions as an insulator, acting within the LCR to override any integration site dependent regulatory influence on a reporter gene. The strategies we employed were similar to those taken by Kellum and Schedl (Kellum and Schedl, 1992) who showed that Specialized Chromatin Structure (SCS) elements from the Drosophila heat shock gene hsp70 are capable of insulating a test gene from chromosomal position effects. We chose the bacterial β-galactosidase gene driven by the mouse hsp68 minimal promoter because this construct is highly susceptible to position effects (Kothary et al., 1989; Tewari et al., 1996). We constructed a plasmid (μZ) in which the expression of the transgene is brought under the influence of the μLCR to direct expression to the erythroid lineage (Fig. 5 A). In a second construct this plasmid is flanked on both sides by the 3kb fragment carrying HS5. If HS5 would function as an insulator it is anticipated that this transgene is expressed in the erythroid lineage but not in any other tissues, irrespective of the site of integration. As a second insulator plasmid we tested the Drosophila scs and scs' elements (Kellum and Schedl, 1992) flanking the μZ construct (μZ/SCS). These gene constructs were introduced in fertilized eggs and the resulting embryos were assayed for β-galactosidase activity at 13 days of gestation. Examples are shown in Fig. 5 C-E. These embryos were chosen to demonstrate erythroid expression, revealed by lacZ staining in the circulation and fetal liver, and ectopic expression revealed by lacZ expression in various other parts of the embryo (Tewari et al., 1996). Ten out of twelve μZ embryos showed staining in the erythroid tissue, but five embryos showed staining in a wide range of tissues such as the nervous system, the limbs and the snout (Fig. 5 C). Two embryos showed ectopic expression of the β-gal gene but did not express in the erythroid tissue possibly due to mosaicism of the transgene (Grosveld et al., 1987). Of the 7 transgenic μZ/HS5 embryos analyzed, all expressed in the erythroid tissue while 3 of these seven animals also showed expression in other tissues (Fig. 5 D). These results suggest that HS5 is not capable of insulating a reporter gene from positive position effects. A similar result was obtained with the μZ/SCS construct,

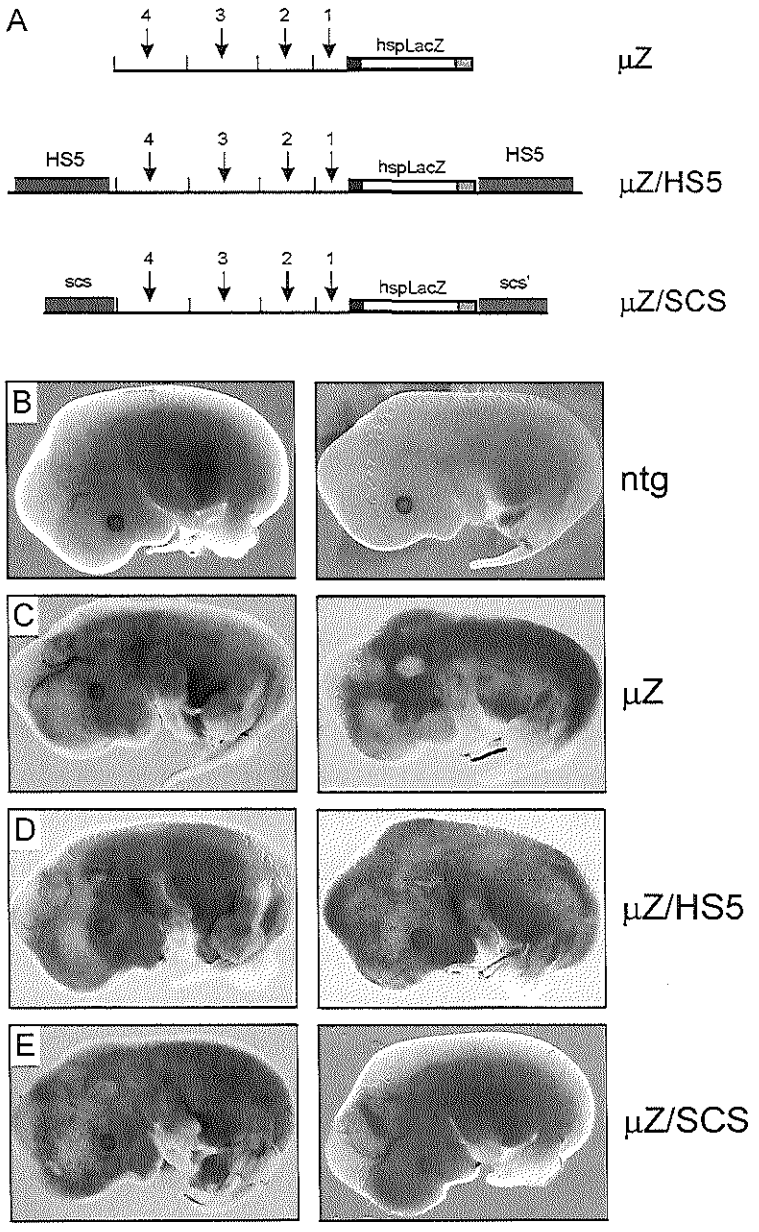

Fig. 5. Position effect assay in transgenic mice.

A. Constructs μZ, μZ/HS5 and μZ/SCS contain the bacterial β-galactosidase gene driven by a 100-bp hsp68 promoter fragment (Kothary et al., 1989). The arrows and numbers indicate the individual hypersensitive sites in the μLCR construct.

B. Examples of transgenic embryos (13.5 dpc) stained for β-galactosidase activity. The transgene is indicated on the right; two different embryos are shown for each construct.. Ntg = non-transgenic control embryos.

again showing that the scs elements do not protect the construct from position effects in transgenic mice (Fig. 5 E).

The second approach we used was the analysis of a β globin gene as the reporter gene cloned upstream of the LCR (Fig. 6 A). In the first two constructs the β globin gene is situated between HS4 and HS5 (5β4) or between a duplicated HS5 (5β5) in the normal sense orientation. The analysis of single and multicopy transgenic animals carrying these constructs (Fig. 6 B and Table 1) demonstrates two points. Firstly the LCR is bi-directional, that is it can fully activate a β globin gene situated upstream of it. Secondly, the presence of HS5 between the LCR (HS1-4) and the gene has no effect. The β globin gene is fully expressed, indicating that in this "enhancer blocking"-type assay HS5 is incapable of negating the effect of the LCR on the gene.

line	Copy no	Expression/copy
5β4 42	1	91 %
5β4 68	1	27 %
5β4 6	3	104 %
5β5 28	1	54 %
5β5 52	1	117 %
5β4 83	2	93 %
5β4 51	5	83 %
1βLCR 29	4	104%
1βLCR 19	3	65 %
1βLCR 25	7	89 %
2βLCR 37	1	67 %
2βLCR 46	2	102 %
2βLCR 11	1	56 %

Table 1. Transgenic lines analyzed in this study. Expression levels of the β-globin are corrected for copy number and shown as percentage of that observed with line 72 (Strouboulis et al., 1992).

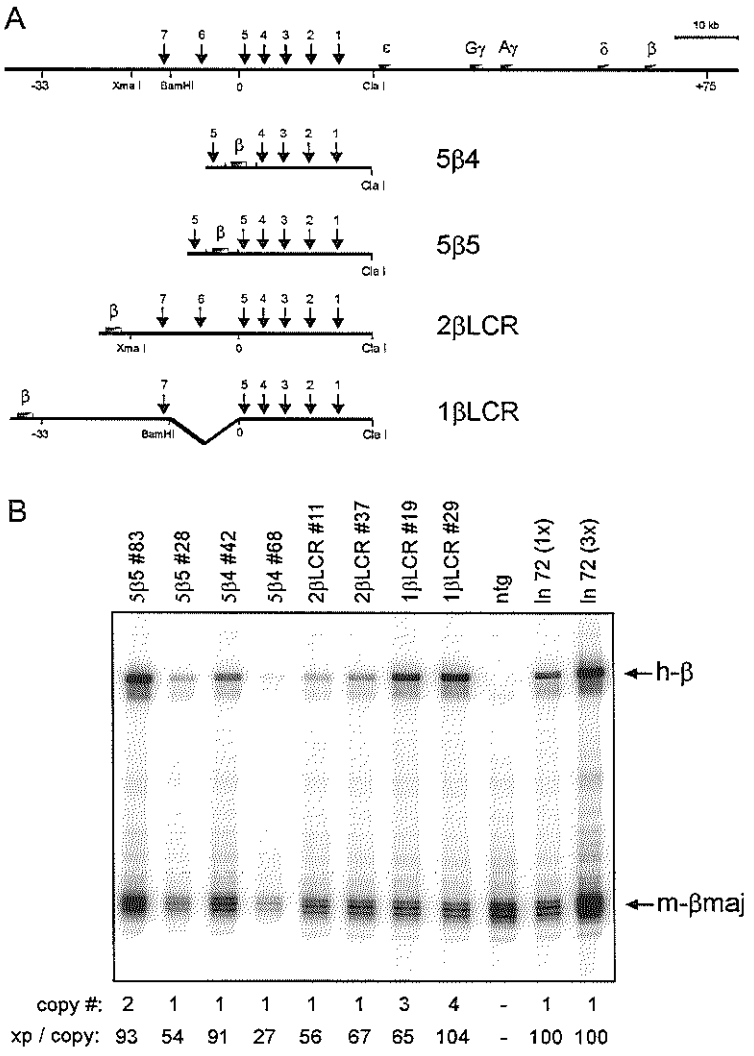

Fig. 6. Quantitative nuclease S1 analysis of HS5 transgenics. Total RNA (approx. 1 μg) from peripheral blood from adult animals was hybridized to a mixture of human β and mouse βmaj end labeled probes adjusted to equal specific activities. Control RNA samples were 1 μg of non-transgenic (nt) and 1 μg (1x) and 3 μg (3x) of 13.5 dpc heterozygous β-locus line 72 adult blood (Strouboulis et al., 1992). Protected fragments specific for the 5' end of the human and mouse β-globin mRNA are indicated. Protected fragments were quantitated and the ratio of human β globin levels over mouse βmaj levels were calculated. The estimated copy number and expression level of the transgene as the percentage of expression of the endogenous mouse βmaj per gene copy is indicated below each lane.

In order to test whether a "border" or insulating sequence is situated further upstream of HS5, we tested two additional constructs. The β globin gene is placed 15 kb upstream of HS5 in 2βLCR (Fig.6 A). In 1βLCR, this 15 kb is replaced with the next 15 kb of genomic sequences at the 5' end of the locus (Fig. 6 A). The analysis showed that the activity of the β globin gene

is not significantly reduced in both 1βLCR and 2βLCR (Fig. 6 B and Table 1). This result suggests that there are no insulating sequences located in the 30 kb upstream of HS5. It also shows that HS5 alone fails to act as an insulator in this transgenic assay. These data are in disagreement with previous reports on insulator properties of the human HS5 (Li and Stamatoyannopoulos, 1994).

Discussion

Here we provide a detailed structural and functional analysis of the hypersensitive site 5 (HS5) of the human β globin LCR. The experiments described in this study are aimed at clarifying the role of this site in the regulation of the β globin domain.

The first set of experiment addressed the question of erythroid specificity and transcriptional activation properties of HS5. We show that HS5 is an erythroid specific site and that its core sequences are bound by erythroid specific activities. However, HS5 does not function as a transcriptional activator. We then turned to investigate whether HS5 functions as an insulator or a boundary. For this purpose we designed two independent series of experiments. The first series tests the ability of HS5 to protect a reporter transgene against position effects at the site of integration (Kellum and Schedl, 1991). The second series of experiments tests HS5's potential to block the influence of a strong activator element on a reporter gene.

HS5 is erythroid specific but it does not have transcriptional activation properties.

The DNaseI mapping and footprint analysis shows that the HS5 core is an erythroid-specific hypersensitive site, in obvious disagreement with the current consensus that considers HS5 to be a non-erythroid hypersensitive site (Li et al., 1999; Tuan et al., 1985). Unlike the previous studies, we have mapped HS5 in human tissues and with nuclear extract derived from these tissues. Therefore we think that the current misconception over the erythroid specificity of HS5 is a consequence of the analysis of this site in cell lines.

We also demonstrate by footprint analysis that HS5 is bound by erythroid-specific activities, although we have yet to establish the identity of these factors. We suggest that these activities are part of a multiprotein complex that is only stable at high protein concentration such as used in the footprint analysis. This may explain our failure to detect these factors in bandshift experiments.

In conclusion, we have determined that HS5 is bound by activities present exclusively in erythroid extracts and that these activities do not have a direct function in transcriptional

activation. Further studies are needed to characterize these factors. We anticipate that unraveling the identity of these factors may provide the necessary clues to the function of HS5 in the β globin locus.

Does HS5 function as an insulator?

The high-level, position independent expression observed with full LCR/globin constructs can be explained by two non-mutually exclusive mechanisms. The LCR may act as a dominant regulator overriding any integration site-dependent influences and/or the LCR may insulate the transgene from such position effects. It has been reported that HS5 is involved in the latter hypothetical function using transfection systems involving selection (Chung et al., 1993). We therefore set up transgenic experiments to test this possibility *in vivo* without a selection procedure. The strategies we employed were similar to those taken by Kellum and Schedl who showed that Specialized Chromatin Structures (scs) elements derived from the Drosophila heat shock locus are capable of insulating a test gene from chromosomal position effects (Kellum and Schedl, 1991; Udvardy and Schedl, 1993). In our constructs we used the bacterial β-galactosidase gene driven by the mouse hsp68 minimal promoter as a reporter which is known to be highly susceptible to position effects (Kothary et al., 1989; Tewari et al., 1996). In the two test constructs the reporter gene is flanked either by HS5 (μZ/HS5) or by the scs elements (μZ/SCS). If HS5 would function as an insulator, it is anticipated that the transgene will be expressed in the erythroid lineage but not in other tissues. In contrast, in the absence of flanking HS5 fragments the transgene will be expressed in the erythroid lineage, and also in other tissues depending on the site of integration. The same rationale applies to the μZ/SCS constructs. Here we have shown that neither HS5 nor the scs elements are able to protect the test gene from position effects in transgenic mice. Furthermore, the analysis of two other constructs (5β4 and 5β5) in transgenic mice demonstrated that the LCR could activate a β globin gene in a bi-directional manner. This result seems quite different from that obtained by Tanimoto et al. (Tanimoto et al., 1999) who show that the LCR activity is orientation dependent. In a very elegant experiment they demonstrate that the inversion of the LCR (HS1-5) within a YAC containing the β globin locus has a dramatic effect on the expression of all the β globin-like gene in transgenic mice. Conversely, in the 5β4 and 5β5 test constructs we show that the LCR can fully activate a β globin gene relocated at a 5' position. Therefore we conclude that in our assay system, a reporter gene placed 5' of the LCR is fully activated and that HS5 is unable to negate the functional activation of the LCR on the reporter gene. Directionality of the LCR, as observed by Tanimoto et al., may be a consequence of the

presence of other regulatory elements present in the large YAC constructs used in their studies.

At present it is unclear why different results were obtained with the chicken and human 5'HS element (Chung et al., 1993). Since human HS5 is only hypersensitive in erythroid cells, it might function as an insulator in these cells exclusively. The chicken and human elements may not be functionally equivalent in this respect. Alternatively, the different results could be due to the assay systems used for each study: Chung et al used a drug selection system while we examined β globin expression in transgenic animals. Another factor that could be of importance is the strength of the properties of the activating element: Chung et al used only part of the LCR while we used the entire LCR. Another difference is that we used HS5 not only in artificial constructs but also in its natural position relative to the LCR. That the properties of the enhancer might be relevant is suggested by experiments on the scs and scs' elements (Vazquez and Schedl, 1994) and on su(Hw) elements (Roseman et al., 1993). These studies indicate that the efficacy of enhancer blocking by these elements depends on the strength of the enhancer.

It is also interesting to note that the Drosophila scs elements do not appear to have any effect in the mouse with respect to position effects, whereas the chicken element was active in both mammalian cells and *Drosophila* (Chung et al., 1993).

Sequences upstream of HS5 do not display insulator properties.

The second series of constructs (5β4, 5β5, 1βLCR and 2βLCR) test HS5 in an enhancer blocking assay. In this assay, if HS5 functions to restrict the action of a strong enhancer, that is if HS5 has insulator properties, then one should observe either a decrease or no expression of the reporter gene.

When the β globin reporter gene was positioned 16 kb 5' of HS5 (2 βLCR) or linked to the LCR via 21 kb of genomic sequences that lie 10 kb 5' of HS5 (1 βLCR), the expression levels of the β globin gene did not change significantly. Since we also show that the β globin gene is fully expressed when positioned directly 5' of the LCR and of HS5 (5β4, 5β5), we conclude HS5 is not an insulator and that 30 kb sequences upstream of HS5 do not have insulator function.

Chicken HS4 marks a boundary between heterochromatin and euchromatin in the chicken locus (Prioleau et al., 1999). Recent experiments performed in the human and mouse β–globin loci have shown that deletions involving the LCR and encompassing HS5 result in the downregulation of the β globin genes, but not in the loss of nuclease sensitivity of the locus (Bender et al., 2000; Epner et al., 1998; Reik et al., 1998). Thus, the LCR is necessary

for the activation of the genes but HS5 does not appear to function as a boundary between heterochromatin and euchromatin.

In conclusion, we have demonstrated that the HS5 of the β globin LCR is erythroid specific but it does not have border or insulator functions. While some experiments indicate that higher eukaryotes have genetic elements that can restrict the action of enhancers and silencers, and perhaps delimit regulatory domains, it is not yet clear whether such elements actually correspond to the boundaries of chromatin domains (Bell and Felsenfeld, 2000; Bell et al., 2001; Hark et al., 2000; Kellum and Schedl, 1992; Kmita et al., 2000; Roseman et al., 1993; Webber et al., 1998; Zhou and Levine, 1999). All of these functional assays for chromatin boundaries were based on the proposal, from the cytological studies done decades ago, that chromatin domains correspond to units of autonomous genetic function (Benyajati and Worcel, 1976; Sedat and Manuelidis, 1978).Whether this proposal is correct has yet to be demonstrated.

Hypersensitive site mapping.

Human lymphocytes were obtained from blood following standard procedures. Nuclei from mouse, human tissues and cultured cells were prepared essentially as described (Forrester et al., 1990). Briefly, tissues were dispersed in ice cold phosphate buffered saline (PBS) with an Ultraturrax homogenizer. Cells were washed in PBS and pelleted at 600g for 5 minutes at 4 °C. Cells were resuspended in ice cold reticulocyte standard buffer (RSB: 10mM TrisHCl pH 7.4; 10mM NaCl; 3mM $MgCl_2$) and Nonidet-P40 was added to 0.1% (v/v) (lymphocytes, 0.2%). Cells were dounced 10-20 times with pestle B. Ten milliliters of NP-40 treated cells were resuspended in 50 ml of RSB and pelleted at 600g for 5 minutes at 4 °C. The nuclei obtained were resuspended at a final nucleic acids concentration of 1 mg/ml and $CaCl_2$ was added to 3mM. DNAseI digestion was carried out on 100-µl aliquots of nuclei. DNAseI concentrations ranged from 0 to 0.24 µg/ml. Digestions were carried out at 37 °C for 4 minutes (fetal liver, thymus and MEL cells) and 8 minutes (human lymphocytes). Reactions were stopped by adding 100 µl of 2x stop buffer (0.6M NaCl, 20mM TrisHCl pH 8.0, 10mM EDTA, 1% SDS) and digested with Proteinase K (100µg/ml) at 55 °C for one hour. After purification, DNA was digested to completion with the appropriate restriction enzyme and analyzed by Southern blotting.

