Characterisation of chicken erythroid nuclear proteins which bind to the nuclease hypersensitive regions upstream of the β^A - and β^H -globin genes

M.A.Plumb⁺, V.V.Lobanenko^{*}, R.H.Nicolas, C.A.Wright, S.Zavou and G.H.Goodwin

Institute of Cancer Research: Royal Cancer Hospital, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

Received 31 March 1986; Revised and Accepted 28 August 1986

ABSTRACT

Chicken erythrocyte sequence-specific nuclear DNA-binding proteins, which bind to the 5'-flanking DNaseI hypersensitive sites of the erythrocyte chromosomal β^A - and β^H -globin genes, have been fractionated by HPLC gel filtration. Three β^A -globin gene DNA binding activities (to sites A, B and B' (10-12)) were separated. The erythroid precursor cell line HD3 has β^A -globin gene sites B and B' binding activities, but binding to site A is detected only after the HD3 cells are induced to differentiate. The fractionated protein binds to a redefined site B', which contains at its center the globin CACCC consensus sequence. The chromosomal β^H -globin gene has two 5'-flanking DNaseI hypersensitive sites which bracket two sequences (H and H') bound by erythrocyte and HD3 nuclear protein in vitro. The β^H - and β^A -globin gene binding sites (H and B) contain variants of the sequences bound by Nuclear Factor 1 and the TGGCA-binding protein, and their protein binding activity(ies) co-purify after HPLC gel filtration.

INTRODUCTION

The observation that the chromatin structure of 5'-flanking gene sequences undergo changes in the nucleus as a consequence of gene activation or repression (1-3) has focused attention on these sequences as likely candidates for binding trans-acting protein factors which modulate gene transcription. Sequence-specific binding of nuclear DNA-binding proteins to 5'-flanking gene sequences which are DNaseI hypersensitive in the nucleus has been shown for a variety of eukaryotic systems including the *Drosophila* hsp70 heat shock locus (4,5), the SV40 early promoter and enhancer sequences (6,7), the human c-myc gene (8), the chicken β^A -globin gene (9-11) and the *Xenopus* 5S (12,14) and ribosomal genes (13). In a few cases it has been possible to ascribe a function to these activities; that is, to stimulate specific gene transcription in vitro.

Chicken erythrocyte nuclei contain proteins which bind to the 5'-flanking sequences of the β^A -globin gene (9-11) and have been shown to be capable of generating the nuclease hypersensitive site in vitro (9). The activity(ies)

is tissue specific, not being present in oviduct (9,11), only at low levels in thymus (10), and is specific for definitive erythrocytes which are expressing β^A -globin (9). In vitro footprint analysis has shown binding to two (11) or three sites (10) on the β^A -globin gene. Two of the sites are bound by protein in erythrocyte nuclei as demonstrated by genomic footprinting (15).

A number of erythroid cell lineages have been identified during red blood cell differentiation in the developing chick embryo (16,17,33,34). In the terminally differentiated definitive erythrocyte appearing at approximately day 12 after fertilisation, two of the developmentally regulated β -globin genes, β^A and β^H , are both activated (16), and the genes are in an active conformation in the nucleus as evidenced by the presence of endonuclease hypersensitive sites (18,19). The 5'-flanking DNaseI hypersensitive site of the coordinately regulated nuclear β^H -globin gene has been described (18), and we have now sequenced the 5'-flanking region of the β^H -globin gene, and fine mapped the erythrocyte nuclear DNaseI hypersensitive sites.

In order to compare the proteins which bind to the 5'-nuclease hypersensitive regions of these two developmentally regulated β -globin genes and to examine how these proteins change when the genes are activated during differentiation, the DNA-footprinting technique has been employed using nuclear proteins extracted from definitive chick red blood cells and from the avian erythroid cell line, HD3, a cell transformed by a temperature sensitive avian erythroblastosis virus (AEVts34) (21) before and after induction of terminal differentiation. The various DNA-binding activities identified in these nuclear extracts were assayed by DNA-footprinting before and after fractionation by HPLC gel filtration. The results confirm that the nuclei of definitive (15 day embryonic and adult) red blood cells have three distinct activities which bind specifically to three sequences within the β^A -globin gene nuclease hypersensitive region (10-11) and show that these activities can be separated by HPLC chromatography. One of these activities is absent in uninduced HD3 cells but appears when the cells are induced to differentiate. Two activities have been detected binding to the β^H -globin gene DNaseI hypersensitive region in both HD3 and definitive red blood cell nuclei. One of these proteins appears to be related to the human nuclear factor 1 (22,23) (or the analogous chicken TGGCA-binding protein (24)), a protein which binds to the origin of replication of adenovirus DNA (23) and which is required for the in vitro replication of the adenovirus genome (22). This protein may also be that binding to one of the β^A -globin sequences. These results show that there are multiple protein binding sequences upstream of the developmentally

regulated β^A - and β^H globin genes. Two of these sequences are bound by the same or closely related erythroid nuclear proteins, and may therefore be involved in the coordinate regulation of these two globin genes.

