TRANSCRIPTIONAL REGULATION OF THE MOUSE β^{MAJOR} GLOBIN GENE IN MURINE ERYTHROLEUKAEMIA CELLS.

Kay Ferguson Macleod.

Thesis submitted to the University of Glasgow
for the degree of Doctor of Philosophy.

The Beatson Institute for Cancer Research, Glasgow.

January 1990.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Mur a biodh ann ach sireadh dh'fhanadh an latha samhraidh 's cha tigeadh am bruadar gu ceann dhuinn; bhiodh an leug luachmhor ri fhaotainn, an t-seasmhachd a meadhon a'chaochlaidh, a'bhoidhchead fo chumadh an aodainn; bhiodh an ceol air bilean na maighdinn, 's an t-oran eile ri chluinntinn, 's a'chlarsach fhathast r'a seinn dhuinn.

Ruaraidh MacThomais.

What is laid down, ordered, factual, is never enough to embrace the whole truth : life always spills over the rim of every cup.

Boris Pasternak.

Come with me into the sunset and travel through the night for sunrise must follow; unless, in a flash of darkness, a man puts out his world and ours eternally. So, hand in hand let us sprint for the dawn of the morning star, incandescent in a rosy-fingered sky, and take red courage to overcome manipulated underworlds of fear, using love's incalculable atoms perpetually.

William Wolfe.

CONTENTS.

LIST OF FIGURES.

LIST OF TABLES.

ABBREVIATIONS.

DNA	deoxyribonucleic acid
cDNA	cloned DNA
RNA	ribonucleic acid
tRNA	transfer RNA
mRNA	messenger RNA
A	adenine
C	cytosine
G	guanine
T	thymine
I	inosine
N	A, C, G or T
dNTP	deoxyribonucleoside triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
ddNTP	dideoxyribonucleotide triphosphate
DNAase	deoxyribonuclease
RNAase	ribonuclease
DHSS	DNAase I hypersensitive site
CAT	chloramphenicol acetyl transferase
hGH	human growth hormone
bp	base pairs
kb	kilobase pairs
kD	kiloDaltons
ml	millilitres
µg	microgrammes
ng	nanogrammes
pmol	picomoles
nmol	nanomoles
V	Volts
mA	milliamperes
OD	optical density
% w/v	weight in grammes per 100 ml water
% v/v	volume in ml per 100 ml water
SDS	sodium dodecyl sulphate
Amp.	ampicillin
Tris	Tris (hydroxymethyl) methylamine
EDTA	Ethylene diamine tetra-acetic acid
DMSO	Dimethylsulphoxide
HMBA	Hexamethylene bis-acetamide
Il-3	Interleukin 3
Epo	Erythropoietin
GM-CSF	Granulocyte/Macrophage Colony Stimulating Factor
CFU-S	Colony Forming Unit (Spleen)
CFU-E	Colony Forming Unit (Erythroid)
BFU-E	Burst Forming Unit (Erythroid)
PMSF	phenylmethyl sulfonyl fluoride
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
MEL	mouse erythroleukaemia

ACKNOWLEDGEMENTS.

I wish to express my sincere thanks to Dr. Mark Plumb for his consistent and thorough guidance in the practical aspects of the scientific training I have obtained while working on this project. I am particularly grateful to Dr. Plumb for the vigorous and highly motivating debates we became involved in which served to develop and discipline my ideas. I am indebted to Dr. Paul Harrison for personal encouragement especially during times of self-doubt and Dr. Dafydd Llewelyn for making me laugh during the other difficult times. I am grateful to Dr. John Pitts whose criticism encouraged me to put my work into a broader biological perspective. I would like to thank all members of R4 and other people at the Beatson Institute for scientific and personal help during my years here.

I am deeply grateful to my parents for their support and for the self-respect they motivated during my years in Glasgow but also for a lifetime of encouragement and love. I shall also remember the close companionship of my sister, Donna during these years. I would like to thank Bob McTear, Billy Wolfe and Canon John Galbraith for spiritual inspiration. I am indebted to my friends, in particular Chris and Judi Poole for making the last stages bearable but also to Douglas, Murdo, Calum, Kenny, Andrew, Chris, Ivan, Marc, David, Stephen, Andrea and Bob for making life fun.

I am grateful to the Medical Research Coucil and to the Cancer Research Campaign for funding.

SUMMARY.

There is considerable evidence that certain features of the mechanism of transcriptional regulation of the β^{major} globin gene are common to other erythroid-specific globin and non-globin genes (reviewed in chapter 1). However, my work suggests that there are aspects of this regulation which are unique.

My work on the binding of nuclear proteins to the mouse β^{major} globin gene promoter has revealed a modular organisation of protein binding sites. These include two binding sites for the erythroid-specific factor, NF-E1 (at -65 and -210 bp) and binding by non-tissue-specific proteins to the CCAAT (-85 bp), CACCC (-95 bp) and Box 1 (-170 bp) motifs and to a nuclear factor 1 (NF-1) binding site (-255 bp).

Mutational analysis of the mouse β^{major} globin gene proximal promoter TATA, CCAAT and CACCC boxes by other researchers had shown that they are required for promoter function in erythroid and non-erythroid cells. Using site-directed mutagenesis, I have confirmed that the CACCC motif is required for promoter function when linked to a reporter gene and assayed in transient and stable transfection assay systems. Furthermore I have shown by mutagenesis that the NF-E1 site at -65 bp is necessary for promoter function.

However mutation of the NF-E1, Box1 and NF-1 sites in the distal region of the promoter results in up-regulation of transcription from the promoter in uninduced MEL cells and it is proposed here that while the proximal promoter is essential for promoter function that the distal promoter has a negative regulatory function in uninduced MEL cells. This negative regulatory effect is derepressed when the MEL

cells are induced to differentiate.

In addition to the role of 5′ promoter sequences in transcriptional regulation of the mouse β^{major} globin gene, I have identified sequences 3′ of the coding region of the gene which contain NF-E1 and Box 1 binding sites and which confer an up-regulatory effect on the transcription of a linked reporter gene from the mouse β^{major} globin gene promoter. The up-regulatory effect of this enhancer is greater when the cells are induced to differentiate.

Work with two different MEL cell lines, F4-12B2 and C88, and with two different inducing agents, DMSO and HMBA, suggests that there are differences in the responses of these cell lines to the same treatment and also that DMSO and HMBA act through different mechanisms to induce differen tiation.

CHAPTER 1 GENERAL INTRODUCTION.

1.1 ERYTHROPOIESIS AS A SYSTEM IN WHICH TO STUDY REGULATED GENE EXPRESSION DURING DIFFERENTIATION.

As a system in which to study differentiation, the erythropoietic system offers several advantages:

(i) the differentiation pathway along which the erythroid cell progresses from the pluripotent haematopoietic stem cell is well characterised at each stage by proliferative capacity and response to the hormone erythropoietin (Stephenson *et al.* (1971); Gregory and Eaves (1977).

(ii) gene expression in the erythroid lineage is well characterised by markers such as spectrin, glycophorins and globins (Reviewed by Harrison (1984).

(iii) there exist erythroleukemic cell lines which can be induced *in vitro* to undergo differentiation in a way that mimics the morphological and biochemical changes taking place *in vivo* during normal erythropoiesis (Friend *et al.* (1966), Beug *et al.* (1982), Lozzio and Lozzio (1975).

1.1.1 Erythropoiesis.

During the development of the mouse, which has a gestation period of 21 days, erythropoiesis occurs in the embryonic yolk sac (days 8-12), the fetal liver (days 12-17) and in the bone marrow after day 16 (Barrowman and Craig (1961); Craig and Russell (1964).

Erythropoiesis also occurs in the spleen starting at day 16 but ceases at birth. Under conditions of anaemia, erythropoiesis is induced in the adult spleen and liver (Conkie et al. (1975). In humans, erythropoiesis takes place in the embryonic yolk sac (2-6 wks post conception), the fetal liver (4 wks post conception until 6 wks after birth), the spleen (15 wks post-conception until 4 wks after birth) and in the bone marrow from 16 wks post-conception through adult life (Karlsson and Nienhuis (1985).

Erythrocytes are derived from pluripotent stem cells which become committed to different lineages in response to growth factors. Growth factor response was first studied using the colony forming unit assay (Gusella et al. (1976); Till and McCulloch (1980); Dexter et al. (1984). When bone marrow from a healthy adult mouse is injected into a mouse whose own haemopoiesis has been ablated by X-ray irradiation, the donor cells colonise the irradiated host. Such colonies develop in the host spleen and the discreteness of the nodules suggests that each colony is a clone of a single founding stem cell, which has been called a colony-forming unit (CFU-S). These colonies can be explanted and grown in culture. In combination with other culture techniques (including the growth of dispersed bone marrow cells on semi-solid medium), the CFU assay has identified the erythrocyte colony forming unit (CFU-E), which is highly sensitive to erythropoietin and gives rise to mature erythrocytes after less than 6 divisions. The erythrocyte burst forming unit (BFU-E) has also been identified and responds only to high concentrations of erythropoietin and divides at least 12 times before becoming mature (Gregory (1976). Thus the BFU-E appears to be a progenitor of the CFU-E and cells intermediate between

BFU-E and CFU-E which have intermediate proliferative capacity and erythropoietin responsiveness have been identified.

Other growth factors, including GM-CSF and interleukin 3 (Il-3 act on earlier erythroid progenitor cells) which influence different cells in the haematopoietic lineage have been purified from culture media and their genes cloned (Clark and Kamen (1987). They may eventually be used in the clinical treatment of disease (Sachs (1988).

1.1.2 Gene Expression During Erythropoiesis.

The erythropoietin receptor is the earliest known erythroid marker expressed, being switched on at the first erythroid committed stage, the BFU-E (Krantz and Goldwasser (1984). The erythropoietin receptor is internalised when it binds erythropoietin and is metabolised within the cell. The receptor has been purified and there seem to be two forms of the receptor which differ in size and affinity for erythropoietin (Krantz et al. (1988). Chemical cross-linking studies show that the high affinity erythropoietin receptor consists of two sub-units (McCaffery et al. (1989) of molecular weight 100kD and 85kD. Protease digestion of these subunits generates identical peptide fragments suggesting that they may be encoded by the same gene. A cDNA clone for the erythropoietin receptor has been obtained from MEL cells. From the sequence of the cDNA, the receptor is predicted to have a molecular weight of 55kD (D'Andrea et al. (1989). There is evidence that the molecular weight observed for the receptor on the surface of erythroblasts is generated by glycosylation and dimerisation of the 55kD polypeptide to yield the biologically important high affinity

receptor (Dube et al. (1988).

Early during erythroid differentiation, at the CFU-E stage, the α and β-spectrin genes are expressed (Eisen and Ikawa (1977). The genes encoding the other components of the red blood cell cytoskeleton are also expressed at this stage. Some of these proteins are not exclusively red blood cell specific and have in fact a wide tissue distribution (Lazarides and Nelson (1982).

The glycophorin proteins are erythroid specific and their genes are first expressed in mid-erythroblast stage (Kasturi and Harrison (1985), before activation of globin gene transcription.

Some of the enzymes of the haem biosynthetic pathway are expressed at higher levels in erythroid tissue where large quantities of haem are required for haemoglobin synthesis, compared with other cells, where the haem requirement for cytochrome synthesis is satisfied by lower levels of expression. The genes encoding the haem biosynthetic enzymes have not all been cloned but evidence suggests that expression of these genes is sequentially induced (Sassa (1987). Up-regulation of the human δ-amino levulinic acid synthetase, in erythroid tissue, is achieved by having two genes, one of which is expressed in all cells and the second of which is only expressed in erythroid tissue (Yamamoto (1985). The human porphobilinogen-deaminase gene has two promoters, one of which is regulated in an erythroid specific fashion (Romeo et al. (1987). Evidence shows that the mouse gene is similarly regulated (Grandchamp et al. (1987).

The five different isozymic forms of carbonic anhydrase show different tissue specificities. Carbonic anhydrase (CA) is the second most abundant protein in erythrocytes after globin. The predominant

form of CA in human erythrocytes is CAI although CA II is also present. (CAII is active in a number of other tissues). CA expression is developmentally regulated in erythrocytes and up-regulation of CAI expression occurs at approximately the same time as the γ to β (foetal to adult) globin switch suggesting possible coregulation of CAI and adult globin genes in development (Boyer et al. (1983). In the differentiation of MEL cells, mouse CAI levels decrease upon induction while those of CAII increase (Fraser and Curtis (1987). The mouse CAI gene has two promoters, one of which is erythroid specific (Fraser et al. (1989). This is analogous to the mouse PBG-D gene.

The globin genes are switched on in late erythroblast stage; the α-globin genes are transcribed before the β-globin genes in the differentiation of MEL cells and β-globin levels accumulate faster due to greater message stability (Orkin et al. (1975) to give equimolar amounts of α and β globin chains.

Later in erythropoiesis, the nucleus is extruded and organelles are broken down, an activity involving the "red blood cell specific" lipoxygenase-15 (Thiele et al. (1979).

1.1.3 Differentiation of Erythroleukemic Cell Lines.

The Friend virus complex was first isolated from a filtrate of mouse spleen tumour cells (Friend (1957) and consists of two C-type retroviruses, a replication defective spleen focus-forming virus (SFFV) and a replication-competent Friend murine leukaemia virus (F-MuLV). The target cell for this virus complex is one of the early erythroid progenitor cells which is erythropoietin-responsive (Mirand (1967).

95% of Friend virus induced murine erythroleukaemias have a common SFFV proviral integration site called Spi-1 (for SFFV proviral integration site) (Moreau-Gachelin et al. (1988). Transformation induced by SFFV occurs in two steps: limited proliferation which is erythropoietin independent followed by immortalisation. An intact SFFV envelope glycoprotein gp55 is necessary for transformation (Linemeyer et al. (1982) and in vitro evidence shows that gp55 binds directly to the erythropoietin receptor to generate erythropoietin and Il-3 independent cell lines (Li unpublished data). This suggests that gp55 causes prolonged proliferation of SFFV infected cells by stimulating the erythropoietin receptor and bypassing the normal requirement for erythropoietin. The role of Spi-1 as a putative oncogene is being investigated although it does not bear homology to any previously identified oncogenes. There is also evidence to suggest that the p53 oncogene is rearranged or mutated in Friend virus induced erythroleukaemias (Mowat et al. (1985).

Friend virus derived erythroleukaemia cell lines will undergo erythroid differentiation in vitro, independent of erythropoietin (Hankins and Krantz (1975) when treated with agents such as DMSO (Friend et al. (1971). Other agents which induce Friend cell differentiation include polar-planar compounds such as hexamethylene bisacetamide (HMBA), actinomycin D, hemin, purines and fatty acids (Preisler et al.(1976), Gusella et al. (1976). HMBA is a particularly strong inducer in that virtually the entire population of cells is induced at low concentrations of HMBA (4 mM) and a higher proportion of total protein synthesis is globin (Reuben et al. (1976).

Different inducing agents probably act via different pathways

since variant cell lines which are resistant to induction by one agent are sensitive to induction by another (Ohta *et al.* (1976). It is possible that inducing agents initiate reactions similar to those functioning *in vivo* but the precise nature of the reactions is not understood.

If Friend cells are incubated for 30-35hrs in DMSO, they will differentiate into erythroid cells even if the inducer is removed. If they are only incubated for 15-25 hrs with DMSO, only a small proportion of the cells will differentiate after the drug is removed. However, if these cells have first been UV-irradiated, a much larger proportion of cells differentiate suggesting that DMSO and UV act synergistically to induce differentiation (Nomura and Oishi (1983). One agent may be inhibiting DNA replication while the other possibly acts via a membrane mediated reaction. A factor has been identified which induces erythroid differentiation when introduced into non-differentiated Friend cells. This factor which is present in the cytoplasm is induced when non-erythroid and Friend cells are treated with agents that inhibit DNA replication (Nomura *et al.* (1986).

Other gene products which may be involved in the regulation of erythroid differentiation include the product of the c-*erb* A gene, the receptor for tri-iodothyronine (Sap *et al.* (1986), a hormone which is known to stimulate erythropoiesis *in vitro* (Dainiak *et al.* (1978). The v-*erb* A oncogene is known to potentiate the transformation properties of the v-*erb* B oncogene (both oncogenes are cotransduced by the avian erythroblastosis virus) to induce acute erythroleukemias and sarcomas *in vivo* (Graf and Beug (1978). A chicken cell line, HD3, generated by infection of chicken bone marrow cells with AEV carrying a temperature

sensitive v-erbB gene, fails to differentiate at 36°C but can be induced to differentiate when shifted to the non-permissive temperature of 42°C (Beug et al. (1982). The c-erb A mRNA is transcribed in late erythroblasts and has a short half-life (Hentzen et al. (1987). These results suggest that c-erb A may play a role in the late stages of erythroid differentiation.

The role of other oncogenes in erythroid differentiation is also being studied . It has been shown that normal expression of the membrane-bound tyrosine kinase product of the c-sea gene is required for normal erythropoietin-responsive erythroid differentiation (Hayman (1985). Down-regulation of expression of the c-myb gene is a prerequisite for commitment to erythropoietin responsive erythroid differentiation (Todokoro et al. (1988). c-myb and c-ets are both DNA-binding proteins located in the nucleus of erythroid progenitor cells but their role in differentiation is not yet understood (Golay et al. (1988).

Recently, a factor called erythroid differentiation factor (EDF), has been isolated from human tumour cell line THP-1 induced with phorbal 12-myristate 13-acetate (PMA) which stimulates differentiation of MEL and K562 cells (Murata et al. (1988). The factor has been purified and its gene cloned. The factor has an identical amino acid sequence to the β_A chain of activin A and inhibin A which respectively enhance and inhibit secretion of follicle-stimulating hormone from pituitary cells in vitro. In addition to this activity, activin A $(\beta_A)_2$ enhances colony formation by stimulated human bone marrow erythroid and multipotential progenitor cells (Yu et al. (1987) while inhibin A $(\alpha\beta_A)$ decreases activin but not Il-3, GM-CSF or Il-4 enhancement of

erythropoietin-stimulated colony formation by erythroid and multipotential progenitor cells (Broxmeyer et al. (1988).

Radiation induced murine erythroleukaemic cell lines have been derived which are erythropoietin responsive (Itoh et al. (1988). These cells differentiate into haemoglobinised cells if cultured in the presence of erythropoietin. Supplementation with Il-3 augments the erythropoietin-mediated differentiation although Il-3 alone does not induce differentiation (Shiozaki et al. (1989).

Various human erythroid cell lines have been established: K562 (Lozzio and Lozzio(1975) which was derived from a pleural effusion of a patient with chronic granulocytic leukaemia in terminal blast crisis; HEL cells established from a patient with Hodgkin's Disease who later developed erythroleukaemia (Martin and Papayannopoulou (1982); and KMOE derived from patients with acute erythremia (Okano (1981). These cells express erythroid markers indicative of commitment to the erythroid lineage but they also express markers of the myeloid and granulocytic lineages suggesting that they arose by transformation of early progenitor cells. The globin gene expression pattern of K562 cells is typical of erythroid cells in the embryonic stage of development expressing ε and γ globin RNAs. HEL and KMOE cells are typical of erythroid cells in the foetal stage of human development expressing both γ and β globin RNAs. Two new human erythroid cell lines have been developed, OC1-M1 (embryonic) and OC1-M2 (foetal) and from studies of expression in these five lines it appears that transformation of erythroid progenitor cells to generate erythroleukaemias activates embryonic and foetal globin gene expression patterns and that there is a failure to set up an adult pattern of

expression in human cell lines (Enver (1988).

1.2 THE ROLE OF CHROMATIN CONFORMATION IN REGULATING GENE EXPRESSION.

Given the large number of inactive genes in the eukaryotic nucleus at any one time, it is likely that the dominant mechanism of repression of gene expression is general as opposed to specific, and evidence suggests that this is mediated by condensation of chromatin in the inactive regions of the genome (Brown (1984), Weisbrod (1982).

1.2.1 Chromatin Structure and Nucleosome Phasing.

The nature of chromatin structure is reviewed at length (Pederson et al. (1986), Eissenberg et al.(1985), Dilworth and Dingwall (1988).

In decondensed chromatin, nucleosomes space out regularly along DNA (Olins (1973) but there is evidence that this positioning of nucleosomes is not random (Ramsey et al. (1984). The regular spacing of a DNA molecule around the nucleosome core is referred to as nucleosome phasing. Even when chromatin is decondensed, nucleosomes could still limit access of regulatory factors to the DNA that is to be transcribed, but there is evidence that in areas of gene activity that nucleosome phasing is disrupted (Simpson and Stafford (1983).

It has been demonstrated by methidiumpropyl-EDTA-Fe(II) cleavage that nucleosomes are phased without interruption along the mouse βmajor globin gene between -3000 and +1500 in L cells where the gene is not expressed but that this phasing is interrupted between -200 and +500 in MEL cells where the gene is expressed. This region is

however protected from MPE-Fe(II) cleavage and is flanked by MPE-Fe(II) hypersensitive sites suggesting that other factors involved in transcription are binding there (Benezra et al. (1986).

Recent work has shown that while nucleosomes assembled on DNA can prevent transcriptional initiation in vitro (Knezetic and Luse (1986), Matsui (1987), they do not prevent elongation of transcription by pre-initiated transcription complexes (Lorch et al. (1987), Knezetic (1988) and that RNA polymerase II will read through nucleosomes, displacing histones. This may explain why altered nucleosome phasing occurs over promoter and enhancer sequences but does not extend right through the gene.

The linker DNA between nucleosomes, which varies in length from 20 to 100bp depending on the organism and tissue (Kornberg (1977), complexes with histone H1 to hold adjacent nucleosomes together and to condense the 10nm fibre into a 30nm solenoid (Felsenfeld (1978), McGhee and Felsenfeld (1980). H1 is a class of polymorphic proteins which are frequently modified and exhibit tissue and organism specificity (Pederson et al. (1986). In nucleated erythrocytes H1 is replaced by H5 during maturation of the erythroblast (Neelin (1974).

Evidence suggests that beyond the 30nm fibre, chromatin is organised into loops which are radially attached to a central protein scaffold called the nuclear matrix (Zehnbauer and Vogelstein (1985). In histone depleted metaphase chromosomes, electron micrography identifies DNA loops emerging from sites on the scaffold and returning to the same site (Paulson and Laemmli (1977) in a similar fashion to lamprush chromosome loops (Callan and MacGregor (1958). Goldman (1988) has suggested that each loop is in fact a chromatin domain which contains

several transcription units and replication origins, is independently supercoilable and may be assembled during replication into an open or closed transcription state depending on the relative abundance of limiting transcription factors and histones. There is evidence suggesting that transcriptionally active DNA is closely associated with the nuclear matrix: experiments show that active proviral sequences in avian sarcoma virus transformed rat fibroblasts associate with the matrix while in the revertant, this association is lost (Cook et al. (1982). Supercoiling may generate altered DNA structures, such as DNA bending, which increase the affinity of transcription factors for regulatory sequences or it may induce altered DNA conformations such as Z-DNA which may reduce the binding affinity of histones (Wu (1988), Garner and Felsenfeld (1987).

Non-histone proteins found associated with chromatin include the high mobility group of proteins although evidence suggesting that HMG14 and 17 are found associated with nucleosomes at active sites in the chromatin (Sandeen et al. (1980), is controversial (Goodwin et al(1985). Histone acetylation and ubiquitination are also associated with gene activation (Allfrey (1977), Hebbes et al. (1988). However the role of HMG proteins, ubiquitination and acetylation in gene regulation remains ambiguous.

1.2.2 DNA Methylation.

In vitro methylation of DNA sequences inhibits their expression *in vivo*, in either mouse fibroblasts or in *Xenopus* oocytes (Vardimon *et al* (1982), Fradin *et al* (1982), Stein *et al* (1982) while 5-aza-cytidine, a hypomethylating agent activates expression of previously silent genes (Jones (1985).

In the genome methylation occurs at the 5 position of the cytosine ring in CpG dinucleotides, which are present in the genome at a frequency of approximately 0.2 times that expected (Russell *et al*. (1976). About 60-90% of all CpGs in the genome are methylated. Non-methylated CpGs tend to be clustered into what have become known as HTF (HpaII Tiny Fragments) islands because they contain many sites for HpaII which recognises the unmethylated site CCGG while its isoschizomer MspI recognises the methylated site $C^{Me}CGG$ (Bird *et al*. (1985). HTF islands are found at the 5′ end of non-tissue specific genes such as the X-linked hypoxanthine phosphoribosyl transferase gene in mouse (Yen *et al*. (1984) and human dehydrofolate reductase gene (Stein *et al*. (1983) but are also found 3′ of genes like the islands 3′ of the glucose-6-phosphate dehydrogenase gene (Wolf *et al*. (1984). Methylation of HTF islands inhibits expression of the associated gene. Bird (1986) has proposed that HTF islands mark out large regulatory sequences which are bound by multiple factors which sterically exclude methylases. Methylation at HTF islands would sterically prevent binding by regulatory factors.

Many tissue specific genes do not have HTF islands partly because CpGs are rarely found near tissue-specific genes. However,

demethylation may be important in regulating some tissue specific expression since treatment of muscle cells with 5-aza-cytidine rapidly induces differentiation (Jones (1985). Treatment of determined red blood cells with 5-aza-cytidine induces globin gene expression but has no effect on globin gene expression in non-erythroid cells implying that methylation is at most a secondary event in tissue-specific gene regulation (Hsiao et al. (1983). Methylation of the γ globin genes in MEL-human foetal erythroid somatic cell hybrids only occurs after the developmental switch to adult β-globin gene expression again suggesting a secondary role for methylation (Enver et al (1988).

The human α globin gene promoter lies in an HTF island which contains multiple copies of Sp1 binding sites which positively regulate transcription of the gene. The α globin gene functions in the absence of an enhancer when transfected into non-erythroid cell lines. In contrast, the mouse α-1 globin gene promoter is not promiscuous and does not lie in an HTF island. In non-replicating systems, the human α-globin gene becomes enhancer dependent like the mouse α-1 globin gene (Whitelaw et al. (1989). This suggests a role for methylation in regulating expression of the human α globin gene in cell-cycle dependent manner, by blocking binding of Sp1 to upstream regulatory sequences.

Methylation plays an important role in stable inactivation of viral sequences (Krucek and Dorfler (1983), in achieving equal gene dosage in mammalian females by inactivation of one X-chromosome (Venolia et al. (1982) and in generating hemizygosity and altered expression at loci in a manner dependent on the gametic origin of the locus (McGowan et al. (1989).

1.2.3 Nuclease Sensitivity of Active Chromatin Domains.

That DNA in active chromatin is preferentially sensitive to digestion by DNAase 1 compared to inactive chromatin or bulk chromatin, was initially shown by Weintraub and Groudine (1976). Digestion of isolated nuclei by DNAase 1 preferentially degrades globin DNA sequences from nuclei obtained from chick red blood cells but not from nuclei obtained from non-erythroid cells or from a population of red blood cell precursors. In contrast, the non-transcribed ovalbumin sequences in nuclei from red blood cells and fibroblasts were resistant to DNAase 1 suggesting that active genes are preferentially sensitive to DNAase 1. Stalder *et al.* (1980) extended this study by identifying a DNAase 1 sensitive region which extended from 7 kb 5'to 8 kb 3' of the β globin gene cluster in chick red blood cells. Moreover by using low concentrations of DNAase 1, he was able to identify DNAase 1 hypersensitive sites (DHSS) which were tissue-specific and some of which were developmental-stage specific. These DHSS are thought to represent sites made accessible to transcription factors by the displacement of nucleosomes. Sites identified by genomic footprinting as being nucleosome free *in vivo* have been further characterised *in vitro* as sites bound by transcription factors. For example the chicken α^D globin gene genomic footprint from -130 to +80 bp relative to the initiation site contains a binding site at -63 to -55 bp for the erythroid specific transcription factor, Eryf I as well as other protein binding sites (Kemper *et al.* (1987). In embryonic red blood cells, there is a DHSS immediately 5' of the embryonic gene, ε and in adult erythrocytes there are two sites, 6 kb and 2 kb 5' of the adult

β^A globin gene. The DHSS close to the 5′ end of the chicken β^A globin gene was more finely mapped to between 60 bp and 260 bp 5′ of the initiation site of the gene and its presence correlates with gene expression (McGhee et al. (1981).

In mouse, the α and the β globin genes are DNAase 1 sensitive in erythroid tissue and in MEL cells. These sensitive domains are present in both uninduced and induced cells even though the globin genes are not being actively transcribed in uninduced cells (Miller et al. (1978). This illustrated that while an open chromatin conformation is necessary, it is not sufficent for transcriptional activation of globin genes. Sheffery et al. (1982), Sheffery et al. (1983) mapped two DHSS in the mouse β^{major} globin gene. One site within the second intron of the gene becomes less hypersensitive upon induction while the other site, immediately 5′ of the gene becomes more hypersensitive upon induction. A third DHSS has been mapped approximately 3 kb upstream of the mouse β^{major} globin gene by Smith and Yu (1984) who also showed that following commitment to differentiation, DNAase 1 hypersensitivity is stably maintained even in the absence of inducer. In MEL cells, which only express the adult globin genes, the adult genes contain DHSS whereas the embryonic genes do not (Smith et al. (1984).

In human foetal liver, DNAasel HSS have been mapped to the 5′ regions of the human $^G\gamma$, $^A\gamma$, δ and β-globin genes whereas in human bone marrow DHSS have only been mapped 5′ of the δ and β globin genes. In the same study, no DHSS could be mapped within the β globin domain in leukocytes (Groudine et al. (1983). Tuan et al. (1985) mapped all the major and minor DHSS within the human β globin domain. All the minor sites mapped close to the 5′ ends of each gene in the domain and

correlated with transcriptional activity of the activity of the downstream gene but the major DHSS map to the boundaries of the DNAase 1 sensitive domain and are present in embryonic, foetal and adult erythroid tissues which are expressing globins. These 5′ and 3′ domain boundaries have been cloned (Grosveld et al. (1987) and brought next to the adult β globin gene in a construct which was used to generate transgenic mice. Previously, globin transgenes were poorly expressed in transgenic animals relative to the endogenous genes and were highly sensitive to the site of integration in the genome (Townes et al. (1985). But when these domains boundaries, which normally map approximately 20 kb 5′ and 3′ of the gene are cloned next to the transgene, 100% expression is observed relative to the endogenous gene. Moreover, expression is completely copy number dependent, is independent of position and orientation and confers absolute tissue specificity on expression. These dominant control regions also operate on β-globin gene promoters in stable transfection assays (Van Assendelft (1989) and will operate on heterologous promoters such as the HSV-TK, Thy-1 and α-1 globin gene (Ryan et al. (1989). It has now been found that only 6.5 kb from the 5′ domain boundary is necessary for dominant control and this region is currently being packaged into retroviral vectors for potential use in gene therapy to treat thalassemias and sickle cell anaemia (Talbot et al. (1989). Dominant control regions conferring T-cell specific, copy-number dependent, position independent expression have also been identified at 4.5 kb 5′ and 9 kb 3′ of the CD2 T-cell marker gene (Greaves et al. (1989).

DHSS marking out regulatory regions have been mapped for non-globin genes. The avian leukosis virus (ALV) integrates in bursal

lymphomas into a region flanked by two DHSS in the 5' region of the unrearranged c-myc gene. Following integration, the two DHSS disappear and a new single DHSS appears in the ALV LTR enhancer reflecting the change in transcriptional regulation of the DNA in the region (Schubach and Groudine (1984).

1.3. REGULATION OF TRANSCRIPTIONAL INITIATION.

Regulation of gene expression by control of initiation of transcription has been widely studied in a number of systems (Reviewed by Maniatis et al. (1987), Jones et al. (1988).

Some promoter elements regulate basal level transcription, such as the Sp1 binding site (Dynan and Tjian (1983) while others control inducible expression, such as the heat shock responsive element (Pelham (1982) or tissue-specific expression, such as the octamer box of immunoglobulin gene promoters in B-cells (Grosschedl and Baltimore (1985).

The distinction between promoter and enhancer elements has become arbitrary but initially enhancers were distinguished from promoters by their ability to stimulate transcription initiation in an orientation and position independent fashion (Serfling et al. (1985). As analysis of enhancer elements has become more detailed it emerges that each enhancer is itself made up of core elements or "enhansons", which can be duplicated or combined in a heterologous fashion to generate new enhancers. Within a given enhancer, the relative spacing of enhansons is critical and determines the overall efficiency of the enhancer (Ondek et al. (1988). Consequently, an element such as the

octamer box can act as a promoter element in the context of the sea urchin histone H2B 5′ flanking sequence where it is position and orientation dependent and as an enhanson in the context of the SV40 enhancer where it interacts with other enhansons to generate the functional enhancer which is orientation and position independent. It has been shown that multimerisation of the octamer motif to generate a tissue-specific enhancer will create a negative regulator which is dominant to the SV40 enhancer (Yu et al. (1989) illustrating that the activity of a particular enhanson depends upon the context in which it is found.

1.3.1 THE MODULARITY OF ENHANCERS AND PROMOTERS EXEMPLIFIED BY THE SV40 ENHANCER.

The 300 bp region between the early and late transcription units of SV40 contains a number of regulatory elements as illustrated in figure 1.1 below.

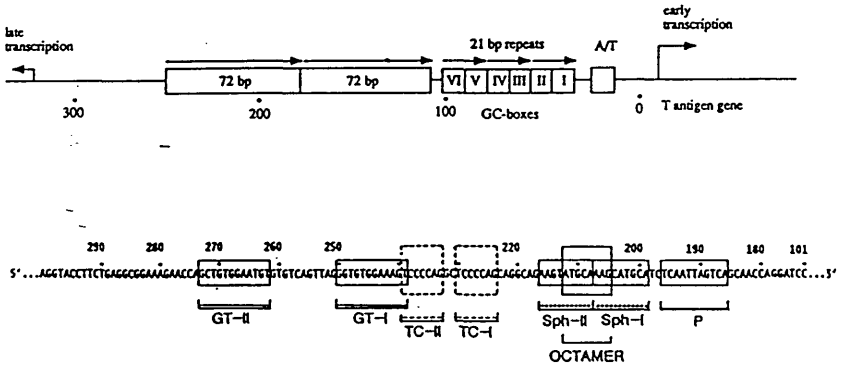


Figure 1.1 The SV40 Promoter and Enhancer.

The top figure illustrates the organisation of SV40 promoter and enhancer elements with respect to the start site of early and late

transcription. The lower figure gives an expanded view of SV40 enhancer elements. The interactions of these elements with their cognate binding factors and with each other is discussed in the text.

Next to the T antigen gene (Tag) is the DNA origin of replication (Myers and Tjian (1980) which overlaps with three Tag binding sites and the initiation site of early transcription (Khoury et al. (1973). Disruption of the replication origin prevents repression of early transcription by eliminating Tag binding sites (Rio and Tjian, (1983) and reduces the number of templates available for late mRNA synthesis reducing late gene transcription (Myers, et al. (1981); Rio, et al. (1980). The AT-rich region adjacent to the DNA origin of replication bears similarity to the TATA box motif described by Goldberg but deletion of this region has little effect on early transcription (Benoist and Chambon (1980). In contrast, deletion of the 3 adjacent GC-rich 21 bp direct repeats severely reduces early transcription and virus viability (Fromm and Berg (1982). Each 21 bp repeat contains two copies of the hexanucleotide GGGCGG which is the core recognition site for the transcription factor Spl (Dynan and Tjian (1983); Briggs et al. (1983). The affinity of Spl for each of the six sites varies depending on sequence flanking the GC core element and this combined with steric constraints determines that only five of the six sites are occupied at any one time (Gidoni et al. (1985).

The 72 bp repeats adjacent to the late transcription start site constitute the SV40 enhancer which confers enhanced transcription, independent of orientation and of position (Banerji et al. (1981). The SV40 enhancer was the first to be discovered and functions in different cell types from different species, including non-mammalian

systems (Nenhaus et al.(1984). The SV40 enhancer can activate transcription from the early or late promoter from anywhere in the viral genome and can also activate transcription from heterologous promoters such as the β-globin promoter (Banerji et al. (1981), the conalbumin and the adenovirus major late promoters (Moreau et al. (1981). The human α-globin gene seems to be resistant to SV40 enhancer activation (Humphries et al. (1982); Treisman et al.(1983).

The SV40 enhancer has been extensively analysed (Zenke et al. (1986); Herr and Clarke (1986); Clarke and Herr (1987) and has been shown to be made up of three domains A, B and C which are 21, 22 and 15 bp respectively. Each domain consists of modules, termed enhansons, such that the activity of one enhanson can be compensated for by the duplication of another, where enhansons cooperate with each other and spacing between enhansons is critical to enhancer activity (Ondek et al. (1988). Individual enhansons are found associated with other viral and cellular enhancers and promoters.

Domain A contains an imperfect direct repeat of a 9 bp element, SphI and II. Mutation of these elements down regulates SV40 enhancer activity in HeLa cells and in embryonal carcinoma cells but not in B cells (Nomiyama et al. (1987). The direct repeat of the Sph elements generates an overlapping octamer box recognition site, homologous to that found in Ig enhancers, the histone H2B promoter and the U2 snRNA promoter (Grosschedl and Baltimore (1985); Sive et al. (1986); Mattaj et al. (1985). Mutation of this octamer box reduces SV40 enhancer activity in B cells but not in HeLa or in embryonal carcinoma cell lines (Nomiyama et al. (1987). The factor binding to this element, OBP 100 may be related to the lymphoid specific octamer

binding factor. Domain A also contains the P element which binds AP-1.

Domain C contains 3 elements GT-11A, B and C which interact with four distinct factors GT-11A, GT-11Bα, GT11Bβ and GT-11C/TEF-1 (Xiao et al. (1987). Factors GT-11A, GT-11Bα and GT-11Bβ are found in HeLa, undifferentiated embryonal carcinoma cell lines and in B cell lines but mutation of their cognate binding sites has little detrimental effect on SV40 enhancer activity in these three cell lines. This suggests that these elements play only a minor role in SV40 enhancer activity. In contrast, mutation of GT-11C reduced the activity of the SV40 enhancer dramatically in HeLa and embryonal carcinoma cell lines but has no effect on activity in B cells. The cognate binding factor, TEF-1 (also called AP4), is absent or inactive in B cells. TEF-1 binds to SphI and II but with weaker affinity, explaining why GT-11C and Sph motifs are interchangeable. TEF1 binds to both Sph and GT-11C despite no obvious sequence homology between the two sites. Sph and GT-11C compete with each other for TEF-1. TEF-1 binds cooperatively to multiple copies of its cognate binding site, GT-11C or Sph (Davidson et al. (1988) and binding to Sph by TEF-1 is augmented by binding of AP1 to the P element of domain A (Mermod et al. (1988).

Integrity of the GT-1 motif of domain B is crucial to full activity of the enhancer in HeLa, and embryonal carcinoma cell lines. Mutations in the TC-11 motif has a down-regulatory effect in HeLa cells and B cells. The activity of these elements in each cell line correlates with the presence of their binding factors (Rosales et al. (1987). While GT-1 bears a high degree of homology to GT-11C (10/12 bp) it binds different factors to closely overlapping sites. GT-1

binds three factors GT-1A, B and C of which GT-1A and B are ubiquitous while GT-1C is absent from B cells. Competition studies indicate that GT-1A is probably Spl while GT-1C/ TEF-2 is the same CACC-binding factor(s) involved in globin gene regulation. Different forms of each factor may be active in different cell lines (Xiao et al. (1987). In HeLa cells TC-I and II bind AP2 (Mitchell et al. (1987). AP2 can bind also to the Spl and Tag binding sites in the early promoter and evidence suggests that Tag mediated repression of transcription may be achieved in part by protein-protein interaction between Tag and AP2, preventing AP2 from binding to TC-I and II of the enhancer.

In certain cell types GT-1 and GT-11C act cooperatively within the enhancer, in a fashion which depends not so much on their stereoalignment but on the overall distance between the elements (Fromenthal et al. (1988). Similarly, there is a distance-dependent cooperativity between SphI and SphII.

From their detailed analysis of the SV40 enhancer, its constituent enhansons and their binding factors, Pierre Chambon and co-workers have proposed a scheme for the functional organisation of enhancers (Fromenthal et al. (1988); Davidson et al. (1988). It is a scheme which extends the original proposal by Ondek et al. (1988) and which accommodates Ptashne's model of how protein activators work (Ptashne (1988). (Figure 1.2).

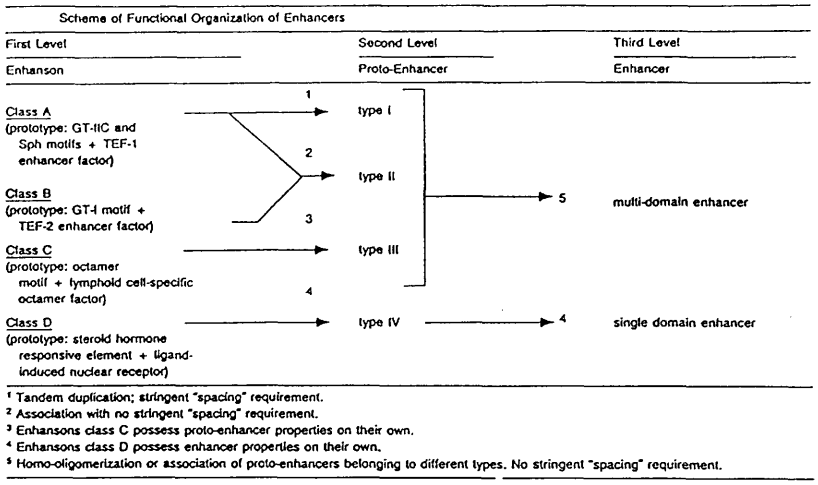


Figure 1.2 Scheme of the Functional Organisation of Enhancers.

This figure was produced by Fromenthal et al. (1988) and summarises their proposals for the organisation of enhancers based on their analysis of the SV40 enhancer and work by others on different systems. Their scheme is discussed in the text.

It is proposed that there are three levels of enhancer organisation. The first level is the level of the enhanson which is categorized as types A, B, C or D depending on the type of factor they bind. Class A factors bind cooperatively to repeats of their cognate binding sites such as TEF-1 binding to GT-11C or Sph. Class B factors do not bind cooperatively to tandem repeats of their cognate binding site and by themselves generate little enhancer activity. However, in association with Class A factors, Class B factors act synergistically to generate enhancer activity, e.g. TEF-2 and GT-1. Class C factors can act synergistically with other factors when the cognate binding sites are not closely appositioned. Examples are OBF and NFKB. Class D factors possess intrinsic enhancer activity when bound to single copies of their recognition site although they can also act

synergistically to increase enhancer activity, for example the human oestrogen receptor and the oestrogen responsive element.

The proto-enhancer which is a combination of enhansons is the second level of enhancer organisation to be proposed. There are four types of proto-enhancer. Type 1 is generated by tandem repetition of Class A enhansons; type 2 by combination of class A and class B enhansons; type 3 composed of a single copy of class C enhansons and type 4 which already has enhancer activity and consists of a class D enhanson.

Finally, the third level of organisation corresponds to combinations of different proto-enhancers.

There are two models to explain the mechanisms by which enhancer sequences work: the scanning model and the looping model (Ptashne (1986), Muller et al. (1988a). Both models attempt to explain how sequences independent of position of the promoter interact with the transcription initiation complex at the promoter. The scanning model proposes that high affinity enhancer sequences act as entry sites for transcription complexes, including RNA polymerase, which then slide along the DNA until they reach and recognise sites within the proximal promoter. The looping model proposes that enhancer sequences interact with proximal promoter sequences where transcription is to be initiated via proteins bound at both sites with the looping out of intervening DNA. Evidence confirms that binding of factors at a distance is cooperative (Cohen and Meselson (1988), Schule et al. (1988). Moreover, the looping model predicts that enhancers should be able to act in trans to stimulate transcription from genes that are not linked to the enhancer sequence. Muller et al. (1989) have shown that the SV40

enhancer will enhance transcription from the rabbit β globin gene *in vitro* when separated from it by a non-covalent bridge.

1.3.2 DNA BINDING DOMAINS AND ACTIVATOR DOMAINS OF TRANSCRIPTION FACTORS.

Protein factor interactions with DNA regulatory sequences are mediated by the DNA binding domain of the protein (Wharton and Ptashne (1985) while the activating domain mediates the protein-protein interactions which modulate initiation of transcription (Bushman and Ptashne (1988). In eukaryotic transcription factors, the DNA-binding domain and the activating domain are usually physically separated (Ptashne (1988).

The nature of DNA binding domains and activator domains of transcriptional regulators is reviewed by Mitchell and Tjian (1989).

1.3.2A DNA Binding Domains.

The Helix-Turn-Helix Domain.

Progress in understanding eukaryotic transcription control mechanisms is due partly to the extensive research that has been carried out on prokaryotic gene control systems. The characterisation of the phage λ repressor and *cro* proteins and how they interact to control the life cycle of λ has revealed features of protein-DNA and protein-protein interactions which have implications for similar interactions in eukaryotes (reviewed by Ptashne (1986). The helix-turn-helix has been identified in prokaryotic proteins including λ repressor and *cro* proteins (Pabo and Sauer (1984); Anderson *et al.* (1981); Pabo and Lewis (1982) but also in eukaryotic proteins such as

the yeast MAT α2 protein (Laughon and Scott (1984) and in the homeodomain of Drosophila segmentation genes (Shepherd et al. (1984).

The helix-turn-helix motif is a pair of protein α-helices, one helix of which lies along the major groove of the DNA to allow specific contacts to be made between the amino acids facing the major groove and the bases lying in the recognition site (Wharton and Ptashne (1985). λ repressor binds as a dimer to operator sites with two-fold symmetry with the recognition helices from each monomer fitting into adjacent major grooves in the DNA and bound to the same side of the DNA helix. The interactions between the amino acids in the recognition helix and base pairs in the major groove are highly specific and the specificity of the repressor for the operator site can be altered by making amino acid changes in the recognition helix facing the major groove (Wharton & Ptashne (1985).

The Homeo-Domain.

The homeo-box domain is a 60-amino acid DNA-binding motif consisting of a helix-turn-helix of the type described above for the λ repressor and was first identified in the developmental regulatory genes of Drosophila (Laughon and Scott (1984). The role of homeo-domain proteins in transcriptional regulation is reviewed by Levine and Hoey (1988). It has since been shown in vitro that the homeo-domain proteins of the fushi-tarazu, paired and zen genes synergistically upregulate expression from the engrailed promoter (Han et al. (1989) and that the products of the even-skipped and engrailed genes repress activity. The biochemical nature of the transcriptional activity of Drosophila homeo-domain proteins is being analysed with

particular attention to protein-protein interactions and synergy.

The B-cell specific transcription of the immunoglobulin genes is critically dependent upon the presence of an intact octamer box. This site is recognised by two factors; OTF-1, which is ubiquitous and OTF-2, which is lymphoid-specific. cDNAs encoding both factors have been obtained and sequence comparisons reveal the presence of a shared highly charged domain of 160 amino acids which is conserved not only between OTF-1 and OTF-2, but also in the unc-86 gene of C.elegans and in the Pit-1 gene (Clerc et al. (1988), Sturm et al. (1988), Muller et al. (1988b), Scheidereit (1988). This domain referred to as the POU (Pit-OTF-Unc) domain is made up of two sub-domains: the homeo-domain and the POU-specific domain. The homeo-domains of these proteins are also conserved in three Drosophila homeo-box proteins, from which they derive their name, Antennapedia, engrailed and even-skipped. The POU domain is necessary for DNA-binding and may alter the sequence specificity of this sub-group of homeo-domain proteins.

Specific mutations in helix 2 of the OTF-1 homeodomain disrupts protein-protein interactions with the viral transcription factor, VP16, in a way that does not affect DNA binding activity. These mutations are analogous to the λrepressor positive control mutations which do not affect DNA binding or negative regulation but disrupt protein-protein interactions with RNA polymerase (Stern et al. (1989).

The Zinc Finger DNA Binding Domain.

The zinc finger DNA binding domain was first identified in TFIIIA of Xenopus and is reviewed by Evans and Hollenberg (1988). TF111A has nine Zn fingers which are formed by the coordination of a Zn^{2+} ion by cytosine and histidine residues at invariant positions (Miller et al. (1985).

A large number of transcription factors have since been found to bear primary sequence which could form zinc finger motifs including the steroid hormone receptors, Sp1, GAL4, testis-determining factor (TDF) and erythroid factor 1 (NF-E1) and several of these factors such as SP1 and the steroid hormone receptors have been shown to do so.

There are two types of zinc finger protein: the C_2H_2 group of which TF111A and SP1 are members in which pairs of cysteines and histidines are separated by loops of 12 amino acids (Berg (1986); and the Cx group in which variable numbers of cysteine residues are arranged periodically and of which the steroid hormone receptors, EF-1 and GAL4 are members. Nuclease analysis of TF111A bound to the 5S RNA gene revealed that each finger bound in the major groove and contacted 5 nucleotides (Fairall et al. (1986).

Sp1 was originally identified in HeLa cells as a factor which bound to the GGGCGG motif of the SV40 promoter and stimulates transcription from a number of cellular promoters in vitro (Dynan and Tjian (1983); Gidoni et al. (1985). Sp1 exists in two active forms of molecular weight 95kD and 110kD and the 95kD form may be a truncated version of the 110kD factor which has been generated by proteolytic cleavage. (Briggs et al. (1986). Both the 95kD and 110kD forms of Sp1 bear equal amounts of this O-linked glycosylation (Jackson and Tjian

(1988). Certain members of the CTF and AP1 family of proteins have been shown to be glycosylated and this differential glycosylation of factors in the same family may alter their response to biologcal signals. Sp1 binds to its cognate DNA binding sites via a COO_2H-terminal domain consisting of 3 C_2H_2 zinc fingers (Kadonaga et al. (1987) but the mechanism by which Sp1 activates transcription and the commonality between Sp1 regulated genes is still not determined.

The glucocorticoid receptor is activated by binding of hormone to its steroid binding domain. This binding elicits a conformational change which increases the affinity of the receptor's DNA binding domain for the glucocorticoid response element (GRE) (Holenberg et al. (1987). The DNA binding domain contains two zinc finger motifs formed from periodically spaced cysteine residues and depends upon Zn^{2+} for its activity (Weinberger et al. (1982). The glucocorticoid receptor bound at its GRE can interact cooperatively via protein-protein interactions with a CACCC binding factor in the tryptophan oxygenase gene promoter (Schule et al. (1988) and also with another glucocorticoid receptor bound at a second GRE in the promoter of the human tyrosine aminotransferase gene promoter (Jantzen et al. (1987); Tsai et al. (1989). The zinc finger domain of the oestrogen receptor is similar to the glucocorticoid receptor and a two amino acid change in the first zinc finger of the glucocorticoid receptor changes its recognition of the glucocorticoid response element to that of the oestrogen response element (Danielson et al. (1989) while a three amino acid change converts the oestrogen receptor into recognising the glucocorticoid response element (Mader et al. (1989).

The zinc finger motif has clearly evolved separately from the

helix-turn-helix motif and it would be interesting to determine whether any bacterial proteins bear this motif.