Protein-DNA binding interaction.

Preparation of nuclear extracts from MEL cells, mouse spleens, adult and fetal livers was performed as described (Wall et al., 1988). A DNA fragment containing HS5 was prepared by PCR amplification of GSE 826 with oligonucleotides

5'-GCCGCCGAATTCCCACTTCAATCAAAAGGGC-3' and

5'-TCGCCGGATCCTCACAGAATAACCTGACTCA-3'.

The amplification products were cleaved with EcoRI and BamHI and cloned in cohesive ended pBluescript (Stratagene) to generate GSE 2122L. The 270 bp EcoRI-BamHI fragment was end labeled by T4 polynucleotide kinase and (γ ^{32}P) ATP. In vitro footprinting and EMSA were performed as described (deBoer et al., 1988; Wall et al., 1988).

HS5 as an activator element in transgenic mice.

To test the activation properties of HS5 in transgenic mice, a 7.5 kb containing the HS5 (2.5 kb) upstream of the human β-globin gene (5.0 kb) was obtained by digestion of GSE 2024 with EcoRV. Transgenic mice were generated by injecting the purified EcoRV fragment in

the male pronucleus of fertilized oocytes and transferred in the oviducts of pseudopregnant (CBA x C57B6) F1 female mice.

Insulator assay constructs.

Three constructs, μZ, μZ/HS5 and μZ/SCS, were prepared to test the properties of HS5 as an insulator. μZ was constructed by linking the bacterial β-galactosidase gene driven by the mouse hsp 68 minimal promoter (Holmgren et al., 1981; Kothary et al., 1989) to the μLCR plasmid (Needham et al., 1992; Tewari et al., 1996). μZ/HS5 was assembled by flanking μZ with the 3kb EcoRI HS5 fragment whereas in μZ/SCS the SCS and SCS' fragments (Kellum and Schedl, 1992) were positioned 5' and 3' with respect to μZ (Fig.).

Enhancer blocking assay constructs.

Four constructs, namely 5β4, 5β5, 1βLCR and 2βLCR, were prepared to test HS5 in the enhancer blocking assay (Fig.)

In 5β4 the 5 kb BglII fragment containing the human β-globin gene and surrounding regulatory regions (Antoniou et al., 1988) was cloned into the unique EcoRI site of a 5.8 kb SalI/SfiI subfragment derived from pLCR. The pLCR plasmid has the 22 kb genomic sequences upstream of the natural ClaI site of the ε-globin gene and contains the HS1-5 of the β-globin LCR. The modified 10.8 kb SalI/SfiI fragment was used to replace the original 5.8 kb SalI/SfiI fragment in the pLCR plasmid to generate 5β4.

5β5 was constructed by cloning an 8 kb SalI/XhoI fragment from 5β4 in the unique SalI site of pLCR. This 8 kb SalI/XhoI fragment contains HS5 linked to the β-globin gene.

5β4 and 5β5 were released from the vectors by digestion with SalI and ClaI. The purified fragments were used to generate transgenic mice as described (Strouboulis et al., 1992).

The last two constructs, 1βLCR and 2βLCR were prepared as follows. First, two plasmids were generated from cosmid HG4 that contains about 38 kb of genomic sequences upstream of HS3 cloned into pTCF (Taramelli et al., 1986). To prepare βHG4A, HG4 was digested to completion with BamHI and re-circularized. This step resulted in the deletion of 16 kb from the 3' end of HG4 and generated HG4A that retains 22 kb of genomic sequences 10 kb upstream from HS5. Following a SalI partial digestion, HG4A was cut to completion with ClaI. Finally the 5 kb BglII β-globin fragment (see above) was cloned as ClaI/XhoI in HG4A to generate βHG4A. The second plasmid, βXmaHG4A, was constructed by subcloning the 5 kb BglII β-globin as a XmaI restriction fragment into the unique XmaI site of HG4B. HG4B was prepared by re-circularizing HG4 after digesting it to completion with XmaI. In HG4B

about 16 kb at the 5' end of HG4 has been deleted. The resulting plasmid retains about 23 kb of contiguous genomic sequences upstream of HS3. Introducing a SalI linker in the BspEI site located between HS3 and HS4 further modified βXmaHG4B. SalI digestion of the linkered plasmid resulted in the deletion of a 1.4 kb at the 3' end of βXmaHG4B. This deleted plasmid was re-circularized resulting in βXmaHG4B.1. A SalI linker was also introduced in the corresponding BspEI site in the pLCR plasmid. Again the linkered plasmid was digested with SalI and circularized to generate pLCR2. Finally, to generate 2βLCR, βXmaHG4B.1 and pLCR2 were released from the vectors by digestion with SalI and ClaI. The fragments were joined at the SalI site and ligated into the ClaI site of pJBF (Grosveld et al., 1982). The same cloning strategy was used to construct 1βLCR from βHG4A and pLCR. Transgenic mice were generated with the purified ClaI inserts from 1βLCR and 2βLCR.

Analysis of transgenic mice.

DNA was obtained from tail biopsies of 10-day old pups and placentas of 13.5 dpc embryos. DNA samples were screened for the presence of the transgene by Southern blotting. Transgene copy number was determined via quantitation of the hybridization signals from probes derived from the transgene and that of an endogenous gene as described previously (Strouboulis et al., 1992).

β-Galactosidase staining.

Embryos were isolated at 13 dpc, washed extensively with PBS/0.02% NP-40 and fixed in 1% Formaldehyde, 0.2% gluteraldehyde, 2mM $MgCl_2$, 5mM EGTA, 0.02% NP-40 in PBS for 60-90 minutes at 4 °C. The fixative was removed by washing the embryos in PBS/ 0.02% NP-40 three times for 30 minutes each time. The embryos were stained with X-gal (1 mg/ml) in a phosphate buffered saline solution containing 5 mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 0.01% sodium deoxycholate and 0.02% NP-40.

Preparation of RNA and S1nuclease protection.

RNA was prepared from adult blood as described. Probes used and quantitative S1 nuclease analysis were as described (Zafarana et al., 2000).

References

Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988) The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *Embo J*, **7**, 377-84.

Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482-5.

Bell, A.C., West, A.G. and Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science*, **291**, 447-50.

Bender, M.A., Mehaffey, M.G., Telling, A., Hug, B., Ley, T.J., Groudine, M. and Fiering, S. (2000) Independent formation of DnaseI hypersensitive sites in the murine beta- globin locus control region. *Blood*, **95**, 3600-4.

Benyajati, C. and Worcel, A. (1976) Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. *Cell*, **9**, 393-407.

Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*, **74**, 505-14.

Collis, P., Antoniou, M. and Grosveld, F. (1990) Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *Embo J*, **9**, 233-40.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *Embo J*, **7**, 4203-12.

Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993) Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *Embo J*, **12**, 127-34.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G. and Groudine, M. (1998) The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*, **2**, 447-55.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*, **4**, 1637-49.

Forrester, W.C., Novak, U., Gelinas, R. and Groudine, M. (1989) Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci U S A*, **86**, 5439-43.

Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res*, **15**, 10159-77.

Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, **83**, 1359-63.

Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res*, **18**, 3503-8.

Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989) Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell*, **56**, 979-86.

Grosveld, F., Greaves, D., Philipsen, S., Talbot, D., Pruzina, S., deBoer, E., Hanscombe, O., Belhumeur, P., Hurst, J., Fraser, P. and et al. (1990) The dominant control region of the human beta-globin domain. *Ann N Y Acad Sci*, **612**, 152-9.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-85.

Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H. and Flavell, R.A. (1982) The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.*, **10**, 6715-32.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486-9.

Holmgren, R., Corces, V., Morimoto, R., Blackman, R. and Meselson, M. (1981) Sequence homologies in the 5' regions of four Drosophila heat-shock genes. *Proc Natl Acad Sci U S A*, **78**, 3775-8.

Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, **64**, 941-50.

Kellum, R. and Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol*, **12**, 2424-31.

Kmita, M., Kondo, T. and Duboule, D. (2000) Targeted inversion of a polar silencer within the HoxD complex re- allocates domains of enhancer sharing. *Nat Genet*, **26**, 451-4.

Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A. and Rossant, J. (1989) Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development*, **105**, 707-14.

Li, Q. and Stamatoyannopoulos, G. (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood*, **84**, 1399-401.

Li, Q., Zhang, M., Duan, Z. and Stamatoyannopoulos, G. (1999) Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics*, **61**, 183-93.

Lindenbaum, M.H. and Grosveld, F. (1990) An in vitro globin gene switching model based on differentiated embryonic stem cells. *Genes Dev*, **4**, 2075-85.

Long, Q., Bengra, C., Li, C., Kutlar, F. and Tuan, D. (1998) A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region. *Genomics*, **54**, 542-55.

Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F. and Hollis, M. (1992) LCR/MEL: a versatile system for high-level expression of heterologous proteins in erythroid cells. *Nucleic Acids Res.*, **20**, 997-1003.

Ogilvy, S., Elefanty, A.G., Visvader, J., Bath, M.L., Harris, A.W. and Adams, J.M. (1998) Transcriptional regulation of vav, a gene expressed throughout the hematopoietic compartment. *Blood*, **91**, 419-30.

Philipsen, S., Pruzina, S. and Grosveld, F. (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *Embo J*, **12**, 1077-85.

Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) The beta-globin dominant control region: hypersensitive site 2. *Embo J*, **9**, 2159-67.

Prioleau, M.N., Nony, P., Simpson, M. and Felsenfeld, G. (1999) An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid- specific folate receptor gene. *Embo J*, **18**, 4035-48.

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. and Philipsen, S. (1991) Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Res.*, **19**, 1413-9.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol*, **18**, 5992-6000.

Roseman, R.R., Pirrotta, V. and Geyer, P.K. (1993) The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *Embo J*, **12**, 435-42.

Sedat, J. and Manuelidis, L. (1978) A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol*, **42**, 331-50.

Strouboulis, J., Dillon, N. and Grosveld, F. (1992) Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev*, **6**, 1857-64.

Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) Detailed analysis of the site 3 region of the human beta-globin dominant control region. *Embo J*, **9**, 2169-77.

Tanimoto, K., Liu, Q., Bungert, J. and Engel, J.D. (1999) Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature*, **398**, 344-8.

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. and Grosveld, F.G. (1986) Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res*, **14**, 7017-29.

Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996) The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development*, **122**, 3991-9.

Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A*, **82**, 6384-8.

Udvardy, A. and Schedl, P. (1993) The dynamics of chromatin condensation: redistribution of topoisomerase II in the 87A7 heat shock locus during induction and recovery. *Mol Cell Biol*, **13**, 7522-30.

Vazquez, J. and Schedl, P. (1994) Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. *Embo J*, **13**, 5984-93.

Wall, L., deBoer, E. and Grosveld, F. (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev*, **2**, 1089-100.

Webber, A.L., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (1998) Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature*, **391**, 711-5.

Whyatt, D.J., deBoer, E. and Grosveld, F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *Embo J*, **12**, 4993-5005.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, **377**, 209-13.

Yu, J., Bock, J.H., Slightom, J.L. and Villeponteau, B. (1994) A 5' beta-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene*, **139**, 139-45.

Zafarana, G., Rottier, R., Grosveld, F. and Philipsen, S. (2000) Erythroid overexpression of C/EBPgamma in transgenic mice affects gamma- globin expression and fetal liver erythropoiesis. *Embo J*, **19**, 5856-63.

Zhou, J. and Levine, M. (1999) A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo. *Cell*, **99**, 567-75.

Chapter 3

Characterization and purification of NF-E6

Gaetano Zafarana, Ali Imam, Frank Grosveld and Sjaak Philipsen

Characterization and purification of NF-E6

Introduction

Tissue specificity and timing of expression of the β-globin genes is achieved, to a great extent, through interactions between the regulatory sequences surrounding each gene and the LCR. In the presence of the LCR, the minimal promoter of the adult β-globin gene is necessary for high level transcription of a reporter gene in mouse erythroleukemia (MEL) cells. The TATA box, together with both the CACC box and the CCAAT box motifs can stimulate activity 10-fold higher than the TATA box alone (Antoniou and Grosveld, 1990). Thus, the CCAAT box and the CACC element are important in regulating the β-globin promoter. Undeniably, factors binding to these elements also contribute to the proper temporal and topological expression of this gene. For example, point mutations that destroy the consensus of the β-globin CACC box or the absence of EKLF, an erythroid-specific factor that binds to these sequences, abolish the expression of the β-globin gene (Thein, 1993; Wijgerde et al., 1996). However, no spontaneous mutations of the β-globin CCAAT box have been reported. Although it is not clear what is the significance for the absence of natural mutation in the CCAAT box of the β- globin gene, mutations that cluster around the CCAAT boxes of the γ-globin genes are responsible for the extemporaneous expression of fetal globin in adult blood (see (Berry et al., 1992). This benign condition is known as hereditary persistence of fetal haemoglobin (HPFH). Thus, it is likely that factors cognate with the γ-globin CCAAT boxes have a distinctive role in the activation of the fetal γ-globin promoters, whereas the CCAAT-binding factors of the adult β-globin promoter may function to complement the obviously necessary CACC factors. Nonetheless, the dual requirement of both CACC and CCAAT boxes for full activation of the β-globin promoter is in the least suggestive that the function of CCAAT factors is somehow instrumental in determining the affinity of the promoter for the LCR (Antoniou and Grosveld, 1990).

Very interesting to note is that the CCAAT consensus sequences of both the γ- and the β-globin genes are very well conserved (Karlsson and Nienhuis, 1985). Their similarity in sequence is also reflected by the fact that these motifs bind, though with different affinities, the same three major complexes in nuclear extracts of erythroid cells. These complexes have been identified as CP1/NF-Y, GATA-1 and NF-E6 (Berry et al., 1992; Delvoye et al., 1993; Wall et al., 1996).

Because of differences in the binding of CP1/NF-Y and GATA-1 in the HPFH promoter of the γ-globin gene these factors have been considered as candidates for a role in γ suppression (Berry et al., 1992; Ronchi et al., 1996; Superti-Furga et al., 1988). In spite of this phenomenon, neither of these two candidates has been shown to be crucial in the HPFH phenotype. The binding of NF-E6 is also not affected by point mutations that result in HPFH indicating that this factor is probably not a repressor of the γ-globin gene (Ronchi et al., 1996). On the contrary, the persistence of NF-E6 on the HPFH promoter may in fact indicate that this factor is in some way implicated in the activation rather than the repression of the fetal genes.

Yet, a more interesting avenue to explore is the relationship between the ability of NF-E6 to bind both fetal and adult globin CCAAT boxes and the regulation of these globin genes during development. Several studies have indicated that the developmental switch from fetal to adult globin is correctly attained only when the adult β-globin gene is *in cis* with the γ-globin gene or in its normal context within the human β-globin gene locus (Hanscombe et al., 1991). Further, it has been shown that the underlying mechanism in operation during the switch is one where the γ- and β-globin genes compete for the LCR (Wijgerde et al., 1995). Thus any factor that can increase the affinity of the promoters for the LCR can in principle favour a productive interaction resulting in sustained transcription of the linked gene (Wijgerde et al., 1996).

NF-E6 is not present in non-erythroid HeLa and F9 cells nuclear extracts but instead binds the CCAAT box in mouse 10.5 days yolk sac as well as MEL nuclear protein extracts, which are regarded as adult-like erythroid cells. On the basis of this finding, NF E6 was first described as an erythroid-restricted factor (Berry et al., 1992).

The work presented in this study addresses two main questions. The first regards the tissue distribution of NF-E6. Further characterization of the NF-E6 binding activity in a wide range of mouse adult tissues can be used to deduce its possible function. The second and clearly more important issue regards the identification of the gene product of NF-E6. Our main interest is to identify the erythroid complex that binds to the globin CCAAT boxes and eventually to study the contribution of this activity to the regulation of the globin genes. For this purpose we have devised a procedure to purify the native complex from MEL cells by means of chromatographic techniques. Here we report the almost homogeneous purification of the NF-E6 activity from MEL cells. The protocol devised is very efficient and therefore we anticipate that this procedure can be successfully applied to purify the limiting amounts of the complex from embryonic tissues that are active during the globin switch.

Results

Tissue distribution of NF-E6.

An activity that binds to the human β CCAAT oligonucleotide and migrates like NF-E6 is present in the nuclear extracts of all the tissue analysed. The NF-E6 shifts are not affected by a GATA-1 oligonucleotide but specifically competed by the rat β CCAAT probe which binds NF-E6 avidly in MEL cell extracts (Pavlovic et al., 1999). Therefore we conclude that NF-E6 is not a factor restricted to erythroid cells as originally postulated, but it appears to be a ubiquitous activity. The activity is highest in both MEL and in the preB cell line 70Z/3 (Fig.1). This could be due to the homogeneity of the nuclear extracts prepared from the cell lines versus the tissue preparations or because NF-E6, although ubiquitous, may be restricted to certain cell types in the various organs. It is also interesting to note that in the two liver samples analysed, the rat β CCAAT probe competes out several bands. However, only one of the bands migrates like NF-E6, whereas the other complexes have a lower mobility. Thus it is possible that NF-E6 dimerises with several other proteins in the liver or that other CCAAT factors with the same binding specificity as NF-E6 are present in liver extracts. No difference in either binding or migration is apparent when comparing extracts derived from 3weeks old with those of adult liver.

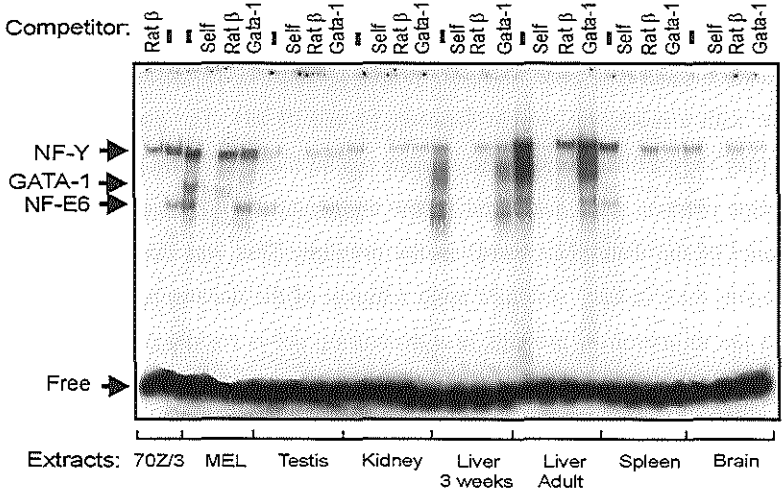

Fig. 1. Tissue distribution of NF-E6.
Five µg of the indicated nuclear extracts were incubated with an oligonucleotide probe covering the CCAAT box of the human beta globin gene. Protein-DNA complexes were fractionated on a 4% native PAA gel. Competitor oligonucleotides were added at a 100-fold molar excess prior to the binding reactions.

