The purification of an erythroid protein which binds to enhancer and promoter elements of haemoglobin genes

Neil D.Perkins, Robert H.Nicolas, Mark A.Plumb[1] and Graham H.Goodwin*

Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB and [1]Beatson Institute for Cancer Research, Glasgow, UK

ABSTRACT

An erythroid nuclear protein (EF1), originally detected as a protein binding within the nuclease hypersensitive site upstream of the chicken β^H-globin gene, has been purified. This protein of 37,000 − 39,000 molecular weight binds to three sites within the hypersensitive region: one between the CCAAT and TATA boxes, the second (further upstream) next to a NF1 binding site, and the third adjacent to a regulatory element found in a number of β-globin genes. The EF1 protein also binds to an erythroid-specific promoter element of the mouse α-globin gene and to two sites within the chicken β^A-globin enhancer. These six EF1-binding sites are related by the consensus sequence $A/_T GATAA/_G G/_C$. A minor protein of molecular weight 72,000 which co-purifies with EF1 also binds to the same sequences.

INTRODUCTION

The four chicken β-globin genes are transcriptionally regulated during embryonic development: the two embryonic β-globin genes, ϱ and ϵ, are expressed first in the earliest primitive red blood cells appearing in the blood islands and then later the β^H and β^A-globin genes are expressed in the definitive red blood cell lineage. The transcription of the latter two genes in red blood cells emanating from the bone marrow is maximal at about days 9 − 14 after fertilization. In the adult chicken the β^H-globin gene is totally inactive whilst β^A-globin gene expression is maintained [1,2].

These changes in expression of the β-globin genes are accompanied by the appearance of chromatin nuclease hypersensitive sites at the 5′ and 3′ ends of the genes and some of these mark important regulatory DNA elements. For example, the hypersensitive site at the 5′ end of the β^A-globin gene is absent in 5 day embryonic red blood cells but appears in 15 day red blood cells and remains in adult chicken erythrocytes [3]. Similarly the β^H-globin 5′ hypersensitive site appears at the same time but it is absent in adult chicken erythrocytes [4, and unpublished data]. The hypersensitive region downstream of the β^A-globin gene marks an enhancer element which is responsible for activating the β^A-globin gene in the adult red blood cell lineage but may also be important for activating the neighbouring downstream promoter of the ϵ-globin gene in the primitive cells [5−8]. Whether this enhancer activates the other globin genes is not certain.

Proteins which bind to DNA sequences within these nuclease hypersensitive sites have been described. Four protein factors, Sp1, Nuclear Factor 1 (NF1), a protein binding to the CACCC motif and a protein binding to a polyG sequence, bind to sequences within the 5′ hypersensitive site of the β^A-globin gene [4,9−11]. Two proteins bind to sequences within the 5′ hypersensitive site of the β^H-globin gene which are associated with protein in vivo [4,12]. One of these proteins is NF1, the other is a factor found at high levels

in erythroid cells which binds to a sequence adjacent to the NF1 binding site. This erythroid factor (termed EF1) has now been purified and shown to be a protein of 37−39 kd molecular weight. A minor 72 kd protein also binds to the same sequence. In this report we demonstrate that the EF1 protein also binds to two sequences within the β^A-globin enhancer described by Emerson *et al.* [10], and to the upstream promoter element of the mouse α-globin gene which is important for erythroid-specific expression. Two additional binding sites for this protein in the β^H-globin nuclease hypersensitive site have also been found.

METHODS
Extraction of red blood cell nuclear proteins and purification of protein EF1
Extraction of nuclear protein. All solutions contained the protease inhibitors PMSF (0.5 mM), benzamidine (0.5 mM), aprotinin (1 μg/ml), pepstatin A (1 μg/ml), leupeptin (2 μg/ml) and bestatin (5 μg/ml) with the phosphatase inhibitors levamisole (2 mM) and β-glycerophosphate (10 mM). Sodium butyrate (10 mM) was present throughout. Buffer E is 20% glycerol (v/v), 20 mM Hepes (pH 7.9), 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.1% Brij-35 with variable salt concentrations.

Blood from twelve dozen 14-day chick embryos was collected in ice cold phosphate buffered saline (PBS) containing 10 mM EDTA (pH 8.0), strained through gauze before being washed once with PBS, 10 mM EDTA, and twice with PBS. The erythrocytes were lysed by homogenising and washing with 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.4% Triton X-100. The nuclei were then washed three times with 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$. The nuclei (175 mg DNA) were resuspended at an approximate DNA concentration of 5 mg/ml in 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$ and 4M NaCl added to a concentration of 0.3M. The nuclei were stirred for 30 mins before being pelleted by centrifugation for 20 mins at 45,000g. The supernatant was adjusted to 100 mM NaCl, 20% glycerol (v/v), 20 mM Hepes (pH 7.9), 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.1% Brij-35 and centrifuged for 40 mins at 45,000g.
DNA-cellulose chromatography. The supernatant from the above extraction was directly loaded onto a calf thymus DNA-cellulose column (2.5g, Sigma) pre-equilibrated with buffer E containing 50 mM $(NH_4)_2SO_4$ and the non-DNA binding fraction eluted with the same buffer. The DNA-binding fraction (the DC protein fraction) was eluted in buffer E containing 250 mM $(NH_4)_2SO_4$ and then dialysed into buffer E containing 100 mM NaCl.
DNA-affinity chromatography. The DC-protein eluted in the previous chromatography was mixed with an approximately equal mass of poly(dI-dC).poly(dI-dC) (3.7 mg), incubated for 45 mins on ice before being centrifuged at 45,000g for 20 mins. The supernatant was loaded onto a 1×2.7 cm affinity column which had immobilised the polymerised AACTGATAAGGAT sequence (80 μg) (see below). The column was washed with 6 ml of buffer E containing 150 mM NaCl followed by a 60 ml gradient from 200 mM to 800 mM NaCl. The column was finally washed with 10 ml of buffer E containing 1M NaCl. The runthrough fractions were frozen at −70°C while the others were dialysed overnight into buffer E containing 100 mM NaCl and 50% (v/v) glycerol before being stored at −20°C. EF1 DNA-binding activity was assayed by gel retardation, assaying 1 μl. Peak EF1 activity eluted between 350 mM and 500 mM NaCl. Peak EF1 activity was diluted to 20% glycerol (v/v), incubated with 25 μg poly(dI-dC).poly(dI-dC) for 30 mins on ice before being centrifuged at 45,000g for 20 mins. The supernatant was loaded back onto the affinity column. The column was washed with buffer E containing sequentially 100 mM NaCl (4 ml), 250 mM NaCl (20 ml), 400 mM NaCl (6 ml), 500 mM NaCl (6 ml) and 1M NaCl (6 ml). EF1 protein free of contaminants elutes at 500 mM NaCl. EF1 protein

was dialysed into buffer E containing 100 mM NaCl and 50% (v/v) glycerol and stored at −20°C.