METHODS

Recombinant Clones

The genomic chicken β^A -globin (pBA650.1) and β^H -globin gene (pB2EH325) subclones have been described (10). The pB2EH325 900 bp EcoRI-HincII and 580 bp HincII restriction fragments were subcloned into the pUC9 (28) vector (pB^HEH4 and pB^HHC3 respectively) as shown in Figure 1A. The 5'-flanking region of the β^H -globin gene was sequenced by the Maxam and Gilbert procedure (25) using restriction fragments end-labelled either with [γ -³²P]-ATP (3000 Ci/mmol) and T4 polynucleotide kinase, or [α -³²P]-ddATP (5000 Ci/mmol) and terminal deoxynucleotidyl transferase (see Fig. 1).

HD3 cell growth and induction

Transformed HD3 cells (21) were subcloned to give a subline which gave 60-70% benzidine-positive cells after incubation (5×10^5 cells/ml) in 1% CO₂ at 42°C and pH 8.0 for three days. Large scale inductions (7×10^{10} cells) were performed in a cell factory (NUNC) (5×10^5 cells/ml) at 42°C, pH 8.0 and 1% CO₂ for three days, essentially as described (26).

Southern Blotting of Nuclear DNaseI digests

Nuclei from 15 day embryonic chicken erythrocytes were prepared and digested with increasing concentrations of DNaseI essentially as described (18,19). DNA was isolated, digested with restriction enzymes, resolved (10 μ g) by agarose gel electrophoresis and transferred to nitrocellulose in 20xSSC. The filters were prehybridised in solutions containing 1x Denhart's, 6xSSC, 2 μ gml⁻¹ BSA, 0.1 mgml⁻¹ carrier DNA, and hybridized with either (a) pB^HEH4 DNA (Fig. 1) nick translated in the presence of [α -³²P]-dCTP (3000 Ci/mmol) at 67°C, or (b) in 50% formamide at 47°C with T7 RNA polymerase transcripts of a pGEM1 (29) subclone of pB^HHC3, transcribed in the presence of [α -³²P]-UTP (3000 Ci/mmol) (29).

End-labelling restriction fragments for footprint assays

Plasmid DNA was linearised with EcoRI or HindIII, end-labelled with T4 polynucleotide kinase and [γ -³²P]-ATP (5000 Ci/mmol) and the insert isolated after secondary digestion (10).

Chicken nuclear protein fraction preparation

Nuclei were prepared from erythrocytes from mature chickens, erythrocytes from 15 day-old embryos, thymus from 3 week old chicks and HD3 cultured cells as

described previously (10). Nuclear proteins were extracted with 0.3M NaCl and fractionated by ammonium sulphate precipitation and DNA-cellulose affinity chromatography as described (9-11). Protein eluted from the DNA-cellulose with 250 mM $(\text{NH}_4)_2\text{SO}_4$ (C4 protein fraction (10)) was used in these studies.

HPLC gel filtration of C4 protein fractions

Protein fractions (~ 0.05 - 0.5 mg) were precipitated with 2 volumes of aqueous 50% (w/v) $(\text{NH}_4)_2\text{SO}_4$ (pH 7.5) at 0°C for 30 min, collected by centrifugation at 40000g for 1 hour and dissolved in 300 μ l of E250 buffer (250mM $(\text{NH}_4)_2\text{SO}_4$ 20mM Hepes (pH 7.9), 5mM MgCl_2 , 0.1mM EDTA, 0.1mM EGTA, 0.5% (w/v) Brij-35, 1mM dithiothreitol, 20% (w/v) glycerol). Gel filtration chromatography was carried out using a Zorbax GF250 (Du Pont Instruments) 25cm HPLC column in E250 buffer at a flow rate of 0.5ml/min. Ten 0.5-0.6ml fractions were collected and dialysed against storage buffer (9). The column was calibrated with BioRad standard molecular weight markers: Thyroglobulin (670000), gamma globulin (158000), ovalbumin (44000), myoglobin (17000) and vitamin B12 (1350).

Footprint assay

Protein (0-25 μ g) and 5'-end labelled DNA (10-40ng) were incubated with or without *E. coli* DNA (0-1 μ g) for 90 min at 4°C in a final volume of 100 μ l in storage buffer (9). The solution was digested with DNaseI, and the DNA resolved by denaturing 6-7% (w/v) polyacrylamide gel electrophoresis and autoradiography (10).

DNA sequencing reactions (25) were carried out and resolved in parallel with the footprint reactions.

RESULTS

Structure of the 5'-flanking region of the chicken β^H -gene.

The 5'-flanking region of the cloned genomic chicken β^H -globin gene was sequenced by the Maxam and Gilbert procedure (25) using the subclones and sequencing strategy shown in Fig. 1A. Partial sequencing from the SstI site of the pB^HEHC4 plasmid (Fig. 1A) confirmed the sequence of the β^H -globin gene second exon reported by Dolan *et al.* (30) (data not shown). The partially sequenced first exon (Fig. 1B) is homologous at both the nucleotide level (98%) and at the level of the N-terminal 20 amino acids encoded (100%), to the chicken β^A -globin gene first exon (31). The 5'-flanking sequence contains a 'TATA' and two 'CAAT' consensus sequences (78 bp, 111 bp and 113 bp, respectively) upstream from the translation initiation codon 'ATG' (underlined in Fig. 1B). Many globin genes (including the chicken β^A -globin gene (20))