Molecular weight of NF-E6

The molecular weight of a DNA binding protein is difficult to estimate from its migration on a band shift gel principally because the shape and charge of the DNA-protein complex formed in the binding assay have a great influence on the mobility of the complex in a native gel. Therefore we sought to analyse extracts which had been first separated according to their molecular weight by two independent methods. The first protocol relied on the capability of NF-E6 to re-fold correctly and still be able to bind DNA after it had been separated on a denaturing gel. The second separation was much gentler as it involved fractionating the activity under native conditions by gel filtration.

NF-E6 renatures form the third slice of the denaturing gel (Fig 2B). The protein extracts derived from this gel slice have molecular weights that range from 19 to 25 kD (Fig 2A). Faster migrating complexes are instead derived from the second gel slice. The nature of these complexes is not known. However, we have noticed that their abundance varies depending on the extract preparation. It is likely that these complexes are partially degraded NF-E6 molecules that retain their DNA binding domain. Furthermore, a slow migrating complex which runs at the position of NF-Y renatures from the fifth gel slice which contains proteins between 30 and 40 kD. Mixing experiments were also performed on the renatured fractions to assess whether any of the activities that bind on the CCAAT box oligonucleotide were derived from two polypeptides of different molecular weight. For the mixing experiment half the amount of each extract was used. However neither difference in binding was observed nor extra complexes were formed (Fig 2B). Therefore it appears that NF-E6 derives from one or more polypeptides with a molecular weight between 19 and 25 kD. The precise molecular weight of NF-E6 may vary a little because the extracts run on the denaturing gel were mildly denatured. The crude extracts were denatured at low temperature with limiting amounts of DDT to retain some of the disulphide bonds in the proteins. This was done to allow the polypeptides to regain their tertiary structure upon removal of the denaturing agents during dialysis. Although the migration of proteins in a denaturing gel is mainly determined from their size because the overall charge is due to the SDS bound to them, we cannot exclude that the presence of disulphide bridges may slightly change the mobility of NF-E6.

To further corroborate the data obtained from the denaturation/renaturation experiment, crude MEL nuclear extracts were separated in their native state by gel filtration and the fractions obtained were subjected to bandshift analysis. As can be seen in Fig. 2C, the NF-E6 binding activity peaks in fraction 18. This elution pattern predicts a protein of an apparent molecular

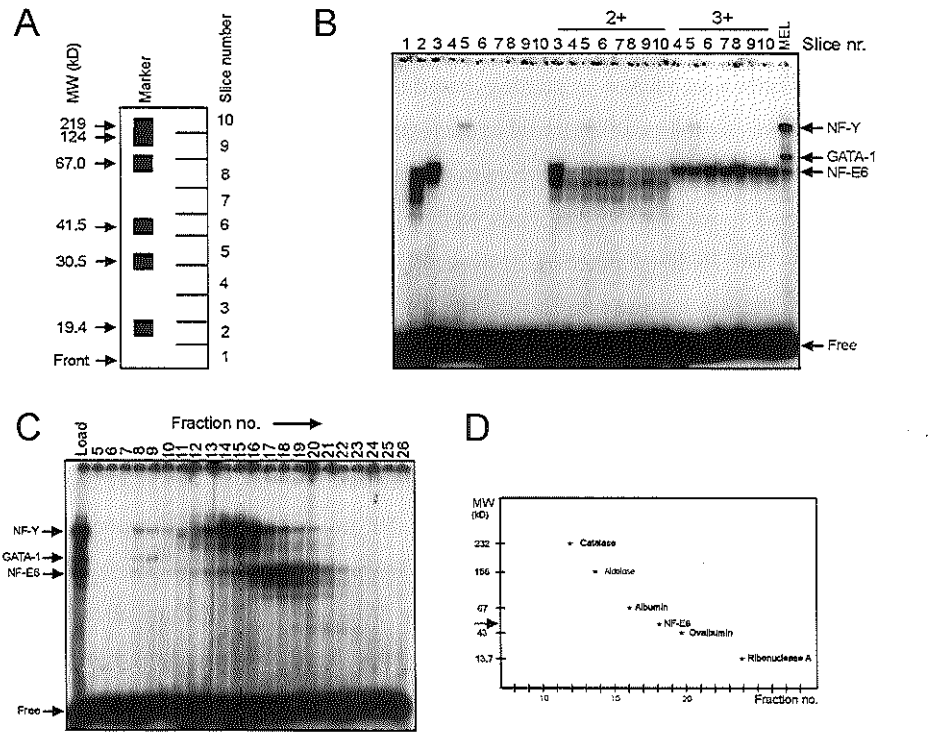

Fig. 2. Determination of the molecular weight of NF-E6.
A). Diagram showing the outline of the SDS-PAGE used to fractionate crude MEL nuclear extract. B). Bandshift assay of the renatured protein fractions obtained from the gel slices depicted in A. Lanes 2+ and 3+ are mixing experiments with the indicated fractions. C). Bandshift analysis of fractions obtained from S 200 gel filtration of crude MEL cell nuclear extracts. D.) Calibration of the S 200 gel filtration column with proteins of known molecular weight. The empirically determined native molecular weight of NF-E6 is indicated.

weight of around 40 to 50 kD. In the light of this result and together with the data obtained from the previous size assessment we predict that NF-E6 is composed from two polypeptides of about 19 to 23 kD. Interestingly, we observe that the extracts in fraction 8 contain NF-Y, GATA-1 and NF-E6. It is very tempting to speculate that the three major activities that bind to the β-globin CCAAT box exist as a pre-formed complex in MEL nuclear extract.

Purification of NF-E6

During the characterisation of NF-E6 we have shown that this activity is ubiquitously expressed, hence its purification can in principle be attempted from several organs of the mouse. Our results also raise the possibility that NF-E6 is composed of two subunits and that at least in the liver several species are present. This suggests that NF-E6 can dimerise with

tissue specific partners and thus form a unique combination of factors with specific functions in a determined tissue. Our main interest is the identification of the erythroid complex. Therefore MEL cells were chosen as starting material for the purification of NF-E6 because they are a homogeneous source of erythroid extracts that can be isolated in large quantities.

Pilot trials were set up with different matrices to determine the binding profile of NF-E6 (data not shown). It was found that NF-E6 binds well to Heparin Sepharose, a commonly used matrix to purify DNA-binding proteins. However, the elution profile of NF-E6 was very broad thus resulting in a modest purification step. In addition, NF-Y and GATA-1 also co-purified with NF-E6, thus making Heparin Sepharose not an attractive choice. Binding tests on ionic matrices suggested that several fold purification could obtained be from the fractionation of NF-E6 on the basis of its charge. NF-E6 binds strongly to S-Sepharose, an ion exchange matrix with a strong positive ionisation potential. In contrast, GATA-1 and NF-Y do not bind efficiently to this matrix. Furthermore, NF-E6 eluted with salt concentrations between 0.6 and 0.8 M and could be recovered in a sharp peak. In a single purification step about 80% of the binding activity was recovered in 5% (25 mg) of the initial amount of nuclear extract. The extract, though enriched in NF-E6, still contained some NF-Y (Fig. 3) and obviously many other proteins that cannot be detected by the bandshift assay. To further purify NF-E6 we exploited the fact that NF-E6 is a sequence-specific DNA-binding protein. The extracts were fractionated on a DNA affinity matrix constructed with the rat β CCAAT oligonucleotide. The rat β CCAAT sequence was used instead of the human CCAAT box because the rat β oligonucleotide binds NF-E6 avidly but it doesn't bind NF-Y In this purification step not only NF-Y was effectively removed (Fig. 3) but also the total amount of protein extract was reduced to 2 mg, with a minimal loss of NF-E6. A second round of DNA affinity purification was performed on a different DNA column (see Materials and Methods). This step differs from the typical DNA affinity protocols where the extracts are fractionated again on the same column (Kadonaga and Tjian, 1986). The second DNA affinity column used a different matrix and a better-defined NF-E6 DNA-binding motif (Zafarana et al., 2000). In this way we sought to remove proteins from the extract that interacted with the previous matrix by electrostatic interaction, and most of the non sequence-specific DNA-binding proteins left after the first round of DNA-affinity purification. More than 90% of NF-E6 bound to the Thio CCAAT column (Fig. 4, FT). NF-E6 bound to the DNA column very strongly and could only be eluted with salt concentrations above 0.8 M. Unfortunately this resulted in tailing of the elution profile. Nonetheless, the majority of the activity eluted in six fractions (lanes 21 to 27, Fig 4). The total amount of protein in these fractions was about 3 micrograms. The fast mobility complex co-purifying with NF-E6 is effectively competed

when a normal amount of dIdC is added to the binding reaction (Fig. 3), suggesting that this shift originates from an a-specific DNA binding protein retained in the purified extract.

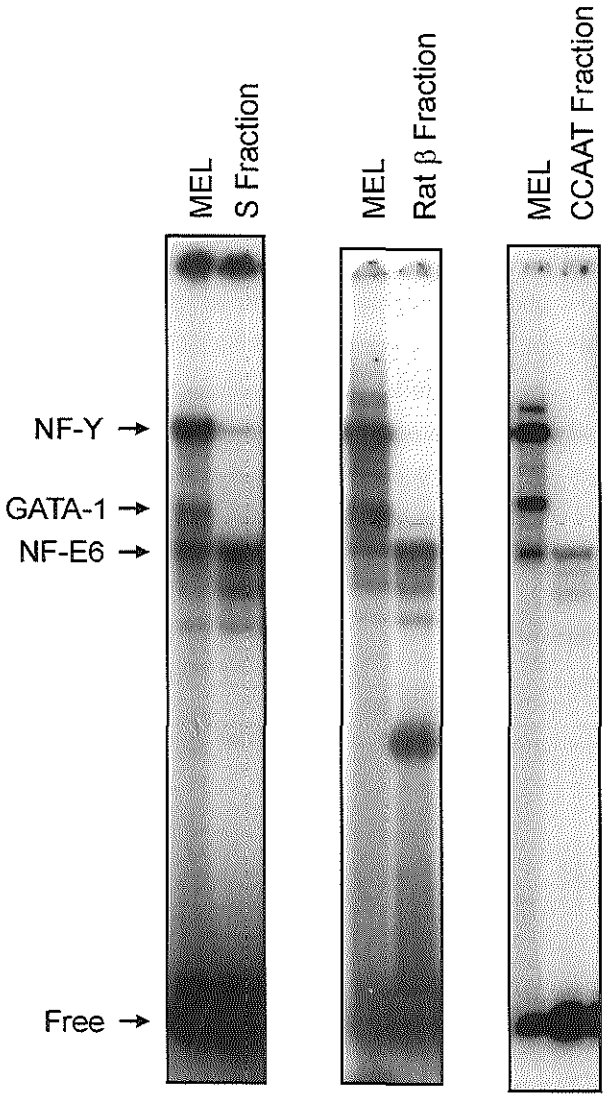

Fig. 3. Purification of NF-E6.
Bandshift assay with the β-globin CCAAT box probe of the purification of NF-E6 after chromatography on S-Sepharose (S Fraction); rat β CCAT DNA affinity column, and the thio CCAAT DNA affinity column. See text for further details.

PAGE purification and sequencing.

To determine the extent of the purification, aliquots of the purified fraction from the Thio CCAAT column were pooled and run on a 15% SDS-PAGE. Although it was found that the purified extract still contained more than one polypeptide, five well-resolved bands could be detected in the region of the gel that separated proteins between 18 and 40 kD. Interestingly, we observed that at the end of the chromatographic purification several distinct polypeptides could be visualised in the region containing NF-E6 (Fig. 2B). The elution profile of NF-E6 followed the appearance of these bands on the gel. This was determined by running one microliter aliquot from each individual fraction obtained from the Thio CCAAT column on a 15% SDS-PAGE (data not shown).

To recover the individual polypeptides, about 1 microgram of the purified extract pooled from fractions 21 to 26 was denatured and run on a preparative 15% SDS-PAGE. After superficially staining the gel with Coomassie, followed by destaining, the bands were visualised and excised from the gel (Fig 5). We have determined that the lower limit of detection is 50 nanograms of a 25-kD protein with this staining protocol (data not shown). By comparison we infer that the amount of protein contained in the excised bands varies between 50 and 150 nanograms. The identity of the excised proteins was determined by

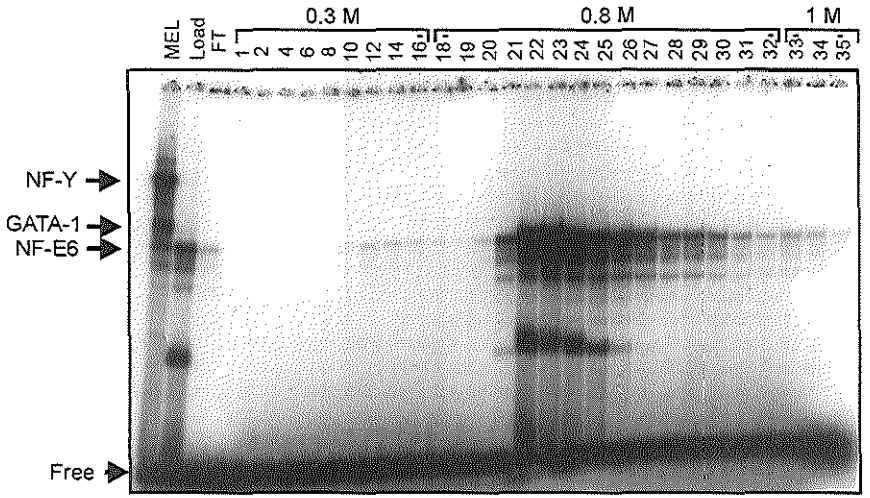

Fig. 4. Purification of NF-E6 on the thio CCAAT DNA affinity column.
One percent of each fraction eluted from the thio CCAAT DNA affinity column was tested for NF-E6 activity in a bandshift assay with the human β-globin CCAAT box probe. NF-E6 elutes from this column at 0.8 M KCl.

Edman degradation of peptides derived from partial *in situ* proteolysis of the bands with an endolysylpeptidase (Totty et al., 1992). Bands 1 and 2 were successfully sequenced. Band 1 was found to contain peptides derived from the ribosomal protein FT3/S3A (Kho and Zarbl, 1992), a known contaminant in DNA affinity purifications (Kadonaga and Tjian, 1986).

Proteolytic degradation of band 2 yielded three peptides that could be deconvoluted from Swissprot to match tsHMG ((Boissonneault and Lau, 1993). This protein has been described as a testis-specific HMG box protein. Interestingly, in one of the peptides sequenced from band 2, thirteen out of the fifteen amino acids are identical to the reported sequence of tsHMG. The sequences of the other two peptides are 100% identical to that found in ts-HMG. Therefore, the protein in band 2 may represent a novel HMG box protein, but it is more likely that the differences are due to ambiguities in peptide sequencing.

Unfortunately the identity of the proteins in bands 3, 4 and 5 could not be determined because the peptide mixture obtained from the partial proteolysis could not be separated by reversed-phase chromatography.

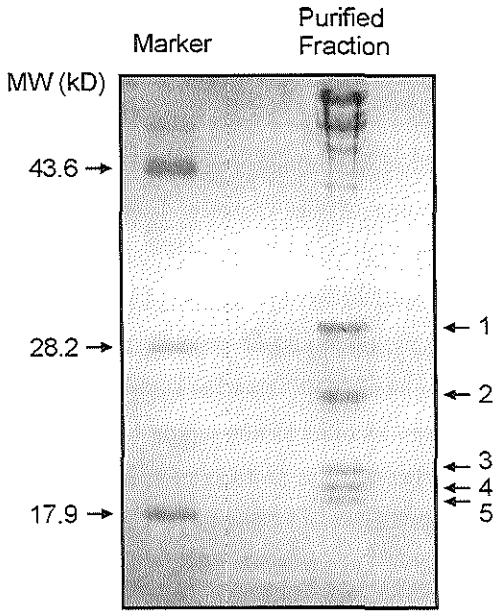

Fig. 5. Preparative SDS-PAGE gel of purified NF-E6.
The fractions eluting from the thio CCAAT column at 0.8 M KCl (Fig. 4) were pooled, precipitated with TCA and 1/3 was loaded on a 15% preparative SDS-PAGE gel. The gel was briefly stained with Coomassie Brilliant Blue G 250 and destained briefly. The indicated bands were excised and further for protein sequencing by Edman degradation.

Discussion

The majority of the known protein sequence has been obtained by the reverse genetic approach. Undoubtedly, this methodology provides a fast route for studying the function of a particular protein under investigation. Thanks to the development of molecular techniques, once the cDNA of a gene has been cloned, recombinant protein can be produced and used for biochemical analysis without the need for lengthy purification protocols. Homogeneous preparations of recombinant protein are also routinely used to produce specific antibodies that are used for further characterization such as tissue distribution and protein interactions. However, in some cases the sequence of the gene coding for a protein is not known and the protein is initially characterized by subcellular fractionation. In the case of NF-E6, the protein is characterized as an activity that binds to the CCAAT box of the β-globin gene.

In this study, we have used the binding specificity of NF-E6 as a coupled assay in an attempt to characterize the protein composition of this binding activity.

The first interesting result comes from the fractionation of crude nuclear erythroid extracts. The data suggests that the factors detected as distinct activities binding to the β–globin CCAAT box in bandshift assays co-exist in erythroid cells as a multi-protein complex. This complex contains, besides NF-E6, at least NF-Y and GATA-1. The functional relevance of this multi-protein complex is not yet known but it does suggest that the CCAAT boxes of the β-globin promoters are needed to provide a platform for bringing multiprotein complexes close to the TATA boxes. The subunit composition of this complex is likely to be of importance in the modulation of gene transcription. The CCAAT box is probably one of the most common elements in the promoter of RNA polII transcribed genes (Mantovani, 1998) and it is sufficient in minimal promoters, such as the mouse hsp68 promoter, to drive gene erythroid specific expression when linked to the β-globin LCR (Tewari et al., 1996).