Affinity columns containing an unpolymerised 37 bp oligonucleotide containing the β^H-globin site B sequence (Fig. 1) or 25 bp oligonucleotide containing the mouse α-globin AACTGATAAGGAT sequence (Table 1) were also initially used in single cycle chromatographies to partially purify EF1 (carried out essentially as described above), giving fractions termed BS protein and AS protein, respectively.

Preparation of affinity column containing the polymerised AACTGATAAGGAT sequence
The oligonucleotides 5′-GATCGCAACTGATAAGGATTC-3′ and 5′-GATCGAA TCCTTATCAGTTGC-3′ (15 nmoles of each) were annealed in 50 mM Tris-HCl (pH 7.6), 10 mM MgCl$_2$ by heating to 90°C for 10 mins cooling to room temperature then incubating at 37°C for one hour. ATP and DTT were added to 1 mM and 10 mM respectively in 100 μl volume before adding 100 units of T4 polynucleotide kinase. After a 2 hr incubation at 37°C the mixture was cooled on ice and 8 units of T4 DNA ligase added before being incubated for 16 hrs at 16°C. The DNA was phenol extracted and passed down a Pharmacia NAP-5 gel filtration column pre-equilibrated with 100 mM potassium phosphate buffer, pH 8.0.

CNBr-activated Sepharose 4B (1g) was washed with 300 ml 1 mM HCl and 100 ml H$_2$0 by filtration. Potassium phosphate buffer, pH 8.0, was added to 10 mM followed immediately by ligated oligonucleotide and then mixed for 20 hrs at room temperature. The Sepharose was then washed with 100 ml 1M ethanolamine-HCl (pH 8.0), 100 ml 10 mM potassium phosphate (pH 8.0), 1mM EDTA, 100 ml 1M KCl, 1 mM EDTA and then finally 100 ml 10 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM EDTA, 0.02% azide before being packed into a column.

Photo-affinity labelling of EF1 protein
An 8 base oligonucleotide primer (1.25 pmoles), complementary to the 3′ end of the top strand of a 42 base oligonucleotide containing the B site, was annealed to 1.25 pmoles of this oligonucleotide (70°C for 10 mins, ice for 2 hrs). Annealing was carried out in 40 mm Tris-HCl (pH 8.0), 10 mm MgCl$_2$, 0.1 mm EDTA and 5 mm DTT (buffer K). Deoxynucleotides dGTP, dCTP and 5-bromodeoxyuridine triphosphate were added to a final concentration of 50 μM with 20 nmoles of 800 mCi/mmole α-^{32}P-dATP in a final volume of 100 μl K buffer. 30 units of Klenow fragment were added and incubated on ice for 2 hrs. The mixture was extracted with phenol and centrifuged through a 1 ml Sephadex G25 (DNA grade) column, prewashed in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 100 mM NaCl, to remove the bulk of the unincorporated nucleotides. The gel filtration was then repeated. This oligonucleotide (50 fmoles) was mixed with approximately 100 ng of partially purified EF1 (the BS-protein fraction) in 25 μl of 20 mM Hepes (pH 7.9), 60 mM NaCl, 1 mM MgCl$_2$, 4% Ficoll, 10 mM DTT, 5 μg poly(dI-dC).poly(dI-dC). Either 1 pmole of unlabelled site B oligonucleotide or 28 ng of HinfI digested PBR 322 were added as competitors. The mixture was incubated for 1 hr on ice, spotted onto Saran Wrap and irradiated on a Tm20 short wave U.V. transilluminator (UVP Inc.) for 4 mins. The mixture was then made 5 mM MgCl$_2$ and 2.5 mM CaCl$_2$ and incubated with 4 μg DNAseI for 10 mins at 37°C. The reaction was stopped with 10 mM EDTA (pH 8) and extracted with one volume of phenol. Bovine serum albumin (3 μg) was added to the phenol phase with 3.7 μl of 0.1M HCl and then acetone precipitated. The precipitate was analysed by SDS-15% polyacrylamide gel electrophoresis with Amersham 'Rainbow' molecular weight protein markers. The gel was dried and autoradiographed.
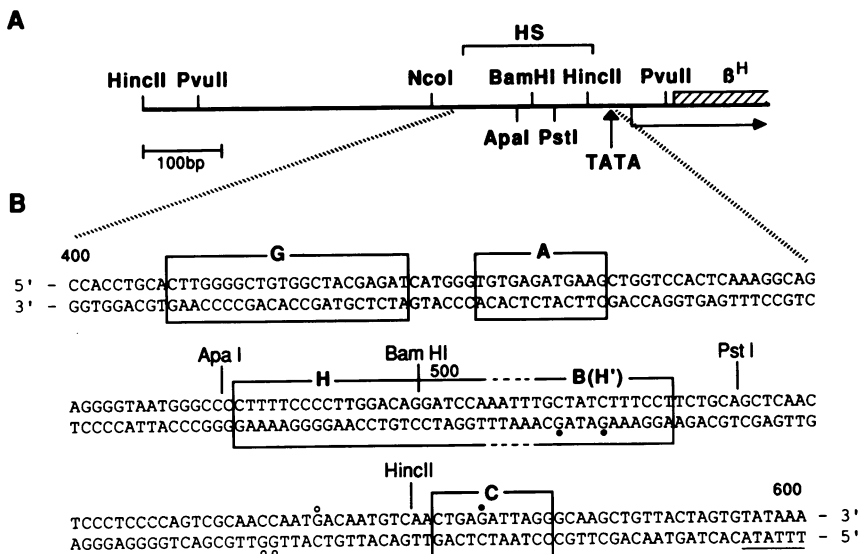
Fig. 1. A. Restriction map of the 5'-flanking region of the β^H-globin gene. HS marks the nuclease hypersensitive region in embryonic red blood cells. B. DNA sequence upstream of the TATA box (underlined) showing the three EF1 binding sites A, B and C. (Filled circles mark the Gs which when methylated inhibit EF1 binding to sites B and C). Open circles mark the Gs which when methylated inhibit the binding of a factor to the CCAAT box. Site H is the binding site for NF1. Site G is the binding site for an uncharacterised factor. This binding site is homologous to a regulatory sequence upstream of the human β-globin gene [20]. The base pairs are numbered from the 5' HincII site. Transcription initiates at approximately bp 620 (G. Partington, unpublished data).

