Volume 17 Number 1 1989 Nucleic Acids Research

GATAAG; a *cis*-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression

M.Plumb, J.Frampton+, H.Wainwright, M.Walker, K.Macleod, G.Goodwin[1] and P.Harrison

Beatson Institute for Cancer Research, Garscube Estate, Bearsden, Glasgow and [1]Chester Beatty Research Laboratories, Fulham Road, London, UK

Received October 25, 1988; Revised and Accepted November 28, 1988

ABSTRACT:

An erythroid-specific nuclear protein factor binds to a sequence motif (GATAAG) which is present in the promoter region of the mouse α and β^{major} globin genes, and in the erythroid-specific promoter of the human porphobilinogen deaminase (PBG-D) gene. The protein activity is conserved across species, being found in mouse erythroleukaemia (MEL) cells, chicken erythrocytes, the human erythroid K562 and KMOE cell lines, but not in a variety of non-erythroid mouse tissues or in HeLa cells. Functional analysis of this element in the α globin gene promoter by stable transfection experiments show that the GATAAG motif resides in a 68 bp sequence which has a stimulatory effect on transcription in mouse erythroleukaemia but not fibroblast cells. The GATAAG motif is conserved in the promoters and 3' enhancers of a variety of globin and non-globin genes implying that it is a cis-element involved in the tissue-specific up-regulation of several genes that are co-expressed during erythroid cell differentiation.

INTRODUCTION:

The tissue-specific expression of genes during erythroid cell differentiation has stimulated the search for common features within the regulatory mechanisms responsible for their transcription. Multiple cis-acting elements involved in the erythroid-specific transcriptional regulation of the β globin genes, and to a lesser extent in the α globin genes and erythroid-specific non-globin genes, have been identified (reviewed in (1)). These functional elements are commonly associated with tissue specific nuclear DNase I hypersensitive sites (DHSS) indicating that the sequences are accessible for binding nuclear trans-acting factors *in vivo*. For example, sequences at the DHSSs flanking the human β globin multigene domain confer normal levels of transcriptional activity on the human β globin gene in transgenic mice (2). Similarly, enhancer elements have been identified downstream of the human β and γ^{A} globin genes (3-6), the chicken β^{A} and α^{A} globin genes (7-9), and the erythroid-specific chicken histone H5 gene (10). The *in vitro* interaction of nuclear trans-acting

factors with these 3' enhancers suggests that both tissue-specific and ubiquitous nuclear proteins can interact with sequences in the enhancer.

Cis-acting regulatory elements act directly or indirectly at the promoter to modulate the rate of transcription initiation. Conserved sequences in globin gene promoters such as the TATA, CCAAT and CACCC motifs have been shown to be essential for transcription (11-13), and their binding by nuclear trans-acting factors in vivo and/or in vitro has been examined (14-22). However, there is evidence that the nuclear protein factors that interact with the CCAAT and CACCC motifs are not tissue-specific (17,21-26).

Erythroid-specific nuclear protein interactions with promoter and downstream globin sequences in vitro have also been identified: For example, a poly(dG) homopolymer sequence in the chicken β^A globin gene promoter (15, 16); a sequence (H') in the chicken $\beta^{hatching}$ promoter (17); the chicken α^D globin gene promoter (19); footprints II, III, and IV in the chicken β^A globin gene 3' enhancer (8); the human γ-globin promoter (site B2, (25)), and site B2 in the mouse β^{major} globin gene second intron (IVS2; (27)). Footprint analysis of the human β (28) and chicken α^A (9) globin gene 3' enhancers have also revealed multiple binding sites for erythroid-specific nuclear DNA-binding protein(s).

In the course of examining the interaction of nuclear proteins with the promoter sequences of the mouse α and β^{major} globin genes and the erythroid-specific promoter of the human porphobilinogen deaminase (PBG-D) gene, it became clear that sequences containing a GATAAG motif are bound by the same erythroid-specific and species-conserved nuclear factor. Furthermore, analysis of the nucleotide sequence required for binding suggests that the protein is very similar and possibly identical to the erythroid-specific protein(s) which bind to the promoters and enhancers of other globin (9, 17, 19, 28) and non-globin erythroid-specific genes. Functional analyses of the α globin gene GATAAG motif cells show that it is within a 68 bp restriction fragment that gives an erythroid-specific stimulation of transcription. We discuss the implications of the presence of the regulatory element in promoter and enhancer elements and its functional dependence on other adjacent cis-elements.

MATERIALS AND METHODS:

Cell lines and tissue culture:

Mouse adherent MEL cells (F4-12B2; a gift from W. Ostertag, Hamburg), suspension MEL cells (clone M707/T), mouse fibroblast (STO) and macrophage

(J774.2; (29)) cell lines and human K562 cells were grown in Eagles MEM with double concentrations of amino acids and vitamins (Gibco) and with 10% foetal calf serum (Biocon). MEL cells were induced to differentiate with 5 mM N,N'-hexamethylene-bisacetamide (HMBA) for 5 days.

Plasmid constructions:

Plasmids were constructed using a BamHI to XbaI fragment from pLW2 (30) containing the CAT coding sequences and Herpes Simplex virus-2 IE5 terminator sequences inserted into pUC12 to give the plasmid p22 (a gift from J. Lang, Beatson Institute).

pαCAT plasmids were constructed by insertion of subcloned DNA fragments of the mouse α globin gene region from the genomic recombinant lambda clone (31) into the p22 vector.

pαCAT1 : The 720 bp NcoI fragment extending 5' of the initiating ATG codon of the α globin gene was blunt ended with S1 nuclease so as to remove the ATG and was then ligated into the SmaI site of p22.

pαCAT1Δ1 : This was generated by removal of α globin sequences between the unique ApaI site at -129 bp and the MspI site at -192 bp in pαCAT1 followed by religation.

pαCAT0 : Sequences in pαCAT1 were deleted up to -52 bp relative to the α globin gene cap site by partial digestion with BstNI and subsequent religation.

pα60(+/-) CAT0: The ApaI-MspI restriction fragment between -129 bp and -192 bp was blunt-end ligated into the SmaI site of pIC20R (32), excised at the flanking EcoRI stes, and then cloned in either orientation into the EcoRI restriction site 5' to the CAT gene in pαCAT0. The orientation of inserts in different clones was confirmed by restriction mapping and sequencing.

Stable transfection and CAT assays:

1 day prior to transfection, 2 X 10^6 cells were plated in 20 ml into 80 cm^2 dishes and then transfected with 40 µg of CAT plasmid, and 0.5 µg of Homer 6 plasmid (33) by the calcium phosphate co-precipitation method (34). Fresh medium was applied after 18 h. 48h after transfection G418 was added to 800 µg/ml. After approximately 14 days, transfectants (greater than 50 colonies) were passaged once at low density (c. 10^4 cells/ cm^2). Where appropriate, 5 mM HMBA was added 2 days later. Cells were harvested 5 days after passaging and split into aliquots for CAT enzyme, DNA or RNA analysis. For CAT enzyme analysis, 2 X 10^6 cells were lysed by three cycles of freezing and thawing, and aliquots analyzed for CAT activity (35).