The NF-E6 binding activity originates from polypeptides of molecular weight between 19 and 25 kD, but it is recovered from a native fraction that contains proteins of at least twice that size. Thus it appears that NF-E6 is composed of at least two polypeptides. The only protein band successfully sequenced from the area of the gel containing NF-E6 was ts-HMG (Boissonneault and Lau, 1993). However, it is unlikely that ts-HMG is a component of NF-E6. Attempts to supershift NF-E6 with antibodies against ts-HMG were unsuccessful and recombinant tsHMG did not bind to the β-globin CCAAT probe (data not shown). We can also conclude that the ribosomal protein identified in this study is a non-specific protein that has copurified with NF-E6. These types of proteins are known contaminants in affinity purifications of DNA binding factors and are usually removed by adding competitors like

poly-dIdC (Kadonaga and Tjian, 1986). However, due to binding kinetics created on a DNA affinity column, addition of dIdC also results in the loss of specific DNA-binding proteins on the column. Therefore, several passages through a DNA affinity columns may be preferable to the addition of large amounts of competitor. Non-specific interactions of proteins with the column matrix are also a common problem observed in many purifications. We reduced this background by using two different matrices to prepare the DNA affinity column. Overall the protocol devised for the purification of NF-E6 was very efficient and resulted in an almost homogeneous purified fraction containing only 5 protein bands below the 40 kD region. Unfortunately, the 3 bands in the 20 kD region, which most likely contain NF-E6, failed to yield any sequencing data. In Chapter 4, we describe that the NF-E6 activity contains the transcription factor C/EBPγ. Since C/EBP proteins function as homo- and heterodimers, it remains possible that NF-E6 is a heterodimer of C/EBPγ and another polypeptide (Lekstrom-Himes and Xanthopoulos, 1998). Therefore, the composition of the NF-E6 complex remains to be elucidated.

Material and Methods

Cell culture.

C88 Mouse Erythroleukaemia (MEL) cells were grown in DMEM supplemented with 10% (v/v) fetal calf serum and 2 mM glutamine. The cells were cultured in roller bottles at a density of approximately $5x10^6$ cells/ml.

Nuclear extracts.

Nuclear extracts were prepared from MEL cells and mouse tissues as described (Wall et al., 1988). The extracts were resuspended in at 10 mg/ml in TM buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM KCl, 20% Glycerol, 1 mM DTT).

Gel mobility shift assays and oligonucleotide probes

Gel mobility shift assays were performed as described by (Wall et al., 1988) using the human β-globin CCCAT box oligonucleotide as probe (deBoer et al., 1988). The following oligonucleotides were used in this study as competitors or to generate DNA affinity columns: GATA-1 5'-GGCGAAGTGATAAGGGGAGGCGCC-3' (Whyatt et al., 1993), rat β CCAAT box 5'-AACCCTCACATTGCCCAATCTGCTCACAC A-3' (Pavlovic et al., 1999) and Thio CCAAT 3x(5'-AGATTGGGCAATGT-3') (based on (Zafarana et al., 2000).

DNA affinity columns

Rat β CCAAT column.

A 2 ml OMNI column with a diameter of 0.5 cm was packed with 1 ml of Ultralink™ Immobilized Neutravidin Plus matrix (Pierce) and fitted to an Econo low pressure system (BioRad). The matrix was equilibrated with sterile 1xPBS, 5 mM EDTA. Following the washing step, an excess of biotinylated double stranded β rat oligonucleotide (22 units, 1mg of oligonucleotide biotinylated at the 3' end of the antisense strand) in 1ml of washing buffer was recirculated at 0.25 ml/min for 3 hours to allow binding to the matrix. A tracer, consisting of radioactively labeled biotinylated rat β CCAAT oligonucleotide (50 nmol) was included to monitor binding. Finally the column was washed extensively with sterile 1xPBS, 5 mM EDTA.

Thio CCAAT column.

The DNA affinity column was prepared by chemical coupling essentially as described (Bernatowicz and Matsueda, 1986; Kang et al., 1995), with a few modifications. The coupling procedure involved two steps. First, the amino groups on an AH-Sepharose matrix were bromoacetylated by reaction with the maleimido functional N-hydroxysuccimamide ester group of the heterobifunctional crosslinking reagent Bromoacetic acid-N-hydroxy-succinamide ester. Then the DNA was loaded onto the matrix by forming a stable thioether linkage between the bromoacetylated residue on the beads and the thiol group on the 3' end of the oligonucleotide. Briefly, 0.25 grams of 6-aminohexylsepharose (AH-Sepharose, Fluka) (10 μmol –NH2/ml of gel), were swollen in water and equilibrated in 0.1 M potassium phosphate buffer (pH 7.0). The coupling reagent (134 μmol in 0.5 ml DMF) was added dropwise to the ice cold matrix resuspended in 5 ml phosphate buffer. The coupling reaction was allowed to continue for 1 hour and then the derivatised matrix was washed with several changes of 0.1 M phosphate buffer to remove the unreacted reagent. The bromoacetyl sepharose matrix (0.5 ml) was then packed in a 2 ml OMNI column with a diameter of 0.5 cm, fitted to a Econo low pressure system (BioRad) and equilibrated with 0.1 M potassium phosphate buffer (pH 7.8). The column was loaded with 1 mg of double-stranded CCAAT oligonucleotide in 0.1 M potassium phosphate buffer (pH 7.8). The sense strand of the oligonucleotides was synthesized with a 3' thiophosphoryl group attached to a $(CH_2)_3$ spacer. A radioactively labeled oligonucleotide tracer was added to monitor the coupling efficiency. The oligonucleotide was recirculated through the column at 0.25 ml/min for 5 hours. At the end of the coupling reaction, the remaining unreacted bromoacetyl groups on the column were blocked by circulating through the column with a 100 molar excess of cold adenosine 5'-O-

(3-thiotriphosphate) in 0.1 M potassium phosphate buffer (pH 7.8). Finally, the column was washed with phosphate buffer to remove the salt and the unbound oligonucleotides.

Molecular weight determination.

To determine the denatured molecular weight of NF-E6, five hundred μg of crude MEL nuclear extract was denatured and separated on a 5-cm long preparative 12.5% Glycine/SDS polyacrylamide gel. The gel was then divided into ten separate 0.5-cm slices cut lengthwise and the proteins were eluted from the gel slices with 1 ml of 50 mM Tris HCl (pH 8.0), 0.1% SDS at 37^0C for 16 hours. Following precipitation with 10 volumes of acetone in the presence of 50 μg of BSA, the eluted proteins from each slice were resuspended in 100 μl of 8 M Guanidine HCl made up in buffer D (20mM HEPES, 20% Glycerol, 0.1M KCl, 0.2 mM DTT, 0.5 mM EDTA, 1 mM PMSF, pH 7.9). Finally the extracts were renatured by slow dialysis against buffer D. Three microliters from each fraction were assayed by EMSA with the human β CCAAT oligonucleotide probe.

The molecular weight of the native NF-E6 activity was established by gel filtration. One hundred μg of MEL crude nuclear extract was fractionated on a calibrated S200 column in a SMART system (Pharmacia). The column was calibrated with known molecular weight standards. Briefly, the extract (2 mg/ml) was applied on a 50 μl loading loop and several 50-μl fractions were eluted with buffer D containing 10% glycerol. Five microliters from each fraction were tested by EMSA with the human β CCAAT oligonucleotide probe.

Protein purification.

The purification protocol consisted of two different steps. In the first step the extracts were purified by ion exchange chromatography. The following step involved separation onto two distinct DNA affinity matrices.

The crude nuclear extracts (50 ml; 10 mg/ml) were loaded in 5 separate aliquots on a 10 ml S Sepharose column (Pharmacia). For each round of purification the bound fraction was eluted in TM buffer with a linear gradient of KCl ranging from 0.1 M to 1.5 M generated by means of an Econo low pressure system (BioRad). Fractions of 5 ml were collected. The fractions containing NF-E6 were identified by testing 5 μl aliquots in bandshifts with the β CCAAT probe. The active extracts were then pooled and dialyzed against TM buffer containing 0.3 M KCl. Finally the buffer was exchanged with buffer D and the volume of the extract reduced to 11.5 ml with a vacuum dialyser (Sartorius).

Before the first DNA affinity step, 500 µg of poly-dIdC in 0.5 ml of buffer D were added to the extracts. After incubation on ice for about 30 min, the extracts were cleared by centrifugation. The cleared extracts were then loaded onto the rat β CCAAT column (see above). The bound proteins were eluted with a linear salt gradient from 0.1 M to 1.5 M KCl in buffer D. Fractions of 500 µl were collected. Aliquots of 2 µl from each fraction were assayed by bandshift. The active fractions were pooled, dyalised against buffer D and the volume reduced to 1.5 ml by vacuum dialysis.

Finally, the extracts recovered from the rat β CCAAT column were loaded onto the thio CCAAT column (see above). After an initial wash with 0.3 M KCl, fractions were eluted by gravity flow with 0.8 M KCl and finally with 1.0 M KCl. The fractions (100 µl) were collected in siliconised Eppendorf tubes. The NF-E6 activity was assayed by bandshift with a 1 µl aliquot of each fraction.

Sequencing.

About 1 µg of purified extract was precipitated with TCA (10% v/v final concentration), the pellet was rinsed with acetone, air dried and solubilised in 50 µl of SDS-PAGE loading buffer. The proteins were denatured at 85^{0}C for 10 min and then run onto a 15 % protein gel. The proteins were visualized by staining the gel with 0.05 % Coomassie brilliant blue in 40% methanol / 1 % acetic acid for 10 min and then destaining it twice in 10% methanol / 5% acetic acid for 5 min each time. Five discrete bands were excised from the gel. The samples were prepared for sequencing as described (Totty et al., 1992) by *in situ* lysylendopeptidase partial digestion, followed by separation of the peptides by reversed-phase chromatography. Finally the resolved peptides were sequenced by N-terminal Edman degradation.

References

Antoniou, M. and Grosveld, F. (1990) beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev*, **4**, 1007-13.

Bernatowicz, M.S. and Matsueda, G.R. (1986) Preparation of peptide-protein immunogens using N-succinimidyl bromoacetate as a heterobifunctional crosslinking reagent. *Anal Biochem*, **155**, 95-102.

Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, **358**, 499-502.

Boissonneault, G. and Lau, Y.F. (1993) A testis-specific gene encoding a nuclear high-mobility-group box protein located in elongating spermatids. *Mol Cell Biol*, **13**, 4323-30.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *Embo J*, **7**, 4203-12.

Delvoye, N.L., Destroismaisons, N.M. and Wall, L.A. (1993) Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol Cell Biol*, **13**, 6969-83.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) Importance of globin gene order for correct developmental expression. *Genes Dev*, **5**, 1387-94.

Kadonaga, J.T. and Tjian, R. (1986) Affinity purification of sequence-specific DNA binding proteins. *Proc Natl Acad Sci U S A*, **83**, 5889-93.

Kang, S.H., Xu, X., Heidenreich, O., Gryaznov, S. and Nerenberg, M. (1995) A new affinity purification procedure for DNA-binding proteins using bromoacetyl agarose. *Nucleic Acids Res*, **23**, 2344-5.

Karlsson, S. and Nienhuis, A.W. (1985) Developmental regulation of human globin genes. *Annu Rev Biochem*, **54**, 1071-108.

Kho, C.J. and Zarbl, H. (1992) Fte-1, a v-fos transformation effector gene, encodes the mammalian homologue of a yeast gene involved in protein import into mitochondria. *Proc Natl Acad Sci U S A*, **89**, 2200-4.

Lekstrom-Himes, J. and Xanthopoulos, K.G. (1998) Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J Biol Chem*, **273**, 28545-8.

Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res*, **26**, 1135-43.

Pavlovic, S., Mitrovic, T., Nikcevic, G., Grujicic, N., Lazic, D., Glisin, V. and Popovic, Z. (1999) The rat beta (b miny)-globin promoter: nuclear protein factors and erythroid-specific induction of transcription. *Cell Mol Life Sci*, **56**, 871-81.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996) Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *Embo J*, **15**, 143-9.

Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988) The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. *Embo J*, **7**, 3099-107.

Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996) The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development*, **122**, 3991-9.

Thein, S.L. (1993) beta-Thalassaemia. *Baillieres Clin Haematol*, **6**, 151-75.

Totty, N.F., Waterfield, M.D. and Hsuan, J.J. (1992) Accelerated high-sensitivity microsequencing of proteins and peptides using a miniature reaction cartridge. *Protein Sci*, **1**, 1215-24.

Wall, L., deBoer, E. and Grosveld, F. (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev*, **2**, 1089-100.

Wall, L., Destroismaisons, N., Delvoye, N. and Guy, L.G. (1996) CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells. *J Biol Chem*, **271**, 16477-84.

Whyatt, D.J., deBoer, E. and Grosveld, F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *Embo J*, **12**, 4993-5005.

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition. *Genes Dev*, **10**, 2894-902.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, 377, 209-13.

Zafarana, G., Rottier, R., Grosveld, F. and Philipsen, S. (2000) Erythroid overexpression of C/EBPgamma in transgenic mice affects gamma- globin expression and fetal liver erythropoiesis. *Embo J*, **19**, 5856-63.

Chapter 4

Erythroid overexpression of C/EBPγ in transgenic mice affects γ-globin expression and fetal liver erythropoiesis

Gaetano Zafarana, Robbert Rottier, Frank Grosveld and Sjaak Philipsen

EMBO J. 19, 5856-5863 (2000)

Erythroid overexpression of C/EBPγ in transgenic mice affects γ-globin expression and fetal liver erythropoiesis

Abstract

The CCAAT boxes of the β-like globin genes interact with three proteins: NF-Y, GATA-1 and NFE-6. We demonstrate that NFE-6 contains C/EBPγ, and address its role in globin gene regulation by erythroid overexpression of C/EBPγ, and a dominant negative form C/EBPγΔB, in mice. Elevated levels of C/EBPγ, but not C/EBPγΔB, increase expression of the (fetal) γ-globin relative to the (adult) β-globin gene. Interestingly, fetal liver erythropoiesis is ablated when the C/EBPγ and C/EBPγΔB levels are further increased in homozygous transgenics. We suggest that targeted expression of dominant-negative leucine zipper proteins is a generally applicable approach to ablate specific tissues in mice.

Key words: C/EBPγ / dominant negative proteins / Locus Control Region / globin gene switching /erythropoiesis

Introduction

Many mRNA coding genes contain functionally important CCAAT boxes in their promoter (Mantovani, 1998). Our interest in this sequence motif stems from the study of the human β-globin gene cluster. CCAAT consensus sequences are present in the minimal promoters of the human β-like globin genes (deBoer *et al.*, 1988; Liberati *et al.*, 1998) between a CACC box and TATA box motif all of which are necessary for efficient transcription (Antoniou *et al.*, 1995; deBoer *et al.*, 1988). The relevance of the TATA box and the CACC box has been deduced from patients with mutations in these motifs (Thein, 1993). In addition, genetic ablation studies in mice have shown that EKLF, a factor binding to the CACC box, is absolutely required for β-globin expression (Wijgerde *et al.*, 1996).

The β-globin CCAAT box binds three major complexes in nuclear extracts from murine erythroleukemia (MEL) cells, a tissue-culture model for adult erythropoiesis. These complexes have been designated NF-Y (previously called CP1), GATA-1 and NF-E6 (Berry

et al., 1992). NF-E6 displays a similar mobility on the β-CCAAT box as a factor denoted DSFr (Delvoye *et al.*, 1993) that was shown through expression cloning and antibody studies to contain C/EBPγ (Wall *et al.*, 1996). C/EBPγ is a member of the CCAAT/enhancer binding protein (C/EBP) family of transcription factors, a class of bZIP proteins characterised by a basic domain followed by a leucine zipper. The leucine zipper serves as a dimerization motif that is necessary for DNA binding by the basic domain. Each member can form homodimers as well as heterodimers with C/EBP family members. In addition, they can form dimers with several other leucine zipper proteins such as jun and fos, raising the possibility that C/EBP proteins exert their biological function by switching partners during differentiation. The C/EBP family consists of six members designated C/EBPα to C/EBPζ. The genes encoding C/EBPα, β, γ, δ and ζ have been disrupted by homologous recombination thus elucidating some of the functions of their gene products. C/EBPα is necessary for the establishment and maintenance of energy homeostasis in neonates (Wang *et al.*, 1995). C/EBPζ is involved in the induction of apoptosis under conditions associated with impaired function of the endoplasmic reticulum (Zinszner *et al.*, 1998). C/EBPα and C/EBPδ have a synergistic role in terminal adipocyte differentiation (Tanaka *et al.*, 1997). C/EBPβ has a critical role in ovarian follicle development (Pall *et al.*, 1997; Sterneck *et al.*, 1997) and in the immune system (Screpanti *et al.*, 1995; Tanaka *et al.*, 1997). C/EBPγ *null* mice show a high mortality rate within 48 hours after birth. The analysis of lymphoid cells of these mice has demonstrated a critical role for C/EBPγ in the functional maturation of NK cells (Kaisho *et al.*, 1999).

C/EBP proteins are expressed in the fetal liver at the onset of definitive hematopoiesis. Therefore, these proteins are likely to have a role in the differentiation and lineage commitment of the developing hematopoietic system (Scott *et al.*, 1992; Zhang *et al.*, 1997., Nerlov *et al.*, 1998).

In this paper, we have explored the relationship between NFE-6 and C/EBPγ, and we have addressed the role of bZIP proteins in globin gene expression and erythropoiesis. C/EBPγ was first cloned from B cells as a factor that binds to the Ig heavy chain promoter and enhancer. Unlike other C/EBP proteins, C/EBPγ is ubiquitously expressed (Roman *et al.*, 1990). The cDNA predicts a small polypeptide of 16.4 kD lacking the consensus N-terminus activation domain experimentally defined in C/EBPα (Friedman *et al.*, 1989) and C/EBPβ. Thus, C/EBPγ might function as LIP, a truncated version of C/EBPβ that operates as a modulator of a series of activators (Cooper *et al.*, 1995; Descombes and Schibler, 1991).

The CCAAT boxes of the β-like globin genes all conform to the C/EBP consensus which is dyad symmetric, consisting of abutted half-sites bearing the CCAAT sequence (Osada *et al.*,

1996). A role for the CCAAT boxes in globin gene expression is best demonstrated by mutations in the duplicated CCAAT box of the Aγ-globin (Berry *et al.*, 1992; Ronchi *et al.*, 1996). Such mutations can result in the extemporary expression of fetal γ-globin in adult life, a condition known as Hereditary Persistence of Fetal Hemoglobin (HPFH) and hence hint that factors binding to the CCAAT boxes have a role in γ-globin silencing. To investigate the potential function of C/EBPγ in globin gene expression and erythroid differentiation, we have generated transgenic mice expressing C/EBPγ under the control of β-globin regulatory sequences. We also produced mice expressing a dominant-negative version of C/EBPγ that lacks the DNA binding domain but retains the leucine zipper (C/EBPγΔB). Such a protein interferes with DNA binding and activity of bZIP transcription factors (e.g. Moitra *et al.*, 1998). We found that moderate overexpression of C/EBPγ, but not C/EBPγΔB, results in increased γ-globin gene expression in transgenic mice carrying the human β-globin locus. Interestingly, higher overexpression of both proteins causes a block in definitive erythropoiesis leading to a severe anemia and fetal lethality.