1.3.2B Activator Domains.

"Domain swap" experiments and analysis of factors mutant in activating potential but still able to bind specifically to DNA have defined three types of activating domain of regulatory proteins:
(i) stretches of negatively charged amino acids (Hope and Struhl (1986) which can form amphipathic α-helices, with the negatively charged groups all on one surface of the helix and hydrophobic groups on the other (Hope et al. (1988) examples of which include GAL4 (Ma and Ptashne (1987); GCN4 (Hope (1988); AP1 (Struhl (1988); the HSV viral protein, VP16 (Sadowski et al. (1988), Triezenberg, S. et al(1988); and the glucocorticoid receptor (Hollenberg and Evans (1988),
(ii) highly basic stretches of amino acids rich in glutamine, examples of which include Sp1 (Courey & Tjian (1988); OTF-1 (Sturm et al. (1988); OTF-2 (Clerc et al. (1988); AP2 (Williams et al. (1988); SRF (Norman et al. (1988);
(iii) proline-rich domains which are functional domains in CTF and NF-1 (Mermod et al. (1989) and which may also play a role in AP1 (Struhl (1988), AP2 (Williams et al (1988), OTF-2 (Clerc et al (1988) and SRF (Norman et al (1988).

There are probably other types of activating domain which have not yet been structurally characterised, for example the A/B domain of the human oestrogen receptor which is not an acidic activating region unlike the A/B region of the glucocorticoid receptor (Tora et al. (1989).

Activating domains of different proteins differ in density of charge, the extent of secondary structure and presumably in their affinity for the target protein. Ptashne has suggested that those proteins with activating regions with highest affinity for the target protein will be the most effective in "acting at a distance" (Ptashne (1988) such that activators bearing the same DNA-binding domain but with different activating regions, will differ markedly in their abilities to activate.

Ptashne has also proposed "squelching" whereby proteins with strong activating domains could bind the target protein off the DNA preventing interaction of the weaker activating domain of the protein bound on the DNA, with the same target protein. This "squelching" phenomenon would effectively inhibit gene expression and may explain why some viral proteins which are strong activators depress expression of cellular genes (Treizenberg et al. (1988). Given the squelching effect of strong activators, each activator must be blocked when not required and this may happen by binding of the activator by an inhibitor, e.g. pGAL4 is blocked by pGAL80 in the absence of galactose (Ma & Ptashne (1987).

Consistent with Ondek et al. who demonstrated that enhansons can be arranged combinatorially to form functional enhancers, Ptashne proposes that multiple weak activators might cooperate to interact efficiently with the target protein when bound to DNA, although none taken separately would be sufficient to interact significantly with the target protein.

Since these activating domains can operate in a number of different organisms (yeast, Drosophila, tobacco plant, mammalian

cells), it is presumed that they all interact with a target protein common to all these cell types, either RNA polymerase or possibly TFIID. Both GAL4 and ATF have been shown to interact with TFIID in such a way as to modify TFIID activity (Horikoshi et al. (1988); Horikoshi et al. (1988), which may explain how GAL4 and ATF cooperate with each other to stimulate transcription by greater than 50 fold while either GAL4 or ATF alone only stimulate transcription by 2 to 4 fold (Lin et al. (1988); Lee et al. (1987). Cooperative interactions have also been demonstrated for factors binding to the CACC motif and the glucocorticoid response element in the tryptophan oxidase gene promoter (Schule (1988).

Ptashne's model predicts that the total number of activators is small compared with the number of regulated genes in a higher eukaryote such that common elements are used in different combinations to confer the many patterns of cell specific gene expression. If the interaction between activator and target was specific, this combinatorial organisation of enhansons would not operate.

Ptashne's model of how protein activators operate is consistent with the structural and functional analysis of multi-domain enhancers such as the SV40 enhancer (figure 1.2). However, the properties of the single domain enhancer, such as the human oestrogen responsive element (hERE) (Fromenthal et al. (1988) do not obviously fit into Ptashne's model. However recent work with the hERE has shown that the human oestrogen receptor (hER) is a multi-domain protein and that in addition to the DNA binding domain, there are two activator domains, the hormone binding domain (TAF-2) and the A/B domain (TAF-1) (Tora et al. (1989). TAF-1, which is cell-type specific, and TAF-2 are

independent activator domains which can act synergistically with each other and with other activators. Thus the hERE is functionally equivalent to the multi-domain enhancer and L.Tora et al. suggest that modular transcription factors would be particularly useful in the combinatorial activation of transcription during development when the supply of transcription factors is limited.

E1a, which is encoded by the Adenovirus genome, is involved in transactivation of a number of different viral and cellular genes as well as playing a role in transformation. E1a does not bind DNA specifically and stimulates genes which do not appear to share a specific promoter element (Berk (1986). The E1a protein region necessary for transcriptional activation has been identified (Green et al. (1988). Designated Region 3, it consists of two domains one of which is necessary for activation and the other which mediates interaction with the promoter (Lillie and Green (1989). The activator domain of E1a can be replaced by the activator domain of VP16 and the promoter binding domain can be replaced by the DNA binding domain of GAL4 to activate only genes containing GAL4 recognition sites. However, since E1a does not bind a specific DNA sequence, it is proposed that E1a is targeted to the promoter by binding of its promoter binding domain to the protein binding sites of DNA bound proteins, for example, ATF. Given the promiscuity of E1a transactivation, it is likely that it recognises many target proteins.

The DNA binding domain and activator domain of GAL4 are separable (Keegan et al. (1986), and fusion proteins between the GAL4 activator domain and the LexA DNA binding domain renders yeast genes with upstream LexA operator sites inducible by galactose (Brent and

Ptashne (1985).

The observation that GAL4 will activate transcription of mammalian genes with a UAS_G (Upstream Activator Sequence of GAL4) and a TATA box in mammalian transcription systems (Lin et al. (1988) and will function synergistically with both the glucocorticoid receptor (Kakidani and Ptashne (1988) and the SV40 enhancer (Webster et al. (1988) illustrates that the mechanisms of transcription activation have been conserved from yeast to man.

The transcription factor TF11D is conserved from yeast to man and may be one of the common "target" proteins bound by the activator domains of both yeast and mammalian transcription factors (Buratowski et al. (1988); Cavallini et al. (1988). In fact, work by Horikoshi et al. (1988) shows that GAL4 mediated activation of Adenovirus E4 gene involves an alteration in TF11D interactions with the TATA box and that these altered TF11D-TATA box interactions are not observed when the activator region of GAL4 is deleted. The yeast TFIID gene has been cloned (Horikoshi et al. (1989) and sequence analysis reveals that it bears homology at the amino acid level to the bacterial sigma factor which determines the specificity of RNA polymerase initiation in prokaryotes and dissociates from the polymerase shortly after initiation . The other factors which associate with the eukaryotic transcription initiation complex (TFIIA, TFIIB and TFIIE) may be functionally homologous to other subunits of the bacterial RNA polymerase (Buratowski et al. (1989).

1.3.3 Modulation of Transcription Factors by Phosphorylation.

Phosphorylation of serine or threonine residues in the activating domain of a transcription factor may affect its activity by

altering protein conformation or by altering hydrogen bonding interactions with RNA polymerase or other trans-activators, and may provide a mechanism by which activators can respond to external stimuli (Hoeffler et al. (1988). Glycosylation of serine or threonine residues of proteins may affect activity by preventing phosphorylation (Jackson et al. (1988). Induction of heat shock genes is not blocked by protein synthesis inhibitors (Zimarino and Wu (1987) and work done to determine how heat shock transcription factors (HSTF) are activated upon heat shock from their pre-existant inactive form has implicated HSTF phosphorylation. In human and Drosophila cells, activation increases DNA binding affinity (Sorger et al. (1987) but in yeast, activation by phosphorylation does not affect DNA binding and presumably works by potentiating the activator domain of the protein to stimulate transcription (Sorger and Pelham (1988).

Phosphorylation is also reported to be responsible for the cAMP mediated hormone induction of the rat somatostatin gene (Yamamoto et al. (1988). Genes respond to cAMP via a 43 kD nuclear phosphoprotein, the CREB factor (also known as ATF) which binds to the cAMP Response element (CRE) (Montminy et al. (1986); Montminy et al. (1987). CREB binds to CRE as both monomer and dimer but transcriptional activity is only associated with dimer binding (Yamamoto et al. (1988). Yamamoto et al. also showed that treatment of CREB with protein kinase C (but not protein kinase A) resulted in increased dimer formation and that phosphatase treatment of CREB reduced its transcriptional activity. Their proposal is that cAMP activated phosphorylation shifts the equilibrium between inactive monomers and active dimers towards dimer formation, causing

transcriptional activation.

The transcription of a number of mammalian genes including proto-oncogene c-fos and the cytoskeletal actins is induced by stimulation with growth factors and whole serum (Kelly et al. (1983). This serum induced response is mediated by the serum response factor which is present but not active in most uninduced cells (Treisman (1987); Prywes and Roeder (1986) and which binds to the serum response element found upstream of serum responsive genes (Treisman (1986). The mechanism by which SRF is activated is not clear but it is probably activated by modification of the activator region, not by altering DNA binding activity. Phosphorylation probably mediates serum responsive induction since agents such as TPA and the calcium ionophore, A23187, which induce the same genes, with the same kinetics as serum and act through the SRE also activate protein kinase C (Greenberg and Ziff (1984); Nishizuka (1986).

1.3.4 The Leucine Zipper Motif.

The leucine zipper is a structural motif which was first described for the CCAAT/Enhancer binding protein (Landschulz et al. (1988) but has since been identified in other protein sequences including Myc (Landschulz et al. (1988); ATF (Hai et al (1989); Fos and Jun (Kouzarides and Ziff (1988).

McKnight et al. originally proposed that the leucine zipper facilitates the dimerisation of proteins which contain periodic repeats of leucine amino acid residues. Specifically, the monomer contains an α-helix with a leucine repeated at every seventh position. It is proposed that the leucine side chains extending from each α-helix

interdigitate with those from another such α-helix.

The leucine zipper mediates the formation of Fos-Jun heterodimers (Ransome et al. (1989); Turner and Tjian (1989); Geinzer et al. (1989); Schuermann et al. (1989). Fos, Jun and GCN4 proteins bind to the AP-1 recognition site/TPA-response element via a basic region which is contributed to by both subunits of the dimer (Kouzarides and Ziff (1989). Domain swap experiments in which the leucine zippers of Fos, Jun and GCN 4 are interchanged indicate that the leucine zipper is critical in determining the DNA binding activity and that the DNA binding domains of these proteins are interchangeable. Fos-Jun heterodimers bind strongly to the AP1 recognition site, Fos-Fos homodimers fail to bind at all while Jun-Jun homodimers bind only weakly (Halazonetis et al. (1988). Scatchard plot analysis indicates that the DNA binding activities of different homo- or heterodimers is determined by the thermodynamic stability of that dimer which is in turn determined by the differing capacities of their leucine zippers to interact (Smeal (1989).

DNA binding by C/EBP also involves a bipartite domain consisting of basic regions juxtaposed to the leucine zipper of each of the two C/EBP polypeptides which make up the functional activator (Landschulz, W.H. et al. (1989).

Figure 1.3 illustrates examples of regulatory domains of transcription factors which are conserved in a number of different factors as discussed above.

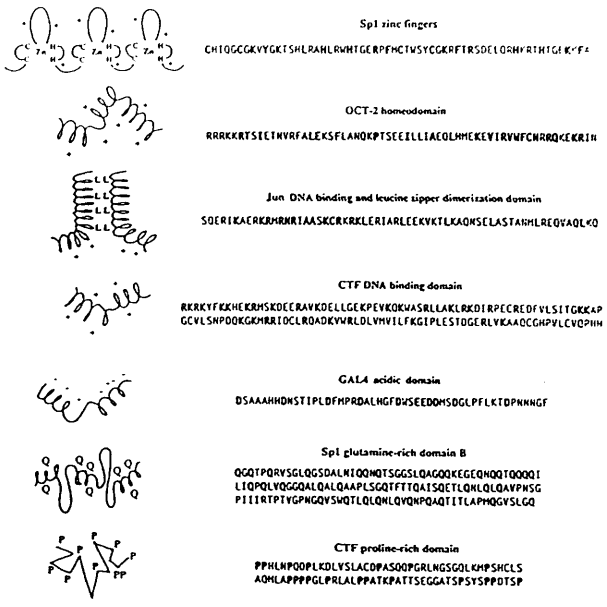


Figure 1.3 Regulatory Domains of Transcription Factors.

This figure illustrates the various structural motifs which have been identified in different transcription factors. These motifs are involved in DNA binding, activator function or in dimerisation of the factors as discussed in the text.

1.3.5 Nuclear Factor 1 and Members of the CCAAT Box Binding Family.

NF-1 was originally identified as a factor necessary for Adenovirus replication (Nagata et al. (1982) which bound to a palindromic sequence TGGCNNNNNGCCAA in the adenovirus replication origin (Rawlins (1984); Hay (1985). NF-1 has since been shown to bind to both viral and cellular gene sequences necessary for the regulation of transcription. In particular NF-1 has been shown to bind to sites within the promoters of several globin genes including the chicken β^H and β^A globin genes (Plumb et al. (1986), Emerson et al. (1985) and at -200 bp in the human β-globin gene (Jones et al. (1987). NF-1 has been affinity purified and forms a family of polypeptides of molecular

weight 60-66kD (Rosenfeld et al. (1986). Jones et al. (1985, 1987) presented evidence that NF1 was identical to CTF which recognises the CCAAT box of the HSV tk promoter. Antibodies to CTF cross-reacted with NF-1. However, more recent work which has attempted to determine the extent of identity between CCAAT box binding factors, has demonstrated that CTF is a family of factors (including NF-1) which all differ in their affinities for different recognition sequences (Chodosh et al (1988). Comparison of the protein sequence of CP1 and NF-1 reveals that while there are stretches of shared protein sequence (this may explain the cross-reactivity of antibodies), there are also considerable differences (Meisternst et al. (1988). Chodosh et al. also identified CP2 which binds to the rat γ-fibrinogen promoter CCAAT box and is distinct from either CP1 or NF-1.

CCAAT/enhancer binding protein (C/EBP) is a heat-stable DNA-binding protein first isolated from rat liver (Graves et al. (1986) and involved in the regulation of a number of genes including the serum albumin gene (Friedman et al. (1989). The cDNA for C/EBP has been obtained (Landschulz et al. (1988) and analysis of the protein reveals the presence of a leucine zipper and dimerisation to generate the functional transcriptional activator as described previously. The gene encoding C/EBP has been localised to chromosome 7 and high levels of C/EBP mRNA have been detected in fully differentiated tissues that metabolise lipids and cholesterol (Birkenmeier et al. (1989). Evidence suggests that C/EBP expression is induced in response to changing physiological conditions during development.

NF-Y recognizes the Y-motif of Major HistoCompatiblity (MHC) Class II genes which contains an inverted CCAAT motif, mutation of

which abolishes NF-Y binding (Dorn et al. (1987). NF-Y has been shown to be present in most cells, is sensitive to Zn^{2+} chelators and has a molecular weight of about 250-300 kD (Hooft van Huijsduijnen et al. (1987). Another factor, NF-Y* also binds to the Y motif but does not cross-compete with NF-Y suggesting that it is a distinct protein.

ACF is a liver specific factor which recognizes and binds to the CCAAT box of the rat albumin gene (Raymondjean et al. (1988) and may be the same factor identified by Lichtsteiner et al. (1987) as binding to half an NF-1 site in the albumin promoter.

What is the function of all these different CCAAT binding factors and to what extent are they conserved? Santoro et al. (1988) presented evidence that a family of CCAAT binding factors are all encoded by a single gene in the human genome and that diversity is generated by differential RNA splicing. Chodosh et al. demonstrated that CP-1, CP-2 and NF-1 are all heteromeric and that the DNA binding domain of each protein was separate from the activator domain. Moreover, they demonstrated that the HAP2 and HAP3 proteins of yeast which activate CYC1 expression from UAS2 could substitute for the CP1B activator subunit and CP1A DNA binding subunit respectively, illustrating that the CCAAT binding factor is functionally conserved in evolution. The multimeric organisation of the CCAAT binding factor could conceivably allow it to respond to a variety of environmental signals with different activator domains operating in response to different signals.

1.4. GLOBIN GENE EXPRESSION.

Mammalian globin proteins are encoded by a family of genes, all derived from a common ancestral gene which has duplicated and diverged over a period of 500 M years (Romero-Herrera et al. (1973). Different globin genes as already discussed are expressed at different stages in mammalian development (embryonic, foetal and adult) but all are expressed exclusively in erythroid tissue. The α and β globin genes are organised into two distinct clusters, which map to different chromosomes (Deisseroth et al. (1977, 1978). α and β globin genes are highly conserved in their coding regions but diverge significantly in flanking and intervening sequence. Both α and β globin genes are interrupted by two introns at similar positions but intron 2 of the β globin gene is considerably larger than that of the α globin gene. (Review by Maniatis et al. (1980). Figure 1.4 summarises the organisation of globin genes in the three species most extensively studied.

There are three related aspects of globin gene expression to be analysed; how different globin genes come to be expressed at different stages of development (Reviewed by Karlsson and Nienhuis (1985), how globin genes are exclusively expressed in erythroid tissue and how globin gene expression is regulated during erythropoiesis.

The regulation of the human α and β globin genes differs in a number of aspects. In humans, the adult α globin gene is transcribed earlier in development than the adult β globin gene (Peschle et al. (1984, 1985). Another observation which illustrates differential α- and β-regulation is the enhancer dependence of the β-globin gene in transient transfection assays (Banerji et al. (1981); Humphries et al.

(1982). This enhancer dependence can be compensated for by trans-acting viral products (Green et al. (1983); Treisman et al. (1983). Following stable integration into the MEL cell genome the human α-globin gene is expressed constitutively while β is expressed only at low and inducible levels (Chao et al. (1983); Charnay et al. (1984); Wright et al. (1983). However, despite these differences in regulation the α and β globin chains come to be expressed in vivo in equimolar amounts in the terminally differentiated cell.

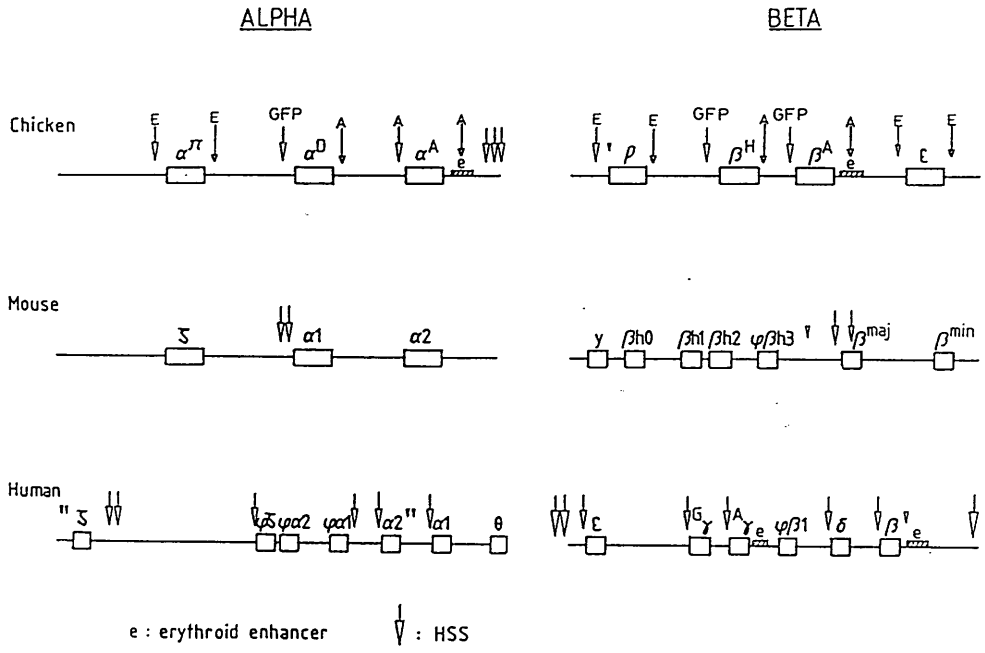


Figure 1.4 The Organisation of Globin Genes.

The figure illustrates the spatial organisation of the globin genes in the α- and β- globin gene loci of chicken, mouse and human. Individual globin genes are represented by open boxes. Arrows indicate the presence of DHSS which have been mapped in vivo. Arrows are superscripted by GFP to indicate the presence of a genomic footprint at that DHSS; "E" to indicate that the DHSS is only present in cells at the embryonic stage of development and by "A" to indicate that the DHSS

is only present in cells at the adult stage of development. Hatched boxes superscribed by "e", indicate the presence of an erythroid specific enhancer.

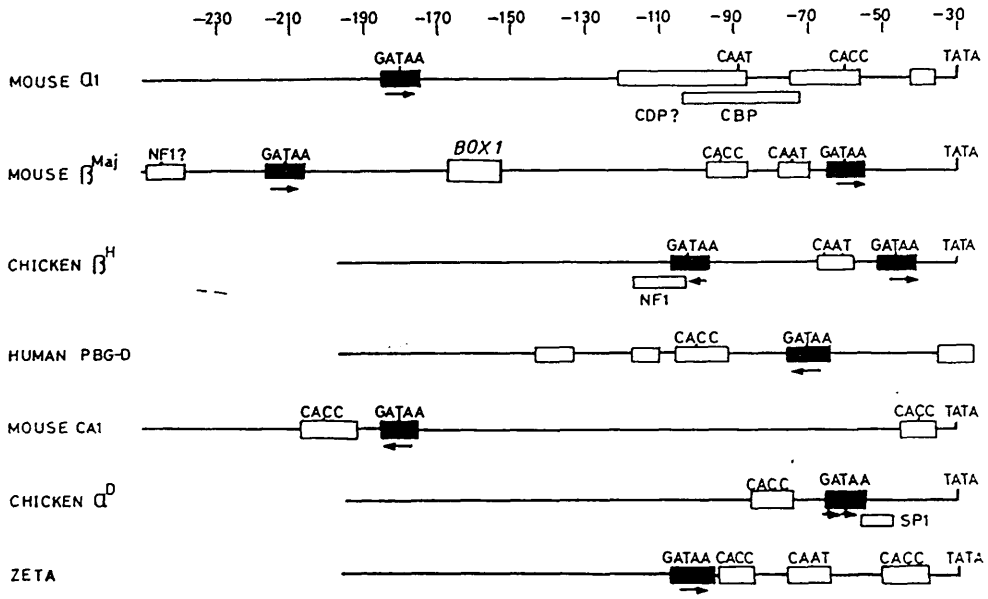


Figure 1.5 Modularity of Globin Gene Promoters.

The figure illustrates the organisation of regulatory elements within the promoters of six different globin genes and within the erythroid promoter of the human porphobilinogen deaminase gene. There are four types of element; the TATA box, CCAAT-like sequences, the CACC box and the NF-E1 recognition element containing the consensus sequence GATAA. The position of the TATA box is conserved in all of the globin gene promoters but the organisation of other elements differs between promoters. It is argued here that these differences in organisation will be reflected in the conformation of the transcription complex and in the efficiency with which the promoter interacts with enhancer or other regulatory sequences.

1.4.1 Globin Gene Promoter Elements.

As previously discussed in 1.3.1, analysis of different promoters and enhancers, including globin gene regulatory elements, has revealed their modular nature.

The demonstration by Wright et al. (1984) that sequences both 5′ and 3′ of the human adult β-globin gene are necessary for regulated expression of the gene was followed by intensive deletion and mutagenic analysis of flanking sequence in the search for regulatory elements. In the rabbit β-globin 5′ region, three elements were shown to be essential for expression in transient assay; the ATA box at -30 bp, the CCAAT box at -75 bp and the CACC box at -85 bp and -100 bp (Dierks et al. (1983). These elements were subsequently shown to be conserved in the 5′ promoter region of all the globin genes studied and their importance to regulation confirmed by mutagenic analysis of each element in the context of each promoter (Myers et al. (1986); Cowie et al (1988). Figure 1.5 summarises the organisation of promoter modules in different globin genes.

A nuclear factor αCP1 which binds to the CCAAT box of the mouse α1-globin gene had been previously partially purified (Cohen et al. (1986) and did not show erythroid specific tissue distribution. Subsequent work (Barnhart et al. (1988) identified a second factor αCP2 binding to a second site overlapping the CCAAT box. The binding activity of αCP2 is activated 2 to 3 fold when MEL cells are induced to differentiate whereas the binding of αCP1 is reduced 3- to 5-fold. Both αCP1 and αCP2 are distinct from CTF/NF1. A third binding factor, α-IRP which interacts with a site, adjacent to that of αCP1, contains the sequence motif, CACCC (Barnhart et al. (1988). Binding of αCP1 to

its cognate binding site prevents binding of αCP2 and αIRP. Barnhart et al. (1988) propose that αCP1 acts as a repressor of α1-globin gene expression during MEL cell differentiation and may be analagous to the CCAAT Displacement Protein (CDP) identified in sea urchin embryo (Barberis et al. (1987).

Cohen et al. (1986) showed that αCP1 binds weakly to the mouse β^{major} CAAT box. My own work indicates that the β^{major} CCAAT box is bound by two factors, one of which is αCP1.

Interestingly, while Myers et al. (1986) showed that mutation at -79 bp and -78 bp of the mouse β^{major} CCAAT box from G-->A, caused up-regulation of expression in HeLa cells, Cowie et al. (1988) have shown that this same mutation has no effect on expression in MEL cells while mutation at -69 bp and -65 bp reduces expression in MEL cells but not in HeLa cells. This work suggests that the β^{major} CCAAT box is bound by different factors in the different cell types.

Cowie et al. compared CCAAT box sequences from α, β and γ sequences from several different species.

α	AG CCAAT GA
β	GG CCAAT CT
γ	GA CCAAT AG

It was suggested the differences in the 5′ and 3′ flanking sequences played a role in determining the differences in timing of expression of the genes in development.

The human $^{A}\gamma$-globin promoter contains two adjacent CCAAT box elements -85 bp and -100 bp which interact with two different factors (Superti-Furga et al. (1988). The proximal element has greater affinity for the ubiquitous CP1 factor which interacts with the human

α-globin gene (Chodosh et al. (1988).

The distal CCAAT box binds an erythroid specific factor, NF-E1, which binds simultaneously with CP1 binding to the proximal site (Superti-Furga et al (1988). Flanking sequence is critical in determining NF-E1 binding. NF-E1 also binds to the CCAAT box of the human β-globin gene. A third factor which has similar properties to the sea urchin CDP also recognizes the proximal CCAAT box and may compete with CP1 for this site, but its role in γ-globin gene regulation is not understood.

Identifying the molecular lesions responsible for human thalassaemias and the mapping of deletion and point mutations led to an early understanding of how human globin genes are developmentally regulated. A G-->A transition at -117 bp of the $^{A}\gamma$-globin gene results in continued expression of this foetal globin gene in the adult organism, a condition known as Hereditary Persistance of Foetal Haemoglobin (HPFH). This mutation at -117 bp alters the binding activity of all three factors CP1, CDP and NF-E1 for their cognate binding sites; the binding activity of CP1 and CDP increases while that of NF-E1 decreases. Superti-Furga et al. suggest that NF-E1 may specifically repress γ-globin gene expression in adult erythroid tissue but stimulate expression of the adult β-globin gene. A variant form of HPFH in which there is a deletion of 13 bp in this region which abolishes the binding of all three factors, induces the same phenotype as the point mutant (Mantovani et al. (1989).

A different form of HPFH involves a T-->C transition at -175bp of the $^{G}\gamma$-globin gene. This change abolishes binding of OTF-1 to the octabox of the $^{G}\gamma$-globin gene promoter and simultaneously

increases binding of the erythroid factor, NF-E1 (Mantovani (1989). Abolishing the NF-E1 site abolishes the mutant overexpression phenotype while inactivation of the OTF-1 site has no effect on overexpression, suggesting that it is the mutation of the NF-E1 site which is responsible for overexpression (Nicolis et al. (1989); Martin et al. (1989).

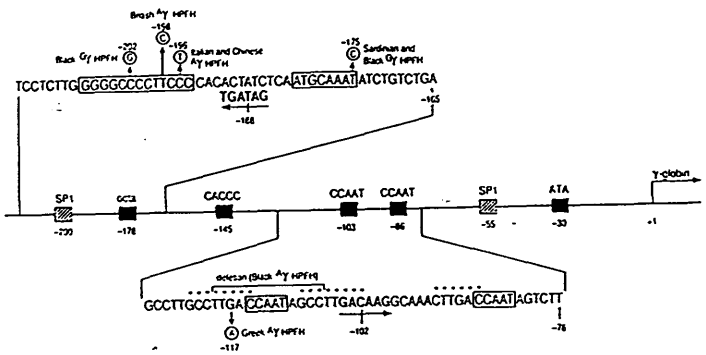


Figure 1.6 Lesions in the Human γ-Globin Gene Promoters which Leads to HPFH.

The figure is taken from "Highlights of Modern Biochemistry:Proceedings of the 14th International Congress of Biochemistry", and a paper published therein by Superti-Furga and Busslinger. It shows the

position of recognition sites for DNA-binding proteins and the position of HPFH mutations which map within or close to these sites.

Analysis of the 5′ promoter of the chicken β^A gene (Emerson et al. (1985); Plumb et al. (1986) has identified four protein binding elements: a G-rich sequence, two overlapping CACC binding sites, a CAAT box and an inverted repeat sequence which bind factors BGP1, CON, CAT and PAL respectively. In in vitro transcription systems, PAL can be substituted for by purified NF-1 (Emerson et al. (1989) and PAL binding activity increases during chick development (Jackson et al (1989). PAL has an inhibitory effect on β^A globin gene transcription and it has been shown that binding of PAL and CON are anti-cooperative. As development progresses and the PAL site becomes occupied, binding of CON to the CACC site is lost. BGP1 is erythroid specific and its activity is correlated with nucleosome displacement from the region of the chicken β^A globin gene promoter (Lewis et al. (1988).

There has been extensive analysis of the important regulatory sequences of the human β adult promoter (Antoniou et al. (1988). The promoter contains a TATA box (-30), a CCAAT (-70), two CACC boxes (-85,-100) and an NF-E1 site (-120) which are necessary for correct expression. The distal region of the promoter contains four elements: a sequence AAGCCAGTG (-155) called Box 1 which is conserved at a similar position in the promoter of the mouse β^{major} globin gene, two copies of NF-E1 (-120,-190) and nuclear factor 1 (-200). The presence of Box 1 and NF-E1 (-120) is necessary for inducible expression (de Boer et al. (1988).

1.4.2 Globin Gene Enhancer Sequences.

As well as sequences found 5′ of globin genes, 3′ sequences have been shown to be critical. These include sequences in the second intron of the mouse β^{major} globin gene (Gabon and Housman (1988), sequences 3′ of the chicken adult α- (Knezetic and Felsenfeld (1989) and β-globin genes (Choi et al. (1986); Hesse et al. (1986), and sequences 3′ of the human γ- and β-globin genes (Bodine and Ley, (1987); Kollias et al. (1987); Wall et al (1988); Behringer et al. (1987).

The chicken β globin gene enhancer lying 307 bp 3′ of the adult gene is required for correct tissue-specific expression but also for correct developmental expression of both the adult β and chicken embryonic ε globin genes (Choi and Engel (1986); Nickol and Felsenfeld (1988). The enhancer upregulates expression from a linked reporter gene in transient transfection experiments into chicken cells by 60-fold (Reitman and Felsenfeld (1988). Functional studies and analysis of the protein factors interacting with the 3′ enhancer reveals that it consists of five enhansons: the TGGCA protein (an NF-1 like factor) binds to Region I (Rupp and Seppel (1987); AP-1 and AP-2 bind to Region II; a CACC binding factor binds to Region III and finally erythroid specific factor, NF-E1, binds to Region IV (Emerson et al. (1987). To date it is the only enhancer detected in the β-globin locus. It is proposed that in definitive erythrocytes expressing adult β globins, the enhancer interacts with positively acting element(s) present between -20 bp and -112 bp in the promoter of the adult β globin gene, called the stage-selector element (SSE).

The SSE of the chicken β^{A} globin gene may be competing with a

similar element in the promoter of the ε globin gene for interaction with the enhancer such that in the adult, the β SSE is dominant over the ε SSE, but in the embryo the interaction is reversed. A trans-acting factor, termed NF-E4, which recognises the sequence AAGAGGA (from -55 to -61 bp) in the chicken β^A globin gene promoter has been shown to be active exclusively in adult erythroid tissue. This suggests that it may be involved in developmental regulation of chicken β globin genes and its recognition site may be the SSE. It remains to be shown that this factor bound at its site in the promoter of the β^A globin gene affects interactions of that promoter or the ε globin gene promoter with the chicken β globin gene enhancer. It has not been reported whether there is any overlap in function between NF-E4 and PAL, which has also been shown to have developmentally regulated activity (Jackson et al (1989); Gallarda et al. (1989).

A tissue-specific enhancer directing correct developmental expression is also present in the 3′ region of the human β globin gene (Kollias et al. (1987) and like the chicken enhancer it is necessary for proper expression of not just the adult genes, but is also important for expression of the foetal γ genes (Trudel and Costantini (1987). Unlike the chicken β-globin gene locus, a second tissue-specific enhancer in the β cluster 3′ of the $^A\gamma$ gene has been identified but the role of this enhancer in developmental regulation is not clear.

Human foetal γ-globin genes when introduced into transgenic mice are expressed as embryonic genes while human β globin is expressed only in the adult mouse. A hybrid γβ globin gene containing γ 5′ sequences and β 3′ sequences is expressed in both embryonic and adult

tissue of the transgenic mice (Kollias et al. (1986). This work points out that although the developmental regulatory sequences and factors have been conserved from mouse to man, the actual timing of the developmental switch has changed, presumably as a result of altered timing of the activity of a regulatory factor. This work also indicates that the adult β 3′ enhancer is sufficient in this context for activation of transcription from the γ promoter and that the γ 3′ enhancer is not essential.

Fusion of uninduced MEL cells with human K562 cells resulted in activation of the mouse embryonic β globin genes and activation of the human adult β globin genes (Baron and Maniatis (1986) illustrating that the mouse embryonic globin genes are not irreversibly repressed in adult tissue and can be activated by human trans activating factors. The switch from human foetal to adult globin gene expression in these hybrids depends upon retention of human chromosome 11 (Papayannopoulou, et al. (1986) suggesting that the trans-acting factors involved in the γ - β switch are linked to the β-globin cluster on chr.11 (Melis et al. (1987). Behringer et al. (1987) have demonstrated that in a similar fashion to the chicken β globin gene enhancer, the human β 3′ enhancer interacts functionally with 5′ sequences in transgenic mice. Grosveld suggests that the expression pattern is determined by the relative affinities of stage-selector elements in the promoter for limiting transcription factors in chicken and in human and that the concentrations of these transcription factors vary through development. Grosveld, F. (pers. comm.) has shown that in transgenic mice, deletion of the adult β globin gene promoter results in up-regulation of tissue specific expression of the γ-globin genes in the adult mouse and

conversely deletion of the γ-globin gene promoter results in tissue specific up-regulation of the adult β gene.

1.4.3 Erythroid Factor 1 (NF-E1).

The recognition site for erythroid factor 1 is present in the promoters and enhancers of human, mouse and chicken globin genes and also in the erythroid promoters of the human and mouse porphobilinogen deaminase gene (Plumb et al. (1989); deBoer et al. (1988); Mignotte et al. (1989); Evans et al. (1988); Perkins et al (1989); Mantovani et al. (1988); Porscher et al. (unpublished results). NF-E1 binding sites are also present within the human β globin gene dominant control regions (F.Grosveld et al. (pers.comm.). This work suggests that NF-E1 has a role not only in modulating the transcription initiation complex of erythroid specific genes but also in activating chromatin domains in an erythroid specific fashion.

The NF-E1 cDNA has been obtained for mouse, human (Tsai et al. (1989b) and chicken (Evans and Felsenfeld (1989) and Southern blot analysis indicates that the NF-E1 gene is present in a single copy in the genome of mouse. The factor has been purified as 20 kD, 38 kD, 50 kD and 115 kD polypeptides for mouse, chicken and human. It is suggested that the 20 and 38 kD peptides, which have retained DNA binding activity, are degradation products of the 50 kD protein and that the 115 kD species is a functional dimer of the 50 kD monomer (Tsai et al (1989b). NF-E1 is a member of the C_x class of zinc finger DNA binding factors. There are two putative zinc finger motifs in the mouse, human and chicken proteins and while there is a high degree of homology between all three proteins within the zinc finger domains

(88-90%), the amino acid sequence of the chicken protein has diverged considerably from the mouse and human proteins at its N-terminal and C-terminal ends. The zinc fingers of each protein are themselves well conserved suggesting that they are duplications of each other (Tsai et al (1989b), Evans & Felsenfeld (1989), Trainor et al (1990).

Preliminary studies of the expression of NF-E1 show that the mRNA is present only in erythroid tissue as expected from previous work on the tissue distribution of NF-E1 protein DNA binding activity. NF-E1 mRNA and DNA binding protein are detected in erythroid cells before globin genes come to be expressed and given the presence of NF-E1 binding sites in the dominant control regions (Fraser pers.comm.), it is likely that NF-E1 first comes to be expressed in one of the earlier erythroid progenitor cells possibly in response to erythropoietin stimulation.

1.5 OUTLINE OF EXPERIMENTAL WORK.

My work began at a time when it had just been shown by saturation and linker scanning mutagenesis that intact TATA (-30 bp), CCAAT (-75 bp) and CACC (-90 bp) boxes were an absolute requirement for correct initiation from the mouse β^{major} globin gene promoter (Charnay et al. (1985), Myers et al. (1986). The function of these mutant promoters had been tested under the control of the SV40 enhancer in HeLa cells, a non-erythroid cell line.

Both the TATA box (-30 bp) and the CCAAT box (-75 bp) are conserved in the sequence and position in many RNA polymerase II transcribed genes. The CACC box (-90 bp) was originally identified in the rabbit β globin gene (Dierks et al. (1983) and is conserved in most globin gene promoters but has since been identified in the promoters of non-erythroid genes. As such it was not obvious how these particular elements could confer erythroid and developmental regulation upon individual globin genes.

I set out to characterise promoter elements necessary for correct tissue-specific and developmental regulation of the mouse β^{major} globin gene. The approach I chose was based on work by Sheffery et al.; they identified a DHSS in MEL cells in the mouse β^{major} globin gene promoter stretching from -250 to -50 bp 5′ of the start site of initiation, so I used DNAase 1 footprinting and gel shift assays to identify any protein binding sites in this promoter sequence (Chapter 3). DNAase 1 footprinting studies of the chicken β^{A} and β^{H} globin genes had revealed the presence of protein binding sites in the promoters of these genes which were important in their transcriptional regulation (Emerson et al. (1985), Plumb et al. (1986).

I expected to observe footprints over the TATA, CCAAT and CACC boxes and hoped to observe additional footprints which may be erythroid specific. I was successful in identifying several new protein binding sites (Chapter 3) and characterised the binding activity to these sites by gel retardation assay (Chapter 4). The function of these elements was then examined in transient (chapter 6) and stable (chapter 7) transfection assays in murine erythroleukaemia cells and in non-erythroid cell lines. The conclusions drawn from my results are summarised and discussed in chapter 8.

Figure 2.1 Recombinant Plasmids.

Figure 2.1 illustrates the recombinant plasmids used in this work.

pHSV-βgal. used as internal control to standardise for transfection efficiency, is derived from the lac Z - containing plasmid pCH110 (Hall et al (1983). Transcription of the lac Z gene is driven by the HSV-2 IE-5 promoter (IE-5), obtained as a 210 bp BamH1 fragment, via cloning into the BamH1 site of the shuttle vector pIC20H and inserted as a 310 bp HindIII fragment into the HindIII site created by removal of the mouse mammary tumour virus promoter used to drive transcription in pCH110.

Homer 6 used to confer neomycin resistance upon cells in the generation of stable transfectants. This plasmid was originally derived as described by Spandidos & Wilkie (1984).

pαCAT1 contains a 700 bp NcoI fragment from the mouse α1-globin gene blunt-end ligated into SmaI site of a pUC12 derived plasmid containing a 900 bp XbaI/BamHI CAT reporter gene fragment and HSV-2 IE-5 transcription termination signals.

pIC20R used in sub-cloning and as carrier plasmid in transfection experiments. Generated as described by Marsh et al (1984).

p0GH used as a negative control plasmid for hGH expression in transient and stable transfection assays.

pTKGH used as a positive control for hGH expression in transient and stable transfection assays.

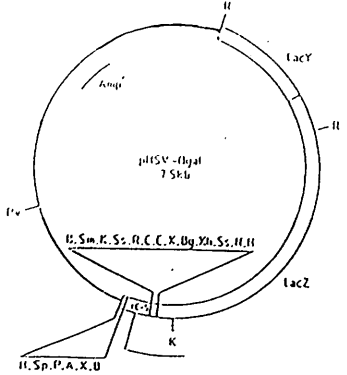
Both p0GH and pTKGH were obtained from Biogenesis Ltd. and were originally described by Selden et al (1986).

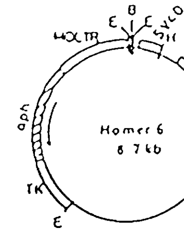
pβ346GH used as a basic test construct and from which all mutant promoter constructs are generated. A 366 bp HindIII/BamHI fragment spanning from -346 to +20 bp of the mouse β^{major} globin gene was cloned into the HindIII/BamHI sites of p0GH.

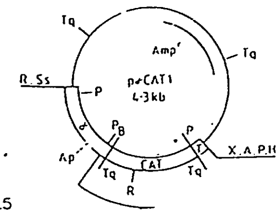
pβ(M)GH was derived from pβ346GH by the blunt-end ligation of a 415 bp PCR generated fragment spanning from +1693 to +2108 3'of the transcription initiation site of the mouse β^{major} globin gene into the Eco RI site of pβ346GH.

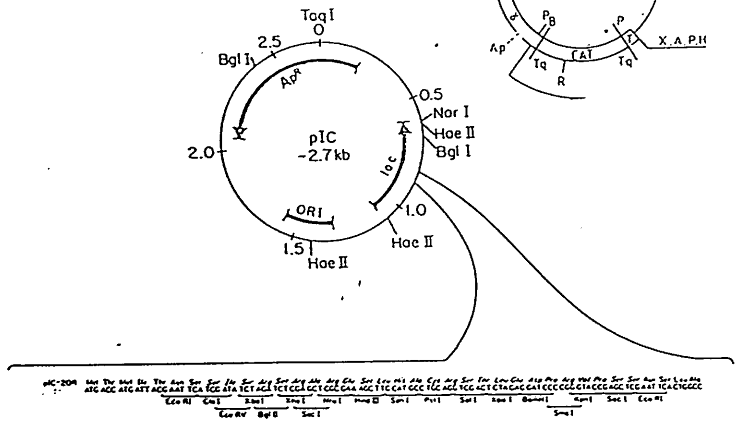
pβ(A)GH was derived from pβ346GH by ligation of a 600 bp Eco RI fragment containing the chicken β globin gene enhancer spanning from +1616 to +2216 3' of the transcription initiation site of the chicken β^{A} globin gene into the Eco RI site of pβ346GH.

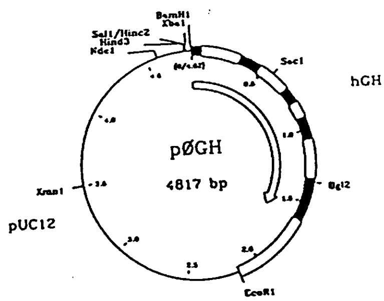
pβ7.2 was generated by ligation of a 7.2 kb Eco RI fragment from -1415 to +5785 bp relative to the transcription initiation site of the mouse β^{major} globin gene into the Eco RI site of pUC18.

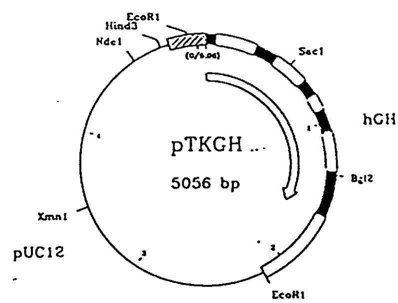

lacY
lacZ
K
Pv

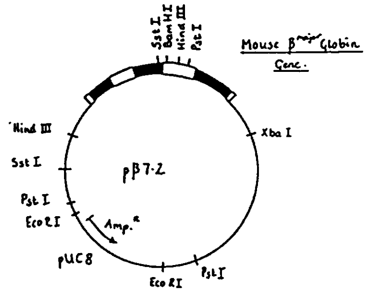

Homer 6
6 7 kb
aph
TK
E

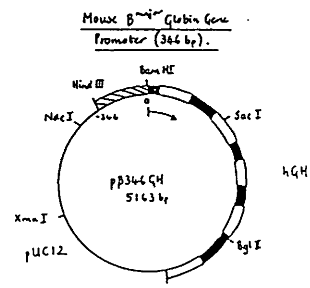

Amp'
Tq
R.Ss
pcCAT1
4.3kb
P
CAT
R
X.A.P.H

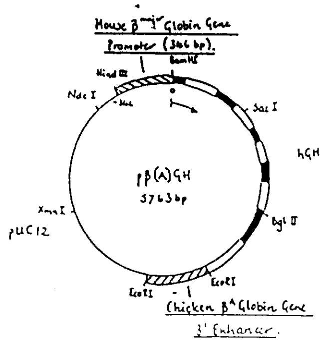

TaqI
2.5
Bgl I
ApR
pIC
~2.7kb
2.0
0.5
Nor I
Hae II
Bgl I
lac
ORI
1.0
1.5
Hae II
Hae II

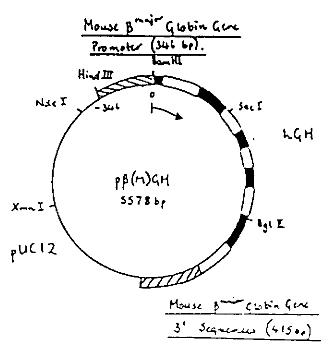

BamH1
Xba1
Sal1/Hinc2
Hind3
Nde1
Sac1
hGH
pØGH
4817 bp
Xmn1
Bgl2
pUC12
EcoR1


EcoR1
Hind3
Nde1
Sac1
hGH
pTKGH
5056 bp
Bgl2
Xmn1
pUC12
EcoR1


Sst I
Bam HI
Hind III
Pst I
Mouse β major Globin Gene.
Hind III
Xba I
Sst I
pβ7.2
Pst I
EcoR I
Amp.R
pUC8
Eco RI
Pst I


Mouse β major Globin Gene
Promoter (346 bp).
Bam HI
Hind III
Nde I
-346
Sac I
hGH
pβ346GH
5163 bp
Xmn I
Bgl II
pUC12
EcoRI


Mouse β major Globin Gene
Promoter (346 bp).
Hind III
BamHI
Nde I
-346
Sac I
hGH
pβ(A)GH
5763 bp
Xmn I
Bgl II
pUC12
EcoRI
EcoRI
Chicken βA Globin Gene
3' Enhancer.


Mouse β major Globin Gene
Promoter (346 bp).
BamHI
Hind III
Nde I
-346
Sac I
hGH
pβ(M)GH
5578 bp
Xmn I
Bgl II
pUC12

PART A : Materials.

2.1 Molecular Clones.

Recombinant plasmids used in this study are listed in Figure 2.1. These clones were stored in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) at -20°C. pUC8 based clones were propagated in E.coli K12-derived bacterial strain JM83 (ara, Δ(lac-proAB), rpsL, Φ80, lacZΔM15).

2.2 Chemicals, Enzymes and Other Materials.

All restriction enzymes, polymerases and other DNA modifying enzymes were obtained from Boehringer Mannheim, Lewes, East Sussex, with the exception of Taq 1 DNA polymerase which was obtained with 10x Taq polymerase buffer from NBL, Cramlington, Northumberland.

Protease inhibitors, protein kinase A catalytic subunit, ribonuclease A, Brij-35, diethylpyrocarbonate, β-Mercaptoethanol, lysozyme, calf thymus DNA cellulose, o-nitrophenyl-β-D-galactopyranoside, Triton X-100, HMBA (N,N'-hexamethylene-bisacetamide) and phosphatase inhibitors were obtained from The Sigma Chemical Company, Poole, Dorset.

Deoxyribonucleotides, deoxyribonuclease 1, yeast t-RNA, DNA size markers, dialysis tubing, agarose, low melting point agarose and urea were obtained from Gibco/BRL, Paisley.

Double-stranded poly(dI)(dC) was obtained from Pharmacia Ltd., Milton Keynes.

Geneclean Kit for DNA fragment purification was obtained from

Stratatech Scientific Ltd.,London.

Elutitip-D columns for DNA fragment purification was obtained from Schleicher and Schuell, Dassel, FRG.

Biogel A-50M agarose beads for gel filtration column chromatography were obtained from Bio-rad Laboratories Ltd., Watford.

Single stranded oligodeoxyribonucleotides were synthesised on an Applied Biosystems 381A automated DNA synthesiser and purified by high performance liquid chromatography (Oswell DNA Service, University of Edinburgh).

Polymerase chain reactions were carried out using the Perkin Elmer Cetus DNA Thermal Cycler.

Phosphate Buffered Saline tablets were supplied by Oxoid Ltd.

Water-saturated phenol was obtained from Rathburn Chemicals Ltd., Walkerburn.

Gauze was obtained from Vernon-Carus Ltd. Preston.

Absolute ethanol was obtained from James Burroughs (FAD) Ltd., Essex.

Nalgene filters were obtained from Techmate Ltd., Luton.

Thin layer chromatography plates were obtained from Camlab, Cambridge.

Maxam-Gilbert sequencing reagents were obtained from Du Pont International.

X-OMAT AR XAR5 Film was obtained from Kodak, Glasgow.

All other chemicals and solvents were obtained from BDH Ltd., Poole, Dorset.

2.3 Media and Antibiotics.

L-broth : 1% w/v bacto-tryptone, 0.5% w/v yeast extract, 1% w/v NaCl. (Ampicillin added where required at 5oug/ml.

2xTY : 1.6% w/v bacto-tryptone, 1% w/v yeast extract, 0.5% w/v NaCl.

L-agar : L-broth containing 1.5% w/v agar.

H-agar : 1% w/v bacto-tryptone, 0.8% w/v NaCl, 1.2% w/v agar.

H-top agar : 1% w/v bacto-tryptone, 0.8% w/v NaCl, 0.8% w/v agar.

Superbroth

Soln. A : 1.2% w/v bactotryptone, 2.4% w/v yeast extract 0.5% v/v glycerol.

Soln. B : 12.5% w/v K_2HPO_4, 3.5% w/v KH_2PO_4

(Mix soln. A and soln. B in ratio 9:1 and pH to 7.2)

Bacto-tryptone, bacto-agar and yeast extract supplied by DIFCO Laboratories, Detroit. Antibiotics supplied by The Sigma Chemical Company. X-GAL (5-bromo-4-chloro-3-indolyl-β-galactoside) and IPTG (isopropyl-β-D-thio galactopyranoside) were supplied by BRL.