Nuclear protein preparation:

Nuclei were prepared and extracted with 0.35 M NaCl, the eluate precipitated with ammonium sulphate (0.35 g/ml) (36), and then either redissolved and dialysed against storage buffer (50 mM NaCl, 20 mM Hepes pH7.9, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, and 20% glycerol) to give the crude nuclear extracts, or fractionated by affinity chromatography on calf thymus DNA-cellulose. The protein fraction eluting from the DNA-cellulose between 100 mM and 250 mM ammonium sulphate (the C4 protein fraction) was collected and dialysed against storage buffer (16, 36). Typically, nuclei containing 100-150 mg of DNA resulted in 50 mg of crude protein or 2 mg of partially purified protein extract each in a volume of 5 ml. All solutions were made with 0.5 mM phenylmethyl-sulfonyl fluoride and benzamidine, and 0.1 µg/ml each of pepstatin A, aprotinin, leupeptin, and bestatin (Sigma). Where indicated, solutions also contained 2 mM levamisole and 10 mM ß-phosphoglycerate.

Footprint analysis:

Restriction fragments were 5' end-labelled with T4 polynucleotide kinase and γ^{32}P-ATP and isolated after secondary restriction as described (16). Markers were prepared by the chemical sequencing reactions (37).

DNase I footprint protection assays were performed in a final volume of 100 µl of storage buffer in the presence of 1 µg of poly (dI-dC):(dI-dC), 2 ng of end-labelled restriction fragment, and up to 98 µl (0-25 µg) of protein in the presence or absence of double-stranded competitor oligonucleotide (100 ng). After DNase I digestion, the nucleic acid was purified and resolved by denaturing 6% polyacrylamide gel electrophoresis and autoradiography (16).

Oligonucleotides:

Oligonucleotides were purchased from Oswell DNA service (Edinburgh University), and complementary single-stranded oligonucleotides annealled. The double-stranded oligonucleotides contained BamHI and BglII restriction sites at their 5' and 3' ends respectively. The sequences of the coding strand of the oligonucleotides containing the (variant) GATAAG motifs are shown in Table II. The oligonucleotide containing the mouse α-globin CCAAT was (αP3a): GATCCAAACCAGCCAATGAGAACTGCTCCA.

Gel shift assays:

Oligonucleotides were 5' end-labelled with T4 polynucleotide kinase and γ^{32}P-ATP, 5' overhangs filled in with the Klenow fragment of DNA polymerase and excess deoxynucleotide triphosphates, and purified by electrophoresis on

a 8% polyacrylamide gel. Labelled oligonucleotide (100 pg) was incubated on ice for 1 h with protein extracts (0-10 µg) in the presence of 5 µg of poly(dI-dC):(dI-dC) in a final volume of 20 µl of storage buffer. Samples were electrophoresed in a 5% polyacrylamide gel in 0.2X TBE for 2 h at 4°C at 150V, and the gel dried for autoradiography.

RNA preparation and S1 nuclease protection analysis:

RNA for CAT RNA analysis was obtained from stably transfected cell pools using the method of Chomczynski and Sacchi (49). A small aliquot of the same cells was simultaneously tested for CAT enzyme activity. Approximately 1-2 x 10^8 cells were used and these yielded 1.5-3 mg RNA. Poly A+ RNA was selected according to Maniatis et al (50).

Single-stranded, uniformly-labelled probe for S1 analysis was obtained by the method of Bentley (51). The probe used here was derived from an M13mp19 clone containing a 430 bp MspI fragment derived from pαCAT1 extending from position-197 bp in the α globin gene sequences to 170 bp downstream of the ATG codon in the CAT gene sequences.

S1 nuclease protection was performed as described by Weaver and Weissman (52). Hybridisation was carried out using probe plus 10 µg polyA+ RNA and 40 µg yeast tRNA in 10 µl 80% formamide/0.4M NaCl/0.04M PIPES pH6.4/1mM EDTA at an optimised temperature of 40°C for 16 h. Hybrids were digested in 250 µl 250mM NaCl/1mM $ZnSO_4$/30mM NaOAc pH4.6 using 20 U S1 nuclease (Boehringer) at 37°C for 60 min. Following ethanol precipitation, protected products were resolved by denaturing 6% polyacrylamide gel electrophoresis.

RESULTS:

Protein binding to the mouse α-globin gene promoter GATAAG motif:

To analyse the in vitro interaction of nuclear trans-acting factors with the promoter sequences of the mouse α and β^{major} globin genes, crude nuclear extracts were prepared from various mouse tissues and from the mouse erythroleukaemia (MEL), macrophage (J774.2) and fibroblast (STO) cell lines and either used directly, or fractionated by non-specific DNA-cellulose affinity chromatography (15). Footprint analysis of the α globin gene promoter between the cap site and -200 bp, revealed a number of binding sites which include the CCAAT and CACCC motifs ((21), and M. Plumb, unpublished results); of these, only one was bound by an MELC-specific activity. As shown in Figure 1a, the sequence between -174 to -189 bp (site αP4) containing a GATAAG motif, is footprinted by an MELC protein factor with a characteristic induction of two hypersensitive cleavage sites at the 5' end, but the

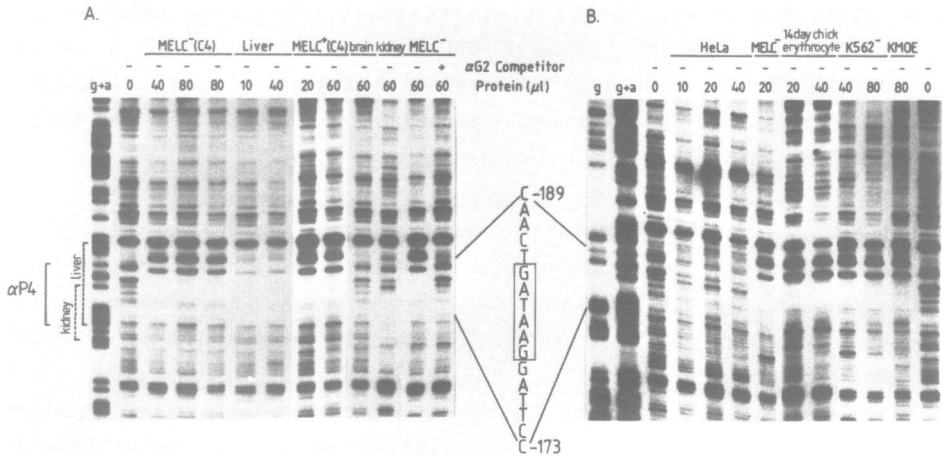

Figure 1.
Footprint analysis of the GATAAG motif.
A 650 bp Pst 1 restriction fragment of the α globin gene promoter (from -15 bp to -665 bp) was cloned into pUC9, 5' end-labelled at the pUC9 HindIII site (the 3' end of the promoter relative to the cap site, labelled on the anti-coding strand), and the insert purified after secondary digestion with EcoRI. Labelled DNA (5 ng) and 1 μg of poly (dI-dC):(dI-dC) was incubated with crude nuclear protein extracts or partially purified (C4) protein fractions in the presence or absence of excess (100 ng) of competitor double-stranded Oligo αG2. Figure 1A, protein from mouse brain, kidney or liver, or from HMBA induced ($MELC^+$) or uninduced ($MELC^-$) M707/T cells; Figure 1B, protein from human HeLa, KMOE or uninduced K562 cells, from 12 d chicken erythrocytes, or from uninduced M707/T ($MELC^-$) cells. After partial DNase I digestion, nucleic acids were resolved by denaturing gel electrophoresis. The G and/or G+A tracks are chemical sequencing markers, and 0 is the zero protein control. Footprint αP4 and its variants in mouse kidney and liver are shown.