A

B

10 20 30 40 50 60
AACACAAAGGCAACTGTGGAGAACTACCGTGGGCAACAACATATGGATNAATTCAAGGG
70 80 90 100 110 120
ACAACAGCTGTGAAACTCTTCATGAGTACTGACAAGAAAGGAGAACCTTGCATTCAAAGC
130 140 150 160 170 180
ATGAGGCCATATTGGGATACGCACTGAGCTCTCGTTTCCGGAAAGCCCTCTGATCATT
190 200 210 220 230 240
CAGTTTGGGGCTGATGATGCTTGGAGCTGTTACGATGGATTCTTAGGAGTGCAGGCAC
250 260 270 280 290 300
CAATGCACCATCAAGAGTCTGAAACAAGGCCAAGTAAAGGCCGTGAGCTCTTTGACATC
310 320 330 340 350 360
ACTGCCTAATCTCACTGGTCTCTCCACAAGTTTCAGCCCCCAACCACTGGCAGTGACACT
370 380 390 400 410 420
GGAAACCTATGGCCATGGAGGAGCCATGCAGGCAGCTCCACCTGCACTTGGGGCTGTGGC
430 440 450 460 470 480
TACGAGATCATGGGTGTGAGATGAAGCTGGTCCACTCAAAGGCAGGGGTAATGGGCC
490 500 510 520 530 540
CTTTTCCCTTGGACAGGATCAAATTGCTATCTTCTCTCTGCAGCTCAACTCCCTCC
550 560 570 580 590 600
CCAGTCGCAACCAATGACAATGTCACTGAGATTAGGGCAAGCTGTTACTAGTGTATAA
610 620 630 640 650 660
AATCCATGGGAAGACTTTGGCCCCACTTCTACTCATTGTTTCAGGGAAGCCAAAGCAAG
670 680 690 700 710 720
CCCTTCAGCTGCCACCATGGTGCCTGCTGAGGAGAAAGCAGCTCATCCGGCCT
730 740
CTGGGGCAAGTCAATGTG
Trp Gly Lys Val Asn Val

Fig.1: Restriction map and sequence of the 5'-flanking region of the chicken β^H -globin gene. (A) Restriction map of the pB2EH325 clone (10) and subclones. The blocked area represents the coding region of the β^H -gene. The horizontal arrows represent the sequencing strategy: E: EcoRI; PI: PstI; PII: PvuII; H: HindIII; HII: HincII; and S: SstI. (B) Sequence of the 680bp 5'-flanking region and 60 bp of the first exon of the β^H -globin gene. Underlined sequences represent putative 'TATA', 'CAAT' and 'CACCC' consensus sequences. The boxed area represents the sequence (sites H and H') bound in footprint analyses (Fig. 4) and the long horizontal arrows above the sequences show the two nuclear DNaseI hypersensitive sites mapped in Fig. 2. ($\rightarrow \leftarrow$) designates the inverted repeat sequence in site H/H'.

have a conserved 5'-CACCC-3' sequence upstream from the 'CAAT' consensus sequence, which has been shown to be important for determining the levels of the transcription of mammalian β -globin genes in vivo (33,34,35). The β^H -globin gene has the sequence GGGTG 123 bp upstream from the "CAAT" consensus sequences (nucleotide 433-437, Fig. 1B); read on the opposite strand this may correspond to an inverted CACCC element.

A 5'-flanking DNaseI hypersensitive site of the nuclear β^H -globin gene has been described (18) but not accurately mapped. Therefore, nuclei from 15-day embryonic chicken erythrocytes were digested with increasing amounts of DNaseI, the DNA isolated, digested with EcoRI and HindIII or PvuII and analysed by Southern blotting using pB^HHC3 plasmid as probe. As shown in Fig. 2A, an EcoRI-HindIII double digest reveals a 3.2 Kbp genomic fragment and, with increasing DNaseI digestion, reveals the appearance of two minor bands of $\sim$ 2.1 (II) and $\sim$ 2.2 (I) Kb in length. A control BamHI-HindIII double digest of total chicken DNA resolved in parallel shows that the two minor bands migrate on either side of the BamHI-HindIII genomic band, and maps the two DNaseI hypersensitive sites within $\sim$ 100 bp on either side of the BamHI

Fig.2: Southern blot analysis of DNaseI treated chicken erythrocyte nuclear DNA. 15-day embryonic erythrocyte nuclei were digested with increasing amounts of DNaseI, the DNA isolated and analysed by Southern blotting. (A) EcoRI and HindIII double digests probed with nick-translated pB^HHC3 DNA. Lane 1: control DNA; lanes 2-7 and 9 increasing DNaseI digestion; lane 8: HindIII and BamHI double digest of control DNA; lane 10: HindIII digested λ DNA molecular weight markers. (B) A PvuII digest probed with a labelled T7 RNA polymerase transcript of a pGEM1 subclone of pB^HHC3. Lane 1: HaeIII digested ϕ x174 DNA molecular weight markers; lane 2: control DNA; lanes 3-5: increasing DNaseI digestion.

site of the genomic pB2EH325 fragment (Fig. 1A). Also as a control, purified total chicken DNA was digested with increasing amounts of DNaseI, then with HindIII, blotted and probed. No 2.1 - 2.2 Kb sub-bands were observed (data not shown), showing that the sub-bands obtained with nuclei are due to chromatin structure. For a more accurate assignment of the DNaseI hypersensitive sites, PvuII digested DNA was hybridised as shown in Fig. 2B. With increasing DNaseI digestion of nuclei the 601 bp genomic PvuII fragment decreases (Fig. 1B) and two broad minor bands appear: one (site 11) 360-380 bp in size, and one (site 1) 440-470 bp in size. We estimate the accuracy of mapping the borders of the sensitive regions to within ± 10 bp, and the DNaseI hypersensitive regions mapped accordingly within the sequenced PvuII fragment shown in Fig. 1B.