2.4 Cell Lines and Cell Culture Media.

Cell Line	Characteristics
F4-12B2	Adherent mouse erythroleukaemia cell line (Ostertag *etal* (1979).
M707/T	Suspension mouse erythroleukaemia cell line. (Harrison *et al* (1978).
C88	Suspension mouse erythroleukaemia cell line.

(Deisseroth & Hendrick (1978).

F4-12B2, M707/T and C88 are Friend virus induced erythroleukaemic cell lines.

STO — Adherent mouse fibroblast cell line. (Martin & Evans (1975).

NIH 3T3 — Adherent mouse fibroblast cell line. (Anderson et al (1979).

J774.2 — Adherent mouse macrophage cell line. (Ralph & Nakoinz (1977).

K562 — Suspension human erythroid cell line. (Lozzio & Lozzio (1975).

HeLa — Adherent human cervical carcinoma cell line. (Gey et al (1952).

F9 — Suspension mouse embryonal carcinoma cell line (Bernstine et al (1973).

Media.

F4-12B2, M707/T, C88, STO, J774.2, Hela and K562 were all grown in Special Liquid Medium (SLM) fortified with 10% foetal calf serum and 4mM L-Glutamine. NIH 3T3 were grown in SLM plus 10% new-born calf serum and 4mM L-Glutamine. F9 cells were grown in Dulbeccos Modified Eagle Medium plus 10% foetal calf serum, 4mM L-Glutamine and 1mM sodium pyruvate.

All media, minerals and GENETICIN/G418-sulphate was supplied by GIBCO/BRL Ltd.

2.5 Radio-labelled Materials.

D-threo-dichloroacetyl-1-^{14}C-chloramphenicol and [α-^{32}P]dCTP, [γ-^{32}P]ATP, [α-^{32}P]ddATP were supplied by Amersham

International. [^{125}I] labelled monoclonal antibody to human growth hormone was supplied by BioGenesis Ltd. as part of their Allegro hGH immunoassay.

2.6 Buffers and Solutions.

10x TBE	10.8% (w/v) Tris-Cl (pH 7.8), 5.5% (w/v) boric acid, 0.93% (w/v) EDTA.
10x TAE	4.84% (w/v) Tris, 1.142% (v/v) acetic acid, 1.8612% w/v EDTA, pH 7.8.
Storage Buffer (SB)	50 mM NaCl, 20 mM HEPES (pH 7.9), 5 mM $MgCl_2$, 0.1 mM EDTA, 20% (v/v) glycerol, 1 mM dithiothreitol.
E_x Buffer	X mM $(NH_4)_2SO_4$, 20 mM HEPES, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, 20% (v/v) glycerol, 0.1% Brij 35, (pH 7.9).
TMS	0.25 M sucrose, 5 mM $MgCl_2$, 10 mM Tris-Cl (pH 7.9).
High Salt Buffer (HSB)	1 M NaCl, 20 mM Tris-Cl (pH 7.4), 1 mM EDTA.
Low Salt Buffer (LSB)	0.2 M NaCl, 20 mM Tris-Cl (pH 7.4), 1 mM EDTA.
STE	0.1 M NaCl, 10 mM Tris-Cl (pH 7.8) 1 mM EDTA.
TE	10 mM Tris-Cl (pH 7.8), 1 mM EDTA.
10x Ligation Buffer (LB)	2.5 mM ATP, 10 mM spermidine, 20 mM DTT, 2 mM EDTA, pH 7.8
5x Ligation Salts (LS)	0.2 M NaCl, 0.2 M Tris-Cl (pH 7.8), 50 mM $MgCl_2$.

10x T4 Buffer	330 mM Tris-acetate (pH 7.9), 660 mM KCH_3CO_2 100 mM $Mg(CH_3CO_2)_2$, 5 mM DTT, 1 mg/ml BSA.
10xKinase buffer	0.5 M Tris-Cl (pH 7.6), 0.1 M $MgCl_2$, 0.05 M DTT.
10x CIP buffer	0.5 M Tris-Cl (pH 9.0), 10 mM MgCl, 1 mM ZnCl, 10 mM spermidine.
5% polyacrylamide	0.25% (w/v) N,N'-methylene bisacrylamide 4.75% (w/v) acrylamide (1:19 bis.:acryl.) 0.2% (v/v) N,N,N',N'-tetramethylethylenediamine, 1x TBE, 0.1% (w/v) $(NH_4)_2S_2O_8$
6% denaturing polyacrylamide	4.75% (w/v) acrylamide, 8 M urea, 0.25% (w/v) N,N'-methylene bisacrylamide 0.2% (v/v) N,N,N',N'-tetramethylethylenediamine, 1xTBE, 0.1% (w/v) $(NH_4)_2S_2O_8$
10x Loading Buffer	50% (v/v) glycerol, 50% (v/v) TE, 0.5% bromophenol blue.
Sequencing Loading Buffer	95% (v/v) deionised formamide, 0.1% (w/v) xylene cyanol FF, 0.1% bromophenol blue.
Proteinase K Buffer	10 mM Tris-Cl (pH7.8), 5 mM EDTA, 0.5% SDS.
Solution 1	50 mM glucose, 25 mM Tris-Cl (pH 7.8), 1 mM EDTA, 5 mg/ml lysozyme.
Solution 2	0.2 M NaOH, 1% SDS.
Footprinting Stop Buffer	83 mM Tris-Cl (pH 7.9), 8.3 mM EDTA, 1% (w/v) SDS, 107 mM NaCl, 0.4 mg/ml proteinase K, 67 µg/ml yeast tRNA.

2x HBS 1 mM Sorensen Salt, 50 mM HEPES (pH 7.1), 0.28 M NaCl. Filter sterilise.

β-galactosidase assay.

Solution A 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgCl_2$, 50 mM β-Mercaptoethanol.

Solution B 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 2 mg/ml o-nitrophenyl-β-D-galactopyranoside (ONPG).

Solution C 1 M Na_2CO_3

PART B : METHODS.

Nucleic Acid Preparation.

2.7 Large Scale Preparation of Plasmid DNA.

(This was carried out by a modified version of the alkaline lysis method for plasmid preparation described initially by Birnboim and Doly (1979).

A fresh 5 ml overnight bacterial culture was used to inoculate 200 ml of superbroth and was incubated in an orbital shaker at 37°C for 30 hours. The bacteria were pelleted at 4000 g for 10 minutes at 4°C. The supernatant was discarded and the pellet washed with 5 ml cold STE pH 8.0. The pellet was then resuspended in 5 ml of Solution 1 and incubated at 19°C for 5 minutes, incubated on ice for 2 minutes before 10 ml of Solution 2 was added. The mixture was incubated on ice for a further 10 minutes before adding 7.5 ml 3 M sodium acetate pH 5.2, vortexing for 10 seconds and leaving on ice for 10 minutes. After centrifugation at 3000 g for 15 minutes at 4°C, the supernatant was filtered through gauze and isopropanol (0.6 volumes) was added. After 15 minutes at 19°C, nucleic acids were pelleted by centrifugation at

3000rpm for 15 minutes at 19°C. The pellet was resuspended in 5ml of TE and incubated with 50 µl of boiled ribonuclease A solution (10 mg/ml) for 30 minutes at 37°C. Nucleic acids were extracted three times with an equal volume of 1:1 phenol:chloroform and once with an equal volume of chloroform. Plasmid DNA was then precipitated with 2 volumes of absolute ethanol at 19°C for 15 minutes, pelleted at 3000 g for 15 minutes at 4°C and resuspended in 1 ml of TE/0.1% SDS. Plasmid DNA was further purified gel filtration chromatography. The DNA solution was loaded onto a 30 cm x1.5 cm Biogel A-50M agarose bead column equilibrated with TE/0.1% SDS. The column was run in TE/0.1% SDS and the absorbance at 260 nm and 280 nm of 1.5 ml fractions collected from the column was determined by spectrophotometry. The first peak in absorbance represents plasmid DNA and these fractions were pooled, the DNA precipitated twice in ethanol and resuspended in TE at a concentration of 1 mg/ml. The second peak, which is well separated from the first, contains small molecular weight impurities and was discarded. It was found that running the column in 0.1% SDS, as well as inhibiting any nuclease activity, also increased the separation of the first plasmid DNA-containing peak from the second peak of low molecular weight impurities. Typically, this method of plasmid DNA preparation yielded 1-2 mg of plasmid DNA. The quality, yield and the extent of plasmid supercoiling was routinely checked by spectrophotometry and agarose gel electrophoresis.

2.8 Small Scale Preparation of Plasmid DNA.

1. 5 ml of L-broth containing selective antibiotic was inoculated

with the plasmid carrying bacteria and incubated overnight at 37°C.

2. 1 ml of overnight culture was pelleted in the microfuge for 30 seconds.

3. The supernatant was discarded and the bacterial pellet resuspended in 100 µl of Solution 1 at 19°C for 5 minutes.

4. 200 µl of Solution 2 was added and incubated on ice for 10 minutes.

5. 150µl of 3M NaAc (pH5.2) was added, vortexed briefly, incubated on ice for 10 minutes and spun in the microfuge for 5 minutes.

6. 400µl of supernatant was transferred to a fresh Eppendorf tube and extracted twice with an equal volume of 1:1 phenol:chloroform and once with an equal volume of chloroform.

7. 800µl of 95% ethanol was added and incubated at 19°C for 5 minutes.

8. Nucleic acids were pelleted in the microfuge for 5 minutes, the pellet dried and resuspended in 50 µl of pure distilled water containing 1 µl of 10 mg/ml RNAase A.

Nucleic Acid Manipulation.

2.9 Restriction Analysis of DNA.

Plasmid DNA was digested with 1-5 units of enzyme/µg DNA for 2 to 3 hours at 37°C under conditions specified by the enzyme manufacturer. Where further manipulation of the restricted DNA was necessary, the fragments were purified as described in 2.14.

2.10 Gel Electrophoresis.

For analysis or preparation of DNA fragments ranging in size from 100 bp to 8 kb flat bed agarose gel electrophoresis in 1x TAE/1x TBE was used. Where purification of the restricted fragment was required, low melting point agarose was used and the gel was run in 1x TAE buffer. Agarose gels were poured at a concentration of 0.8%-2% w/v depending on the size of fragments to be resolved. Samples were loaded in 1x loading buffer and electrophoresed for 20 minutes to 16 hours and at 10-100 mA depending on the size of the gel and the resolution required. Gels were stained for 20 minutes in 1 µg/ml ethidium bromide solution followed by 10 minutes destaining in distilled water to remove excess ethidium bromide. Nucleic acids were visualised by illumination with short wave ultra violet light and photographed through a red filter on Polaroid 57 high speed film. When a fragment identified on a gel was to be purified, it was excised from the gel on the UV illuminator and purified as described in 2.14

For analysis or purification of fragments less than 150 bp, 5-10% polyacrylamide 1x TAE/1x TBE vertical gel electrophoresis were used. The gels were electrophoresed at 10-100 mA for 1-4 hours. Gels were stained and fragments visualised as described for agarose gels above.

For analysis and separation of single stranded DNA/RNA, 0.4 mm thick vertical 6% denaturing polyacrylamide gels were used which were poured between siliconised glass plates. Gels were electrophoresed for 1-4 hours at 40 mA. Gels were pre-electrophoresed for 1 hour prior to the loading of samples

and samples were loaded in sequencing loading buffer and heated to 90°C for 2 minutes before loading. After electrophoresis, the gel was transferred to Whatmann 3mm paper and dried for 1-2 hours under vacuum at 80°C.

2.11 DNA Fragment Purification.

Two methods were used to purify DNA fragments from low melting agarose. For fragments greater than 500 bp, fragments were purified using GeneClean:

1. 2.5 volumes of 5M NaI were added to the excised block of agarose containing DNA, heated at 55°C for 5 to 10 minutes and vortexed occasionally.

2. 5 µl of glassmilk for up to 5 µg DNA were added, vortexed and incubated on ice for 5 minutes before being spun down for 5 minutes in the microfuge.

3. The pellet was resuspended in 450 µl NewWash and spun down 30 seconds. Repeat three times.

4. The pellet was taken up in 10 µl of TE, incubated for 3 minutes at 50°C and spin down for 30 seconds.

5. 10 µl of supernatant was transferred to fresh tube and step 4 above was repeated with a second aliquot of TE.

6. The supernatants were pooled and a sample analysed on a mini-gel.

For fragments less than 500 bp, purification from agarose is more efficient using Elutitip-D columns which are equilibrated with 3 ml of HSB followed by 5 ml of LSB. The agarose block containing the DNA fragment was melted in 1 ml of LSB at 65°C for 15 minutes, mixed with a further 8 ml of LSB and vortexing, and

incubated at 42°C for 15 minutes. 5 ml of the solution was pushed through an Elutitip-D column using a 5ml syringe. The column is washed quickly with 5 ml of LSB at 42°C and the DNA fragment is then eluted with 0.6 ml of HSB and precipitated with 30 µg of carrier yeast tRNA and 0.6 ml of absolute ethanol at -20°C overnight. This procedure is repeated for the remaining 5 ml of molten agarose solution.

DNA fragments are purified from acrylamide by elution overnight into 0.2 ml of 0.5 M $NH_4CH_3CO_2$, 10 mM $Mg(CH_3CO_2)_2$, 0.1 mM EDTA, 0.1% SDS at 37°C followed by ethanol precipitation in the presence of 30 µg carrier yeast tRNA.

2.12 Sub-cloning of DNA Fragments.

DNA fragments were subcloned into a vectors either by:

(a) ligation into a restriction site containing complementary sticky ends to those of the DNA fragment, or

(b) blunt-end ligation carried out by generating blunt ends on both the insert and the vector, phosphatasing the 5′ ends of the vector (to reduce the possibility of the vector religating to itself) and then blunt-end ligating the insert into the vector. Blunt ends were generated by treating DNA fragments with the Klenow fragment of DNA polymerase 1/T4 DNA polymerase in the presence of 0.2 mM deoxyribonucleotides, in 1x T4 buffer and incubating at 19°C for 30 minutes. Reactions were stopped by incubating at 70°C for 5 minutes. Vector fragments were phosphatased with 1 unit of calf intestinal phosphatase/50 µg DNA in 1x CIP buffer at 37°C for 30 minutes.

Ligation reactions were typically carried out in a volume of 20

µl with 200 ng of insert and 200 ng of vector in the presence of 1x ligation buffer, 1x ligation salts and 1 unit of T4 DNA ligase. In addition, 1 unit of T4 RNA ligase was added to blunt end ligation reactions. The ligation reactions were incubated overnight at 14°C for ligations involving sticky ends and at 4°C for blunt end ligations.

2.13 Transformation of DNA into E.coli.

Bacterial strain JM83 was made competent for tranformation 24 hours prior to transformation by the following procedure:

1. 1 ml of a fresh overnight of JM83 was inoculated into 100 ml of L-broth and incubated at 37°C in an orbital shaker until the culture reached an optical density of 0.6 at A_{600} when the bacteria should be growing exponentially.

2. The bacteria were pelleted by centrifugation at 1500 g at 4°C for 15 minutes and resuspended in 20 ml of transformation buffer at 4°C, and incubated on ice for at least 30 minutes.

3. The bacteria were pelleted at 1200 g for 15 minutes at 4°C, resuspended in 2 ml of transformation buffer and stored overnight at 4°C.

Transformation of DNA into competent JM83 is carried out as follows:

1. 10 µl of the ligation reaction described in 2.15 above was mixed gently on ice with 100 µl of competent JM83 bacteria.

2. The suspension was incubated on ice for 30 minutes and then incubated at 42°C for 5 minutes before being immediately returned to ice.

3. 2 ml of warm L-broth was added to the suspension, the culture

was incubated at 37°C for 1 hour, and the bacteria were pelleted at 1500 g for 15 minutes.

5. The bacterial pellet was resuspended in 250 µl of L-broth and 50 µl was spread out onto 9 cm diameter L-agar plates containing selective antibiotics and incubated at 37°C overnight.

6. Single colonies were then picked and inoculated into 5 ml of L-broth plus selective antibiotic. Plasmid DNA was prepared as described in 2.8 above and restriction analysis carried out to determine whether the sub-cloning and transformation had been successful.

End-labelling of DNA Probes.

2.14 5′ end labelling of double-stranded oligonucleotides.

Single-stranded complementary oligonucleotides were annealed at a concentration of 0.1 mg/ml in 0.1 M NaCl, 1x TE by boiling and then cooled slowly to 19°C. The 5′ end-labelling reaction was carried out as follows:

1. The reaction was set up in a total volume of 10 µl with 2 µl double-stranded oligonucleotide (200 ng), 1 µl 10x kinase buffer, 1 µl 50 mM DTT, 1 µl 20 mM spermidine, 2 µl [γ-^{32}P]ATP (370 MBq/ml), 2 µl water and and 1 µl T4 polynucleotide kinase.

2. The reaction mix was incubated at 37°C for 30 minutes, 70°C for 5 minutes, 37°C for 10 minutes and 19°C for 5 minutes.

3. End-labelled double stranded oligonucleotides for use in gel retardation assays were treated with Klenow fragment in a total volume of 20 µl made up of 10 µl of end-labelling reaction, 1 µl 10x T4 buffer, 4 µl 5mM deoxyribonucleotides, 2 units Klenow and 4 µl water and incubated on ice for 2 hours. This filling-in of

sticky ends on oligonucleotide probes reduces non-specific interactions with DNA binding proteins.

4. The end-labelled probes were purified by gel electrophoresis on an 8% polyacrylamide 1x TBE gel. The probes were eluted from acrylamide overnight into 1 ml TE at 37°C. The probes were labelled to a specific activity in the range 2-8x10^8cpm/µg.

2.15 5′ end labelling of DNA restriction fragments.

50 µg of plasmid DNA was cut with the appropriate restriction enzyme to generate 5′overhanging ends, extracted, precipitated and resuspended at 1 µg/µl. The restricted plasmid was then treated with 0.5 units of calf intestinal phosphatase in 1x CIP buffer in a total volume of 50 µl and incubated at 37°C for 30 minutes when a second aliquot of 0.5 units of phosphatase was added and the incubation extended for a further 30 minutes. The reaction was stopped by the addition of 50 µl water and 5 µl 10% SDS and incubated for 15 minutes at 68°C. Following this the DNA solution was organically extracted, precipitated twice and resuspended at 1 µg/µl. The end-labelling reaction was carried out in a total volume of 50 µl with 10 µl restricted, phosphatased DNA, 7 µl [γ-^{32}P]ATP (370 MBq/ml), 5µl 10x kinase buffer, 2.5 µl 20 mM spermidine, 2.5 µl 100 mM DTT, 3 µl T4 polynucleotide kinase (1 unit/µl) and 20 µl water.

2. The reaction was incubated at 37°C for 30 minutes and then stopped by the addition of 2 µl 0.2 M EDTA and 150 µl 10x TE.

4. The DNA was extracted twice with an equal volume of 1:1 (v/v) phenol:chloroform and once with an equal volume of chloroform.

The DNA was ethanol precipitated twice, first in 0.4 M NaCl and then in 0.1 M KCH_3CO_2.

5. Where end-labelling was required on one strand of the probe (for DNAase 1 footprinting and for Maxam-Gilbert sequencing), the pellet was resuspended in 35 µl of water, 5 µl of BSA (10 mg/ml), 5 µl 10x restriction buffer and 5 µl of an appropriate restriction enzyme (5 units/µl) and incubated at 37°C for 3 hours.

6. The end-labelled probe was purified by gel electrophoresis on a 1.5% low melting point agarose gel, fragments visualised by autoradiography and the fragment purified from agarose using Elutitip-D column as described in 2.14. Probes were generated at a specific activity in the range of $2\text{-}8\text{x}10^6$ cpm/µg.

2.16 5′ end labelling of single-stranded oligonucleotides.

The end-labelling reaction was set up with 0.2 µg of single-stranded oligonucleotide, 1 µl 10x kinase buffer, 1 µl 100 mM DTT, 2 µl [γ-^{32}P]ATP, 1 unit T4 polynucleotide kinase and made up to a total volume of 10 µl with water. The reaction mixture was incubated at 37°C for 30 minutes and then the reaction was stopped by incubation at 65°C for 10 minutes. The labelled oligonucleotide was used directly for generating 5′end labelled DNA fragments by PCR amplification.

2.17 Polymerase Chain Reaction (PCR) Mediated End Labelling of DNA Fragments.

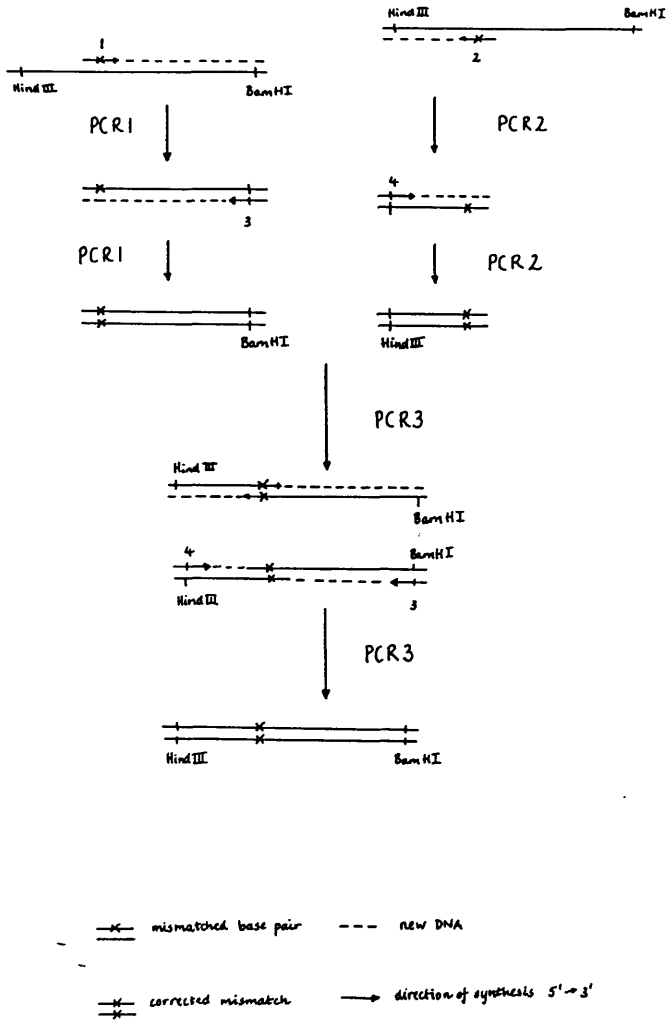
The conditions for polymerase chain reactions are discussed in more detail in 2.24. Where no suitable restriction enzyme sites existed to generate the desired end-labelled DNA fragments, PCR

mediated extension on the chosen template was used. The reaction was primed with an end-labelled oligonucleotide primer complementary to the region of the template from which the probe was generated and with an unlabelled primer complementary to the opposite strand at the distal end of the template. The PCR reaction contained 10 µl of the single-stranded oligonucleotide labelling reaction described above, 0.2 µg of the second oligonucleotide primer, 10 ng of template, 5 µl of 10x Taq polymerase buffer, 5 µl of dimethylsulphoxide (DMSO), 4 µl 10 mM deoxyribonucleotides, 1 µl 100mM $MgCl_2$ and made up to a total volume of 50 µl with water. Two drops of mineral oil were applied to the surface of the reaction mix before the reaction was boiled for 5 minutes and then allowed to cool slowly to 19°C. 2 units of Taq DNA polymerase were added to the reaction mix and the desired DNA fragment end-labelled at one end only was amplified in the Perkin Elmer Cetus DNA Thermal Cycler. Typically 30 cycles of amplification were carried out overnight at a denaturation temperature of 91°C for 90 seconds/cycle, an annealing temperature of 50°C for 90 seconds/cycle and a polymerisation temperature of 72°C for 180 seconds/cycle. The DNA probe is extracted once with an equal volume of chloroform and purified by gel electrophoresis on a 1.5% low melting point agarose gel. The DNA probe is extracted from the agarose using an Elutitip-D column. Typically the specific activity of fragments generated in this way was in the range $2\text{-}8\text{x}10^8$cpm/µg.

2.18 Maxam-Gilbert Sequencing.

Maxam-Gilbert sequencing reagents were provided by Du Pont

Figure 2.2 PCR Mediated Site Directed Mutagenesis.
The figure summarises the method described in section 2.19 to generate desired mutations in sequences as originally described by Higuchi et al (1988).



International. End-labelled probes were generated as described above and Maxam-Gilbert sequencing reactions were carried out according to the manufacturer's specifications.

2.19 PCR Mediated Site Directed Mutagenesis.

The use of polymerase chain reaction as developed by Saiki *et al.* (1987) to generate specific mutations in genes of interest was first described by Higuchi *et al.* (1988) and further described by Vallette *et al.* (1988) (The general applications of polymerase chain reaction for nucleic acid analysis and manipulation is reviewed by White *et al.*(1989). In this study, the technique was modified in several ways until such time as the yield and fidelity of amplification of the specifically mutated DNA fragment was maximised. The procedure for generating mutant DNA fragments for sub-cloning is summarised in Figure 2.2. The reactions were carried out as follows:

PCR Reactions 1/2

0.3 µg mutant primer 1/2 (mutant primers 1 and 2 are complementary)

0.3 µg linker primer 3/4 (linker primers 3 and 4 contain restriction sites which will be used for sub-cloning)

10 ng wild-type template

5 µl 10x Taq polymerase buffer

5 µl DMSO

4 µl 10 mM dNTP

1 µl 100 mM $MgCl_2$

water up to a total volume of 50 µl.

Two drops of mineral oil were added to the surface of the reaction mix and boiled for 5 minutes, and then cooled slowly to 19°C. 2 units of Taq DNA polymerase were added and amplification carried out in the Perkin Elmer Cetus DNA Thermal Cycler at a denaturation temperature of 90°C for 90 seconds/cycle, an annealing temperature of 40-45°C (depending on the extent of mismatch between the mutant primer and the wild-type template) for 90 seconds/cycle and a polymerisation temperature of 72°C. Typically 30 cycles of amplification were carried out.

The products of PCR reactions 1 and 2 were separated by gel electrophoresis on a 1.7% low melting point agarose gel and purified on an Elutip-D column. The yield of the desired product is typically 1-2 µg.

PCR Reaction 3

5 µg PCR reaction 1 products
5 µg PCR reaction 2 products
0.3 µg linker primer 3
0.3 µg linker primer 4
5 µl 10x Taq polymerase buffer
5 µl DMSO
4 µl 10 mM dNTP
1 µl 100 mM $MgCl_2$
Water to a total volume of 50 µl.

2 drops of mineral oil were added to the surface of the reaction mix, boil for 5 minutes, cooled slowly to 19°C and 2 units of Taq polymerase were then added. 30 amplification cycles were carried out in the Thermal cycler at a denaturation temperature of 90°C for 90 seconds/cycle, an annealing temperature of 50°C for 90

seconds/cycle and a polymerisation temperature of 72°C for 180 seconds/cycle. 4 units of the restriction enzyme appropriate to the restriction site contained within the linker primers at each end of the PCR product were added. The PCR products were separated by gel electrophoresis on a 1.7% low melting point agarose gel. The desired PCR generated fragment was subcloned into a suitable vector, positive recombinants screened for by restriction analysis and Maxam-Gilbert sequencing carried out to confirm successful mutagenesis. Where appropriate a unique restriction site was generated at the site of mutagenesis to facilitate the screening for successful mutagenesis and to obviate the need for DNA sequencing as a screening procedure.

Cell Culture and Expression Analysis.

2.20 Growth of Cells in Culture.

The cells used in this study are listed in 2.4. Cells grown in suspension were kept growing exponentially by subculturing every three to four days maintaining the cell density within the range $5x10^4$ to $2x10^6$ cells/ml. Cells grown in an adherent monolayer were also passaged every three to four days such that they were always maintained, depending on the cell line, within the range 30-80% confluent, that is $3x10^4$cells/cm^2 to $2x10^6$cells/cm^2. Adherent cells were removed from culture flask or dishes by removing the culture medium, washed with an appropriate volume of phosphate buffered saline and then removed from the flask with 0.025% (w/v) trypsin in citrate buffer (pH 7.8). The trypsin was then inactivated with an equal volume of fresh culture medium.

Cell number and integrity was determined by haemocytometry. Stocks of cell lines at 10^7 cells/ml in 10% (v/v) glycerol were stored in liquid nitrogen. Frozen stocks were made by slow freezing overnight to -70°C before transfer to liquid nitrogen. Cells were taken up from liquid nitrogen by rapid thawing at 37°C and immediate transfer to culture medium also at 37°C. Cells in culture were discarded after no more than 20 passages in favour of fresh cells taken up from frozen stocks.

2.21 Calcium Phosphate Mediated Transfer of DNA into Cells in Culture for Functional Analysis Under Transient Transfection Conditions.

10^6 exponentially growing cells in 10 ml of fresh medium were placed in 9 cm Petri dishes and incubated overnight at 37°C, in a humid atmosphere and at 5% CO_2.

DNA/calcium phosphate precipitates were generated by making a solution of 0.25 M $CaCl_2$ containing 40 µg of DNA (test plasmid+reference plasmid+carrier plasmid) up to 1ml with Tris-Cl (pH7.1). This solution was then added slowly while gently shaking to 1 ml of 2xHBS (pH 7.1) and mixed by aeration. The calcium phosphate precipitate was allowed to form at 19°C for at least 30 minutes before being gently added to the cells set up in culture 24 hours earlier.

Approximately 16 hours after addition of the precipitate, the medium on the cells was replaced with fresh medium and the cells returned to the incubator at 37°C, in 5% CO_2.

Culture medium/cell lysates (depending on the nature of the

functional assay) were assayed for activity of the reporter gene product 24 hours following change of medium.

The efficiency of transfection varies between cell lines and in some cases it may be more efficient to use different methods of transfection, for example lipofection or electroporation. However, in this study, although there were differences in transfection efficiencies between cell lines and between experiments, these differences were controlled for by cotransfecting the test plasmid with the pHSV-IE-β-galactosidase reference plasmid and standardising levels of expression from the test plasmid to that from the reference plasmid.

2.22 Generation of Stable Transfectants for Functional Analysis.

The DNAs to be assayed were transfected into cells in culture dishes as described above for transient transfection except that instead of the pHSV-IE-β-galactosidase reference plasmid, the test plasmid was cotransfected with HOMER 6 (Spandidos & Wilkie) which encodes the neomycin resistance gene. On day 3 when the culture medium was changed and fresh culture medium added, 80 μg/ml of G418 is also added to the culture medium. Three days later, only those cells which had stably incorporated and expressed the neomycin resistance gene from HOMER 6 survived the G418 selection. It was assumed that those cells which had stably incorporated the neomycin gene into their genome will also have incorporated the test plasmid sequences. The medium on these cells was replaced with fresh medium containing 80 μg/ml of G418.

Three days later colonies of the cells that had survived selection were seen with the naked eye. Each colony represented a different clone each having incorporated different numbers of copies of the resistance gene and the test plasmid sequences at different locations in the genome. More than 50 clones were pooled by trypsinising them from the culture dish and reseeding in fresh medium at a density of $4x10^4$cells/ml. The cells were then cultured and passaged as for other cells as described above except that 80 µg/ml of G418 was added to the culture medium. Cells could then be induced to differentiate with 4 mM HMBA or with 2% DMSO and changes in expression from the reporter gene monitored at 24 hour periods. The cells were set up as normal and allowed to grow for 24 hours before the inducing agent was added.

2.23 Human Growth Hormone Assay.

This technique was first described by Selden et al. (1986). The human growth hormone reporter gene was driven by the test promoter and/or enhancer sequences. When the expression construct was transfected into cells in culture in either transient or stable transfection experiments, expressed human growth hormone (hGH) was secreted from the cells and the culture medium was assayed directly for growth hormone activity. The radioimmunoassay, supplied by Biogenesis Ltd., was carried out according to their specifications. hGH in the culture medium was bound by a mouse anti-hGH [^{125}I]-labelled monoclonal antibody. A second biotin-linked mouse monoclonal antibody raised against a different epitope of hGH also binds hGH and the hGH sandwiched

between two antibodies is recovered from the reaction mix by incubation for 4 hours at 19°C with avidin coated glass bead. The beads bind the biotin linked monoclonal antibodies and having washed off unbound antibody, the amount of [^{125}I] indirectly linked to the glass bead was taken as a measure of the ammount of hGH present in the sample of culture medium. The amount of [^{125}I] on the glass beads is measured by its gamma emission in a Beckman Gamma Counter. The hGH assay is more sensitive than the CAT assay and as such is useful for the functional analysis of weak promoter sequences. Similarly, the assay is easy and fast and since hGH is secreted from the cells, stable transfectants can be continuously cultured and culture medium assayed at different time points.

2.24 Generating Cell Lysates.

Following transient transfection, cell lysates were made to assay for β-galactosidase and CAT activity within the cells.

Medium was removed from the culture dish, cells washed in 10 ml of PBS, and then scraped off the dish in 1 ml of PBS. The cells were pelleted at 2500 g for 30 seconds and the cell pellet resuspended in 200 µl of 0.25 M Tris-Cl (pH 7.8). The cells were lysed by repeated rapid freezing and thawing. Cell debris was pelleted at 2500 g for 2 minutes and the supernatant assayed for β-galactosidase/CAT activity.

2.25 β-Galactosidase Assay.

60 µl of cell lysate, 1 ml of solution A and 0.2 ml of solution B were mixed by pipetting and then incubated at 37°C for 30 to 60 minutes. The reaction was stopped by addition of 0.5 ml of

solution C. The absorbance at 420nm is determined as an indirect measure of β-galactosidase activity. Absorbance values in the range of 0.15-1.00 were obtained.

2.26 Chloramphenicol Acetyl Transferase Assay.

This technique was carried out as described by Gorman et al. (1982) and modified as follows:

1. 10mM stock (8.1mg/ml) of acetyl Co-A was made and mixed with D-threo-dichloroacetyl-1-^{14}C-chloramphenicol (2.11GBq/ml) in the ratio 10:1.

2. Cell lysate was made up to 89µl where necessary with 0.25 M Tris-Cl (pH 7.8) heated to 60^{o}C for 7 minutes before adding the acetyl Co-A mix in order to inactivate endogenous cellular chloramphenicol acetyl transferase activities. 11 µl of the acetyl Co-A mix was added to the lysate.

3. The reaction was incubated at 37^{o}C for 30 minutes, stopped by placing on ice and 300 µl of ethyl acetate was added. After vortexing, the organic and aqueous phases were separated by centrifugation in the microfuge at 2000 rpm for 2 minutes.

4. The organic phase was removed to a clean Eppendorf tube, dried under vacuum and the pellet resuspended in 20 µl of ethyl acetate.

5. The 20 µl samples are spotted slowly onto a thin layer chromatography plate and reaction products separated by running in a 95:5 $CHCl_3$:CH_3OH liquid phase.

6. The plate was then allowed to air dry and exposed overnight against Kodak XAR-5 film.

Analysis of DNA-Protein Interactions.

2.27 Preparation of Crude Nuclear Protein.

Crude nuclear protein was prepared from animal tissue or from cells grown in culture in a modified version of the procedure initially described by Emerson et al. (1982). The protocol for preparation from cells in culture is essentially the same as for preparation from animal tissue after step 1 below which is for animal tissue only. All solutions used in the preparation were kept at 4°C and contained 0.5 mM PMSF, 0.1 mM EGTA, 1 µg/ml of each of aprotinin, leupeptin, pepstatin A, 0.1 µM sodium vanadate, 10 mM sodium butyrate, 0.5 mM benzamidine and 10 mM β-glycerophosphate.

1. Typically 100 NIH mice aged 2-4 months were used for each preparation. The tissue was collected immediately into 200 ml of PBS on ice and minced into 200 ml of TMS/0.25% Triton X-100, and homogenised in a Dounce homogeniser and filtered through gauze. The nuclei were pelleted at 2000rpm, 4°C for 20 minutes.

3. The nuclei were washed three times with TMS. During the third wash 100 µl of resuspended nuclei were added to 0.9 ml of water and then added to 9 ml of 1 M NaOH. Following sonication of this 10 ml solution the concentration of nucleic acid was determined by spectrophotometry to give an indirect measure of the concentration of nuclei in the preparation (A_{260} 20 = 1 mg/ml).

4. After the third wash in TMS, the nuclei were resuspended in fresh TMS at a nucleic acid concentration of 5 - 10 mg/ml.

5. Non-histone proteins were extracted from the nuclei stirring with 0.3 M NaCl on ice for 15 minutes and the supernatant

clarified by centrifugation first at 15 000 g for 15 minutes and then at 100 000 g for 1 hour at 4°C.

6. Solid $(NH_4)_2SO_4$ was added to the solution at 0.35 mg/ml and stirred on ice for 30 minutes. Precipitated proteins were recovered by centrifugation at 10 000 g for 30 minutes at 4°C.

7. The protein pellet was resuspended in 3-5 ml of E_{50} buffer and dialysed overnight at 4°C against SB (or against E_{50} buffer for partial purification by DNA affinity chromatography).

8. The crude nuclear protein preparation dialysed overnight against SB was then clarified at 100 000 g for 1 hour at 4°C and stored at -70°C.

2.28 Partial Purification of Nuclear Protein by DNA Affinity Chromatography.

5g of calf thymus DNA cellulose was equilibrated in 50ml of E_{500} buffer and used to pack a Pharmacia C 10 column. The column was washed through with a small volume of E_{500} buffer. The column was equilibrated overnight with 200 ml E_{50} buffer at 4°C. 5 ml of the crude nuclear preparation generated in 2.27 which had been dialysed against E_{50} buffer was loaded on to the column. 1 ml fractions were collected and the protein content measured by spectrophotometry. Once the first peak of activity came off the column, DNA binding prteins were eluted with step-wise increases in NaCl concentration; first E_{100}, then E_{250} and fraction collection continued until the second peak comes off the column. Fractions from the E_{250} peak were pooled and dialysed overnight against SB at 4°C. The pooled fractions were stored at -70°C for use in DNAase 1 footprinting assays.

2.29 DNAase 1 Footprinting.

This technique was used to identify protein binding sites within DNA fragments of interest and was carried out as described by Plumb and Goodwin in Chapter 12 of Methods in Molecular Biology Chapter 4.

End-labelled probe was generated as described in section 2.15 and less than 20 ng was incubated in storage buffer on ice in the presence of non-specific DNA competitor (1 µg poly (dI)(dC), 0-800 µg of nuclear protein extract in a total volume of 100 µl. Sequence specific DNA oligonucleotide competitor was added where necessary at a 250 M excess. The reaction was incubated on ice for 60 minutes before DNAase I was added at an appropriate concentration. The concentration of DNAase I to be added depended on the protein extract and competitor DNA concentration and was determined by carrying out a set of preliminary digestion reactions varying the DNAase I concentration from 1-10 µl of serial dilutions in SB/0.1 mM $CaCl_2$ of a stock DNAase I solution of 2 µg/µl. The concentration of DNAase I was selected such that only one nick per end-labelled probe molecule was made. After 30 seconds, the digestion reaction was stopped with 100 µl of stop buffer and then incubated at 37°C for 30 minutes and 90°C for 2 minutes. The DNA probe was extracted by repeated phenol/chloroform extraction and precipitated with 10 µl 5 M LiCl and 600 µl of absolute ethanol on dry ice for 15 minutes. The DNA precipitate was pelleted, washed with 1 ml ethanol, dried under vacuum and resuspended in 8 µl of sequencing loading buffer. The samples were denatured at 90°C for 2 minutes and resolved by 6%

denaturing polyacrylamide gel electrophoresis for 90-180 minutes depending on the region of the probe to be resolved. Maxam-Gilbert sequencing reactions with the same DNA probe were run along side the footprinting reactions. The gel was dried under vacuum and exposed against Kodak XAR-5 film overnight at -70°C in a cassette with intensifying screens.

2.30 Gel Retardation Assays.

This technique was used to analyse the binding characteristics of DNA-binding proteins in different tissue and cell extracts to different double-stranded oligonucleotide probes. The technique was carried out in a modified version of that described by Garner and Revzin (1981).

The binding reaction was set up in a total volume of 20 µl with 1 µl of double-stranded end-labelled oligonucleotide probe (100 pg), 2 µl poly(dI)(dC) (3 µg/µl), 1 µl 0.1 mg/ml specific oligonucleotide competitor, 2 µl of nuclear extract and 14 µl of SB at a final NaCl concentration of 50 mM or 150 mM. The reaction was incubated on ice for 60 minutes and then protein-DNA complexes resolved by gel electrophoresis on a 0.2x TBE/0.5xTBE 5% polyacrylamide gel pre-electrophoresised for 90 minutes on ice at 150V. The gel was electrophoresised at 150V constant voltage for 2 hours on ice. The gel was then dried under vacuum at 80°C for 2 hours and autoradiographed.

CHAPTER 3 IDENTIFICATION OF PROTEIN BINDING SITES IN THE PROMOTER AND 3' FLANKING REGION OF THE MOUSE β^{MAJOR} GLOBIN GENE.

INTRODUCTION.

A DHSS has been mapped over the mouse β^{major} globin gene promoter and the nuclease sensitivity of this site in MEL cells increases when these cells are induced to differentiate (Sheffery et al. (1982), (1983). Mutagenesis analysis of the mouse β^{major} globin gene promoter has shown that there is an absolute requirement for intact TATA (-30 bp), CCAAT (-75 bp) and CACC (-90 bp) boxes to ensure correct initiation of transcription in transient and stable transfection assays (Charnay et al. (1985), Myers et al. (1986). These elements map within the DHSS identified over the β^{major} globin gene promoter. From this and other evidence that nucleosome phasing is disrupted in this region in MEL cells (Benezra et al. (1986), it seemed likely that, in common with the chicken α^{D}, β^{H} and β^{A} globin gene promoters (Kemper et al. (1987), Emerson et al. (1985), Plumb et al. (1986), that non-histone proteins bind to these sites, possibly displacing histones and are involved in regulating the initiation of transcription from the mouse β^{major} globin gene promoter. Using in vitro DNAase 1 footprinting as described in section 2.29, I set out to identify sites in the mouse β^{major} globin gene promoter which bind factors present in crude and partially purified nuclear extracts prepared as described in sections 2.27 and 2.28. Such sites might be involved in the regulation of transcriptional initiation and by analysing the sequences bound by nuclear protein factors and the tissue specificity of binding activity, I hoped to determine whether any of the identified sites were conserved in the

promoters of other globin or non-globin genes.

End-labelled probes for DNAase1 footprinting were prepared by the PCR technique, as described in section 2.17 or by 5'end kinase labelling, as described in section 2.15. The PCR technique uses plasmid pβ745GH as template and oligonucleotides FP1-6 as primers to generate probes β730, β600, β460, β280 and β160. pβ7.2 was used as template and oligonucleotides FP7-8 as primers to generate probe β415. The recombinant plasmids used here are illustrated in figure 2.1 and oligonucleotides used are illustrated in Table 3.1. Figure 3.1 summarises the organisation of probes and primers on the promoter and 3'sequences.

3.1 The Proximal Promoter.

The proximal promoter can be defined as the region 5' of the mouse β^{major} globin gene spanning from the initiation site of transcription to a site 100 bp upstream. This region contains the TATA (-30 bp), CCAAT (-75 bp) and CACC (-90 bp) boxes and Myers et al. (1986) have suggested that sequences contained within this region are sufficient to confer erythroid specificity and inducibility upon transcription from this promoter in transient transfection assays.

To determine whether the functionally important TATA, CCAAT and CACC boxes bind nuclear factors and whether any additional protein binding sites exist in the proximal promoter region, footprinting probe β730 was incubated in vitro with crude or partially purified nuclear extract prepared from a variety of cell lines and mouse tissues, in the presence or absence of sequence specific oligonucleotide competitors, (as described in section 2.29) except that

incubations with partially purified nuclear extracts were carried out in the absence of poly(dI)(dC) non-specific competitor. The results are shown in Figure 3.3A and B. Table 3.1 illustrates the oligonucleotides used for both DNAase 1 footprinting and gel retardation assays.

The interaction of the proximal promoter region with crude nuclear protein extracts (Fig.3.3A) results in the generation of a strong DNAase 1 hypersensitive site at -100 bp 5′ of the initiation site. This DHSS can be competed out with excess of the βCC oligonucleotide competitor (lanes 8 and 16). To determine whether any other nuclear factors were involved in the generation of this DHSS at -100 bp, competition with other sequence specific oligonucleotide competitors was carried out. Incubating probe β730 in the presence of crude nuclear extracts with oligonucleotide competitors βCT (lanes 7, 15), βG215 (lane 9), NF-1 (lanes 10, 17) or βNF-1 (lanes 11,18) did not affect formation of the DHSS.

No footprints are observed in the proximal promoter region. The same observation has been made in DNAase 1 footprinting of the chicken β^H globin gene proximal promoter (Goodwin pers.comm.). It is surprising that this region does not footprint since it contains the functionally important TATA, CCAAT and CACC boxes. The same crude nuclear extracts generate footprints over the proximal promoter region of the mouse α-1 globin gene (lanes 19-24). Specifically, footprints are observed over the CACC box (-60 bp) of the mouse α-1 globin gene which can be competed out by a 250 M excess of oligonucleotide competitors, βCACC (lane 22) and αCACC (lane 24) and at a site at -40 bp 5′ of the transcription initiation site previously identified by Plumb et al. This suggests that the failure to obtain a footprint over

the functionally important TATA, CCAAT and CACC boxes of the mouse β^{major} globin gene promoter does not reflect degradation or inactivity of the binding factors. Binding to the proximal promoter region of the mouse β^{major} globin gene may depend either upon a critical DNA:protein ratio or upon protein:protein interactions which do not occur in this in vitro assay system. Failure to obtain a footprint over this region may be caused by low binding affinity of these proximal promoter sequences for their cognate binding factors.

Partial purification of crude nuclear extracts by DNA affinity chromatography (as described in section 2.28) enriches the extracts for DNA binding proteins and to determine whether any additional protein binding to the mouse β^{major} globin gene promoter could be detected, DNAase 1 footprinting using partially purified extracts was carried out.

DNAase 1 footprinting of β730 incubated with partially purified extract from MEL cells (lane 3) generates a large footprint spanning from -56 bp to -100 bp 5′ of the transcription initiation site (Figure 3.3B). This region contains the CCAAT (-75 bp) and CACC (-90 bp) boxes and also a recognition site for NF-E1 (-60 bp). Binding to this region also results in the generation of the strong DHSS at -100 bp described above. Competition with a 250 M excess of oligonucleotide βCT containing the mouse β^{major} globin gene CCAAT box sequence resulted in a loss of the footprint over this region but the DHSS at -100 bp remained (lane 8). However, competition with 250 M excess of oligonucleotide βCC (lane 9) containing the mouse β^{major} globin gene CACC box sequence resulted in the loss of the DHSS site at -100 bp in addition to loss of the footprint over this region. Competition with a

250 M excess of oligonucleotide βG215 (lane 10) results in a footprint reduced in size spanning from -70 to -90 bp but which still covers the CCAAT and CACC box sequences. The DHSS at -100 bp persists. Competition with a 250 M excess of oligonucleotide NF-1 (lane 4) does not alter the size of the footprint in the region or the intensity of the DHSS at -100 bp.

These results suggest that at least three different MEL factors bind to this region: a CACC box binding factor which is responsible for generating the DHSS at -100 bp and whose binding is also necessary for binding of NF-E1 and the CCAAT box binding factor; a CCAAT box binding factor whose binding is not altered by competition with NF-1 or βG215 oligonucleotide competitor but binding of which is competed out with excess of the βCC oligonucleotide; and NF-E1 which recognises the sequence AGGATAGA and binding of which is competed out by both βG215 and βCC. The ability of the βCC oligonucleotide to compete out binding not only to the CACC box but also the CCAAT box and to the NF-E1 binding site suggests that the CACC box binding factor is involved in a cooperative interaction with the factors binding to these elements. This is consistent with work by Schule et al (1988a, 1988b) that the CACC box is involved in a cooperative interaction with the GRE in the promoter of the tryptophan oxygenase gene and that the CACC box cooperates with the NF-E1 binding site in the promoter of the human PBG-D gene (Walker et al. Unpublished results).

Footprinting of this region with partially purified nuclear extracts from mouse liver (lanes 5-7) reveals a different pattern of binding and competition. In addition to the DHSS at -100 bp, binding by factors present in the liver extracts generates a second strong DHSS at

-90 bp and a footprint spanning from -56 to -90 bp (lane 5). Competition with a 250 M excess of oligonucleotide βCT (lane 6) results in loss of the DHSS at -90 bp and a weaker footprint over the region -56 bp -90 bp. Competition with a 250 M excess of oligonucleotide competitior βCC (lane 7) results in the loss of the DHSS at -100 bp but in an increase in the intensity of the DHSS at -90 bp and a new but weak DHSS at -66 bp. These results suggest that binding to the CCAAT box in liver does not depend upon binding to the CACC box unlike the situation with MEL extracts. The pattern of DHSS generated by the binding to the CCAAT box in liver extracts suggests that a different CCAAT box binding factor is active in liver to MEL cells.

In addition to footprints observed in the proximal promoter region, footprints are also observed in the distal region when footprinted with partially purified extracts which are not observed in footprinting experiments with crude nuclear extracts.

Footprints are observed over the regions -118 to -134 bp and -160 to -169 bp when β730 is footprinted with partially purified extracts from both MEL and liver nuclei (lanes 3, 5). Binding to both these regions is competed out by an excess of NF-1 (lane 4) and an excess of βCT (lanes 6, 8) oligonucleotide competitors. The sequence bound between -160 to -169 bp is a Box 1 recognition sequence AAGCCAGTG as described by Antoniou et al. (1988). in the human β globin gene promoter and is a CCAAT box like recognition sequence which explains why binding to this site is successfully competed by NF-1 and βCT oligonucleotides. Although this site is bound by factors present in both MEL and liver extracts, the pattern of binding in each extract is different; the DHSS present at -165 bp in MEL cell nuclear extracts is

absent in liver extracts. This suggests that different factors are binding to this region in MEL and liver nuclear extracts.

Protection of this region -118 to -134 bp with MEL and liver nuclear extracts occurs between two DHSS: at -100 bp caused by binding to the CACC box and at -135 bp which is probably generated by the poly purine tract in this region and not by protein binding since this DHSS is also present in the zero protein control track (lane 2). The protected region -118 to -134 bp does not contain any known protein binding recognition sequence. It may represent a novel binding site or alternatively, the protection observed in the region may be a conformational effect of DNA bending at the two DHSS at -100 and -135 bp.