footprint is not detected with the brain protein extract. A footprint over this region is detected with both the kidney and liver protein extracts: however, the kidney protein binds to an adjacent sequence which does not include the GATAAG motif indicating that it is bound by a distinct binding activity; and, as the liver extract footprint lacks the 5' hypersensitive cleavage sites, it suggests that there is a distinct (or modified) GATAAG-binding protein(s) in liver (this is also confirmed by the gel shift experiments described below). A double stranded 27 bp synthetic oligonucleotide homologous to the footprinted sequence (oligo αG2, Table II) specifically competes out the MELC-specific footprint (Figure 1A).

The GATAAG motif bound in site αP4 is homologous to the GATAAGATAAG direct repeat in the chicken α^D globin gene promoter which is bound by chicken erythrocyte protein *in vivo* and *in vitro* (19), implying that the

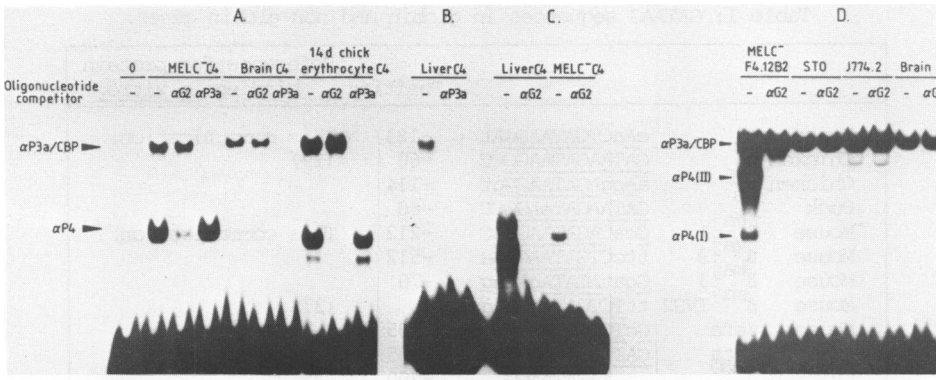

Figure 2.
Gel shift analysis of the GATAAG binding protein.
End-labelled oligonucleotides αG2 and/or αP3a (see Methods) were incubated with 3-5 µl of partially purified C4 fraction or crude nuclear extracts from uninduced MELC M707/T or F4.12B2 (MELC⁻) cells, fibroblasts (STO), macrophage (J774.2), brain, or chicken erythrocyte in the presence of 6 µg of poly (dI-dC):(dI-dC) and the presence (+) or absence (-) of 100 ng of cold competitor oligonucleotide DNA. (Figure 2A) Labelled oligo αG2 and αP3a were incubated with partially purified C4 protein fractions. (Figures 2B and 2C) Partially purified C4 protein fractions from liver or M707/T cells were incubated with either labelled oligo αG2 (Figure 2C) or αP3a (Figure 2B). (Figure 2D) Crude nuclear extracts were incubated with both labelled oligos βG2 and αP3a. Samples were resolved by non-denaturing gel electrophoresis and autoradiography.

GATAAG-binding protein is conserved across species. Crude nuclear protein extracts were therefore prepared from 14 d chicken erythrocytes, two human erythroid cell lines (K562 and KMOE) and HeLa cells, bound to the α-globin gene promoter sequence and assayed by footprinting (Figure 1B). No footprint is detected with the HeLa nuclear extract, but all the erythroid cell extracts give the same footprint with the characteristic 5' hypersensitive cleavage sites, indicating that the mouse, human and chicken proteins are highly related and erythroid-specific.

The protein interactions with this sequence were further examined by gel shift assays using radiolabelled oligo αG2 incubated with crude or partially purified C4 protein fractions (see Materials and Methods). As a control for the protein preparations, protein was also bound to a double-stranded oligonucleotide (oligo αP3a, see Materials and Methods) containing the α-globin gene CCAAT motif, which is bound by the ubiquitous CCAAT Binding Protein (CBP; (23)). All the partially purified protein preparations, including one from 14 day chicken erythrocytes, bind the CCAAT oligo (Figure 2A), whereas only the MELC, chicken erythrocyte and mouse liver proteins

Table I: GATAAG sequences in globin and non-globin genes.

			Position	Reference for protein binding in vitro.
Mouse	$\alpha 1_D$	cAaCTGATAAGGAt	-183	This communication.
Chicken	α_D	GATAAGATAAGGCC	-60	(19)
Chicken	α_D^{TT}	aAggAGATAAGGgt	-114	
Duck	α_D	GATAAGATAAGGCC	-60	
Mouse	$\beta^{maj}1$	GcaCAGATAAGGAC	-212	This communication.
Mouse	$\beta^{maj}2$	ttcCTGATAAGAAa	-512	
Mouse	$\beta^{maj}3$	GgatAGATAgAGAg	-70	
Mouse	β^{maj}IVS2	ttTCTGATAgGAAg		(27)
Mouse	zeta	GcTCTGATAAcAgg	-105	
Human	zeta	GATCTGATAAGAAa	-105	
Chicken	β^{rho}	AAGATAAGG	-200	
Chicken	β_H 1	GgaAAGATAgcAAa	-120	(17)
Chicken	β_H 2	acTgAGATtAGGgC	-55	
Human	A_γ	GgTgTGATAgAGtt	-190	(20, 40)
Human	δ_A	ttTAAGATAAGcAg	-80	
Chicken	$\beta_A$enh	GtctTGATAgcAAa	3'	(8)
Chicken	$\beta_A$enh	ttgCAGATAAAcAt	3'	(8)
Mouse	CA1	aAcCTGATAAGGgg	-180	
Human	PBG-D	aAagAGATAAGGCC	-70	This communication.
Mouse	GSHPx	GgcCgGATAAGGCg	-110	
Human	GSHPx	GgcCgGATGAGGCg	-110	
Rat elastase 1		aGctgCTGATAAGA	-40	(38)
Rat elastase 2		aaTAtCAGATAAAt	-155	(38)

Sequences are compared to the GATAAGATAAGG direct repeat in the promoter of the chicken α_D globin gene (19), and homologous sequences are in capitals and underlined. The position of the sequence relative to the cap site is shown, and references are given for reports of protein binding in vitro. Abbreviations are: CA1- Carbonic anhydrase; GSHPx - glutathione peroxidase; PBG-D - erythroid-specific promoter of the porphobilinogen deaminase gene promoter. $\beta_A$enh represents the two inverted variant GATAAG motifs of the chicken β_A globin gene 3' enhancer within site IV (8)

bind oligo αG2. The nucleoprotein complexes are specifically competed out by an excess of the corresponding unlabelled oligo αP3a or αG2. The electrophoretic mobility of the chicken erythrocyte αG2 nucleoprotein complex is reproducibly slightly faster than the MELC (Figure 2A) or K562 (data not shown) complexes, but it is unclear whether this is due to a difference in size or post-translational modifications of a related protein. There appear to be multiple liver αG2-nucleoprotein complexes whose electrophoretic mobilities are clearly different to those observed with the erythroid nuclear extracts, consistent with the differences in footprint patterns described above. Thus there are distinct GATAAG DNA-binding activities in liver and erythroid nuclei, but it is not yet possible to distinguish between the

Table II.