Renaturation of EF1 protein eluted from SDS polyacrylamide gels

Semi-purified EF1 protein that has been passed down an affinity column once (the AS protein fraction) was concentrated by 10% TCA precipitation, washed with acetone and electrophoresed on a SDS 12% polyacrylamide gel as described by Hunkapiller *et al.* [13]. The proteins were electroblotted onto an Immobilon-P (Millipore) membrane as described below. The protein bands were visualised with 0.2% w/v Amido Black 10B and the proteins eluted with 1% (v/v) triton X100, 2% (w/v) SDS, 50 mM Tris-HCl (pH 8.8) 0.1 mg/ml BSA and concentrated by acetone precipitation. The protein samples were redissolved in 20 μl of buffer E containing 100 mM NaCl and incubated with thioredoxin for 48 hrs at 4°C as described by Szewczyk and Sommers [14].

Two-dimensional analysis of proteins associated with gel retarded DNA-protein complexes

Partially purified EF1 protein (400 ng of the AS protein fraction) was incubated for 1 hr with 20 pmoles of the 25 bp oligonucleotide containing the α-globin AACTGATAAGAT sequence and 32 μg poly(dI-dC).poly(dI-dC) and a trace of ^{32}P-labelled oligonucleotide. The mixture (64 μl) was loaded into two wells of a 3 mm thick TBE gel retardation electrophoresis gel. After electrophoresis one track was used for autoradiography. The other track was equilibrated with SDS electrophoresis loading buffer (two 30 min incubations with 15 ml buffer) and run into a second dimension SDS 12% polyacrylamide electrophoresis gel (3h, 30 mA) as described by O'Farrell [15]. The proteins were then electroblotted onto an Immobilon-P membrane (Millipore) prewetted with methanol. Transfer was carried out in 10 mM CAPS buffer (pH 11), 10% methanol at 0.5 amps for 2 hr [16]. The membrane was then washed three times with PBS, 0.3%
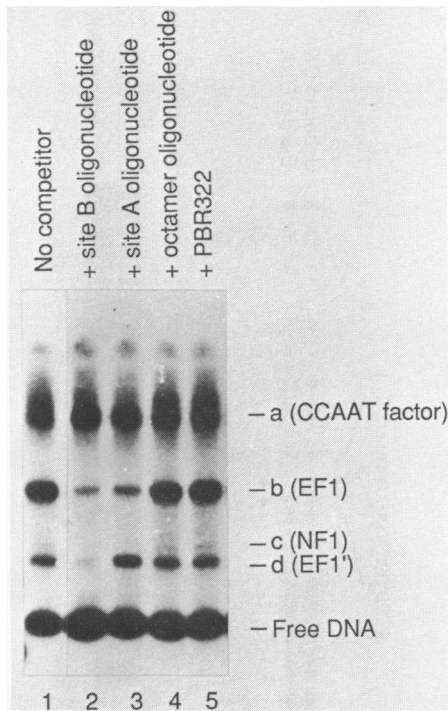
Fig. 2. Gel retardation analysis of RBC nuclear proteins binding to the 5'-flanking sequences of the β^H-globin gene. Embryonic RBC nuclear proteins (~ 2 µg of the total 0.3M NaCl nuclear extract) were incubated with the end-labelled NcoI-HincII fragment (~ 5 fmoles) and various competitor DNA fragments prior to electrophoresis: (lane 1) no competitor added; (lane 2) plus oligonucleotide containing the EF1 binding site B (300 fmoles); (lane 3) plus oligonucleotide containing the EF1 binding site A (300 fmoles); (lane 4) plus oligonucleotide containing the binding sequence for the octamer transcription factor (300 fmoles); (lane 5) plus PBR322 cut with HinfI (10 ng).

Tween 20 (once at 37°C and twice at room temperature) and stained with Aurodye Forte (Janssen).

DNA-binding assays

DNAse-I footprinting, gel retardation and methylation interference assays were carried out as described previously [4,17] except that gel retardation incubations were carried out in 1 mM MgCl$_2$. Footprinting was carried out in 2 mM MgCl$_2$ except where stated. End-labelling of genomic DNA fragments was carried out using PUC plasmid subclones of the β^H-globin 5'-flanking sequences [4] and the SphI-SphI β^A-globin 3' enhancer fragment. After cutting with the appropriate restriction enzyme, polynucleotide kinase or reverse transcriptase were used to label 5'- or 3'- ends respectively. Oligonucleotides were labelled with reverse transcriptase.

RESULTS

Using footprinting techniques, we previously described the binding of two proteins to adjacent sequences (H and H') within the nuclease hypersensitive region upstream of the β^H-globin gene [4]. One of these factors was the ubiquitous transcription factor NF1, the other a factor present in erythroid cells but not in thymocytes. When a DNA fragment
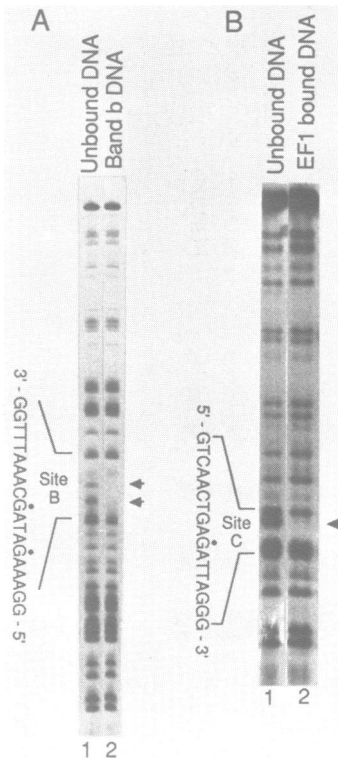
Fig. 3. Methylation interference of the binding of EF1 to (A) the site B sequence and (B) the site C sequence. The methylated NcoI-HincII fragment 5' end-labelled on the lower strand or the ApaI-PvuII fragment 3' end-labelled on the upper strand were incubated with partially purified EF1 protein (the BS protein fraction). EF1-bound DNA and unbound DNA were separated by electrophoresis cleaved with piperidine and the fragments resolved on a sequencing gel. Lanes 1 and 2 show DNA from unbound and EF1 bound DNA respectively. Arrows show positions where cleavage is reduced.

encompassing this region (the NcoI-HincII fragment, Fig 1) was used in gel retardation electrophoresis experiments, four retarded bands (a−d) were seen following incubation of nuclear extracts from 15 day chick embryo red blood cells with the labelled DNA fragment (Fig. 2). As described in detail below, using the gel retardation technique with specific competitor DNA fragments or oligonucleotides, and using the methylation interference technique, we have determined the binding sites for the proteins responsible for the gel-retarded DNA−protein complexes. Band c is due to the binding of NF1 to the H sequence shown in Fig. 1; this is the subject of a separate publication and will not be discussed further here [12]. (It should, however, be pointed out that band c is formed from a degraded or dissociated form of NF1 since when proteolysis is completely inhibited during nuclear protein preparation by the addition of all the protease inhibitors as described in Methods, this band disappears and a slower mobility smear is seen near band a.) Band a is formed by the binding of a CAAT box factor to the CCAAT motif 40 bp upstream of the TATA box (Fig. 1). Band b is formed by the binding of the erythroid factor to the previously described footprint sequence H' [4]. In this paper this footprint sequence
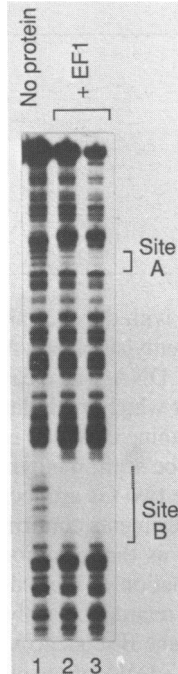
Fig. 4. DNAseI footprint analysis of partially purified EF1 protein. The NcoI-HincII fragment 5'-end labelled on the bottom strand was incubated with: (lane 1) no protein; (lane 2) 20 μl BS protein fraction; (lane 3) 40 μl BS protein fraction. The MgCl$_2$ concentration was 5 mM.