In conclusion, crude nuclear extracts from MEL cells do not generate footprints over the mouse β^{major} globin gene proximal promoter while partially purified nuclear extracts do. There are several technical reasons for this; for example, the DNAase1 footprinting reactions are carried out in the absence of non-specific polynucleotide competitor poly (dI)(dC) which may reduce the amount of non-specific interactions of the factors specifically recognising the proximal promoter binding sites and release them for binding to the DNA probe. This is unlikely since even when crude nuclear extract incubations are carried out in the absence of poly (dI)(dC), footprints are not obtained in this region. Similarly, since partially purified nuclear extracts have been separated from residual DNA or RNA present in crude nuclear extracts, again non-specific interactions of binding factors may be reduced. Partial purification of nuclear protein extract may result in the dilution, loss or inactivation of some factor which may

be inhibiting the proximal promoter binding factors from binding the DNA. Alternatively, partial purification may be altering the relative concentration of factors such that functional DNA binding complexes can now form. Other protein-DNA interactions occurring elsewhere in the promoter which prevent complex formation in the proximal promoter, may be altered by fractionation of the factors involved. The binding activities of these different factors and their tissue and sequence specificities is further investigated by gel retardation assay and is described in chapter 4.

Figure 3.1 Generation and Organisation of Footprinting Probes.
Figure 3.1A illustrates the probes which have been generated by PCR mediated end-labelling as described in section 2.17 which have been purified on a 1.5% low melting point gel.

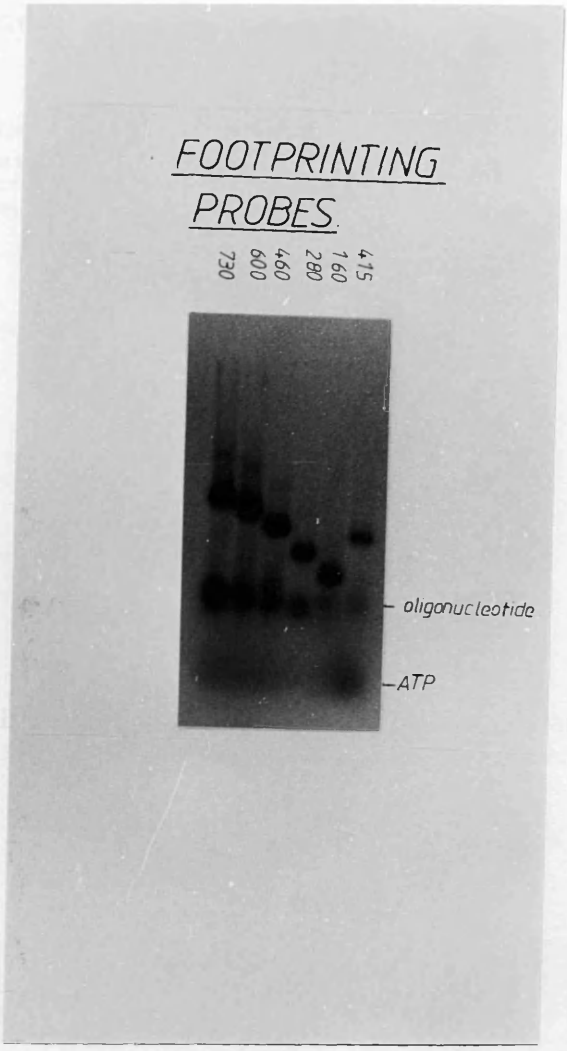


Figure 3.1B illustrates the organisation of primers FP1-6 on template pβ745GH relative to the transcription initiation site. Primer FP6 is common to all amplification reactions and is unlabelled. Primers FP1-5, which are 5' end-labelled, are used to generate probes β730, β600, β460, β280 and β160 respectively.

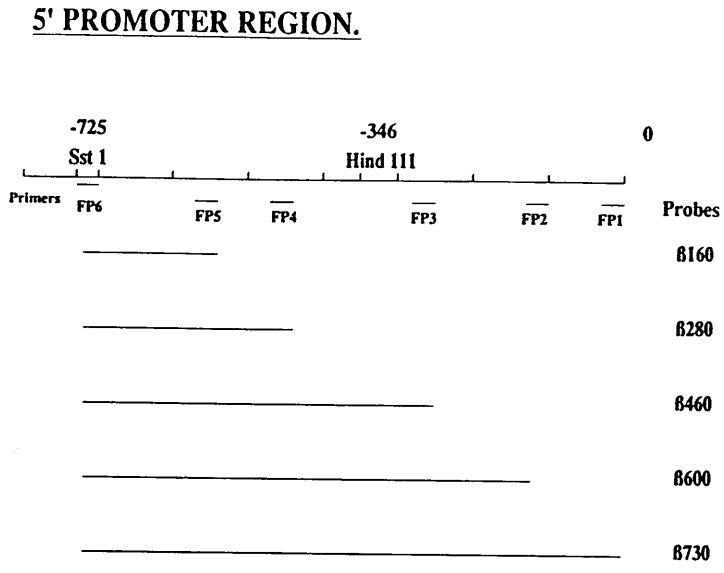


Figure 3.1C illustrates the organisation of primers FP7 and 8 on template pβ7.2 to generate β415. Either FP 7 or 8 can be end-labelled.

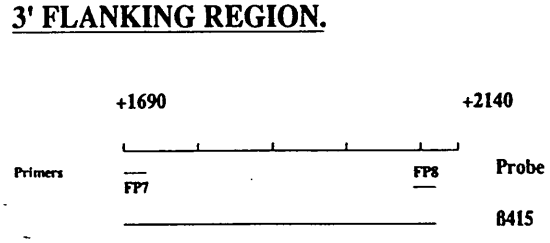


Table 3.1 Oligonucleotides Used as Primers, Competitors and Probes in DNAase I Footprinting and Gel Retardation Assays.

The table illustrates the single stranded oligonucleotide primers (top) used to prime the PCR mediated generation of probes for DNAase I footprinting as described in section 2.17 and illustrated in figure 3.1.

The table (below) also illustrates the single stranded oligonucleotides which were annealed with their complementary oligonucleotides to generate double stranded oligonucleotides which were then used as described in the text as sequence specific competitors in DNAase I footprinting reactions or as probes and/or competitors in gel retardation assays. The recognition sequences for respective cognate binding factors are underlined.

FOOTPRINTING OLIGONUCLEOTIDE PRIMERS.

GCA AAT GTC AGG AGC AAC TGA TCC TAC	FP1
GAA GCG ATA TTT CTC CCT GGA CAT GCT	FP2
GCT CAA TCA ACT ACT GAA TTG TGT T	FP3
AGC ATA CAT GAT GAT GAT TCC ATT	FP4
GTT TGT GTG GCT TAC TTG TAT ATA TGT	FP5
AGC TCA GGG TTT ACT TGA GAG ATC CTG	FP6
CAG TTC TCA AGC CAA TAA TTT TTC	FP7
CCC ACA CTG GAT ATT CAA GAT AGA T	FP8

COMPETITOR OLIGONUCLEOTIDES.

GAT CGC AAA TGC GTT CGC CAA AAA GGA	ßNF1
GAT CTT ATT TTG GCT TGA AGC CAA TAT G	NF1
CAT TTT TCT GAT TGG TTA AAA GTT GAG	NF-Y
GAT CCT GGT AAG GGC CAA TCT GCT CAC	ßCT
GAT CCA AAC CAG CCA ATG AGA ACT GCT CCA	aP2a
GAT CCT GAG ACG TCC TAA GCC AGT GAG	ßB1
GCC TGA TTC CGT AGA GCC ACA CCC TGG TAA	ßCC
GAT CCG GGC GTG TCC ACC CTG CCT GGA	aCC
GAT CCT CTG CAC AGA TAA GGA CAA ACA	ßG215
GAT CCA CAC AGG ATA GAG AGG GCA GGA	ßG60
GAT CCT TTC TGA TAG GAA GGT TGA GCA	ßGIVS2
GAT CCT TGC AGA TAA ACA TTT TGC TAT CAA GAC TTG CA ßAG	

Figure 3.2 The Proximal Promoter.

The proximal promoter region is defined here as the region from the start site of transcription (0) to 100bp upstream (-100). This region, as discussed in the text is extensively characterised and elememts TATA (-30), CCAAT (-75) and CACC (-90) have been shown to be essential for initiation of transcription. In addition to these three elements, there is a sequence GATAGA, at -65bp relative to the start site, which shows close homology to the recognition site for NF-E1.

The Proximal Promoter.

-107
GTA GAG CCA CAC CCT GGT AAG GGC CAA TCT GCT CAC
ACA CAG GAT AGA GAG AGG GCA GGA GCC AGG GCA GAG
CAT ATA AGG TGA GGT AGG ATC AGT TGC TCC TGA CA
0

Figure 3.3A
The Binding of Factors Present in Crude Nuclear Protein Preparations to the Proximal Promoter Regions of the Mouse α-1 and β^{major} Globin Genes.

Lanes 1-4 Maxam-Gilbert sequencing reactions.
Lanes 6-12 Probe β730 (10ng) is incubated with non-specific competitor DNA, polydIdC (1μg) and 60 units of crude MEL cell nuclear extract made up to a total volume of 100μl with Storage Buffer and a final salt concentration of 50 mM.
Lanes 5, 13 As for lanes 6-12 except that no protein extract is added.
Lanes 14-18 As for lanes 6-12 except using STO crude nuclear extracts.
Lanes 20-24 As for lanes 6-12 except using α650 as probe.
Lane 19 As for lanes 20-24 except that no protein extract is added.
Specific competitor is added as a synthetic double stranded oligonucleotide containing the recognition site for the binding factor to be competed off the DNA probe. 100ng of specific competitor is added to the incubation reaction to give a final volume of 100μl. This concentration of competitor gives a 250 M excess of competitor sites over probe sites. Table 3 1 illustrates all of the double stranded oligonucleotide competitors used in this work.
Lanes 6, 14 No specific competitor added.
Lanes 7, 15, 21 βCT oligonucleotide competitor added.
Lanes 8,16,22 βCACC, Lane 9 βG215, Lanes 10,17 NF1, Lanes 11,18 NF1, Lanes 12, 23 αP2a, Lane 24 αCACC

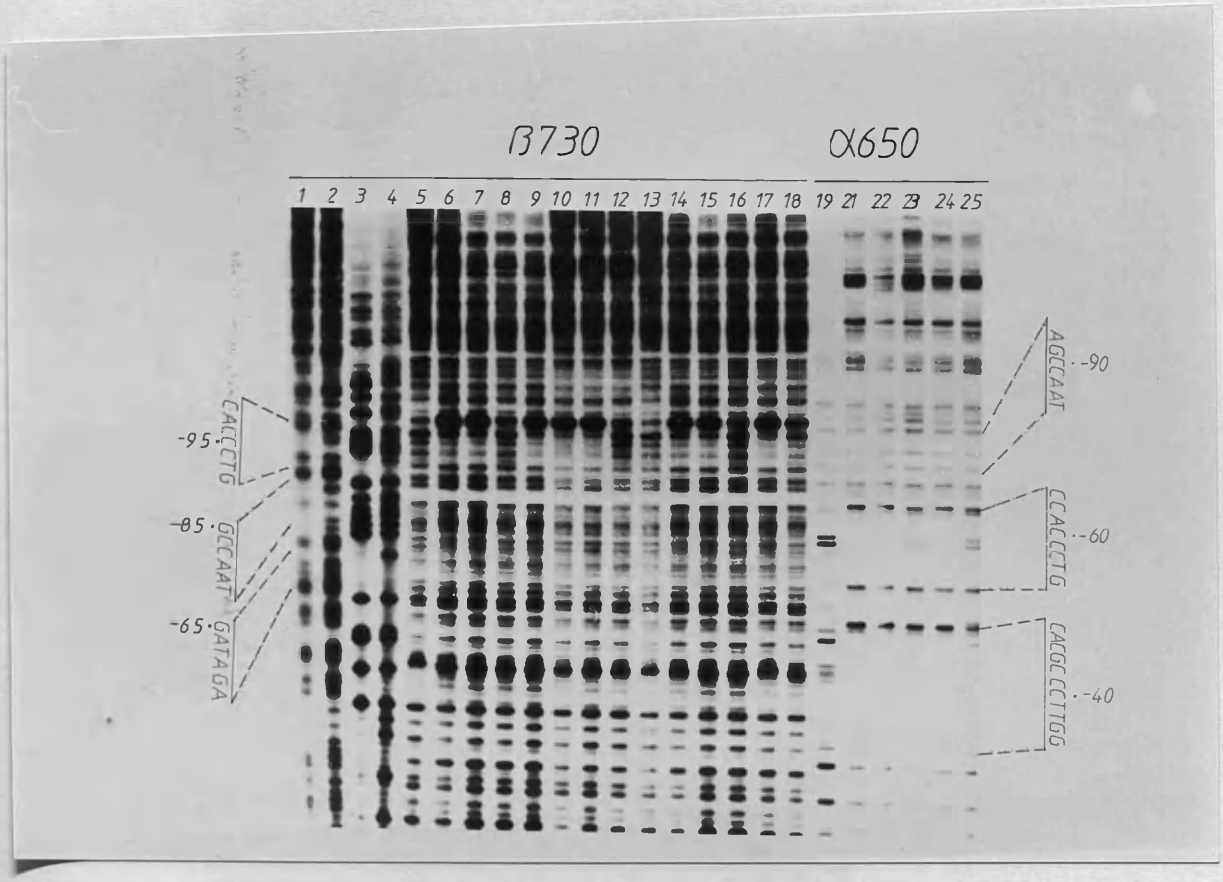


Figure 3.3B

The Binding of Factors Present in Partially Purified Nuclear Protein Preparations to the Proximal Promoter Region of the Mouse β^{major} Globin Gene.

Lane 1 Maxam-Gilbert G reaction.

Lane 2 Probe β730 (10ng) incubated with polydIdC (1μg) and made up to a total volume of 100μl with SB at a final salt concentration of 50 mM. No protein extract added.

Lane 3-4, 8-10 Probe β730 (10ng) incubated with polydIdC (1μg) and 60 units MEL partially purified nuclear extract and made up to a total volume of 100μl with SB and a final salt concentration of 50 mM.

Lane 5-7 Probe β730 (10ng) incubated with 1 μg of polydIdC and 60 units of mouse liver partially purified nuclear extract and made up to a final volume of 100μl with SB and a final salt concentration of 50 mM.

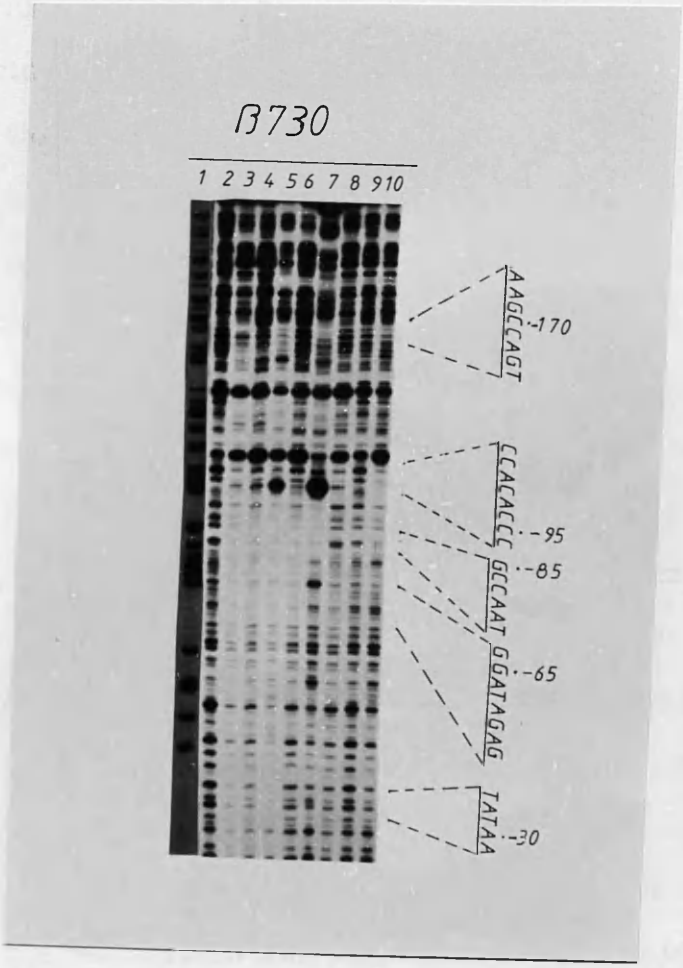
100ng of specific oligonucleotide competitor added as described for figure 3.3A as follows:

Lanes 6, 8 βCT

Lanes 7, 9 βCACC

Lane 4 NF-1

Lane 10 βG215



3.2 The Distal Promoter.

The distal region of the β^{major} globin gene promoter is defined as that region spanning from -100 to -250 bp 5' of the transcription initiation site, that is to the 5' boundary of the DHSS which has been mapped *in vivo* (Fig.3.4). This region is much less well characterised than the proximal region. Footprinting reactions were carried out as before but using probe β600.

Two strong footprints are observed in this region with crude nuclear extracts from different cell lines (fig. 3.5). Binding to the AGATAAGG motif at -212 bp is only observed with nuclear extracts from erythroid cell lines, either MEL or K562 (lanes 12-16 and 22-24). Binding to this site can be specifically competed out with 250 M excess of oligonucleotide competitor βG215 (lanes 7, 15 and 24), αG2 containing the mouse α-1 globin gene NF-E1 recognition site (lanes 8, 16) or β^{A}G containing the NF-E1 recognition site from the chicken β^{A} globin gene enhancer (lanes 9 and 17). This implies that the erythroid factor binding to the AGATAAG motif at -212 bp of the β^{major} globin gene promoter is the same as the factor that recognises and binds to the NF-E1 recognition site in the mouse α-1 globin gene promoter and the chicken β^{A} globin gene enhancer.

Binding to the GCCAA motif at -249 bp is observed in all of the nuclear extracts used (MEL, STO, K562 and HeLa) and can be specifically competed out with excess of oligonucleotide competitor containing either the recognition sequence for nuclear factor 1 (NF 1) (lanes 5, 14 and 21) or containing its own recognition sequence (lanes 4, 13, 18, 23 and 27). As it is not competed out by an excess of oligonucleotides containing the recognition site for either the mouse α1 or the β^{major}

globin gene CCAAT boxes (lanes 6, 19, 20 and 26), the factor binding to this sequence is a member of the NF-1 class of transcription factors but is not apparently related to the CCAAT box binding factors.

A sequence AAGCCAGT (-160 bp), called Box 1, which is conserved in a similar position in the human β globin gene and has been shown to be necessary for inducible expression from the human promoter in MEL cells (de Boer et al. (1988) is not seen to be bound by sequence-specific factors in this DNAase 1 footprinting system. However, as shown in figure 3.3B, this sequence is bound by factors present in partially purified MEL cell and liver nuclear extracts and gel retardation assays show that Box 1 binding activity is present in crude nuclear extracts from MEL cells (Chapter 4). This suggests that the DNAase 1 footprinting conditions used are not sensitive enough to detect it.

Figure 3.4 The Distal Promoter.

The distal promoter is defined here as the region immediatley upstream of the proximal promoter region spanning from -100bp to -250bp. This region has not been previously characterised either for protein binding sites or for function in initiating transcription.

The Distal Promoter.

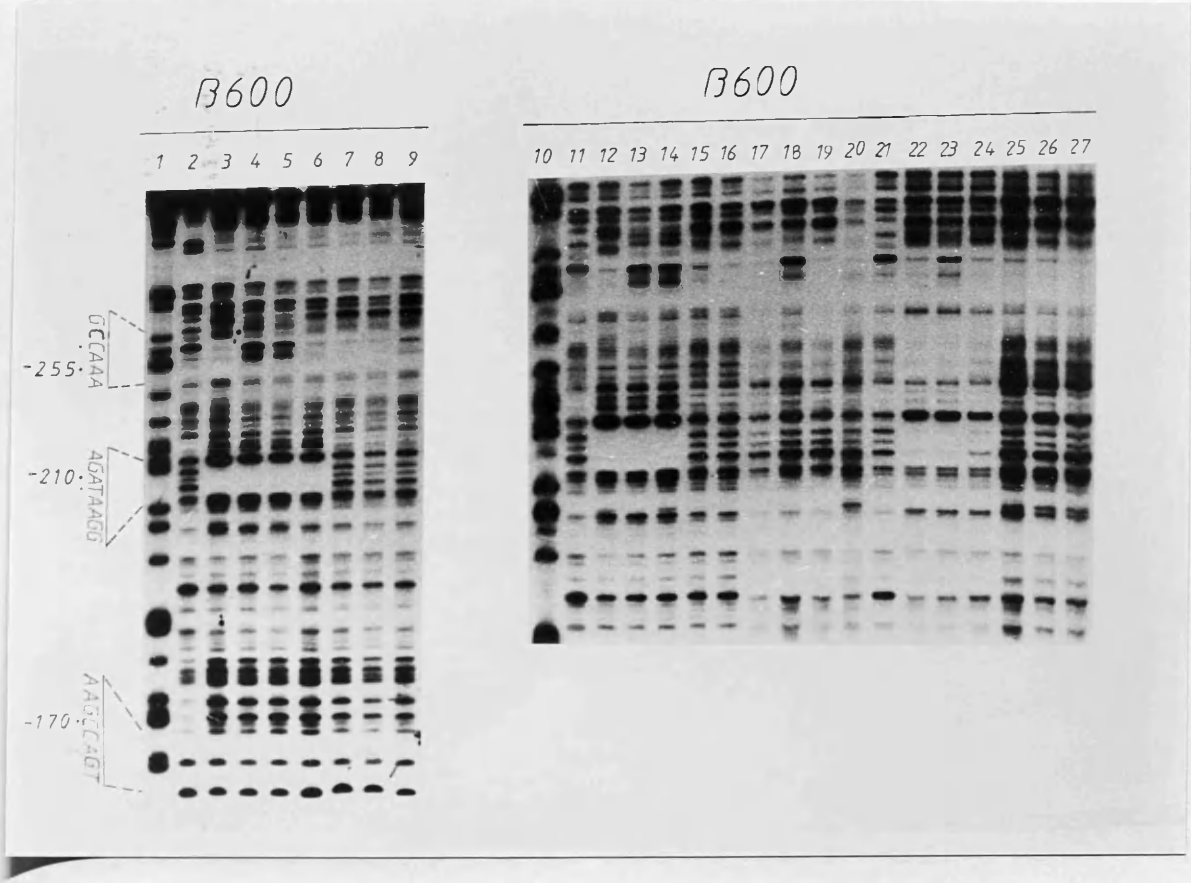
-244

CGT TCG CCA AAA AGG ATG CTT TAG AGA CAG TGT TCT
CTG CAC AGA TAA GGA CAA ACA TTA TTC AGA GGG AGT
CCC AGA GCT GAG ACG TCC TAA GCC AGT GAG TGG CAC
AGC ATG TCC AGG GAG AAA TAT CGC TTC GTC CTC ACC
GAA GCC TGC TTC CGT

-107

Figure 3.5 The Binding of Factors Present in Crude Nuclear Extracts to the Distal Promoter Region of the Mouse β^{major} Globin Gene.

Lanes 1, 10 Maxam-Gilbert G reaction.
Reactions were set up as described previously for figure 3.3.
60 units of crude nuclear extract was added to each reaction as follows:
Lanes 2, 11 Zero protein control.
Lanes 3-9, 12-16 MEL cell crude nuclear extract.
Lanes 17-21 STO crude nuclear extract.
Lanes 22-24 K562 crude nuclear extract.
Lanes 25-27 HeLa crude nuclear extract.
Specific oligonucleotide competitor was added (at a concentration of 100ng in 100µl of reaction mix, that is a 250 M excess) as follows:
Lanes 4,13,18,23,27 βNF1, Lanes 5,14,21 β^{A}NF1, Lanes 6,19,26 βCT, Lanes 7,15,24 βG215, Lanes 8,16 αG, Lanes 9 β^{A}G, Lanes 20 αCT



3.3 The Upstream Region.

In addition to promoter sequences which map immediately 5′ of a gene, other regulatory sequences which act at a distance have been identified both 5′ and 3′ of different mammalian genes. For this reason *in vitro* DNAase1 footprinting was used to determine whether these sites actually bind their cognate factors. However, given that this region lies outside the DHSS mapped *in vivo* and whose 5′ boundary is at -250 bp, it is unclear whether this region is functional *in vivo*. This is dealt with further in chapter 6. Footprinting of this region is carried out using probes β280 and β160 and is shown in figure 3.7.

A second high affinity NF-E1 recognition site is present at -520 bp 5′ of the initiation site (fig. 3.7A). This site binds a factor present in MEL and K562 crude nuclear extracts (lanes 3 and 7) and binding is competed out by excess of βG215 competitor (lanes 4 and 8). Binding at this site is not detected with STO, HeLa or J774.2 crude nuclear extracts.

An additional footprint is observed at -455 to -485 bp with STO crude nuclear extracts (lanes 5, 6) and the sequence bound contains a perfect TGGAATCA direct repeat. No footprint was observed in this region with MEL, K562, HeLa or J774.2 cell nuclear extracts and the sequence does not appear to be conserved in promoter/enhancer sequences of other globin or non-globin genes. For this reason, binding to this site was not followed up.

A site present at -550 bp contains the sequence AAGGCAGT which is recognised by the Box 1 factor (fig. 3.7B). Binding to this site is erythroid specific; binding is present with crude MEL nuclear extracts

(lanes 3-8, 16-19) but not with extracts from mouse kidney (lanes 9-11) or STO (lanes 12-13), although mouse liver extracts were not tested. Binding to this site by factor(s) present in MEL nuclear extracts can be competed out by 250 M excess of specific double-stranded oligonucleotide competitor βB1 (lanes 5, 10, 13, 18) but not by βCT (lanes 4, 17) or by βNF1 (lane 8). The figure also shows binding to the NF-E1 site at -520 bp, described in figure 3.7A.

In addition to a second high affinity NF-E1 binding site (-520 bp), the upstream region bears a Box 1 site (-550 bp) which binds nuclear factors present in crude nuclear extracts from MEL cells. This binding activity is not present in crude nuclear extracts from mouse STO or kidney. Figure 3.3A and 3.3B described DNAase 1 footprinting of the proximal region of the mouse β^{major} globin gene with crude and partially purified nuclear extracts respectively. It was shown that a footprint was obtained over the Box 1 sequence, AAGCCAGT (-169 bp), with partially purified MEL and liver nuclear extracts but not with crude extracts. The Box 1 sequence at -550 bp, AAGGCAGT, is however bound when footprinted with crude MEL extracts. This may be explained if the Box 1 recognition sequence at -550 bp has a higher affinity for Box 1 binding factor than the site at -160 bp. There is only a single base pair difference (G to C) between them but this may explain the difference in binding activity. This would need to be tested out by gel retardation assay and comparison of the ability of oligonucleotides containing these different sequences to compete for binding by the Box 1 recognition factor. Given that this region lies outwith the DHSS mapped *in vivo*, it is unclear whether these sites are active *in vivo*.

Figure 3.6 The Upstream Region.

This region maps from -250 at the 5' end of the distal promoter region to -745bp upstream of the initiation site.

The Upstream Region.

-730

GAG CTC AGG GTT TAC TTG AGA GAT CCT GAC TCA ACA

ATA AGG TAT AGA GCA ATC AAG GGC GAT TCT CTG AAG

GCA GTT ATT GAA CTC CTT GTA CAC TCT TCC CAC ACA

CAC GTG TTC AGA GAA CTT GAT AAA ACA CAT ATA TAC

AAG TAA GCC ACA CAA ACA TTC ACA TGA GAG AAG AAA

AAC AAG AGC AAA CTA AGT AAG ATG CAT TTT CTT ATC

AGG AAG TTT AGT TGA CAC CAG AAA GAA GTC ATA TTT

GGA ATC AAA ATG GAA TCA TCA TCA TGT ATG CTA AAG

ATG TTT TTT TCA CAT TCT TGA GCA ATG TGG ACA GAG

AAG GAG ATT CAT CCA TGC ACT CAA ACT GGG AAA CAA

AGA AAA GAA ATC CTC TTC TAA GCT TTG CTT CTC AAT

TTC TTA TTT GCA TAA TGA GAA AAA AAG GAA AAT TAA

TTT TAA CAC AAT TCA GTA GTT GAT TGA GCA AAT GCG

-264

Figure 3.7 Protein Binding Sites Upstream of the Mouse β^{major} Globin Gene.

Figure 3.7A The Proximal Upstream Region.

Footprinting using β280 as probe identifies protein binding sites in the region -346 to -550.

Lane 1 Maxam-Gilbert G reaction.

Lane 2 Zero Protein Control.

60 units of crude nuclear extracts were added to reactions as follows:

Lanes 3-4 MEL cell crude nuclear extract

Lanes 5-6 STO

Lanes 7-8 K562

Lanes 9-10 HeLa

Lanes 11-12 J774.2

All the above footprinting incubations are carried out as before in a total volume of 100µl with SB and a final salt concentration of 50 mM. Specific oligonucleotide competitor βG215 was added to reactions 4, 6, 8, 10 and 12 at a concentration of 100ng in 100µl, that is in 250 M excess.

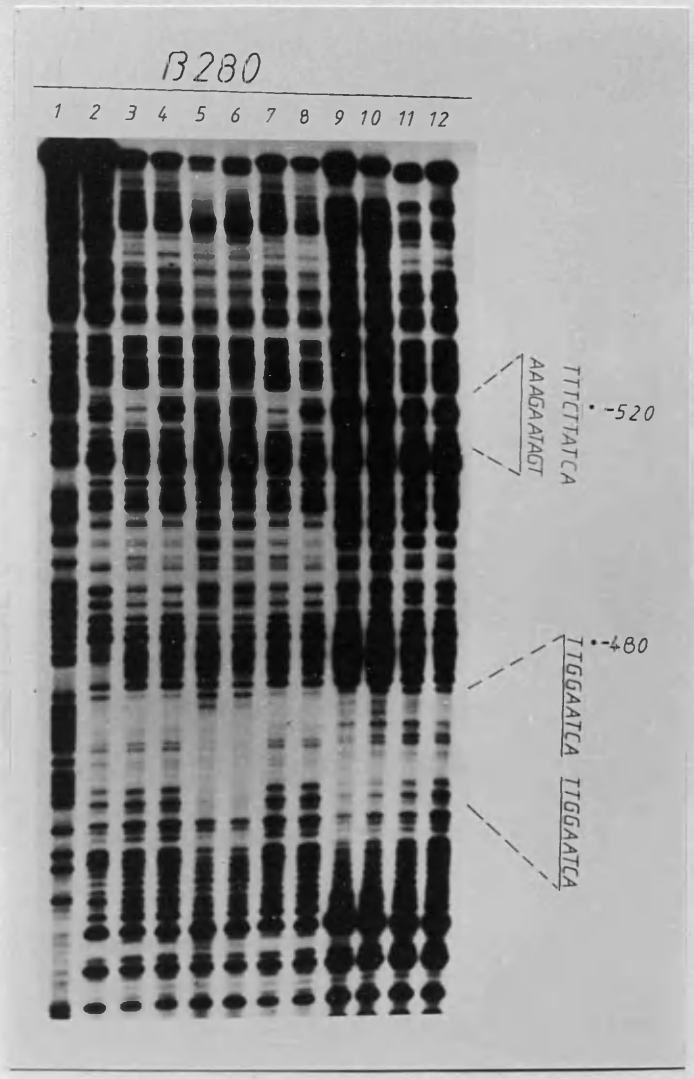


Figure 3.7B The Distal Upstream Region.

To identify protein binding sites in the distal part of the upstream region, probe βl60 was used in footprinting reactions. Footprinting reactions were carried out as described previously.

Lanes 1, 14 Maxam-Gilbert G reaction.
Lanes 2, 15 Zero Protein Control.

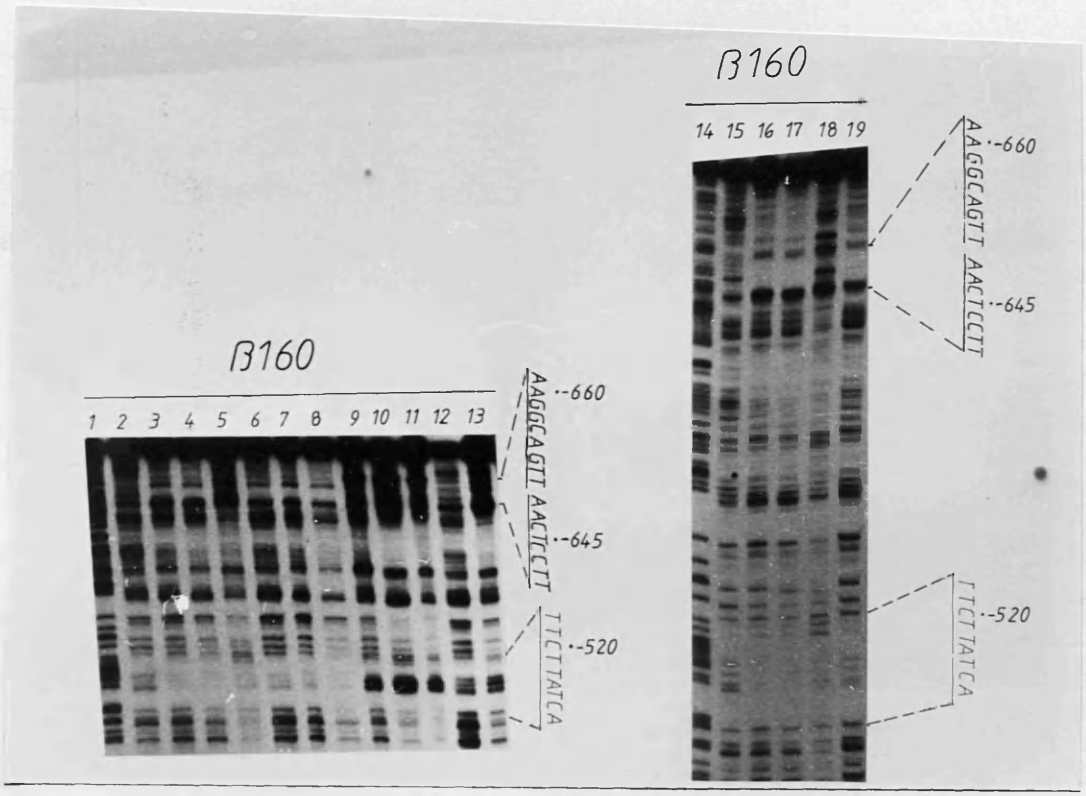
Crude nuclear extracts were added to reactions as follows:

Lanes 3-7, 15-19 60 units of MEL cell crude nuclear extract
Lanes 9-11 60 units of mouse kidney crude nuclear extract
Lanes 12-13 40 units of brain crude nuclear extract.

Footprinting reactions were carried out as described previously.

Specific double-stranded oligonucleotide competitor was added to reactions as follows:

Lanes 4, 7 βCT
Lanes 5, 10 13, 18 βB1
Lanes 6, 11, 19 βG215
Lane 7 αG
Lane 8 βNF1



of the chicken α^A β^A and the human γ and β globin genes were

During the course of my study, sequences in the 3′ flanking region of the chicken α^A β^A and the human γ and β globin genes were identified (Knezetic and Felsenfeld (1989); Choi et al. (1986); Bodine and Ley (1987); Kollias et al. (1987) which upregulate transcription from linked genes in an erythroid and developmental specific fashion. The chicken β enhancer sequences have since been characterised in terms of protein binding sites and it has been demonstrated that these enhancers as well as containing variable combinations of AP-1, CACC boxes and other previously characterised binding sites, possess multiple copies of NF-E1 binding sites. Figure 3.8A illustrates diagrammatically the binding sites present in β globin gene 3′ enhancers and their relative positions within that enhancer. The chicken β^A enhancer has been shown to act upon both the adult gene (β^A) and the embryonic gene (ε). No such enhancer sequence had been identified for the mouse β globin gene locus.

By analysing DNA sequence 3′ of the mouse β^{major} globin gene at a position equivalent to that where the chicken and human adult β globin gene 3′ enhancer sequences had been found, I was able to identify a number of potential binding sites for sequence specific DNA binding proteins (figure 3.8B). These sequences included several potential NF-E1 binding sites, an AP-1 site, an AP-2 site and a Box 1 binding site. I was interested firstly in determining whether these potential binding sites actually bound factors and secondly whether these sequences were involved in the transcriptional regulation of the gene. DNAase 1 footprinting was used to look for protein binding to these

sites.

A factor present in MEL cell nuclear extracts binds to the sequence AAGCCAAT present at +1705 relative to the start site of transcription (fig.3.9B). This sequence contains the recognition sequence for both CCAAT box binding factors, for βNF-1 and for Box 1 binding factor. However binding to this sequence is competed out by oligonucleotide competitor βB1 (lane 6) but not by βCT (lane 9) or by βNF1 (lane 7). This indicates that this site is bound by Box 1 binding factor, the same factor that binds to sites in the promoter and upstream region of the gene, and is not bound by CCAAT box or βNF-1 binding factors.

By DNAase 1 footprinting (fig.3.9A), I was able to show that only one of the NF-E1 binding sites at +2088, bound NF-E1 and that this binding could be competed out by excess of oligonucleotide competitor, βG215 containing the NF-E1 binding site from -212 bp in the β^{major} promoter, suggesting that the same factor binding to the promoter sequence is also binding to this 3′ NF-E1 recognition site. There are 3 additional potential NF-E1 recognition sites and their failure to bind NF-E1 in the DNAase 1 footprinting assay suggests that they may be similar to the NF-E1 site at -60 bp within the 5′ promoter region which also fails to bind NF-E1 in DNAase 1 footprinting assays with crude nuclear extracts. DNAase 1 footprinting of the proximal promoter with partially purified nuclear extracts shows a footprint over the NF-E1 binding site at -60 bp (fig. 3.3). However, footprinting of the 3′ region was not carried out with partially purified extracts and it is therefore not possible to make further comparisons between these sites. The potential AP-1 and AP-2 sites in the 3′ region also did not bind

although they are not perfect recognition sequences and this may account for their failure to bind.

The Box 1 site at +1705 bp binds factor which can be specifically competed out by oligonucleotide competitor containing the Box 1 recognition site from -160 bp in the promoter but not by oligonucleotide containing the β^{major} globin gene CCAAT box sequence.

Figure 3.8A Comparison of the Position of Regulatory Elements in the 3′ Flanking Region of Chicken β^A, Human β and Chicken α^A Globin Genes.

The figure shows the presence in the human and chicken adult β and the chicken adult α globin gene enhancers of multiple copies of the NF-E1 recognition sites. In addition, there are AP1, AP2, CACC and NF-1 elements. The human and chicken enhancers map to a positions 400 to 900bp 3′ of the polyadenylation signal of each gene. Looking in this region 3′ of the mouse adult β globin gene, the β^{major} globin gene (figure 3.8B), I was able to identify several copies of the NF-E1 recognition site, a Box 1 binding site and potential AP1 and AP2 binding sites.

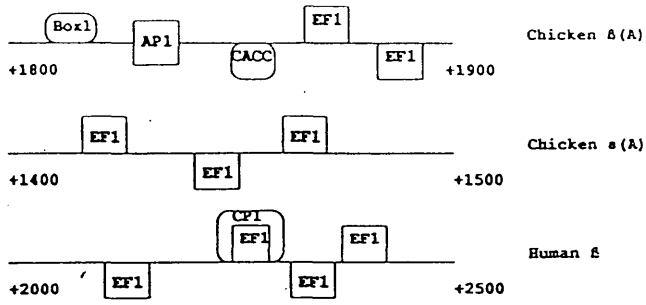


Figure 3.8B Comparison of the Sequence of the Mouse β^{major} Globin Gene 3′ Region and the Chicken β^{A} Globin Gene Enhancer.

The figure shows the sequence of the mouse β^{major} globin gene 3′ region (A) and the chicken β^{A} globin gene enhancer (B).

Sequences bearing homology to previously described protein binding sites are underlined for the mouse β^{major} globin gene 3′ region as are characterised binding sites in the chicken β^{A} globin gene enhancer.

The mouse β^{major} globin gene 3′ region contains a Box 1 recognition site (AAGCCAAT), five NF-E1 recognition sites (AGATAA, AATAAG, GATAT, AGATAGA & GATAT), a weak AP1 recognition site (TGACTG) and an AP2 site (ATGGGATTTC).

The chicken β^{A} globin gene enhancer contains an NF-1 like sequence (TGGCA), an AP1 site (TGACTC), a weak AP2 site (CTGGGC), a CACCC box and two NF-E1 sites (AGATAA, GATAG).

The organisation of these elements within the mouse β^{major} globin gene 3′ region is similar to that within the chicken β^{A} globin gene enhancer.

A.

+1693

CAGTTCTCAA GCCAATAATT TTTCTTTTGT AAAACTACCA

TTATTCTCTA AACTTTTCCC TCTGTGTTTA CCAAGCAACA

TTATTTATCT TTTCATAAAT CCTGTTGCCT TAGACAGCTT

CAGTAGCAAT AGAGGTAGGA TTAAGGAGAG AATAGAAGTT

CATACCATGC CTGCACAGTC AATAGTCACT ATGGGATTTC

AAATGGCACT TTGCCTGGGA CCTTTACACT TCACACCATA

CTCTGGCTTG AGTTAGGAGT TAAGAATGAG AGAAATATAA

TCTAGAGAGA ATAAGAATAT CTAGTTTTTA AGGCTCATTA

CTGGGGTCTT ATGAAATTTC CATAATACCC TGTAAATGGA

AGCATTTATT TTTTCAATAA ATCTATCTTG AATATCCAGT

+2102

B.

+1820

TCG ACC TCG AGG ATC ATT TCT GGC ATT CAG

CCT CCC CGA AAG GAG CTG ACT CAT GCT AGC

CCA GCA GCC AGC TGG GTG GGG GCA GGT TGC

AGA TAA ACA TTT TGC TAT CAA GAC TTG CAC

AGA CCT TGT TTC TAG +1920

Figure 3.9 DNAase 1 Footprinting of 3′ Sequences.

Probe β415 is generated as decribed in figure 3.1, using primers FFP7 and 8, such that one primer is 5′ end-labelled and the other is unlabelled.

Figure 3.9A illustrates footprinting with β415 generated using end-labelled FP8 and figure 3.9B illustrates footprinting with β415 generated using end-labelled FP7.

Figure3.9A

Lane 1 Zero protein control.

Crude nuclear extracts were added to reactions as follows:

Lanes 2-6 60 units MEL cell crude nuclear extract

Lanes 7-9 80 units STO nuclear extract.

Specific oligonucleotide competitor was added in 250 M excess as follows:

Lanes 2, 8 $\beta_{G}215$

Lane 5 β^{A}

Lane 3 β_{CT}

Lane 6, 9 β_{B1}

Footprinting with crude nuclear extracts from MEL cells generates a footprint st +2088 over the sequence AGATAGA (lanes 2-6) which is not obtained over the same sequence with STO nuclear extracts (lanes 7-9). Binding to this site is competed out by 250 fold excess of specific oligonucleotide competitor $\beta_{G}215$ (lane 3) and by 250 M excess of oligonucleotide competitor $\beta^{A}G$ (lane 5).

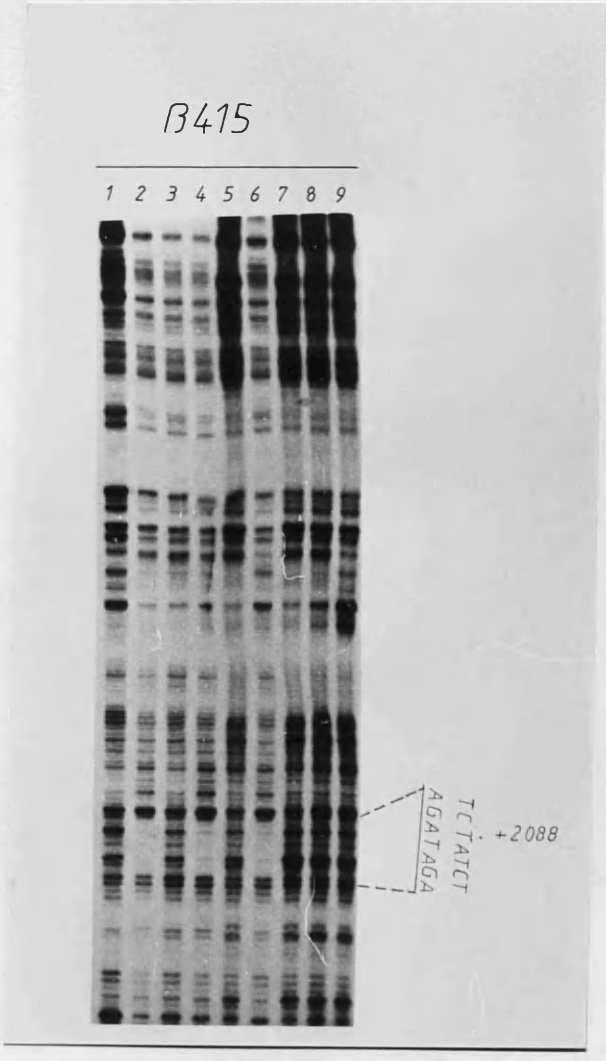


Figure 3.9B
Lane 1, 2 Maxam-Gilbert G, G+A Reactions.
Lane 3 Zero Protein Control.
Lanes 4-10 60 units of crude MEL nuclear extract
Specific oligonucleotide competitor was added in 250 M excess as follows:
Lane 5 βG215
Lane 6 βB1
Lane 7 βNF1
Lane 8 βCACC
Lane 9 βCT
Lane 10 NF-1

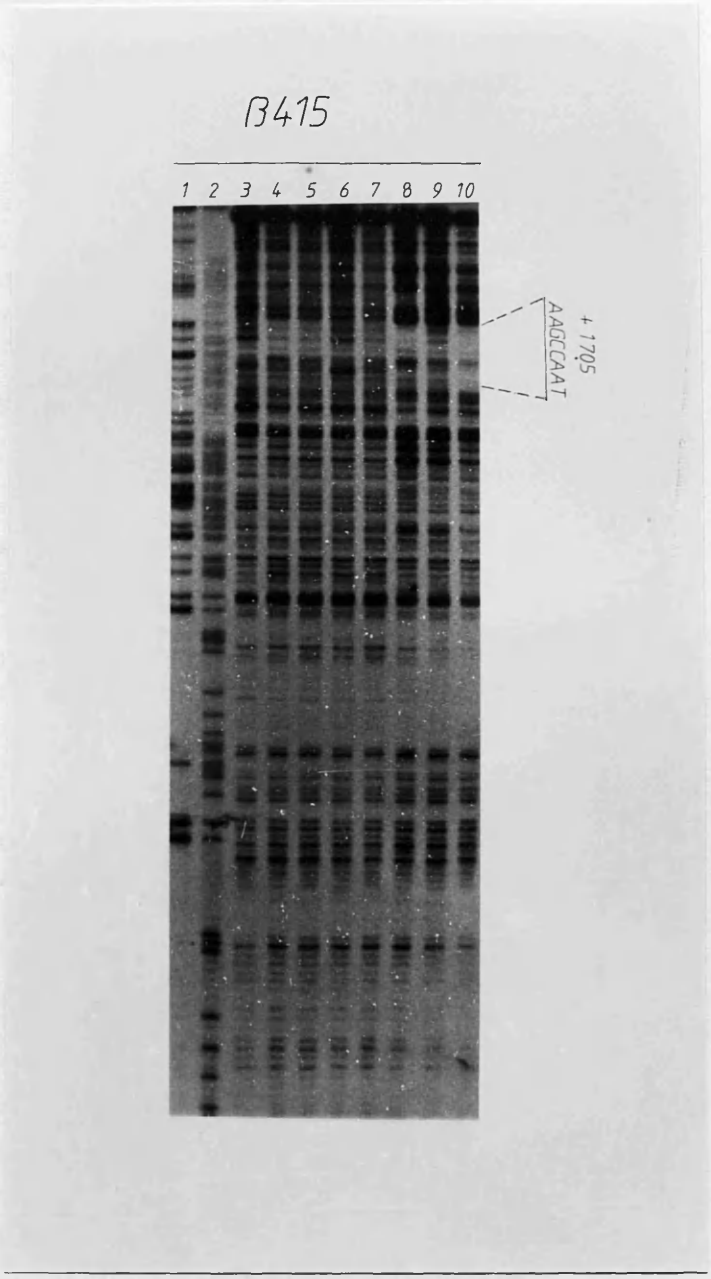


Figure 3.10 Summary of Footprinting Data.

The figure summarises the results of DNAase I footprinting of the mouse β^{major} globin gene 5′ and 3′ flanking regions.
Conserved sequences are underlined and those sequences actually protected from DNAase I cleavage by protein binding are boxed. A dotted box represents a footprint which is not obtained with MEL cell nuclear extracts.

The Proximal Promoter.

-107

GTA GAG CCA CAC CCT GGT AAG GGC CAA TCT GCT CAC
ACA CAG GAT AGA GAG AGG GCA GGA GCC AGG GCA GAG
CAT ATA AGG TGA GGT AGG ATC AGT TGC TCC TGA CA

0

The Distal Promoter.

-264

CGT TCG CCA AAA AGG ATG CTT TAG AGA CAG TGT TCT
CTG CAC AGA TAA GGA CAA ACA TTA TTC AGA GGG AGT
CCC AGA GCT GAG ACG TCC TAA GCC AGT GAG TGG CAC
AGC ATG TCC AGG GAG AAA TAT CGC TTC GTC CTC ACC
GAA GCC TGC TTC CGT

-107

The Upstream Region.

-730

GAG CTC AGG GTT TAC TTG AGA GAT CCT GAC TCA ACA
ATA AGG TAT AGA GCA ATC AAG GGC GAT TCT CTG AAG
GCA GTT ATT GAA CTC CTT GTA CAC TCT TCC CAC ACA
CAC GTG TTC AGA GAA CTT GAT AAA ACA CAT ATA TAC
AAG TAA GCC ACA CAA ACA TTC ACA TGA GAG AAG AAA
AAC AAG AGC AAA CTA AGT AAG ATG CAT TTT CTT ATC
AGG AAG TTT AGT TGA CAC CAG AAA GAA GTC ATA TTT
GGA ATC AAA ATG GAA TCA TCA TCA TGT ATG CTA AAG
ATG TTT TTT TCA CAT TCT TGA GCA ATG TGG ACA GAG
AAG GAG ATT CAT CCA TGC ACT CAA ACT GGG AAA CAA
AGA AAA GAA ATC CTC TTC TAA GCT TTG CTT CTC AAT
TTC TTA TTT GCA TAA TGA GAA AAA AAG GAA AAT TAA
TTT TAA CAC AAT TCA GTA GTT GAT TGA GCA AAT GCG

-264

The 3' Region.

+1693

CAGTTCTCAA GCCAATAAT TTTCTTTTGT AAAACTACCA
TTATTCTCTA AACTTTTCCC TCTGTGTTTA CCAAGCAACA
TTATTTATCT TTTCATAAAT CCTGTTGCCT TAGACAGCTT
CAGTAGCAAT AGAGGTAGGA TTAAGGAGAG AATAGAAGTT
CATACCATGC CTGCACAGTC AATAGTCACT ATGGGATTTC
AAATGGCACT TTGCCTGGGA CCTTTACACT TCACACCATA
CTCTGGCTTG AGTTAGGAGT TAAGAATGAG AGAAATATAA
TCTAGAGAGA ATAAGAATAT CTAGTTTTTA AGGCTCATTA
CTGGGGTCTT ATGAAATTTC CATAATACCC TGTAAATGGA
AGCATTTATT TTTTCAATAA ATCTATCTTG AATATCCAGT

+2102

CHAPTER 4 CHARACTERISATION OF FACTORS BINDING TO SITES IN THE PROMOTER AND 3′ FLANKING REGION OF THE MOUSE β^{MAJOR} GLOBIN GENE.