Results

C/EBPγ is an integral part of NFE-6.

The CCAAT box of the human β-globin gene spans from position -62 to -87 of the start site (Fig. 1A). At least three activities bind these sequences in MEL cell extracts. Two of these have been identified as NF-Y and GATA 1; the third was called NFE-6 (Berry *et al.*, 1992). A complex migrating with similar mobility to NFE-6 was shown to contain C/EBPγ by another group (Wall *et al.*, 1996). To characterise NFE-6 further we analysed its contact points on the human β-globin CCAAT box by depurination analysis (Wall *et al.*, 1988). The contact points of NFE-6 (Fig. 1A) are in agreement with the consensus binding site of C/EBP proteins (Osada *et al.*, 1996). We therefore determined which members of this family are contained in NFE-6. For this purpose, we performed bandshifts on the β-globin CCAAT box with antibodies raised against C/EBP family members. NFE-6 is supershifted by antibodies against C/EBPγ (Fig. 1B, lane 4) and not by any of the other sera used. In addition, the reduced intensity of the slowest migrating complex in lane 3 correlates with the C/EBPβ supershift. The NF-Y antibodies also remove this shift (Fig. 1B, lane 10) and we conclude that this activity is composed of both NF-Y and C/EBPβ, in agreement with Wall *et al.* (1996).

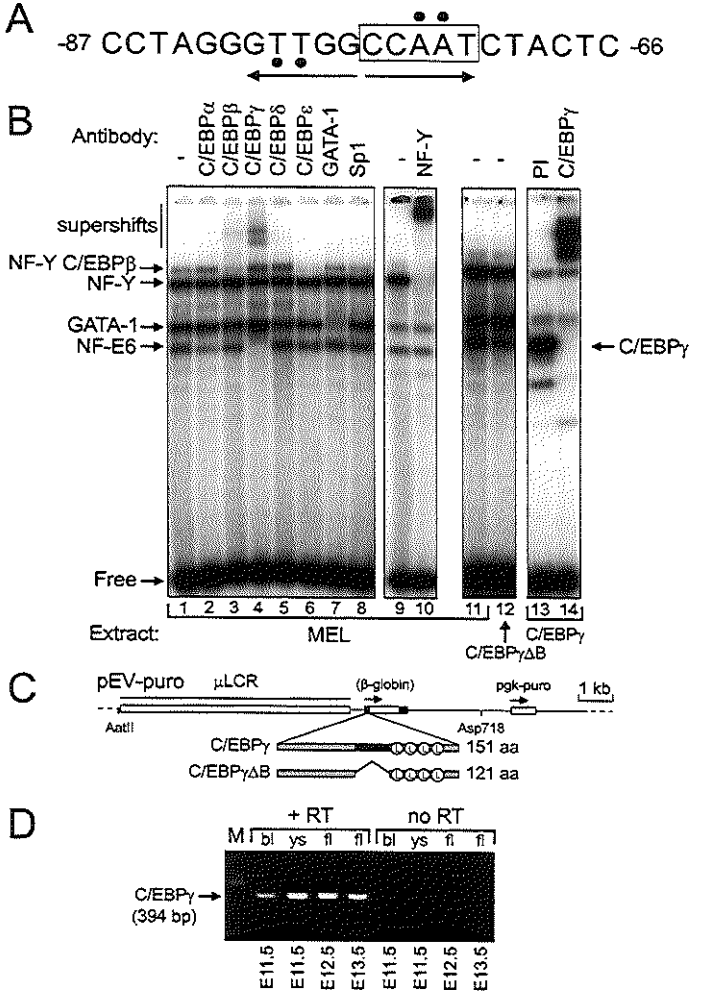

Fig. 1. Proteins interacting with the CCAAT box of the human β-globin gene. A) Sequence of the CCAAT box with the contact sites of NFE-6 indicated by black circles. The imperfect dyad symmetric C/EBP binding site is shown by arrows. B) Bandshift analysis of proteins binding to the β-globin CCAAT box. An oligonucleotide probe covering the CCAAT box was incubated with 4 μg MEL cell nuclear extract plus the indicated antibodies; PI = pre-immune serum. Lanes 12-14 contain extracts from MEL cells overexpressing the indicated EBP proteins. C) Constructs used to express C/EBPγ and C/EBPγΔB. Restriction sites used to isolate the microinjection fragments are indicated. D) RT-PCR analysis of C/EBPγ expression in embryonic and fetal hematopoietic tissues. Two μg of RNA from the indicated tissues (bl = blood; ys = yolk sac, fl = fetal liver) was used for RT-PCR (30 cycles). No RT: Reverse transcriptase was omitted to exclude amplification of contaminating DNA.

To further confirm that NFE-6 contains C/EBPγ, we cloned the cDNA by RT-PCR and expressed two versions of C/EBPγ under the control of the β-globin promoter and locus control region (LCR) in MEL cells (Needham *et al.*, 1992). The first construct, pEV-C/EBPγ, contains

the 0.54 kb open reading frame of C/EBPγ. The second construct, pEV-C/EBPγΔB, is a dominant-negative version of the former, in which we deleted the DNA sequences coding for amino acids Met58 to Glu89. This removes the basic DNA binding domain but leaves the leucine zipper dimerization domain intact (Fig. 1C). We then analysed nuclear extracts of transfected populations in bandshifts with the human β-globin CCAAT box oligonucleotide. This shows that C/EBPγΔB causes a reduction in the intensity of the NF-E6 band (Fig. 1B, lanes 11, 12). Furthermore, C/EBPγ protein displays the same mobility as NFE-6 (Fig 1B, lanes 13, 14). These data are in agreement with the notion that C/EBPγ is an integral part of NFE-6. Although it appears that NFE-6 is a homodimer of C/EBPγ, our results do not formally exclude that NFE-6 activity is a heterodimer of C/EBPγ and another polypeptide.

Overexpression of C/EBPγ in transgenic mice.

It has been suggested that C/EBPγ could be involved in the regulation of the β-globin gene in MEL cells (Wall *et al.*, 1996). However, we did not observe a change in the expression of the mouse adult globin genes upon MEL cell differentiation when C/EBPγ or C/EBPγΔB were overexpressed (data not shown). The outcome of this experiment suggests either that C/EBPγ may not be directly involved in the transcription of the globin genes or that the role of this factor is important at another stage during the differentiation of erythroid cells than that represented by MEL cells. C/EBPγ is ubiquitously expressed in adult tissues and in most tissues of fetuses at 14 days of gestation (E14) (Cooper *et al.*, 1995). Because we are interested in the function of C/EBPγ in erythropoiesis, we extended these studies to the erythroid organs of E11.5 to E13.5 mouse fetuses through an RT-PCR analysis of C/EBPγ expression. C/EBPγ mRNA is detected in E11.5 yolk sack and E12.5/E13.5 fetal liver (Fig. 1D) demonstrating that C/EBPγ is already expressed before the onset of adult erythropoiesis. Thus, C/EBPγ may have a role in erythroid cells of embryonic and fetal origin, and in globin gene switching. To investigate this, we generated transgenic mice with the pEV-C/EBPγ and pEV-C/EBPγΔB constructs (Fig. 1C). These expression vectors are active in embryonic, fetal and adult erythropoiesis. Several founders were generated and bred to obtain germline transmission (Table 1). Transgenic F1 pups reached adulthood with no obvious phenotype. Collectively, these mice will be referred to as EBP transgenics in this paper.

Expression of C/EBPγ, but not C/EBPγΔB, results in upregulation of γ-globin expression.

The EBP transgenics were bred with mice harbouring the 70 kb human β-globin locus (line 72, Strouboulis *et al.*, 1992), in order to analyse the effects of transgene-derived C/EBPγ and

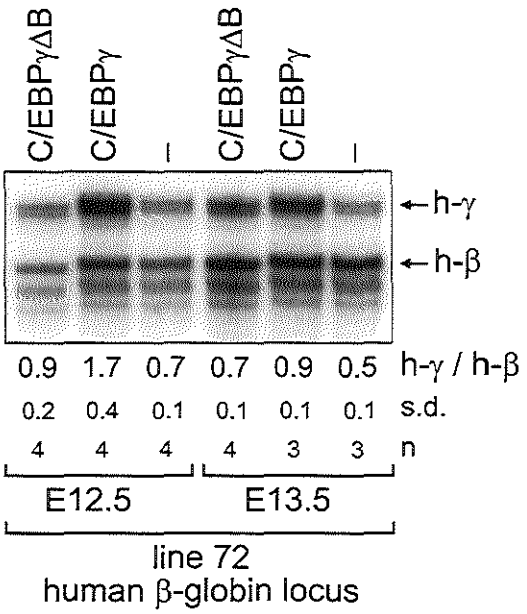

Fig. 2. Overexpression of C/EBPγ affects globin gene switching.
Mice harbouring the 70 kb human β-globin locus (line 72) were crossed with the lines C/EBPγΔB line 1 and C/EBPγ line 21, heterozygous for the EBP transgene. RNA was isolated from E12.5 and E13.5 livers and 1 μg was used for quantitative S1 nuclease analysis of human globin gene expression (Wijgerde *et al.*, 1996). Arrows indicate human γ-globin (h-γ) and β-globin (h-β) signals. The average of the h-γ/h-β signals with standard deviations (s.d.) and the number of fetuses analyzed (n) are shown below the lanes.

C/EBPγΔB on the regulation of the human β-like globin genes. This is particularly useful to study the switch from fetal (γ-) to adult (β-) globin expression which is absent in the endogenous mouse β-globin locus.

Fetal liver globin mRNA levels of compound transgenic fetuses dissected at E12.5 and E13.5 were assayed by S1 nuclease protection (Fig. 2). For each transgene combination, several fetuses from independent litters were analyzed. The RNA samples were also assayed for expression of the murine βH1 and εy genes to assess the contribution of primitive erythrocytes derived from the circulation; this was found to be minimal and hence the γ-globin mRNA measured is derived from fetal liver cells (data not shown).

Interestingly, the γ-globin gene is upregulated in the presence of the pEV-C/EBPγ transgene, but not in the presence of the pEV-C/EBPγΔB transgene (Fig. 2). This is particularly evident in the E12.5 samples when we observe 2-3-fold higher γ-to-β ratio in the presence of the pEV-C/EBPγ transgene (Fig. 2). There was no significant difference in the upregulation of γ-globin

Founder	Copy number	Remarks	γ-globin expression
PEV-C/EBPγ			
7	(10)	No transmission	n.a.
8 *	6	Homozygous alive	γ up in het. and hom.
13	4	Transmission	n.d.
14	(42)	No transmission	n.a.
19	26	Homozygous lethal	γ up in het.
21 *	20	Homozygous lethal	γ up in het.
22	(22)	No transmission	n.a.
28	12	Homozygous alive	n.d.
pEV-C/EBPγΔB			
1 *	22	Homozygous lethal	γ normal in het.
2	(16)	No transmission	n.a.
10	(20)	No transmission	n.a.
12	6	Homozygous alive	γ normal in het. and hom.
19	5	Homozygous alive	γ normal in het. and hom.
21	(20)	No transmission	n.a.
22 *	40	Heterozygous lethal	γ normal in het.
14	(< 1)	No transmission	n.a.

Table 1. EBP transgenic mice generated for this study.
n.a. = not applicable; n.d. = not determined; het. = heterozygous; hom. = homozygous for the EBP transgene. Expression of γ globin refers to γ mRNA levels in E12.5 fetal liver in crosses with mice carrying the human β-globin locus (line 72, Strouboulis *et al.*, 1992). There was no significant difference in the upregulation of γ-globin between the different C/EBPγ transgenics.
* Data obtained with these lines are shown in the Figures.

expression between the different C/EBPγ transgenic lines (Table 1). In adult animals, the EBP transgenes do not activate γ-globin expression, and these animals have normal numbers of erythrocytes, lymphocytes, granulocytes and platelets in their circulation (data not shown).

We conclude that overexpression of C/EBPγ in the fetal liver results in increased γ-globin expression. This prolongation is dependent on the presence of the basic DNA-binding domain in the C/EBPγ protein, suggesting that C/EBPγ acts preferentially at the γ-globin CCAAT boxes. These results are in contrast with data obtained from transfection experiments that suggested a function for C/EBPγ in transcriptional enhancement through the β-globin CCAAT box (Wall *et al.*, 1996).

High-level expression of the EBP transgenes blocks fetal liver erythropoiesis.

Animals heterozygous for the EBP transgenes are viable and show no morphological abnormalities in blood and fetal liver (see Fig. 3). However, one of the pEV-C/EBPγΔB founder males gave rise to small litters and the pups analysed were non-transgenic. We extended the analysis of this line to E11.5-E15.5 fetuses and we found normal-sized litters up to E13.5 and re-absorbed fetuses from E14.5 onwards. At E13.5, 30% of the fetuses were severely anemic.

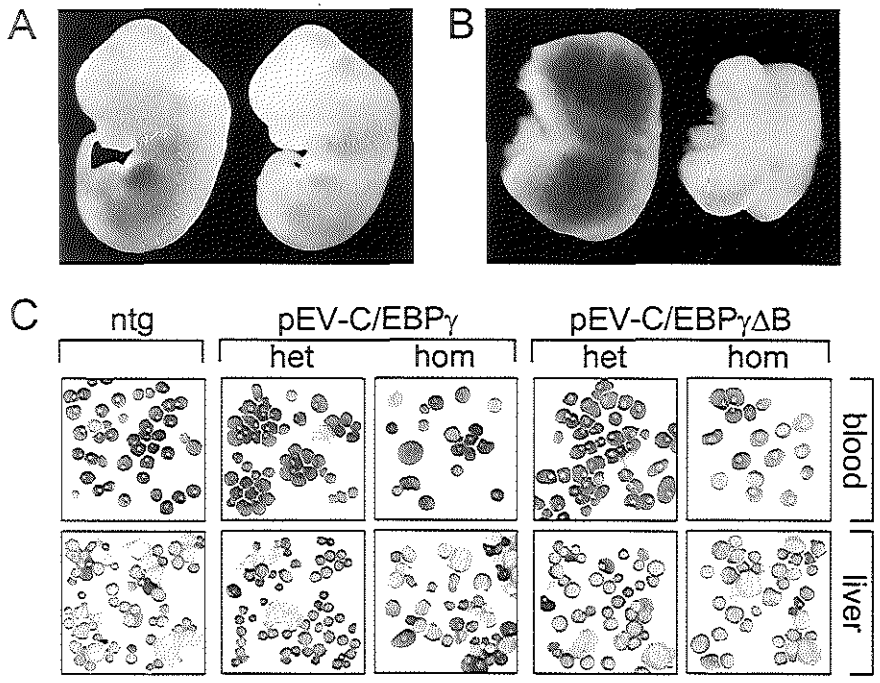

Fig. 3. Fatal anemia caused by high level expression of the EBP transgenes.
A) Wildtype (left) and C/EBPγΔB-overexpressing line 22 (right) E12.5 littermates. Note the pallor of the transgenic fetus that has an otherwise normal gross morphological appearance. B) Fetal livers dissected from the fetuses shown in A. The transgenic liver (right) is small and pale relative to the control liver (left), indicative of hampered erythropoiesis. C) Cytospins of peripheral blood and fetal liver single cell suspensions prepared from fetuses with the indicated genotype. Ntg = non-transgenic; het = heterozygous; hom = homozygous. Data shown are obtained with C/EBPγΔB line 1 and C/EBPγ line 21. The slides were stained with a combined histological and neutral benzidine stain.

The affected fetuses all carried the transgene and had otherwise developed normally (Fig. 3A). The fetal livers of these animals were much smaller and paler than those of non-transgenic littermates (Fig. 3B). Since this phenotype could be due to high levels of transgene-derived protein, it might be reproduced in other EBP transgenic lines if the transgene would be expressed at sufficiently high levels. We therefore raised the expression levels by breeding the EBP lines to homozygosity. Western blot analysis of fetal liver nuclear extracts shows that the levels of the transgenic EBP proteins are raised in homozygous transgenics, that the expression levels correlate with the transgene copy number, and that the C/EBPγ wildtype and ΔB proteins are expressed at comparable levels (Fig. 4A). We found that homozygous fetuses of the high copy number lines were anemic and died around E13-E14, confirming that the phenotype observed is due to the overexpression of the EBP proteins, and not a secondary effect of transgene integration. Furthermore, there is a good correlation between transgene copy number and fetal

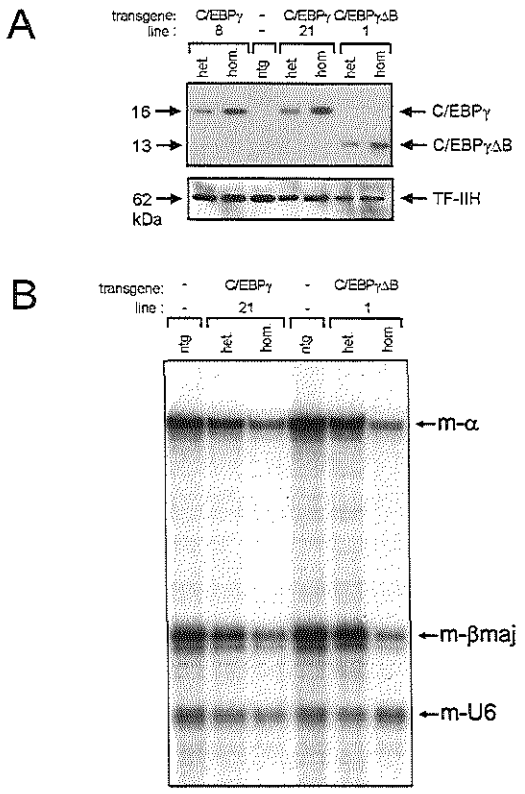

Fig. 4. High levels of the EBP proteins result in reduced expression of the α- and β-globin genes.
A) Ten μg E12.5 fetal liver nuclear extract was used for Western blot analysis of EBP protein expression. Results are shown for non-transgenic (ntg), heterozygous (het.) and homozygous (hom.) EBP transgenics of the lines indicated. The polyclonal rabbit antiserum used was raised against the N-terminal part of C/EBPγ that is shared by the C/EBPγ and C/EBPγΔβ proteins. Note that the levels of endogenous C/EBPγ, although present in the RT-PCR analysis (Fig.1D), are too low for detection with this antiserum. A mouse monoclonal antibody recognizing the p62 subunit of TF-IIH was used as a loading control.
B) E12.5 livers were dissected from non-transgenic (ntg), heterozygous (het.) and homozygous (hom.) EBP transgenics of the lines indicated and used for RNA isolation. Approximately 1 μg of total RNA was used for quantitative S1 nuclease analysis of mouse α- (m-α) and β_{major}-globin gene expression (m-βmaj); a probe detecting mouse U6 RNA (m-U6) was used as an internal loading control (Wijgerde *et al.*, 1996).

lethality (Table 1). Surprisingly, high levels of both C/EBPγ and C/EBPγΔB proteins cause anemia and fetal lethality. The paleness of the fetal livers prompted us to determine by S1 nuclease analysis if overexpression of the EBP proteins interfered with globin gene expression. We observe a general reduction in the levels of the mouse (Fig. 4B) and human (data not shown) globin genes if the fetuses are homozygous for the high copy number EBP transgenes. The ratio between the α- and β-like genes is comparable to the controls, demonstrating that the block in

erythroid differentiation is accompanied by reduced expression of the globin genes from both loci (Fig.4B). Thus, we conclude that the anemia is not due to chain imbalance caused by a defect in either α- or β-globin expression, but more likely by an early block in definitive erythropoiesis resulting in fewer cells transcribing the globin genes.