is henceforth referred to as site B and the erythroid factor which binds to it as EF1. Band d is formed by the binding of a modified form of EF1 (EF1'), since comparison of nuclear extracts prepared in the presence of only phenyl-methyl sulphonyl fluoride to inhibit proteolysis and extracts prepared in the presence of all the protease inhibitors indicate that the protein giving band d may be a degraded or dissociated form of the EF1 giving band b. We have found that inclusion of phosphatase inhibitors also inhibits conversion of band b EF1 to band d EF1' during nuclear preparations and chromatography of the EF1 protein. Thus it appears that EF1 is rather labile to proteases and phosphatases. One possibility is that native EF1 is a dimer which can dissociate to a monomer EF1' (which then gives band d) by dephosphorylation or proteolytic trimming of N-or C-terminal ends. We have also noticed that addition of sodium butyrate to buffers during the collection of blood and the subsequent manipulation results in elevated levels of the EF1 band b.

In order to identify the sequences bound by the proteins responsible for the gel retarded bands, the red blood cell nuclear extract was first fractionated by chromatography on a calf thymus DNA-cellulose column followed by affinity chromatography using a column composed of a DNA fragment containing the B sequence immobilized on Sepharose 4B. Gel retardation was used to analyse the fractions eluting from the affinity column (not shown). The protein giving gel retard band a elutes in the unbound fraction whilst proteins giving bands b and d elute with 0.3M NaCl (the BS protein fraction). Methylation interference on the formation of band b was carried out by mixing the above partially

Table 1. Sequences bound by EF1.

Mouse α-globin promoter	AACTGATAAGGAT
β^H-globin site A	GTGAGATGAAG
β^H-globin site B	GAAAGATAGCAAA
β^H-globin site C	CTGAGATTAGG
β^A-globin enhancer site IVa	TGCAGATAAACAT
β^A-globin enhancer site IVb	TCTTGATAGCAAA
Putative core consensus sequence:	AGATAAG
	T̲ ̲G̲C̲

purified protein in the bound fraction with the end-labelled NcoI-HincII fragment that had been randomly methylated with dimethylsulphate. After electrophoresis of the mixture, protein-bound (band b) and unbound DNA was cleaved with piperidine and analysed on a sequencing gel to determine sites at which methylation inhibits binding. Fig. 3A shows the positions at which methylated guanine cleavage is reduced in the protein bound DNA as compared with free DNA. As can be seen, methylation of two Gs on the bottom strand inhibit binding of protein EF1. These two Gs are located within the previously described footprint B (Fig. 1). Footprinting experiments confirmed that this partially purified protein preparation gives the same footprint as that described previously (Fig. 4). Methylation interference experiments on the formation of band d showed that EF1′ also binds to the same sequence (not shown). In the gel retardation analysis shown in Fig. 2, added unlabelled oligonucleotide competitor containing the B sequences competes for EF1 protein since band b (and d) is decreased (lane 2). PBR322 DNA or an oligonucleotide containing the sequence bound by the 'octamer' transcription factor [18] do not compete for EF1.

The binding site for the protein giving the gel retardation band a was similarly identified by methylation interference. The methylated Gs that inhibit specific binding are marked (open circles) in the sequence of Fig. 1. Methylation of three Gs within and adjacent to the CCAAT box inhibit protein binding, showing that the protein giving gel retardation band a is a CCAAT-box factor. Competition gel retardation experiments confirm that DNA fragments containing the CCAAT motif (e.g. the Pst1-HincII fragment) inhibits binding of the factor. However, DNA containing the NF1 consensus inverted repeat [12] does not compete strongly for the factor (data not shown). Thus the factor may not be NF1/CTF.

The EF1 factor has been found to bind to a number of other sites associated with erythroid expressed genes. We noticed that the β^H-globin site B sequence is very similar to a sequence within a footprint in the 3′ enhancer of the β^A-globin gene described by Emerson et al. [10]. Emerson et al. [10] demonstrated that nuclear extracts gave multiple footprints in this enhancer; one of the footprint sequences (footprint IV) was shown to bind an erythroid factor and to be important for enhancer activity. The 3′ half of this sequence (designated here IVb) is very similar to the β^H-globin site B (Table 1). As described in more detail below EF1 binds to this sequence and to the neighbouring sequence IVa which is also homologous to site B (Table 1). Similarly EF1 binds strongly to a homologous sequence in the upstream promoter region of the mouse α-globin gene (bases −173 to −189; Table 1) which we have shown to be important for erythroid specific gene expression [19]. Thus EF1 binds to sequences implicated in erythroid-specific gene expression. Two additional EF1 binding sites in the 5′-flanking sequences of the β^H-globin gene have been detected (sites A and C, Fig. 1). Mixing the partially purified EF1 (the BS-protein fraction) with the NcoI-HincII fragment gives a weak footprint on site A upstream of footprint B (Fig.
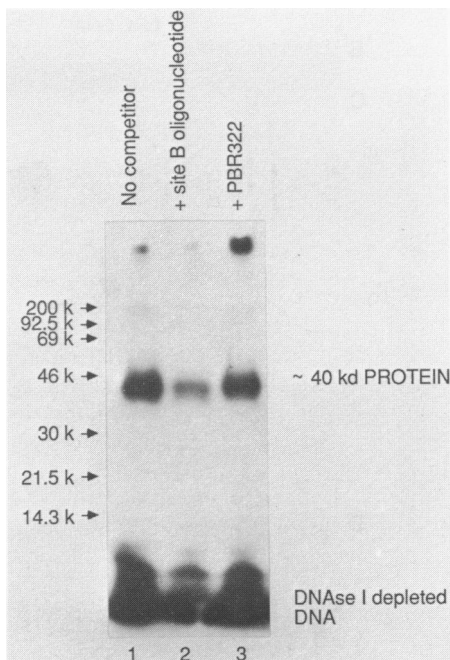
Fig. 5. SDS gel electrophoresis of photoaffinity labelled EF1. Partially purified EF1 protein (the BS protein fraction) was UV crosslinked to a labelled oligonucleotide containing site B in the presence of: (lane 1) no competitor DNA; (lane 2) a 50 fold molar excess of unlabelled oligonucleotide containing site B or (lane 3) PBR322 DNA (28 μg). Excess DNA was removed by DNAseI digestion, the labelled protein electrophoresed through an SDS-15% polyacrylamide gel and detected by autoradiography. The position of the 14.3−200 kd 'Rainbow' molecular weight markers are shown on the left.