INTRODUCTION.

DNAase 1 footprinting of the 5′ and 3′ flanking sequence of the mouse β^{major} globin gene has revealed a limited number of sites which specifically bind factors present in nuclear extracts prepared from cell lines and mouse tissue. Gel retardation assays carried out as described in section 2.30 and using oligonucleotide probes end-labelled as described in section 2.14 were then used to further characterise binding activities. DNAase 1 footprinting with crude nuclear extracts had failed to show protein binding activity to DNA sequences such as the CCAAT box, the CACC box, the NF-E1 recognition site at -60 bp and the Box 1 binding site at -160 bp. Based on previous work on these sites in other genes, it was expected that these sites would bind factors. I investigated the binding to these sequences by *in vitro* gel retardation assay which is a more sensitive technique than DNAase 1 footprinting for determining protein binding to specific sequences. By analysing the mobility of shifted bands, competition patterns and tissue specificities, it was possible to distinguish binding activities.

4.1 Cross Competition Studies to Characterise Binding Activities.

DNAase 1 footprinting revealed that binding to certain sites within the promoter and 3′ region of the mouse β^{major} globin gene could be competed out by an excess of oligonucleotide containing the

recognition sequence from other sites. For example, binding to the GCCAA motif at -250 bp could be competed out by oligonucleotide containing the recognition site for NF1 from the Adenovirus origin of replication. Gel retardation assays are used here to further investigate competition between different protein binding sites in uninduced MEL cell crude nuclear extracts.

Figure 4.1A shows that four different NF-E1 binding sites differ in their ability to compete for binding to the high affinity NF-E1 site present at -212 bp in the mouse β^{major} globin gene promoter. Oligonucleotide βG215 is 5' end-labelled and incubated with 2 units of MEL crude nuclear extract in the presence or absence of a 1000 M excess of oligonucleotide competitors βG60, β^{A}G or βG^{IVS2}. The strongest competitor is βG215 (lane 2), followed by the NF-E1 binding site from the chicken β^{A} globin gene enhancer (lane 4). The NF-E1 site from the DHSS of the mouse β^{major} globin gene second intron is a less effective competitor still (lane 5) but the NF-E1 binding site at -60 bp (βG60) hardly competes at all (lane 6). The ability of these oligonucleotides to compete for binding to βG215 reflects their relative affinities for NF-E1 where βG215 is bound most strongly and βG60 weakly. βG60 will bind NF-E1 sufficiently well to observe a retarded complex in a gel retardation assay (Data not shown). β^{A}G and βG^{IVS2} bind NF-E1 with intermediate affinity.

Figure 4.1B shows the existence of two β CCAAT box binding activities, βCT1 and βCT2 (lane 1). βCT1 is competed out by the αP2a oligonucleotide containing the CCAAT box from the mouse α1 globin gene (lane 5) and βCT2 which is partially competed out by binding to βNF1. This suggests that one of the β^{major} globin gene CCAAT box binding

factors also binds the α1 globin gene CCAAT box and that the other CCAAT box binding factor binds to the βNF-1 binding site at -250 bp. Binding to the βCT oligonucletide by factors present in crude nuclear extracts in gel retardation assays illustrates that the failure to obtain a footprint over the CCAAT box using crude nuclear extracts is not due to the absence of these factors in the extracts but as suggested in Chapter 3, is probably due to the technical parameters of the DNAase 1 footprinting assay system.

Figure 4.1C shows that binding to the NF1 recognition sequence is partially competed out by binding to βNF1 (lane 3) while binding to βNF1 is completely competed out by binding to NF1 (figure 4.1D, lane 3). Binding to the NF-1 oligonucleotide generates a shifted complex which is smeared over a range of mobility unlike the βNF-1 complex which is a tight band. This is consistent with published results from gel retardation assays using radio-labelled NF-1 probes (Schneider et al. 1986). The NF-1 smear suggests that there is a family of factors recognising this sequence of which βNF-1 is a member. This may explain why the NF-1 oligonucleotide competes completely for binding to the βNF-1 recognition site while binding to the NF-1 probe is only partially competed out by βNF-1 oligonucleotide competitor. Binding to βNF1 is not competed out by βCT (fig.4.1D, lane 4). Binding to NF1 is not competed out by 1000 M excess of βCT, αP2a, BB1 or NF-Y oligonucleotides (fig.4.1C, lanes 4-7). These results suggest that binding by the NF-1 family of factors is distinct from the binding activity of the CCAAT box binding factors, βCT and αP2a or the Box 1 binding activity and the NF-Y binding activity.

Binding to the Box 1 recognition sequence is competed out by the

βCT oligonucleotide (fig.4.1E, lane 5) but not by βNF1 (lane 3), NF1 (lane 4), αP2a (lane 6) or NF-Y (lane 7). That binding to the Box 1 motif from position -160 bp of the promoter, can be competed out by βCT but not by αP2a suggests that the factor binding to the Box 1 motif may be homologous to the βCT2 but not the βCT1 binding activity (fig.4.1B). These competition studies reveal that several of the factors shown to bind elements within the β^{major} globin gene promoter cross compete and share aspects of their sequence recognition. This may have significance for functional interactions between elements and will be discussed in Chapter 8. The results of competition binding studies are summarised in Table 4.1 below.

Table 4.1 Competition for DNA Binding Activities.

	1000 M excess of competitor oligonucleotide								
	βG215	βG60	$\beta^{A}G$	$\beta^{IVS2}G$	βCT	αP2a	Box 1	NF-1	βNF-1
βG215	+++	+	++	++					
βG60	+++	+++	+++	+++					
$\beta^{A}G$	+++	+	+++	++					
$\beta^{IVS2}G$	+++	+	+++	+++					
βCT1					+++	+++		-	-
βCT2					+++	-		-	+
Box 1					+++	-	+++	-	-
NF-1					-	-	-	+++	++
βNF-1					-	-	-	+++	+++

- no competition; + weak competition;

++ moderate competition; +++ complete competition.

Figure 4.1 Competition Studies to Compare the Roles of Different DNA Binding Factors.

Figure 4.1A.

100pg of probe βG215 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of MEL cell crude nuclear extract in a total volume of 20 µl with SB and salt concentration of 50 mM.
Specific double stranded oligonucleotide competitor is added at 100ng per reaction, that is in 1000 M excess as follows:
Lane 1 No specific competitor.
Lane 2 βG215
Lane 3 No specific competitor.
Lane 4 $\beta^{A}G$, Lane 5 $\beta^{IVS2}G$, Lane 6 βG60

Figure 4.1B.

100pg of probe βCT is incubated with 6 µg of non-specific competitor polydIdC, 2 units of MEL cell crude nuclear extract in a total volume of 20 µl with SB and salt concentration of 50 mM.
Specific double stranded oligonucleotide competitor is added at 100ng per reaction, that is in 1000 M excess as follows:
Lane 1 No specific competitor.
Lane 2 βNF1
Lane 3 NF1
Lane 4 βCT
Lane 5 αP2a

Figure 4.1C.

100pg of probe NF-1 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of MEL cell crude nuclear extract in a total volume of 20 µl with SB and salt concentration of 50 mM.
Specific double stranded oligonucleotide competitor is added at 100ng per reaction, that is in 1000 M excess as follows:
Lane 1 No specific competitor.
Lane 2 NF1
Lane 3 βNF1
Lane 4 βCT
Lane 5 αP2a
Lane 6 βB1
Lane 7 NF-Y

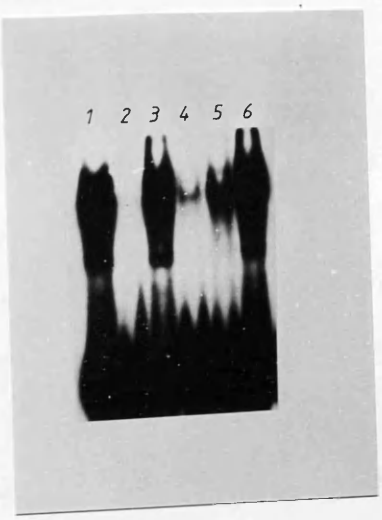
Figure 4.1D.

100pg of probe βNF1 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of MEL cell crude nuclear extract in a total volume of 20 µl with SB and salt concentration of 50 mM.
Specific double stranded oligonucleotide competitor is added at 100ng per reaction, that is in 1000 M excess as follows:
Lane 1 No specific competitor.
Lane 2 βNF1
Lane 3 NF1
Lane 4 βCT

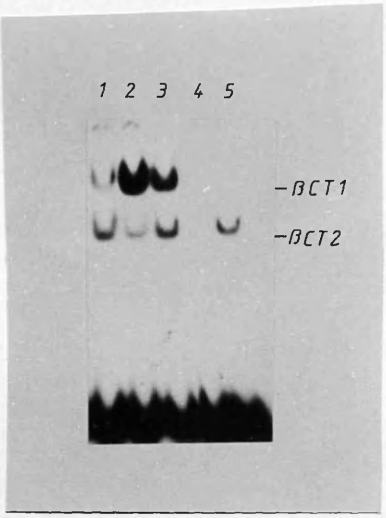
Figure 4.1E.

100pg of probe βB1 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of MEL cell crude nuclear extract in a total volume of 20 µl with SB and salt concentration of 50 mM.
Specific double stranded oligonucleotide competitor is added at 100ng per reaction, that is in 1000 M excess as follows:
Lane 1 No specific competitor.
Lane 2 βB1, Lane 3 βNF1, Lane 4 NF1
Lane 5 βCT, Lane 6 αP2a, Lane 7 NF-Y

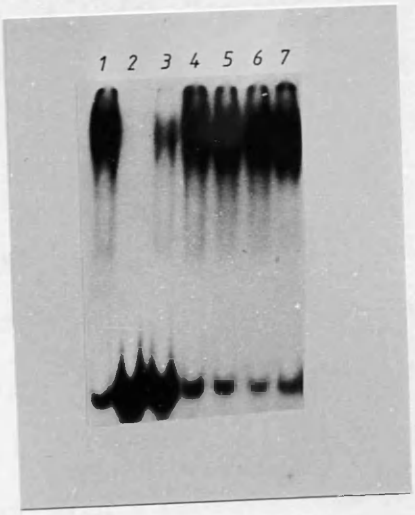
A.



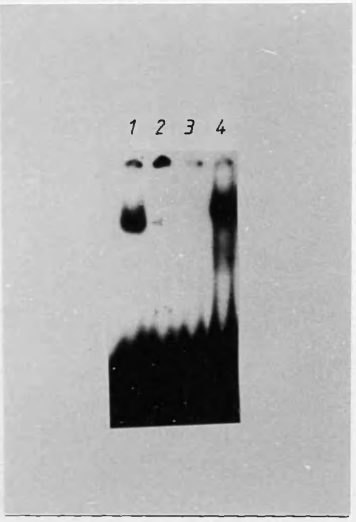
B.



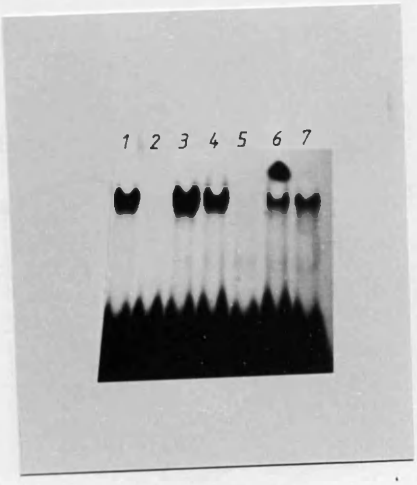
C.



D.



E.



4.2 An Analysis of the Tissue Specificity of Binding to Mouse β^{major} Globin Gene Promoter and 3' Sequences.

Different transcription factors have been shown to be developmentally and tissue-specifically regulated. To further characterise the factors which bind the mouse β^{major} globin gene flanking sequences, I looked at binding activity to oligonucleotide probes in crude nuclear extracts from a variety of mouse tissues and cell lines.

Crude nuclear extracts were prepared from several mouse tissues including liver and brain as well as from uninduced and induced MEL cells, a fibroblast cell line (STO), a macrophage cell line (J774.2) and an embryonal carcinoma cell line (F9) and two human cell lines K562 and HeLa. Figure 4.2 shows the results of gel retardation assays carried out in the presence of these different crude nuclear protein extracts.

Figure 4.2A illustrates that while CCAAT box binding activity to αP2a is present in both MEL (lane 1) and mouse brain (lane 4) nuclear extracts, binding activity to the NF-E1 recognition site is present only in MEL nuclear extracts (lane 1). NF-E1 binding can only be competed out with a 1000 M excess of βG215 competitor, (lane 2) whereas binding to αP2a, which contains the mouse α-1 globin gene CCAAT box recognition site, is only competed out by a 1000 M excess of competitor αP2a (lane 3).

Figure 4.2B shows that the activity which binds to the GCCAA motif at -250 bp is present in MEL, STO, J774.2, K562 and Hela extracts and is also present in mouse brain and liver extracts. This is consistent with the known tissue distribution of NF1.

Binding to oligonucleotide probe βCC containing the CACC recognition site from -90 bp of the mouse β^{major} globin gene is shown in figure 4.2C. There are two binding activities, βCC1 the slower mobility complex and βCC2 the faster mobility complex. βCC1 binding activity is present in all the nuclear extracts tested but βCC2 binding activity is present only in induced MEL, K562 and J774.2 crude nuclear extracts. Jackson et al. (1989) describe two different chicken erythrocyte nuclear binding activities which recognise the β^{A} globin gene CACC box: the slower mobility complex is present in chicken erythrocyte extracts at different stages of development whereas the faster mobility complex becomes less abundant in extracts from progressively later stages in development. They suggest that the slower mobility complex is an Sp1 bound complex suggesting that the slower mobility complex present in all mouse cell types looked at contains Sp1. Competition studies with Sp1 oligonucleotide competitor have not yet been carried out. The tissue specificity of the higher mobility complex present only in extracts of haematopoietic origin suggests that it may play a role in regulation of differentiation.

Figure 4.2D shows the presence of two CCAAT binding activities that differ in their tissue-specific activity. βCT1 is present in all crude nuclear extracts looked at except F9 nuclear extracts. βCT2 is absent in induced MEL cell, mouse brain and liver nuclear extracts but present in all other extracts looked at. The absence of βCT2 binding activity in induced MEL cell extracts suggests that it may play a role in the induction of β^{major} globin gene expression during differentiation. If it is an inhibitory binding factor, possibly acting by preventing binding of βCT1 and preventing the formation of a

functional transcription complex, then its absence in induced cells would lead to derepression of transcription. It would be acting in an analogous fashion to the CDP, (CCAAT Displacement Protein) involved in the regulation of human $^{A}\gamma$ globin gene (Superti-Furga et al (1989) and sea urchin histone H2B gene (Barberis et al (1987).

βCT2 is present in the embryonal cell line, F9, while βCT1 is absent suggesting a role for βCT2 in repression of globin gene expression during development.

Box 1 binding activity is present in all tissues looked at except induced MEL cells, F9 and J774.2 cells (figure 4.2E). Box 1 activity is present in uninduced MEL cells but absent in induced MEL cells suggesting a role for this factor in regulating the induction of β^{major} globin gene expression during differentiation. The significance of this pattern of binding activity will be further discussed in chapters 6 and 7 when the functional role of the Box 1 sequence is assessed.

Table 4.2 summarises the results described here.

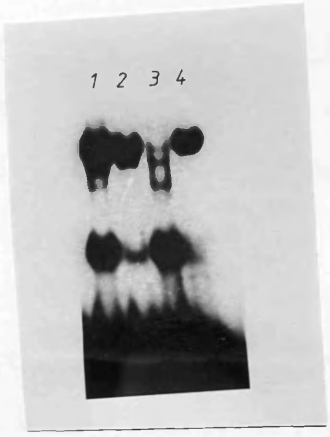
Table 4.2 TISSUE SPECIFICITY OF DNA BINDING.

	Fr^{-}	Fr^{+}	Br.	Liv.	K562	HeLa	STO	F9	J774.2
βG215	+	+	-	-	-	-	-	-	-
βNF1	+	?	+	+	+	+	+	?	+
βCACC	+	+	+	-	+	+	+	?	+
βCT1	+	+	+	+	+	+	?	-	+
βCT2	+	-	-	-	+	+	?	+	+
βB1	+	-	+	+	+	+	?	-	-

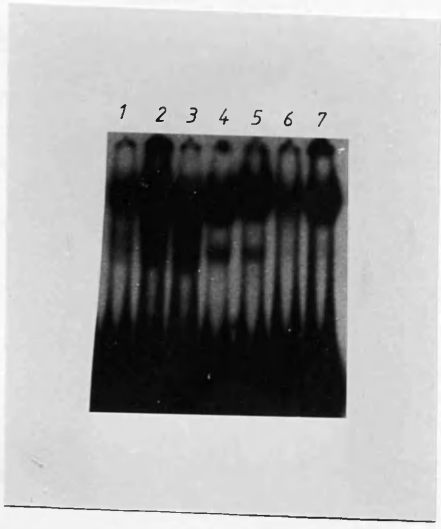
Figure 4.2 Tissue Specific Binding Activities .
Binding assays were carried out as desribed previously in section 2.35.
Figure 4.2A
Lane 1 100pg αP2a probe and 100pg βG215 probe are coincubated with 2 units of MEL nuclear extract, 6 µg non-specific competitor polydIdC, to a total volume of 20µl with SB and a final salt concentration of 50 mM.
Lane 2 As for lane 1, except 100ng of specific oligonucleotide competitor is added.
Lane 3 As for lane 1, except that 100ng of specific oligonucleotide competitor αP2a is added.
Lane 4 As for lane 1, except that 2 units of mouse brain crude nuclear extract is added.
Figure 4.2B
100pg of probe βNF1 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of crude nuclear extract in a total volume of 20µl with SB and a final salt concentration of 50 mM.
2 units of crude nuclear extract are added as follows:
Lane 1 Uninduced MEL cell nuclear extract
Lane 2 Mouse brain, Lane 3 Mouse liver, Lane 4 K562
Lane 5 HeLa, Lane 6 STO, Lane 7 J774.2
Figure 4.2C
100pg of probe βCACC is incubated with 6 µg of non-specific competitor polydIdC, 2 units of crude nuclear extract as described below, in a final volume of 20 µl with SB and a final salt concentration of 50 mM.
Lane 1 Uninduced MEL cell nuclear extract
Lane 2 Induced MEL cell
Lane 3 K562, Lane 4 HeLa, Lane 5 Mouse brain
Lane 6 Mouse liver, Lane 7 STO, Lane 8 J774.2

Figure 4.2D
100pg of probe βCT is incubated with 6µg of non-specific competitor polydIdC, 2 units of crude nuclear extract as described below, in a total volume of 20µl of SB and a final salt concentration of 50 mM.
Lane 1 uninduced MEL cell nuclear extract
Lane 2 induced MEL, Lane 3 K562 , Lane 4 HeLa, Lane 5 Mouse brain,
Lane 6 Mouse liver, Lane 7 STO, Lane 8 J774.2
Figure 4.2E
100pg of probe βB1 is incubated with 6 µg of non-specific competitor polydIdC, 2 units of crude nuclear extract as described below, in a total volume of 20 µl with SB and a final salt concentration of 50 mM.
Lane 1 uninduced MEL cell crude nuclear extract
Lane 2 induced MEL cell , Lane 3 J774.2, Lane 4 Mouse brain
Lane 5 Mouse liver, Lane 6 F9 cell, Lane 7 K562, Lane 8 HeLa

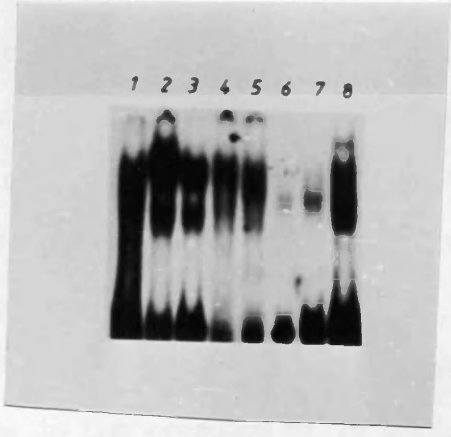
A.



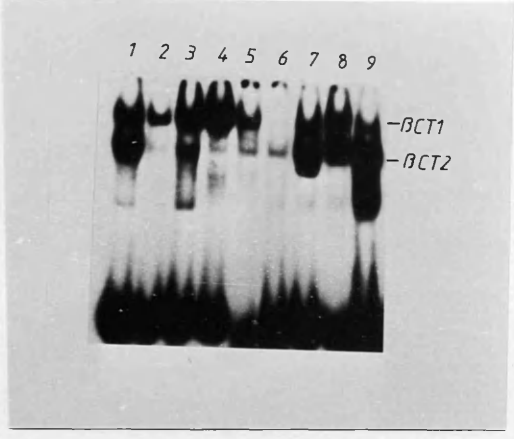
B.



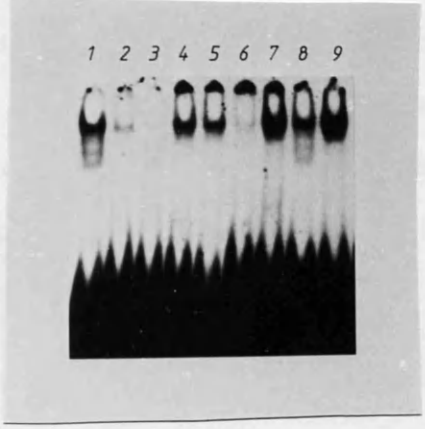
C.



D.



E.



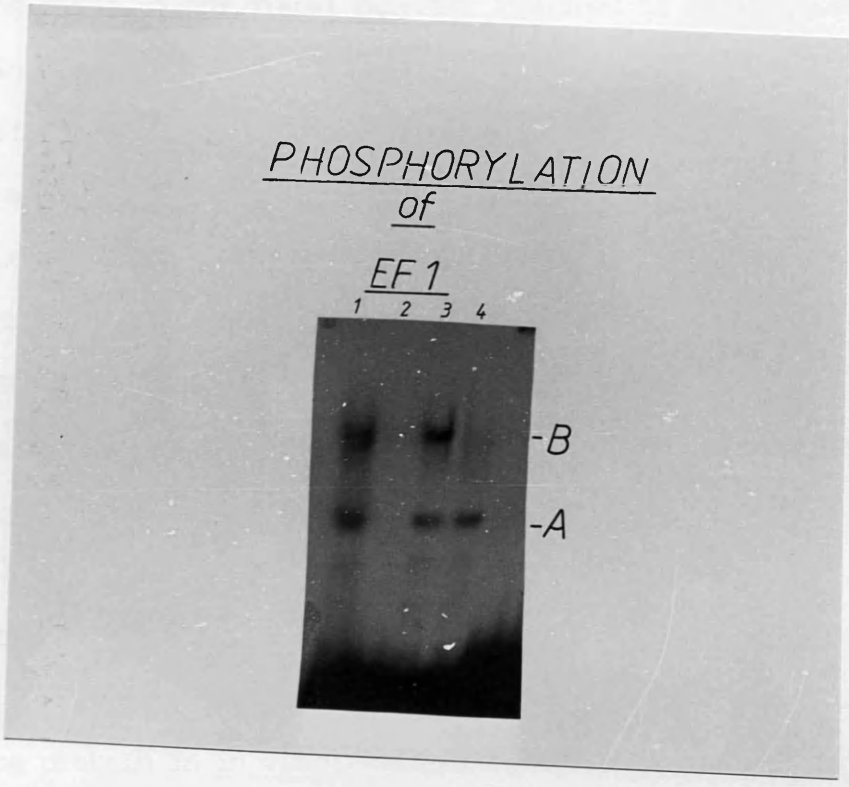
4.3 The Phosphorylation State of Erythroid Factor 1.

The electrophoretic mobility of the NF-E1 nucleoprotein complex in gel retardation assays varied depending on whether it was performed with nuclear extracts prepared from MEL cells in the presence or absence of phosphatase inhibitors, (sodium orthovanadate and β-glycerophosphate). Compare figure 4.2A where the MEL nuclear extracts were prepared in the absence of phosphatase inhibitors with figure 4.1A where the MEL extracts were prepared in the presence of phosphatase inhibitors. In figure 4.1 A, there are two shifted bands, one of high mobility (A) and one of lower mobility (B). In figure 4.2A, only the high mobility complex (A) is present. This suggests a role for phosphorylation in altering the binding activity and therefore possibly the function, of NF-E1. The role of phosphorylation in modulating the binding activity of yeast ATF (Jones and Jones (1989) had been investigated *in vitro* using gel retardation assays and suggested a method by which to investigate the role of phosphorylation in the binding activity of NF-E1.

Crude nuclear extracts were therefore incubated at 37^{o}C for 15 minutes in the presence of 10 units of calf intestinal phosphatase and the dephosphorylation reaction arrested by the addition of 16 mM sodium orthovanadate. Control incubations were carried out in the absence of phosphatase. The nuclear extracts were then used in gel retardation assays as described in section 2.30. Two retarded complexes which differ in mobility are observed in gel retardation assays carried out with nuclear extracts prepared in the presence of phosphatase inhibitors (lane 1). Both complexes are specifically competed out by excess of the β215 oligonucleotide competitor (lane 2). Extracts which

have been incubated at 37°C for 15 minutes in the absence of calf intestinal phosphatase as a control for degradation still show binding activity for both forms of the factor (lane 3). When extract which has been incubated for 15 minutes with CIP is used in the assay only the high mobility complex is observed (Figure 4.3). The slow mobility complex disappears. However there is not a subsequent increase in the amount of the high mobility complex formed as might be expected if the slow mobility complex formed by the dimerisation of the low molecular weight form of the protein. It is possible that the high molecular weight complex is a heterodimer consisting of NF-E1 and some other unidentified protein and these results suggest that phosphorylation is necessary for the formation of this complex.

Figure 4.3 Phosphorylation of EF1.
Lane 1 Untreated extract prepared in the presence of phosphatase inhibitors.
Lane 2 1000 M excess of specific oligonucleotide competitor βG215.
Lane 3 Control extract incubated at 37°C for 15 minutes in the absence of phosphatase
Lane 4 Extract treated with calf intestinal phosphatase for 15 minutes at 37°C.



CHAPTER 5 GENERATION OF MUTANT PROMOTER CONSTRUCTS.

INTRODUCTION.

Previous work has only analysed the functional importance of proximal promoter elements and the hitherto unidentified protein binding sites in the distal promoter remained to be assessed for their functional significance. Methods for studying the relative importance of different elements have in the past included taking individual elements out of context and testing their function in heterologous promoters or in isolation in artificial promoter constructs. These methods are less than satisfactory since it is now clear from work on the SV40 enhancer and other regulatory sequences that the context of the element determines its overall function. In many cases no one element is sufficient to ensure correct regulation of transcription. Where possible it is preferable to analyse the function of individual promoter elements in the context of their own promoters or enhancers.

I set out to analyse whether the elements I had identified as binding protein in in vitro assays had a functional role in modulating transcription from the mouse β^{major} globin gene promoter. In particular I was interested in the distal promoter elements which I had identified and characterised. As described in section 2.19, I used a PCR-mediated site-directed mutagenesis method to introduce carefully selected point mutations. Figure 5.1 summarises the method used.

All the mutant promoter constructs are based on the wild-type promoter construct, pβ346GH, which contains 5' flanking sequence of the mouse β^{major} globin gene spanning from the transcription start site to 346 bp upstream. Figure 5.2 illustrates the mutations introduced which

were predicted to abolish protein binding and shown by gel retardation assay and DNAase 1 footprinting to do so and figure 5.3 illustrates the PCR products obtained by the method described in figure 5.1. Figure 5.4 illustrates schematically the organisation of elements in the wild-type and mutant promoters.

Figure 5.1 PCR-Mediated Site Directed Mutagenesis.

All the mutant promoter constructs are based on the wild-type promoter construct, pβ346GH. Figure 5.2 illustrates the mutations introduced which were predicted to abolish protein binding. The figure below illustrates the mutant and linker oligonucleotide primers used in the method described in section 2.19 to generate mutant promoter constructs.

PCR SITE DIRECTED MUTAGENESIS OLIGONUCLEOTIDE PRIMERS.

LINKER OLIGONUCLEOTIDES

ACG ACG GCC AGT GCC AAG CTT TGC TTC TCA	LK5'U
AGT GGT TCG GGG AGT TGG GCC TTG GGA TCC	LK3'L

MUTANT OLIGONUCLEOTIDES.

TGT TCT CTG CAC AGC TCC GGA CAA ACA TTA	B2(A)
GAA TAA TGT TTG TCC GGA GCT GTG CAG AGA	B2(B)
GAG CTG AGA CGT CCT CCT AAA GTG AGT	B3(A)
GTG CCA CTC ACT TTA GGA GGA CGT CTC	B3(B)
TGA GCA AAT GCG TTC GCC GGT TAG GAT GCT	B1(A)
TAA AGC ATC CTA ACC GGC GAA CGC ATT TGC	B1(B)
CCT GAT TCC GTA GAG CTG CAG GAA GGT	B4(A)
ACC TTC CTG CAG CTC TAC GGA ATC AGG	B4(B)
TGC TCA CAC AGC TGC AGG AGG GCA GGA	B5(A)
TCC TGC CCT CCT GCA GCT GTG TGA GCA	B5(B)

Figure 5.2 Mutant β^{major} Globin Gene Promoter Constructs.

The figure illustrates the mutations introduced (as described in figure 5.1) into the mouse β^{major} globin gene promoter.
The recognition sites for the binding factors identified and characterised in chapters 3 and 4 were mutated as described below:

β1	GCCAA	GCCGG
β2	GATAAG	GCGCCG
β3	AAGCCAGT	CCTAAAGT
β4	CACCCTG	CTGCAGG
β5	AGGATAGA	CTGCAGGA

Figure 5.3 PCR Products.

The figure illustrates the predicted size of PCR products from PCR reactions 1 and 2, and those actually obtained. The main products of each PCR reaction are of the expected size. These fragments are gel purified and used as templates for PCR reaction 3.

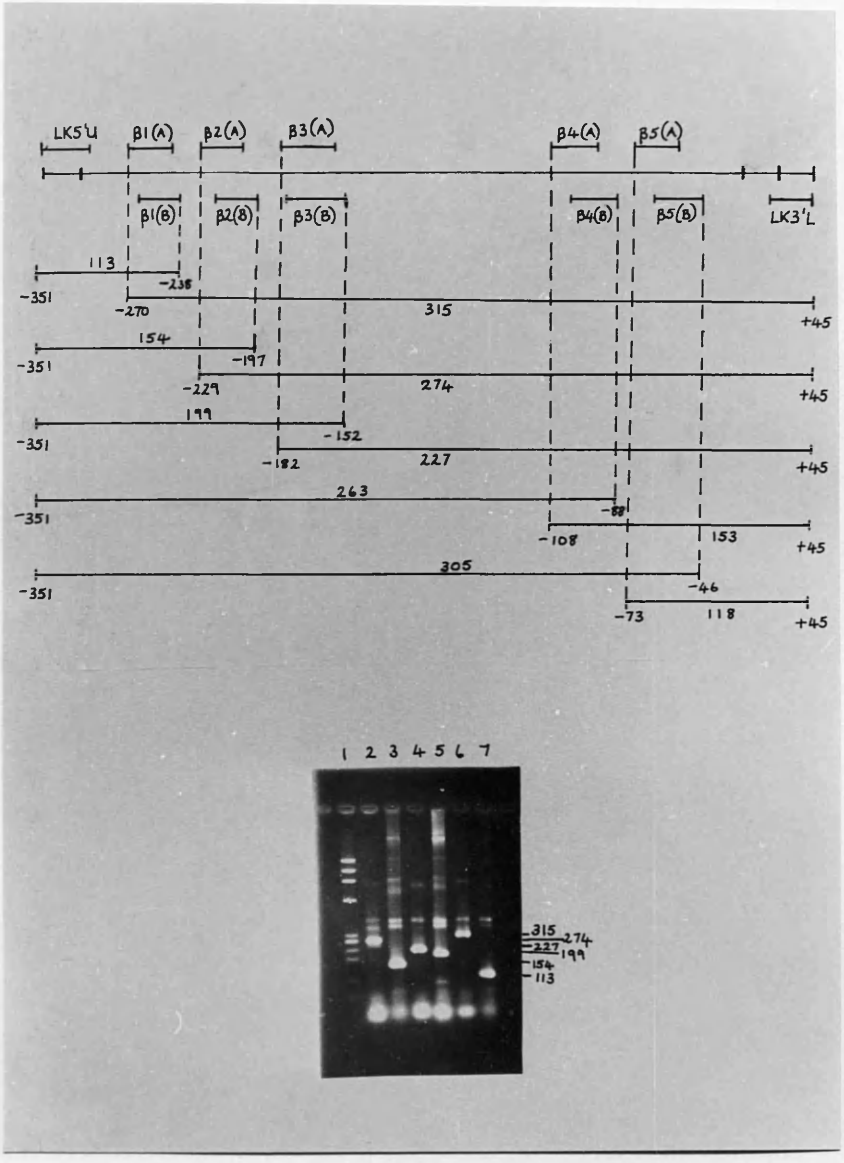
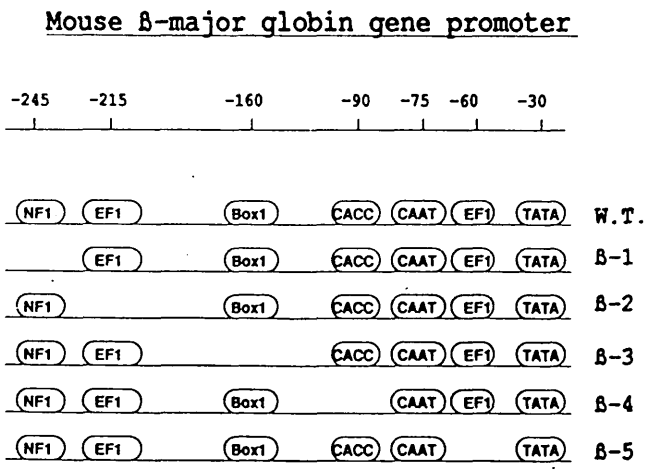


Figure 5.4 Mutant Promoter Constructs.
The figure illustrates the organisation of elements in the mutant promoters compared to the wild-type promoter following PCR mediated mutagenesis.



CHAPTER 6 FUNCTIONAL ANALYSIS OF THE MOUSE β^{MAJOR} GLOBIN GENE PROMOTER AND 3' SEQUENCES IN TRANSIENT TRANSFECTION ASSAYS.

INTRODUCTION.

Each of the systems which have been developed to analyse the regulatory role of DNA sequences in the initiation of transcription, have their own specific merits and disadvantages. This chapter describes the results of work using a transient transfection assay system to analyse the regulatory role of the protein-binding sites in the mouse β^{major} globin gene promoter identified in Chapter 3 and further characterised in Chapter 4. The test constructs (Figure 2.3) were introduced into MEL or other non-erythroid cell lines by calcium phosphate precipitation as described in section 2.21. Expression from the test construct is assayed 24-48 hours after transfection as described in section 2.23 or 2.26. The benefit of transient transfection assay systems such as this is the speed with which a result can be obtained (5 days), unlike stable transfection assay systems which take 3-4 weeks to set up. The main disadvantage of this system is the inability to look at changes in expression over longer culture periods. In particular, in order to look at changes in expression as the MEL cells are induced to differentiate it is necessary to generate stable transfectants as described in Chapter 7.

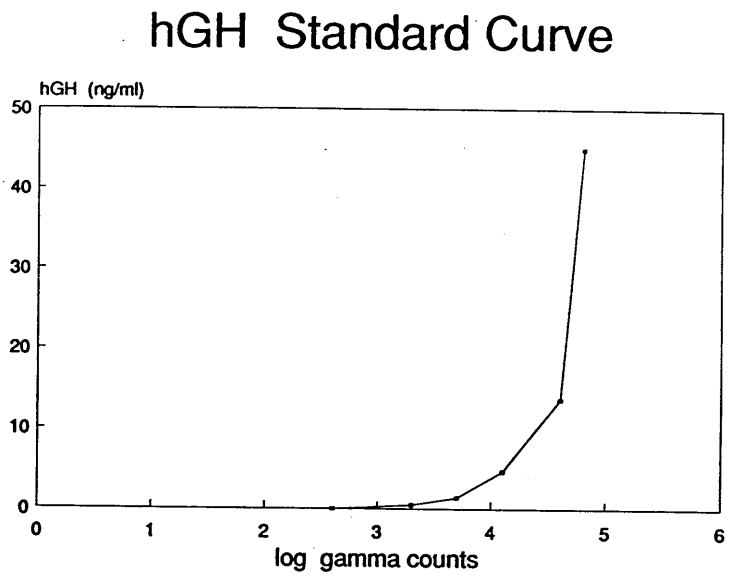
It is difficult to determine how closely results obtained in transient transfection assays approximate to the regulation of the same gene in vivo. By introducing a large copy number of test plasmid into cells, artificial conditions are created in which the ratio of protein: DNA binding site is reduced below that observed in vivo. Also, as

extra-chromosomal DNA, test plasmid sequences are not under the *cis* regulatory effects of flanking chromosomal DNA and are not subject to the regulatory effects of changing chromatin conformation. However it is possible to compare the relative importance of different DNA sequences in these assay systems and this is what is attempted here.

6.1 Characterisation of the Assay System.

The human growth hormone (hGH) reporter gene and the hGH radioimmunoassay were used for functional analysis of promoter/enhancer sequences (as discussed in section 2.23) because of the sensitivity and facility of the assay. The constructs used in the assays described here are listed in Table 6.1. The amount of hGH secreted by cells transfected by the appropriate test construct is measured indirectly as the amount of ^{125}I labelled hGH antibody bound covalently to an avidin coated glass bead via hGH and a biotin-linked hGH antibody. The amount of γ radiation emitted by ^{125}I bound to the glass bead can be measured. The relationship between concentration of hGH and γ counts is exponential as illustrated in figure 6.1.

Figure 6.1 hGH Standard Curve.



The figure illustrates the relationship between γ counts/minute measured in the assay and the amount of hGH activity that these counts represent for data presented in this chapter.

Known quantities of hGH were incubated under the assay conditions described in 2.28 and the amount of ^{125}I bound indirectly to the glass bead was measured in a Beckman 20 Gamma Counter. By plotting hGH values against γcounts, a standard curve can be obtained as above from which the amount of hGH present in experimental samples can be determined from the amount of γemission measured.

Table 6.1 Test Constructs used in Expression Assays.
The table illustrates promoter and enhancer constructs used in expression studies described in chapters 6 and 7 and based on plasmids described in figure 2.1. The table describes the reporter gene, the promoter, the promoter mutation and the enhancer present in each construct. Mutant promoter constructs were generated as described in chapter 5.

TEST CONSTRUCTS.

CONSTRUCT	REPORTER GENE	PROMOTER	MUTATION	ENHANCER
p0GH	hGH	-	-	-
pTKGH	hGH	HSV-TK	-	-
pß346GH	hGH	Mouse ßmajor globin gene 0 t0-346 bp	-	-
pß1GH	hGH	"	ßNF1	-
pß2GH	hGH	"	G215	-
pß3GH	hGH	"	Box 1	-
pß4GH	hGH	"	CACC	-
pß5GH	hGH	"	G60	-
pß(A)GH	hGH	"	-	Chicken ß(A) globin gene enhancer
pß(M)GH	hGH	"	-	Mouse ß(M) globin gene 3' sequences
pß745GH	hGH	Mouse ß(M) globin gene 0 to -745bp	-	-
paCAT0	CAT	Mouse a-1 globin gene 0 to -52bp	-	-
pCAT1	CAT	Mouse a-1 globin gene 0 to -700bp	-	-

6.2 The Relationship Between hGH Production and Concentration of Transfected DNA.

The efficiency with which a given cell line can be transfected with plasmid DNA is dependent upon a number of parameters including concentration of DNA, transfection method, pH of medium, CO_2 concentration of medium, cell density etc. (Chen and Okayama (1987).

A calcium phosphate precipitation method has been established for the efficient transfection of MEL cells (Plumb et al. (1989), which involves the transfection of 20µg of DNA into $1\text{-}2x10^6$ cells as described in 2.21.

I was interested to determine whether there was a concentration of test plasmid as a proportion of the total 20µg of transfected DNA which gave an optimum expression level. Transfected DNA was made up to a total amount of 20µg with pIC20R carrier plasmid, keeping the amount of reference plasmid, pHSV-βgal., constant at 5µg. As described in section 2.21, the pHSV-βgal. reference plasmid is cotransfected with the test plasmid to allow the assay results to be standardised for transfection efficiencies.

Figure 6.2 illustrates the results of experiments in which MEL cells are transfected with 20µg of plasmid DNA as follows:-

	test plasmid	pHSV-IE-βgal.	pIC20R	TOTAL
A.	2.5µg	5.0µg	12.5µg	20.0µg
B.	5.0µg	5.0µg	10.0µg	20.0µg
C.	10.0µg	5.0µg	5.0µg	20.0µg

All the results are standardised for β-galactosidase activity expressed from reference plasmid pHSV-βgal.

There is an approximately linear relationship between hGH production and the concentration of test plasmid DNA used in transfection when the hGH reporter gene is under the control of the Herpes simplex virus thymidine kinase (HSV-TK) gene promoter. However, when the hGH reporter gene is under the control of the mouse β^{major} globin gene promoter, hGH production increases with increase in plasmid copy number until an apparent threshold level is reached. Above this threshold, increasing plasmid copy number does not appear to result in increased hGH production. This does not reflect the ability of the cell to take up plasmid DNA nor its ability to secrete hGH since the levels of expression from the HSV-TK promoter are higher. It probably reflects saturation of the transcriptional machinery necessary to effect transcription from the mouse β^{major} globin gene.

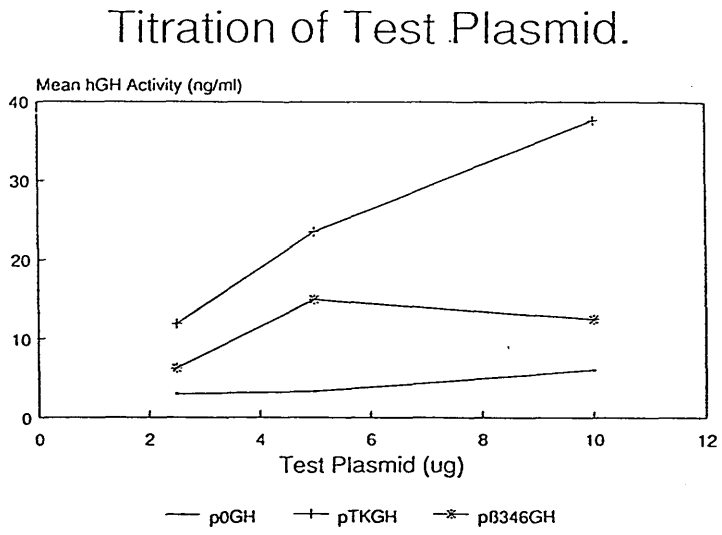
Subsequently, all transient transfections were carried out at the following DNA concentrations:-

test plasmid	pHSV-IE-βgal.	pIC20R	TOTAL
5.0μg	5.0μg	10.0μg	20.0μg

These concentrations allow maximal expression from test plasmids without saturating the system.

Figure 6.2 The Relationship Between hGH Production and Concentration of Test Plasmid DNA.

The figure shows the relationship between hGH production and the amount of test plasmid transfected into the MEL cell line used in these experiments, F4-12B2.



6.3 hGH Expression in Different Cell Lines.

The mouse β^{major} globin gene is tightly regulated *in vivo* as it is expressed exclusively in adult mouse erythroid tissue. Having set up a transient transfection system to study transcriptional regulation of this gene in MEL cells, I considered it important to show that the basic test promoter that I was using was poorly expressed in non-erythroid cell lines.

I looked at hGH expression from the β346 promoter fragment in seven different cell lines and compared expression in each with expression from the 0 promoter (negative control) and the HSV-TK promoter (positive control).

Optimal transfection conditions differed for each cell line used but to make comparisons as meaningful as possible, I tried to keep the conditions used similar to those used for transfection of the F4B12 which was the cell line that I had worked with most extensively. Table 6.3A illustrates the results obtained.

Measurement of β-galactosidase activity gives a measure of the efficiency of transfection for each experiment. Clearly, the efficiency of transfection in three of the seven cell lines (C88, J774.2, F9) is low making the interpretation of these particular transfections less valid. Levels of β-galactosidase activity were in the range 0.064-0.148 and subsequently, I judged all experiments in which transfection was less than 0.150 as measured by β-galactosidase activity to be null experiments. Clearly this is an arbitrary determination of what is a significant level of β-galactosidase activity and it may be argued that the results obtained with F9 cells whose β-galactosidase activity

values are in the range 0.089-0.148, are indeed significant. However, to obtain these results, the cell lysates were incubated with the β-galactosidase assay reagents at 37°C for up to two hours which is up to four times the length of time required to obtain a "significant" β-galactosidase activity value with extracts from transfected MEL, HeLa, 3T3 or STO cell lines. Using this parameter for efficient transfection, it is clear that transfection into four of the seven cell lines (F4B12, HeLa, 3T3 and STO) is efficient.

However, the hGH levels present in the media containing transfected STO and 3T3 cells, is very low even in the cells transfected with pTKGH. There are several possibile explanations, including: 1. the TK promoter is not promiscuous in fibroblast cell lines, 2. the hGH message is unstable in these cell lines or 3. hGH itself is not being secreted from these cells into the medium.

It was possible to test this last explanation by generating cell lysates of transfected 3T3 and/or STO (as described in section 2.24) and carrying out the hGH assay on the cell lysates (Figure 6.3B).

From Table 6.3B, it is clear that hGH activity is present in cell lysates of STO/3T3 transfected cells and that these cells fail to secrete hGH efficiently.. The ability of fibroblast-derived cell lines to secrete hGH had not been previously tested by the manufacturers and if other cell lines lack this ability, it may represent a drawback to the use of this assay system.

Analysis of expression from the β346 promoter fragment in 3T3 or STO cell lines shows that expression is not above background level as measured by hGH expression from the promoter-less construct p0GH. In contrast, expression from the promiscuous HSV-TK promoter is well above

background levels.

Unlike fibroblast derived cell lines, HeLa which is an epithelial derived cell line, and F4B12, which is an MEL cell line, secrete hGH at efficient levels into the growth medium of transfected cells. Comparison of hGH expression from the HSV-TK promoter in F4B12 and HeLa shows that expression levels are similar suggesting that HeLa is a good non-erythroid control cell line for F4B12.

To allow comparison of expression of different promoter/enhancer sequences in different cell types, the relative expression level (E) in different cell lines is defined as:

$$E = \frac{\text{expression above background from test promoter}}{\text{expression above background from control promoter}}$$

$$= \frac{x - x_0}{x_{TK} - x_0}$$

It is clear from table 6.3C that the relative expression level from the mouse β^{major} globin gene promoter fragment β346 is at least 3 times greater in F4-12B2 cells than it is in STO cells and 30-60 times greater than in HeLa or 3T3 cells. This result is comparable to that obtained by Plumb *et al.* (1989) who showed that the mouse α1 globin gene promoter is 3-4 times more active in F4-12B2 cells than in STO cells. The results, supported by the protein binding data, confirm that there are elements in the promoter which confer erythroid specificity upon expression.

Table 6.3A hGH Expression in Different Cell Lines.

Cell Line	γcounts	hGH (ng/ml)	β-gal.act.	hGH/β-gal.	x
F4B12					
p0GH	1736	0.9	0.297	3.0	3.1
	2017	1.0	0.315	3.2	
pTKGH	8976	3.1	0.264	11.7	12.55
	11321	3.8	0.283	13.4	
pβ346GH	5734	2.1	0.224	9.4	8.4
	6218	2.3	0.309	7.4	
C88.					
p0GH	912	0.6	0.123	4.9	5.05
	776	0.6	0.115	5.2	
pTKGH	4351	1.7	0.098	17.3	14.6
	3289	1.4	0.118	11.9	
pβ346GH	1767	0.9	0.137	6.6	6.4
	1583	0.8	0.130	6.2	
HeLa.					
p0GH	1310	0.7	0.187	3.7	3.5
	1059	0.7	0.211	3.3	
pTKGH	8762	3.1	0.237	13.1	15.25
	12378	4.2	0.242	17.4	
pβ346GH	1002	0.7	0.231	3.0	3.3
	1536	0.8	0.192	3.6	
F9.					
p0GH	782	0.6	0.108	5.6	5.3
	897	0.6	0.121	5.0	
pTKGH	2356	1.1	0.148	7.4	10.5
	4812	1.8	0.132	13.6	
pβ346GH	2185	1.0	0.089	11.2	10.0
	1976	1.0	0.113	8.8	
J774.2.					
p0GH	873	0.6	0.087	6.9	7.4
	912	0.6	0.076	7.9	
pTKGH	1234	0.7	0.102	6.9	8.9
	2013	1.0	0.092	10.9	
pβ346GH	784	0.6	0.064	9.4	8.15
	892	0.6	0.087	6.9	
STO.					
p0GH	345	0.4	0.321	1.2	1.45
	491	0.5	0.298	1.7	
pTKGH	367	0.5	0.315	1.6	1.7
	512	0.5	0.276	1.8	
pβ346GH	713	0.6	0.287	2.1	2.0
	412	0.5	0.269	1.9	
3T3.					
p0GH	634	0.5	0.287	1.7	2.05
	721	0.6	0.255	2.4	
pTKGH	591	0.5	0.247	2.0	1.85
	666	0.5	0.293	1.7	
pβ346GH	683	0.6	0.311	1.9	1.7
	647	0.5	0.325	1.5	

Table 6.3B hGH Expression in Cell Lysates from 3T3/STO Cell Lines.

Plasmid	γcounts	hGH (ng/ml)	β-gal.act.	hGH/βgal.	x
STO					
p0GH	1576	0.83	0.321	2.6	2.8
	1789	0.89	0.298	3.0	
pTKGH	9541	3.3	0.315	10.5	10.15
	7603	2.7	0.276	9.8	
pβ346GH	2317	1.1	0.287	3.8	3.95
	2584	1.1	0.269	4.1	
3T3					
p0GH	1935	0.94	0.287	3.3	3.05
	1201	0.71	0.255	2.8	
pTKGH	8550	3.0	0.247	12.1	11.5
	9129	3.2	0.293	10.9	
pβ346GH	2035	0.97	0.311	3.1	2.95
	1857	0.92	0.325	2.8	

Table 6.3C Relative Expression Levels in Erythroid and Non-erythroid Cell Lines.