Cytological analysis of fetal livers and blood.

Fetal livers of mice expressing high levels of EBP proteins are pale and small (Fig. 3B). We therefore analysed the morphology and distribution of the cells in the fetal livers and peripheral blood of wildtype- and EBP transgenic fetuses. Cytospins of blood and E12.5 livers were prepared and stained with histological dyes and neutral benzidine (Beug *et al.*, 1982). The pattern found in E12.5 control livers is shown in Fig. 3C (ntg). A predominance of small, benzidine positive, polychromatic and orthochromatic erythroblasts is observed, followed by fewer basophilic erythroblasts and proerythroblasts. A similar pattern is observed in the livers of heterozygous fetuses (Fig. 3C, het). In contrast, liver preparations from homozygous fetuses are composed mainly of benzidine-negative blast-like cells. The absence of small, benzidine positive, cells is immediately apparent (Fig 3C, hom). Thus, it appears that the transgenic fetal livers are hypocellular and pale because erythropoiesis is impeded at a stage preceding haemoglobinized polychromatic erythroblasts. The failure of liver erythropoiesis was further confirmed by *in vitro* progenitor assays with pEV-C/EBPγ line21 and pEV-C/EBPγΔB line1 fetal livers. E12.5 livers of homozygous fetuses formed virtually no CFU-Es (>15-fold reduction) and a drastically reduced number of BFU-Es (6-7-fold reduction). These results indicate that the block in erythropoiesis occurs at the level of relatively early precursor represented by the BFU-E colonies. Since the colony numbers are not significantly different between wild type and heterozygous fetal livers, we conclude that there is a critical threshold level for the dominant negative effect of the EBP proteins. Finally, the blood shows a large number of blue nucleated cells distinct from the primitive erythrocytes (compare Fig. 3C, ntg and het, with Fig. 3C, hom). These abnormal cells are presumably of fetal liver origin, and are probably prematurely released in circulation in response to the severe anemia.

We conclude that fetal liver erythropoiesis is severely impaired in homozygous EBP transgenics. This defect in erythropoiesis is independent of the presence of the DNA-binding domain of C/EBPγ, suggesting that C/EBPγ might indeed act as a modulator of the activity of other bZIP proteins that are essential for the proper execution of the erythropoietic program.

Discussion

C/EBPγ and the developmental regulation of the β-globin gene cluster.

The work described in this paper was initiated to study the role of the previously described CCAAT box binding protein NFE-6 in the regulation of the human β-globin gene cluster. Several naturally occurring mutations in the promoters of the globin genes denote the involvement of factors binding to these sequences for the proper temporal and topological expression of these genes. For example, mutations in the CACC box of the human β-globin gene or ablation of EKLF, an erythroid restricted factor that binds to this motif, result in impaired expression of the β-globin gene (Thein, 1993; Wijgerde *et al.*, 1996). Intriguingly, no patients with mutations in the β-CCAAT box are known, suggesting that the β-CCAAT box is perhaps not important for β gene expression. The relevance of the globin CCAAT boxes *in vivo* is best demonstrated by naturally occurring mutations clustered around the distal CCAAT box of the fetal γ-globin gene that result in an HPFH phenotype (Poncz *et al.*, 1988). Consequently it would appear that factors binding to these sequences are causative in the proper regulation of the fetal genes and have a role in the switch from γ- to β-globin expression. Reactivation of γ-globin expression in adults is clinically important because this would be beneficial to large numbers of β-thalassemia and sickle cell anemia patients.

We show in this paper that NFE-6 contains C/EBPγ, and that it is most likely the same factor as the DSFr activity described by (Delvoye *et al.*, 1993). It has been postulated that C/EBPγ could play a role in the developmental upregulation of the adult β globin gene (Wall *et al.*, 1996). However, our data in mice overexpressing C/EBPγ do not support this model, because we observe an increase, rather than a decrease, in γ-globin expression during the switching period. If C/EBPγ would have been a β-promoter specific activator, γ-globin transcription should have decreased. For instance, a robust positive correlation between the levels of the β-promoter transcription factor EKLF and β gene expression in the fetal liver has been reported, and higher β levels are accompanied by lower γ levels (Wijgerde *et al.*, 1996). Thus, our data strongly suggest that C/EBPγ is not an activator of the β-globin promoter. We can only speculate about the mechanism by which increased C/EBPγ levels favour γ-globin expression. One obvious possibility is that it exerts this effect directly through the γ-globin CCAAT boxes. Alternatively, it could affect the developmental progression of fetal liver erythropoiesis resulting in prolonged maintenance of a transcription factor environment favouring γ-globin expression. In this respect, it is interesting to note that C/EBP proteins are thought to be part of a developmental transcription factor cascade in hematopoiesis (Sieweke and Graf, 1998).

After the fetal liver stage, the γ-globin genes are silenced normally and we do not observe an HPFH phenotype in our EBP transgenic lines. Furthermore, the γ-globin expression levels in adult transgenic mice with the –117 Greek HPFH mutation in the distal γ-globin CCAAT box (Berry *et al.*, 1992) remain unchanged in the presence of the EBP transgenes (data not shown). This is consistent with the notion that C/EBPγ is not an activator of the β-globin promoter, because in that case a shift in favour of β expression should occur in the presence of elevated C/EBPγ levels. However, the molecular mechanism underlying Greek HPFH remains elusive since neither point mutagenesis of the distal CCAAT box (Ronchi *et al.*, 1996) nor manipulation of one of the factors binding to it (this study) has provided clear-cut insight into the mode of action of the –117 mutation. It appears that either more than one factor is involved, or that the –117 mutation precipitates a secondary effect like localized chromatin remodelling.

High-level expression of the EBP transgenes blocks fetal erythropoiesis.

We have found that overexpression of the EBP transgenes in MEL cells does not interfere with DMSO-induced MEL cell differentiation. Since MEL cells represent the proerythroblast stage of erythroid development, this suggests that the differentiation program beyond this stage can be efficiently executed in the presence of high levels of EBP proteins. In sharp contrast, the effects of such high levels on erythropoiesis in mice are dramatic. It results in a developmental block of early erythroid progenitors in the fetal liver. C/EBP family members have distinct pivotal roles in lineage commitment in the hematopoietic system (Cooper *et al.*, 1994; Kaisho *et al.*, 1999; Nerlov *et al.*, 1998; Screpanti *et al.*, 1995; Zhang *et al.*, 1997). These proteins are known to interact with several other factors in large multiprotein complexes that regulate the tissue-specific expression of several hematopoietic-specific genes (reviewed in Sieweke and Graf, 1998). It has been proposed that C/EBPγ is a modulator of the activity of other bZIP proteins (Cooper *et al.*, 1995). Our data are consistent with this proposal, since we observe defective erythropoiesis in mice with high levels of C/EBPγ and the dominant-negative version C/EBPγΔB. Presumably, provided they are expressed at sufficiently high levels, both EBP proteins interfere with activator functions of bZIP proteins required for fetal liver erythropoiesis.

Tissue-specific expression of dominant-negative leucine zipper proteins.

Inactivation of genes by knockout strategies is a frequently chosen approach to address the function of gene products in whole organisms. In the case of C/EBP family members, much

has been learnt about the specific roles of individual family members in the development and physiology of mice (LekstromHimes and Xanthopoulos, 1998). Functional redundancy and early lethality are potential problems of gene knockouts. To some extent, these issues can be tackled by intercrossing knockout lines and by using conditional knockout alleles. However, the number of combinations that can be made is limited for practical reasons. The dominant-negative transgenic approach does not suffer from these limitations. Expression of the transgene can be directed to specific tissues and unlike toxin genes, leaky expression in other tissues should not adversely affect these tissues because low level expression is unlikely to interfere with the functioning of the cells. Furthermore, dimerization specificity of the leucine zipper can be modulated by in vitro mutagenesis (Vinson *et al.*, 1993), and fusion with steroid receptor ligand binding domains would give an additional level of control over the spatio-temporal activity of the dominant-negative protein (Littlewood *et al.*, 1995).

Dominant-negative leucine zippers have been applied successfully in gene expression and differentiation studies of tissue-culture cells (i.e. Nerlov *et al.*, 1998; Olive *et al.*, 1997), and this approach was used to study the function of adipose tissues in mice (Moitra *et al.*, 1998). Expression of a dominant-negative leucine zipper-containing protein under the control of the adipose-specific aP2 enhancer/promoter resulted in the complete ablation of white adipose tissue with dramatic consequences for the physiology and behaviour of the transgenic mice. In our EBP transgenics we have inactivated the erythroid system, but development of the transgenic fetuses is otherwise normal. Thus, we suggest that targeted expression of dominant negative leucine zipper proteins might be a generally applicable approach to ablate specific cell types and tissues in order to study their role in the development, physiology and behaviour of mice.

Materials and Methods

Cloning of C/EBPγ and expression constructs.

Total RNA from MEL cells was isolated as described (Antoniou, 1991) and 5 μg reverse transcribed with the γ-r primer 5'-GGCCTGGAAGGAGATCTACT-3'. One-fifth of the first strand product was subjected to 30 cycles of PCR with the γ-f primer 5'-ACGTGCCCAAATGAGCAAGC-3'. The PCR product was cloned and sequenced. To delete the DNA binding domain, the 3' end of C/EBPγ was amplified with the C/EBPγ reverse primer and the gamma-1 primer 5'-ATGCCATGGATACACTGCAA-3'. The product was digested with NcoI and BglII, ligated to the C/EBPγ plasmid digested with NcoI and BglII.

EcoRI/BglII inserts of C/EBPγ- and C/EBPγΔB-plasmids were ligated to pEV3-puromycin^r (Needham *et al.*, 1992) to generate pEV-C/EBPγ and pEV-C/EBPγΔB.

Cell culture and protein preparation.

PEV-C/EBPγ and pEV-C/EBPγΔB were linearized with ScaI and transfected into MEL cells as described (Antoniou, 1991) with the exception that puromycyin (1 μg/ml) was used for the selection. RNA and nuclear extracts were prepared as described (Antoniou, 1991; Wall *et al.*, 1988).

Production of antibodies and gel mobility shift assays.

An NcoI site was introduced at the first ATG of C/EBPγ by PCR. The following primers were used: 5'-CCATGGGCAAGCTGTCGCAGCCAG-3' and the γ-r primer. The PCR product was cloned and sequenced. The plasmid was digested with NcoI and the insert coding for the unique N terminus of C/EBPγ was cloned into the NcoI site of pET15 and used for the production of recombinant protein in E. coli BL2 (DE3). Gel purified recombinant protein was used to immunise rabbits.

Gel mobility shift assays were performed as described by (Wall *et al.*, 1988) using the human β-globin CCAAT box oligonucleotide as probe (deBoer *et al.*, 1988). For supershifts, appropriately diluted antibodies were added to the reactions followed by a further 20 minutes incubation at room temperature prior to loading of the gel.

Generation of transgenic mice and DNA analysis.

PEV-C/EBPγ and pEV-C/EBPγΔB were digested with AatII and Asp718 .The fragments were isolated and used for microinjection as described (Kollias et al. 1986). Transgenics were screened by Southern blotting of EcoRI-BamHI digested DNA using the 0.54 kb EcoRI C/EBPγ fragment as probe.

RNA analysis.

RNA was prepared from yolk sacs, fetal livers and adult blood as described (Strouboulis *et al.*, 1992). RT-PCR was performed with the γ-f and the γ-r2 primer 5'-GGCTGTGCGCATGCTCAAGAAAC-3' (394 bp product) to detect C/EBPγ expression. Globin and U6 RNA levels were analysed by S1 nuclease protection (Wijgerde *et al.*, 1996). The S1 signals were quantitated by Phosphorimage analysis.

Protein analysis.

Nuclear protein was isolated from E12.5 liver (Andrews and Faller, 1991) and 10 μg was used for Western blot analysis with the rabbit anti-C/EBPγ polyclonal antibodies. A mouse monoclonal antibody recognizing the 62kD subunit of THII-H was used as a loading control; this antibody was generously provided by Dr. E. Citterio (Rotterdam).

Colony assays.

Colony assays were performed essentially as previously described (Wong *et al.*, 1986). Fetal livers were disaggregated into single cells by passage through a 100μm mesh and plated at a density of $3x10^5$ cells per ml in methyl cellulose containing 1U/ml Epo. The appearance of CFU-E and BFU-E colonies was scored after 2 and 14 days respectively.

Cytospin preparations and staining.

E12.5 fetuses were dissected from the uterus and bled on a dish to collect fetal blood. Then, the fetal livers were dissected and rinsed several times in cold PBS. A single cell suspension of the fetal livers was prepared by gentle pipetting with a yellow tip. Aliquots of blood and fetal liver cell suspensions were loaded on a cytofunnel and spun at 400 rpm for 5 min on microscope slides. The preparations were left to air dry and stained with a combined histological and neutral benzidine stain (Beug *et al.*, 1982).

References

Andrews, N.C. and Faller, D.V. (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.*, **19**, 2499.

Antoniou, M. (1991) Induction of erythroid-specific expression in murine erythroleukemia (MEL) cell lines. *Methods Mol Biol.* **7**, 421-434.

Antoniou, M., de Boer, E., Spanopoulou, E., Imam, A. and Grosveld, F. (1995) TBP binding and the rate of transcription initiation from the human beta-globin gene. *Nucleic Acids Res.* **23**, 3473-3480.

Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, **358**, 499-502.

Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H. and Graf, T. (1982) Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. *Cell*, **28**, 907-919.

Cooper, C.L., Berrier, A.L., Roman, C. and Calame, K.L. (1994) Limited expression of C/EBP family proteins during B lymphocyte development. Negative regulator Ig/EBP predominates early and activator NF-IL-6 is induced later. *J Immunol*, **153**, 5049-5058.

Cooper, C., Henderson, A., Artandi, S., Avitahl, N. and Calame, K. (1995) Ig/EBP (C/EBP gamma) is a transdominant negative inhibitor of C/EBP family transcriptional activators. *Nucleic Acids Res.* **23**, 4371-4377.

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

Delvoye, N.L., Destroismaisons, N.M. and Wall, L.A. (1993) Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol Cell Biol.* **13**, 6969-6983.

Descombes, P. and Schibler, U. (1991) A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell*, **67**, 569-579.

Friedman, A.D., Landschulz, W.H. and McKnight, S.L. (1989) CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells. *Genes Dev.* **3**, 1314-1322.

Kaisho, T., Tsutsui, P., Tanaka, T., Tsujimura, T., Takeda, K., Kawai, T., Yoshida, N., Nakanishi, K. and Akira, S. (1999) Impairment of natural killer cytotoxic activity and interferon γ production in CCAAT/enhancer binding protein γ-deficient mice. *J. Exp. Med*, **190**, 1573-1581.

LekstromHimes, J. and Xanthopoulos, K.G. (1998) Biological role of the CCAAT/ enhancer-binding protein family of transcription factors. *J Biol Chem*, **273**, 28545-28548.

Liberati, C., Ronchi, A., Lievens, P., Ottolenghi, S. and Mantovani, R. (1998) NF-Y organizes the gamma-globin CCAAT boxes region. *J Biol Chem*, **273**, 16880-16889.

Littlewood, T.D., Hancock, D.C., Danielian, P.S., Parker, M.G. and Evan, G.I. (1995) A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res*, **23**, 1686-1690.

Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res.* **26**, 1135-1143.

Moitra, J., Mason, M.M., Olive, M., Krylov, D., Gavrilova, O., MarcusSamuels, B., Feigenbaum, L., Lee, E., Aoyama, T., Eckhaus, M., Reitman, M.L. and Vinson, C. (1998) Life without white fat: a transgenic mouse. *Genes Dev.* **12**, 3168-3181.

Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F. and Hollis, M. (1992) LCR/MEL: a versatile system for high-level expression of heterologous proteins in erythroid cells. *Nucleic Acids Res.* **20**, 997-1003.

Nerlov, C., McNagny, K.M., Doderlein, G., Kowenzleutz, E. and Graf, T. (1998) Distinct C/EBP functions are required for eosinophil lineage commitment and maturation. *Genes Dev.* **12**, 2413-2423.

Olive, M., Krylov, D., Echlin, D.R., Gardner, K., Taparowsky, E. and Vinson, C. (1997) A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J Biol Chem.* **272**, 18586-18594.

Osada, S., Yamamoto, H., Nishihara, T. and Imagawa, M. (1996) DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. *J Biol Chem.* **271**, 3891-3896.

Pall, M., Hellberg, P., Brannstrom, M., Mikuni, M., Peterson, C.M., Sundfeldt, K., Norden, B., Hedin, L. and Enerback, S. (1997) The transcription factor C/EBP-beta and its role in ovarian function; evidence for direct involvement in the ovulatory process. *EMBO J.* **16**, 5273-5279.

Poncz, M., Henthorn, P., Stoeckert, C. and Surrey, S. (1988) Globin gene expression in hereditary persistence of fetal haemoglobin and (delta beta) naught-thalassaemia. *Oxf Surv Eukaryot Genes.* **5**, 163-203.

Roman, C., Platero, J.S., Shuman, J. and Calame, K. (1990) Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP. *Genes Dev.* **4**, 1404-15.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996) Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *EMBO J.* **15**, 143-149.

Scott, L.M., Civin, C.I., Rorth, P. and Friedman, A.D. (1992) A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells. *Blood.* **80**, 1725-1735.

Screpanti, I., Romani, L., Musiani, P., Modesti, A., Fattori, E., Lazzaro, D., Sellitto, C., Scarpa, S., Bellavia, D., Lattanzio, G. and et al. (1995) Lymphoproliferative disorder and imbalanced T-helper response in C/EBP beta-deficient mice. *EMBO J.* **14**, 1932-1941.

Sieweke, M.H. and Graf, T. (1998) A transcription factor party during blood cell differentiation. *Curr Opin Gen Dev.* **8**, 545-551.

Sterneck, E., Tessarollo, L. and Johnson, P.F. (1997) An essential role for C/EBP beta in female reproduction. *Genes Dev.* **11**, 2153-2162.

Strouboulis, J., Dillon, N. and Grosveld, F. (1992) Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev.* **6**, 1857-1864.

Tanaka, T., Yoshida, N., Kishimoto, T. and Akira, S. (1997) Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. *EMBO J.* **16**, 7432-7443.

Thein, S.L. (1993) beta-Thalassaemia. *Baillieres Clin Haematol.* **6**, 151-175.