The 3' boundary of site II and the 5'-boundary of site I are relatively sharp (Fig. 2B) and bracket a sequence (Fig. 1A) of approximately 50 bp in length which contains a TTGGA....TCCAA inverted repeat. This is homologous to the inverted repeat sequences bound by the chicken TGGCA-binding protein, which has been described in several chicken tissues (24,36) (see below, Fig. 7). This strongly suggests that the nuclear DNaseI-protected 50 bp are bound in the nuclei by a non-histone protein, possibly the chicken TGGCA-binding protein.

Tissue specificity of proteins binding to the β^A and β^H -globin genes in vitro.

The isolation of a chicken DNA-binding nuclear protein fraction (C4 fraction) enriched in sequence-specific DNA-binding proteins has been described (9-11). To investigate the tissue specificity of proteins binding to the 5'-flanking regions of the β^A - and β^H -globin genes, C4 proteins from chicken mature erythrocytes (MEC4), 15 day embryonic erythrocytes (EEC4), 3 week old chick thymus (TC4) and the transformed avian erythroid cell line HD3 (HDC4), were prepared and analysed for sequence specific binding by the DNA footprinting technique.

Chicken erythrocyte and chick embryo red blood cell nuclear extracts have previously been shown to contain several activities which bind to the 5'-flanking region of the β^A -globin gene. One activity binds to the poly(G)₁₈CGGGTGG sequence (footprint A, bases -177 to -201) and another activity to a sequence containing an inverted repeat (footprint B, bases -137 to -161). At higher protein/DNA ratios in the footprinting assay, footprint B enlarges in the 3' direction (to -115) suggestive of a third protein present in lower quantities binding next to it to give a third footprint (footprint

Fig.3: Footprint analysis of C4 protein bound to the chicken β^A -globin gene. The pBA650.1 β^A -globin gene subclone insert (10) was 5'-end-labelled at the 5'-PvuII site (the pUC9 HindIII site) incubated (~ 10 ng) with C4 protein, digested with DNaseI and the DNA resolved by denaturing polyacrylamide gel electrophoresis (10). (A) Lanes 1,2,3: 2.5, 5 and 10 μ g of 15 day embryonic erythrocyte C4 protein (EEC4); lanes 4,5,6: 2, 10 and 20 μ g of uninduced HD3 cell C4 protein (HDC4); lanes 7,8,9: 1, 2.5 and 5 μ g of mature erythrocyte C4 protein (MEC4); lane 10: no protein control. (B) Lanes 1 and 3: (-) 5 and 25 μ g of uninduced HD3 cell C4 protein. Lanes 2 and 4: (+) 5 and 25 μ g of induced HD3 cell C4 protein. Footprints A,B,B' were located by G+A sequencing reactions.

B') (10). This is reproduced in Fig. 3A, showing erythrocyte activities binding to sites A and B strongly and to a lesser extent with B'. Protein from HD3 cells (HDC4) on the other hand exhibit no footprint A activity, a strong site B binding activity, and no extension into the B' site (Fig. 3A,

lanes 4-6). In addition to the above footprints, erythrocyte proteins occasionally give protected regions 5' to site A, but these do not vary with protein concentration in the footprint assays and may therefore be due to differences in the degree of DNaseI digestion.

The HD3 cells are an erythroid precursor cell transformed by a temperature sensitive strain of the avian erythroblastosis virus (AEVts34) (21,26). When grown at 37°C, less than 1% of the HD3 cells are benzidine positive and β -globin gene expression is repressed as defined by cellular globin mRNA levels and the inactive β -globin gene chromosomal conformation (20). At the non-permissive temperature (42°C), the HD3 cells terminally differentiate and globin gene expression is elevated (20,21). In order to correlate the binding activities to sites A, B and B' and the activation of β^A -globin gene expression, an HD3 cell subline (see Methods) was incubated at 42°C for three days, and C4 protein prepared. After induction, approximately 30% of the cells were benzidine-positive, a low induction efficiency possibly due to the tendency of cells to revert to an uninducible phenotype after the large number of passages required to culture the 10^{10} cells required for a C4 protein preparation (21). C4 protein from induced and uninduced HD3 cells was bound to the β^A -globin gene fragment for footprint analysis and, as shown in Figure 3B, binding to site A occurs with induced but not uninduced C4 protein. This indicates that the binding to site B may be erythroid cell specific, whereas binding to site A occurs only after terminal differentiation and β^A -globin gene activation.