4). Also an oligonucleotide containing site A sequence competes for EF1 binding to the NcoI-HincII fragment (Fig. 2, lane 3). Similarly, using the ApaI-PvuII fragment, a footprint on site C is seen (not shown) and methylation interference experiments showed that one G within site C inhibits EF1 binding (Fig. 3B). Methylation interference has not been carried out on site A. Using the gel retardation binding assay with varying amounts of competitor oligonucleotides containing the above EF1 binding site (A,B,C,IV, mouse α-globin promoter) approximate relative binding affinities of EF1 for the binding site could be obtained. This data (not shown) demonstrates that sites B and IV have approximately the same affinities whilst site C is 10-fold higher. The mouse α-globin EF1 binding site is the strongest (approximately 5-fold higher affinity than site C). Site A is the weakest, being about 2-fold weaker than site B.

The degraded or demodified form of EF1 (EF1′) appears not to bind to site A since band d in Fig. 2, lane 3 is not competed out by site A oligonucleotide. This may be due to our observation (not shown) that EF1′ elutes from DNA-affinity columns at a lower ionic strength than intact EF1 indicating that EF1′ binds with a weaker affinity to specific sequences.

In order to facilitate identification of the EF1 protein during its purification the molecular weight of the protein was determined by photo-affinity labelling. A labelled oligonucleotide
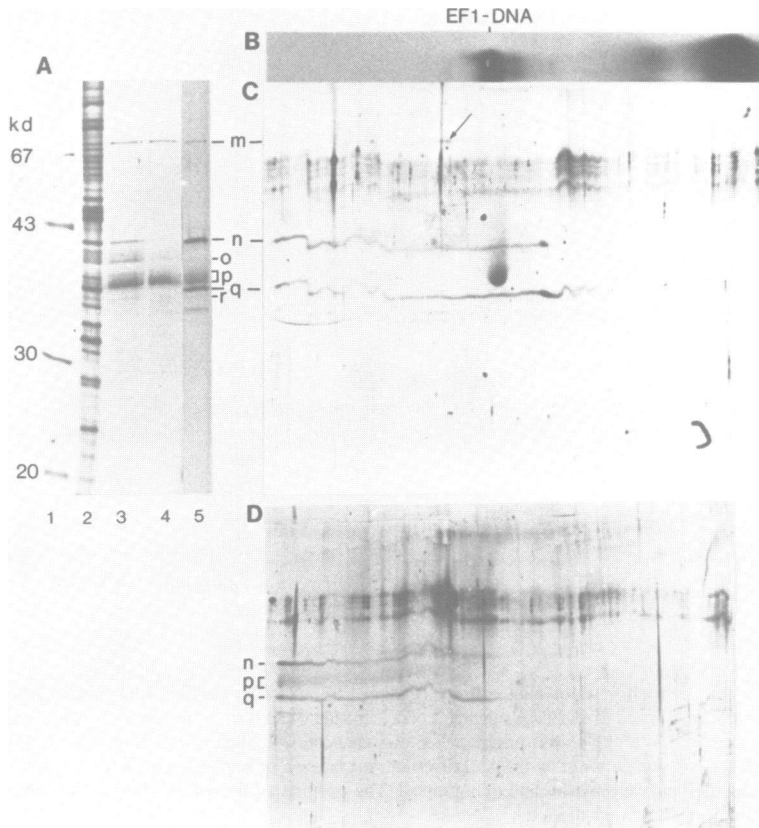
Fig. 6. Electrophoretic characterisation of EF1. A. One-dimensional SDS 12.5% polyacrylamide gel electrophoresis of (lane 1) 20—67 kd molecular weight markers; (lane 2) total DC protein fraction; (lane 3) 1st cycle affinity chromatography protein eluted from polymerised AACTGATAGGAT column; (lane 4) 2nd cycle affinity chromatography protein eluted from polymerised AACTGATAAGGAT column; (lane 5) the AS protein fraction. B. Gel retardation electrophoretic separation following mixing of AS protein fraction with labelled AACTGATAAGGAT oligonucleotide (migration left to right). C. Second dimension SDS polyacrylamide gel electrophoresis of proteins associated with gel retarded bands of Fig. 6B. The major spot is the 37—39 kd EF1 protein (band p); the minor arrowed spot is the 72 kd protein (band m). D. Control for the two dimensional analysis of panels B & C in which the AS protein was mixed and analysed as described above but without labelled oligonucleotide. (The four bands of ~60 kd which appear in the second dimension SDS gel and which run completely across the gel are contaminants, probably skin keratins, which accumulate during the analysis).

containing the EF1 B binding site was synthesised containing deoxybromouridine in the bottom strand. The oligonucleotide was mixed with partially purified EF1 with and without unlabelled competitor DNA fragments. The samples were cross-linked with U.V. light, digested with DNAseI and labelled protein analysed by SDS-gel electrophoresis. The autoradiograph of the gel (Fig. 5) shows a labelled protein of molecular weight ~40 kd. The autoradiographic signal of the band is reduced if the specific oligonucleotide competitor is included in the incubation prior to cross-linking but not if PBR322 DNA is included.

Thus denatured EF1 is a protein of molecular weight ~40 kd. This was confirmed using a two-dimensional electrophoretic technique similar to that described by Rupp and Sippel [20] and using protein partially purified by affinity chromatography as follows. DNA-cellulose fractionated nuclear protein was loaded onto a Sepharose column which had immobilised an oligonucleotide containing the strong EF1 binding site AACTGAT AAGGAT found in the mouse α-globin promoter. The column was eluted with a salt gradient and the fractions assayed for EF1 binding activity and the proteins analysed by SDS-electrophoresis. Most of the proteins elute in the flow through peak and the salt gradient line elutes a series of bands between 35 and 40 kd plus a 72 kd protein. The fractions with peak EF1 activity were pooled. An SDS gel electrophoresis of this material (termed AS protein fraction) is shown in Fig. 6A, lane 5. The concentrated pooled protein was mixed with a labelled oligonucleotide containing the AACTGATAAGGAT sequence and the protein−DNA complex electrophoresed through two tracks of a TBE gel. The whole lengths of both tracks were excised; one was autoradiographed (Fig. 6B), the proteins in the other were run into a second-dimension SDS-polyacrylamide gel (Fig. 6C). The proteins were then detected by blotting onto Immobilon-P and gold-staining. Fig. 6C shows that a protein of 37−39 kd is associated with the oligonucleotide in the major first dimension gel retarded band. This protein smears upwards slightly (suggestive of postsynthetic modification) and corresponds to the slightly fuzzy band p of the one-dimensional SDS electrophoresis of the proteins (Fig. 6A, lane 5). The two major contaminants n and q appear as streaks in the first dimension, presumably due to their binding to low molecular weight poly(dI-dC).poly(dI-dC) and these complexes migrate into the gel. When no labelled oligonucleotide is added to the two-dimensional analysis, proteins n, p and q give streaks, but no discrete spot of protein p is seen (Fig. 6D). Thus EF1 corresponds to the 37−39 kd protein p. This was further confirmed by eluting each of the bands m−r from a one-dimension SDS gel electrophoresis such as that shown in Fig. 6A, lane 5. The eluted proteins were renatured in the presence of thioredoxin as described in Methods and assayed for EF1 binding activity. Only band p protein gave the EF1 gel retardation band (data not shown).