Vinson, C.R., Hai, T. and Boyd, S.M. (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.* **7**, 1047-1058.

Wall, L., deBoer, E. and Grosveld, F. (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev.* **2**, 1089-1100.

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

Wang, N.D., Finegold, M.J., Bradley, A., Ou, C.N., Abdelsayed, S.V., Wilde, M.D., Taylor, L.R., Wilson, D.R. and Darlington, G.J. (1995) Impaired energy homeostasis in C/EBP alpha knockout mice. *Science.* **269**, 1108-1112.

Wong, P.M.C., Chung, S.W., Chui, D.H.K. & Eaves, C.J. (1986) Properties of the earliest clonogenic hemopoietic procursors to appear in the developing murine yolk sac. *Proc. Natl. Acad. Sci. USA.* **83**, 3851-3854.

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition. *Genes Dev.* **10**, 2894-2902.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J. and Tenen, D.G. (1997) Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A.* **94**, 569-574.

Zinszner, H., Kuroda, M., Wang, X.Z., Batchvarova, N., Lightfoot, R.T., Remotti, H., Stevens, J.L. and Ron, D. (1998) CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes Dev.* **12**, 982-995.

Chapter 5

General Discussion

General Discussion

The human genome is composed of $3x10^9$ base pairs of DNA. If unraveled, the genome would extend for about two meters. However, it is compacted into a nucleus of about 10 μm. It would seem obvious that the folding of DNA presents many impediments to any metabolic processes requiring access to the double helix. Remarkably, functions such as DNA replication, transcription, recombination and repair are very efficiently carried out *in vivo* and thus the genetic information stored in the linear sequence of DNA are competently utilized. The dynamics of these processes are possible because the DNA is not folded in exactly the same way, and the manner in which a region of the genome is packaged in different cells influences the activity of the genes it contains.

Differences in the packaging of the chromatin fibre are reflected by differences in sensitivity to nucleases. In particular, regions in an open chromatin conformation are detected as sites that are hypersensitive to DNaseI cleavage. These regions are associated with actively transcribed genes or poised, competent genes. For example, DNaseI readily digests globin genes in red blood cells nuclei but not in fibroblasts (Weintraub and Groudine, 1976). DNaseI hypersensitive sites appear before a gene is transcribed (Litt et al., 2001) and they may contribute to establish domains of general sensitivity to nucleases. Such domains and their formation are an important pre-requisite for transcription.

The establishment of domains of gene regulation

A series of experiments involving DNaseI digestion, sedimentation, fluorescent in situ hybridization and electron microscopy measurements have established a strong case for the presence of large independent loops (50-100 kb) of chromatin fibre (Benyajati and Worcel, 1976; Cook and Brazell, 1975; Sedat and Manuelidis, 1978). The loops have an average length of about 0.5 μm. If organized predominantly into looped domains, a typical human chromosome might contain more than 2000 of them. This notion has lead to the almost generally accepted concept that in the eukaryotic nucleus these physical structures contain the transcribed regions of the genome. Thus, transcribed regions are organized into independently organized domains, each containing the necessary regulatory elements for the correct expression of the genes located within (Bell and Felsenfeld, 1999; Dillon and Grosveld, 1994). Several biochemical assays have been used to describe the role of chromatin in the regulation of eukaryotic gene expression. On the other hand the only direct evidence that links loop formation and gene transcription has come from cytological and genetic studies of the fruit fly *Drosophila melanogaster*. Almost 40 years ago Ritossa showed that in the polytene

chromosomes of *Drosophila melanogaster* the chromatin fibre of a specific band decondenses and forms puffs when high levels of transcription are induced (Ritossa, 1962). Since then chromatin has gone in and out of fashion and only in the last ten years a more robust scientific effort has been made to understand the mechanisms underlying the development of such a remarkable phenomenon. A direct link between chromosome architecture and gene expression has been recently reported. An 880-base pair deletion, responsible for the altered expression of the *facet-strawberry* allele of *Notch* also eliminates an interband and fuses 3C7 and 3C6 in polytene chromosomes (Vazquez and Schedl, 2000). This finding is consistent with a large body of evidence that is accumulating for the existence of genetic elements that are able to create and delimit functional domains of gene expression. The presence of a domain of independent gene activity protects genes with their own pattern of programmed expression from other DNA sequences that have the potential to affect their expression. If the β-globin locus is used as an example, one can say that the localized chromatin changes and the exclusive expression of the globin genes in the erythropoietic tissue are a consequence of epigenetic events that happen solely in the erythroid lineage. This chain of events is possible because there are genetic elements in the locus that set up, delimit and organize a domain of independent gene activity. It is important to point out at this stage that these elements called borders and insulators are able to protect a test gene against position effects and to block the influence of an enhancer on a promoter. However, most of the properties of these elements have been determined using artificial constructs and interpretation of the data has much suffered from the design of the experiments. Fortunately, new evidence has shed some light on the role of borders and insulators in their native genomic sites. Evidence points to an important role for nuclear compartmentalization in the events leading to gene activation. This evidence derives mainly from studies in *Drosophila melanogaster* genetic on the transposable element, *gypsy*. Insertion of this *gypsy* element is found near the *yellow* gene (Geyer and Corces, 1992). The presence of a *gypsy* element upstream of the *yellow* promoter prevents the enhancer located 5' of this insertion from activating *yellow*. The protein suppressor of hairy wing Su(Hw) is essential to enhancer-blocking properties. This effect is modulated by another protein, mod(mdg4) that binds to the Su(Hw) protein (Gerasimova et al., 1995). It has been shown that the Su(Hw)/mod(mdg4) complex tends to be arranged in clusters near the nuclear periphery of interphase diploid cells (Gerasimova and Corces, 1998) and the complex may be tethered to the nuclear lamina, creating a series of separate loop domains. Additional evidence for this model shows not only that the nuclear localization of a DNA sequence near a *gypsy* element overlaps with the mod-(mdg4) protein cluster but more interestingly that copies of the *gypsy* element normally found at separate nuclear locations

now co-localize at the nuclear periphery. The effects are dependent on the expression of Su(Hw) protein (Gerasimova et al., 2000). In contrast to the known insulating properties of single Su(Hw) elements, insertion of a pair of Su(Hw) elements between an enhancer and a promoter neutralizes insulator activity and the enhancer-promoter interaction is instead facilitated (Cai and Shen, 2001; Muravyova et al., 2001). The implication is that two insulators interact, probably through the protein complexes bound to them. Therefore, insulator function could in principle be mediated by bound cofactors to either repress or facilitate transcription.

In the loop domain model, an enhancer and promoter can only interact when they are within the same loop. When proteins like the bound Su(Hw)/md-(mdg4) create two separate loops, the positional enhancer blocking effect takes place. The loop model can also explain protection against position effects because the presence of such a structure also interferes with the extension of adjacent condensed chromatin structure. Significant in this context is that the normal role of mod(mdg4) might be the formation of heterochromatin (Gerasimova et al., 1995). The loop model would also leave an open avenue for genes to be regulated by transvection. Cross talk between two spatially adjacent active loci on different loops may be modulated by facilitators like the proteins *Chip* and *Nipped-B* that can promote dimerization of DNA bound proteins (Dorsett, 1999) and perhaps stabilize the chromatin loop domains (Bulger and Groudine, 1999). The looping model can also be used to explain the regulation of the β globin locus. This model has been invoked several times to justify how the genes are transcribed within this active domain (Dillon and Grosveld, 1994; Dillon et al., 1997; Wijgerde et al., 1995). Although there is a plethora of data suggesting that this may be the case, direct proof that loops are formed is still lacking.

The β globin locus is bordered by DNaseI hypersensitive elements that separate this erythroid gene cluster from other genes with different tissue expression (Bulger et al., 1999). The biggest clue on how erythroid specific expression of the globin genes is achieved came from the finding that in several thalassemias, deletions encompass about 100 kb upstream of the locus. This area was shown to contain DNaseI hypersensitive sites spread over 20 kb 5' of the ε globin gene (Tuan et al., 1985). We now know that this region contains the LCR. In transgenic animals, a gene linked to the LCR is expressed at high levels and erythroid expression is achieved in most of the integration sites (Grosveld et al., 1987). However, ectopic expression is often seen too (Chapter 2). Therefore, it is likely that other elements besides the LCR are present in the native locus that restrict the action of the LCR and hence the expression of the globin genes to the erythroid tissue.

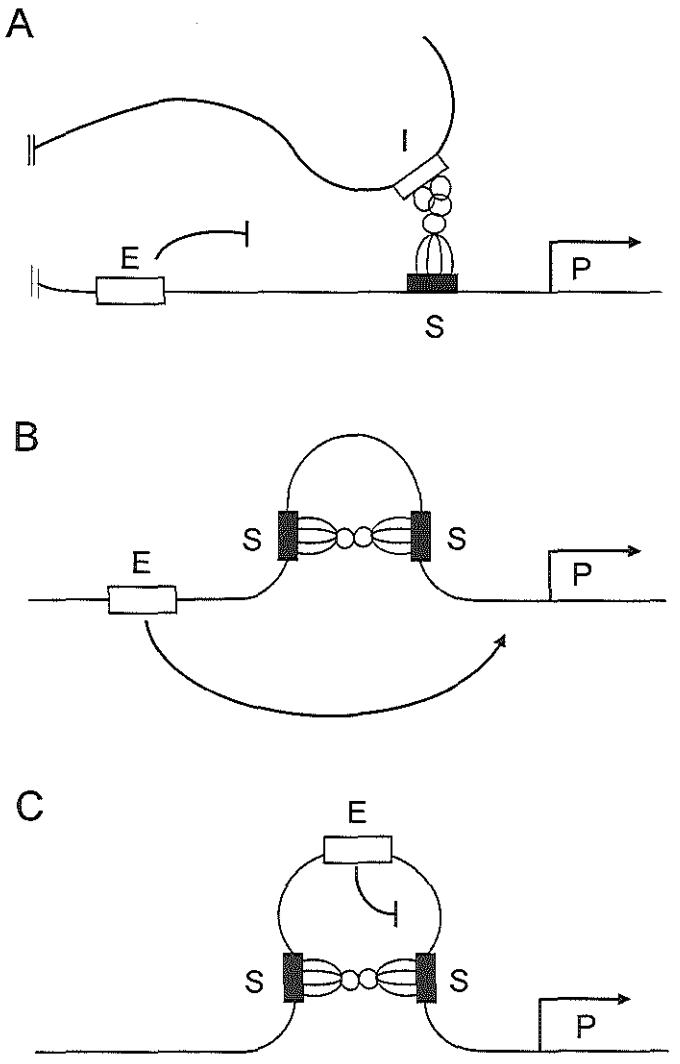

Fig. 1. Insulator-mediated loop formation.
(**A**) A su(Hw) insulator (S) may interact with other sites/insulators (I), separating the enhancer (E) and the promoter (P) into distinct domains and blocking their interaction. (**B**) Interactions between two tandem su(Hw) insulators fail to sequester the enhancer and may even facilitate enhancer-promoter interaction by looping out the intervening DNA. (**C**) Enhancer blocking may be strengthened by the preferred interaction between two su(Hw) insulators flanking the enhancer. (Adapted from Cai and Shen, 2001)

It is noteworthy in this regard that the smallest LCR deletions, such as those found in Hispanic and γβ thalassemias remove an additional 30 to 100 kb upstream of the LCR (Forrester et al., 1990; Kioussis et al., 1983; Taramelli et al., 1986; Vanin et al., 1983).

The 3-kb sequences immediately upstream of the LCR do contain another hypersensitive site, HS5. We show that this site is erythroid specific but that it does not behave like a classical insulator. Unlike a *gypsy* element it does not protect a linked transgene from position effect. The same result holds when the rest of the 30-kb sequences are interposed between the β globin test gene and the LCR. No significant downregulation of the β globin gene is achieved. Recent reports have suggested that these sequences contain two additional hypersensitive sites, HS6 and HS7 (Bulger et al., 1999). However, only the presence of HS6 has been independently confirmed. This site probably corresponds to the retrovirus ERV-9 Long Terminal Repeat (LTR) (Long et al., 1998). The LTR could in principle act as a "factors sink" and thus interfere with LCR function. The data presented in chapter 2 does not support this simple model. Moreover, additional experiments should be performed to strengthen the data. Rather than increasing the number of transgenic animal for the constructs used in the study, it would be preferable to use larger constructs like PACs or BACs containing the LCR and at least 100-kb of upstream sequences. The β globin gene, or another reporter gene replacing the β globin coding sequences, could be placed via homologous recombination (Imam et al., 2000) at the same positions 5' of the LCR like those in the constructs described in chapter 2. Even more valuable insight in the role of insulators could be gathered if a human globin gene would be placed at homologous positions upstream of the LCR in the endogenous mouse β globin locus. The advantage of such an experiment is that the location of the domain of gene activity can be determined, at least for the mouse β globin locus for in its natural chromosomal environment. Finally, it would be interesting to find out whether a mammalian insulator would also function in a different location. For example, this element could be placed between the γ globin and the β globin gene to test whether the β globin gene is excluded from the active domain.

Regulation of globin transcription through the CCAAT boxes.

The β globin promoter is sufficient, albeit at low levels, to drive expression of a linked gene in transgenic mice. Inducible and high level expression is only achieved when these sequences are linked to the LCR (Grosveld et al., 1987). Deletion of either the CCAAT or the CCAC boxes lowers expression of the β globin gene 7-fold and deters induction. Deletion of both elements reduces expression to almost baseline (Antoniou et al., 1988; Antoniou and Grosveld, 1990). The requirement for both elements has been shown from a different angle. The TATA box region of the β globin gene binds only TFIID. However, the association of TFIID with the TATA box is not the rate-limiting factor in the rate of initiation of

transcription. Only mutations that drastically affect the binding of TBP result in decreased levels of transcription. A threshold value of TBP binding of 15-30 % is sufficient to ensue normal levels of transcription (Antoniou et al., 1995). This suggests that the presence of the CCAAT and the CCAC boxes are required for positioning the initiation complex. In vivo studies have demonstrated that EKLF is the functional factor on the β globin CCAC box (Nuez et al., 1995; Wijgerde et al., 1996).

NF-E6 was originally detected as a binding activity on the β-globin CCAAT box in MEL cells (Berry et al., 1992). This activity is also found on the CCAAT boxes of the γ-globin genes (Ronchi et al., 1996). It has been shown that NF-E6 displays the same mobility on the β CCAAT box as DSFr (Delvoye et al., 1993) and that it is immunologically related to C/EBPγ (Wall et al., 1996), and Chapter 4. C/EBPγ is a member of the CCAAT/enhancer binding protein (C/EBP) family of transcription factors (Lekstrom-Himes and Xanthopoulos, 1998; Yamanaka et al., 1998) for review. The active form of these proteins comprises two polypeptide chains held together by short amphipathic α helices. The dimer interface, termed the leucine zipper, serves as a dimerization motive. A polypeptide segment immediately amino terminal to the leucine zipper is rich in basic amino acids. Dimerization brings two such 'basis regions' into close apposition, forming a bivalent DNA contact surface (Landschulz et al., 1988). The basic regions are structurally disordered in the absence of DNA but become α helical once bound (Shuman et al., 1990). C/EBPγ was identified by its ability to bind the Ig heavy chain promoters and the Ig heavy chain enhancer and originally named Immunoglobulin/enhancer binding protein (Ig/EBP). Its mRNA encodes a 16.4-kD protein that lacks an activation domain (Roman et al., 1990). C/EBPγ alone neither activates nor represses transcription. C/EBPγ is a transdominant negative inhibitor of the other C/EBP family activators, such as C/EBPα and C/EBPβ. C/EBPγ is ubiquitously expressed and its widespread presence suggests it may play a role in modulating gene activation resulting from induction of other partners that are activators (Cooper et al., 1995).

C/EBP proteins all heterodimerize with one another and in general do not dimerize with non-C/EBP bZip proteins. However some of them, like C/EBPβ, can form heterodimers with members of the activating transcription factor (ATF) family of proteins (Vinson et al., 1993). Although ATF proteins have a different binding specificity, they are related to C/EBP proteins in that they are bZip proteins too. Perhaps not surprisingly, heterodimers of ATF members with C/EBP proteins bind to a different set of DNA sequences.

Our data show that the binding activity of NF-E6 on the β-globin CCAAT box decreases in chemically induced MEL cells, while RNA levels of C/EBPγ do not change during this

differentiation process (Chapter 4 and unpublished data). These observations suggest that either post-transcriptional modification prevent C/EBPγ from forming homodimers or that C/EBPγ forms a heterodimer complex with a different factor that alters its binding specificity. Data presented in Chapter 3 demonstrate that a fraction of NF-E6 is present in a large multiprotein complex and that the free fraction probably contains another polypeptide. Determination of the identity of the ubiquitous and erythroid specific components of the NF-E6 complex will help understand the function of this complex in the regulation of globin gene transcription and the role it plays in the development of the haematopoietic system.

In conclusion, the observation that there is a correlation between the level of expression of C/EBPγ and the γ-globin genes further strengthens the hypothesis that factors binding to the CCAAT boxes play an important role in the regulation of the globin genes (Chapter 4). Recently C/EBPγ knockout mice have been generated (Kaisho et al., 1999). The mice show a high mortality rate within 48 hours after birth, suggesting that C/EBPγ is important for neonatal survival. Furthermore, C/EBPγ KO mice have impaired Natural Killer (NK) cytotoxic activity and IFN-γ production, revealing that this factor is also critical in the functional maturation of NK cells. Therefore, although C/EBPγ is ubiquitously expressed, a specific function of this factor is performed in the haematopoietic tissue. In the light of these results, it would be interesting to test whether the switch from foetal genes to adult globin genes is correctly timed in C/EBPγ KO mice that carry the human β-globin locus transgene. If C/EBPγ is involved in regulating the transition between foetal and adult globin gene expression, the levels of these genes or their switching time should be affected in the C/EBPγ knockout background.

References

Antoniou, M., de Boer, E., Spanopoulou, E., Imam, A. and Grosveld, F. (1995) TBP binding and the rate of transcription initiation from the human beta-globin gene. *Nucleic Acids Res*, **23**, 3473-80.

Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988) The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *Embo J*, **7**, 377-84.

Antoniou, M. and Grosveld, F. (1990) beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev*, **4**, 1007-13.

Bell, A.C. and Felsenfeld, G. (1999) Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev*, **9**, 191-8.

Benyajati, C. and Worcel, A. (1976) Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. *Cell*, **9**, 393-407.

Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, **358**, 499-502.

Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, **13**, 2465-77.

Bulger, M., van Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R., Groudine, M. and von Doorninck, J.H. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A*, **96**, 5129-34.

Cai, H.N. and Shen, P. (2001) Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science*, **291**, 493-5.

Cook, P.R. and Brazell, I.A. (1975) Supercoils in human DNA. *J Cell Sci*, **19**, 261-79.

Cooper, C., Henderson, A., Artandi, S., Avitahl, N. and Calame, K. (1995) Ig/EBP (C/EBP gamma) is a transdominant negative inhibitor of C/EBP family transcriptional activators. *Nucleic Acids Res*, **23**, 4371-7.