The chicken β^A - and β^H -globin genes are coordinately regulated during erythropoiesis (16) so the binding of C4 protein to the 5'-flanking DNaseI region of the β^H -globin gene was investigated. Footprint analysis using the pB^HHC3 insert DNA end-labelled at the 3' HincII site with erythrocyte (Fig. 4, lane 2) and uninduced HD3 (Fig. 4, lane 2) protein shows a protected region H-H' from approximately nucleotide 481 to 520 of Fig. 1B. No footprint was obtained with thymus protein (Fig. 4, lane 5). With protein from induced HD3 cells (lane 4) a smaller footprint H is seen; its 3' boundary is not definable since DNaseI does not cut within this region. We have found that the precise C4 protein:DNA ratio in the footprinting reaction mix determines whether footprint H or the larger H-H' footprint is observed, and we have detected both types of footprints with erythrocyte and HD3 (both induced and uninduced) C4 protein in other experiments. As demonstrated below, this is due to the presence of two proteins in the extracts binding to H and H'. As shown in Figure 1B, binding sites H/H' map to the 50 bp sequence protected in DNaseI

Fig.4: Footprint analysis of C4 protein bound to the chicken β^H -globin gene. The pB^HHC3 insert (Fig. 1A) was 5'-end-labelled at the 3' HincII site (the pUC9 HindIII site) incubated (10ng) with C4 protein and analysed by footprinting. Lane 1: No protein control; lane 2: 10 μ g of MEC4 protein; lane 3: 5 μ g of control uninduced HD3 cell C4 protein; lane 4: 10 μ g of 42°C induced HD3 cell C4 protein; lane 5: 5 μ g of thymus C4 protein.

nuclear digests, and suggests the same protein(s) may bind the sequence both in vitro and in the nucleus.

Certain similarities between the pattern of C4 protein binding to sites B and B' of the β^A -globin gene, and binding to sites H and H' of the β^H -globin gene, are evident: (1) Binding to sites B/B' and H/H' is specific for both HD3 and definitive erythrocyte C4 proteins (Figs. 3 and 4, and (11)), although binding of HD3 protein to B' is only observed after further purification (see below); (2) Binding to sites B' and H' occur at a different erythroid C4 protein:DNA ratio compared to that required for the footprint analysis of sites B and H (Figs. 3 and 4, and see below); (3) The binding sites B and B' are contiguous (10) as are sites H and H'; (4) sites B/B' and H/H' occur in the erythrocyte chromosomal DNaseI hypersensitive sites of co-

Fig.5: Footprint analysis of fractionated C4 protein bound to the chicken β^A -globin gene. Plasmid pB^A650.1 (10) insert end-labelled at the 5' PvuII site was incubated with protein for footprint analysis. (A) 5 μ g, 12 μ g and 25 μ g of MEC4, HDC4 and EEC4 protein respectively; lanes 1-9: 50 μ l each of the first nine 0.5ml EEC4 HPLC gel filtration fractions in order of decreasing molecular weight. The overlapping sequence of binding sites B and B' are shown. (B) Lanes 1-10: 50 μ l of each of the ten HD3C4 (uninduced) HPLC gel filtration fractions in order of decreasing molecular weights.

ordinately regulated β -globin genes (Fig. 1B and (9-11)); and (5) Sites B/B' and H/H' both contain variants of the sequence bound by Nuclear Factor 1 and the TGGCA-binding protein (Fig. 7). These observations suggest that erythroid cell-specific protein(s) are binding to related sequences in the 5'-flanking DNA sequence of coordinately regulated genes.

Separation of sequence-specific DNA-binding activities by HPLC gel filtration.

In order to further resolve the different sequence-specific DNA-binding activities, the C4 protein fractions were fractionated by Zorbax GF250 HPLC gel filtration.

The factors binding to the 5'-flanking region of the chicken β^A -globin gene (sites A, B and B') are resolved as shown in the footprint analysis of the embryonic erythrocyte C4 protein gel filtration fractions (Fig. 5A).

Fig. 6: Footprint analysis of fractionated C4 protein bound to the chicken β^H -globin gene. The pB^HHC3 insert end-labelled at the 3' HincII site was incubated with protein for footprint analysis. Lanes 1-9: 40 μ l of each of the first nine 0.6ml EEC4 HPLC gel filtration fractions; lane 10: 12 μ g of total EEC4 protein; lane 11: undigested DNA. Footprints H and H' are located from G + A sequencing reaction (see Fig. 1B).

Binding to site A elutes in fractions 2-4, followed by a weak binding to site B (fraction 4). A strong footprint labelled B' then elutes in fraction 5-7. Binding to site B by the fractionated HD3 protein elutes in fractions 4 and 5 (Fig. 5B, lane 4). Although binding to site B' is not detected in the total HD3 protein (Fig. 3), it is detectable after chromatography, eluting after the B-binding activity (fractions 6 and 7). This suggests that in both erythrocytes and HD3 cells there are two protein factors which bind to sites B and B' (Fig. 5). Examination of the sequences protected by B and B' activities separated by HPLC gel filtration reveal that footprint B is as described previously covering nucleotides (-137 to -161, shown on the side of Fig. 5), but footprint B' (nucleotides -115 to -150) is not simply an extension of footprint B but rather overlaps into the B footprint region as well as covering a region 3' to footprint B. The estimated molecular weight of the footprint A-binding activity was ~100-200000, footprint B activity ~100000, and footprint B' activity ~20-50000.