	Variant GATAAG oligonucleotides.	Molar excess to compete αG2
αG2	5'-GATCCGGGCAACTGATAAGGATTCCCA-3'	50
αG3	.....GATAAaGA.....	50
αG5	.....GATAAcGA.....	50
αG7	.....GATAgGGA.....	<200
αG4	.....GgTAAGGA.....	<4000
αG6	.....GATgAGGA.....	<1000
αG8	.....GAcAAGGA.....	<3000
αG9	.....GATAtaGA.....	200
αG10	.....GATAcaGA.....	400
$\beta^{H}H'$	5'-gatccgaaggaaaGATAgcaAatttta-3'	300
βG1	5'-gatcctctgcacaGATAAGGAcaaaca-3'	50
βIVS2	5'-gatcctttctGATAgGaAggttgagca-3'	1000
β^{A}en	5'-gatccttgcAGATAAacAttTtgcTATCaagacttgca-3'	200

Synthetic double stranded oligonucleotides are either variants (αG3-αG11) of the mouse α1 globin gene GATAAG sequence αG2, or correspond to sequences bound in vitro in the promoters of the chicken $\beta^{Hatching}$ ($\beta^{H}H'$) and mouse β^{major} globin (βG1) genes; or in the second intervening sequence of the mouse β^{major} globin gene (βIVS2); or in the 3' enhancer of the chicken β^{Adult} globin gene (β^{A}en) (see Table I). Relative binding affinities are estimated as the molar excess of unlabelled oligonucleotide required to compete for binding to radiolabelled αG2 oligonucleotide in gel shift assays as shown in Figure 5B (and data not shown).

possibility that they represent different proteins, or modified forms of a single protein. The liver proteins may be related to the pancreas- and liver-specific protein(s) which bind to at least two GATAAG motifs in the rat elastase gene promoter (see Table I, and (38)).

In contrast, when crude nuclear extracts are used in gel-shift assays, two oligo αG2 nucleoprotein complexes are detected with protein from MEL (both M707/T and F4.12B2 cell lines), but not with those from fibroblast (STO), macrophage (J774.2) or brain (Figure 2D). During fractionation of the crude protein extract by non-specific DNA-cellulose affinity chromatography, there is a tendency to lose the more slowly migrating complex II (Figures 2A and 2B). A similar effect is observed with chicken erythrocyte nuclear protein preparations (data not shown). Since the addition of several protease inhibitors (see Materials and Methods) during preparation and fractionation does not have a significant effect on the ratio of the two complexes, it is unlikely that complex I is a protease degradation product of complex II. However, the presence of the phosphatase inhibitor β-phosphoglycerate during protein preparation does appear to increase the yield of complex II in crude nuclear extracts.

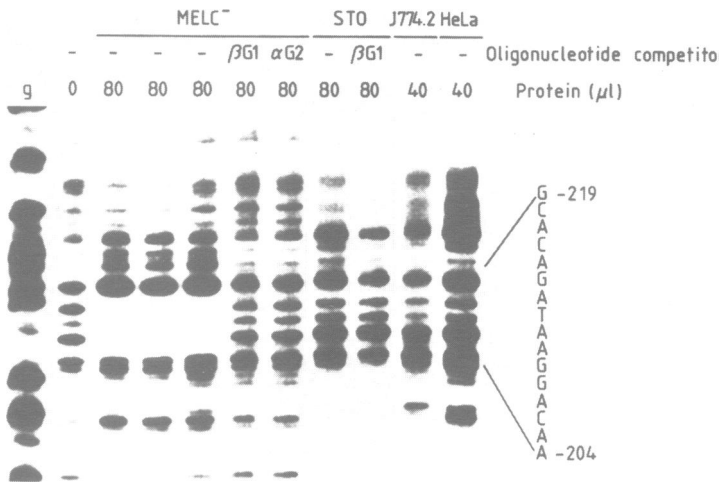

Figure 3.
Footprint analysis of the mouse β^{Major} gene promoter.
A 365 bp HindIII-HincII restriction fragment (-345 bp to +20 bp) of the mouse β^{major} globin gene promoter was subcloned, 5' end-labelled at the HindIII site and purified after secondary digestion with HincII. Labelled DNA (5 ng) was incubated with crude nuclear extracts (0-80 µl) from uninduced M707/T (MELC⁻), fibroblast (STO), macrophage (J774.2) or HeLa cells, and with 3 µg of poly(dI-dC):(dI-dC), in the presence (+) or absence (-) of 100 ng of cold competitor oligonucleotide (αG2 or βG1). After partial digestion with DNase I, nucleic acids were resolved by denaturing gel electrophoresis. G is a chemical sequencing markers, and 0 is the zero protein control. The footprint over the GATAAG motif at -204 to -212 bp is indicated.

The GATAAG-binding protein binds to the mouse β^{major} and human PBG-D promoters:

Analysis of the mouse β^{major} globin gene promoter also revealed an MELC-specific footprint at -206 to -222 bp upstream from the cap site (Figure 3) as well as other non-tissue specific footprints (data not shown). The bound sequence contains a GATAAG motif which is homologous to the α globin gene GATAAG sequence (Table I). Cross-competitions with oligonucleotides containing either the α globin (oligo αG2) or β^{major} globin gene GATAAG (oligo βG1) motifs (Table II), revealed that the binding patterns were indistinguishable in either footprint (Figure 3) or gel shift (see below, Figure 5A) analyses: this strongly suggests that the same erythroid specific factor binds to both GATAAG elements. Additional variant GATAAG motifs are present in the mouse β^{major} globin gene promoter, at positions -70 bp (GATAGAGA) and -512 bp (GATAAG) as well as in the second intervening sequence (IVS2, GATAGGAA). We have detected protein binding to the site at -512 bp