Cell Line	E
F4B12	0.56
HeLa	0.02
STO	0.16
3T3	0.01

6.4 Comparison of Wild-type and Mutant Promoters.

To determine the role in modulation of transcription of individual elements within the promoter, I generated a series of mutant constructs as described in Chapter 5. Each mutant construct is mutant in one particular protein binding site. The mutant promoters are illustrated in figure 6.3 and are mutated either in the distal region of the promoter or in the proximal region of the promoter.

The results of transient transfection of wild-type and mutant constructs (described in fig.6.1) into uninduced F4-12B2 cells are given in table 6.4A and figure 6.4 presents these values in graph form. Table 6.4B illustrates the expression values (E) for each construct relative to the pTKGH construct as defined in the previous section.

The results show that hGH expression from the distal promter mutants, pβ1(NF1$^-$)GH, pβ2(NF-E1$^-$)GH and pβ3(Box1$^-$)GH, is 5.6, 3.7 and 6.7 fold greater respectively than expression from the wild-type promoter construct, pβ346GH. However, hGH expression from the proximal promoter mutants, pβ4(CACC$^-$)GH and pβ5(NF-E1$^-$)GH, is 3.3 and 52 fold less respectively than that from the wild-type promoter construct pβ346GH. hGH expression from the longer wild-type promoter fragment β745 is 2.5 fold greater than that from β346, the shorter promoter fragment.

These results indicate that the distal promoter mutations upregulate transcription by 4-7 fold, that the proximal promoter mutations down-regulate transcription by 3-52 fold, and that there are sequences upstream between -346 and -745 bp which upregulate transcription by 2.5 fold. This suggests that CACC, NF-E1, CCAAT and

TATA elements in the proximal promoter are essential for promoter function. While demonstrating for the first time that there is an NF-E1 binding site in the proximal promoter, this data confirms the work of Myers et al. (1986) which showed an essential role for the TATA, CCAAT and CACC boxes in mouse β^{major} globin gene promoter function. This work identifies NF-E1, Box 1 and NF-1 elements in the distal promoter which exert a negative regulatory effect on expression in uninduced MEL cells. These elements may constitute a silencer element of the kind described for the human ε globin gene (Cao et al (1989). There are additional elements upstream which exert a small positive regulatory effect on transcription.

NF-E1 exerts a different effect on transcriptional initiation depending on the context of the NF-E1 binding site; the NF-E1 site in the distal promoter is a negative regulatory element while that of the proximal promoter is a positive regulatory element. This illustrates the importance of context in determining the role of individual enhansons as described for the SV40 enhancer. The role of these regulatory elements in modulating transcription of the mouse β^{major} globin gene as MEL cells are induced to differentiate is examined in stably transfected cell lines and is described in chapter 7.

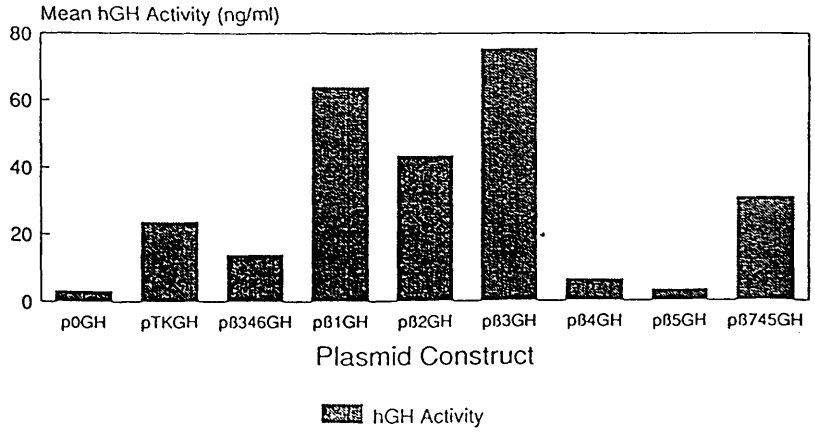
Table 6.4A Comparison of Wild-Type and Mutant Promoters.

Plasmid	x	S.E.	n
p0GH	3.35	0.65	8
pTKGH	24.2	1.74	8
pβ346GH	14.2	1.33	8
pβ1GH	64.2	2.85	8
pβ2GH	43.7	2.34	8
pβ3GH	75.7	3.08	8
pβ4GH	6.6	1.48	3
pβ5GH	3.5	1.08	3
pβ745GH	31.2	3.22	3

Table 6.4B Relative Expression Values for Mutant and Wild-Type Promoter Constructs.

plasmid	E
pTKGH	1.00
pβ346GH	0.52
pβ1GH	2.9
pβ2GH	1.9
pβ3GH	3.5
pβ4GH	0.16
pβ5GH	0.01
pβ745GH	1.3

Figure 6.4 Expression from Wild-type and Mutant Promoters.



6.5 Comparison of hGH Expression from the Wild-type Promoter Modulated by the Chicken β^A and Mouse β^{major} Globin Gene 3′ Sequences.

Section 3.4 described the identification of recognition sites for the NF-E1 and Box 1 DNA binding proteins in a region 400 bp 3′ of the poly adenylation site of the mouse β^{major} globin gene. To determine whether these sequences had a functional role in the modulation of transcription, this region was cloned immediately 3′ of the hGH gene controlled by the wild-type promoter fragment, β346, as described in Table 6.1 to generate construct $p\beta^M$. Expression from this construct was compared with expression from the analogous construct, $p\beta^A$GH, containing the chicken β^A globin gene enhancer 3′ of the hGH gene. These plasmid constructs are illustrated in Table 6.1. Table 6.5A and figure 6.5A illustrate the results of these experiments and table 6.5B gives the relative expression value for each construct.

Clearly expression from the wild-type promoter is enhanced by 6.0 fold by sequences from the 3′ region of the mouse β^{major} globin gene in MEL cells. This enhancement is greater than that obtained for the chicken β^A globin gene enhancer 4.8 fold but not significantly greater. The introduction of 3′ enhancer sequences generates an up-regulation of transcription that is equivalent to that generated by mutation of distal promoter sequences. These functional results, combined with the protein binding data, suggest that these 3′ mouse β^{major} globin gene sequences do indeed constitute an enhancer element. Further work to determine whether the element can operate in either orientation is theoretically necessary to confirm that this element is a "classical" enhancer.

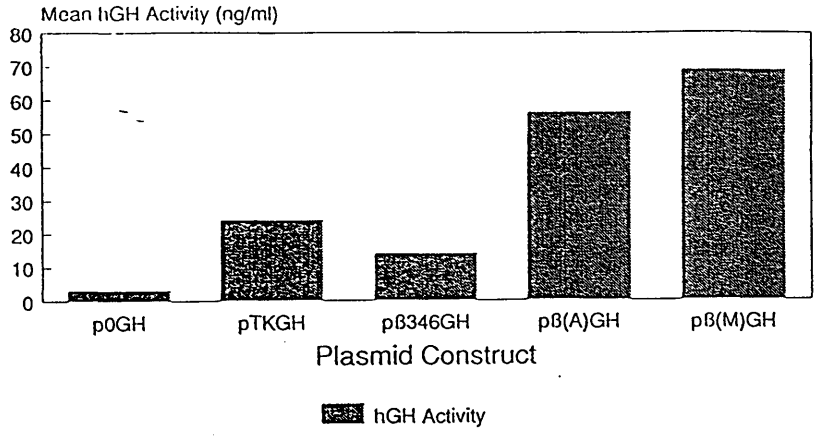
Table 6.5A Comparison of hGH Expression from the Wild-Type Promoter Modulated by the Chicken β^{A} Enhancer or by Mouse β^{major} Globin Gene 3′ Sequences.

Plasmid	x	S.E.	n
p0GH	3.35	0.65	8
pTKGH	24.2	1.72	8
pβ346GH	14.2	1.33	8
$p\beta^{A}$GH	56.4	2.66	8
$p\beta^{M}$GH	69.0	4.80	3

Table 6.5B Relative Expression Levels from Wild-type Promoter Sequences modulated by Chicken and Mouse 3′ Sequences.

Plasmid	E
pTKGH	1.00
pβ346GH	0.52
$p\beta^{A}$GH	2.5
$p\beta^{M}$GH	3.1

Figure 6.5 Enhancer Modulated Expression.



6.6 α-1 and β^{major} Globin Gene Promoter Interactions.

As discussed in section 1.4, the regulation of human α and β globin genes differ in a number of respects. Differences in the regulation of mouse α and β globin genes are less well characterised. Certainly, the regulation of the human α and the mouse α1 globin genes differs in that the human α globin gene promoter lies in an HTF island and can function in the absence of an enhancer when transfected into non-erythroid cell lines whereas the mouse gene is not promiscuous (section 1.2.2) and does not lie in an HTF island (Whitelaw *et al.* (1989). It was interesting to look for evidence of α-β promoter interaction at the level of transcription by cotransfection of α-1 promoter constructs and β^{major} promoter constructs in the transient transfection assay systems described above. The α1 and β^{major} globin genes are not linked *in vivo* and in this study they are cotransfected on separate plasmid vectors; the chicken ε and β^{A} globin genes are linked *in vivo* and in work carried out by Choi & Engel (1988) to compare the regulation of the chicken ε and β^{A} globin genes, the two genes were linked in *cis* and cotransfected on the same plasmid vector. The activity of the α-1 promoter could be distinguished from that of the β^{major} promoter on the basis that the α-1 promoter regulated initiation of transcription from the CAT reporter gene whereas the β^{major} promoter regulated initiation of transcription from the hGH reporter gene. Figure 6.1 describes the constructs used. Table 6.6A and 6.6B describe the results of these cotransfection experiments.

% CAT activity was determined by cutting spots A, B and C from the TLC plates, determining emission from them due to ^{14}C in a scintillation counter and expressing CAT activity as:

$$\%\text{CAT activity} = \frac{A+B}{A+B+C}$$

where A = 1, 3 - diacetyl ^{14}C-Chloramphenicol

B = 1 or 3 acetyl ^{14}C-Chloramphenicol

C = ^{14}C-Chloramphenicol

The % CAT activities were standarised to the β-galactosidase activities for each experiment.

The results show that CAT activity increases linearly with increase in CAT plasmid construct transfected (lanes 2, 3, 4 of figure 6.6A).

The effect of cotransfection is to down-regulate transcription from the α-1 promoter and to up-regulate transcription from the β^{major} promoter. When pβ346GH is cotransfected with pαCAT1 there is a 9.1 fold down-regulation in CAT activity (figure 6.6B) and a 3.3 fold increase in hGH activity (figure 6.6C).

It is known that during differentiation of the MEL cell that mouse α-1 globin gene is expressed earlier than the β^{major} gene. It is therefore interesting to note that initiation of transcription from the α-1 promoter in these cotransfection experiments is down-regulated by expression from the mouse β^{major} globin gene promoter. If a transcription factor which is necessary for initiation from both promoters is in limiting supply, this observation could be explained if the β^{major} promoter had a greater affinity for this factor than the α-1 promoter. It is known that NF-E1 is required for modulation of transcription from both of these promoters. It is also known that the β^{major} globin gene promoter has two NF-E1 binding sites, at -60 and

-212 bp whereas the α-1 promoter has only one, at -180 bp (Plumb et al (1989). When the NF-E1 (pβ2GH) or Box 1 (pβ3GH) binding sites in the β^{major} distal promoter are mutated, there is a much smaller down regulation of CAT expression from the α1 promoter of 1.3-1.8 fold (figure 6.6B).

It is observed that mutation of βNF1 (pβ1GH) in the β^{major} promoter also gives a smaller down-regulation, compared with pβ346GH, of transcription from the α1 promoter by 4.3 (Figure 6.6B). Unlike NF-E1, NF-1 and Box 1 are not known to be involved in regulating transcription of the α-1 globin gene and it is initially difficult to explain why mutation of NF-1 or Box 1 should affect transcription from the α1-globin gene promoter. However, if NF-E1 interaction with the distal region of the mouse β^{major} globin gene promoter depends upon a cooperative interaction with βNF1 and Box 1, then mutation of either βNF1 or Box 1 will release NF-E1 to allow interaction with the α-1 promoter.

These results suggest that the α-1 and β^{major} globin gene promoters are competing for NF-E1. It should be realised however that this transient transfection assay system creates artificial conditions, such as the high copy number of both the α-1 and the β^{major} promoter constructs. This may be responsible for the competition effects observed since at low plasmid copy number it is unlikely that NF-E1 or other binding factors would be limiting.

Cotransfection of β346GH with equimolar amounts of pαCAT1 results in a 3.3 fold stimulation of transcription from the β^{major} promoter (Figure 6.6C). The upregulatory effect of cotransfection on the β^{major} promoter is in contrast to the down-regulatory effect on the

α-1 promoter. If it is assumed, as for the effect on CAT activity, that the α-1 and β^{major} promoters are competing for a limiting transcription factor, probably NF-E1, then the upregulatory effect on hGH expression may be explained as follows. The distal promoter region of β^{major} globin gene contains an NF-E1 site and mutation of this NF-E1 site results in up-regulation from β^{major} of 6.3 fold in transient transfection assays (section 6.4). This leads to the hypothesis that this NF-E1 site is part of a negative regulatory element which also includes βNF1 and Box 1. By competing out the binding of NF-E1 to this site, the α-1 promoter is derepressing transcription from the β^{major} promoter.

Cotransfection of pαCAT1 with pβ1GH, pβ2GH and pβ3GH results in a 4.3, 6.0 and 5.2 fold upregulation of hGH expression compared with pβ346GH (figure 6.6C). This is equivalent to the effect observed when the mutant constructs are transfected alone (section 6.4) and suggests that, unlike its up-regulatory effect on wild-type β^{major} promoter sequences, that cotransfection with pαCAT1 has no effect on expression from the mutant promoter sequences. This is as expected if the mutant promoters have already been derepressed by mutation of NF-E1, Box 1 or NF-1.

Figure 6.6 α-1 and β^{major} Globin Gene Promoter Interactions.

α-1-CAT and β^{major}-hGH promoter constructs were cotransfected as described in section 2.26 into F4-12B2 cells, cell lysates generated as described in 2.24 and CAT activity measured as described in 2.26. The results of the CAT assays are shown below.

Lane 1 5µg pαCAT0, 5µg pHSV-IE-βgal., 10µg pIC20R.

Lane 2 1µg pαCAT1, 5µg pHSV-IE-βgal., 14µg pIC20R.

Lane 3 2.5µg pαCAT1, 5µg pHSV-IE-βgal., 12.5µg pIC20R.

Lane 4 5µg pαCAT1, 5µg pHSV-IE-βgal., 10µg pIC20R.

Lane 5 5µg pαCAT1, 5µg pHSV-IE-βgal., 5µg pIC20R.
5µg pβ346GH

Lane 6 5µg pαCAT1, 5µg pHSV-IE-βgal., 5µg pIC20R
5µg pβ1GH

Lane 7 5µg pαCAT1, 5µg pHSV-IE-βgal., 5µg pIC20R
5µg pβ2GH

Lane 8 5µg pαCAT1, 5µg pHSV-IE-βgal., 5µg pIC20R
5µg pβ3GH

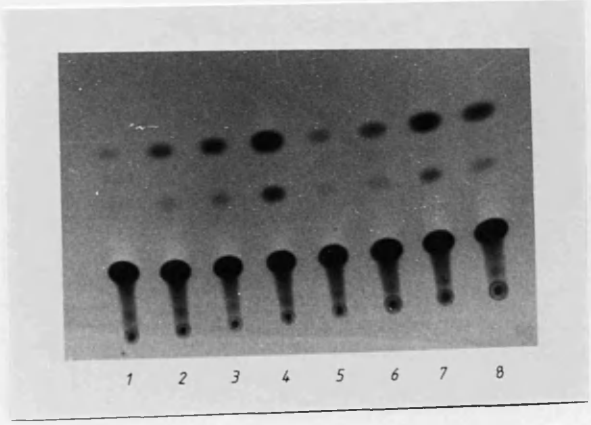


Table 6.6A α-1 and β^{major} Globin gene Promoter Interactions.

Plasmid	γcounts	hGH	β-gal.act.	hGH/β-gal.	% CAT	%CAT/β-gal.
1. αCAT0	729	0.6	0.253	2.4	2.45	9.7
	990	0.6	0.276	2.2	2.40	8.7
2. αCAT1	985	0.6	0.311	1.9	9.70	31.2
(1.0μg	765	0.6	0.257	2.3	7.09	27.6
3. αCAT1	885	0.6	0.302	2.0	14.65	48.5
(2.5μg)	929	0.6	0.251	2.4	13.1	52.3
4. αCAT1	774	0.6	0.262	2.3	23.42	89.4
(5.0μg)	834	0.6	0.271	2.2	24.72	91.2
5. αCAT1	9520	3.3	0.253	13.0	2.58	10.2
β346	12379	4.2	0.207	20.3	2.01	9.7
6. αCAT1	17414	5.7	0.266	21.4	6.04	22.7
β1	20440	6.7	0.302	22.2	5.89	19.5
7. αCAT1	23250	7.5	0.231	32.5	15.5	67.3
β2	24325	7.9	0.256	30.9	17.97	70.2
8. αCAT1	24913	8.1	0.294	27.6	14.5	49.5
β3	23165	7.5	0.302	24.8	16.1	53.6
9. β346	3486	1.4	0.275	5.1	-	-
	4211	1.6	0.319	5.0	-	-
10.p0GH	1762	0.89	0.287	3.1	-	-
	1235	0.72	0.271	2.7	-	-

Table 6.6B.

	hGH x	CAT x
1.	2.3	9.2
2.	2.1	29.4
3.	2.2	50.4
4.	2.25	90.3
5.	16.65	9.95
6.	21.8	21.1
7.	31.7	68.75
8.	26.2	51.55
9.	5.05	-
10.	2.9	-

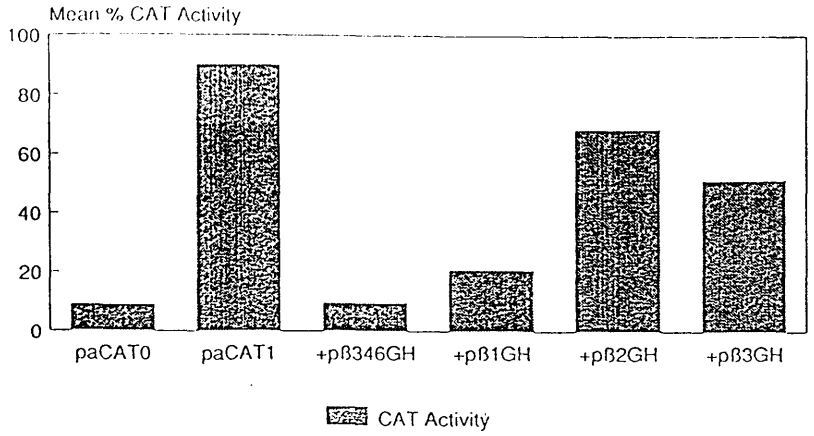


Figure 6.6B Changes in CAT Expression After Cotransfection.
Cotransfection of wild-type promoter construct pβ346GH with equimolar amounts of pαCAT1 results in a 9.1 fold down-regulation in CAT activity. Cotransfection of pαCAT1 with pβ1GH, pβ2GH and pβ3GH results in a down-regulation in CAT activity of 4.3, 1.3 and 1.8 fold respectively.

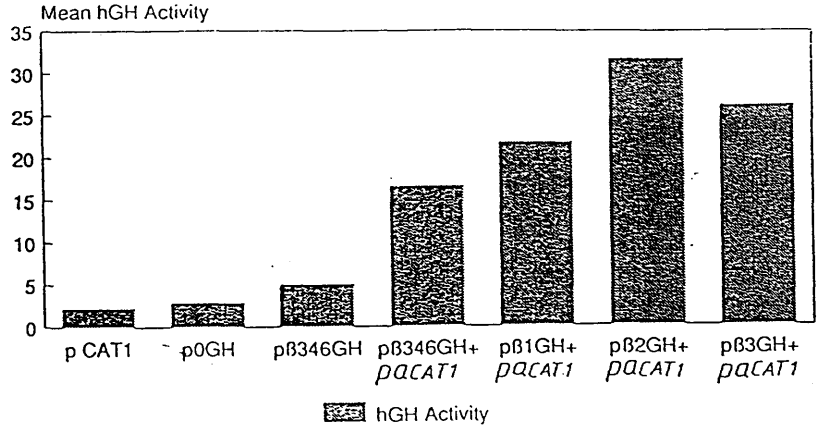


Figure 6.6C Changes in hGH Expression After Cotransfection.
Cotransfection of wild-type promoter construct pβ346GH with equimolar amounts of pαCAT1 results in a 3.3 fold increase in transcriptional initiation from the β^{major} promoter. Cotransfection of pαCAT1 with pβ1GH, pβ2GH and pβ3GH results in an up-regulation of 4.3, 6.3 and 5.2 fold respectively.

CHAPTER 7 FUNCTIONAL ANALYSIS OF THE MOUSE β^{MAJOR} GLOBIN GENE PROMOTER AND 3' SEQUENCES IN STABLE TRANSFECTION ASSAYS.

INTRODUCTION

As mentioned in Chapter 6, stable transfection assay systems are useful as they allow expression to be monitored over longer culture periods than that which transient transfection assay systems allow. It is therefore possible to monitor changes in reporter gene expression as the MEL cells are induced to differentiate. Stable transfectants for the promoter/enhancer constructs to be analysed were generated by cotransfection of test plasmid and a selection plasmid conferring neomycin resistance on cells (described in section 2.22). The stable integration of the test plasmid into the host cell genome along with a selective marker gene allows expression to be monitored through successive cell divisions. One of the main disadvantages of this system is that although the test DNA sequences are stably integrated, there are often many copies integrated. Multiple copies may be integrated at different sites throughout the genome such that each copy is under different cis-acting influences such as differences in chromatin conformation or other flanking regulatory elements. While it is possible to determine the copy number of a given construct in the genome, it is not easy to determine which copies are being actively transcribed or how many are inactive as a consequence of chromatin condensation at that locus.

However, the stable transfectants described here are generated such that for each test DNA introduced, a pool of cells is generated from at least one hundred different neomycin resistant colonies as

described in section 2.22. Each clone probably contains different plasmid copy numbers integrated at different sites in the genome. Differences in copy number and integration sites between clones should average out such that different pools of cells have approximately equivalent copies of integrated test plasmid.

7.1 Cell Growth Rate Under Different Inducing Conditions.

Stable transfectants were generated to analyse the effect that induction to differentiate had on hGH expression from the different promoter and enhancer constructs being studied. Two different types of Friend virus induced mouse erythroleukaemic cell lines were used; F4-12B2 and C88. Study of changes in transcriptional regulation as C88 cells are induced to differentiate could be compared to the work in the laboratory of F.Grosveld on the transcriptional regulation of the human β globin gene in C88 cells (Antoniou et al. (1988).

F4-12B2, which was isolated in the laboratory of Ostertag (1979) has a high transfection efficiency under the conditions used in this system and can be induced to differentiate rapidly (3-5 days) with either 2% DMSO or 4 mM HMBA. C88, which is deficient in adenine phosphoribosyl transferase (Deisseroth and Hendrick, (1978), transfects poorly in this system, and responds more slowly to induction (5-8 days) with 2% DMSO or 4 mM HMBA. Although the C88 cell line transfects poorly, this does not adversely affect its use in stable transfections since stringent selection procedures are used to confer growth advantage on successful transfectants. After selection, C88 transfectants grow as well as F4-12B2 transfectants.

To assay changes in hGH production during the induction of differentiation, each stable transfectant was set up in triplicate in 5ml of medium at $1x10^5$cells/ml, 80 µg/ml neomycin and incubated at 37^oC and 5% CO_2. Twenty four hours after setting the cells up, 2% DMSO and 4mM HMBA were added to two of the three plates for each transfectant. The third plate was not treated and the cells therein were allowed to grow in the uninduced state. The medium was assayed for hGH activity at 24 hour intervals and is therefore a measure of the accumulation of hGH in the media over the 7 day period.

Control plates were set up in parallel with each transfectant and grown under identical conditions. At twenty four hour intervals, the cell number in each control plate was determined for uninduced, DMSO induced and HMBA induced cells. Cell number was measured in control plates rather than in the plates from which medium was being removed for assay since determining cell number involved treating the cells with trypsin. By the end of the seven day period, both the F4-12B2 and the C88 transfectants had differentiated as judged by the extent of haemoglobin present in these cells and the resulting red to purple colour of the cell pellets.

The results of measuring cell growth rate for C88 and F4-12B2 cell lines under uninduced, DMSO induced and HMBA induced conditions are illustrated in figures 7.1A and B. The results show that both C88 and F4-12B2 cells grow exponentially over a 7 day period and that induction to differentiate with either DMSO or HMBA causes the rate of proliferation to decrease dramatically over the same period. The growth rate of F4-12B2 cells decreases more rapidly in response to induction from a doubling time of 16 hours prior to induction to 30 hours 4 days

post induction, than the growth rate of C88 cells which decreases from a doubling time of 16 hours prior to induction to 30 hours measured 5 days post induction. This is consistent with the more rapid rate of differentiation of F4-12B2 cells in response to chemical induction.

Figure 7.1A F4-12B2 Growth Curve.

F4-12B2 cells were set up at $1x10^5$ cell/ml in 5 ml of medium and allowed to grow for 7 days under uninduced conditions, DMSO induced conditions or HMBA induced conditions as described in the text.

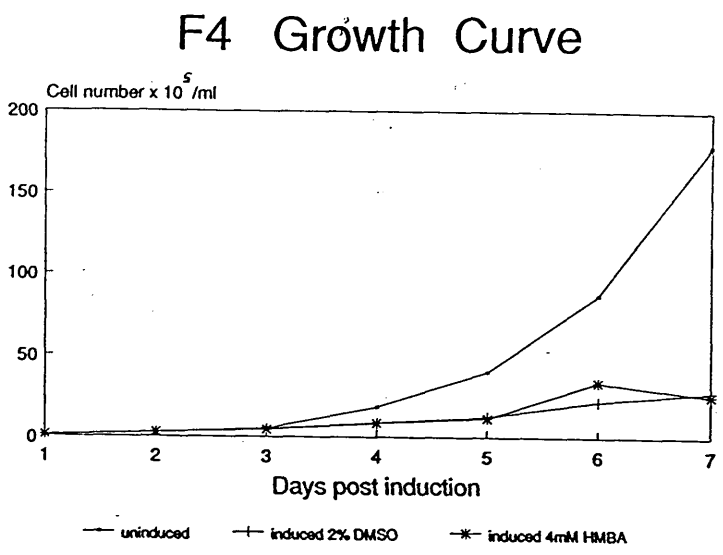
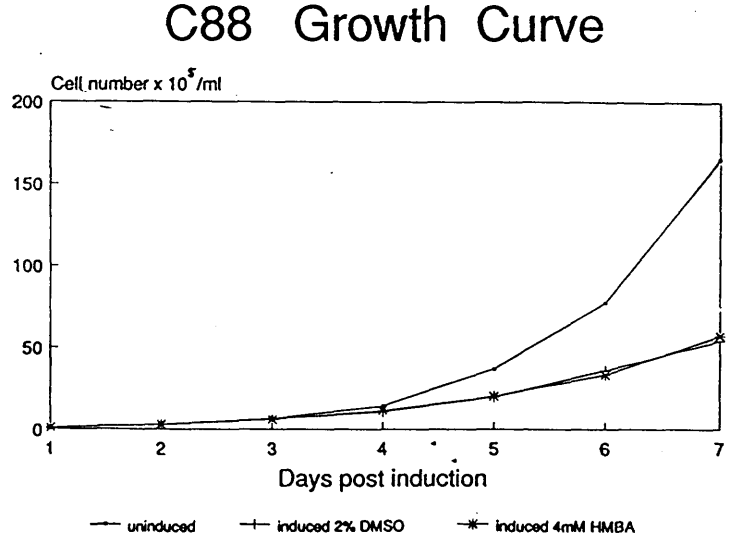


Figure 7.1B C88 Growth Curve.

C88 cells were set up at $1x10^5$ cell/ml in 5 ml of medium and allowed to grow for 7 days under uninduced conditions, DMSO induced conditions or HMBA induced conditions as described in the text.



7.2 The Effect of Induction on Expression from Wild-Type and Mutant Promoters.

In section 6.4 , results were presented which showed that mutation of βNF1, NF-E1 or Box 1 recognition sites in the distal region of the mouse β^{major} globin gene promoter resulted in transcriptional upregulation from that promoter of 4 - 7 fold. This suggested that the distal promoter region is a negative regulatory element, whereas the proximal promoter region is absolutely required for transcription from the mouse β^{major} globin gene.

These results were obtained in transient transfection assay systems and did not take account of the effect that induction to differentiate might have on the rate of transcriptional initiation from mutant promoters. It was therefore of interest to look at the role of the negative regulatory element during differentiation. In particular, it was of interest to determine whether the activity of this element was repressed or altered following induction to differentiate.

By generating cell lines stably transfected with the test constructs, it was possible to examine changes in hGH expression over a 7 day period as these cell lines were induced to differentiate. The cells were set up as described in section 7.2 and 2x100 µl aliquots of growth medium were harvested at 24 hour intervals and assayed for hGH activity.

To compare changes in hGH production per cell between stable transfectants and to obtain a measure of the level of induction under different conditions and at different time points, an induction ratio is defined as:

$$I = \frac{\text{hGH/cell on day x}}{\text{hGH/cell on day 2}}$$

The cells are induced to differentiate on day 2, thus I gives a measure of the change in hGH expression following induction and is independent of copy number and of integration site since hGH levels are being compared within each pool of transfectants. For the purposes of this study, which is to compare the effects of different mutations and the role of enhancer sequences in the reponse to induction, I gives a meaningful comparison of expression levels.

The effects of induction to differentiate upon expression from the wild-type and mutant promoter constructs stably transfected into F4-12B2 are shown in figure 7.2. There is a 10-20 fold increase in hGH production from the wild-type promoter construct but only a 5-10 fold increase in expression from the mutant promoter constructs (figure 7.2A). When the levels of hGH production are standardised for cell number and expressed relative to expression on day 2, (figure 7.2B) it can be seen that while the induction value for the wild-type promoter in F4-12B2 cells induced to differentiate increases 2-4 fold, that the induction value under uninduced conditions is less and decreases to less than 0.5 on day 5. The induction values for the mutant promoter constructs under all three growth conditions are similar to those obtained for the wild-type promoter construct in cells grown under uninduced growth conditions.

The responses of the C88 cell lines to the three growth conditions (uninduced, induced 2% DMSO, induced 4 mM HMBA) are shown in figure 7.3A. There is a 12-25 fold increase in hGH production from the

wild-type promoter construct and only a 5-10 fold increase in hGH production from the mutant promoter constructs in the C88 cell lines under all three growth conditions. However, there is a 20 fold increase in hGH production from mutant construct pβ1GH in uninduced C88 cells.

When hGH production by all stable transfectants was standardised for cell number and expressed relative to hGH production per cell prior to induction (Figure 7.3B), hGH production from the mutant promoter constructs in C88 cells are equivalent, including that from promoter construct pβ1GH. The induction values for these constructs are less than one and up to 40 fold less than that obtained for the wild-type promoter construct under all three growth conditions.

It was observed in figure 7.3A that hGH production from the wild-type promoter construct in C88 cells, grown under uninduced growth conditions, is as high as that obtained under induced growth conditions (either 2% DMSO or 4 mM HMBA induced). When hGH production is standardised for cell number and expressed relative to expression on day 2, the induction value obtained for the wild-type promoter construct on days 3, 4 and 5 under uninduced growth conditions is greater than that obtained on those days for cells induced with DMSO or HMBA. However, it was observed that the rate of decrease in hGH production after day 2 was greater in uninduced cells than in those induced to differentiate.

The effects of the two inducing agents on each cell line transfected with the wild-type promoter construct are different. The induction effect of 2% DMSO on both C88 and F4-12B2 stably transfected with pβ346GH appears to be less than the effect of 4 mM HMBA on the same cells. Neither inducing agent succeeds in inducing hGH expression

from the mutant promoter constructs in either MEL cell line. The absolute hGH expression levels from the mutant promoters are as high as the induced level of hGH expression from the wild-type promoter for both F4-12B2 and C88. This leads to the suggestion that mutation of the negative regulatory element in the distal region of the mouse β^{major} globin gene results in constitutive expression from that promoter. This is consistent with the observation that this element is responsible for down-regulating expression from the wild-type mouse β^{major} globin gene promoter in uninduced MEL cells.

The Effect of Induction on Expression from Wild-type and Mutant Promoters.

Figure 7.2A F4-12B2 hGH Production.

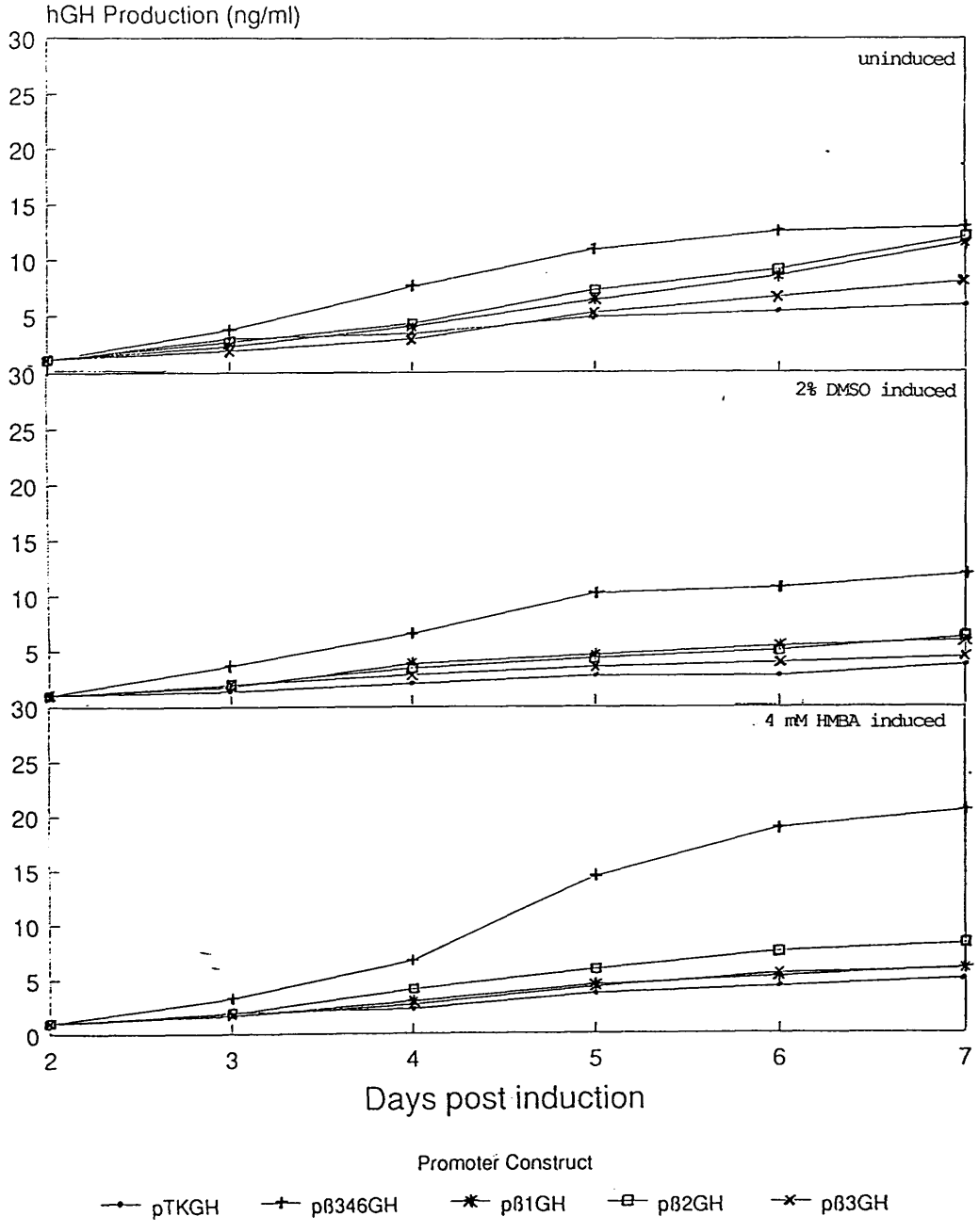


Figure 7.2B F4-12B2 I Values.

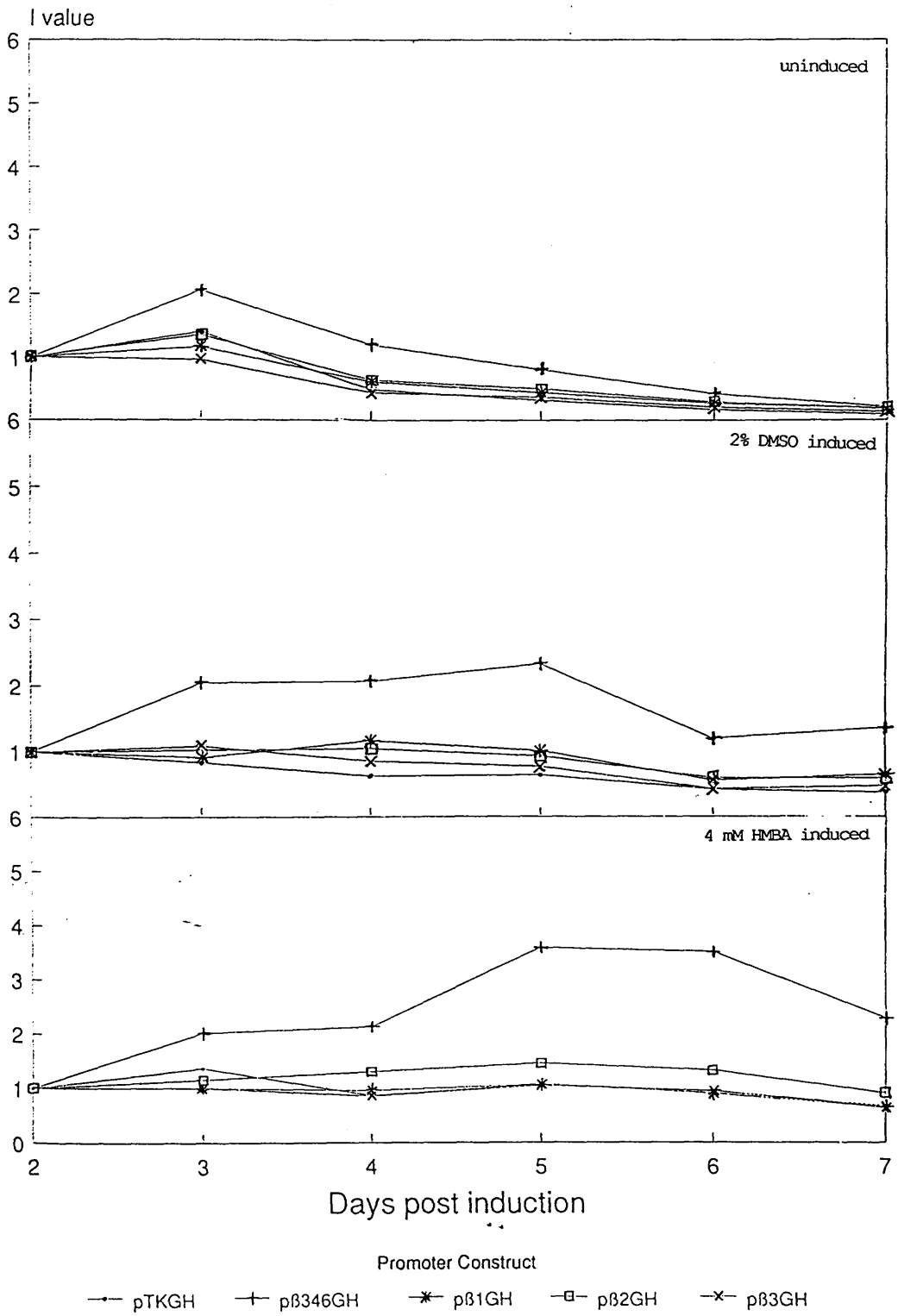


Figure 7.3A C88 hGH Production.

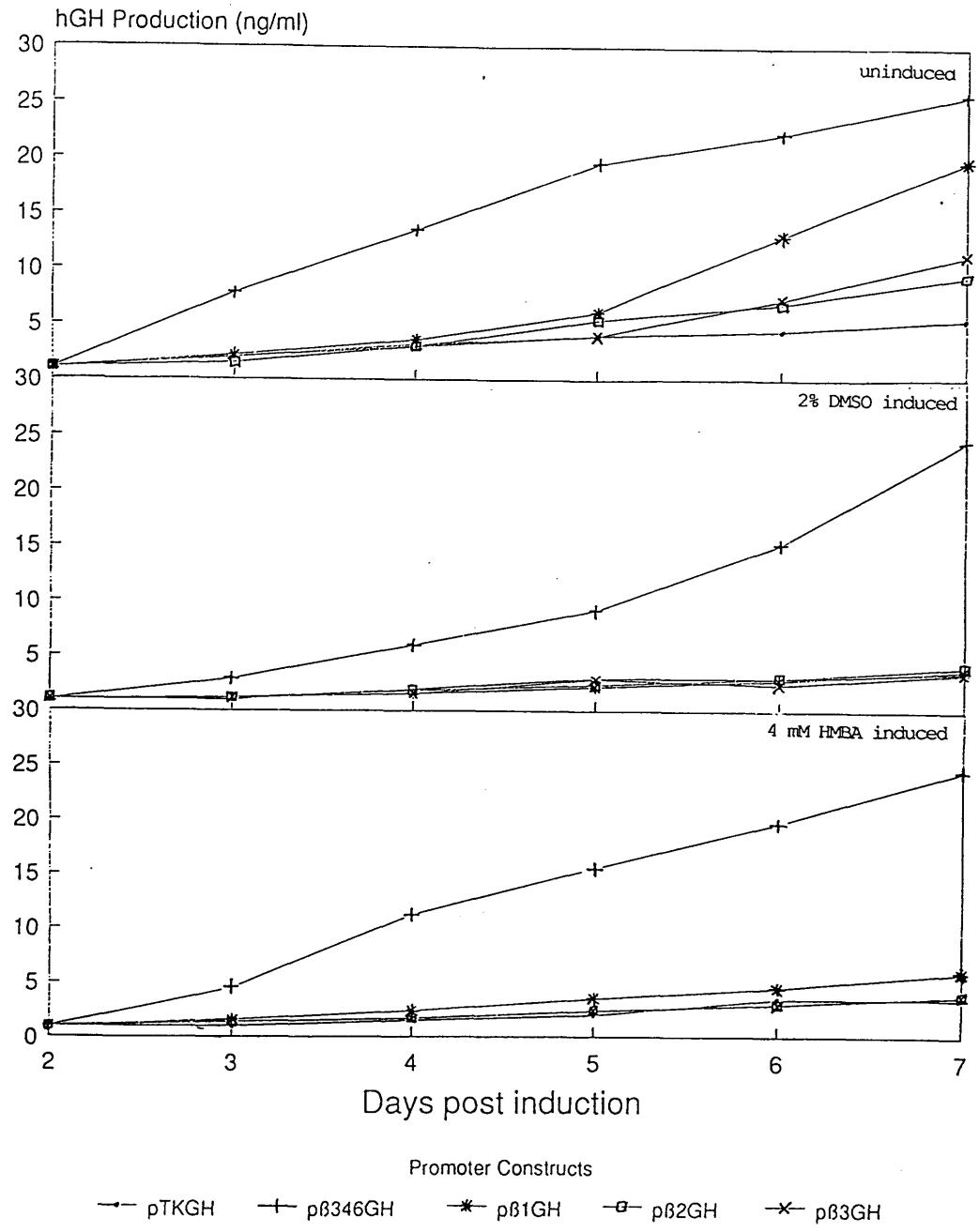
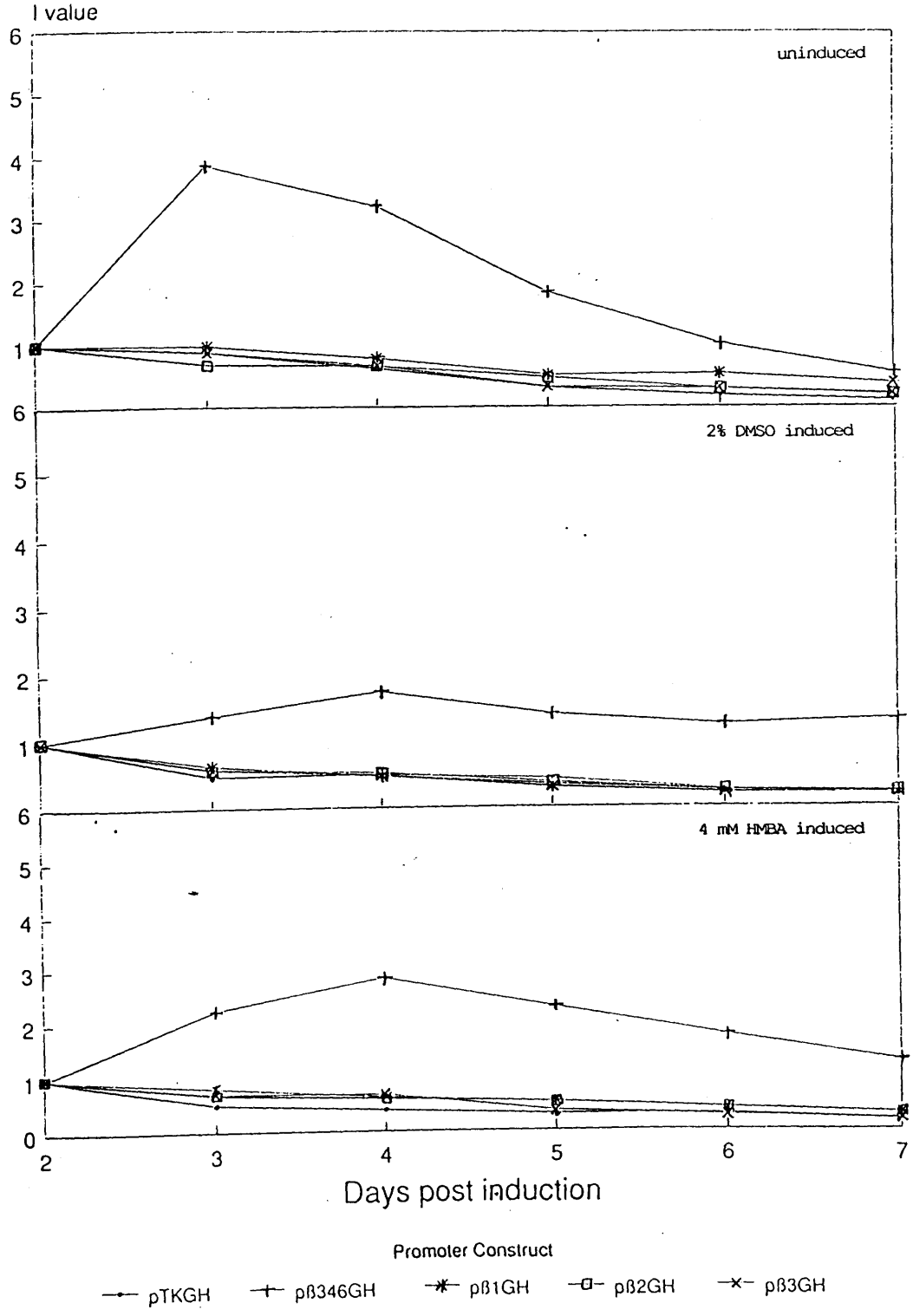


Figure 7.3B C88 I Values.



7.3 Modulation of Induction of Expression from the Mouse β^{major} Globin Gene Promoter by Chicken β^{A} and Mouse β^{major} Globin Gene Enhancers.

As mentioned at the end of section 7.2, the observed level of induction from the wild-type mouse β^{major} globin gene promoter in transfected F4-12B2 and C88 cells induced with either 2% DMSO or 4mM HMBA is not as high as would be expected based on the observed 12-fold induction by 2% DMSO of endogenous haem production in Friend cells (Paul and Hickey (1974).

Sequences found 3′ of the mouse β^{major} globin gene have been shown to bind NF-E1 and Box 1 by DNAase 1 footprinting and are suspected of constituting an enhancer element based on transient transfection assays descibed in section 6.5. These sequences were cloned 3′ of the hGH reporter gene as described previously and stably transfected into F4-12B2 and C88 cells to determine whether they enhanced the level of induction with DMSO or HMBA of hGH expression observed from the wild-type mouse β^{major} globin gene promoter fragment β346. This experiment was also carried out with the chicken β^{A} globin gene enhancer cloned 3′ of the hGH gene and the effect of the mouse sequences on inducible expression was compared with that of the chicken β^{A} globin gene enhancer.

The hGH production results are illustrated in figure 7.4 and figure 7.5. There is no appreciable increase in hGH production from either of the enhancer constructs compared with the wild-type promoter construct, in F4-12B2 cells induced to differentiate with either DMSO or HMBA (figure 7.4). However with C88 transfectants, a 70-80 fold induction of hGH expression is observed with both DMSO and HMBA

compared with a 20-fold induction from the wild-type promoter construct (Figure 7.5). This difference in induction between the enhancer and the promoter constructs is not observed in the uninduced cells.

When the hGH values are standardised for cell number and to expression on day 2 prior to induction, it can be seen that the response to induction in F4-12B2 cell transfectants is greater with HMBA induced cells than for DMSO cells (Figure 7.4). This is consistent with work with F4-12B2 cells transfected with the mutant promoter constructs and suggests that DMSO induction is less efficient than HMBA induction for this cell line. The 2 to 3-fold induction by DMSO of hGH expression observed with the enhancer constructs is not significantly greater than the induction value obtained for expression from the wild-type promoter without any linked enhancer sequences. When these cells are induced with 4 mM HMBA, the induction values for expression from the enhancer constructs range from 2 to 6-fold compared with 2 to 3.5 fold for the wild-type promoter construct. These results suggest that the enhancer sequences are enhancing the effect of HMBA mediated induction of expression from the mouse β^{major} globin gene promoter but that this effect is not significant with DMSO mediated induction.

In contrast, the effect of a 3′ linked enhancer sequence on expression from the β^{major} globin gene promoter in C88 cells is significant with both DMSO and HMBA mediated induction (Figure 7.5). With DMSO induced expression, induction levels from the enhancer constructs is 4 fold greater than that from the promoter alone. With HMBA induced cells, the effect is similar to that of DMSO induced cells with an induction level almost 5-fold higher for cells transfected with the enhancer constructs compared with the cells transfected with the

wild-type promoter construct.

In conclusion, it is apparent that the induction of expression from the mouse β^{major} globin gene depends upon which cell line is used and also upon which inducing agent is used to analyse the effect. However, the up-regulatory effect upon induction of the enhancer sequences is consistently observed except in DMSO induced F4-12B2 transfected cells. These results together with the results in section 7.3 present a picture of the induction of transcription from the mouse β^{major} globin gene dependent upon activity of 3′ enhancer sequences and upon derepression of the negative regulatory element in the distal promoter.

Derepression of transcription from this promoter in vivo will be occurring in response to natural inducing agents, developmental and differentiation signals. How far the observations made here with DMSO and HMBA mediated induction are related will be discussed in Chapter 8 but clearly the variability, both qualitative and quantitative suggests that a better model system is required.