HPLC gel filtration fractions of erythroid C4 protein were compared for binding to the β^H -globin gene 5'-flanking sequences. As shown in Figure 6, when 15-day embryonic red blood cell C4 proteins are fractionated by gel

Fig.7: Comparison of the inverted repeat sequences bound by chicken and human nuclear proteins. (a) The chicken β^H -globin gene inverted repeat sequence within site H/H' sequence bound by avian erythroid C4 protein. (b) The adenovirus-5 sequence bound by Nuclear Factor 1 (38). (c) Consensus sequence bound by Nuclear Factor 1 (27). (d) Consensus sequence bound by the avian TGGCA-binding protein (36). (e) The 5'-flanking sequence (BS2) of the chicken lysozyme gene bound by the TGGCA-binding protein (24). (f) The β^A -globin gene site B sequence bound by avian erythroid C4 protein.

filtration, protein binding to site H elutes in fractions 3 and 4 (molecular weight ~100,000) followed by binding to site H' (fractions 6 and 7, molecular weight 20-50,000). Similar elution profiles were obtained for C4 proteins from mature erythrocytes and HD3 cells before and after induction (demonstrating the presence of both H and H' activities in all erythroid extracts) except that the H activity appeared to elute over a broader molecular weight range (20-100,000), suggesting that the H binding protein is multimeric and can dissociate during chromatography (data not shown).

Comparison of these results with those observed for the β^A -globin gene binding activities suggest that the B and H binding activities may be the same since they co-elute and bind to similar inverted repeat sequences (Fig. 7 and see Discussion below). The B' and H' activities also co-elute but since their binding sequences are different it suggests that the two genes could be utilising two proteins of similar size with different sequence specifications.

DISCUSSION

The nuclease hypersensitive domain in the 5'-flanking region of the chicken erythrocyte nuclear β^A -globin gene is a 200 bp sequence, 60 to 260 bp upstream from the mRNA CAP site (19). The domain contains the 'CAAT' and CACCC consensus sequences which have been shown to be important for mammalian β -globin gene expression (for review see refs. 34,35), the CACCC motif being found in many β -globin gene promoters (33-35). Sequence-specific nuclear DNA-binding proteins from definitive erythrocytes bind to three sites in this region of the β^A -globin gene in vitro (sites A, B and B' (9-11)), and two of the sites are bound by non-histone protein in the nucleus (15). As the β^A - and β^H -globin genes are coordinately regulated during chicken erythrocyte erythrocyte development, both at the levels of transcription (16) and of upstream chromosomal nuclease hypersensitive sites (18,19), we have analysed the structure of the nuclear β^H -globin gene in detail, and determined some of the properties of erythroid sequence-specific DNA-binding proteins which bind to the 5'-flanking DNaseI hypersensitive domain in vitro.

Fine mapping of the erythrocyte nuclear β^H -globin gene 5'-flanking DNaseI hypersensitive site (first described by Larsen and Weintraub (18)), revealed a 150 bp domain from 10-160 bp upstream from the two putative CAAT consensus sequences (Fig. 1B). However, the nuclease sensitive region is interrupted in the middle by a 40-50 bp nuclease insensitive sequence which, because it is too small to be bound by a nucleosome, is probably bound in situ by a non-histone protein factor. Confirmation for this comes from experiments in which

nuclei, extracted with 0.3M NaCl prior to DNaseI digestion, exhibit a nuclease hypersensitive region without the protected internal region, suggesting that proteins have been dissociated from this site by the salt extraction (manuscript in preparation). The same definitive erythrocyte nuclear DNA-binding protein fraction (C4 protein fraction (10)) which binds specifically to three sites in the 5'-flanking region of the β^A -globin gene in vitro (9-11) also binds to two contiguous sites in the 5'-flanking region of the β^H -globin gene in vitro (sites H and H', Fig. 1B). As sites H and H' map to the protected sequence within the nuclear DNaseI hypersensitive domain (Fig. 1B), this suggests that site H/H' is bound by the same proteins in vitro and in the nucleus, and furthermore implies some relationship between the factors which bind to the hypersensitive domains of the β^A - and β^H -globin genes.

It has been suggested that specific binding in the nucleus to the β^A -globin gene sites A and/or B-B' are responsible for inducing the 5'-flanking DNaseI hypersensitive site in terminally differentiated definitive erythrocytes (9,11,15). To test this, we have analysed the binding activities to the β^A - and β^H -globin genes in nuclear protein extracts from a temperature sensitive avian erythroblastosis virus (AEVts34) transformed erythroid precursor cell line (HD3 (21,26)) before and after the induction of terminal differentiation. Uninduced HD3 cells at the permissive temperature (37°C) express little globin (21,26), do not have the upstream globin gene DNaseI hypersensitive sites (20), but do have nuclear sequence specific DNA-binding activities to sites B/B' and H/H' (but not to site A) of the β^A - and β^H -globin genes. After HD3 cell induction at the non-permissive temperature (42°C), globin gene expression is activated (21,26) concomitantly with the appearance of the nuclear globin gene upstream DNaseI hypersensitive sites (20), and although the in vitro DNA-binding activities to sites B/B' and H/H' remain unchanged binding to the β^A -globin gene site A is now detectable. This indicates that although binding to site A may be involved in β^A -globin (but not β^H -globin) gene activation, binding activities to sites B/B' and H/H', which are erythroid specific, may be necessary but are not sufficient for β -globin gene activation.