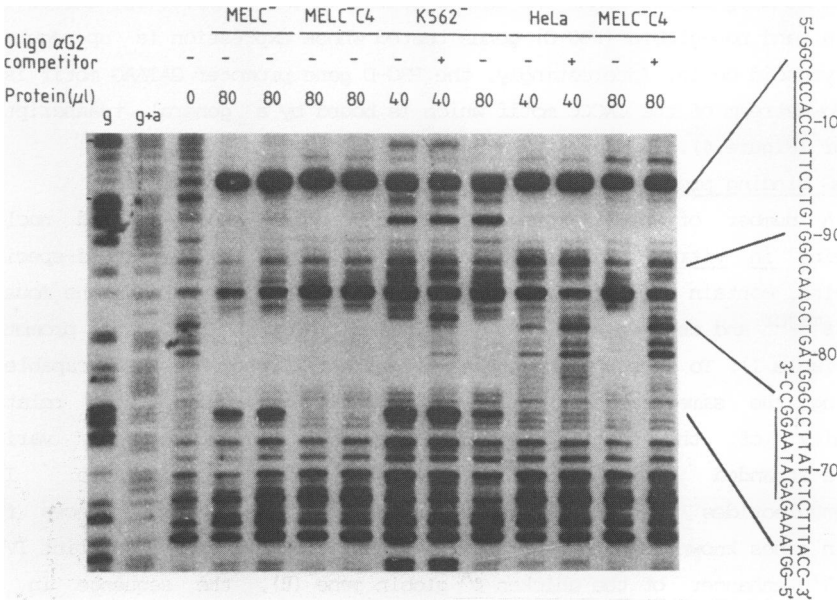

Figure 4.
Footprint analysis of the erythroid-specific promoter of the human PBG-D gene.
A subclone containing the erythroid-specific promoter of the PBG-D gene (48) was 5' end-labelled at an artificial BamHI site which corresponds to nucleotide +25 of the PBG-D promoter, and the insert isolated after secondary digestion with EcoRI (corresponding to -800 bp) and the BamHI-EcoRI fragment isolated. DNA (5 ng) was incubated with 0-80 μl of crude or partially purified (C4) protein from uninduced M707/T (MELC⁻) or K562 (K562⁻) cells, or from HeLa cells, in the presence (+) or absence (-) of cold competitor oligonucleotide αG2 (100 ng). After limited DNase I digestion the DNA was isolated and resolved by denaturing polyacrylamide gel electrophoresis and autoradiography. 0 is the zero protein control, and the sequences protected by protein are shown in the right hand margin as deduced from the chemical sequencing G+A and G tracks.

(data not shown), and others have shown protein binding to the IVS2 variant GATAAG motif (27) indicating that multiple copies of this element may be involved in tissue-specific gene expression.

A similar analysis of the erythroid-specific promoter of the human porphobilinogen deaminase gene (PBG-D) also revealed a footprint over a GATAAAG motif on the non-coding strand (Figure 4) which is bound by erythroid nuclear protein extracts from either mouse (MEL) or human (K562), but not from HeLa cells. This footprint is specifically competed out with the α globin gene GATAAG oligo αG2 sequence. Thus the same erythroid specific

GATAAG-binding factor binds the GATAAG motifs in the promoter regions of the globin and non-globin (PBG-D) genes tested whose expression is up-regulated in erythroid cells. Interestingly, the PBG-D gene promoter GATAAG motif is 20 bp downstream of the CACCC motif which is bound by a general transcription factor (Figure 4).

GATAAG binding protein interactions with variant GATAAG sequences:

A number of globin gene cis-acting control elements bind nuclear proteins *in vitro*, and several, which are bound by erythroid-specific proteins, contain similar sequences to the GATAAG motifs found in the mouse α and β^{major} and chicken α^{D} globin genes, and the human PBG-D gene promoters (see Table I). To test whether these variant GATAAG sequences were capable of binding the same factor, and more importantly to estimate the relative affinity of the protein for the sequences, a number of variant double-stranded oligonucleotides were synthesised (Table II). Oligonucleotides synthesised included variant GATAAG-like sequences from globin genes known to be footprinted *in vitro*; for example, footprint IV in the 3' enhancer of the chicken β^{A} globin gene (8), the sequence in the promoter of the chicken β^{H} globin gene which is bound by protein *in vitro* and *in vivo* (Footprint H'; (17)), and a sequence within the mouse β^{major} globin gene IVS2 (site B2, (27)). In gel shift assays the variant oligonucleotides were either tested as competitors against labelled oligo αG2 (Figure 5A and 5B), or labelled and bound to crude or partially purified nuclear MEL extracts (Figure 5C and 5D). The competition titrations permit an estimate of the relative affinity of the MEL protein for a sequence to be determined (Figure 5, summarised in Table II). On occasion, less complex (II) is detected in the control samples in the absence of competitor than when competitor oligo αG4 is added (Figures 5A and 5C), but this is not reproducible (see Figure 5B).

As shown in Table II, of the sequences tested, base substitutions of the ATA sequence (Oligos αG4, 6, and 8) severely reduce the relative affinity of the protein binding. Although the chicken β^{A} globin gene enhancer (footprint IV, (8)) contains two inverted variant GATAAG sequences, gel shift experiments suggest that the same number of protein molecules bind to it as to oligo αG2, since the predominant nucleoprotein complexes have the same electrophoretic mobilities (data not shown). There is no evidence (see Tables I and II) that sequences flanking the GATAAG motif in 30-40 bp oligonucleotides are involved in modulating the binding site's affinity for the GATAAG-binding protein. In conclusion, these binding studies indicate