Modulation of Induction of Expression from the Mouse β^{major} Globin Gene Promoter by Chicken β^{A} and Mouse β^{Major} Globin Gene Enhancers.
Figure 7.4A F4-12B2 hGH Production.

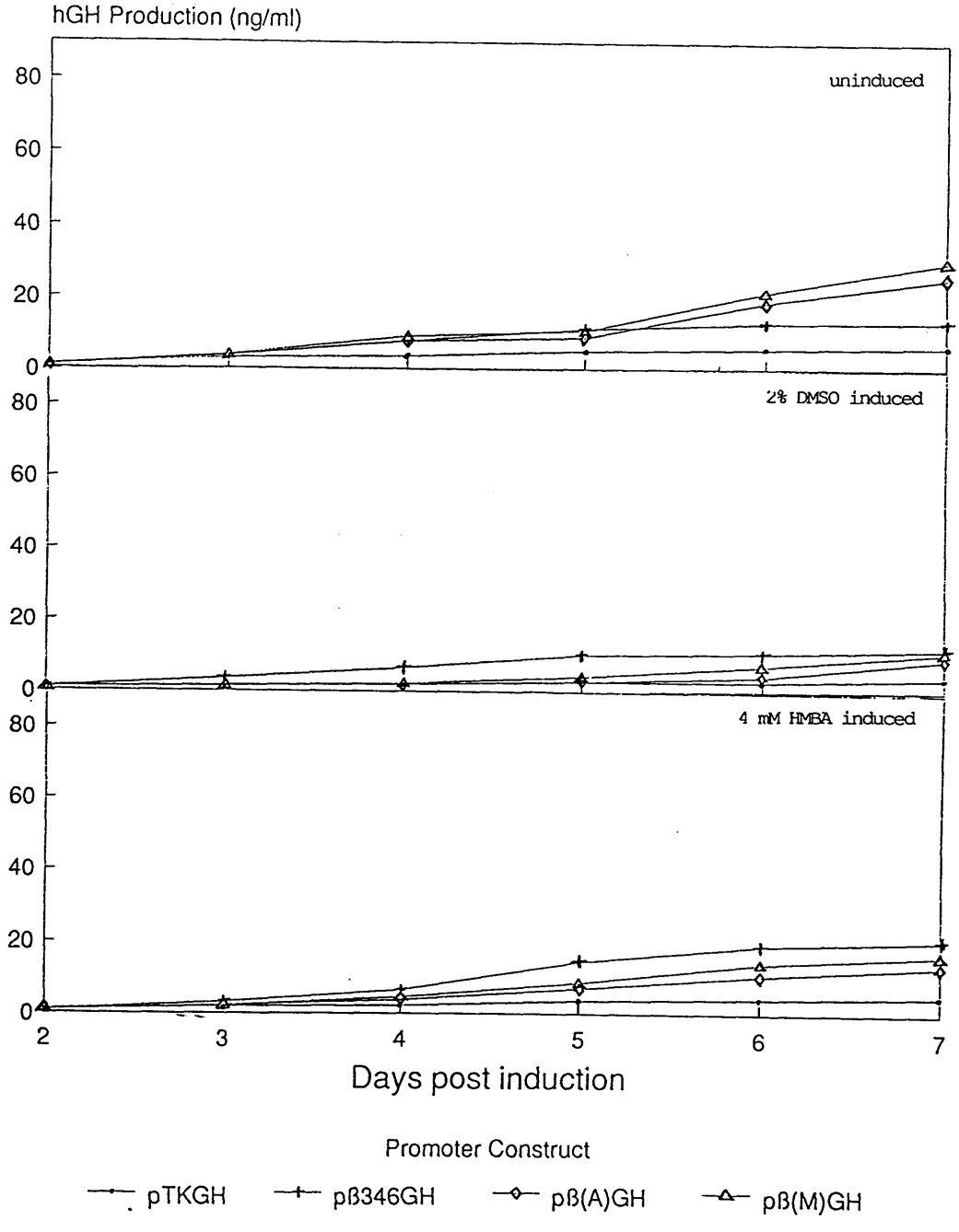


Figure 7.4B F4-12B2 I Values.

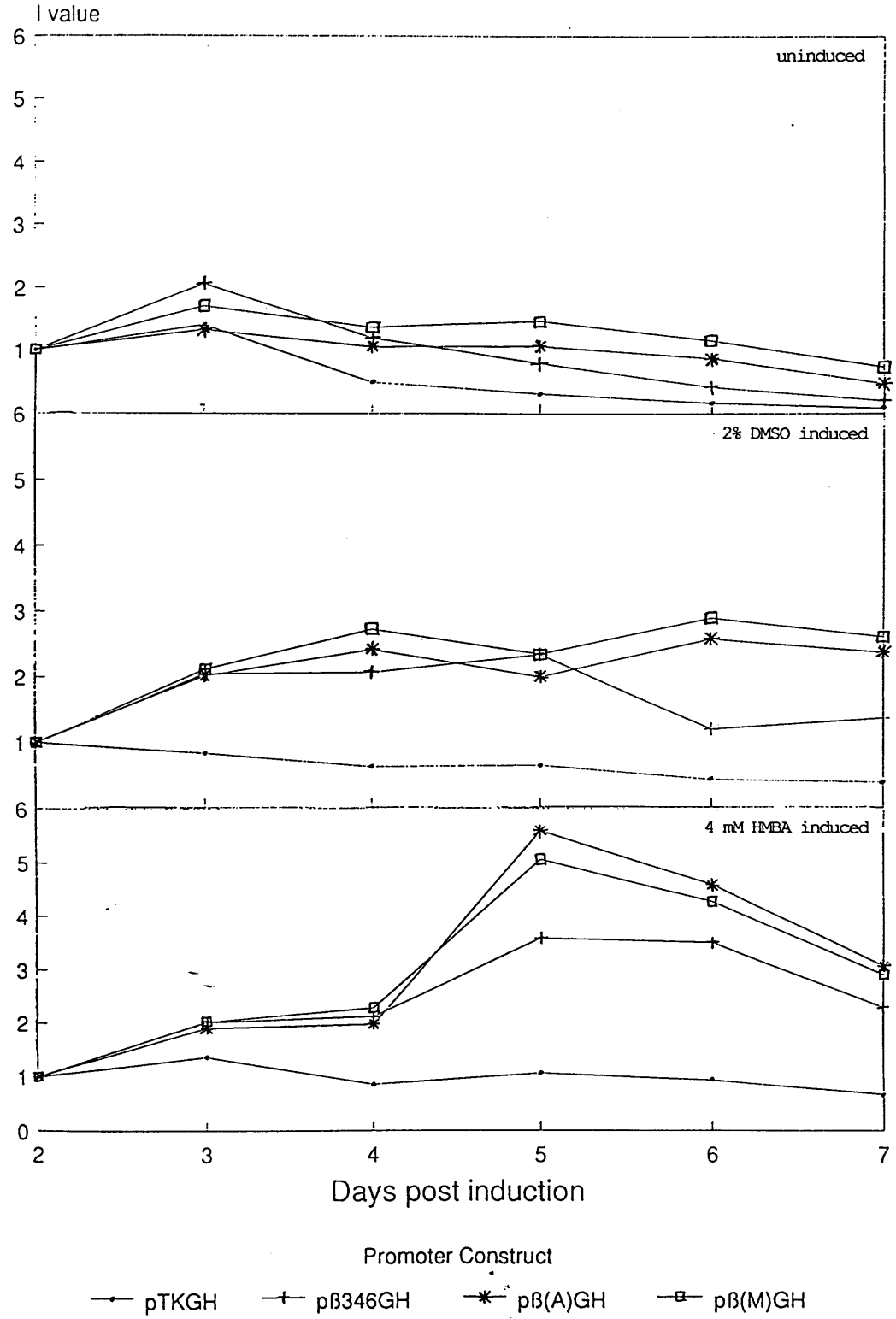


Figure 7.5A C88 hGH Production.

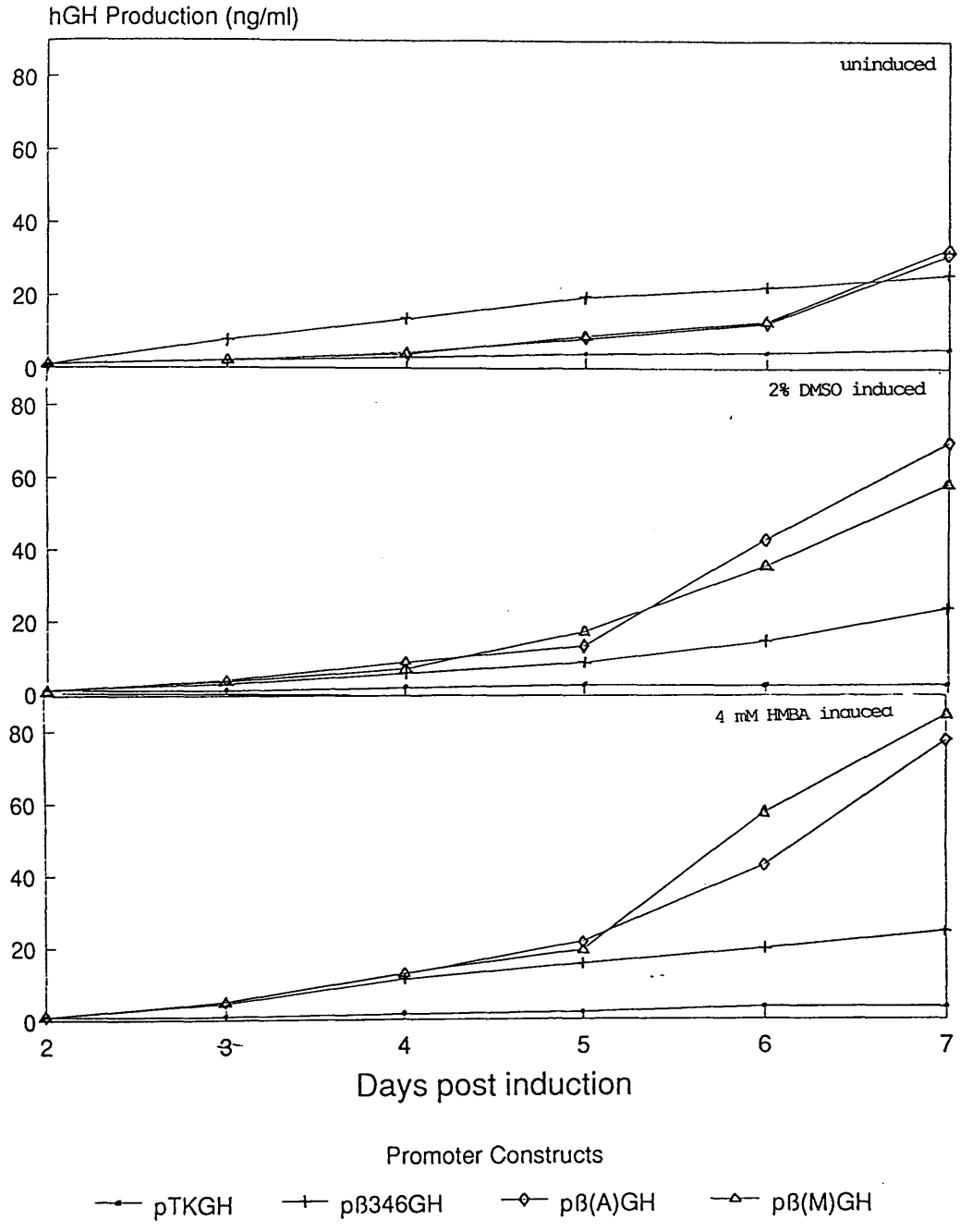
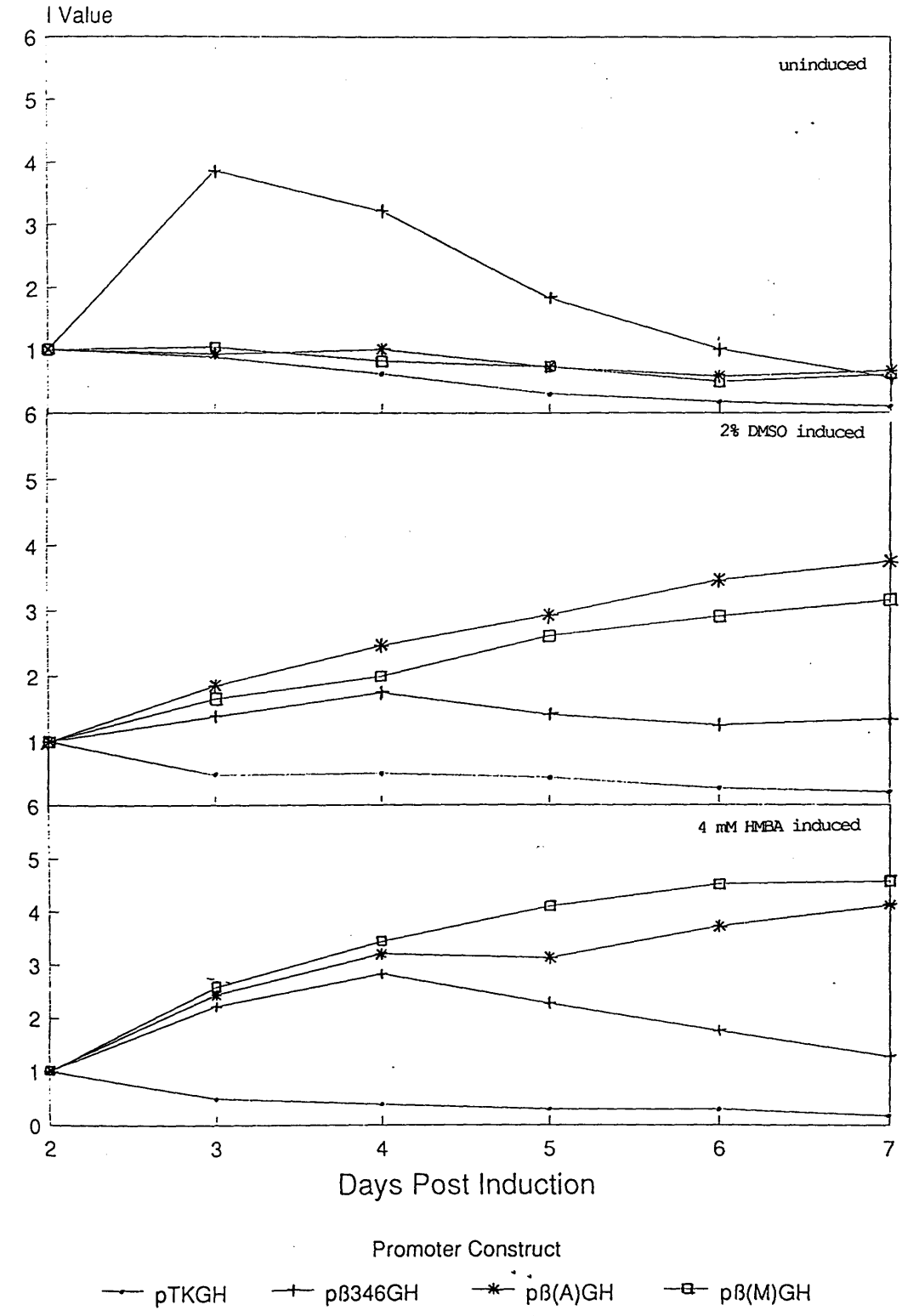


Figure 7.5B C88 I Values.



8.1 Summary of Results.

The results presented demonstrate that transcriptional regulation of the mouse β^{major} globin gene shares features with the regulation of other globin genes but that there are additional features to its regulation that have not been previously described. Before these features are discussed, I shall summarise the results which have been presented.

In vitro DNAase 1 footprinting revealed that elements in the proximal promoter region, TATA, CCAAT, CACCC, known to be essential for transcription, fail to bind their cognate DNA binding factors present in MEL cell crude nuclear extracts. However, when these extracts are partially purified, footprints over these sites are observed. In contrast, sites within the distal promoter region did bind their cognate binding factors in footprint experiments when incubated with crude nuclear extracts. These distal promoter sites had not been previously described and sequence analysis and competition studies revealed that these sites bound an NF-1 like factor (-249 bp) and the erythroid specific factor NF-E1 (-220 bp). A third site, known as Box 1 (-165 bp), and identified as being potentially important on the basis of sequence homology to the Box 1 sequence of the human β globin gene (Antoniou et al (1988) failed to bind factor in crude nuclear extracts. Sequences further upstream contain a second NF-E1 binding site (-550 bp) and a Box 1 site (-650 bp) which do bind in footprint experiments, factors present in MEL cell nuclear extracts but not those of STO or mouse kidney. Sequences 3′ of the mouse β^{major} globin gene coding region were also shown to contain NF-E1 and Box 1 binding site.

Gel retardation assays were used to further characterise the factors binding to sites present in the β^{major} globin gene 5′ or 3′ flanking sequence. This revealed that the NF-E1 binding site present in the distal promoter has a greater affinity for NF-E1 factor than the sites present in the chicken β^{A} globin gene enhancer (Emerson et al (1987), the mouse β^{major} globin gene intron 2 (Galson & Housman (1988), or the NF-E1 site in the proximal region of the β^{major} globin gene. Evidence was presented which suggested that NF-E1 binding activity is altered by phosphorylation. Band shift assays also revealed that there are two CCAAT box binding activities which show different tissue specificities : βCT1, is competed out by the sequence recognised by the factor αCP1 while binding by βCT2 is partially competed out by the sequence recognised by βNF1. However neither βNF1 or NF1 binding is competed out by βCT competitor. Binding to the Box 1 recognition sequence is competed out by binding to βCT. Both the Box 1 and βCT2 binding activities are present in uninduced MEL cell crude nuclear extracts and absent in induced MEL cell extracts. There are two CACCC box binding activities; one of which is present in all nuclear extracts looked at and the second of which is present only in induced MEL cell, K562 and J774.2 nuclear extracts.

The relative importance in transcriptional regulation of sequences identified and characterised in protein binding studies was determined in transient and stable transfection assays. Transient assays revealed that site directed mutagenesis of elements in the distal promoter, βNF1, NF-E1 and Box 1, resulted in up-regulation of expression from the β346 promoter fragment and that mutation of the NF-E1 and CACCC site in the proximal promoter resulted in down-regulation. This suggests that the distal promoter element is a

negative regulator of transcriptional regulation. It was shown that there are sequences upstream of -346 bp (Box 1, NF-E1) which may be involved in upregulating transcription and sequences 3′ of the gene (Box 1, NF-E1) which are analogous to the 3′ enhancer of the chicken β^A globin gene.

Cotransfection of α-1 and β^{major} globin gene promoter constructs revealed promoter competition could be an artefact of the transfection system but could equally reflect genuine promoter competition occurring in vivo.

Induction to differentiate leads to upregulated transcription from the wild-type promoter but not from promoter constructs mutant in distal promoter sequences. The greatest effect of induction is observed with HMBA induced expression from the wild-type promoter, modulated either by the 3′ enhancer of the β^A globin gene or by sequences 3′ of the β^{major} globin gene itself. DMSO mediated induction is less efficient than that of HMBA.

Together these results introduce two new elements : 1. a potential silencer element in the distal promoter which down regulates transcription from the β^{major} globin gene promoter and which is derepressed upon induction to differentiate, and 2. a putative 3′ enhancer which up-regulates transcription and whose activity in addition to derepression of the distal promoter in uninduced MEL cells is necessary to achieve maximal expression following induction to differentiate.

8.2 Organisation of Regulatory Elements in the 5' and 3' Flanking Region of the Mouse β^{major} Globin Gene.

The modular nature of promoter and enhancer sequences has been extensively described for viral and cellular systems (Fromenthal et al (1988) and my analysis of the mouse β^{major} globin gene 5' and 3' flanking sequence has shown that it too is organised on a modular basis. Figure 1.5 illustrates the organisation of regulatory elements in a number of different globin genes and shows that the same types of element are present in their promoter regions but that the organisation of these elements within each promoter is unique.

The organisation of elements within the mouse β^{major} and the human β globin gene promoters are similar. Both promoters appear to be organised into distal and proximal regions: the distal region containing an NF-1 like binding activity at the 5' boundary, an adjacent high affinity NF-E1 binding site and a Box 1 site while the proximal promoter region contains the TATA box, CCAAT box, CACC box and a low affinity NF-E1 binding site.

The human promoter differs from the mouse promoter in that at high protein concentration binding to the NF-1 like sequence prevents binding to the adjacent NF-E1 site in MEL cells (de Boer et al (1988); this competition does not occur in the mouse promoter. However, DNAase1 footprinting of the human promoter with lower protein concentrations shows simultaneous binding to both sites suggesting that the competition effect observed at high protein concentrations is due to protein-protein interactions off the DNA and is not because binding to the NF-1 like sequence and the NF-E1 site is mutually exclusive.

The human promoter has two CACC boxes in its proximal

promoter region while the mouse gene has only one. It has been demonstrated that the CACC box cooperates with the NF-E1 box in the promoter of the human PBG-D gene (Walker et al pers.comm.) and with the GRE in the tryptophan oxidase gene promoter (Schule et al (1988b). The CACC box may be mediating the interaction of the distal promoter region with the proximal promoter region in both the mouse β^{major} and the human β globin genes by cooperating with the NF-E1 site in the distal promoter. It is not clear what function the extra CACC box of the human β promoter plays; it may be mediating interaction with the 3' enhancer of the human β globin gene. The single CACC box of the mouse β^{major} globin gene may mediate interaction with both the distal promoter, which in the case of the mouse gene has a negative regulatory effect, and with 3' sequences which have a positive regulatory effect. If these interactions are mutually exclusive, it suggests a mechanism by which the mouse β^{major} globin gene is transcriptionally regulated.

The proximal region of the β^{major} promoter fails to footprint with crude nuclear extracts from a number of different cell types but the same region does footprint with partially purified extracts. As discussed in section 3.1, this has also been observed for the chicken β^{H} globin gene and may be explained if the relative protein concentrations (5-10 mg/ml) in the crude extracts favours protein-protein interactions off the DNA which prevent binding to the DNA whereas the comparatively dilute protein concentrations (0.3 - 0.6 mg/ml) in the partially purified extracts may favour the formation of the DNA-bound complex. It is unlikely that failure to obtain a footprint with crude nuclear extracts over the proximal region is due to low binding affinity of the sites since in gel retardation assays the same sites bind strongly.

The CCAAT box of the mouse β^{major} globin gene binds two factors in gel retardation assays, βCT1 and βCT2. βCT1 binding activity can be competed out by the mouse α-1 globin gene CCAAT box sequence, αP2a, and is present in both uninduced and induced MEL cell crude nuclear extracts. βCT2 is absent from induced MEL cell crude nuclear extracts. One of these activities may therefore be a CCAAT displacement factor in a manner analogous to the displacement factors of the mouse α-1 globin gene (Barnhardt et al (1988) and the human $^{A}\gamma$ globin gene (Superti-Furga et al (1989). Given that βCT2 disappears upon induction, it is more likely to be that activity.

βCT2 is partially competed out by binding to the β^{major} globin gene NF-1 like sequence. However it is unlikely that the βCT2 binding activity is the same as the βNF1 binding activity since the latter is present in induced MEL cell nuclear extracts and binding to βNF1 cannot be competed out by binding to βCT.

βNF1	TTCGCCAAAAA
BOX 1	TAAGCCAGTGA
βCT	GGGGCCAATCT

βCT2 binding activity does not share its pattern of tissue specificity with BOX 1 binding activity either although they are both absent in induced MEL cell extracts and although Box 1 binding activity can be competed out by βCT. Thus, while the βCT, βNF1 and Box 1 activities share similar recognition sequences, it is unlikely that they are the same factors.

De Boer et al (1988) have shown that an NF-E1 site and the Box 1 sequence from the distal promoter region of the human β globin gene cooperate in the up-regulation of transcription upon induction to differentiate. Work presented here involving the mutation of the Box 1,

NF-E1 and NF-1 recognition sequences in the distal promoter of the mouse β^{major} globin gene shows that these elements play the same role in this promoter as Box 1 and NF-E1 do in the human β globin gene promoter. Both results imply that modulation of Box 1 activity is necessary but not sufficient for the up-regulation of transcription following induction to differentiate. In uninduced MEL cells, Box 1 in the distal promoter is part of a modular negative regulatory element and mutation of this sequence or induction to differentiate, overcomes this negative effect. As already mentioned Box 1 binding activity disappears upon induction of MEL cells. Thus induction may derepress transcription of the mouse β^{major} globin gene by altering Box 1 binding activity. In addition to the Box 1 binding site in the distal promoter region, there are also sites in the upstream region and in the 3′ region. What their role is in these different contexts, is not known.

The NF-E1 binding site in the distal region (-220 bp) of the mouse β^{major} globin gene has a different role from the NF-E1 site in the proximal region as shown by the effect on transcription of mutating these different sites. Mutating the distal NF-E1 site has an up-regulatory effect on transcription whereas mutating the proximal site has a down-regulatory effect in uninduced MEL cells. The two elements differ in their affinities for NF-E1, the distal element has a higher affinity than the proximal element and has an equivalent affinity to the NF-E1 site in the mouse α-1 globin gene promoter (Plumb et al (1989). There are also NF-E1 binding sites in the second intron of the gene and in the 3′ region. Consequently there are four NF-E1 sites which all differ in their affinity for NF-E1.

$$G215 > \beta^{A}G > \beta^{IVS2} > G60$$

From the functional work presented here, the role of the NF-E1 sites in

the promoter is clear: the distal NF-E1 site is part of a negative regulatory element which represses transcription prior to induction, while the proximal NF-E1 is part of the basal transcription complex. Ptashne has proposed that differential usage of regulatory elements which differ in affinity for the same factor may be an important way of regulating a gene. When a given factor is in low abundance or modified such that a small proportion of available molecules can bind DNA, only the high affinity binding sites will be bound. As the factor becomes more abundant or its protein-protein interactions alter, the other low affinity binding sites will be bound. The roles of different affinity sites is likely to be important in determining which gene in a locus is to be expressed at a given time during development. Choi & Engel (1988) first suggested the idea of stage-selector elements from their work on the chicken β globin gene locus where it was shown that the chicken embryonic (ε) and adult (β^A) gene promoters compete with each other for productive interaction with the chicken β globin gene enhancer.

The chicken β globin gene enhancer contains four protein binding regions, I - IV. Regions I and III can be mutated and the function of the enhancer maintained. However, both regions II and IV are necessary for function of the enhancer and mutation of either of these regions reduces the function of the enhancer. Region II contains an AP-1 and an AP-2 binding site and mutation of either site is as effective in down-regulating as mutating both sites suggesting that factors bound at these sites are interacting cooperatively. Region IV of the enhancer contains an inverted repeat of the NF-E1 binding site and while mutation of either NF-E1 site reduces the function of the enhancer, mutation of both NF-E1 sites reduces the activity of the enhancer to a greater extent. Thus, unlike region II, the enhansons of

region IV do not interact cooperatively (Reitman & Felsenfeld (1988).

The mouse β^{major} globin gene 3′ flanking sequence contains five putative NF-E1 binding sites, one Box 1 site, an AP-1 site and an AP2 site as described in figure 3.8B. However, DNAase 1 footprinting as shown in figure 3.9, reveals binding only to one of the NF-E1 sites and to the Box 1 site. Further footprinting needs to be carried out since it is possible that improved footprinting conditions may reveal more extensive binding to this region. The functional assays described in chapters 6 and 7 certainly indicate that this region 3′ of the coding region is important in regulating expression of the mouse β^{major} globin gene. At the level of DNA sequence there are great deal of similarities between the chicken and the mouse enhancers: they both possess an inverted repeat of the NF-E1 binding sites and adjacent AP1 and AP2 sites. It is interesting to note that Reitman and Felsenfeld (1988) mention without describing a negative regulatory element that they have identified 3′ of the chicken β globin enhancer. This element may be analogous to the negative regulatory element found in the distal region of the mouse β^{major} globin gene. It may also suggest that the region that I have identified 3′ of the mouse β^{major} globin gene may include negative regulatory elements which when deleted from this region generate an enhancer which has greater activity than that measured in the assays described in chapters 6 and 7.

Unpublished work on the dominant control regions (DCRs) of the human β globin gene has shown that they contain multiple NF-E1 binding sites which may be involved in setting up an active globin chromatin domain in erythroid tissue prior to the expression of specific globin genes. These sites may also be involved in maintaining an open chromatin domain and in interacting with sites within the

promoter and enhancer regions of individual genes of the domain during development (Fraser et al pers. comm.).

Evidence is presented which demonstrates that phosphorylation of NF-E1 alters its DNA binding activity. Dephosphorylation results in the loss of a higher molecular weight binding form which may be either a homodimer or a heterodimer of the NF-E1 protein and some other unknown activity. Clearly phosphorylation does not affect DNA binding since the dephosphorylated low molecular weight form is still active in this respect. It is possible therefore that phosphorylation is modulating the activator domain of the protein by altering the interaction of DNA-bound NF-E1 with other factors to generate an altered conformation of transcription complex. It will be interesting to determine whether this phosphorylation of NF-E1 occurs in vivo and whether it correlates with response to an extracellular signal transmitted through the signal transduction pathway of cells inducing them to differentiate.

8.3 Negative Regulation of Transcription.

Mutating the Box1, NF-E1 and the βNF1 sites in the distal promoter such that they no longer bind protein, derepresses transcription and these constitutively active promoters do not then respond to induction. That mutation of any of the three elements has this same effect suggests that they interact with each other to form a complex which blocks transcription by some as yet unknown mechanism.

Until recently negative regulation was not considered to be a major feature of transcriptional regulation in higher eukaryotes unlike

the situation in prokaryotes and yeast. However, recent work has revealed that negative regulation is a more commonly employed mechanism than was thought.

Mutation and deletion of sequences flanking the immunoglobulin heavy chain enhancer have identified two distinct negative regulatory elements which lead to up-regulation of the heavy chain gene in non-lymphoid cells and down-regulation in lymphoid cells (Kadesch et al (1986), Wasylyk & Wasylyk (1986), Imler et al (1987), Schuermann & Chen (1989). Transcriptional stimulation by such a mutated enhancer transiently transfected into non-lymphoid cells was weaker when transfection efficiency was low and it was suggested that negative regulatory factors were being titrated out at higher efficiencies (Wasylyk & Wasylyk (1987). The factor, which binds to these negative regulatory sequences, has been identified: NF-μNR (Schuermann & Chen (1989). NF-μNR binding activity is present in immature B-cells, T-cells, macrophage and fibroblast but is absent from mature B-cells; that is, NF-μNR is absent from cells which are expressing the immunoglobulin heavy chain. The function of the immunoglobulin enhancer can be restored in non-expressing cells by deletion of the NF-μNR binding sites. This suggests that the developmental and tissue specificity of the immunoglobulin enhancer is in part determined by negative regulation in non-expressing cells. It is suggested that there are three types of enhanson in the immunoglobulin enhancer:

1. sites like μE 1-4, which are necessary for functional activity in any cell type and which have limited functional redundancy,
2. the octabox, which is necessary for positive activation in the appropriate cell-type, and
3. flanking NF-μNR sequences, which repress enhancer function in

inappropriate cell types.

Negative regulation is dominant to positive regulation since fusion of mature B-cells with non-B-cells extinguishes immmunoglobulin enhancer activated transcription (Junker et al (1988). Zaller et al (1988) demonstrated that immunoglobulin heavy chain expression in myeloma cell lines was suppressed after fusion with a T lymphoma or fibroblast cell line and that the immunoglobulin enhancer suppressed expression in T-cells, from a heterologous promoter which is normally expressed in T-cells.

Transcriptional silencer elements have been identified in the flanking sequence of several genes, including the rat insulin-I gene where a silencer element was identified 2-4 kb downstream (Laimins et al (1986), Nir et al (1986) but the reporting of a transcriptional silencer in the 5′ flanking region of the human ε globin gene was of particular interest since it represented a precedent for the use of negative regulation in the regulation globin genes. Cao et al (1989) identified a region from -177 to -392 bp of the human ε globin gene which had a negative regulatory effect on expression of a linked CAT reporter gene in transient transfection assays in K562 and HeLa cells. The repression was 3 fold greater in HeLa cells than in K562 and the element was shown to operate in a position and orientation independent manner and on a heterologous promoter. Sequence homology to the silencer of the chicken lysozyme gene was made but no homology is apparent between this region and the distal promoter of the mouse β^{major} globin gene. It would be interesting to determine whether the distal promoter of the β^{major} gene can be taken out of context and shown to operate in an orientation independent fashion and on heterologous promoters.

Basal expression of the chicken erythroid specific histone H5 gene is under negative control in addition to positive and the upstream negative regulatory element has been mapped to a position -95 to -115 bp upstream of the gene (Rousseau et al (1989). The negative regulatory element can down-regulate heterologous promoters and is non-cell type specific suggesting a role for repression in non-erythroid tissue. The negative regulatory element contains a polypyrimidine tract and an indirect repeat of an element AGGCA, which bears homology to the Box 1 sequence of the mouse β^{major} globin gene distal promoter negative regulatory element but also to the Box 1 sequence in the 3' enhancer of the same gene.

Interestingly, site 3 of the human β-globin gene 5' DCR contains, in addition to positive regulatory elements, a negative regulatory region made up of NF-E1 sites and other unidentified elements (Fraser pers.comm.). Again this illustrates that the NF-E1 site can act in a negative or a positive regulatory role depending on its context. What the role of a negative regulatory element in the DCR is unclear.

Cell fusion experiments involving the fusion of globin gene expressing human K562 cells and a non-erythroid myeloma cell line, (P3HR-1), demonstrated that globin gene expression was not suppressed by factors in the non-erythroid cell line and Grosveld et al (1988) suggested that this is because of the dominance of the dominant control regions at the human β-globin gene locus boundaries. However, fusion of adherent fibroblast cell lines with suspension Friend cell lines resulted in two types of hybrid: suspension hybrids, which could be induced to synthesise haemoglobin, and adherent hybrids which could not (Allan et al (1980). Meanwhile, work demonstrating that erythroid

differentiation in a Friend erythroleukaemic x lymphoma hybrid cell line is limited suggested that the mechanism of suppression was post-transcriptional since globin mRNA was inducible although haemoglobin synthesis was not taking place.

The use of cell fusion experiments to identify negative regulatory elements is exemplified in work by Poole et al (pers.comm.) who have correlated the suppression of the transformed phenotype in RSV transformed rat embryo fibroblast x human mononucleocyte hybrids with down regulation of proviral transcription which is in turn determined by the site of retroviral integration. It is proposed that DNA sequences flanking the provirus are exerting a negative regulatory effect on proviral transcription and retroviral probes may be used to identify cellular regulatory elements and trans-acting factors that have not been identified by other means.

Fusion of non-growth hormone producing mouse L cells with growth hormone producing rat pituitary GH3 cells, resulted in three hybrids (McCormick et al (1988). Two of the hybrids were non-producers of growth hormone and one hybrid was a producer but only after mouse chromosomal loss. The extinction of growth hormone expression was associated with loss of trans activator, GHF-1 in the hybrids.

Viral genomes, such as SV40 and MuLV are not efficiently expressed in undifferentiated embryonal carcinoma cells (F9), but are in the differentiated derivatives. However, when SV40 is transfected into undifferentiated F9 cells, expression of viral genes is observed and it is proposed that the higher copy number of viral genome introduced by transfection is titrating out negative regulatory factors (Gorman et al (1985). The MSV LTR promoter does not operate in undifferentiated F9 cells although when cotransfected with the RSV LTR,

MSV LTR promoter activity is observed. Again it is suggested that the RSV LTR is competing with the MSV LTR for negative regulatory factors. It was observed that mutants of adenovirus which were defective in E1a function, cannot replicate in differentiated F9 cells but are able to do so in undifferentiated cells suggesting that undifferentiated F9 cells have a cellular homologue of E1a which down-regulates viral enhancers and is involved in the maintenance of the undifferentiated state. Enhansons of the SV40 enhancer are individually repressed by E1a and it has been suggested that E1a mediated phosphorylation of transcription factors may be responsible for their repression (Rochette-Egly et al (1990).

Deletion of the 3′ end of the β-interferon enhancer results in up-regulation of transcription of the human β-interferon gene (Goodbourn et al (1986). Negative regulation of the human β-interferon gene is relieved by induction of fibroblasts with virus or double stranded RNA and characterisation of the changes in DNA-protein and protein-protein interactions as transcription of the gene is induced reveals features which may be shared by other inducible genes, such as the β-globin genes in the MEL cell system. The factors which bind to the β-interferon enhancer have been identified (Zinn et al (1986), Goodbourn & Maniatis (1988) and the pattern of binding to the enhancer changes when transcription of the gene is induced. In uninduced cells, a positive regulatory factor is bound, PRD I which allows basal levels of expression. A negative regulatory factor, NRD I is also bound and prevents binding by PRD II, which is necessary for induced levels of expression. Following induction, NRD I binding is inactivated to allow PRD II to bind and PRD I binding activity increases. The binding sites for NRD I and PRD II overlap and thus, negative regulation of

β-interferon can be explained in terms of competition between transcription factors for DNA-binding sites. However, negative regulation has been shown to operate using non-overlapping DNA-binding sites which are effective in repression even when placed at a distance upstream or downstream of a gene. This is the case for the negative regulation of "a" type genes in α-type cells in yeast (Johnson & Herskowitz (1985) where the MAT α2 gene product, α2 interacts cooperatively with the GRM protein, to repress transcription from the operator site in the promoters of "a" type genes (Keleher et al (1988). In "a" type cells, GRM activates "a" type genes and in α type cells, α2 activates α type genes. Two models for repression were proposed: the masking model, where α2 blocks contact between GRM and the transcription machinery, and the locking model, where the α2/GRM complex interacts too strongly with the transcription machinery.

The negative regulatory activity of Fos and Jun has been described for some genes. In vitro translated Fos and Jun bind cooperatively to the negative regulatory element of the human c-myc promoter (Hay et al (1989). Fos trans-represses c-fos transcription and this repression appears to be mediated by the serum response element since repression can occur in the absence of the TPA response element and the cAMP response element of the c-fos promoter. The leucine zipper of c-fos is necessary for its repressor activity and although the serum response factor does not have a leucine zipper, it is suggested that a Fos-Jun complex may be modifying an already existing SRE bound complex (Lucibello et al (1989).

Jun-B, c-jun and jun-D all bind to the TRE as Jun homodimers or as Fos-Jun heterodimers but while c-Jun activates, Jun-B inhibits activation of promoters containing single TREs. Promoters containing

multiple TREs are activated by both c-Jun and Jun-B suggesting that Jun-B requires cooperative interactions between adjacently bound factors in order to activate transcription (Chui *et al* (1989). Cotransfection of Jun-B and c-Jun into fibroblast cell lines decreases AP-1 *trans*-activation and it has been suggested that the Jun-B/c-Jun heterodimer is inactive since cotransfection of c-Jun and Fos does not have this effect. It is also possible that Jun-B and c-Jun are competing for AP-1 binding sites. The transformation activity of Jun-B is less than that of c-Jun and it cooperates less well with Ras to transform rat fibroblasts and cotransfection of *jun*-B with *ras* and c-*jun* results in decreased transformation compared with *ras* and c-*jun* alone (Schutte *et al* (1989). How the inhibitory activity of Jun-B in transcription correlates with its reduced transformation capacity remains to be determined.

Negative regulation is widely used in the regulation of developmental genes of *Drosophila* where the products of the *engrailed* and *even-skipped* genes down-regulate the up-regulatory effect of the *fushi-tarazu* and *paired* gene products. The *ultrabithorax* gene is down-regulated by the *even-skipped* gene product but unlike the negative regulation of *fushi-tarazu* and *paired* which is mediated by protein-protein interactions, the negative regulation of *ultrabithorax* is effected by the binding of the *even-skipped* gene product to a site in the DNA 5′ of the *ultrabithorax* gene (Biggin & Tjian (1989).

These studies of eukaryotic systems has allowed a theory of repression to be proposed in which repression is postulated to occur by one of three mechanisms:

1. Competition between factors for DNA binding sites.
2. Alteration in the function of the activator domain.

3. Squelching.

All three mechanisms have the effect of preventing the formation of a functional transcription initiation complex. This complex can be visualised as an allosteric protein complex consisting of primary transcription factors which undergoes conformational change in response to the binding of secondary transcription factors which may be cell-specific or stage specific. The function of this complex is to interact with RNA polymerase II to initiate transcription in a tightly regulated fashion from the transcription initiation site. The complex will include factors bound to DNA at different sites, the DNA linking these sites, and non-DNA bound factors. The nature of the interaction with RNA polymerase is not as well characterised as the interactions of prokaryotic factors, such as λ repressor with prokaryotic RNA polymerase where the amino acid contacts with the major groove of the DNA binding site have been established (Pabo & Sauer (1984). The transcription complex may be modulated like other allosteric proteins; its activity altered by covalent modification or by interaction with inhibitory factors which block interaction with other factors or block conformational change. Whether a given factor is a repressor or an activator will depend upon the other factors in the initiation complex and its interaction with them.

8.4 The Effects of Induction to Differentiate.

Friend erythroleukaemic cell lines are used to study the transcriptional regulation of the mouse β^{major} globin gene. As discussed in the general introduction, this cell line and its

derivatives are arrested in the erythroid differentiation pathway and can be induced to continue differentiating by treatment with 4 mM HMBA or 2% DMSO amongst other agents. This method of induction is effective in inducing globin gene expression but since these agents are not used in vivo, this system and all others involving chemical induction are not optimal. To study the effect of induction to differentiate upon globin gene expression, it would be preferable to use a cell line which could be induced with erythropoietin. Such a cell line has recently been produced and any future work on induction will involve its use. The results presented in chapter 7 show that induced expression of the hGH reporter gene is observed from the wild-type promoter fragment β346 and higher levels are observed from this promoter when its activity is modulated by either the chicken β^A globin gene or by mouse 3' sequences. This effect is seen with induction by both 4 mM HMBA and 2% DMSO and for both Friend cell lines, F4B12 and C88. However the effect observed for HMBA induced cells is 3 fold greater than that observed for DMSO. Clearly there are differences in the way in which these agents operate. It has already been reported that HMBA is more efficient in inducing globin gene expression than DMSO. Maximal globin gene expression in MEL cells is observed after 2-3 days in HMBA but not until 3-4 days in DMSO and only 60% of cells become committed to differentiate with DMSO compared to 95% with HMBA (U.Nudel Cell 12 463-469 (1977). In this way it is possible that the general stress responses of the cells to these chemical agents could be distorting the changes in globin gene expression observed as the cells differentiate.

Another problem with analysis of data obtained from this system is the problem with standardisation to allow comparison with the regulation of expression of other genes measured in other systems.

Different laboratories use different cell lines, different inducing agents and different methods of analysing their results. This work compares two different cell lines and shows that the effect of induction to differentiate on globin gene expression varies depending upon the agent used. To analyse the results obtained here, expression levels in terms of hGH produced were first standardised for cell number since clearly uninduced and induced cells grow at different rates and then the expression level per cell was standardised to expression level per cell prior to induction. In this way the effect upon transcriptional activity of the mouse β^{major} globin gene promoter and flanking sequence could be analysed. It is difficult to compare these results with work done by some other groups since they have not made it clear how they standardised their results.

The results obtained following induction of MEL cells stably transfected with different promoter and enhancer constructs shows that mutation of any of the three elements in the distal promoter results in constitutive expression from the promoter that does not respond to induction but that induction generates a 2-3.5 fold increase in expression from the wild-type promoter and a 2-6 fold increase from the wild-type promoter modulated by chicken β^{A} enhancer sequences or by mouse β^{major} 3′ flanking sequence. These results show the presence of a "silencer" in the distal promoter of the gene which down-regulates transcription and whose effect is overcome by mutation or by induction to differentiate. These results and the results obtained from transient transfection assays show that the sequences from 3′ of the mouse β^{major} globin gene used in these assays effect the same level of enhancement upon expression as the chicken β^{A} enhancer. It is necessary to show that the activity of these sequences is position and

orientation independent and that they act on promoters other than that of the mouse β^{major} globin gene. The activity of a 3′ "enhancer" sequence is required to achieve optimal levels of expression. It was suggested earlier that the CACC box in the proximal region of the promoter may be interacting with both the "silencer" and the "enhancer" in a mutually exclusive fashion and that this suggested a mechanism by which the mouse β^{major} globin gene was regulated. Early in differentiation and in development when the mouse β^{major} globin gene is not yet expressed transcription factor bound at the CACC box interacts with EF1, Box 1 and βNF1 in the distal promoter region which results in repression of expression. As the erythroid cell differentiates, the activity of the Box 1 binding factor decreases such that the distal promoter repressor complex cannot form. The CACC box factor then cooperates with the proximal promoter - 3′ enhancer complex to up-regulate transcription from the mouse β^{major} globin gene promoter. This mechanism is purely speculative but work with the chicken β^{A} globin gene promoter has revealed a central role for the CACC element in developmental regulation through its interactions with other elements in the promoter (PAL) and in the enhancer (NF-E4).

Allan, M. & Harrison, P.R. (1980). Co-expression of differentiation markers in hybrids between Friend cells and lymphoid cells and the influence of the cell shape. Cell 19 437-447.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. (1981). Structure of the cro repressor from bacteriophage λ and its interactions with DNA. Nature 290 754-758.

Andersson, P., Goldfarb, M.P. & Weinberg, R.A. (1979). A defined subgenomic fragment of in vitro synthesised Moloney sarcoma virus DNA can induce cell transformation upon transfection. Cell 16 63-75.

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

Balberis, A., Superti-Furga, G. & Busslinger,M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. Cell 50 347-359.

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

Barnhart, K.M., Kim, C.G., Banerji, S.S., & Sheffery, M. (1988). Identification and Characterisation of multiple erythroid cell proteins

that interact with the promoter of the murine α-globin gene. Mol.Cell Biol. 8 3215-3226.

Baron, M. & Maniatis, T. (1986). Rapid reprogramming of globin gene expression in transient heterokaryons. Cell 46 591-602.

Barrowman, J. & Craig, M. (1961). Haemoglobins of foetal C57BL/6 mice. Nature 190 818-819.

Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. & Townes, T.M. (1987). Two 3′ sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc.Natl.Acad.Sci.USA 84 7056-7060.

Benezra, R., Cantor, C.R. & Axel, R. (1986). Nucleosomes are phased along the mouse β-major globin gene in erythroid and nonerythroid cells. Cell 44 697-704.

Benoist, C. & Chambon, P. (1980). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. Proc.Natl.Acad.Sci.USA 77 3865-3869 (1980).

Berg, J.M. (1986). Potential metal-binding domains in nucleic acid binding proteins. Science 232 485-487.

Berk, A.J. (1986). Adenovirus promoters and E1A transactivation. Ann.Rev.Gen. 20 45-79.

Bernstine, E.G., Hooper, M.L., Grandchamp, S. & Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. Proc.Natl.Acad.Sci. USA 70 3899-3903.

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

Biggin, M.D. & Tjian, R. (1989). A purified Drosophila homeodomain protein represses transcription in vitro. Cell 58 433-440.

Bird, A., Taggart,M., Frommer, M., Miller, O.J. & Macleod, D. (1985). A Fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. Cell 40 91-99.

Bird, A.P. (1986). CpG-rich islands and the function of methylation. Nature 321 209-213.

Birkenmeier, E.H., Gwynn, B., Howard, S., Terry, J., Gordon, J.I., Landschulz, W.H. & McKnight, S. (1989). Tissue-specific expression, developmental regulation and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes & Dev. 3 1146-1156.

Bodine, D.M. & Ley, T.J. (1987). An enhancer element lies 3′ to the human $^{A}\gamma$ globin gene. EMBO J. 6 2997-3004.

Boyer, S.H., Siegel, S. & Noyes A.N. (1983). Developmental changes in human erythrocyte carbonic anhydrase levels: coordinate expression with adult hemoglobin. Dev. Biol, 97 250-253.

Brent, R. & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43 729-736.

Briggs, M.R., Kadonaga, J.T. Bell, S.P. & Tjian, R. (1986). Purification and biochemical characterisation of the promoter-specific transcription factor, Sp1. Science 234 47-52.

Brown, D.D. (1984). The role of stable complexes that repress and activate eukaryotic genes. Cell 37 359-365.

Broxmeyer, H.E., Lu, L., Cooper, S., Schwall, R.H., Mason, A.J. & Nikolics, K. (1988). Selective and indirect modulation of human multipotential and erythroid hematopoietic progenitor cell proliferation by recombinant human activin and inhibin. Proc.Natl.Acad.Sci.USA 85 9052-9056.

Buratowski, S., Hahn, S., Sharp, P.A. & Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature 334 37-42.

Buratowski, S., Hahn, S., Guarente, L. & Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase

II. Cell 56 549-561.

Bushman, F.D. & Ptashne, M. (1988). Turning λ cro into a transcriptional activator. Cell 54 191-197.

Callan, H.G. & Macgregor, H.C. (1958). Action of deoxyribonucleases on lamprush chromosomes. Nature 181 1479-1480.

Cao, S.X., Gutman, P.D., Dave, P.G. & Schechter, A.N. (1989) Identification of a transcriptional silencer in the 5'-flanking region of the human ε-globin gene. Proc.Natl.Acad.Sci USA 86 5306-5309.

Cavallini, B., Huet, J., Plassat, J.L., Sentenac, A., Egly, J.M. & Chambon, P. (1988). A yeast activity can substitute for the HeLa cell TATA box factor. Nature 334 77-80.

Chao, M.V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. (1983). The regulated expression of β-globin genes introduced into mouse erythroleukemia cells. Cell 32 483-493.

Charache, S., Dover, G., Smith, K., Conover Talbot,Jnr., C., Moyer, M. & Boyer, S. (1983). Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the γ- δ- β-globin gene complex. Proc.Natl.Acad.Sci.USA 80 4842-4846.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. & Maniatis,

T. (1984). Differences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell 38 251-263.

Charnay, P., Mellon, P. & Maniatis, T. (1985). Linker scanning mutagenesis of the 5′ flanking region of the mouse β-major globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol.Cell Biol. 5 1498-1511.

Chen, C. & Okayama, H. (1987). High efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell Biol.7 2745-2752.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. & Sharp, P.A. (1988). Human CCAAT-binding proteins have heterologous subunits. Cell 53 11-24.

Choi, O.R. & Engel, J.D. (1986). A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. Nature 323 731-734.

Choi, O.R.B. & Engel, J.D. (1988). Developmental Regulation of β-globin Gene Switching. Cell 55 17-26.

Chui, R., Boyle, W.J., Meek, J., Smeal, T., Hunter, T. & Karin, M. (1988). The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. Cell 54 541-552.

Clark, S.C. & Kamen, R. (1987). The human hemopoietic

colony-stimulating factors. Science 236 1229-1237.

Clarke, J. & Herr, W. (1987). Activation of mutated simian virus 40 enhancers by amplification of wild-type enhancer elements. Mol.Cell Biol. 6 3536-3542.