Site A of the β^A -globin gene contains a deoxyguanosine homopolymer sequence ((G)₁₈CGGGTGGG (10,11)) which has an altered DNA secondary structure as evidenced by an S1 nuclease sensitivity in supercoiled plasmid DNA (32). In a previous report we showed that the erythrocyte C4 fraction also had an activity which bound to an S1 nuclease sensitive AT-rich sequence upstream of the human β -globin gene (10). However, in a filter binding analysis of the

erythrocyte C4 protein HPLC fractions, it was found that chicken β -globin binding activity eluted in fractions 2-4 whilst the human β -globin activity eluted much later showing that the activities are in fact not the same (unpublished results).

The three β^A -globin gene binding activities have been partially separated by HPLC gel filtration. Binding activities to sites A, B and B' elute with apparent molecular weights of 100-200000, ~100000 and 20-50000 respectively. The partial separation of sites B and B' binding activities reveals that although site B is as reported (137-161 bp upstream from the mRNA CAP site (10,11), site B' overlaps into site B, 115-150 bp upstream from the mRNA CAP site (see Fig. 4, and (10)). This places the conserved globin gene CACCC motif in the middle of binding site B', and the importance of this motif for globin gene expression (33-35) suggests it is bound by a transcription factor.

Emerson and Felsenfeld (personal communication) have also detected a chicken red blood cell activity which binds to the CACCC sequence of the chicken β^A -globin gene and which also binds to one of the CACCC motifs of the rabbit β -globin gene.

The two β^H -globin gene binding activities (to sites H and H') can be partially separated by HPLC gel filtration. The protein binding to site H' has an apparent molecular weight of 20-50000. The protein binding to site H elutes from the HPLC column in the 100,000 molecular weight size range. There are significant similarities between the binding sites and the binding activities to sites B/B' and H/H'. The sites occur 40-100 bp upstream of the CAAT consensus sequences within developmentally regulated nuclease hypersensitive domains, and in both cases, two protein activities bind overlapping or contiguous sequences. Binding activities to sites B and H, and to sites B' and H' co-elute after HPLC gel filtration, and are found in definitive erythrocyte and in HD3 cell nuclei before and after induction of terminal differentiation.

Although the β^H -globin gene does not have a bona fide CACCC motif, site H/H' has a TTGGA...TCCAA inverted repeat which resembles the adenovirus 5 sequence which binds the human nuclear factor 1 (NF1) protein of molecular weight 47,000 (22,23) (Fig. 7). Binding of NF1 has been implicated in the stimulation of DNA replication (22), has been found to bind a number of such inverted repeat sequences, and a consensus binding sequence deduced (Fig. 7 and (27)). A similar chicken nuclear protein, the TGGCA-binding protein has been described (24). These proteins have been shown to bind in vitro to inverted repeat sequences found in the nuclease hypersensitive sites in the 5'-flanking region of the chicken lysozyme gene (24), and the human c-myc gene

(8) and within the human IgM gene (37). Site B of the β^A -globin gene also contains an inverted repeat sequence, TCTGGGCA...TGCCCTGA (Fig. 4), and this is homologous to the sequence bound by the TGGCA-binding protein in the upstream region of the chicken lysozyme gene (see Fig. 7 and (24)). The sequence comparisons and the chromatography data can be used to argue that the binding activities to site B and site H of the β^A and β^H -globin genes are the same, and that they are related to the TGGCA-binding protein (or the NF1 protein). However, the tissue specificity of the binding activities are not in accordance with this proposal. For example, Sippel and co-workers (24) have found the TGGCA-binding protein in a number of non-erythroid chicken tissues (e.g. oviduct). On the other hand, we have found the level of the activity(ies) binding to sites B and H to be high in erythroid cells (definitive erythrocytes and HD3 cells), but much lower in thymus (see Fig. 3, and (10)). Similarly, Emerson and Felsenfeld (9,11) were unable to detect binding to the β^A -globin gene with oviduct or brain nuclear extracts, and were unable to find specific binding to the β^A -globin gene with a partially purified NF1 protein preparation. The tissue distribution of the globin gene binding activities would suggest that they are not obviously related to the TGGCA-binding protein. This, however, remains to be formally demonstrated.

The function of inverted repeat sequences in the β -globin gene nuclease 5'-flanking nuclease hypersensitive domains is not yet known. As proteins that bind to these sequences are present in erythroid cells (HD3) which do not express globin and do not have the globin gene nuclear hypersensitive sites, it is unlikely that the proteins are sufficient for globin gene activation or for the induction of nuclease hypersensitivity; however, they could stabilise nuclease hypersensitive sites induced by other gene-specific factor(s).

In summary the results show that the β^H - and β^A -globin genes may be regulated by a similar mechanism involving the binding of common factor to the inverted repeats and this may in turn affect positively or negatively the binding of a second factor to a contiguous site; the second factor for the β^H -globin gene (the H'-binding activity) is probably different from the β^A -globin gene factor (the CACCC-binding factor B').