In order to purify EF1 to homogeneity using affinity chromatography, it was necessary to use a column containing the polymerised AACTGATAAGGAT sequence immobilized to Sepharose 4B essentially as described by Kadonaga and Tjian [21]. Using this column the DNA-cellulose fractionated EF1 was loaded onto it in the presence of poly (dI-dC).poly(dI-dC) and eluted with a salt gradient. The fractions were assayed for EF1 binding activity (Fig. 7B) and the proteins analysed by SDS electrophoresis (Fig. 7A). As can be seen from Fig. 7 the elution of EF1 activity (fractions 16−26) correlates with the elution of protein band p. Protein band q elutes ahead of p. Band n trails back into the EF1 fractions which also contain minor bands m, o and r. To remove contaminants the pooled EF1 fractions (Fig. 6A, lane 3) were recycled through the affinity column and eluted with salt steps. Fig. 6A, lane 4 shows the SDS gel electrophoretic analysis of the peak EF1 binding activity eluting with 0.5M NaCl. The 37−39 kd band p protein is seen to be purified by the two cycling steps. A two-dimensional gel retardation followed by SDS electrophoresis analysis of this protein fraction as described above for Fig. 6B and C revealed just one major spot of the 37−39 kd p protein (not shown). Approximately 1 μg of EF1 protein was obtained from 40 mg of nuclear protein extract; approximately 1% of the DNA-binding activity was recovered.
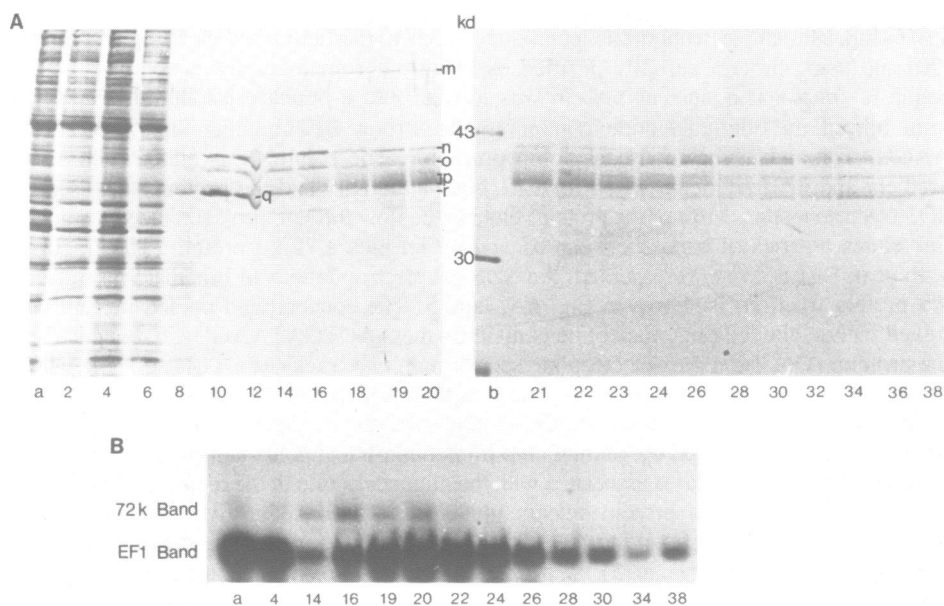
Fig. 7. Affinity chromatographic purification of EF1 using a polymerised AACTGATAAGGAT column. A. SDS-polyacrylamide gel electrophoresis of: (lane a) initial DC-protein fraction; (lane b) 30 and 43 kd molecular weight markers; (lanes 2, 4, 6) run-through fractions 2,4 and 6 from the affinity column; (lanes 8−38) fractions (2 ml) eluted from the column with the salt gradient − the lane numbers refer to fraction numbers. B. Gel retardation analysis of: (lane a) initial DC-protein fraction; (lane 4) run-through fraction 4 from the affinity column; (lanes 14−38) fraction numbers 14−38 from salt gradient elution of the affinity column.

A minor 72 kd protein (band m) is also seen to be co-purified by the affinity column; from densitometry measurements this protein is about 5% of EF1 in the final preparation. The protein gives a specific gel retardation complex visible as a slower migrating band in Fig. 7B and gives a faint protein spot (arrowed) in the two-dimensional analysis of Fig. 6C. Competition gel retardation analyses demonstrate that the binding of this protein to EF1 DNA binding sites is also specific and that it binds to sites A,B,C and AACTGATAAGGAT with the same order of affinities as EF1 (data not shown). This minor component therefore appears to be related to EF1. Gel retardation analysis of fibroblast nuclear extacts did not detect this protein; thus it may be erythroid-specific.