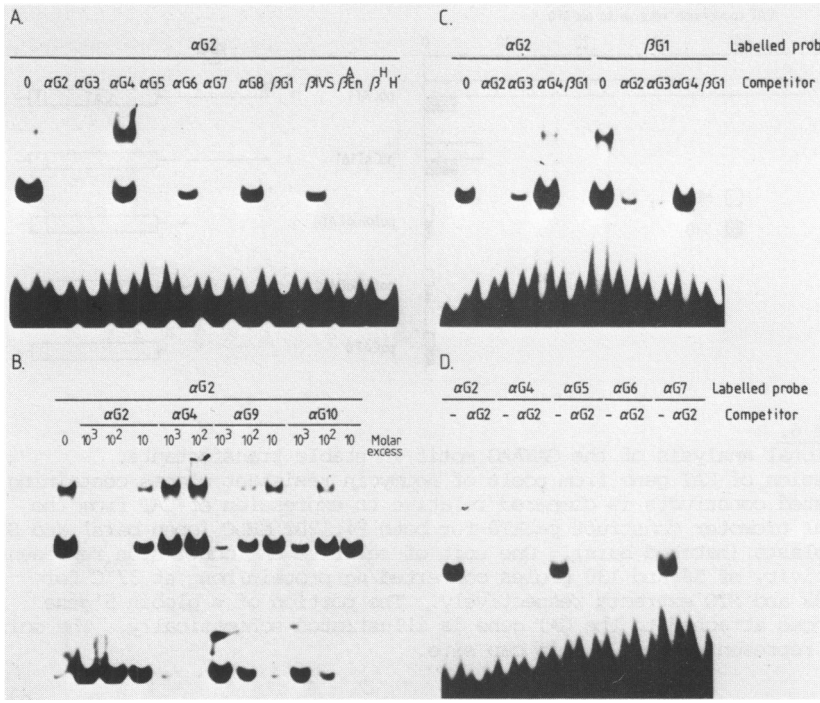

Figure 5.
Competition gel shift assays of variant GATAAG motifs.
A). Oligonucleotide αG2 (100 pg) was 5' end-labelled, incubated with uninduced M707/T (MELC$^-$) crude nuclear extracts in the presence of 6 µg of poly (dI-dC):(dI-dC), and in the absence (0) or presence of 100 ng of competitor oligonucleotides as shown (Table II).
B). Labelled oligo αG2 (100 pg) was bound to uninduced M707/T (MELC$^-$) crude nuclear extract (3 µl) in the presence of 6 µg of poly (dI-dC):(dI-dC) and increasing amounts of cold competitor oligonucleotides (Table II) expressed as a molar excess compared to the 100 pg of labelled αG2.
C). Labelled oligo αG2 or βG1, and 6 µg of poly (dI-dC):(dI-dC) was incubated with MELC crude nuclear extract (3 µl) in the presence or absence (0) of 100 ng of cold competitor oligonucleotide (Table II).
D). Labelled oligos αG2, αG4, αG6 or αG7 (100 pg) were incubated with partially purified M707/T C4 protein (3 µl) in the presence of 6 µg of poly (dI-dC):(dI-dC), and in the presence (+) or absence (-) of 100 ng of cold competitor oligo αG2.
Nucleoprotein complexes were resolved by non-denaturing gel electrophoresis and autoradiography.

that the same GATAAG binding protein is capable of binding to a variety of GATAAG-like sequences contained within globin gene promoters and enhancers (Table I), although the binding sites have different affinities for the protein.

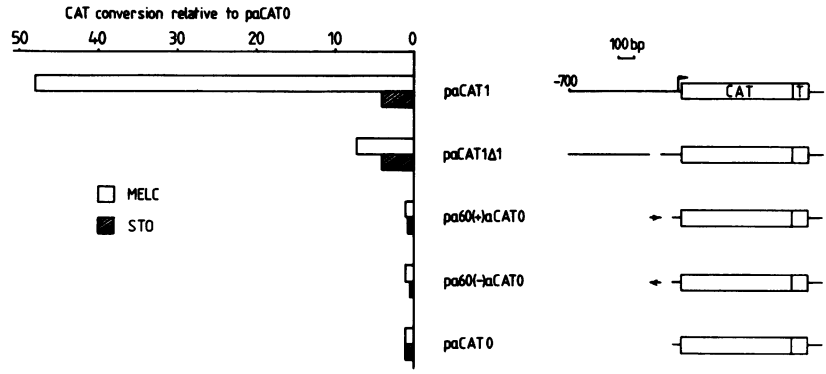

Figure 6.
Functional analysis of the GATAAG motif in stable transfectants. Expression of CAT gene from pools of neomycin resistant clones containing the indicated constructs is compared relative to expression of CAT from the minimal promoter construct pαCAT0 for both F4.12B2 MELC (open bars) and STO fibroblasts (hatched bars). One unit of relative CAT conversion represents an activity of 36 and 130 pmoles converted/μg protein/hour at 37°C for F4.12B2 and STO extracts respectively. The portion of α globin 5'gene sequences attached to the CAT gene is illustrated schematically. The solid arrow represents the α globin cap site.

Functional analysis of the GATAAG motif:

To test the effect of the GATAAG motif on transcription, α globin gene promoter sequences were linked to the bacterial chloramphenicol acetyltransferase (CAT) gene (Figure 6) and co-transfected into MEL or STO cells with a construct expressing the neomycin resistance gene driven by the Moloney murine sarcoma virus LTR promoter (Homer 6, (33)). Stable transfectants were selected by growth in G418 , and pools of at least 50 G418-resistant colonies assayed for CAT gene expression. A minimal promoter construct, pαCAT0, which contains promoter sequences from 1 bp 5' to the ATG to -52 bp relative to the cap site gives a low level of transcription in both STO and MEL cells. pαCAT1 which contains 700 bp of promoter sequence, gives a stimulation of transcription in both MEL and STO cells relative to pαCAT0, but this stimulation is 12-fold higher in MEL cells. To test the functional activity of the sequence containing the GATAAG erythroid-specific footprint observed at -183 bp, a derivative of pαCAT1 containing an internal deletion of sequences between the MspI site at -197 bp and the ApaI site at -129 bp was constructed (pαCAT1Δ1). Except for the footprint over the GATAAG motif at -183 bp, no other footprint was detected in the region between -129 bp and -197 bp using crude nuclear extracts from a variety of mouse and human cell

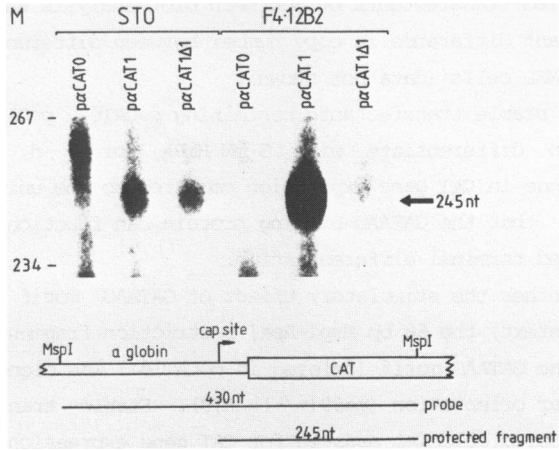

Figure 7.
S1 nuclease protection analysis of CAT RNAs expressed by stably transfected cell pools.
10 µg of polyA+ RNA from the STO fibroblast or F4.12B2 MELC stably transfected cell pools containing the indicated constructs was hybridised to the 430 nt single-stranded probe indicated in the line diagram. Following S1 nuclease digestion the products were resolved on a 6% denaturing polyacrylamide gel. M represents pBR322/HaeIII markers. The arrow indicates the 245 nt protected fragment corresponding to correct initiation at the α globin gene cap site.

lines (Figure 1 shows sequences from -150 bp to - 200 bp, and data not shown). As shown in Figure 6, pools of stable transfectants containing pαCAT1Δ1 have much lower levels of CAT gene expression than pαCAT1: the 12-fold erythroid specificity of gene expression of pαCAT1 is essentialy abolished, indicating that the 68 bp restriction fragment containing the GATAAG motif is largely responsible for the erythroid-specific stimulatory effect of the 700 bp globin gene promoter fragment on transcription in stable transfectants.

To test whether the effects observed at the level of CAT enzyme activity correspond to changes in transcription from the α globin gene cap site, S1 nuclease protection experiments were carried out using polyA+ RNA from the respective stably transfected pools and a continuously labelled single-stranded probe derived from pαCAT1 (Figure 7). A band representing the expected length of protection of 245 nt due to transcription initiating at the α globin gene cap site and extending into the CAT gene sequences is clearly visible and exhibits a steady state level in agreement with the CAT enzyme levels.

Determination of the average copy number of non-rearranged integrated CAT

genes in the pooled transfectants by Southern blot analysis showed that there was no significant difference in copy number between different plasmids or between STO and MEL cells (data not shown).