Nuclear extracts of avian erythroid cells also bind to multiple elements within the hypersensitive sites of the avian c-myc gene [39]. One of these footprints has the sequence CTC(G)₉ and the protein that binds to it coelutes with β^A -globin A-binding activity, suggesting that the two activities may be the same. Thus, the A-binding activity may not be a β^A -globin specific regulatory protein.

ACKNOWLEDGEMENTS

We thank D. Engel for chicken β -globin recombinants and H. Beug for helpful discussions on HD3 cell induction, and Dr. H. Paterson for advice on cell cloning. These studies were supported by grants from the Medical Research Council and the Cancer Research Campaign. V.V.L. acknowledges fellowships from the International Agency for Research on Cancer and the Institute of Cancer Research.

*Present address: The Beatson Institute for Cancer Research, Switchback Road, Bearsden, Glasgow G61 1BD, UK

*Permanent address: All-Union Cancer Research Centre, 24 Kashirskoye Shosse, 115478 Moscow, USSR

REFERENCES

1. Weisbrod, S. (1982) *Nature* 297, 289-295.
2. Bird, A.P. (1984) *Nature* 307, 503-504.
3. Elgin, S.C.R. (1984) *Nature* 309, 213-214.
4. Parker, C.S. and Topol, J. (1984) *Cell* 37, 273-283.
5. Wu, C. (1984) *Nature* 309, 229-234.
6. Dynan, W.S. and Tjian, R. (1983) *Cell* 35, 79-87.
7. Wildeman, A.G., Sassone-Corsi, P., Grundstrom, T., Zenke, M. and Chambon, P. (1984) *EMBO J.* 3, 3129-3133.
8. Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1984) *Cell* 37, 381-391.
9. Emerson, B.M. and Felsenfeld, G. (1984) *Proc. Natl. Acad. Sci. USA* 81, 95-99.
10. Plumb, M.A., Nicolas, R.H., Wright, C.A. and Goodwin, G.H. (1985) *Nucl. Acids Res.* 13, 4047-4065.
11. Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985) *Cell* 41, 21-30.
12. Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983) *Science* 222, 740-748.
13. Dunaway, M. and Reeder, R.H. (1985) *Mol. Cell Biol.* 5, 313-319.
14. Shastry, B.S., Honda, B.M. and Roeder, R.G. (1984) *J. Biol. Chem.* 259, 11373-11382.
15. Jackson, P.D. and Felsenfeld, G. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2296-2300.
16. Landes, G.M., Villeponteau, B., Pribyl, T.W. and Martinson, H.G. (1982) *J. Biol. Chem.* 257, 11008-11014.
17. Kenne, R.W. and Abbot, U.K. (1980) *Dev. Biol.* 75, 442-453.
18. Larsen, A. and Weintraub, H. (1982) *Cell* 29, 609-622.
19. McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. and Felsenfeld, G. (1981) *Cell* 27, 45-55.
20. Weintraub, H., Beug, H., Groudine, M. and Graf, T. (1982) *Cell* 28, 931-940.
21. Beug, H., Palmieri, J., Freudenstein, C., Zentgraf, H. and Graf, T. (1982) *Cell* 28, 907-919.
22. Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichey, J.H. and Hurwitz, J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6438-6442.
23. Rawlins, D.R., Rosenfeld, P.J., Wides, R.S., Challberg, M.D. and Kelly, T.J. (1984) *Cell* 37, 309-319.
24. Borgmeyer, U., Nowock, J. and Sippel, A.E. (1984) *Nucl. Acid Res.* 12, 4295-4311.
25. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
26. Beug, H., Doederlein, G., Freudenstein, C. and Graf, T. (1982) *J. Cell Physiol. Supplement* 1, 195-207.
27. Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1985) *Mol. Cell Biol.* 5, 964-971.

-
28. Vieira, J. and Messing, J. (1982) *Gene* 10, 259-268.
 29. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.* 12, 7035-7056.
 30. Dolan, M., Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1981) *Cell* 24, 669-677.
 31. Dolan, M., Dodgson, J.B. and Engel, J.D. (1983) *J. Biol. Chem.* 258, 3983-3990.
 32. Nickol, J.M. and Felsenfeld, G. (1983) *Cell* 35, 467-477.
 33. Dierks, P., von Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissman, S.M. (1983) *Cell* 32, 695-706.
 34. Harrison, P.R. (1984) *Exp. Cell Res.* 155, 321-344.
 35. Collins, F.J. and Weissman, S.M. (1984) *Prog. Nucl. Acid Res. and Mol. Biol.* 31, 315-462.
 37. Henninghausen, L., Siebenlist, U., Danne, D., Leder, P., Rawlins, D., Rosenfeld, P. and Kelly, T. (1985) *Nature* 314, 289-292.
 38. Nagata, K., Guggenheimer, R.A. and Harwitz, J. (1983) *Proc. Natl. Acad. Sci. USA* 80, 6177-6181.
 39. Lobanekov, V.V., Nicolas, R.H., Plumb, M.A., Wright, C.A. and Goodwin, G.H. (1986) *Eur. J. Biochem.* 159, 181-188.