Footprinting of the purified EF1 protein on the β^H-globin upstream sequences and the β^A-globin 3'enhancer is shown in Fig. 8. The two footprints IVa and IVb in the enhancer (Fig. 8C) (identical to the footprint IV described by Emerson et al. [10]) and the two footprints B and C in the β^H-globin upstream region (Fig. 8B) are seen. Surprisingly footprint A, which was previously observed using partially purified EF1 (Fig. 4), was not seen using the purified protein (Fig. 8A, lane 7) despite the fact that this binding site is apparently not much weaker than site B when assayed using crude EF1 preparations (see above). This may be simply because the purified protein is more dilute or being assayed under sub-optimal conditions (although this has not been fully investigated, footprint A appears to be sensitive to divalent metal ion concentrations since this footprint was observed with the DC protein incubated in 1 mM $MgCl_2$ (Fig. 8A, lane 6) but not in 2 mM $MgCl_2$
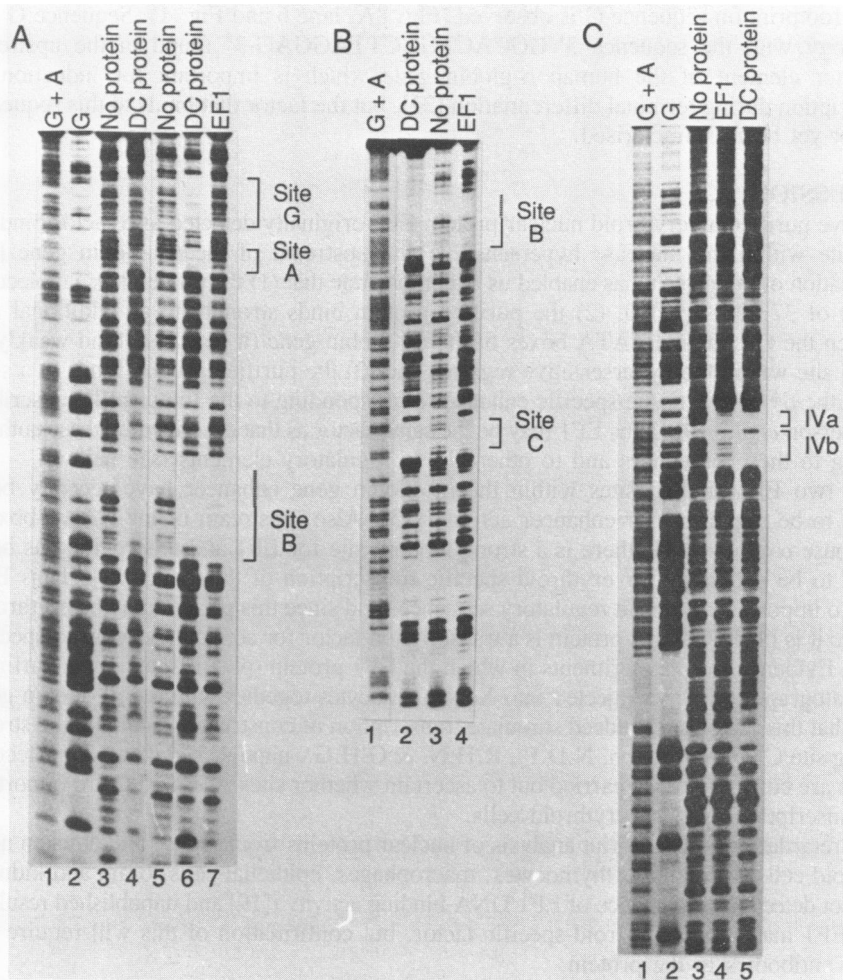
Fig. 8. Footprints of the factor EF1 on the 5'-flanking sequences of the β^H-globin gene and the β^A-globin 3'-enhancer. A. The NcoI-HincII 5'-β^H-globin fragment (~2 ng) 5'-end labelled on the bottom strand was mixed with: (lanes 3 and 5) no protein; (lanes 4 and 6) 5 μg DC-protein and 5 μg poly (dI-dC).poly(dI-dC); (lane 7) 40 ng affinity purified EF1. The samples were digested with DNAseI (5−100 ng) and analysed on a sequencing gel. All samples were incubated in the presence of 2 mM MgCl₂ except that shown in lane 6 which contained 1 mM MgCl₂. B. The ApaI-PvuII 5'-β^H-globin fragment (2 ng) 5'-end-labelled at the PvuII site was mixed with: (lane 2) 5 μg DC-protein plus 5 μg poly(dI-dC).poly(dI-dC); (lane 3) no protein; (lane 4) 40 ng affinity purified EF1. The samples were digested with DNAseI and analysed on a sequencing gel together with a G + A sequencing reaction (lane 1). C. The SphI fragment encompassing the enhancer 3' of the β^A-globin gene was 5'-end-labelled on the bottom strand and mixed with: (lane 3) no protein; (lane 4) 40 ng affinity purified EF1 protein; (lane 5) 5 μg DC-protein plus 5 μg Poly(dI-dC).poly(dI-dC). The samples were digested with DNAseI and analysed on sequencing gels together with G + A and G sequencing reactions (lanes 1 and 2).

(lane 4)). Alternatively, there may be a different factor to EF1 which binds to site A which is lost during the final purification using the AACTGATAAGGAT affinity column.

It is of interest to point out that using the DC protein fraction incubated in 1 mM MgCl₂

a new footprint on sequence G is observed (Fig. 8A, lane 6 and Fig. 1). Sequence G has homology with the sequence 5'-GGCACTGGCTTAGGAG-3' found in the upstream promoter element of the human β-globin gene which is important for induction of transcription during terminal differentiation [22], but the factor that binds to this sequence has not yet been characterised.

DISCUSSION

We have purified an erythroid nuclear protein EF1 originally detected as a factor binding to a site within the nuclease hypersensitive site upstream of the β^H-globin gene [4]. Purification of the protein has enabled us to demonstrate that (1) the protein has a molecular weight of $37,000-39,000$, (2) the purified protein binds strongly to an additional site between the CAAT and TATA boxes of the β^H-globin gene (it may also bind weakly to a third site within the hypersensitive region), and (3) the purified protein binds two sites within the β^A-globin tissue-specific enhancer corresponding to the footprint IV described by Emerson et al. [10]. Thus EF1 may be the same factor as that described by these authors binding to these sequences and to other globin regulatory elements (see below).

The two EF1-binding sites within the β^A-globin gene enhancer have recently been shown to be important for enhancer activity [23]. Also, upstream of the CAAT-box of the mouse α-globin gene there is a strong binding site for EF1 and this region has been shown to be necessary for erythroid specific transcription of this gene [19]. Thus EF1 binds to important erythroid regulatory sequences and since this protein may be erythroid-specific it is likely that this protein is a transcription factor for activating erythroid-specific genes. Evidence from experiments in which the EF1 protein (purified by a single affinity chromatography cycle) is injected into Xenopus oocytes together with the β^H-globin gene show that this factor does indeed stimulate transcription of constructs containing the strong binding site C (G. Partington, N.D.P., R.H.N. & G.H.G., unpublished data). Transfection studies are currently being carried out to ascertain whether sites A, B and C are important for transcription in avian erythroid cells.

Gel retardation and footprint analysis of nuclear proteins from mouse and chicken non-erythroid cells (fibroblasts, thymocytes, macrophages, epithelial cells, brain and kidney) have not detected the presence of EF1 DNA-binding activity ([19] and unpublished results). Thus EF1 may be an erythroid-specific factor, but confirmation of this will require the use of antibodies to the protein.

The factor that we have purified is likely to be the same as or related to the erythroid-specific factor (Eryf1) described by Felsenfeld and co-workers binding to the chicken α- and β-globin regulatory elements [24,25] (Felsenfeld, personal communication) and by Galson and Housman [26] binding within the hypersensitive domain of the mouse β-major globin intervening sequence since the binding sites described are all related to the consensus sequence given in Table 1. Binding sites for this factor have been identified near a number of other globin and non-globin erythroid expressed genes [19] suggesting that the factor may be important for activating all erythroid-specific genes.