Delvoye, N.L., Destroismaisons, N.M. and Wall, L.A. (1993) Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. *Mol Cell Biol*, **13**, 6969-83.

Dillon, N. and Grosveld, F. (1994) Chromatin domains as potential units of eukaryotic gene function. *Curr Opin Genet Dev*, **4**, 260-4.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997) The effect of distance on long-range chromatin interactions. *Mol Cell*, **1**, 131-9.

Dorsett, D. (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev*, **9**, 505-14.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*, **4**, 1637-49.

Gerasimova, T.I., Byrd, K. and Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol Cell*, **6**, 1025-35.

Gerasimova, T.I. and Corces, V.G. (1998) Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*, **92**, 511-21.

Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O. and Corces, V.G. (1995) A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell*, **82**, 587-97.

Geyer, P.K. and Corces, V.G. (1992) DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev*, **6**, 1865-73.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-85.

Imam, A.M., Patrinos, G.P., de Krom, M., Bottardi, S., Janssens, R.J., Katsantoni, E., Wai, A.W., Sherratt, D.J. and Grosveld, F.G. (2000) Modification of human beta-globin locus PAC clones by homologous recombination in Escherichia coli. *Nucleic Acids Res*, **28**, E65.

Kaisho, T., Tsutsui, H., Tanaka, T., Tsujimura, T., Takeda, K., Kawai, T., Yoshida, N., Nakanishi, K. and Akira, S. (1999) Impairment of natural killer cytotoxic activity and interferon gamma production in CCAAT/enhancer binding protein gamma-deficient mice. *J Exp Med*, **190**, 1573-82.

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983) Beta-globin gene inactivation by DNA translocation in gamma beta- thalassaemia. *Nature*, **306**, 662-6.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759-64.

Lekstrom-Himes, J. and Xanthopoulos, K.G. (1998) Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J Biol Chem*, **273**, 28545-8.

Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N. and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *Embo J*, **20**, 2224-35.

Long, Q., Bengra, C., Li, C., Kutlar, F. and Tuan, D. (1998) A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region. *Genomics*, **54**, 542-55.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V. and Georgiev, P. (2001) Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science*, **291**, 495-8.

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316-8.

Ritossa, F. (1962) *Experientia*, **18**, 571-73.

Roman, C., Platero, J.S., Shuman, J. and Calame, K. (1990) Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP. *Genes Dev*, **4**, 1404-15.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996) Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *Embo J*, **15**, 143-9.

Sedat, J. and Manuelidis, L. (1978) A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol*, **42**, 331-50.

Shuman, J.D., Vinson, C.R. and McKnight, S.L. (1990) Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science*, **249**, 771-4.

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. and Grosveld, F.G. (1986) Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res*, **14**, 7017-29.

Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A*, **82**, 6384-8.

Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983) Unexpected relationships between four large deletions in the human beta- globin gene cluster. *Cell*, **35**, 701-9.

Vazquez, J. and Schedl, P. (2000) Deletion of an insulator element by the mutation facet-strawberry in Drosophila melanogaster. *Genetics*, **155**, 1297-311.

Vinson, C.R., Hai, T. and Boyd, S.M. (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev*, 7, 1047-58.

Wall, L., Destroismaisons, N., Delvoye, N. and Guy, L.G. (1996) CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells. *J Biol Chem*, **271**, 16477-84.

Weintraub, H. and Groudine, M. (1976) Chromosomal subunits in active genes have an altered conformation. *Science*, **193**, 848-56.

Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition. *Genes Dev*, **10**, 2894-902.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, **377**, 209-13.

Yamanaka, R., Lekstrom-Himes, J., Barlow, C., Wynshaw-Boris, A. and Xanthopoulos, K.G. (1998) CCAAT/enhancer binding proteins are critical components of the transcriptional regulation of hematopoiesis (Review). *Int J Mol Med*, **1**, 213-21.

Summary

About 300 years ago Antoni van Leeuwenhoek, using a good quality simple magnifying lens observed nuclei and unicellular organisms, including bacteria. However it wasn't until the middle of the nineteenth century that the declaration '*omnis cellula e cellula*' (every cell originates from a cell) postulated that the cell is the basic unit of structure and function in living organisms.

Nowadays it is common knowledge that DNA codes for the blueprint of cellular organisms as well as many viruses. The information is stored in the linear sequence of DNA in the form of genes. A gene can be defined as the smallest unit of an organism that is capable of transmitting genetic information and expressing genetic information.

Classical genetics has demonstrated that all somatic cells of an organism carry the same genetic complement, that is they contain the same number of chromosomes. Despite this, cells in multicellular organisms show a wide variation in structure and function. In other words, because the number of characteristics of any individual cell vastly outnumbers the chromosomes, each chromosome must carry many genes. Genes, as we understand them nowadays, are DNA sequences of an organism which encode all the RNA and protein molecules that are needed to construct its cells and regulate its metabolism. A number of these genes are functional, or as we say expressed, in different cells or at different times, so that their product is synthesised according to circumstances and demand.

In this thesis I have explored some of the mechanisms that direct specificity and temporal expression of genes. The studies presented are focused on the regulation of expression of the β globin genes in the erythropoietic tissue.

In humans, the β globin locus contains five functional genes, which are arranged in the same order as they are expressed during development. The first gene in the cluster is the ε-globin gene. Its expression is restricted mainly to the embryonic stage. Two γ-globin genes, expressed during the foetal stage, follow the ε-globin gene. The last two functional genes in the locus, δ- and β-globin, are expressed from birth into adulthood. Regulation of the β–globin genes is accomplished through the gene's own regulatory sequences and the LCR, an element that resides 5 to 25 kb upstream of the ε–globin gene. The LCR is composed of five hypersensitive sites (HS). Detailed analysis of each site HS1-4 in transgenic animals has shown that each site has distinct functions but that all the sites work together as a complex. Integrity of the LCR is necessary for its correct function.

The first part of thesis, presented in Chapter 2, is a detailed molecular study of the most 5' hypersensitive site of the LCR, HS5. The results obtained demonstrate that this site is

erythroid specific, but unlike the other hypersensitive sites of the LCR, it does not provide high levels of erythroid specific expression of a linked globin gene. Because of the peripheral location of HS5 in the β-globin locus, it was hypothesised that the function of this hypersensitive site was to mark the 5' border of the β-globin domain possibly by restricting the action of the LCR in a directional manner. Therefore we also tested this site for border and insulator activity. However we found that neither of these functions is associated with HS5. These result are in clear disagreement with the currently published data that suggests that HS5 is not erythroid specific and that this element has insulator activity.

Inducible activation by the LCR requires that the minimal promoter of the globin genes contain a TATA box, a CCAAT box and a CACC box. In Chapter 3 and 4 we describe and then determine that C/EBPγ is one of the components of NF-E6, a factor that binds to the CCAAT boxes of the globin genes. In Chapter 4 we investigate the function of C/EBPγ in the regulation of globin gene expression. We demonstrate that C/EBPγ is involved in the molecular mechanisms that are active in the foetal liver during the switch of expression from foetal to adult haemoglobin. Interference with the physiological levels of C/EBPγ by overexpression of either wild type or a dominant negative isoform of C/EBPγ also results in a fatal haematopoietic defect. Furthermore, we show in Chapter 3 that the purification of the NF-E6 DNA binding activity from erythropoietic cells yields at least two polypeptides. This adds evidence to the hypothesis that the active NF-E6 complex contains at least another polypeptide in addition to C/EBPγ. We suggest that C/EBPγ dimerises with another protein present in the blood tissue to form an erythroid specific activity.

Sommario

Circa 300 anni fa' Antoni van Leewenhoek, tramite l'uso di una semplice lente d'ingrandimento di buona qualita', osservo' nuclei e organismi unicellulari, batteri compresi. Ciononostante, soltanto alla meta' del diciannovesimo secolo con la dichiarazione '*omnis cellula e cellula*' (ogni cellula origina da una cellula) fu postulato che la cellula e' l' unita' basilare, strutturale e funzionale di un organismo vivente.

Oggigiorno e'nozione comune che il DNA contiene il codice genetico degli organismi cellulari e della maggioranza dei virus. Le informazioni sono depositate nella sequenza lineare del DNA nella forma di geni. Un gene puo' essere definito come la piu' piccola unita' di un organismo capace di trasmettere ed esprimere informazioni genetiche.

La Genetica classica ha dimostrato che tutte le cellule somatiche di un organismo hanno lo stesso complemento genetico, cioe' contengono lo stesso numero di cromosomi. Nonostante questo, le cellule negli organismi multicelullari mostrano una grande varieta' in struttura e funzione. In altre parole, poiche' le caratteristiche che una cellula possiede sono in numero maggiore rispetto al numero di cromosomi, ogni cromosoma deve contenere molti geni. I geni, secondo la corrente interpretazione, sono sequenze di DNA presenti in ogni organismo le quali codificano tutte le molecole di RNA e proteine necessarie per costruire le sue cellule e regolare il suo metabolismo. Alcuni geni sono funzionali, o come diciamo noi espressi, in cellule diverse oppure in tempi diversi, cosicche' il loro prodotto e' sintetizzato a seconda delle circostanze e richieste.

In questa tesi ho studiato alcuni tra i meccanismi che dirigono l'espressione specifica e temporale dei geni. Gli studi presentati sono focalizzati sulla regolazione dei geni della β globina nel tessuto eritropoietico.

Negli esseri umani, il locus della β globina contiene cinque geni funzionali che sono posizionati nello stesso ordine in cui sono espressi durante lo sviluppo. Il primo gene nel locus e' il gene della ε-globina. L'espressione di questo gene e' ristretta principalmente al periodo embrionale. I due geni della γ-globina, espressi durante il periodo fetale, seguono il gene della ε-globina. Gli ultimi due geni funzionali nel locus, il δ- e la β-globina, sono espressi sin dalla nascita. I geni della β-globina sono regolati tramite le sequenze regolatorie del gene proprio e la LCR, un elemento che risiede all'incirca da 5 a 25 kb a monte del gene della ε-globina. La LCR (Regione che Controlla il Locus) e' composta da cinque siti ipersensitivi (HS). L'analisi dettagliata dei siti 1,2,3 e 4 in animali transgenici ha dimostrato che ogni sito ha funzioni distinte ma che tutti i siti lavorano insieme come un complesso. L'integrita' della LCR e' necessaria per la sua corretta funzione.

La prima parte della tesi, presentata nel Capitolo 2, e' un dettagliato studio molecolare del piu' 5' sito ipersensitivo della LCR, HS5. I risultati ottenuti dimostrano che questo sito e' eritroide specifico, ma al contrario degli altri siti ipersensitivi della LCR, non provvede alti livelli di espressione specifica eritroide ad un gene della globina. A causa della locazione periferale di HS5 nel locus della β-globina, e' stato ipotizzato che la funzione di questo sito ipersensitivo e' di demarcare il confine 5' del dominio della β-globina. Probabilmente questo sito delimita l'azione della LCR in maniera direzionale. Percio' abbiamo indagato l'attivita' demarcante e isolatrice di questo sito. Tuttavia abbiamo constatato che nessuna di queste due funzioni e' associata con HS5. Questi risultati sono in chiaro disaccordo con i dati correntemente pubblicati che suggeriscono che HS5 non e' eritroide specifico e che questo elemento funziona come un isolatore.

L'attivazione inducibile tramite la LCR richiede che il promotore minimo dei geni della globina contenga una cassetta TATA, una cassetta CCAAT e una cassetta CACC. Nei Capitoli 3 e 4 descriviamo e quindi determiniamo che C/EBPγ e' uno dei componenti di NF-E6, un fattore che si lega alle cassette CCAAT dei geni della globina. Nel Capitolo 4 investighiamo la funzione di C/EBPγ nella regolazione dell'espressione dei geni della globina. Chiaramente dimostriamo che C/EBPγ e' coinvolto nei meccanismi molecolari che sono attivi nel fegato fetale durante il cambiamento di espressione da emoglobina fetale ad emoglobina adulta. L'interferenza con i livelli fisiologici di C/EBPγ tramite l'espressione forzata di entrambe la forma normale o di una isoforma che funziona da dominante negativa di C/EBPγ risulta fatale al sistema ematopoietico. Inoltre dimostriamo nel Capitolo 3 che la purificazione di NF-E6 da cellule eritropoietiche produce almeno due polipeptidi. Questo risultato conferma l'ipotesi che il complesso attivo di NF-E6 contiene almeno un'altro polipeptide in aggiunta a C/EBPγ. In conclusione il modello da noi proposto prevede che C/EBPγ dimerizza con un'altra proteina presente nel tessuto sanguigno e forma un'attivita' specifica eritroide.

Samenvatting

Zo'n 300 jaar geleden nam de grote Nederlandse wetenschapper Antoni van Leeuwenhoek voor het eerst celkernen en eencellige organismen waar met niet veel meer dan een simpele lens van goede kwaliteit. Pas in het midden van de 19e eeuw leidde dit tot de hypothese "*omnis cellula e cellula*" (elke cel komt voort uit een andere cel), met als consequentie dat de cel de basiseenheid is van structuur en functie in alle levende organismen.

Nu is het algemeen bekend dat het DNA codeert voor alle informatie in cellen en vele virussen. Deze informatie is opgeslagen in de lineaire volgorde van het DNA in de vorm van genen. Een gen kan beschreven worden als de kleinste eenheid in een organisme dat erfelijke informatie door kan geven en tot expressie kan brengen.

De klassieke erfelijkheidsleer heeft aangetoond dat met uitzondering van de geslachtscellen alle cellen dezelfde genetische informatie bevatten, en dus hetzelfde aantal chromosomen hebben. Niettemin hebben de cellen in meercellige organismen een enorme variatie aan structuur en functie. Omdat het aantal variaties veel groter is dan het aantal chromosomen, volg hieruit dat elk chromosoom vele genen bevat. De genen bevatten alle informatie voor de RNA en eiwitmoleculen die nodig zijn om al die verschillende celtypen op te bouwen en te laten functioneren. Veel van die genen functioneren alleen in specifieke celtypen of tijdens bepaalde tijdsramen, zodat hun product alleen op gepaste wijze gemaakt wordt in die cellen en op die tijdstippen.

In dit proefschrift heb ik een gedeelte van de mechanismen die hieraan ten grondslag liggen onderzocht. Ik heb met name de expressie van de beta-globine genen in rode bloedcellen bestudeerd. De menselijke beta-globine locus bevat vijf functionele genen, die op het chromosoom liggen in de volgorde waarin ze tot expressie komen gedurende de ontwikkeling. Het eerste gen is het epsilon-globine gen dat tijdens de embryonale periode actief is. Naast het epsilon-globine gen liggen de twee gamma-globine genen, die in het foetale stadium tot expressie komen. Deze worden gevolgd door de delta- en beta-globine genen die na de geboorte geactiveerd worden en tijdens de rest van het leven actief blijven. De activiteit van deze genen wordt gecontroleerd door hun eigen lokale regulerende DNA volgordes, en door de Locus Control Region (LCR), een gebied dat zich 5 tot 25 kilobasen voor het epsilon-globine gen bevindt. De LCR is samengesteld uit vijf sleutelelementen, de hypersensitive sites (HS). Een gedetailleerde analyse van HS1-4 in transgene muizen heeft laten zien dat elke HS zijn eigen karakteristieken heeft, maar dat alle HS samen werken als een complex. De integriteit van de LCR is nodig voor het correct functioneren van de LCR.

In het eerst deel van mijn proefschrift, in hoofdstuk 2, beschrijf ik een gedetailleerde moleculaire studie van HS5 van de LCR. De resultaten laten zien dat deze HS specifiek gevormd wordt in rode bloedcellen, maar geen meetbare invloed heeft op de expressie van een gekoppeld globine gen. Omdat HS5 aan de rand van de beta-globine locus ligt, wordt verondersteld dat HS5 zou kunnen functioneren als een grenspost of isolator van het globine domein, om zo te voorkomen dat de activerende werking van de LCR zich buiten de globine genen uit zou kunnen strekken. Ik heb daarom experimenten ontworpen om deze veronderstelde activiteit van HS5 in transgene muizen te testen. Met deze experimenten heb ik geen enkele aanwijzing gevonden dat HS5 inderdaad als grenspost of isolator zou kunnen werken. Dit is in tegenspraak met enkele publicaties in de literatuur die suggereren dat HS5 niet specifiek gevormd wordt in rode bloedcellen en dat HS5 als een isolator kan werken.

Activering van de globine genen door de LCR is afhankelijk van DNA elementen in deze genen zelf, zoals TATA box, CCAAT box en CACC box elementen in de promotors. In hoofdstuk 3 en 4 beschrijf ik dat de transcriptiefactor C/EBP-gamma een component is van NF-E6, een factor die bindt aan de CCAAT boxen van de globine genen. In hoofdstuk 4 beschrijf ik de mogelijke functie van C/EBP-gamma in de controle van globine genexpressie. Ik laat zien dat C/EBP-gamma een rol speelt bij de moleculaire mechanismen die tijdens de ontwikkeling resulteren in de overgang van foetale naar volwassen globine expressie. Transgene overexpressie van wildtype en een dominant-negatieve vorm van C/EBP-gamma die niet langer aan DNA kan binden resulteert in een fataal defect in de ontwikkeling van rode bloedcellen. Het werk beschreven in hoofdstuk 3 geeft aan dat NF-E6 waarschijnlijk uit minstens twee eiwitketens bestaat. Dit suggereert dat NF-E6 naast C/EBP-gamma nog een andere eiwitpartner bevat, en dat het functioneren van dit complex verstoord wordt door transgene overexpressie van C/EBP-gamma. Ik stel een model voor waarin C/EBP-gamma in rode bloedcellen een dimeer vormt met een ander eiwit, resulterend in een complex dat een specifieke functie vervult in de deze cellen.

Curriculum vitae

Name:	Gaetano Zafarana
Date of birth	17 December 1963
Place of birth	Siracusa (Italy)
1977-1982	Secondary education: Liceo Scientifico
1983-1984	Law studies: Constitutional Law, Administrative Law
1985-1993	London, England
1986-1987	Callan School, Hammersmith and West London College. English language studies: Cambridge curriculum
1987-1989	Hammersmith and West London College: Advanced Levels in Biology and Chemistry.
1989-1992	University College London Department of Biochemistry and Molecular Biology: Honors Degree in Biochemistry and Molecular Biology
1990	Charing Cross Hospital Department of Biochemistry: Muscular Dystrophy
1991-1992	Ludwig Institute for Cancer Research Department of Transformation studies: Senescence and Immortalization
1992-1993	National Institute for Medical Research Department of Gene Structure and Expression: MRC Ph.D. studentship
1993-present	Rotterdam, The Netherlands
1993-1998	Erasmus University Rotterdam Department of Cell Biology and Genetics A.I.O. studentship
1998-present	Josephine Nefkens Institute Department of Pathology Laboratory of Experimental Patho-Oncology KWF Postdoctoral researcher.