When the MEL stable transfectants containing pαCAT0, pαCAT1 and pαCAT1Δ1 were induced to differentiate with 5 mM HMBA for 3 d, there was no significant change in CAT gene expression compared to the uninduced levels, which indicates that the GATAAG-binding protein can function in MEL cells throughout induced terminal differentiation.

To test whether the stimulatory effect of GATAAG motif operates when placed out of context, the 68 bp MspI-ApaI restriction fragment (-197 to -129 bp) containing the GATAAG motif (deleted in pαCAT1Δ1) was cloned upstream of pαCAT0 in either orientation (pα60(+/-)αCAT0). Stable transfectants were generated and pooled colonies assayed for CAT gene expression. As shown in Figure 6, neither of the constructs showed any significant effect on CAT gene expression compared to pαCAT0 in either MEL or STO cells, indicating that the restriction fragment is non-functional when placed out of context of other cis-regulatory elements, and that it is possibly a subunit of a larger regulatory cis-element which requires the interaction of two or more element(s).

DISCUSSION:

In this paper we describe an erythroid-specific nuclear factor which binds to a GATAAG sequence motif in the mouse α and β^{major} globin gene promoters, and in the erythroid-specific promoter of the human PBG-D gene. The internal deletion of a 68 bp restriction fragment which contains the GATAAG motif from the α globin gene promoter reduces by 10 fold its ability to confer erythroid-specific transcription of a linked CAT gene in stable transfection assays. Competition gel shift analyses with variant GATAAG sequences indicate that this species-conserved protein can bind to a variety of similar motifs known to bind erythroid-specific nuclear protein(s) *in vivo* and/or *in vitro*: in the chicken β^{Adult} globin gene 3' enhancer and, α^{D} and $\beta^{Hatching}$ promoters (8, 17, 19) and human γ^{A} globin gene promoter (20,25,40). Multiple copies of variant GATAAG motifs are also present in the 3' enhancers of the chicken α^{A} (9) and human β (28) globin genes, and these are bound by erythroid-specific nuclear proteins. It remains to be seen whether the chicken histone H5 gene 3' enhancer (10) contains a GATAAG motif. The GATAAG motif is present in the promoter sequences of a variety of globin and non-globin genes (see Table I) whose expression is up-regulated in erythroid cells such as the mouse GSHPX gene and mouse CA1 gene

(P.Butterworth, personal communication) genes, although binding of the factor to all these sequences has yet to be demonstrated. These results suggest that the GATAAG binding-protein is involved in the tissue-specific up-regulation of globin and possibly several non-globin genes.

Several conserved cis-elements (the TATA , CCAAT and CACCC) motifs have been identified in the promoters of globin and non-globin genes which have been implicated in gene transcription (11-13) and although several distinct forms of the trans-acting factors interact with these elements, they are not tissue-specific (17,21-26,41). This implies that additional elements which interact with tissue-specific factors may be necessary for regulated transcriptional activation, and the interactions of the GATAAG motif with the erythroid-specific trans-acting factor seems to be one such component. However, as the binding protein is detected in erythroid cells (MEL, K562 and KMOE cells) prior to the initiation of active globin gene transcription, it appears to be necessary but not sufficient for globin gene activation; this may reflect a direct role in the transcriptional activation of genes expressed at earlier stages of differentiation (such as the glutathione peroxidase, carbonic anhydrase I and PBG-D genes).

GATAAG motifs appear to be conserved in promoter and enhancer sequences of many erythroid-specific genes, but they vary with respect to their context of surrounding sequences. For example, the chicken α^D globin gene promoter contains two direct repeats (GATAAGATAAG, (19)), the chicken β^A globin gene enhancer has two variant inverted repeats separated by 3 bp (Table II, (8)) and the chicken $\beta^{Hatching}$ (17) and mouse β^{major} globin gene promoters contain multiple GATAAG-like motifs (Table I). GATAAG motifs are also found within 15 bp of either the CACCC or CCAAT motifs in several promoters including the human PBG-D, the mouse CA1 genes, and the mouse β^{major}, chicken $\beta^{hatching}$ and several zeta globin genes. Thus the role of the GATAAG motif very likely depends on a number of parameters: i) whether it is present in the promoter or 3' enhancer of the gene; ii) its copy number; iii) the spatial relationship between multiple copies; iv) its orientation relative to the promoter and/or relative to other copies; v) the relative affinity of binding to the GATAAG binding protein; and vi) its location and orientation relative to other cis-control elements such as the TATA, CCAAT, CACCC sequences or other trans-acting factor binding sites.

Functional analysis of the mouse α globin gene promoter in stable transfection experiments demonstrates that a 68 bp internal deletion of a restriction fragment essentially abolishes the 12-fold stimulatory effect on

transcription in mouse erythroid (MEL) compared to non-erythroid (fibroblast) cell lines. In footprint experiments with crude nuclear extracts from MEL and fibroblast cells, the only sequence-specific binding detected to this 68 bp DNA sequence was over the GATAAG motif, strongly suggesting that the GATAAG motif is involved in the erythroid-specific stimulatory effect on transcription. However, when the GATAAG motif is placed on its own upstream of a minimal promoter (containing only the TATA motif) linked to the CAT gene, no erythroid specific effect on CAT gene expression is observed in stable transfectants, indicating that the motif is non-functional when placed out of context. Similarly, in transient transfection studies no effect on CAT gene transcription is observed when up to 5 head-to-tail copies of the GATAAG motif are placed in either orientation upstream of the minimal promoter (J. Frampton, unpublished results). This can be interpreted according to the proposal (42) that regulatory sequences such as enhancers are composed of multiple interacting cassettes ("enhansons"), each of which is not functional when taken out of context. A tissue-specific regulatory region could therefore be composed of cassettes bound by general transcription factors which would be inactive in the absence of binding to one or more additional cassettes by tissue-specific factors. This has been described for the three mouse α-fetoprotein gene enhancers (43), and the chicken β^{Adult} 3' enhancer, which is composed of five protein binding sites, three of which are bound by erythroid specific factors, but only one resembles the GATAAG sequence (8). On the other hand, the chicken α^{A} (9) and human β globin (28) gene 3' enhancers have multiple variant GATAAG motifs, indicating that the cassette composition of the enhancers is flexible.

There is increasing evidence for cooperative protein-protein interactions between proteins binding to cis-acting elements (42-46); for example, the interactions between the glucocorticoid receptor and the CACCC binding factor in the tryptophan oxygenase gene promoter (26). Competition between proteins for binding to a site have also been described. For instance, a displacement protein competes for binding with the CCAAT binding protein in the sea urchin histone H2B-1 gene promoter (47), and the TGGCA-binding protein competes for binding with distinct proteins in the promoters of the chicken $\beta^{Hatching}$ and β^{Adult} globin genes (17). Alternatively, a non-DNA binding protein can interact with the trans-acting factor directly as described for the cytosolic NF-kB inhibitor (39), and inhibition of the AP-2 dependant stimulation of transcription of the human metallothionein enhancer by T-antigen (44). The location of the GATAAG motifs identified so far indicate that a complex

number of interactions are possible between protein bound to the GATAAG sequence and to protein bound to CCAAT, CACCC, TATA or NF1 binding sites.