Does the presence of EF1 correlate with the presence of the globin hypersensitive site and the expression of the genes? EF1 is present in the avian erythroid precursor cell HD3 (an erythroid cell line transformed with a temperature-sensitive avian erythroblastosis virus) [27] which expresses globin genes at very low levels and do not exhibit the globin hypersensitive sites ([28], and unpublished data). Nevertheless, on induction of the globin transcription by incubation of the cells at 42°C, the levels of EF1 rise $5-10$ fold to a

level comparable to that found in 11−14 day embryonic red blood cells (unpublished results). Thus, a sufficiently high level of EF1 may be important for transcriptional activation of chromosomal genes. It is of interest to note that Kretsovali et al. [7] found that the β^A-globin enhancer was functional when transfected into uninduced HD3 cells but was apparently less active than when transfected in fully differentiated embryonic red blood cells [6]. These results would be consistent with the low (but still functional) level of EF1 in uninduced HD3 cells. Comparison of the EF1 levels in 14-day embryonic red blood cells with adult chicken erythrocyte reveals that there is approximately three-fold less EF1 in adult chicken red blood cells which are transcribing the β^A-globin at very low levels. In adult erythrocytes the β^H-globin is completely inactive and its 5' hypersensitive site is absent (unpublished data). Thus the decline in transcription of the adult genes does not strictly correlate with EF1 levels.

The purified EF1 preparation also contains a 72 kd protein as a minor contaminant (5% of EF1). This protein is difficult to detect by gel retardation analysis in unfractionated nuclear extracts but becomes more apparent during the DNA-affinity chromatographic purification. Although we have not yet investigated in detail the sequences to which it binds and its tissue specificity, the 72 kd protein appears to bind to the EF1-binding sequences with about the same affinities as EF1 and it seems quite probable that the two proteins are related in some way; but there is no evidence to suggest that EF1 is a degradation product of the 72 kd protein or that EF1 can be artefactually converted to a larger molecular weight entity. There are a number of precedents for a single sequence binding several different proteins; for example the same sequence upstream of the adenovirus major late promoter can bind a 45 kd protein and a 116 kd protein [29]. Finally, the footprint G sequence within the DNAse I hypersensitive is of interest since it lies adjacent to the weak A site (and the protein binding to it may therefore assist binding of EF1 or a related protein to this site) and it is homologous to a sequence within a region upstream of the human β-globin gene which has been shown to be important for induction of the transfected gene in MEL cells [22]. Very similar sequences are found in a number of other β-globin genes. The factor binding for this sequence is currently being characterized.

ACKNOWLEDGMENTS

*To whom correspondence should be addressed

REFERENCES
1. Roninson,I.D. and Ingram,V.M. (1983) J. Biol. Chem. 258, 802−809.
2. Landes,G.M., Villeponteau,B., Pribyl,T.M. and Martinson,H.G. (1982) J. Biol. Chem. 257, 11008−11014.
3. Larsen,A. and Weintraub,H. (1982) Cell 29, 609−622.
4. Plumb,M.A., Lobanenkov,V.V., Nicolas,R.H., Wright,C.A., Zavou,S. and Goodwin,G.H. (1986) Nucl. Acids Res. 14, 7675−7693.
5. Hesse,J., Nickol,J.M., Lieber,M.R. and Felsenfeld,G. (1986) Proc. Natl. Acad. Sci. USA 83, 4312−4316.
6. Choi,O-R. and Engel,J.D. (1986) Nature 323, 731−734.
7. Kretsovali,A., Muller,M.M., Weber,F., Marcaud,L., Farache,G., Schreiber,E., Schaffner,W. and Scherrer,K. (1987) Gene 58, 167−175.

8. Nickol,J.M. and Felsenfeld,G. (1988) Proc. Natl. Acad. Sci. USA **85**, 2548−2552.
9. Emerson,B.M., Lewis,C.D. and Felsenfeld,G. (1985) Cell **41**, 21−30.
10. Emerson,B.M., Nickol,J.M., Jackson,P.D. and Felsenfeld,G. (1987), Proc. Natl. Acad. Sci. USA **84**, 4786−4792.
11. Evans,T., Dechiara,T., Efstratiadis,A. (1988) J. Mol. Biol. **199**, 61−81.
12. Rupp,R.A.W., Nicolas,R.H., Borgmeyer,L., Lobanenkov,V.V., Plumb,M.A., Sippel,A.E. and Goodwin,G.H. (1988) Eur. J. Biochem. **177**, 505−511.
13. Hunkapiller,M.W., Lujan,E., Ostrander,F. and Hood,L.E. (1983) Methods in Enzymology **91**, 227−236.
14. Szewczyk,B. and Sommers,D.F. (1988) Anal. Biochem. **168**, 48−53.
15. O'Farrell,P.H. (1975) J. Biol. Chem. **250**, 4007−4021.
16. Matsudaira,P. (1987) J. Biol. Chem. **262**, 10035−10038.
17. Goodwin,G.H. (1988) J. Virology **62**, 2186−2190.
18. Falkner,F.G., Mocikat,R. and Zachau,H.G. (1986) Nucleic Acids Res. **14**, 8819−8831.
19. Plumb,M.A., Frampton,J., Wainwright,H., Walker,M., Macleod,K., Goodwin, G. and Harrison,P. (1989) Nucleic Acids Res, in press.
20. Rupp,R.A.W. and Sippel,A.E. (1987) Nucleic Acids Res. **15**, 9707−9719.
21. Kadonaga,J.T. and Tjian,R. (1986) Proc. Natl. Acad. Sci. USA **83**, 5889−5893.
22. Antoniou,M., de Boer,E., Habets,G. and Grosveld,F. (1988) EMBO J. **7**, 377−384.
23. Reitman,M. and Felsenfeld,G. (1988) Proc. Natl. Acad. Sci. USA **85**, 6267−6271.
24. Kemper,B., Jackson,P.D. and Felsenfeld,G. (1987) Mol. Cell Biol. **8**, 2059−2067.
25. Evans,T., Reitman,M. and Felsenfeld,G. (1988) Proc. Natl. Acad. Sci. USA **85**, 5976−5980.
26. Galson,D.L. and Housman,D.E. (1988) Mol. Cell Biol. **8**, 381−392.
27. Beug,H., Palmieri,J., Freudenstein,C., Zentgraff,H. and Graf,T. (1982) Cell **28**, 907−919.
28. Weintraub,H., Beug,H., Groudine,M. and Graf,T. (1982) Cell **28**, 931−940.
29. Safer,B., Cohen,R.B., Garfinkel,S. and Thompson,J.A. (1988) Mol. Cell Biol. **8**, 105−113.