Clerc, R.G., Corcoran, L.M., LeBowitz, J.H., Baltimore, D. & Sharp, P.A. (1988). The B-cell specific Oct-2 protein contains POU box- and homeo-box type domains. Genes Dev. 2 1570-1581.

Cohen, R.S. & Meselson, M. (1988). Periodic interactions of heat shock transcriptional elements. Nature 332 856-858.

Cohen, R.B., Sheffery, M. & Kim, C.G. (1986). Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α1-globin gene. Mol.Cell Biol. 6 821-832.

Conkie, D., Kleiman, L., Harrison, P.R. & Paul, J. (1975). Increase in the accumulation of globin mRNA in immature erythroblasts in response to erythropoietin in vivo or in vitro. Exp.Cell Res. 93 315-324.

Cook, P.R., Lang, J., Hayday, A., Lania, L., Fried, M., Chiswell, D.J. and Wyke, J.A. (1982). Active viral genes in transformed cells lie close to the nuclear cage. EMBO J. 1 447-452.

Courey, A.J. & Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich

activation motif. Cell 55 887-898.

Cowie, A. & Myers, R.M. (1988). DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol.Cell Biol. 8 3122-3128.

Craig, M.L. & Russell, E.S. (1964). A developmental change in hemoglobins correlated with an embryonic red cell population in the mouse. Dev.Biol. 10 191-201.

Dainiak, N., Hoffman, R., Maffei, L.A. & Forget, B.G. (1978). Potentiation of human erythropoiesis in vitro by thyroid hormone. Nature 272 260-262.

D'Andrea, A.D., Lodish, H.F. & Wong, G.G. (1989). Expression cloning of the murine erythropoietin receptor. Cell 57 277-285.

Danielson, M., Hinck, L. & Ringold, G.M. (1989). Two amino acids within the knuckle of the first zinc finger specify DNA response element activation by the glucocorticoid receptor. Cell 57 1131-1138.

Davidson, I., Xiao, J.H., Rosales, R., Staub, A. & Chambon, P. (1988). The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell 54 931-942.

deBoer, E., Antoniou, M. Mignotte, V., Wall, L. & Grosveld, F. (1988). The human β-globin promoter; nuclear protein factors and erythroid

specific induction of transcription. EMBO J. 7 4203-4212.

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., French Anderson, W., Ruddle, F., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977). Localisation of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridisation assay. Cell 12 205-218.

Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. & Ruddle, F., (1978). Chromosomal localisation of human β globin gene on human chromosome 11 in somatic cell hybrids. Proc.Natl.acad.Sci. USA 75 1456-1460.

Deisseroth, A. & Hendrick, D. (1978). Human α-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell 15 55-63.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32 695-706.

Dilworth, S.M. & Dingwall, C. (1988). Chromatin assembly in vitro and in vivo. Bioessays 9 44-49.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. & Mathis, D. (1987). A multiplicity of CCAAT box-binding proteins. Cell 50 863-872.

Dube, S., Fisher, J.W. & Powell, J.S. (1988). Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion and biological function. J. Biol.Chem. 263 17516-17521.

Dynan, W.S. & Tjian, R. (1983). Isolation of transcription factors that discriminate between differnet promoters recognised by RNA polymerase II. Cell 32 669-680.

Eissenberg, J.C., Cartwright, I.L., Thomas G.H. & Elgin, S.C.R. (1985). Selected Topics in Chromatin Structure. Ann.Rev.Biochem. 19 485-536.

Eisen, H., Bach, R. & Emery, R. (1977). Induction of spectrin in erythroleukemic cells transformed by Friend virus. Proc.Natl.Acad.Sci. USA 3898-3902.

Emerson, B.M. & Felsenfeld, G. (1984). Specific factor conferring nuclease hypersensitivity at the 5′ end of the chicken adult β globin gene. Proc.Natl.Acad.Sci. USA 95-99.

Emerson, B.M., Lewis, C.D. & Felsenfeld, G. (1985). Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: nature of the binding domain. Cell 41 21-30.

Emerson, B.M., Nickol, J., Jackson, P.D. & Felsenfeld, G. (1987). Analysis of the tissue-specific enhancer at the 3′ end of the chicken adult β-globin gene. Proc.Natl.Acad.Sci. USA 84 4786-4790.

Emerson, B.M., Nickol, J. & Fong, T.C. (1989). Erythroid-specific activation and derepression of the chick β-globin promoter in vitro. Cell 57 1189-1200.

Enver. T., Zhang, J.W., Anagnou, N.P., Stamatoyannapoulos, G. & Papayannopoulou, T. (1988). Developmental programs of human erythroleukemia cells: globin gene expression and methylation. Mol. Cell Biol. 8 4917-4926.

Enver. T., Zhang, J.W., Papayannopoulou, T. & Stamatoyannapoulos, G. (1988). DNA methylation: a secondary event in globin gene switching? Genes & Dev. 2 698-706.

Evans, R.M. & Hollenberg, S.M. (1988). Zinc Fingers: Gilt by Association. Cell 52 1-3.

Evans, T., Reitman, M. & Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognises a regulatory sequence common to all chicken globin genes. Proc.Natl.Acad.Sci. USA 85 5976-5980.

Evans, T. & Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: A new zinc finger protein. Cell 58 877-885.

Fairall, L., Rhodes, D. & Klug, A. (1986). Mapping of the sites of protection on a 5 S RNA Gene by the Xenopus transcription factor IIIA. A model for the interaction. J.Mol.Biol. 192 577-591.

Felsenfeld, G. (1978). Chromatin. Nature 271 115-122.

Fradin, A., Manley, J.L. & Prives, C.L. (1982). Methylation of simian virus 40 Hpa II site affects late, but not early viral gene expression. Proc.Natl.Acad.Sci. USA 79 5142-5146.

Fraser, P.J. & Curtis, P.J. (1987). Specific pattern of gene expression during induction of mouse erythroleukemia cells. Genes Dev. 1 855-861.

Fraser, P.J., Cummings, P. & Curtis, P.J. (1989). The mouse carbonic anhydrase I gene contains two tissue-specific promoters. Mol. Cell Biol. 9 3308-3313.

Friedman, A.D., Landschulz, W.H. & 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.

Friend, C. (1956). Cell-free transmission in adult Swiss mice of a disease having the character of a leukaemia. J.Exp.Med. 105 307-325.

Friend, C., Patuleia, M.C. & de Harven, E. (1966). Erythrocytic Maturation in vitro of murine (Friend) virus-induced leukemic cells. National Cancer Institute Monograph 22 505-522.

Friend, C., Scher, W., Holland, J.G. & Sato, T. (1971). Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation

of erythroid differentiation by dimethyl sulfoxide. Proc.Natl.Acad.Sci. USA 68 378-382.

Fromenthal, C., Kanno, M., Nomiyama, H. & Chambon, P. (1988). Cooperativity and hierarchical levels of functional organisation in the SV40 Enhancer. Cell 54 943-953 .

Gallarda, J.L., Foley, K.P., Yang, Z. & Engel, J.D. (1989). The β-globin stage selector element factor is erythroid -specific promoter/enhancer binding protein NF-E4. Genes Dev. 3 1845-1859.

Galson, D.L. & Housman, D.E. (1988). Detection of two tissue-specific DNA-binding proteins with affinity for sites in the mouse β-globin intervening sequence 2. Mol.Cell Biol.8 381-392.

Garner, M.M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions : applications to components of the Escherichia coli lactose operon system. Nucl.Acids Res. 9 3047-3060.

Garner, M.M. & felsenfeld, G. (1987). Effect of Z-DNA on nucleosome placement. J. Mol. Biol. 196 581-590.

Gentz, R., Rauscher III, F.J., Abate, C. & Curran, T. (1989). Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. Science 242 1695-1699.

Gey, G.O., Coffman, W.D. & Kubcek, M.T. (1952). Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. Cancer Res. 12 264-265.

Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. & Tjian,R. (1985). Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science 230 511-517.

Goodbourn, S., Burstein, H. & Maniatis, T. (1986). The human β-interferon gene enhancer is under negative control. Cell 45 601-610.

Goodbourn, S. & Maniatis, T. (1988). Overlapping positive and negative regulatory domains of the human β-interferon gene. Proc.Natl.Acad.Sci. USA 85 1447-1451.

Goodwin, G.H., Nicolas, R.H., Cockerill, P.N., Zavou, S. & Wright, C.A. (1985). The effect of salt extraction on the structure of transcriptionally active genes; evidence for a DNAase I sensitive structure which could be dependent on chromatin structure at levels higher than the 30 nm fibre. Nucl.Acids Res. 13 3561-3579.

Golay, J., Introna, M. & Graf, T. (1988). A single point mutation in the v-ets oncogene affects both erythroid and myelomonocytic cell differentiation. Cell 55 1147-1158.

Goldman, M.A. (1988). The chromatin domain as a unit of gene regulation. Bioessays 9 50-55.

Gorman, C.M., Moffat, L.F. & Howard, B.H. (1982). Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell Biol. 2 1044-1051.

Gorman, C., Rigby, P. & Lane, D. (1985). Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell 42 519-526.

Grandchamp, B., De Verneuil, Beaumont, C., Chretien, S., Walter, O. & Nordmann, Y. (1987). Tissue-specific expression of porphobilinogen deaminase. Eur.J. Biochem. 162 105-110.

Greaves, D.R., Wilson, F.D., Lang, G. & Kioussis, D. (1989). Human CD2 3'-flanking sequences confer high-level, T-cell specific, position independent gene expression in transgenic mice. 56 979-986.

Green, M., Loewenstein, P.M., Pusztai, R. & Symington,J.S. (1988). An adenovirus E1A protein domain activates transcription in vivo and in vitro in the absence of protein synthesis. Cell 53 921-926.

Green, M.R., Treisman, R. & Maniatis, T. (1983). Transcriptional activation of cloned human β-globin gene by viral immmediate early gene products. Cell 35 137-148.

Greenberg, M.E. & Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature 311 433-438.

Gregory, C.J. and Eaves, A.C. (1977). Human marrow cells capable of erythropoietic differentiation in vitro: definition of three erythroid colony responses. Blood 49 855-864.

Grosschedl, R. & Baltimore, D. (1985). Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell 41 885-897.

Grosveld, F., Van Assendelft, G.B., Greaves, D.R. & Kollias, G. (1987). Position independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51 975-985.

Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, g. & Papayannopoulou, T. (1983). Human fetal to adult hemoglobin switching: Changes in chromatin structure of the β-globin gene locus. Proc.Natl.Acad.Sci. USA 80 7551-755.

Gusella, J.F. & Housman, D. (1976). Induction of erythroid differentiation in vitro by purines and purine analogues. Cell 8 263-269.

Hai, T., Liu, F., Coukos, W.J. & Green, M.R. (1989). Transcription factor ATF cDNA clones : an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. 3 2083-2090.

Halazonetis, T.D., Georgopoulos,K., Greenberg, M.E. & Leder, P. (1988).

cJun dimerises with itself and with c-Fos, forming complexes of different DNA binding affinities. Cell 55 917-924.

Hall, C.V., Jacob, P.E., Ringold, .M. & Lee, F. (1983). Expression and Regulation of Escherichia coli lac Z Gene Fusions in Mammalian Cells. J. of Molecular and Applied Genetics 2 101-107.

Han, K., Levine, M.S. & Manley, J.L. (1989). Synergistic activation and repression of transcription by Drosophila homeobox proteins. Cell 56 573-583.

Hankins, W.D. & Krantz, S.B. (1975). In vitro expression of erythroid differentiation induced by Friend polycythaemia virus. Nature 253 731-732.

Harrison, P.R., Rutherford, T., Conkie, D., Affara, N., Sommerville, J., Hissey, P. & Paul, J. (1978). Analysis of erythroid differentiation in Friend cells using non-inducible variants. Cell 14 61-70.

Harrison, P.R. (1984). Molecular Analysis of Erythropoiesis. Exp.Cell Res. 155 321-344.

Hay, N., Takimoto, M. & Bishop, J.M. (1989). A FOS protein is present in a complex that binds a negative regulator of MYC. Genes Dev. 3 293-303.

Hayman, M.J., Kitchener, G., Vogt, P.K. and Beug, H. (1985). The

putative transforming protein of S13 avian erythroblastosis virus is a transmembrane glycoprotein with an associated protein kinase activity. Proc.Natl.Acad.Sci. USA 82 8237 8241.

Hebbes, T.R., Thorne, A.W. & Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. 7 1395-1402.

Hentzen, D. , Renucci, A., Le Guellec, D. Benchaibi, M., Jurdic, P., Gandrillon, O. & Samarut, J. (1987). The chicken erb-A proto-oncogene is preferentially expressed in erythrocytic cells during late stages of differentiation. Mol.Cell Biol. 7 2416-2424.

Herr, W. & Clarke, J. (1986). The SV40 enhancer is composed of multiple functional elements that can compensate for one another. Cell 45 461-470.

Hesse, J.E., Nickol, J.M., Lieber, M.R. & Felsenfeld, G. (1986). Regulated gene expression in transfected primary chicken erythrocytes. Proc.Natl.Acad.Sci. USA 83 4312-4316

Higuchi, R., Krummel. B. & Saiki, R.K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucl.Acid Res. 16 7357-7373.

Hoeffler, W.K., Kovelman, R. & Roeder, R.G. (1988). Activation of transcription factor IIIC by the adenovirus E1A protein. Cell 53

907-920.

Hollenberg, S.M., Giguere, V., Segui, P. & Evans, R.M. (1987). Colocalisation of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. Cell 49 39-46.

Hollenberg, S.M. & Evans, R.M. (1988). Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. Cell 55 899-906.

Horikoshi, M., Carey, M.F., Kakidani, H. & Roeder, R,G. (1988). Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. Cell 54 665-669.

Horikoshi, M., Hai, T., Lin, Y.S., Green, M.R. & Roeder, R.G. (1988). Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. Cell 54 1033-1042.

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. & Roeder, R.G. (1989). Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature 341 299-303.

Humphries, R.K., Ley, T., Turner, P., Moulton, A.D. & Nienhuis, A.W. (1982). Differences in human α-, β- and δ-globin gene expression in monkey kidney cells. Cell 30 173-183.

Imler, J.L., Lemaire, C., Wasylyk, C. & Wasylyk, B. (1987). Negative regulation contributes to tissue specificity of the immunoglobulin heavy-chain enhancer. Mol.Cell Biol. 7 2558-2567.

Itoh, K., Sasaki, R., Ono, K., Tezuka, H., Sakoda, H., Sawada, H. Hitomi, K., Nakane, H., Uchiyama, T., Uchino, H. & Mori, K.J. (1988). Stromal cell dependent growth of the leukemic cells from murine erythroblastic leukemia. Jpn. J. Cancer 79 931-937.

Jackson, P.D., Evans, T., Nickol, J. & Felsenfeld, G. (1989). Developmental modulation of protein binding to β-globin gene regulatory sites within chicken erythrocyte nuclei. Genes Dev. 3 1860-1873.

Jackson, S.P. & Tjian, R. (1988). O-glycosylation of eukaryotic transcription factors: implications for mechanisims of transcriptional regulation. Cell 55 125-133.

Jantzen, H.M., Strahle, U., Gloss, B., Stewart, F., Schmid, W. Boshart, M., Miksicek, R. & Schutz, G. (1987). Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. Cell 49 29-38.

Johnson, A.D. & Herskowitz, I. (1985). A repressor (MAT α2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell 42 237-247.

Jones, K.A., Yamamoto, K.R. & Tjian, R. (1985). Two distinct

transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42 559-572.

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. & Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48 79-89.

Jones, N.C., Rigby, P.W.J. & Ziff, E.B. (1988). Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. Genes Dev. 2 267-281.

Jones, P.A. (1985). Altering gene expression with 5-azacytidine. Cell 40 485-486.

Jones, R.H. & Jones, N.C. (1989). Mammalian cAMP-responsive element can activate transcription in yeast and binds a yeast factor(s) that resembles the mammalian transcription factor ATF. Proc.Natl.Acad.Sci. USA 86 2176-2180.

Junker, S., Nielson, V., Matthias, P. & Picard, D. (1988). Both immunoglobulin promoter and enhancer sequences are targets for supppression in myeloma-fibroblast hybrid cells. EMBO J. 7 3093-3098.

Kadesch, T., Zevros, P. & Ruezinsky, D. (1986). Functional analysis of the murine IgH enhancer : evidence for negative control of cell-type specificity. Nucl.Acids Res. 14 8209-8221.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. & Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51 1079-1090.

Kakidanis, H. & Ptashne , M. (1988). GAL4 activates gene expression in mammalian cells. Cell 52 161-167.

Karlsson, S. & Nienhuis, A.W. (1985). Developmental regulation of human globin genes. Ann.Rev. Biochem. 54 1071-1108.

Kasturi, K. & harrison, P.R. (1985). The cell specificity and biosynthesis of mouse glycophorins studied with monoclonal antibodies. Exp.Cell Res. 157 253-264.

Keegan, L., Gill, G. and Ptashne, M. (1986). Separation of DNA-binding from the transcription activating function of a eukaryotic regulatory protein. Science 231 699-704.

Keleher, C.A., Goutte, C. & Johnson, A.D. (1988). The yeast cell-type specific repressor α2 acts cooperatively with a non-cell-type specific protein. Cell 53 927-936.

Kelly, K., Cochran, B.H., Stiles, C.D. & Leder, P. (1983). Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet derived growth factor. Cell 35 603-610.

Kemper. B., Jackson, P.D. & Felsenfeld, G. (1987). Protein-binding

sites within the 5′ DNAase I-hypersensitive region of the chicken α^D-globin gene. Mol.Cell Biol. 7 2059-2069.

Knezetic, J.A. & Luse, D.S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell 45 95-104.

Knezetic, J.A., Jacob, G.A. & Luse, D.S. (1988). Assembly of RNA polymerase II preinitiation complexes before assembly of nucleosomes allows efficient initiation of transcription on nucleosomal templates. Mol.Cell Biol. 8 3114-3121.

Knezetic, J.A. and Felsenfeld, G. (1989). Identification and characterisation of a chicken α-globin gene enhancer. Mol.Cell Biol. 9 893-901.

Kollias, G., Wrighton, N., Hurst, J. & Grosveld, F. (1986). Regulated expression of human $^A\gamma$-, β-, and hybrid γβ-globin genes in transgenic mice: manipulation of the developmental expression patterns. Cell 46 89-94.

Kollias, G. , Hurst, J., deBoer, E. & Grosveld, F. (1987). A tissue and developmental specific enhancer is located downstream from the human β-globin gene. Nucl.Acids Res. 15 5739-5747.

Kornberg, R.D. (1977). Structure of chromatin. Ann.Rev.Biochem. 46 931-954.

Kouzarides, T. & Ziff, E. (1988). The role of the leucine zipper in the Fos-Jun interaction. Nature 336 646-651.

Kouzarides, T. & Ziff, E. (1989). Leucine zippers of Fos, Jun and GCN4 dictate dimerisation specificity and thereby control DNA binding. Nature 340 568-571.

Krantz, S.B. & Goldwasser, E. (1984). Specific binding of erythropoietin to spleen cells infected with the anemia strain of Friend virus. Proc.Natl.Acad.Sci. USA 81 7574-7578.

Krantz, S.B., Sawyer, S.T. & Sawada, K.I. (1988). Purification of erythroid progenitor cells and characterisation of erythropoietn receptors. Br. J. Cancer 58 Suppl. IX 31-35.

Kruczek, I. & Doerfler, W. (1983). Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters; effect of promoter methylation on gene expression. Proc.Natl. Acad.Sci. USA 80 7586-7590.

Laimins, L., Holmgren-Konig, M. & Khoury, G. (1986). Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin I gene locus. Proc.Natl.Acad.Sci. USA 83 3151-3155.

Laughon, A. & Scott, M.P. (1984). Sequence of a Drosophila segmentation gene: protein structure homolgy with DNA-binding proteins. Nature 310

25-31.

Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J. & McKnight, S.L. (1988). Isolation of a recombinant copy of the genes encoding C/EBP. Genes Dev. 2 786-800.

Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240 1759-1764.

Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 243 1681-1688.

Lazarides, E. & Nelson, W.J. (1982). Expression of spectrin in nonerythroid cells. Cell 31 505-508.

Lee, K.A.W., Hai, T.Y., Sivaraman, L., Thimmappaya, B., Hurst, H.C., Jones, N.C. & Green, M.R. (1987). A cellular protein, activating transcription factor, activates transcription of multiple E1A-inducible adenovirus early promoters. Proc.Natl.Acad.Sci. USA 84 8355-8359.

Levine, M. & Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors.Cell 55 537-540.

Lewis, C.D., Clark, S.P., Felsenfeld, G. & Gould, H. (1988). An erythrocyte-specific protein that binds to the poly(dG) region of the

chicken β-globin gene promoter. Genes Dev. 2 863-873.

Li, J.P., D'Andrea, A.D., Lodish, H.F. & Baltimore, D. (1990). The Friend spleen focus-forming virus gp55 glycoprotein binds to the erythropoietin receptor and activates cell growth. (Unpublished results).

Lichsteiner, S., Wuarin, J. & Schibler, U. (1987). The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell 51 963-973.

Lillie, J.W. & Green, M.R. (1989). Transcription activation by the adenovirus E1A protein. Nature 338 39-44.

Lin, Y.S., Carey, M.F., Ptashne, M. & Green, M.R. (1988). GAL4 derivatives function alone and synergistically with mammalian activators in vitro. Cell 54 659-664.

Linemeyer, D.L., Menke, J.G., Ruscetti, S.K., Evans, L.H. & Scolnick, E.M. (1982). Envelope gene sequences which encode the gp52 protein of spleen focus-forming virus are required for the induction of erythroid cell proliferation. J.Virol. 43 223-233.

Lorch, Y., Lapointe, J.W. & Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49 203-210.

Lozzio, C.B. & Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood 45 321-334.

Lucibello, F.C., Lowag, C., Neuberg, M. & Muller, R. (1989). Trans-repression of the mouse c-fos promoter: a novel mechanism of Fos-mediated trans-regulation. Cell 59 999-1007.

Ma, J. & Ptashne, M. (1987). A new class of yeast transcriptional activators. Cell 51 113-119.

Mader, S., Kumar, V., de Verneuil, H. & Chambon, P. (1989). Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid responsive element. Nature 338 271-274.

Mantovani, R., Malgaretti, N., Nicolis, S., Ronchi, A., Giglioni, B. & Ottolenghi, S. (1988). The effects of HPFH mutations in the human γ -globin promoter on binding of ubiquitous and erythroid specific nuclear factors. Nucl.Acids Res. 16 7783-7797.

Mantovani, R., Superti-Furga, G., Gilman, J. & 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. Nucl.Acids Res. 17 6681-6691.

Maniatis, T., Fritsch, E.F., Lauer, J. & Lawn, R.M. (1980). The molecular genetics of human hemoglobins. Ann.Rev.Gen. 14 145-178.

Maniatis, T., Goodbourn, S. & Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science 236 1237-1245.

Marks, P.A. & Rifkind, R.A. (1972). Protein synthesis: its role in erythropoiesis. Science 175 955-961.

Marsh, J.L., Erfle, M. & Wykes, E.J. (1984). The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene 32 481-485.

Martin, D.I.K., Tsai, S.F. & Orkin, S.H. (1989). Increased γ-globin expression in a non-deletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature 338 435-438.

Martin, G.R. & Evans, M.J. (1975). Differentiation of clonal lines of teratocarcinoma cells : formation of embryoid bodies in vitro. Proc.Natl.Acad.Sci. USA 72 1441-1445.

Martin, P. & Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science 216 1233-1235.

Matsui, T. (1987). Transcription of adenovirus 2 major late and peptide IX genes under conditions of in vitro nucleosome assembly. Mol. Cell Biol. 7 1401-1408.

Mattaj, I. W., Lienhard, S., Jiricny, J. & de Robertis, E.M. (1985). An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes. Nature 316 163-167.

McCaffery, P.J., Fraser, J.K., Lin, F.K. and Berridge, M.V. (1989). Subunit structure of the erythropoietin receptor. J.Biol.Chem. 264 10507-10512.

McCormick, A., Wu, D., Castrillo, J.L., Dana, S., Strobl, J., Thompson, E.B. & Karin, M. (1988). Extinction of growth hormone expression in somatic cell hybrids involves repression of the specific trans-activator GHF-1. Cell 55 379-389.

McGhee, J.D. & Felsenfeld, G. (1980). Nucleosome structure. Ann. Rev. Biochem. 49 1115-1156.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. & Felsenfeld, G. (1981). A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell 27 45-55.

McGowan, R., Campbell, R., Peteron, A. & Sapienza, C. (1989). Cellular mosaicism in the methylation and expression of hemizygous loci in the mouse. Genes Dev. 3 1669-1676.

Meisterernst, M., Gander, I., Rogge, L. & Winnacker, E.L. (1988). A quantitative analysis of nuclear factor 1/DNA interactions. Nucl.Acids Res. 16 4419-4313.

Melis, M., Demopulos, G., Najfeld, V., Zhang, J.W., Brice, M., Papayannopoulou, T. & Stamatoyannopoulos, G. (1987). A chromosome 11-linked determinant controls fetal hemoglobin expression and fetal-to-adult globin switch. Proc.Natl.Acad.Sci. USA 84 8105-8109.

Mermod, N., Williams, T. & Tjian, R. (1988). Enahncer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription in vitro. Nature 332 557-561.

Mermod, N., O'Neill, E.A., Kelly, T.J. & Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell 58 741-753.

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

Mignotte, V., Eleouet, J.F., Raich, N. & Romeo, P.H. (1989). Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. Proc.Natl.acad.Sci. USA 86 2-5

Miller, D.M., Turner, P., Nienhuis, A.W., Axelrod, D.E. & Gopalakrishnan, T.V. (1978). Active conformation of the globin genes in uninduced and induced mouse erythroleukemia cells. Cell 14 511-521.

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

Mirand, E.A. (1967). Virus-induced erythropoiesis in hypertransfused polycythemic mice. Science 156 832-833.

Mitchell, P.J., Wang, C. & Tjian, R. (1987). Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell 50 847-861.

Mitchell, P.J. & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245 371-378.

Montminy, M.R. & Bilezikjian, L.M. (1987). Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature 328 175-178.

Moreau, P., Hen, R., Wasylyk, B., Everatt, R., Gaub, M.P. & Chambon, P. (1981). The Sv40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl.Acids Res. 9 6047-6068.

Moreau-Gachelin, F., Tavitian, A. & Tambourin, P. (1988). Spi-1 is a putative oncogene in virally induced murine erythroleukemias. Nature 331 277-280.

Mowat, M., Cheng, A., Kimura, N., Bernstein, A. & Benchimol, S. (1985). Rearrangements of the cellular p53 gene in erythroleukemic cells transformed by Friend virus. Nature 314 633-636.

Muller, H.P., Sogo, J.M. & Schaffner, W. (1989). An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. Cell 58 767-777.

Muller, M.M., Gerster, T. & Schaffner, W. (1988a). Enhancer sequences and the regulation of gene transcription. Eur. J. Biochem. 176 485-495.

Muller, M.M., Rupppert, S., Schaffner, W. & Matthias, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature 336 544-551.

Murata, M., Eto, Y., Shibai, H., Sakai, M. & Muramatsu, M. (1988). Erythroid differentiation factor is encoded by the same mRNA as that of the inhibin β_A chain. Proc.Natl.Acad.Sci. USA 85 2434-2438.

Myers, R.M. & Tjian, R. (1980). Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc.Natl.Acad.Sci. USA 77 6491-6495.

Myers, R.M., Rio, D.C., Robbins, A.K. & Tjian, R. (1982). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25 373-384.

Myers, R.M., Tilly, K. & Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science 232 613-618.

Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H. & Hurwitz, J. (1982). Adenovirus DNA replication in vitro: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc.Natl.Acad.Sci. USA 79 6438-6442.

Neuhaus, G., Neuhaus-Uri, G., Gruss, P. & Schweiger, H.G. (1984). Enhancer controlled expression of the simian virus 40 T-antigen in the green alga Acetabularia. EMBO J. 3 2169-2172.

Neuberg, M., Adamkiewicz, J., Hunter, J.B. & Muller, R. (1989). A Fos protein containing the Jun leucine zipper forms a homodimer which binds to the AP1 binding site. Nature 341 243-245.

Nickol, J.M. & Felsenfeld, G. (1988). Bidirectional control of the chicken β- and ε-globin genes by a shared enhancer. Proc.Natl.Acad.Sci. USA 85 2548-2552.

Nicolis, S., Ronchi, A., Malgaretti, N., Mantovani, R., Giglioni, B. & Ottolenghi, S. (1989). Increased erythroid-specific expression of a mutated HPFH γ-globin promoter region requires the erythroid factor NFE-1. Nucl.Acids Res. 17 5509-5516.

Nir, U., Walker, M.D. & Rutter, W.J. (1986). Regulation of rat insulin

I gene expression : evidence for negative regulation in nonpancreatic cells. Proc.Natl.Acad.Sci. USA 83 3180-3184.

Nishizuka, Y. (1986). Studies and Perspectives of Protein Kinase C. Science 233 305-312.

Nomura, S. & Oishi, M. (1983). Indirect evidence of erythroid differentiation in mouse Friend cells: evidence for two intracellular reactions invoved in the differentiation. Proc.Natl.Acad.Sci, USA 80 210-124.

Nomura, S., Yamagoe, S., Kamiya, T. & Oishi, M. (1986). An intracellular factor that induces erythroid differentiation in mouse erythroleukemia (Friend) cells. Cell 44 663-669.

Norman, C., Runswick, M., Pollock, R. & Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell 55 989-1003.

Nudel, U., Salmon, J., Fibach, E., Terada, M., Rifkind, R., Marks, P.A. & Bank, A. (1977). Accumulation of α- and β- globin messenger RNAs in mouse erythroleukemia cells. Cell 12 463-469.

Ohta, Y., Tanaka. M., Terada, M., Miller, O.J., Bank, A., Marks, P.A. & Rifkind, R.A. (1976). Erythroid cell differentiation: murine erythroleukemia cell variant with unique pattern of induction by polar compounds. Proc.Natl.Acad.Sci. USA 73 123222-1236.

Olins, D.E. & Wright, E.B. (1973). Glutaraldehyde fixation of isolated eukaryotic nuclei. J.Cell Biol. 59 301-317.

Ondek, B., Gloss, L. & Herr, W. (1988). The SV40 enhancer contains two distinct levels of organisation. Nature 333 40-45.

Orkin, S.H., Swan, D. & Leder, P. (1975). Differential expression of α- and β-globin genes during differentiation of cultures erythroleukemic cells. J. Biol. Chem. 250 8753-8760.

Ostertag, W., Pragnell, I., Jovin-Arndt, D. & Eisen, H. (1979). The Friend virus during Friend cell differentiation. "Oncogenic Viruses and Host Cell Genes" Edited by Ikawa, Y. & Odaka, T. Published by Academic Press, New York.

Pabo, C.O. & Lewis, M. (1982). The operator-binding domain of λ repressor: structure and DNA recognition. Nature 298 443-447.

Pabo, C.O. & Sauer, R.T. (1984). Protein-DNA recognition. Ann.Rev. Biochem. 53 293-321.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L.G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. Cell 51 1091-1104.

Papayannopoulou, T., Brice, M. & Stamatoyannapoulos, G. (1986). Analysis of human hemoglobin switching in MEL x human fetal erythroid cell hybrids. Cell 46 469-476.

Paul, J. & Hickey, I. (1972). Molecular pathology of the cancer cell. J.Clin. Path. 27 Suppl.(Roy.Coll.Path.) 7, 4-10.

Paulson, J.R. & Laemmli, U.K. (1977). The structure of histone-depleted metaphase chromosomes. Cell 12 817-818.

Pederson, D.S., Thoma, F. & Simpson, R.T. (1986). Core particle, fber, and transcriptionslly active chromatin structure. Ann. Rev. Cell Biol. 2 117-147.

Pelhamn H.R.B. (1982). A regulatory upstream promoter elemenet in the Drosophila Hsp 70 heat shock gene. Cell 30 517-528.

Perkins, N.D., Nicolas, R.H., Plumb, M.A. & Goodwin, G.H. (1989). The purification of an erythroid protein which binds to enhancer and promoter elements of hemoglobin genes. Nucl.Acids Res. 17 1299-1314.

Peschle, C., Migliaccio, A.R., Migliaccio, G., Petrini, M., Calandrini, M., Russo, G., Mastroberardino, G., Presta, M., Gianni, A.M., Comi, P., Giglioni, B. & Ottolenghi, S. (1984). Embryonic to fetal Hb switch in humans: studies on erythroid bursts generated by embryonic progenitors from yolk sac and liver. Proc.Natl.Acad.Sci. USA 81 2416-2420.

Peschle, C., Mavilio, F., Care, A., Migliaccio, G., Migliaccio, A.R., Salvo, G., Samoggia, P., Petti, S., Guerriero, R., Marinucci, M, Lazzaro, D., Russo, G. & Mastroberardino, G. (1985). Haemoglobin switching in human embryos: asynchrony of ζ --> α and ε --> γ globin switches in primitive and definitive erythropoietic lineage. Nature 313 235-238.

Plumb, M., Lobanenkov, Nicolas, Wright, C.A., Zavou, S. & Goodwin, G.H. (1986). Characterisation of chicken erythroid nuclear proteins which bind to the nuclease hypersensitive regions upstream of the β^A- and β^H-globin genes. Nucl.Acids Res. 14 7675-7693.

Plumb, M.A. & Goodwin, G.H. (1988). Detection of sequence specific protein DNA interactions by the DNA-Footprinting Technique. Chapter 12 in Methods in Molecular Biology, Volume 4. Edited by J.M. Walker. Published by Humana Press, New Jersey.

Plumb, M., Frampton, J., Wainwright, H., Walker,M., Macleod, K., Goodwin, G.H. & Harrison, P.R. (1989). GATAAG; a cis-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. Nucl.Acids Res. 17 73-92.

Preisler, H.D., Christoff, G. & Taylor, E. (1976). Cryoprotective agents as inducers of erythroleukemic cell differentiation in vitro. Blood 47 363-368.

Prywes, R. & Roeder, R.G. (1986). Inducible binding of a factor to the

c-fos enhancer. Cell 47 777-784.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322 697-701.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature 335 683-689.

Ralph, P. & Nakoinz, I. (1977). Direct effects of immunopotentiators on monocytic, myelomonocytic and histiocytic or macrophage tumor cells in culture. Cancer Res. 37 546-550.

Ramsay, N., Felsenfeld, G., Rushton, B.M. & McGhee, J.D. (1984). A 145-base pair DNA sequence that positions itself precisely and asymmetrically on the nucleosome core. EMBO J. 3 2605-2611.

Ransome, L.J., Visvader, J., Sassone-Corsi, P. & Verma, I. (1989). Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins. Genes Dev. 3 770-781.

Rawlins, D.R., Rosenfeld, P.J., Wides, R.J., Challberg, M.D. & Kelly,Jnr., T.J. (1984). Structure and function of the adenovirus origin of replication. Cell 37 309-319.

Raymondjean, M., Cereghini, S. & Yaniv, M. (1988). Several distinct "CCAAT" box binding proteins coexist in eukaryotic cells. Proc.Natl.Acad.Sci. USA 85 757-761.

Reitman, M. & Felsenfeld, G. (1988). Mutational analysis of the chicken β-globin enhancer reveals two positive-acting domains. Proc.Natl.acad.Sci. USA 85 6267-6271.

Reuben, R.C., Wife, R.L., Breslow, R., Rifkind, R.A. & Marks, P.A. (1976). A new group of potent inducers of differentiation in murine erythroleukemia cells. Proc.Natl.Acad.Sci. USA 73 862-866.

Rio, D., Robbins, A., Myers, R. & Tjian, R. (1980). Regulation of simian virus 40 early transcription in vitro by a purified tumor antigen. Proc.natl.Acad.Sci. USA 77 5706-5710.

Rio, D.C. & Tjian, R. (1983). SV40 T antigen binding site mutations that affect autoregulation. Cell 32 1227-1240.

Rochette-Egly, C., Fromenthal, C. & Chambo, P. (1990). General repression of enhanson activity by the adenovirus-2 E1a proteins. Genes Dev. 4 137-150.

Romero-Herrera, A.E., Lehmann, H., Joysey, K.A. & Friday, A.E. (1973). Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. Nature 246 389-395.

Rosales, R., Vigneron, M., Macchi, M., Davidson, I., Xiao, J. H. & Chambon, P. (1987). In vitro binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of the SV40 enhancer and reltaed

motifs present in other promoters and enhancers. EMBO J. 6 3015-3025.

Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. & de Crombrugghe, B. (1988). A Nuclear Factor 1 binding site mediates the transcriptional activation of a type 1 collagen promoter by transforming growth factor-β. Cell 52 405-414.

Rousseau, S., Renaud, J. & Ruiz-Carillo, A. (1989). Basal expression of the histone H5 gene is controlled by positive and negative cis-acting sequences. Nucl. Acids Res. 17 7495-7511.

Rupp, R.A.W., & Sippel, A.E. (1987). Chicken liver TGGCA protein purified by preparative mobility shift electrophoresis (PMSE) shows a 36.8 to 29.8 kd microheterogeneity. Nucl.Acids Res. 15 9707-9726.

Russell, G.J., Walker, P.M.B., Elton, R.A. & Subak-Sharpe, J.H. (1976). Doublet frequency analysis of fractionated vertebrate nuclear DNA. J.Mol. Biol. 108 1-23.

Ryan, T.M., Behringer, R.R., Townes, T.M., Palmiter, R.D. & Brinster, R.L. (1989). High-level expression of human α-globin genes in transgenic mice. Proc.Natl.Acad.Sci. USA 86 37-41.

Sachs, L. (1987). The molecular control of blood cell devlopment. Science 238 1374-1379.

Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. (1988). GAL4-VP16

is an unusually potent transcriptional activator. Nature 335 563-564.

Saiki, R.K., Gelfand, D.H., Stoffel, Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 487-491.

Sandeen, G., Wood, W.I. & Felsenfeld, G. (1980). The interaction of high mobility proteins HMG14 and 17 with nucleosomes. Nucl.Acids Res. 8 3757-3778.

Santoro, C., Mermod, N., Andrews, P.C. & Tjian, R. (1988). A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature 334 218-224.

Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H. & Vennstrom, B. (1986). The c-erb-A protein is a high-affinity receptor for thyroid hormone. Nature 324 635-640.

Sassa, S. (1980). Control of heme biosynthesis in erythroid cells. "In vivo and in vitro erythropoiesis: the Friend system." 219-228. Edited by Rossi, G.B. Published by Elsevier /North-Holland Biomedical press.

Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C.G., Currie, R.A. & Roeder, R.G. (1988). A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox

protein. Nature 336 551-557.

Scheuermann, R.H. & Chen, U. (1989). A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. Genes Dev. 3 1255-1266.

Schubach, W. & Groudine, M. (1984). Alteration of c-myc chromatin structure by avian leukosis virus integration. Nature 307 702-708.

Schuermann, M., Neuberg, M., Hunter, J.B., Jenuwein, T., Ryseck, R.P., Bravo, R. & Muller, R. (1989). The leucine repeat motif in Fos protein mediates complex formation with Jun/AP-1 and is required for transformation. Cell 56 507-516.

Schule, R., Muller, M., Kaltschmidt, C. & Renkawitz, R. (1988a). Many transcription factors interact synergistically with steroid receptors. Science 242 1418-1420.

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

Schutte, J., Viallet, J., Nau, M., Segal, S., Fedorko, J. & Minna, J. (1989). Jun-B inhibits and c-Fos stimulates the transforming and trans-activating activities of c-Jun. Cell 59 987-997.

Selden, R.F., Howie, K.B., Rowe, M.E., Goodman, H.M. & Moore, D.D.

(1986). Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Mol. Cell Biol. 6 3173-3179.

Serfling, E., Jasin, M. & Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. TIG 1 224-230.

Sheffery, M., Rifkind, R.A. & Marks, P.A. (1982). Murine erythroleukemia cell differentiation: DNAase 1 hypersensitivity and DNA methylation near the globin genes. Proc.Natl.Acad.Sci. USA 79 1180-1184.

Sheffery, M., Rifkind, R.A. & Marks, P.A. (1983). Hexamethylenebisacetamide-resistant murine erythroleumia cells have altered patterns of inducer-mediated chromatin changes. Proc.Natl.Acad.Sci. USA 80 3349-3353.

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., de Robertis, E.M. & Gehring, W. (1984). Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. Nature 310 70-71.

Shiozaki, M., Itoh, K. & Mori, K.M. (1989). Cooperative Effects of erythropoietin and interleukin 3 on differentiation of responsive mouse erythroblastic leukemia cells. (In press).

Simpson, R.T. & Stafford, D.W. (1983). Structural features of a phased nucleosome core particle. Proc.Natl.Acad.Sci. USA 80 51-55.

Sive, H.L., Heintz, N. & Roeder, R.G. (1986). Multiple elements are required for maximal in vitro transcription of a human histone H2B gene. Mol.Cell Biol. 6 3329-3340.

Smeal, T., Angel, P., Meek, J. & Karin, M. (1989). Different requirements for formation of Jun;Jun and Jun:Fos complexes. Genes Dev.3 2091-2100.

Smith, R.D., Yu, J. & Seale, R.L. (1984). Chromatin structure of the β-globin gene family in murine erythroleukemia cells. Biochemistry 23 785-790.

Smith, R.D. & Yu, J. (1984). Alterations in globin gene chromatin conformation during murine erythroleukemia cell differentiation. J.Biol.Chem. 259 4609-4615.

Sorger, P.K., Lewis, M.J., Pelham, H.R.B. (1987). Heat shock factor is regulated differently in yeast and HeLa cells. Nature 329 81-84.

Sorger, P.K. & Pelham, H.R.B. (1988). Yeast shock factor is an essential DNA-binding protein that exhibits temperature -dependent phosphorylation. Cell 54 855-864.

Spandidos, D.A. & Wilkie, N.W. (1984). Malignant transformation of early passage rodent cells by a single mutated human oncogene. Nature 310 469-474.

Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M., & Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase 1. Cell 20 451-460.

Stein, R., Razin, A. & Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyl transferase gene inhibits its expression in mouse L cells. Proc.Natl.Acad.Sci. USA 79 3418-3422.

Stein, R., Sciaky-Gallili, N., Razin, A. & Cedar, H. (1983). Pattern of methylation of two genes coding for housekeeping functions. Proc.Natl.Acad.Sci. USA 80 2422-2426.

Stephenson, J.R., Axelrad, A.A., Mcleod, D.L. & Shreeve, M.M. (1971). Induction of colonies of hemoglobin-synthesising cells by erythropoietin in vitro. Proc.Natl.Acad.Sci. USA 68 1542-1546.

Stern, S., Tanaka, M. & Herr, W. (1989). The Oct-1 homeodomain directs formation of multiprotein-DNA complex with the HSV transactivator VP16. Nature 341 624-630.

Struhl, K. (1987). The DNA-binding domains of the Jun oncoprotein and the yeast GCN 4 transcriptional activator protein are functionally homologous. Cell 50 841-846.

Sturm, R.A., Das, G. & Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev. 2 1582-1599.

Superti-Furga, G, Barberis, A., Schaffner, G. & Busslinger, M. (1988). The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the γ-globin gene. EMBO J. 3099-3107.

Superti-Furga, G., Barberis, A., Schreiber, E. & Busslinger, M. (1989). The protein CDP, but not CP1, footprints on the CCAAT region of the γ -globin gene in unfractionated B-cell extracts. Biochim.Biophys.Acta 1007 237-242.

Superti-Furga, G. & Busslinger, M. (1989). Nuclear factors interacting with the promoter of human γ-globin genes. Published in Highlights of Modern Biochemistry: Proceedings of the 14th International Congress of Biochemistry, VSP, Zeist, The Netherlands.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. & Greaves, D.R. (1989). A dominant control region from the human β-globin locus conferring integration site-independent gene expression. Nature 338 352-355.

Thiele, B., Belkner, J., Andree, H., Rapoport, T.A. & Rapoport, S. (1979). Synthesis of non-globin proteins in rabbit-erythroid cells. Eur. J. Biochem. 96 563-569.

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

Till, J.E. & McCulloch, E.A. (1980). Hemopoietic Differentiation. Biochim.Biophys.Acta 605 431-459.

Todokoro, K., Watson, R.J., Higo, H., Amanuma, H., Kuramochi, S., Yanagisawa, H. & Ikawa, Y. (1988). Down-regulation of c-myb gene expression is a prerequisite for erythropoietin-induced erythroid differentiation. Proc.Natl.Acad.sci. USA 85 8900-8904.

Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E. & Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. Cell 59 477-487.

Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. & Palmiter, R.D. (1985). Erythroid-specific expression of human β-globin genes in transgenic mice. EMBO J. 4 1715-1723.

Trainor, C., Stamler, S.J. & Engel, J.D. (1987). Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3' enhancer. Nature 328 827-830.

Trainor, C.D., Evans, T., Felsenfeld, G. & Boguski, M.S. (1990). Structure and evolution of a human erythroid transcription factor. Nature 343 92-96.

Treisman, R., Green, M.R. & Maniatis, M.R. (1983). cis and trans activation of globin gene transcription in transient assays.

Proc.Natl.Acad.Sci. USA 80 7428-7432.

Treisman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell 46 567-574.

Treisman, R. (1987). Identification and purification of a polypeptide that binds to the c-fos serum response element. EMBO J. 6 2711-2717.

Treizenberg, S.J., Kingsbury, R.C. & McKnight, S. (1988). Functional dissection of VP16, the trans-activator function of herpes simplex virus immediate early gene expression. Genes Dev. 2 718-729.

Trudel, M. & Costantini, F. (1987). A 3' enhancer contributes to the stage-specific expression of the human β-globin gene. Genes Dev. 1 954-961.

Tsai, S.F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G. & Orkin, S.H. (1989a). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature 339 446-451.

Tsai, S.Y., Tsai, M.J. & O'Malley. B.W. (1989b). Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. Cell 57 443-448.

Tuan, D., Solomon, W., Li, Q. & London, I.M. (1985). The

"β-like-globin" gene domain in human erythroid cells. Proc.Natl.Acad.Sci. USA 82 6384-6388.

Turner, R. & Tjian, R. (1989). Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. Science 243 1689-1694.

Vallette, F., Mege, E., Reiss, A. & Adesnik, M. (1989). Construction of mutant and chimeric genes using the polymerase chain reaction. Nucl.Acids Res. 17 723-733.

Van Assendelft, G., Hanscombe, O., Grosveld, F. & Greaves, D.R. (1989). The β-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. Cell 56 969-977.

Van Huijsduijnen, R.A.M.H., Bollekens, J., Dorn, A., Benoist, C. & Mathis, D. (1988). Properties of a CCAAT box-binding protein. Nucl.Acids Res. 16 7265-7282.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M. & Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. Proc.Natl.Acad.Sci. USA 79 1073-1077.

Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T. & Shapiro, L.J. (1982). Transformation with DNA from 5-aza-cytidine-reactivated X chromosomes. Proc.Natl.Acad.Sci. USA 79 2352-2354.

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

Wasylyk, C. & Wasylyk, B. (1986). The immunoglobulin heavy-chain B-lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. EMBO J. 5 553-560.

Webster, N., Jin, J.R., Green, S., Hollis, M. & Chambon, P. (1988). The yeast UAS_G is atranscriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. Cell 52 169-178.

Weinberger, C., Hollenberg, S.M., Rosenfeld, M.G. & Evans, R.M. (1985). Domain structure of human glucocorticoid receptor and its relationship to the v-erb-A oncogene product. Nature 318 670-672.

Weintraub, H. & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science 193 848-856.

Weisbrod, S. (1982). Active chromatin. Nature 297 289-295.

Wharton, R.P. & Ptashne, M. (1985). Changing the binding specificity of a repressor by redesigning an α-helix. Nature 316 601-605.

White, T.J., Arnheim, N. & Erlich H.A. (1989). The polymerase chain reaction. TIG 5 185-189.

Whitelaw, E. Hogben, P., Hanscombe, O. & Proudfoot, N.J. (1989). Transcriptional promiscuity of the human α-globin gene. Mol.Cell Biol. 9 241-251.

Williams, T., Admon, A., Luscher, B. & Tjian, R. (1988). Cloning and expression of AP-2, a cell-type -specific transcription factor that activates inducible enhancer elements. Genes Dev. 2 1557-1569.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T. & Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyl transferase locus on the X chromosome: implications for X-chromosome inactivation. Proc.Natl.Acad.Sci. USA 81 2806-2810.

Wright, S., deBoer, E., Grosveld, F. & Flavell, R.A. (1983). Regulated expression of the human β-globin gene family in murine erythroleukaemia cells. Nature 305 333-336.

Wright, S., Rosenthal, A., Flavell, R. & Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell 38 265-273.

Wu, H.Y., Shyy, S., Wang, J.C. & Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. Cell 53 433-440.

Xiao, J.H., Davidson, I., Ferrandon, D., Rosales, R., Vigneron, M.,

Macchi, M., Ruffenach, F. & Chambon, P. (1987). One cell-specific and three ubiquitous nuclear proteins bind in vitro to overlapping motifs in the domain B1 of the SV40 enhancer. EMBO J. 6 3005-3013.

Xiao, J.H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. & Chambon, P. (1987). In vitro binding of several cell-specific and ubiquitous nuclear proteins to the GT-1 motif of the SV40 enhancer. Genes Dev. 1 794-807.

Yamamoto, K.K., Gonzalez, G.A., Biggs III, W.H. & Montminy, M.R. (1988). Phosphorylation-induced binding and transcriptional regulation efficacy of nuclear factor CREB. Nature 334 494-498.

Yamamoto, M., Yew, N.S., Federspiel, M., Dodgson, J.B., Hayashi, N. & Engel, J.D. (1985). Isolation of recombinant cDNAs encoding chicken erythroid δ-aminolevulinate synthase. Proc.Natl.Acad.Sci. USA. 82 3702-3706.

Yen, P.H., Patel, P., Chinault, A.C., Mohandas, T. & Shapiro, L.J. (1984). Differential methylation of hypoxanthine phosphoribosyl transferase genes on active and inactive human X chromosomes. Proc.Natl.Acad.Sci.USA. 81 1759-1763.

Yu, H., Porton, B., Shen, L. & Eckhardt, L.A. (1989). Role of the octamer motif in hybrid cell extinction of immunoglobulin gene expression: extinction is dominant in a two enhancer system. Cell 58 441-448.

Zaller, D.M., Yu, H. & Eckhardt, L.A. (1988). Genes activated in the presence of an immunoglobulin enhancer or promoter are negatively regulated by a T-lymphoma cell line. Mol.Cell Biol. 8 1932-1939.

Zehnbauer, B.A. & Vogelstein, B. (1985). Supercoiled loops and the organisation of replication and transcription in eukaryotes. Bioessays 2 52-54.

Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. & Chambon, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. 5 387-397.

Zimarino, V. & Wu, C. (1987). Induction of sequence-specific binding of Drosophila heat shock activator protein without protein synthesis. Nature 327 727-730.

Zinn, K. & Maniatis, T. (1986). Detection of factors that interact with the human β-interferon regulatory region in vivo by DNAase I footprinting. Cell 45 611-618.