There is some indirect evidence that the activity of the GATAAG binding protein is modulated by protein-protein interactions which are themselves modulated by phosphorylation, as noted earlier. Thus, the protein-protein interactions, the role of phosphorylation in the GATAAG-binding protein's activity, and the functional relationship between the GATAAG motif and other cis-elements is currently being investigated.

ACKNOWLEDGEMENTS:

We thank R. Nicolas for his technical advice; A. Lang and C. Lang for the KMOE nuclear extract; and M. Goossens and D. Llewellyn for the PBG-D gene clones. We would also like to thank J. Wyke and N. Wilkie for critical reading of the manuscript. This research was supported by the Cancer Research Campaign. K.M. is supported by a Medical Research Council research Studentship.

+Present address: EMBL Differentiation Programme, Postfach 10.2209, Meyerhofstrasse 1, 69 Heidelberg, FRG

REFERENCES:

1. Harrison, P.R. (1988) In J. Garland, T.D. Dexter and N. Testa (eds), Cellular and Molecular Biology of Colony Stimulating Factors. Marcell Dekker, in press.
2. Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kolias, G. (1987) Cell, **51**: 975-985.
3. Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) Nucleic Acids Res., **15**: 5739-5747.
4. Trudel, M., and Costantini, F. (1987) Genes and Development, **1**: 954-961.
5. Antoniou, M., deBoer, E., Habets, G., and Grosveld, F. (1988) EMBO J., **7**: 377-384.
6. Bodine, D.M., and Ley, T.J. (1987) EMBO J., **6**: 2997-3004.
7. Choi, O-R., and Engel, J.D. (1986) Nature, **323**: 731-734.
8. Emerson, B.M., Nickol, J.M., Jackson, P.D., and Felsenfeld, G. (1987) Proc.Natl.Acad.Sci.USA., **84**: 4786-4790.
9. Evans, T., Reitman, M., and Felsenfeld, G. (1988) Proc. Natl. Acad. Sci. USA, **85**: 5976-5980.
10. Trainor, C.D., Stamler, S.J., and Engel, J.D. (1987) Nature, **328**: 827-830.
11. Myers, R.M., Tilly, K., and Maniatis, T. (1986) Science, **232**: 613-618.
12. Maniatis, T., Goodburn, S., and Fischer, J.A. (1987) Science, **236**: 1237-1245.
13. Cowie, A., and Myers, R.M. (1988) Mol. Cell Biol. **8**: 3122-3128.
14. Jackson, P.D., and Felsenfeld, G. (1985) Proc.Natl Acad.Sci.USA. **82**: 2296-2300.
15. Emerson, B.M., Lewis, C.D., and Felsenfeld, G. (1985) Cell, **41**: 21-30
16. Plumb, M.A., Nicolas, R.H., Wright, C.A., and Goodwin, G.H. (1985) Nucleic Acids Res., **13**: 4047-4065.
17. Plumb, M.A., Lobanenkov, V.V., Nicolas, R.H., Wright, C.A., Zavou, S., and Goodwin, G.H. (1986) Nucleic Acids Res., **14**: 7675-7693.

18. Benezra, R., Cantor, C.R., and Axel, R. (1986) Cell, **44**: 697-704.
19. Kemper, B., Jackson, P.D., and Felsenfeld, G. (1987) Mol.Cell Biol., **7**: 2059-2069.
20. Mantovani, R., Malgaretti, N., Giglioni, B., Comi, P., Cappellini, N., Nicolis, S., and Ottolenghi, S. (1987) Nucleic Acids Res., **15**: 9349-9364.
21. Barnhart, K.M., Kim, C.G., Banerji, S.S., and Sheffery, M. (1988) Mol. Cell Biol. **8**: 3215-3226.
22. Superti-Furga, G., Barberis, A., Schaffner, G., and Busslinger, M. (1988) EMBO J. **7**: 3099-3107.
23. Cohen, R.B., Sheffery, M., and Kim, C.G. (1986) Mol. Cell Biol., **6**:821-832
24. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) Cell, **48**: 79-89.
25. Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M.T., Caligaris-Cappio, F., and Ottolenghi, S. (1988) Nucleic Acids Res., **16**: 4299-4313.
26. Schule, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) Nature, **332**: 87-90.
27. Galson, D.L., and Housman, D.E.(1988) Mol.Cell Biol., **8**: 381-392.
28. Wall, L., de Boer, E., and Grosveld, F. (1988) Genes and Develop., **2**: 1089-1100.
29. Ralph, P., and Nakoinz, I. (1977) Cancer Research, **37**: 546-550.
30. Gaffney, D.F., McLauchlan, J., Whitton, J.L., and Clements, J.B. (1985) Nucleic Acids Res., **13**: 7847-7863.
31. Nishioka, Y., and Leder, P. (1979) Cell, **18**: 875-882.
32. Marsh, J.L., Erfle, M., and Wykes, E.J. (1984) Gene, **32**: 481-485.
33. Spandidos, D.A., and Wilkie, N.M. (1984) Nature, **310**: 469-475.
34. Graham, F.L., and van der Eb, A.J. (1973) Virology, **52**: 456-467.
35. Gorman, C., Moffat, L., and Howard, B. (1982) Mol. Cell Biol., **2**: 1044-1053.
36. Emerson, B.M., and Felsenfeld, G. (1984) Proc. Natl. Acad. Sci. USA, **81**: 95-99.
37. Maxam, A.M., and Gilbert, W. (1980) Methods Enzymol., **65**: 499-560.
38. Kruse, F., Komro, C.T., Michnoff, C.H., and MacDonald, R.J. (1988) Mol.Cell.Biol., **8**: 893-902.
39. Baeuerle, P.A., and Baltimore, D. (1988) Cell, **53**: 211-217.
40. Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., and Ottolenghi,S. (1988) Nucleic Acids Res. **16**: 7783-7797.
41. Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987) Cell, **50**: 863-872.
42. Ondek, B., Gloss, L., and Herr, W. (1988) Nature, **333**: 40-45.
43. Godbout, R., Ingram, R.S., and Tilghman, S.M. (1988) Mol. Cell Biol., **8**: 1169-1178
44. Mitchell, P.J., Wang, C., and Tjian, R. (1987) Cell, **50**: 87-861.
45. Rauscher, F.J., Cohen, D.R., Curran, T., Bos, T.J., Vogt, P.K., Bohmann, D., Tjian, R., and Franza, B.R. (1988) Science, **240**: 1010-1016.
46. Sigler, P.B. (1988) Nature, **333**: 210-212.
47. Barberis, A., Superti-Furga, G., and Busslinger, M. (1987) Cell, **50**: 347-359.
48. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., and Romeo, P-H. (1988) Proc. Natl. Acad. Sci. USA, **85**: 6-10.
49. Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**: 156-159.
50. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Laboratory Manual of Molecular Cloning. Cold Spring Harbor Publications, New York.
51. Bentley, D.L. (1984) Nature **307**: 77-80.
52. Weaver, R.F. and Weissmann, C. (1979) Nucleic Acids Res. **7**: 1175-1193.